SPECIALIST PERIODICAL REPORTS

Amino-acids Peptides and Proteins

VOLUME 4

THE CHEMICAL SOCIETY

A Specialist Periodical Report

Amino-acids, Peptides, and Proteins

Volume 4

A Review of the Literature Published during 1971

Senior Reporter

G. T. Young, *The Dyson Perrins Laboratory, University of Oxford*

ISBN: 0 85186 034 6

Library of Congress Catalog Card No. 72-92548

The Chemical Society
Burlington House, London, W1V 0BN

Reporters

P. M. Bayley, *National Institute of Medical Research, Mill Hill, London*
T. L. Blundell, *University of Oxford*
J. R. Brocklehurst, *University of Pennsylvania, U.S.A.*
B. W. Bycroft, *University of Nottingham*
J. S. Davies, *University College of Swansea, University of Wales*
C. S. Johnson, *University of Liverpool*
J. H. Jones, *University of Oxford*
G. L. Kellett, *University of York*
R. H. Pain, *University of Newcastle upon Tyne*
R. N. Perham, *Cambridge University*
N. C. Price, *University of Pennsylvania, U.S.A.*
H. W. E. Rattle, *Portsmouth Polytechnic*
B. Ridge, *University of Exeter*
B. Robson, *University of Newcastle upon Tyne*
R. M. Stephens, *Portsmouth Polytechnic*
J. O. Thomas, *Cambridge University*

Organic formulae composed by Wright's Symbolset method

PRINTED IN GREAT BRITAIN BY JOHN WRIGHT AND SONS LTD., AT THE STONEBRIDGE PRESS, BRISTOL

Preface

This fourth Report reviews papers relevant to the chemistry of amino-acids, peptides, and proteins, appearing in the main journals during 1971; the literature continues to increase, and this volume contains nearly 3,000 references. Metal derivatives are reviewed biennially in this series and are not covered this year; volume 5 will survey papers in this field appearing in 1971 and 1972. As in Volume 3, work on the structure and synthesis of cyclic peptides will be found in Chapter 4 (Peptides with Structural Features Not Typical of Proteins). In Chapter 5 we reprint for the convenience of readers the 1971 revision of the recommendations of the I.U.P.A.C.–I.U.B. Commission on Biochemical Nomenclature, 'Symbols for Amino-acid Derivatives and Peptides', together with their recommendations 'Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains'.

For the new reader, we would note that in place of a subject index (the preparation of which would delay publication unduly) there is an extended list of contents from which the sections relevant to a search can be ascertained. Some overlap between sections will be found, and within limits is no doubt desirable.

Finally, I express my gratitude here to the contributors who in these first four volumes have established this series as a service to their colleagues in research.

G. T. YOUNG

Contents

Chapter 2 Structural Investigation of Peptides and Proteins

I: Primary Structure and Chemical Modification

By R. N. Perham and J. O. Thomas

III: Conformation and Interaction of Peptides and Proteins in Solution
Edited by R. H. Pain

Abbreviations

Abbreviations for amino-acids and their use in the formulation of derivatives follow with some exceptions the revised (1971) Recommendations of the I.U.P.A.C.–I.U.B. Commission on Biochemical Nomenclature, which are reprinted in Chapter 5 of this volume.

Other abbreviations which have been used without definition are:

Adoc	adamantyloxycarbonyl
Aoc	t-amyloxycarbonyl
Asu	α-aminosuberic acid
Asx	aspartic acid or asparagine (not yet determined)
ATP	adenosine 5′-triphosphate
Bpoc	2-(4-biphenylyl)-isopropoxycarbonyl
BSA	bovine serum albumin
c.d.	circular dichroism
Cha	cyclohexylamine
Cm	carboxymethyl
Cmc	*S*-carboxymethylcysteine
Dce	2,2-diethoxycarbonyl
Dcha	dicyclohexylamine
DMF	*NN*-dimethylformamide
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
Dnp	2,4-dinitrophenyl
Dns	1-dimethylaminonaphthalene-5-sulphonyl (dansyl)
Dopa	3,4-dihydroxyphenylalanine
DP	degree of polymerization
Ec	ethylcarbamoyl
edta	ethylenediamine tetra-acetate
e.p.r.	electron paramagnetic resonance
e.s.r.	electron spin resonance
Gal	galactose
g.l.c.	gas–liquid chromatography
Glc	glucose
Glp or Pca	pyrrolid-2-one-5-carboxylic acid
Glx	glutamic acid or glutamine (not yet determined)
GTP	guanosine 5′-triphosphate
i.r.	infrared

Man	mannose
NAD	nicotinamide-adenine dinucleotide (NAD^+, oxidized; NADH, reduced)
n.m.r.	nuclear magnetic resonance
ONSu	succinimido-oxy
OPcp	pentachlorophenoxy
OPic	4-picolyloxy
o.r.d.	optical rotatory dispersion
OTcp	2,4,5-trichlorophenoxy
Pca	*see* Glp
Pipoc	piperidino-oxycarbonyl
Pth-Gly	the phenylthiohydantoin derived from glycine, *etc.*
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
Ser(P)	*O*-phosphorylserine
t.l.c.	thin-layer chromatography
u.v.	ultraviolet
Ztf	1-benzyloxycarbonylamino-2,2,2-trifluoro-ethyl

1
Amino-acids

BY B. W. BYCROFT

The past twelve months have witnessed interesting developments in amino-acid chemistry, but pride of place must be taken by the isolation, characterization, and synthesis of the amino-acid derivative from phenylalanine t-RNA. This work also reflects the growing sophistication in the application of physical methods in general. The interesting advances reported last year on asymmetric synthesis have been extended, and a staggering number of new amino-acids continues to be synthesized for a variety of reasons. The established pattern of coverage for this chapter is maintained, attention being focused on a broad selection of topics with the significant developments highlighted where necessary.

1 Naturally Occurring Amino-acids

A. Occurrence of Known Amino-acids.—A large number of papers which are concerned with the free amino-acid content of a wide variety of living organisms is published annually. Since the emphasis for the majority of this work is on biological aspects, it is not considered appropriate to cover them in this section and only those amino-acids which are rarely encountered or are interesting from the chemical viewpoint have been included.

The presence of amino-acids in extra-terrestrial material as well as their synthesis under simulated prebiotic conditions (see Section 2B) has attracted a considerable amount of attention. Examination of samples from the Apollo 11 and Apollo 12 missions has revealed extremely low concentrations of amino-acids (*ca.* 20—70 p.p.b.);[1-3] ultra-sensitive analytical techniques were employed for these investigations and, because of the presence of a considerable number of non-protein amino-acids, the investigators maintain that they are not due to terrestial contamination. Similar conclusions have been made concerning the presence of amino-acids in the Murray [4] and

[1] B. Nagy, J. E. Modzeleski, V. E. Modzeleski, M. A. J. Mohammad, L. A. Nagy, W. M. Scott, C. M. Drew, J. E. Thomas, R. Ward, P. B. Hamilton, and H. C. Urey, *Nature*, 1971, **232**, 94.
[2] K. Harada, P. E. Hara, C. R. Windsor, and S. W. Fox, *Science*, 1971, **173**, 433.
[3] C. W. Gehrke, R. W. Zumwalt, D. L. Stalling, D. Roach, W. A. Aue, C. Ponnamperuma, and K. A. Kvenvolden, *J. Chromatog.*, 1971, **59**, 305.
[4] J. G. Lawless, K. A. Kvenvolden, E. Peterson, C. Ponnamperuma, and C. Moore, *Science*, 1971, **173**, 626.

Murchinson [5] meteorites. In the case of the Murray meteorite, seventeen amino-acids were identified of which seven were conclusively shown to be racemic and eleven to be non-protein in origin. An earlier report [6] that the amino-acid content of the Orgueil meteorite was due solely to terrestrial contamination has been questioned, and it is now suggested that there are amino-acids indigenous to the meteorite in addition to those present as contaminants.[7]

The presence of amino-acids in the North Atlantic ocean has been the subject of a detailed examination and the distribution appears to be non-uniform and varies qualitatively with depth.[8] In marine sediments the degree of racemization of free amino-acids shows a progressive increase with the age of the sediment,[9, 10] and the ratio of *allo*-isoleucine to leucine is a reliable indicator of age for samples less than 400 000 years old. In samples older than about 15×10^6 years the amino-acids are completely racemic.

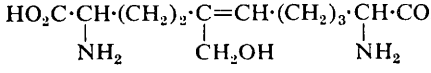
The stereochemistry of the α-hydroxy-analogue of cysteine present in the urine of certain mentally retarded patients has been established by comparison with synthetic material.[11] The work reported initially last year on the isolation of the methylated derivatives of arginine and lysine has been extended and it has been noted that, in patients with malignant tumours, the relative urinary level of *guanidino-NN*-dimethylarginine to that of arginine is markedly increased.[12] Bovine and rat brain tissue has been shown to contain appreciable amounts of N^G-monomethylarginine as well as $N^G N'^G$-dimethylarginine.[13]

Further spectral evidence on the aldol condensation product isolated on alkaline hydrolysis of elastin provides strong support [14] for the previously assigned structure (1).

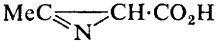
Interest in plants containing L-Dopa continues, and a widespread investigation has shown that several *Mucuna* species contain up to 5% of this important amino-acid.[15] The neurotoxin α-amino-β-oxalylaminopropionic acid has been isolated from *Lathyrus sativus* [16] and the major alkaloid of *Aotus subglauca* has been identified as *S*-(+)-N^α-methyltryptophan methyl ester.[17] The previously unidentified amino-acid from *Peganum harmala* is

[5] J. R. Cronin and C. B. Moore, *Science*, 1971, **172**, 1327.
[6] J. Oro, S. Nakaparksin, H. Lichenstein, and E. Gil-Av, *Nature*, 1971, **230**, 107.
[7] J. G. Lawless, K. A. Kvenvolden, E. Peterson, and C. Ponnamperuma, *Nature*, 1972, **234**, 66.
[8] R. Pocklington, *Nature*, 1971, **230**, 374.
[9] R. O. Brinkhurst, K. E. Chua, and E. Batoosingh, *Limnol. Oceanogr.*, 1971, **16**, 555.
[10] J. Wehmiller and P. E. Hare, *Science*, 1971, **173**, 907.
[11] M. Wälti and D. B. Hope, *J. Chem. Soc.* (*C*), 1971, 2326.
[12] S. Akazawa, *Osaka Daigaku Igaku Zasshi*, 1970, **22**, 461.
[13] T. Nakajima, Y. Matsuoka, and Y. Kakimoto, *Biochim. Biophys. Acta*, 1971, **230**, 212.
[14] G. Crombie, B. Faris, P. M. Gallop, and C. Franzblau, *Biochemistry*, 1971, **10**, 4145.
[15] E. A. Bell, J. R. Nulu, and C. Cone, *Phytochemistry*, 1971, **10**, 2191.
[16] K. Bahadar and S. P. Billa, *Indian J. Appl. Chem.*, 1970, **33**, 168.
[17] S. R. Johns, J. A. Lamberton, and A. A. Sioumis, *Austral. J. Chem.*, 1971, **24**, 439.

established as L-4-hydroxypipecolic acid.[18] D-Pipecolic acid, as well as L-*threo*- and D-*erythro*-$\alpha\beta$-diaminobutyric acid, has been isolated on acid hydrolysis of the antibiotic amphomycin,[19] and griselimycin afforded, *inter alia*, 4-*trans*-4-methyl-L-proline and *N*-methyl-D-leucine on hydrolysis.[20]

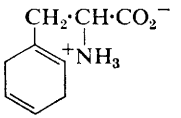


(1)

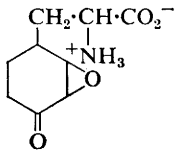


(2)

B. New Natural Free Amino-acids.—The novel azirine amino-acid (2) isolated from a strain of *Streptomyces aureus*, and appropriately named azirinomycin, exhibits broad-spectrum activity *in vitro* against both Gram positive and negative bacteria.[21] It is unstable especially in concentrated form and was identified by spectral analysis and by conversion on catalytic hydrogenation into L-α-aminobutyric acid.[22] L-2,5-Dihydrophenylalanine (3), which had previously been synthesized by a Birch reduction of L-phenylalanine,[23] is an antimetabolite of L-phenylalanine produced by an unidentified *streptomyces*.[24] A related compound, anticapsin (4), is produced by a strain of *Streptomyces griseoplanus* [25] and is presumably derived from L-tyrosine by a similar reduction with subsequent epoxidation.



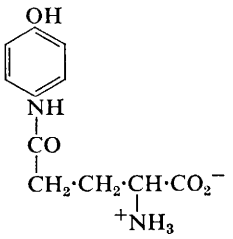
(3)



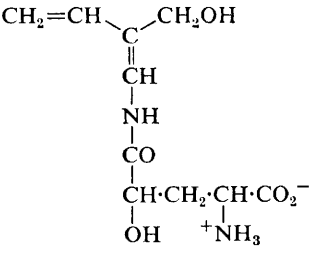
(4)

The increasing application of mass spectrometry as a valuable tool for characterization is apparent and has been employed for many of the amino-acids described in this section. Trimethylsilylation is still the most commonly employed procedure for enhancing the volatility of amino-acids containing a number of polar substituents and has been successfully applied in characterizing L-*threo*-α-amino-$\beta\gamma$-dihydroxybutyric acid.[26] The new compound (5), which has been isolated from the common mushroom, affords on oxidation an unstable quinone, also present in the organism.[27]

[18] V. U. Ahmad and M. A. Khan, *Phytochemistry*, 1971, **10**, 3339.
[19] M. Bodanszky, N. C. Chaturvedi, J. A. Scozzie, R. K. Griffith, and A. Bodanszky, *Antimicrobial Agents and Chemotherapy*, 1969, 135.
[20] B. Terlain and J. P. Thomas, *Bull. Soc. chim. France*, 1971, 2349.
[21] E. O. Stapley, D. Hendlin, M. Jackson, and A. K. Miller, *J. Antibiotics*, 1971, **24**, 42.
[22] T. W. Miller, E. W. Tristram, and F. J. Wolf, *J. Antibiotics*, 1971, **24**, 48.
[23] M. L. Snow, C. Lawinger, and C. Ressler, *J. Org. Chem.*, 1968, **33**, 1774.
[24] J. P. Scannell, D. L. Pruess, T. C. Demny, T. Williams, and A. Stempel, *J. Antibiotics*, 1970, **23**, 618.
[25] R. Shah, N. Neuss, M. Gorman, and L. D. Boeck, *J. Antibiotics*, 1970, **23**, 613.
[26] J. W. Westley, D. L. Pruess, L. A. Volpe, T. C. Demny, and A. Stempel, *J. Antibiotics*, 1971, **24**, 330.
[27] R. F. Weaver, K. V. Rajagopalan, P. Handler, D. Rosenthal, and P. W. Jeffs, *J. Biol. Chem.*, 1971, **246**, 2010.

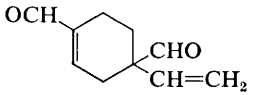


(5)

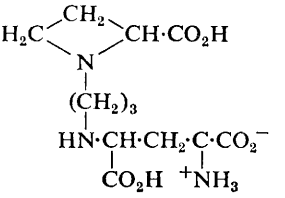


(6)

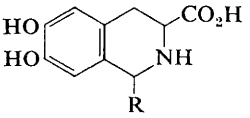
Glutamic acid derivatives have also been isolated from a number of plant sources. Pinnatanine (6), a new amino-acid isolated from the European bladder nut [28] (*Staphylea pinnata*), affords on acid hydrolysis L-*allo*-γ-hydroxyglutamic acid, ammonia, and compound (7), the structure of which has been confirmed by synthesis. It is suggested that (7) arises *via* a Diels–Alder dimerization of the transient 2-methylenebut-3-enal initially formed in the hydrolysis. In the course of an investigation on nitrogen metabolism in tobacco plants a new amino-acid, nicotianamine, was isolated and the structure (8) was assigned on the basis of extensive chemical degradation and spectral analysis.[29] Azetidine-2-carboxylic acid had previously been obtained from a considerable number of plant species.[30]



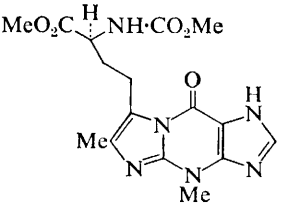
(7)



(8)



(9) R = H
(10) R = Me



(11)

The isoquinoline derivative (9) has been isolated from the plant *Mucuna mutisiana* [15] which also contains appreciable amounts of L-Dopa, and it

[28] M. D. Grove, M. E. Daxenbichler, D. Weisleder, and C. H. VanEtten, *Tetrahedron Letters*, 1971, 4477.
[29] M. Noma, M. Noguchi, and E. Tamaki, *Tetrahedron Letters*, 1971, 2017.
[30] L. Fowden, *Adv. Enzymol.*, 1967, **29**, 89.

appears likely that (9) is derived from L-Dopa by condensation with formaldehyde. The corresponding derivative (10), formed from L-Dopa and acetaldehyde, had previously been isolated.[31]

The urine of homocystinuric patients has been shown to contain *S*-(3-hydroxy-3-carboxy-n-propylthio)- and *S*-(2-hydroxy-2-carboxyethylthio)-homocystine[32] in addition to the α-hydroxy-analogue of cystine. Further work on the phaemelanin pigments has been reported, and an isomer of trichosiderin C has been isolated and characterized.[33]

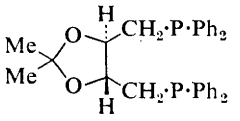
C. New Amino-acids from Hydrolysates.—The most notable achievement in this field is undoubtedly the characterization and subsequent synthesis (see Section 2C) of the fluorescent Y base present in phenylalanine t-RNA derived from yeast, wheat germ, and rat liver. It represents a significant triumph for the application of physical methods; the structure (11) was established on 300 μg of material mainly on the interpretation of the results obtained from high-resolution mass spectrometry and n.m.r. spectroscopy.[34]

Full details on the structures of the novel piperazic acid derivatives from the antibiotic monamycin[35] and the guanidine amino-acid viomycidine from viomycin[36] have been reported. β-Hydroxyhistidine has been isolated on the acid hydrolysis of the antibiotic bleomycin.[37]

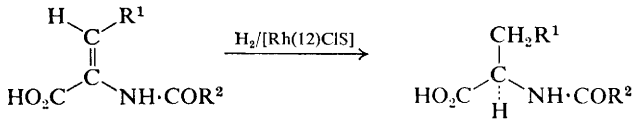
2 Chemical Synthesis and Resolution of Amino-acids

A. Introduction and General Methods.—The interest in asymmetric syntheses of amino-acids continues unabated and a valuable article covering the literature up to 1969 has been published.[38] In addition those syntheses which involve enantioselective catalytic hydrogenation have also been reviewed.[39] In this respect it is of considerable interest to note that this year has seen the first significant application of homogeneous asymmetric catalysis to amino-acid synthesis.[40] The rhodium complex of the chiral diphosphine (12) readily reduces *N*-acyldehydroamino-acids (13) to the corresponding (*R*)-*N*-acylamino-acids in high yield and with an optical efficiency of about 70%. The relatively high stereoselectivity is ascribed to the conformational rigidity of the diphosphine chelating the rhodium, together with the participation of the acid function of the substrate.

[31] P. Müller and H. R. Schütte, *Z. Naturforsch.*, 1968, **236**, 491.
[32] H. Kodama, S. Ohmori, M. Suzuki, S. Mizuhara, T. Oura, G. Isshiki, and I. Uemura, *Physiol. Chem. Phys.*, 1971, **3**, 81.
[33] G. Prota, A. Suarato, and R. A. Nicolaus, *Experientia*, 1971, **27**, 1145.
[34] K. Nakanishi, N. Furutachi, M. Funamizu, D. Grunberger, and I. B. Weinstein, *J. Amer. Chem. Soc.*, 1970, **92**, 7617.
[35] C. H. Hassall, Y. Ogihara, and W. A. Thomas, *J. Chem. Soc.* (*C*), 1971, 522.
[36] G. Büchi and J. A. Raleigh, *J. Org. Chem.*, 1971, **36**, 871.
[37] T. Takita, T. Yoshioka, Y. Muraoka, K. Maeda, and H. Umezawa, *J. Antibiotics*, 1971, **24**, 795.
[38] J. D. Morrison and H. S. Mosher, 'Asymmetric Organic Reactions', Prentice-Hall, New Jersey, 1971.
[39] Y. Izumi, *Angew. Chem., Internat. Edn.*, 1971, **12**, 871.
[40] T. P. Dang and H. B. Kagan, *Chem. Comm.*, 1971, 481.

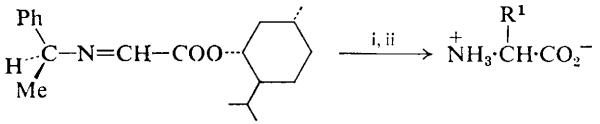


(12)



(13)

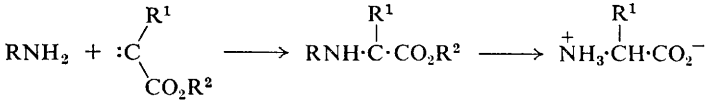
The inherent problem of conformational mobility of the substrate in all asymmetric syntheses which was referred to in last year's Report is probably the cause of the relatively low optical efficiency of the new synthesis [41] outlined in Scheme 1 and of the asymmetric reduction of dehydroamino-acid



Reagents: i, R^1MgX; ii, H^+

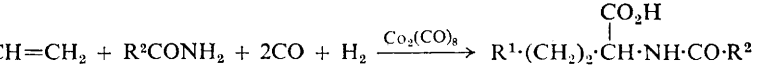
Scheme 1

peptides.[42] Unfortunately, the ingenious synthesis of amino-acids by the insertion of a carbene into an N—H bond of an optically active amine [43] (Scheme 2) also gives low optical efficiency, presumably for the same reason.



Scheme 2

A general synthesis has been reported [44] which employs a carbonylation reaction and is essentially a variation of the hydroformylation reaction and the Oxo process (Scheme 3). Although the products are completely racemic



Scheme 3

[41] J. Fiaud and H. B. Kagan, *Tetrahedron Letters*, 1971, 1019.
[42] M. Nakayama, G. Maeda, T. Kaneko, and H. Katsura, *Bull. Chem. Soc. Japan*, 1971, **44**, 1150.
[43] J. F. Nicoud and H. B. Kagan, *Tetrahedron Letters*, 1971, 2065.
[44] H. Wakamatsu, J. Uda, and N. Yamakami, *Chem. Comm.*, 1971, 1540.

the yields are reasonable and it is claimed that the reaction has a wide application to amino-acid synthesis. The starting materials are readily available and the synthesis reflects the growing commercial interest in the large-scale chemical production of amino-acids.

B. Synthesis under Simulated Prebiotic Conditions.—It is now well established that amino-acids are formed when a simulated primitive atmosphere is subjected to high temperature, u.v. irradiation, or high-frequency discharge, or any combination of these conditions. Recent work has centred on more rigorous analysis of the reaction mixtures and identification of amino-acids with functional groups in the side-chain.[45, 46] Irradiation of mixtures of methane, ammonia, hydrogen sulphide, and water produces mixtures which contain either cysteine or cystine depending on the reaction conditions[47] and similar experiments using high-frequency discharge produce mixtures in which histidine has been conclusively identified.[48, 49] The preferential adsorption of the L-isomer of racemic amino-acids on Kaolinite templates has been demonstrated [50] and is of particular interest in connection with the natural predominance of L-amino-acids.

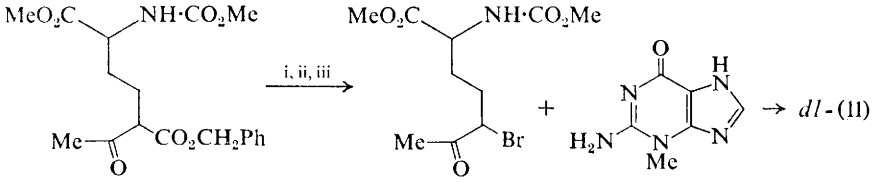
C. Protein and Other Naturally Occurring Amino-acids.—Many of the syntheses described in this section have been achieved by standard procedures, and therefore only the salient features of the more important will be presented.

The various methods for the synthesis of glycine have been reviewed [51] and a large-scale preparation of ornithine from glutamic acid has been reported.[52] A new synthesis of threonine from the glycine copper complex and acetaldehyde has been described [53] together with a detailed investigation of the course of this reaction. Interest in the synthesis of L-Dopa and related compounds (Section 2F) continues, and two further syntheses [54, 55] are now available.

The full details of the syntheses of capreomycidine,[56] indospicine,[57] and the piperazic acid derivatives [58] from the antibiotic monamycin, initially reported in preliminary form, have now been published.

[45] D. Yoshino, R. Hayatsu, and E. Anders, *Geochim. Cosmochim. Acta*, 1971, **35**, 927, 939.
[46] V. Marshall and A. Bennett, *Proc. Indian Acad. Sci.*, 1970, **80**, 369.
[47] B. N. Khare and C. Sagan, *Nature*, 1971, **232**, 577.
[48] S. Yuasa, M. Ishigami, Y. Honda, and K. Imahori, *Sci. Rep. Osaka Univ.*, 1970, **19**, 33.
[49] S. Yuasa, M. Yamamoto, Y. Honda, M. Ishigami, and K. Imahori, *Sci. Rep. Osaka Univ.*, 1970, **19**, 7.
[50] T. A. Jackson, *Experientia*, 1971, **27**, 242.
[51] B. P. Thacker, *Indian J. Chem.*, 1971, **5**, 42.
[52] V. Gut and K. Poduška, *Coll. Czech. Chem. Comm.*, 1971, **36**, 3470.
[53] P. Maldonado, C. Richaud, J. P. Aune, and J. Metzger, *Bull. Soc. chim. France*, 1971, 2933.
[54] K. Ogura and G. Tsuchihashi, *Tetrahedron Letters*, 1971, 3151.
[55] H. Nakamoto, M. Aburatuni, and M. Inagaki, *J. Medicin. Chem.*, 1971, **14**, 1021.
[56] B. W. Bycroft, D. Cameron, and A. W. Johnson, *J. Chem. Soc.* (*C*), 1971, 3040.
[57] C. C. J. Culvenor, M. C. Foster, and M. P. Hegarty, *Austral. J. Chem.*, 1971, **24**, 371.
[58] K. Bevan, J. S. Davies, C. H. Hassall, R. B. Morton, and D. A. S. Phillips, *J. Chem. Soc.* (*C*), 1971, 514.

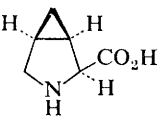
The structure of the Y base from phenylalanine t-RNA, as well as the stereochemistry at the single asymmetric centre, has been conclusively established [59] by the synthesis outlined in Scheme 4. The novel cyclopropyl



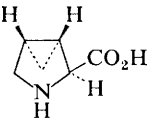
Reagents: i, H_2–Pd/C; ii, Br_2–$CHCl_3$; iii, H_2O

Scheme 4

amino-acid (14) from horse chestnuts has been synthesized by carbene addition to 3,4-dehydroproline.[60] The reaction gives rise to a mixture of the *cis*- and *trans*-isomers (14) and (15) in the ratio 1 : 3.5. The stereochemistry of the natural amino-acid (14) was established by *X*-ray crystallographic analysis.

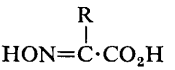


(14)



(15)

D. *C*-Alkyl- and Substituted *C*-Alkyl-α-amino-acids.—An improved method is claimed for the preparation of unsaturated α-amino-acids by reduction with aluminium amalgam of the unsaturated α-hydroxyimino intermediate (16) derived by the normal malonate route.[61] A convenient method for



(16) R = alkenyl

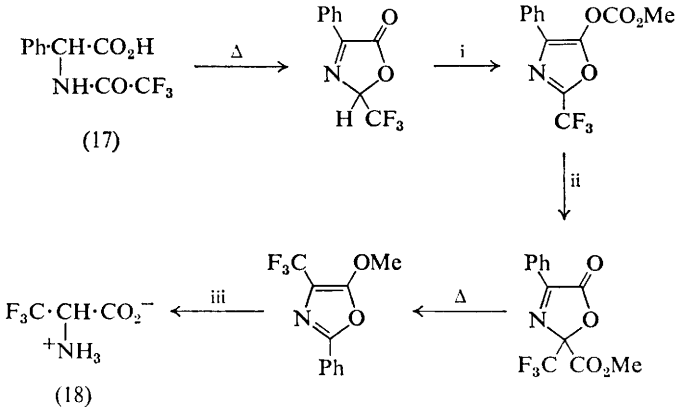
direct conversion of *N*-acyl-α-amino-acids into the *N*-acyl-αβ-dehydro-amino-acids has been reported.[62] (*S*)-α-Methyl-α-amino-acids have been obtained in high yield by a modified Strecker synthesis [63] and new syntheses of α-aminosuberic and α-aminosebacic acids have been described.[64]

Considerable interest exists in cyanoamino-acids because of the strong neurotoxic properties of the naturally occurring β-cyano-L-alanine.

[59] M. Funamizu, A. Terahara, A. M. Feinberg, and K. Nakanishi, *J. Amer. Chem. Soc.*, 1971, **93**, 6708.
[60] Y. Fujimoto, F. Irreverre, J. M. Karle, I. L. Karle, and B. Witkop, *J. Amer. Chem. Soc.*, 1971, **93**, 3471.
[61] D. J. Drinkwater and P. W. G. Smith, *J. Chem. Soc.* (*C*), 1971, 1305.
[62] J. M. Riordan and C. H. Stammer, *Tetrahedron Letters*, 1971, 4969.
[63] K. Weinges, G. Graab, D. Nagel, and B. Stemmle, *Chem. Ber.*, 1971, **104**, 3594.
[64] G. M. Shakhnazaryan, L. A. Voskanyan, and M. T. Dangyan, *Armyan. khim. Zhur.*, 1970, **23**, 709.

Dehydration of *N-o*-nitrophenylsulphenyl (Nps) asparagine and glutamine with dicyclohexylcarbodi-imide affords the corresponding protected ω-cyanoamino-acid.[65] The free amino-acids were obtained by removing the Nps group under mild conditions which did not effect the cyano-group. α-Cyanoglycine has been prepared by enzymic deacetylation of the readily available acetamidocyanoacetic acid.[65] Attempts to prepare it by chemical hydrolysis were unsuccessful because it undergoes rapid decarboxylation in hot aqueous solution.

N-Trifluoroacetylphenylglycine (17) has been converted by an interesting and unusual series of thermal rearrangements into trifluoroalanine (18).[66] The proposed reaction sequence is outlined in Scheme 5. The overall yield is acceptable and it is probable that this novel synthesis could be extended to other fluorinated amino-acids.



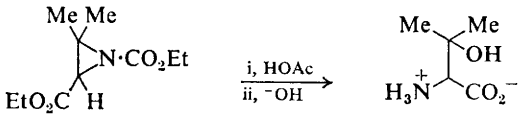
Reagents: i, $ClCO_2Me$–NEt_3; ii, 4-dimethylaminopyridine: iii, HBr–HOAc

Scheme 5

E. α-Amino-acids with Aliphatic Hydroxy-groups in the Side-chain.—The considerable interest in β-hydroxyvaline in relation to penicillin chemistry is reflected in the large number of syntheses already available for this compound. A further synthesis involves the addition of ethoxycarbonylnitrene to ethyl ββ-dimethylacrylate to give the aziridine intermediate (19). Ring opening of (19) with acetic acid and subsequent base hydrolysis affords (20) in good yield.[67]

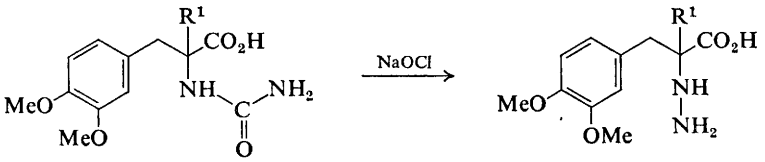
Both *erythro-* and *threo*-β-hydroxyleucine have been synthesized from β-isopropylglycidic acid,[68] and a detailed account of a number of unsuccessful routes to α-hydroxyamino-acids has appeared.[69]

[65] A. Chimiak and J. J. Pastuszak, *Chem. Ind. Internat.*, 1971, 427.
[66] G. Hofle and W. Steglich, *Chem. Ber.*, 1971, **104**, 1408.
[67] C. Berse and P. Bessette, *Canad. J. Chem.*, 1971, **49**, 2610.
[68] S. Futagawa, M. Nakahara, T. Inui, H. Katsura, and T. Kaneko, *Nippon Kagaku Zasshi*, 1971, **92**, 374.
[69] Y. Liwschitz, A. Singerman, Y. Wiesel, M. Michman, P. Braun, S. Kassel, and D. Perera, *Israel J. Chem.*, 1971, **9**, 89.

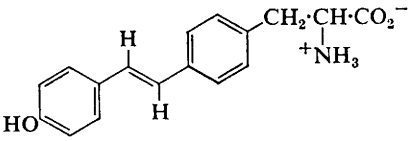


(19) (20)

F. Aromatic and Heterocyclic α-Amino-acids.—Aromatic amino-acids continue to be synthesized as potential enzyme inhibitors. A series of 6-alkyl-Dopa derivatives [70] and 3-(2,5-dimethoxy-4-methyl)phenylalanine [71] have been prepared by standard routes, and an improved synthesis of 6-hydroxy-Dopa is claimed.[72] A variety of tyrosine derivatives [73, 74] and 2-aminoindan-2-carboxylic acids [75] have been prepared as possible inhibitors of tyrosine hydroxylase. A novel method for the synthesis of α-hydrazino-acids related to L-Dopa [76] has been reported. The key intermediate was the hydantoic acid (21) (prepared from the amino-acid and potassium cyanate) which on treatment with sodium hypochlorite afforded the hydrazino-acid (22) in good yield. The stereochemistry at the α-centre was retained, but in all the cases so far investigated the α-centre was fully substituted. It is possible that this method may have a broader application to the synthesis of α-hydrazino-acids.



(21) (22)



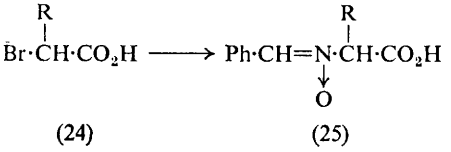
(23)

[70] A. P. Morgenstern, C. Schuijt, and W. Th. Nauta, *J. Chem. Soc.* (*C*), 1971, 3706.
[71] K. Brewster and R. M. Pinder, *J. Medicin. Chem.*, 1971, **14**, 650.
[72] F. Lee, G. H. Dickson, E. Donald, and A. A. Manian, *J. Medicin. Chem.*, 1971, **14**, 266.
[73] Y. H. Caplan, N. Zenker, D. A. Blake, and E. M. Johnson, *J. Medicin. Chem.*, 1971, **14**, 405.
[74] R. E. Counsell, P. Desai, A. Ide, P. G. Kulkarni, P. A. Weinhold, and V. B. Rethy, *J. Medicin. Chem.*, 1971, **14**, 789.
[75] R. M. Pinder, B. H. Butcher, D. A. Buxton, and D. J. Howells, *J. Medicin. Chem.*, 1971, **14**, 892.
[76] S. Karady, M. G. Ly, S. H. Pines, and M. Sletzinger, *J. Org. Chem.*, 1971, **36**, 1946, 1949.

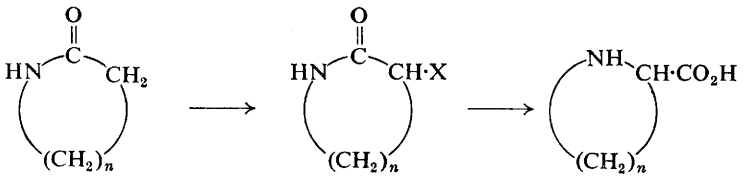
A number of fluoro-[77, 78] and amino-phenylalanines,[79] as well as *p*-azidophenylalanine,[80] designed as a photoaffinity probe in synthetic peptides, has been reported. Interest continues in structural analogues of thyroxine; the stilbene derivative (23) has been synthesized by means of a Wittig reaction.[81]

A wide range of pyridyl,[82, 83] pyridimyl,[84–86] and purinyl [87, 88] amino-acids have been synthesized mainly by well-established routes for a variety of specific reasons too diverse to enumerate. The first example of a ring-fluorinated histidine derivative has been obtained [89] by a new route involving the photochemical decomposition of diazonium fluoroborates, and the method promises to offer a general route to aromatic and heterocyclic fluorination.

G. N-Substituted α-Amino-acids.—A further method for the preparation of *N*-methylamino-acids, which is also claimed to give high yields and optical purity, has been reported.[90] The reaction of the L-bromo-acid (24) with *anti*-benzaldoxime afforded the *N*-oxide (25) which, on hydrogenation and hydrolysis, gave D-phenylalanine,[91] thus demonstrating that the reaction



(24) (25)



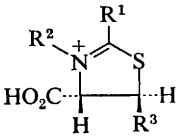
(26) (27) (28)

[77] A. T. Prudchenko, *Izvest. sibirsk. Otdel. Akad. Nauk, Ser. khim. Nauk*, 1970, 95.
[78] J. L. Fauchere and R. Schwyzer, *Helv. Chim. Acta*, 1971, **54**, 2078.
[79] I. Straukas, N. Dirvianskyte, and J. Degutis, *Zhur. org. Khim.*, 1971, **7**, 1390.
[80] R. Schwyzer and M. Caviezel, *Helv. Chim. Acta*, 1971, **54**, 1395.
[81] G. Jones and S. Wright, *J. Chem. Soc.* (*C*), 1971, 141.
[82] S. J. Scotty, P. T. Sullivan, and C. B. Sullivan, *J. Medicin. Chem.*, 1971, **14**, 211.
[83] P. T. Sullivan and S. J. Norton, *J. Medicin. Chem.*, 1971, **14**, 557.
[84] M. Y. Lidak, R. A. Paegle, M. G. Plata, and Y. P. Shvachkin, *Khim. geterotsikl. Soedinenii*, 1971, 530.
[85] R. A. Paegle, M. G. Plata, M. Y. Lidak, and Y. Y. Popel, *Khim. geterotsikl. Soedinenii*, 1971, 258.
[86] I. Y. Ulane and M. Y. Lidak, *Khim. geterotsikl. Soedinenii*, 1971, 527.
[87] D. S. Letham and H. Young, *Phytochemistry*, 1971, **10**, 23.
[88] M. Y. Lidak, Y. Y. Shluke, S. E. Poritere, and Y. P. Shvachkin, *Khim. geterotsikl. Soedinenii*, 1971, 427.
[89] K. L. Kirk and L. A. Cohen, *J. Amer. Chem. Soc.*, 1971, **93**, 3060.
[90] J. R. Coggins and N. L. Benoiton, *Canad. J. Chem.*, 1971, **49**, 1968.
[91] B. Liberek and Z. Palacz, *Roczniki Chem.*, 1971, **45**, 1173.

had proceeded with inversion of configuration at the α-centre. A novel route to α-hydrazino-acids from the corresponding amino-acid has been described in the foregoing section.[76]

An interesting variation of the Favorski reaction has been employed to synthesize a series of ring homologues of proline.[92] α-Halogenation of the readily available ω-aminolactams (26) afforded the α-halogeno-compounds (27) which underwent ring contraction on treatment with base to give the α-imino-acids (28). The already extensive programme on the synthesis of *N*-bis-2-halogenoethyl derivatives of amino-acids as potential antimetabolites has been extended.[93–95]

H. α-Amino-acids containing Sulphur.—Thioamides react smoothly with α-bromo-αβ-unsaturated acids to form the thiazolinium derivatives (29) which on hydrolysis yield N-substituted cysteines. The rate of thiazolinium



(29)

formation decreases with increasing number of substituents owing to unfavourable steric interactions; nevertheless the reaction offers an interesting and versatile route to substituted cysteine derivatives.[96] The syntheses of a series of *S*-alkyl-2-methyl cysteine [97] and *S*-alkylhomocystine derivatives [98] employing standard procedures have been reported.

I. α-Amino-acids which have been Synthesized for the First Time

Compound	*Ref.*
L-3-Carboxy-6,7-dihydroxy-1,2,3,4-tetrahydroisoquinoline	15
'γ-L-Glutaminyl-4-hydroxybenzene' (*N*-γ-L-glutamyl-*p*-aminophenol)	27
S-(3-Hydroxy-3-carboxyethylthio)homocystine	32
S-(2-Hydroxy-2-carboxyethylthio)homocysteine	32
cis- and *trans*-3,4-methylene-L-proline	60
3-(2-Methyl-4,5-dihydroxyphenyl)-DL-alanine	70
3-(2-Ethyl-4,5-dihydroxyphenyl)-DL-alanine	70
3-(2-Isopropyl-4,5-dihydroxyphenyl)-DL-alanine	70
3-(2-t-Butyl-4,5-dihydroxyphenyl)-DL-alanine	70
3-(2,5-Dimethoxy-4-methylphenyl)-DL-alanine	71
3-Ethyl-α-methyl-DL-tyrosine	74

[92] H. T. Nagasawa, J. A. Elberling, P. S. Fraser, and N. S. Mizumo, *J. Medicin. Chem.*, 1971, **14**, 501.
[93] G. G. Blinova and E. G. Sochilin, *Zhur. obshchei Khim.*, 1970, **40**, 2748.
[94] L. V. Alekseeva, N. L. Burde, and Z. V. Pushkareva, *Zhur. org. Khim.*, 1971, **7**, 647.
[95] G. A. Davydova, B. M. Karmanskaya, I. A. Redkin, and A. I. Tochilkin, *Khim. Farm. Zhur.*, 1971, **5**, 11.
[96] A. Eidem, K. Undheim, and K. R. Reistad, *Acta Chem. Scand.*, 1971, **25**, 1.
[97] S. Tahara and Y. Obata, *Agric. and Biol. Chem. (Japan)*, 1971, **35**, 53.
[98] A. Rinaldi and C. De Marco, *Ital. J. Biochem.*, 1971, **20**, 1.

Compound	*Ref.*
3-Isopropyl-α-methyl-DL-tyrosine	74
3-t-Butyl-α-methyl-DL-tyrosine	74
DL-2-Amino-5-hydroxy-indan-2-carboxylic acid	75
DL-2-Amino-5-methoxy-indan-2-carboxylic acid	75
DL-2-Amino-5-carboxy-indan-2-carboxylic acid	75
DL-2-Amino-5-chloro-indan-2-carboxylic acid	75
DL-2-Amino-5-bromo-indan-2-carboxylic acid	75
DL-2-Amino-5-iodo-indan-2-carboxylic acid	75
L-α-(3,4-Dihydroxybenzyl)-α-hydrazinopropionic acid	76
L-α-(3,4-Dimethoxybenzyl)-α-hydrazinopropionic acid	76
L-α-(3,4-Dihydroxybenzyl)-α-hydrazinobutyric acid	76
L-α-(3,4-Dimethoxybenzyl)-α-hydrazinobutyric acid	76
3-(2,4-Difluorophenyl)-DL-alanine	77
3-(3,4-Difluorophenyl)-DL-alanine	77
3-(3,5-Difluorophenyl)-DL-alanine	77
3-(2,5-Difluorophenyl)-DL-alanine	77
3-(2,6-Difluorophenyl)-DL-alanine	77
3-(2,3,5,6-Tetrafluorophenyl)-DL-alanine	77
3-(3,5-Dichloro-2,4,6-trifluorophenyl)-DL-alanine	77
3-(2,3,4,5,6-Pentafluorophenyl)-DL-alanine	78
3-(*m*-Bis-2-chloroethylaminophenyl)-DL-serine (*threo* and *erythro*)	79
3-(*p*-Bis-2-chloroethylaminophenyl)-DL-serine (*threo* and *erythro*)	79
p-Azido-L-phenylalanine	80
trans-β-4-*p*-hydroxystyryl-DL-phenylalanine	81
β-(1,2-Dihydro-2-oxo-3-pyridyl)-DL-alanine	82
β-(1,2-Dihydro-2-oxo-4-pyridyl)-DL-alanine	82
β-(1,2-Dihydro-2-oxo-5-pyridyl)-DL-alanine	82
β-(1,2-Dihydro-2-oxo-6-pyridyl)-DL-alanine	82
β-(2-Fluoro-3-pyridyl)-DL-alanine	82
β-(2-Fluoro-5-pyridyl)-DL-alanine	82
β-(2-Fluoro-6-pyridyl)-DL-alanine	82
β-(2-Bromo-3-pyridyl)-DL-alanine	83
β-(2-Bromo-4-pyridyl)-DL-alanine	83
β-(2-Bromo-5-pyridyl)-DL-alanine	83
β-(2-Bromo-6-pyridyl)-DL-alanine	83
β-(2-Chloro-3-pyridyl)-DL-alanine	83
β-(2-Chloro-4-pyridyl)-DL-alanine	83
β-(2-Chloro-5-pyridyl)-DL-alanine	83
β-(2-Chloro-6-pyridyl)-DL-alanine	83
β-(Thymin-1-yl)-DL-alanine	84
β-(5-Fluorouracil-1-yl)-DL-alanine	84
β-(5-Bromouracil-1-yl)-DL-alanine	84
β-(5-Chlorouracil-1-yl)-DL-alanine	84
β-(Cytosin-1-yl)-DL-alanine	84
N-(2-Chloro-5-fluoro-4-pyrimidyl)glycine	85
N-(2-Chloro-5-fluoro-4-pyrimidyl)-DL-phenylalanine	85
N-(2-Chloro-5-fluoro-4-pyrimidyl)-DL-leucine	85
N-(2-Chloro-5-fluoro-4-pyrimidyl)-DL-valine	85
N-(2-Chloro-5-fluoro-4-pyrimidyl)-DL-tryptophan	85
N-(2-Ethylthio-5-fluoro-4-pyrimidyl)glycine	85
N-(2-Ethylthio-5-fluoro-4-pyrimidyl)-DL-phenylalanine	85
N-(2-Ethylthio-5-fluoro-4-pyrimidyl)-DL-leucine	85
N-(2-Ethylthio-5-fluoro-4-pyrimidyl)-DL-valine	85
N-(2-Ethylthio-5-fluoro-4-pyrimidyl)-DL-tryptophan	85

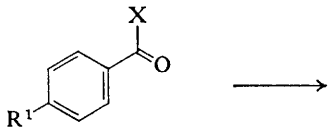
Compound	*Ref.*
N-(2-Chloro-5-bromo-4-pyrimidyl)-DL-alanine	86
N-(2-Chloro-5-bromo-4-pyrimidyl)-DL- and L-leucine	86
N-(2-Chloro-5-bromo-4-pyrimidyl)-DL- and L-valine	86
N-(2-Chloro-5-bromo-4-pyrimidyl)-DL-tryptophan	86
N-(2-Chloro-5-bromo-4-pyrimidyl)-L-isoleucine	86
N-(2-Chloro-5-iodo-4-pyrimidyl)-DL-leucine	86
N-(2-Chloro-5-iodo-4-pyrimidyl)-DL-valine	86
N-(2-Chloro-5-iodo-4-pyrimidyl)-DL-alanine	86
N-(Purin-6-yl)-DL-α-phenylglycine	87
N-(Purin-6-yl)-L-leucine	87
N-(Purin-6-yl)-L-valine	87
DL-2-(Purin-6-yl-amino)-5-methylhex-4-enoic acid	87
N^{ε}-(5-Amino-6-chloro-4-pyrimidyl)lysine	88
α-Amino-ε-(6-chloro-9-purinyl)caproic acid	88
4-Fluoro-DL-histidine	89
DL-Hexahydro-1*H*-azepine-2-carboxylic acid	92
DL-Octahydro-2-azocine carboxylic acid	92
DL-Octahydro-2-azonine-2-carboxylic acid	92
DL-Decahydro-2-azecine carboxylic acid	92
DL-Azacycloundecane-2-carboxylic acid	92
S-Methyl-2-methyl-DL-cysteine	97
S-Ethyl-2-methyl-DL-cysteine	97
S-Propyl-2-methyl-DL-cysteine	97
S-Isopropyl-2-methyl-DL-cysteine	97
S-Butyl-2-methyl-DL-cysteine	97
S-Isobutyl-2-methyl-DL-cysteine	97
S-t-Butyl-2-methyl-DL-cysteine	97
S-Amyl-2-methyl-DL-cysteine	97
S-Isoamyl-2-methyl-DL-cysteine	97
S-Allyl-2-methyl-DL-cysteine	97
S-(β-Aminoethyl)homocysteine	98

J. Labelled Amino-acids.—The commercial availability of a wide range of ^{14}C-labelled amino-acids is reflected in the noticeable drop in the publications relating to their syntheses, and those described have been performed by standard procedures such as the Strecker synthesis.[99] The chemical emphasis appears to have shifted to stereospecific syntheses of tritiated and deuteriated amino-acids and this has led to detailed examinations of the mechanisms of some of the general amino-acid syntheses.

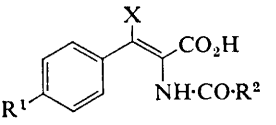
In last year's Report it was pointed out that the synthesis of labelled aldehydes of the type (30) had considerable potential for specific labelling of amino-acids at the prochiral β-centre. This potential has now been realized by two groups working independently.[100, 101] Condensation of (30) with an *N*-acylglycine derivative gave an oxazolone which opened with alkali to yield the acylaminocinnamic acid (31) of established *trans*-configuration. Catalytic hydrogenation of (31) was expected, and observed,

[99] L. Pichat, P. N. Liem, and J. P. Guermont, *Bull. Soc. chim. France*, 1971, 837.
[100] K. R. Hanson, R. H. Wightman, J. Staunton, and A. R. Battersby, *Chem. Comm.*, 1971, 185.
[101] G. W. Kirby and J. Michael, *Chem. Comm.*, 1971, 415.

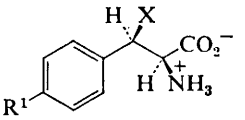
to proceed in a *syn*-stereospecific fashion. Resolution of the resulting acylamino-acid by enzymic deacylation afforded the (3*R*)-L- and (3*S*)-D-amino-acids (32) and (33) respectively. So far this method has only been applied to derivatives of phenylalanine and tyrosine but it undoubtedly has wider potential.



(30) X = ^{2}H or ^{3}H

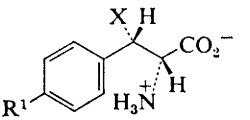


(31)



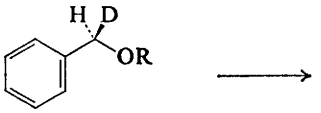
(32)

X = ^{2}H or ^{3}H

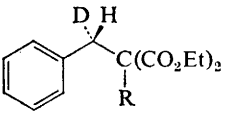


(33)

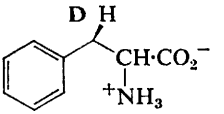
An alternative route[101,102] to some of the deuteriated derivatives described above, which has also confirmed the stereochemical assignments, employs (1*S*)-[1-^{2}H]benzyl alcohol (34). Enzymic reduction of benzaldehyde under established conditions afforded (34) and reaction of the corresponding tosylate (35) with diethyl malonate or with diethyl *N*-acetamidomalonate gave (36) and (37) respectively, which were converted by standard methods into (3*R*)-DL-[3-^{2}H]phenylalanine (38). The reaction proceeds with inver-



(34) R = H
(35) R = tosyl



(36) R = H
(37) R = NH·COMe



(38)

sion of configuration at the β-centre and the high degree of stereoselectivity has been established by physical methods and by comparison with the corresponding products from the azlactone route described above.

[102] R. Ife and E. Haslam, *J. Chem. Soc.* (*C*), 1971, 2818.

A convenient and specific general method for the preparation of α-deuteriated and α-tritiated amino-acids involves incorporation of solvent isotopic hydrogen concomitant with decarboxylation of synthetic substituted aminomalonate precursors.[103] Reaction of substituted pyruvic acids with ^{15}N-labelled ammonia and sodium cyanohydridoborate affords a facile route to ^{15}N-labelled amino-acids. The use of the deuteriated or tritiated cyanohydridoborate and ^{15}N-labelled ammonia allows the synthesis of doubly labelled amino-acids.[104]

K. Resolution of α-Amino-acids.—A general review which analyses the problems associated with enzymic, chemical, and chromatographic methods for resolving optical isomers has appeared.[105] The separation of racemates on asymmetric sorbents continues to attract attention. A co-polymer of a chloromethylated styrene-*p*-divinylbenzene and L-proline is claimed to resolve quantitatively DL-amino-acids when used as sorbent in the presence of transition-metal ions.[106, 107] The basis of the technique is stated to be the preferential transition-metal complex formation between the L-amino-acid bound in the resin and the D-isomer in solution.

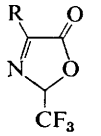
It has been known for some time that optically active stationary phases, usually dipeptides, are effective in resolving amino-acid derivatives by gas–liquid chromatography, but so far this method has been limited to analytical applications rather than preparative use; further investigations in this area have been reported.[108]

The use of carrier-bound enzymes, *i.e.* water-insoluble enzyme systems, for the resolution of amino-acids is now well established. A continuous process using carrier-bound hog acylase has been developed which is claimed to be more efficient than the normal process, since the inhibition of the enzyme is reduced by the continual removal of the reaction products.[109] The papain-catalysed reaction between racemic *N*-acylamino-acids and phenylhydrazines has been shown to proceed preferentially with the L-isomer to give the corresponding optically active hydrazide.[110] Practically 100% optical efficiency was achieved with *o*-fluorophenylhydrazine, and the method represents a novel application of the reverse of the normal proteolytic action of papain. Chymotrypsin-catalysed reactions have been employed to resolve a series of ring-substituted phenylalanine esters.[111]

Chemical methods of resolution of synthetic racemates are still commonly

[103] J. W. Thanassi, *J. Org. Chem.*, 1971, **36**, 3019.
[104] R. F. Borch, M. D. Bernstein, and H. D. Durst, *J. Amer. Chem. Soc.*, 1971, **93**, 2897.
[105] S. V. Rogozhin, *Vestnik Akad. Nauk S.S.S.R.*, 1971, **40**, 56.
[106] S. V. Rogozhin, V. A. Davankov, V. V. Korshak, V. Vesa, and A. L. Belchich, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1971, 502.
[107] S. V. Rogozhin and V. A. Davankov, *Chem. Comm.*, 1971, 490.
[108] W. Parr, J. Pleterski, C. Yang, and E. Bayer, *J. Chromatog. Sci.*, 1971, **9**, 141.
[109] T. Barth and H. Mašková, *Coll. Czech. Chem. Comm.*, 1971, **36**, 2398.
[110] J. L. Abernethy, E. Albano, and J. Comyns, *J. Org. Chem.*, 1971, **36**, 1580.
[111] J. H. Tong, C. Petitclerc, A. Diorio, and N. L. Benoiton, *Canad. J. Biochem.*, 1971, **49**, 877.

employed for routine work and details on the use of ephedrine [112] and menthol [113] have been described. An obvious limitation to the normal methods of resolution is that, if a specific enantiometer is required, then the maximum yield is only 50%. An interesting preliminary report on a possible method of converting racemates completely into either the D- or the L-enantiomer has been outlined.[114] The reaction of a racemic amino-acid with trifluoroacetic anhydride leads to the oxazolinone (39) which with dimethyl L-glutamate affords the LL-dipeptide derivative (40). Hydrolysis of the dipeptide, in the case of t-leucine, the only example so far reported, gave *ca.* 80% chemical and optical yields of the L-isomer. Whether or not this method has a wider application must await further evaluation.



(39)

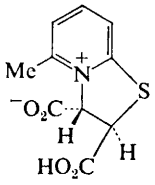
$$CF_3CO{\cdot}NH{\cdot}\underset{(L)}{CH(R)}{\cdot}CO{\cdot}NH{\cdot}\underset{(L)}{CH(CO_2Me)}{\cdot}(CH_2)_2{\cdot}CO_2Me$$

(40)

3 Physical and Stereochemical Studies of α-Amino-acids (See also Chapter 2, Part II, Section 2, and Part III)

A. Crystal Structures of Amino-acids.—The crystal structures of L-isoleucine,[115] L-tyrosine,[116] L-arginine,[117] L-cysteine,[117] DL-lysine,[117] DL-phenylalanine,[117] L-α*γ*-diaminobutyric acid [118] and the hydrochloride,[119] L-Dopa hydrochloride,[120] and DL-tryptophan formate,[121] have been described, and also those of the amides, *N*-acetyl-L-prolyl methylamide,[122] *N*-acetyl-L-methionyldimethylamide,[123] and 3,5-di-iodo-L-thyroninyl-methylamide.[124] The structure, configuration, and conformation in the crystal state of the novel amino-acid (14) have been established by an *X*-ray crystallographic analysis.[60] An analysis of the addition compound derived from bromomalic anhydride and 6-methylpyrid-2-thione has confirmed the structure (41) proposed on the basis of chemical and physical data.[125]

[112] H. Kinoshita, M. Shintani, T. Saito, and H. Kotake, *Bull. Chem. Soc. Japan*, 1971, **44**, 286.
[113] V. M. Belikov, T. F. Saveleva, and E. N. Safonova, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1971, 1461.
[114] W. Steglich, E. Frauendorfer, and F. Weygand, *Chem. Ber.*, 1971, **104**, 687.
[115] K. Torii and Y. Iitaka, *Acta Cryst.*, 1971, **B27**, 2237.
[116] A. Mostad, H. M. Nissen, and C. Rømming, *Tetrahedron Letters*, 1971, 2131.
[117] B. Khawas, *Acta Cryst.*, 1971, **B27**, 1517.
[118] P. S. Naganathan and K. Venkatesan, *Acta Cryst.*, 1971, **B27**, 2159.
[119] H. Hinazumi and T. Mitsui, *Acta Cryst.*, 1971, **B27**, 2152.
[120] R. J. Jandacek and K. M. Earle, *Acta Cryst.*, 1971, **B27**, 841.
[121] E. Bye, A. Mostad, and C. Rømming, *Acta Chem. Scand.*, 1971, **25**, 364.
[122] T. Matsuzaki and Y. Iitaka, *Acta Cryst.*, 1971, **B27**, 507.
[123] A. Aubry, M. Marraud, J. Protas, and J. Neel, *Compt. rend.*, 1971, **273**, *C*, 959.
[124] V. Cody, W. L. Duax, and D. A. Norton, *Chem. Comm.*, 1971, 683.
[125] P. Groth, *Acta Chem. Scand.*, 1971, **25**, 118.



(41)

B. Nuclear Magnetic Resonance Spectra.—N.m.r. spectroscopy is now an indispensable tool for structural elucidation and has played a decisive role in the characterization of many of the new amino-acids described in Section 1 (see, for example, refs. 22, 25, 28, 29, and 34). A detailed n.m.r. study on the conformation of valine and phenylalanine derivatives in different solvents, using the variation of the vicinal coupling constants to calculate changes in the population of side-chain rotamers, suggests that intramolecular interactions are more important than the dielectric constant of the medium in determining the relative energies of the rotamers.[126]

Carbamate formation in solutions of α-amino-acids and carbonate-bicarbonate, as well as the conformation of the products, has been studied by n.m.r. techniques.[127] Also conformational studies have been reported on *N*-nitroso-derivatives of sarcosine and proline,[128] and *N*-acylproline derivatives.[129]

The use of lanthanide complexes to produce large differential shifts in the ^{1}H and ^{13}C n.m.r. spectra of a wide range of organic compounds has greatly simplified the interpretation of the spectra of complex molecules. Unfortunately the use of these shift reagents has, until now, been limited to solutions of non-co-ordinating organic solvents, thus excluding the majority of amino-acids. It has now been reported that the hydrated perchlorates of europium or praeseodymium afford considerable differential shifts on the spectra of compounds in deuterium oxide.[130] So far this technique has had limited application but developments in this area could be of considerable interest.

Interest continues in the ^{13}C n.m.r. spectra of amino-acids,[131] but the most significant development in this area is the application of pulsed Fourier transform techniques to make use of the natural abundance of this isotope.[132] The spectra were determined with the amino-acids bound to cationic resins, and deuterium-decoupled spectra of deuteriated amino-acids showed ^{13}C linewidths which were significantly narrower than those

[126] R. A. Newmark and M. A. Miller, *J. Phys. Chem.*, 1971, **75**, 505.
[127] R. U. Lemieux and M. A. Barton, *Canad. J. Chem.*, 1971, **49**, 767.
[128] F. H. C. Stewart, *Austral. J. Chem.*, 1971, **24**, 1949.
[129] H. L. Maia, K. G. Orrell, and H. N. Rydon, *Chem. Comm.*, 1971, 1209.
[130] F. A. Hart, G. P. Moss, and M. L. Staniforth, *Tetrahedron Letters*, 1971, 3389.
[131] W. Voelter, G. Jung, E. Breitmaier, and E. Bayer, *Z. Naturforsch.*, 1971, **26**, 213.
[132] H. Sternlicht, G. L. Kenyon, E. L. Packer, and J. Sinclair, *J. Amer. Chem. Soc.*, 1971, **93**, 199.

from proton-decoupled, protonated amino-acids. The resin method has the advantage of shortening the spin–lattice relaxation time for quaternary carbons and other carbons that do not have hydrogens bonded to them, thus enabling these to be distinguished from carbons which are bonded to hydrogens.

The recent advances in pulsed Fourier transform techniques have allowed n.m.r. studies of ^{15}N in natural abundance to be undertaken, and the ^{15}N chemical shifts of some amino-acid methyl ester hydrochlorides have been reported.[133] There are as yet insufficient data to give more than broad generalizations, but it appears that a γ-alkyl substituent effect operates similar to that observed in ^{13}C spectra, and the method could have considerable potential. The abundant ^{14}N isotope is less useful since it possesses an electric quadrupole moment which results in considerable line-broadening, but studies on amino-acids in the solid state have established the ^{14}N nuclear coupling constants.[134] A further study in the solid state has determined the proton spin–lattice relaxation times as a function of temperature and has related the results to the activation energies for reorientation of the various $^{+}NH_3$ groups.[135] The first routine measurements on amino-acids of tritium magnetic resonance spectra have been reported.[136] The problems, including self-radiolysis, were discussed in detail but it is difficult, at present, to envisage any significant advantages of this technique to amino-acid chemistry.

C. Optical Rotatory Dispersion and Circular Dichroism.—Surveys of the o.r.d. and c.d. curves of protein amino-acids and the c.d. curves of less-common amino-acids[137] have revealed that all compounds with the L-configuration at the α-centre give positive Cotton effects, provided that there is no other chromophoric system present. A general sector rule for α-amino-acids has now been proposed[138] which relates the sign and amplitude of the Cotton effect with the conformation and the absolute configuration at the α-centre. The rule is based on the octant rule and derives from the sector principle originally proposed for lactones. It depends on the basic assumption, for which there is considerable evidence, that in solution the $N—C^{\alpha}—COO$ atoms are co-planar. For the purpose of analysis the two C—O bonds in the carboxylate ion are considered as two equivalent ketone groups and the plane bisecting the carboxylate ion is taken as a symmetry plane (Figure 1, plane A). By assigning two additional planes P_1 and P_2 through the carboxylate carbon atom, each perpendicular to a C—O bond, the octant rule can be applied to each separately. The

[133] P. S. Pregosin, E. W. Randall, and A. I. White, *Chem. Comm.*, 1971, 1602.
[134] D. T. Edmonds and P. A. Speight, *Phys. Letters* (*A*), 1971, **34**, 325.
[135] R. G. C. McElroy, R. Y. Dong, M. M. Pintar, and W. F. Forbes, *J. Magn. Resonance*, 1971, **5**, 262.
[136] J. Bloxsidge, J. A. Elvidge, J. R. Jones, and E. A. Evans, *Org. Magn. Resonance*, 1971, **3**, 127.
[137] L. Fowden, P. M. Scopes, and R. N. Thomas, *J. Chem. Soc.* (*C*), 1971, 833.
[138] E. G. Jorgensen, *Tetrahedron Letters*, 1971, 863.

summation of the two effects leads to the cancellation of contributions in some 30° sectors and reinforcement in others, as illustrated in Figure 1. Since the carboxylate ion has a true plane of symmetry the corresponding sectors below the plane are of opposite sign.

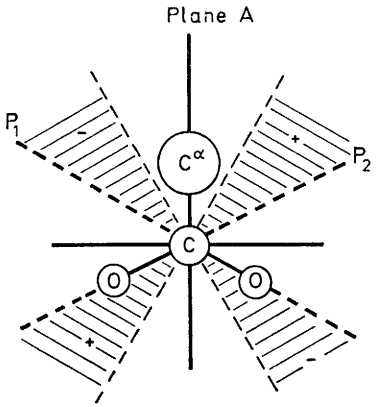


Figure 1

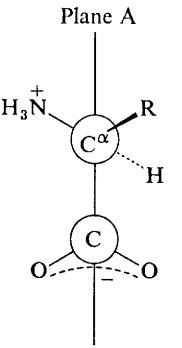


Figure 2

It can be seen from Figure 2 that, assuming the above-mentioned conditions apply, only the side-chain group and the α-hydrogen atom make significant contributions to the rotatory effect and that for L-α-amino-acids a greater amount of the positive sector is occupied. The rule appears to be generally applicable and rationalizes the low rotation of L-proline and the increase in rotation with increasing bulk of the side-chain, and it can be modified to account for the greater positive amplitude of L-α-amino-acids in acidic solution.

The c.d. curves for a range of α-aryl-α-amino-acids and their derivatives have been recorded. Those with the *S*-configuration all give strong positive Cotton effects and the conformations of these compounds have been discussed in relation to other α-amino-acids.[139] The o.r.d. and c.d. have also been reported for L-tyrosine,[140] L-tryptophan,[141] mono- and oligo-nucleotides[142] of L-α-amino-acids, and dithioethoxycarbonyl-L-α-amino-acids.[143]

[139] W. Klyne, P. M. Scopes, R. N. Thomas, and H. Dahn, *Helv. Chim. Acta*, 1971, **54**, 2420.

[140] E. H. Strickland, M. Wilchek, J. Horwitz, and C. Billups, *J. Biol. Chem.*, 1971, **247**, 572.

[141] H. Umeyama, T. Nagai, and H. Nogami, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 441.

[142] E. S. Gromova, B. V. Tyaglov, and Z. A. Shabarova, *Biochim. Biophys. Acta*, 1971, **240**, 1.

[143] K. Ishikawa, K. Achiwa, and I. S. Yamada, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 912.

D. Mass Spectrometry.—The continuing interest in the application of mass spectrometry to peptide sequence determination has led to further investigations on the mass spectra of amino-acid derivatives, and the spectra of methyl-,[144] phenyl-,[144] and phenylthio-hydantoins[145] as well as 2-anilino-5-thiazolinine[146] derivatives have been reported. For the same reason, chemical ionization mass spectrometry has been extended to phenythio-hydantoins. The technique leads to enhanced stability of the molecular ion and markedly reduces fragmentation, thus increasing the analytical potential of the method.[147]

Trimethylsilylation is still commonly employed to increase the volatility of amino-acids (see ref. 26) and the mass spectra of a series of trimethylsilyl derivatives of deuteriated and ^{13}C-enriched amino-acids have been determined,[148] in relation to their possible application to biosynthetic studies. Details of the fragmentation patterns of simple amino-acids have also been recorded.[149, 150]

E. Other Physical and Stereochemical Studies.—A considerable amount of detail on the conformation of free amino-acids[151, 152] and their derivatives,[153–155] both in solution and the solid state, has been deduced from i.r. and Raman spectroscopy. The heats of ionization of all the commonly occurring amino-acids have been calculated from calorimetric data[156] and the heats of solution of a number determined both in light and heavy water.[157]

The determination of the relative and absolute stereochemistry is an integral part of structure elucidation of amino-acids, and physical methods have played a predominant role in establishing the chirality of a considerable number of synthetic and naturally occurring amino-acids. The separation and characterization of the diastereoisomers of threonine,[158] γ-fluoroglutamic acid,[159] and β-methyltryptophan[160] have been reported.

[144] T. Sun and R. E. Lovins, *Analyt. Biochem.*, 1971, **45**, 176.
[145] F. Weygand and R. Obermeier, *European J. Biochem.*, 1971, **20**, 72.
[146] T. Fairwell and R. E. Lovins, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 1280.
[147] H. M. Fales, Y. Nagai, G. W. A. Milne, H. B. Brewer, J. T. Bronzert, and J. J. Pisano, *Analyt. Biochem.*, 1971, **43**, 288.
[148] W. J. A. Van den Heuvel, J. L. Smith, and J. S. Cohen, Proceedings of the Sixth International Symposium on Advances in Chromatography, 1970, p. 293.
[149] J. G. Lawless and M. S. Chadha, *Analyt. Biochem.*, 1971, **44**, 473.
[150] E. Stenhagen and B. A. Andersson, *Arch. Mass Spectral Data*, 1971, **2**, 146.
[151] P. K. Ponnuswamy and V. Sasisekharan, *Internat. J. Protein Res.*, 1971, **3**, 1, 9.
[152] T. Akimoto, M. Tsuboi, M. Kainosho, F. Tamura, A. Nakamura, S. Muraishi, and T. Kajuira, *Bull. Chem. Soc. Japan*, 1971, **44**, 2577.
[153] Y. Koyama, T. Shimanouchi, M. Sato, and T. Tatsumo, *Biopolymers*, 1971, **10**, 1059.
[154] J. Smolikovă, A. Vitek, and K. Bláha, *Coll. Czech. Chem. Comm.*, 1971, **36**, 2474.
[155] S. Boehm and B. Ruestow, *Studies Biophys.*, 1969, **13**, 169.
[156] M. A. Marini, R. L. Berger, D. P. Lam, and C. J. Martin, *Analyt. Biochem.*, 1971, **43**, 188.
[157] A. I. Klimov and V. I. Deshcherevskii, *Biofizika*, 1971, **16**, 556.
[158] Y. Ariyoshi and N. Sato, *Bull. Chem. Soc. Japan*, 1971, **44**, 2787.
[159] J. C. Unkeless and P. Goldman, *Mol. Pharmacol.*, 1971, **7**, 293.
[160] K. F. Turchin, M. N. Preobrazhenskaya, Yu. N. Sheinker, and N. N. Suvorov, *Zhur. org. Khim.*, 1971, **7**, 1290.

An additional chemical correlation between viomycidine and capreomycidine has provided further confirmation for the relative and absolute stereochemistry of these guanidine amino-acids.[161]

4 Chemical Studies of Amino-acids

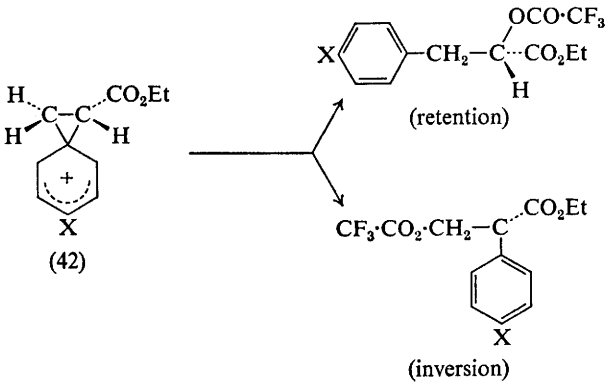
A. Introduction.—The first volume of what is intended as a continuous series of reviews on the general chemistry and biochemistry of amino-acids, peptides, and proteins has been published.[162] Although the first issue contained only a limited amount on amino-acids (assignment of configuration), the series offers a further source of information on amino-acid chemistry. Each year a great deal of research is published on the chemistry of amino-acids and their derivatives, much of it in relation to peptide synthesis, and this is presented elsewhere in this Report; the remainder is undertaken for a wide variety of reasons and it is therefore inevitable that this section is something of a miscellany.

B. General Reactions.—When amino-acids are heated at 1000 °C the predominant pyrolysis product is hydrogen cyanide;[163] in some cases the amino-nitrogen is almost quantitatively converted into hydrogen cyanide. At lower temperatures phenylalanine and tryptophan afford a mixture of polynuclear hydrocarbons and heterocyclic compounds.[164] A detailed investigation on the thermal decomposition employing differential scanning calorimetry has allowed thermolysis pathways to be proposed for a number of amino-acids in terms of the resolved thermograms.[165]

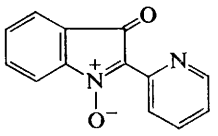
The stereochemistry and product distribution on nitrous acid deamination of L-phenylalanine, and of *p*-substituted phenylalanine ethyl esters in trifluoroacetic acid, suggests[166] that the reaction proceeds predominantly through the phenonium ion (42) as outlined in Scheme 6.

The deamination of a series of aminocycloalkane carboxylic acids has been reported;[167] the results were discussed in relation to the possible conformation of the ring systems. Strecker degradation of amino-acids with benzil afforded the expected aldehydes and tetraphenylpyrazine, which is claimed to arise by self-condensation of the intermediate 2-amino-2-phenylacetophenone.[168] Isatogen derivatives of the type (43) react with α-amino-acids in a similar manner to ninhydrin and isatin, causing oxidative deamination and decarboxylation to the aldehydes and isatogen reduction products.[169]

[161] C. Gallina, C. Marta, C. Colombo, and A. Romeo, *Tetrahedron*, 1971, **27**, 4681.
[162] 'Chemistry and Biochemistry of Amino-acids, Peptides, and Proteins', ed. B. Weinstein, Marcel Dekker, New York, 1971.
[163] W. R. Johnson and J. C. Kang, *J. Org. Chem.*, 1971, **36**, 189.
[164] J. M. Patterson, W. Y. Chen, and W. J. Smith, *Tobacco Sci.*, 1971, **15**, 41.
[165] P. G. Olafsson and A. M. Bryan, *Geochim. Cosmochim. Acta*, 1971, **35**, 327.
[166] K. Koga, C. C. Wu, and S. Yamada, *Tetrahedron Letters*, 1971, 2283.
[167] R. J. W. Cremlyn, R. M. Ellam, and T. K. Mitra, *J. Chem. Soc.* (*C*), 1971, 1647.
[168] A. F. Al-Sagyab, A. T. Atto, and F. Y. Sarah, *J. Chem. Soc.* (*C*), 1971, 3260.
[169] M. Hooper and J. W. Robertson, *Tetrahedron Letters*, 1971, 2139.



Scheme 6



(43)

The generally accepted mechanism for the photo-decarboxylation of 2,4-dinitrophenyl-α-amino-acids,[170] which involves decarboxylation with intramolecular oxygen transfer from the nitro-group, has been questioned.[171] It is pointed out that such a mechanism cannot explain the photo-decarboxylation in the solid state when the nitro-group is unaffected. Furthermore, it has been demonstrated that *N*-*p*-nitrophenylvaline is rapidly photo-decarboxylated to give isobutyraldehyde and *p*-nitro-aniline,[171] which suggests that the initial process is decarboxylation without oxygen transfer.

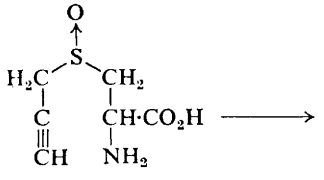
A generally efficient azeotropic method for the esterification of amino-acids has been reported.[172] The reduction of amino-acid esters proceeds more readily with lithium dimethoxyaluminium hydride than with lithium aluminium hydride itself,[173] and the esters also react with borane, trichloro-borane, and trifluoroborane to form adducts which can be converted into the corresponding N-substituted borazines.[174] A kinetic investigation extending further the programme on the Dakin–West reaction has provided evidence for the proposed pathway for the rearrangement of *N*-acyl-*s*-amino-acids.[175]

[170] O. Meth-Cohn, *Tetrahedron Letters*, 1970, 1235.
[171] P. H. MacFarlane and D. W. Russell, *Tetrahedron Letters*, 1971, 725.
[172] A. K. Saund and N. K. Mathur, *Indian J. Chem.*, 1971, **9**, 936.
[173] E. F. Rothgery and L. F. Hohnstedt, *Inorg. Chem.*, 1971, **10**, 181.
[174] M. Dymicky, E. F. Mellon, and J. Naghski, *Analyt. Biochem.*, 1971, **41**, 487.
[175] R. Knorr and G. K. Staudinger, *Chem. Ber.*, 1971, **104**, 3621, 3633.

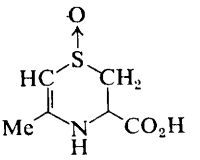
Amino-acids continue to be used as convenient chiral reagents both in synthesis and chromatography. The asymmetric synthesis of the alkaloid (+)-mesembrine has been achieved *via* an intermediate L-proline derivative.[176] Enantioselective Raney nickel catalysts have been prepared with amino-acids [39, 177] and amino-acid derivatives have been employed as stationary phases in ion-exchange [106, 107] and gas–liquid chromatography.[108]

C. Specific Reactions.—A systematic investigation of the desulphurization of sulphur-containing amino-acids with Raney nickel has established that cysteine and cystine can be completely desulphurized under relatively mild conditions, whereas methionine is essentially unchanged.[178] Similar results have been obtained using phosphorous acid, which has been used successfully with glutathione and oxytocin.[179]

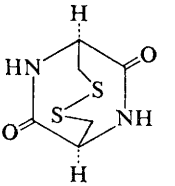
The alkynyl cysteine-*S*-oxide (44) and the corresponding dioxide, in the presence of base, undergo an internal addition of the amino-function to the triple bond to give a cyclic sulphoxide (45) and a cyclic sulphone respectively.[180] Bisulphite has been shown to catalyse the aerial oxidation of methionine to the *S*-oxide [181] and a detailed investigation of thiazolidine formation from L-cysteine and formaldehyde has been reported.[182] *S*-Acetamido-methylcysteine can be directly oxidized to cystine derivatives with iodine,[183] and the method has been applied to the synthesis of the previously unknown cyclo-L-cystine (46).



(44)



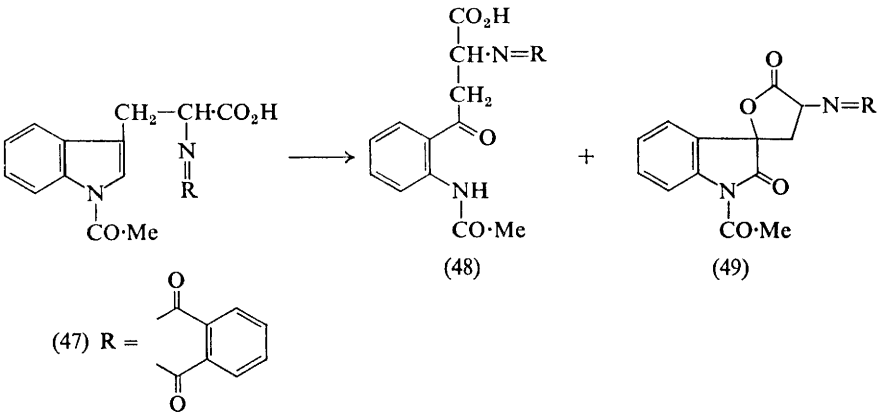
(45)



(46)

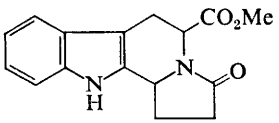
1-Acetyltryptophan has been synthesized from tryptophan *via* the *N*-phthaloyl derivative (47); surprisingly this compound had not previously been reported in the literature.[184] Oxidation of (47) with chromium trioxide afforded the kynurenine derivative (48) and the unusual dioxindole lactone (49).

[176] S. Yamada and G. Otani, *Tetrahedron Letters*, 1971, 1133.
[177] F. Higashi, T. Ninomiya, and Y. Izumi, *Bull. Chem. Soc. Japan*, 1971, **44**, 1333.
[178] M. T. Perlstein, M. Z. Atassi, and S. H. Cheng, *Biochem. Biophys. Acta*, 1971, **236**, 174.
[179] C. Ivanov and C. O. Ivanov, *Doklady Bolg. Akad. Nauk*, 1971, **23**, 1365.
[180] J. F. Carson and L. E. Boggs, *J. Org. Chem.*, 1971, **36**, 611.
[181] M. Inoue and H. Hikoya, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 1286.
[182] R. G. Kallen, *J. Amer. Chem. Soc.*, 1971, **93**, 6227, 6236.
[183] B. Kamber, *Helv. Chim. Acta*, 1971, **54**, 927.
[184] S. Ohki and T. Nagasaka, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 545, 603.



(47) R =

Tryptophan methyl ester reacts smoothly with 3-pyrrolin-2-one to yield (50); the stereochemical course of the reaction is not defined but it is suggested that the reaction could be employed for modifiying peptide structures.[185]



(50)

A new mild reductive cleavage of acyl-guanidines to amines has been demonstrated by the conversion of arginine into ornithine.[186] The reaction of nitromalondialdehyde with arginine in aqueous alkaline media results in the quantitative formation of δ-(5-nitro-2-pyrimidyl)ornithine.[187] The reaction may have potential in peptide sequence work and in mass spectrometry for increasing the volatility of guanidine-containing amino-acids.

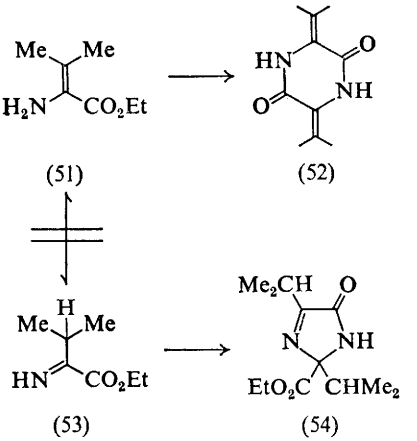
The *N*-oxides of *NN*-alkylamino-acids readily decarboxylate when heated with toluene-*p*-sulphonyl chloride in pyridine to yield the secondary amine and formaldehyde,[188] presumably by the mechanism outlined in Scheme 7.

The αβ-dehydrovaline ester (51), prepared by reduction of the corresponding α-nitro-ester, dimerizes on heating[189] to give the diketopiperazine (52), whereas the α-imino-ester (53), prepared from the α-keto-ester by the

[185] V. Bocchi, G. Casnati, and G. P. Gardini, *Tetrahedron Letters*, 1971, 683.
[186] J. S. Bland and J. F. W. Keana, *Chem. Comm.*, 1971, 1024.
[187] A. Signor, G. M. Bonora, L. Biondi, D. Nisato, A. Marzotto, and E. Scoffone, *Biochemistry*, 1971, **10**, 2748.
[188] Y. Ikutani, *Bull. Chem. Soc. Japan*, 1971, **44**, 271.
[189] C. Shin, M. Masaki, and M. Ohta, *Bull. Chem. Soc. Japan*, 1971, **44**, 1657.

addition of *N*-phenyltriphenylphosphinimine, cyclizes at room temperature to the imidazolidone (54). It is suggested that under the conditions of the two reactions there is no interconversion between (51) and (53).

Scheme 7

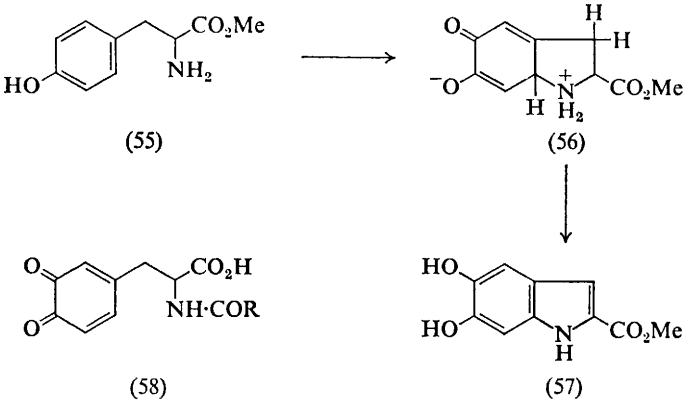


A novel synthesis of D-ribose has been reported which employs as the starting material L-glutamic acid, the asymmetric α-centre of which is subsequently converted into C-4 of D-ribose.[190]

D. Non-enzymic Models of Biochemical Processes Involving Amino-acids.— The chemical oxidation of tyrosine with potassium nitrosodisulphonate (Fremy's salt) has been found to follow the proposed enzymic pathways. Tyrosine methyl ester (55) is oxidized to 2-carboxy-5,6-dihydroxyindole (57) through the intermediacy of (56), which possesses the characteristic spectral properties of the enzyme-catalysed intermediate for amino-terminal tyrosine peptides. Carboxy-terminal peptides are oxidized by Fremy's salt to the *o*-quinone intermediate (58) which on reduction affords 3,4-dihydroxyphenylalanine derivatives, again paralleling the enzymic pathway.[191]

The stability of organically bound nitrogen in soil humic acids has been attributed to combination of amino-acids, peptides, and proteins with quinones. In order to investigate the chemical ability of this type of compound the synthesis of a number of benzoquinones containing α-amino-acid esters has been described.[192] Model reactions involving the

[190] K. Koga, M. Taniguchi, and S. Yamada, *Tetrahedron Letters*, 1971, 266.
[191] S. Dukler, M. Wilchek, and D. Lavie, *Tetrahedron*, 1971, **27**, 607.
[192] P. A. Cranwell and R. D. Haworth, *Tetrahedron*, 1971, **27**, 1831.



(55) (56) (58) (57)

direct cleavage of diphenyl ether linkages, in studies related to the metabolism of thyroxine, have been documented.[193] The non-enzymic reaction of L-serine with indole and hydrogen sulphide, in the presence of pyridoxal, to give DL-tryptophan and DL-cystine respectively, is reported to proceed through an intermediate pyridoxylidene amino-acrylic acid derived by β-elimination of the Schiff base.[194]

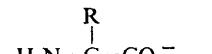
A synthetic chiral biphenyl derivative is stated to act as a stereospecific catalyst for the racemization of amino-acids.[195] Tropic acid with the *S*-configuration, *i.e.* naturally occurring tropic acid, has been obtained from L-phenylalanine by nitrous acid deamination, and the reaction is claimed to proceed by a route similar to the biochemical process.[196]

E. Effects of Electromagnetic Radiation on Amino-acids.—The radicals produced from amino-acids continue to be studied under various conditions. Both radiolysis and chemical techniques have been employed for the formation of radicals which were subsequently studied by e.s.r. spectroscopy, and it is important to note that in cases in which a comparison of the results from both methods is possible the e.s.r. spectra obtained are essentially identical. E.s.r. studies on radicals produced by radiolysis support the generally established pathways for the reaction of amino-acids with •OH radicals and hydrated electrons,[197–200] and in the aqueous alkaline solution of all the amino-acids investigated, radicals of the type (59) were observed.[197] Similar observations were made on the radicals produced with Ti^{3+}–H_2O_2 at pH 9—12, and in this case the temperature-dependent proton hyperfine splitting from the amino-hydrogens was noted and interpreted in terms of

[193] T. Matsuura, T. Nagamachi, and A. Nishinaga, *J. Org. Chem.*, 1971, **36**, 2016.
[194] K. Korte and U. Schmidt, *Monatsh.*, 1971, **102**, 207.
[195] K. Hirota and Y. Izumi, *Bull. Chem. Soc. Japan*, 1971, **44**, 2287.
[196] K. Koga, C. C. Wu, and S. Yamada, *Tetrahedron Letters*, 1971, 2287.
[197] R. W. Fessenden and P. Neta, *J. Phys. Chem.*, 1971, **75**, 738.
[198] H. C. Box and E. E. Budzinski, *J. Chem. phys.*, 1971, **55**, 2446.
[199] G. Lassman and W. Damerau, *Studies Biophys.*, 1969, **17**, 195.
[200] V. T. Srinivasan and A. Van de Vorst, *Internat. J. Radiation Biol.*, 1971, **19**, 133.

the non-coplanarity of the amino-group with the nodal plane of the π-system.[201]

The radicals (60) and (61), which have different e.s.r. spectra, are formed from the $\cdot OH$ radical and glycine in aqueous solution; exchange between the two radicals can be induced by addition of phosphoric acid.[202]



(59)

(60)

(61)

Radiolysis of oxygenated cysteine[203] and cysteine formate[204] solutions has provided further evidence for the radiolytic pathways involving sulphur radicals. The major volatile products formed by radiolysis of *S*-n-propyl-L-cysteine sulphoxide and *S*-allyl-L-cysteine sulphoxide have been identified by a combination of g.l.c. and mass spectrometry.[205]

Flash spectrophotometry has been employed to follow the reactions of the hydrated electron with aromatic amino-acids and the results accord well with those obtained from pulse radiolysis.[206]

The photo-luminescence of aromatic amino-acids is still being extensively investigated,[207–210] usually with the object of applying the results to the elucidation of protein structure. The photolysis of tryptophan and tryptophan derivatives has attracted considerable attention; evidence has been presented which suggests that the primary photochemical reaction is N—H bond fission.[211] Photo-oxidation of tryptophan in aqueous solution at pH 6—9 affords *N*-formyl-kynurenine,[212, 213] but in dilute ammonia the main product is 4-(2-amino-2-carboxyethyl)quinazoline, which is not formed *via* formyl-kynurenine.[213] The rates of the flavin-sensitized photo-oxidation of tryptophan and of tyrosine have been determined,[214] and the role of the triplet state in their photo-ionization has been discussed.[215]

The general photochemically induced deamination of amino-acids has been investigated in detail.[216]

[201] R. Poupko, A. Loewenstein, and B. L. Silver, *J. Amer. Chem. Soc.*, 1971, **93**, 580.
[202] H. Paul and H. Fischer, *Helv. Chim. Acta*, 1971, **54**, 485.
[203] J. P. Barton and E. J. Packer, *Internat. J. Radiation Phys. Chem.*, 1970, **2**, 159.
[204] M. Morita, K. Sasai, M. Tajima, and M. Fujimaki, *Bull. Chem. Soc. Japan*, 1971, **44**, 2257.
[205] H. Nishimura, J. Mizutani, Y. Obata, and M. Namiki, *Tetrahedron*, 1971, **27**, 307.
[206] B. Finnstrom, *Photochem. and Photobiol.*, 1971, **13**, 375.
[207] E. Leroy, H. Lami, and G. Laustriat, *Photochem. and Photobiol.*, 1971, **13**, 411.
[208] P. G. Bogach, V. L. Zima, and A. M. Filenko, *Biofizika*, 1971, **16**, 340.
[209] I. P. Vinogradov and N. Ya. Dodonova, *Biofizika*, 1971, **16**, 343.
[210] J. Tournon and M. A. El-Bayoumi, *J. Amer. Chem. Soc.*, 1971, **93**, 6396.
[211] M. T. Pailthorpe and C. H. Nicholls, *Photochem. and Photobiol.*, 1971, **14**, 135.
[212] R. S. Asquith and D. E. Rivett, *Biochim. Biophys. Acta*, 1971, **252**, 111.
[213] W. E. Savige, *Austral. J. Chem.*, 1971, **24**, 1285.
[214] F. Y. H. Wu and D. B. McCormick, *Biochim. Biophys. Acta*, 1971, **236**, 479.
[215] E. E. Fesenko, *Studies Biophys.*, 1968, **9**, 155.
[216] A. Riva and R. Cultrera, *Ann. Chim. (Italy)*, 1971, **61**, 624.

5 Analytical Methods (See also Chapter 2, Part I, Section 2A)

The pattern established in previous Reports for this section is maintained; the majority of references are cited under the appropriate heading without discussion and only the pertinent advances are presented in more detail. The improvements and developments of technique for amino-acid analysis in relation to structural studies on proteins and peptides are also covered. A book describing recent developments in analytical methods has been published.[217]

A. Gas–Liquid Chromatography.—The wealth of literature relating to g.l.c. of derivatized amino-acids emphasizes the developing potential of this truly quantitative method. The great advantages of the technique are speed and sensitivity. The instrumentation is simple and the resolution is such that even optical isomers can be separated on an asymmetric support.[218–220] The problems encountered with derivatization have now been largely overcome and more convenient methods for sample handling and quantitative preparation with a minimum of manipulation are being developed.[221] It is perhaps significant that the methods employed to investigate the amino-acid content of extra-terrestrial material were predominantly g.l.c. methods (see Section 1A), and arising out of this work a g.l.c. technique for nanogramme amounts of amino-acids is now available.[222] Several papers report that the effective resolution of all twenty protein amino-acids has been documented,[223–225] establishing that the problems encountered with resolving histidine, arginine, tryptophan, and cysteine have now been largely overcome. A number of other papers relating to g.l.c. of amino-acids in general have been published [226–230] and the increasing application of the method for biological materials is evident.[231–234] Chromatography of phenyl- and methyl-hydantoin derivatives also would appear to be promising [235] and the

[217] A. Niederwieser and G. Pataki, 'New Techniques in Amino-acid, Peptide, and Protein Analysis', Ann Arbor-Humphrey Sci. Publishers, Ann Arbor, Michigan, 1971.
[218] B. Feibush, *Chem. Comm.*, 1971, 544.
[219] W. Parr and P. Howard, *Chromatographia*, 1971, **4**, 162.
[220] J. C. Dabiowiak and D. W. Cooke, *Analyt. Chem.*, 1971, **43**, 791.
[221] J. M. L. Mee and C. C. Brooks, *J. Chromatog.*, 1971, **62**, 138.
[222] R. W. Zumwalt, K. Kuo, and C. W. Gehrke, *J. Chromatog.*, 1971, **55**, 267; **57**, 193.
[223] C. W. Gehrke, K. Kuo, and R. W. Zumwalt, *J. Chromatog.*, 1971, **57**, 209.
[224] C. W. Gehrke, R. W. Zumwalt, and K. Kuo, *J. Agric. Food Chem.*, 1971, **19**, 605.
[225] C. W. Moss, M. A. Lambert, and F. J. Diaz, *J. Chromatog.*, 1971, **60**, 134.
[226] B. C. Pettit and J. E. Stouffer, ref. 148, p. 273.
[227] E. Gil-Av, W. Parr, C. Yang, and E. Bayer, ref. 148, p. 287.
[228] J. Metz, W. Ebert, and H. Weicker, *Chromatographia*, 1971, **4**, 259.
[229] D. A. Shearer and R. M. Warner, *Internat. J. Environ. Analyt. Chem.*, 1971, **1**, 11.
[230] K. Bergstrom and J. Gurtler, *Acta Chem. Scand.*, 1971, **25**, 175.
[231] F. J. Diaz, C. W. Moss, and M. A. Lambert, *Rev. Asoc. bioquim. argentina*, 1971, **36**, 67.
[232] J. L. Laseter, J. D. Weete, A. Albert, and C. H. Walkinshaw, *Analyt. Letters*, 1971, **4**, 671.
[233] J. E. R. Schultz, *Column*, 1971, **13**, 5.
[234] E. D. Pellizzari, C. Rising, J. H. Brown, R. W. Farmer, and L. F. Fabre, *Analyt. Biochem.*, 1971, **44**, 312.
[235] J. J. Pisano, T. J. Bronzert, and H. B. Brewer, *Analyt. Biochem.*, 1972, **45**, 43.

radiochromatographic assay of the optical purity of ^{14}C-labelled amino-acids has been described.[236]

B. Ion-exchange Chromatography.—Two review articles on ion-exchange techniques have been published [237, 238] and various aspects of quantification of automatic analyser data reported.[239–242] The effect of conditions of preparation of resins on their resolution efficiency,[243] the use of Chelex-100,[244] and the separation of acidic amino-acids have also been investigated.[245] A modified pyridine–formic acid solvent system is recommended for use with radioactive compounds; no initial desalting is necessary but the resolution is not as good as with citrate buffers.[246] The hydrolysis of proteins and peptides in the presence of tritiated hydrochloric acid affords a simple and novel means of estimating the degree of racemization.[247]

It has been suggested that in order to avoid errors in quantitative work it is necessary to determine the colour constants with hydrindatin-ninhydrin for every amino-acid of interest.[248] The use of titanous chloride with ninhydrin instead of stannous chloride is claimed to give better colorimetric analysis [249] and further details on the detection of amino-acids with ninhydrin have been reported.[250, 251] Further work on the use of 2,4,6-trinitrobenzenesulphonic acid for the quantitative determination of amino-acids has been described.[252]

C. Thin-layer Chromatography.—Amino-acid derivatives which do not react readily with ninhydrin can be detected on t.l.c. by exposing the chromatogram to bleaching powder and hydrochloric acid and subsequently spraying with starch–potassium iodide solution.[253] Further improved techniques for the t.l.c. of free amino-acids have been

[236] A. V. Barooshian, M. J. Lautenschleger, and W. G. Harris, *Analyt. Biochem.*, 1971, **44**, 543.
[237] Y. Watanabe and N. Hamada, *Kagaku Tokogyo*, 1971, **45**, 17.
[238] A. N. Kosharov and M. A. Marina, *Nov. Metody Modif. Biokhim. Issed. Zhivotnovod*, 1970, 36.
[239] W. F. Kwolek and J. F. Cavins, *J. Assoc. Offic. Analyt. Chemists*, 1971, **54**, 1283.
[240] L. S. Bates, *Analyt. Biochem.*, 1971, **41**, 158.
[241] E. Welte, E. Przemeck, and M. C. Nuh, *Z. Pflanzennernaehr. Bodenk.*, 1971, **128**, 243.
[242] J. P. Ellis and J. M. Prescott, *J. Chromatog.*, 1971, **61**, 152.
[243] J. Rahm, H. Weinova, and Z. Prochazka, *J. Chromatog.*, 1971, **60**, 256.
[244] J. Boisseau and P. Jouan, *J. Chromatog.*, 1971, **54**, 231.
[245] A. Smith, P. J. Peterson, and L. Fowden, *J. Chromatog.*, 1971, **62**, 144.
[246] M. Reford-Ellis and M. N. Kelson, *J. Chromatog.*, 1971, **59**, 434.
[247] J. M. Manning, *J. Amer. Chem. Soc.*, 1970, **92**, 7449.
[248] M. Yamamoto and J. L. Lowell, *J. Chromatog.*, 1971, **57**, 152.
[249] L. B. James, *J. Chromatog.*, 1971, **59**, 178.
[250] K. Samejima, W. Dairman, and S. Udenfriend, *Analyt. Biochem.*, 1971, **42**, 222.
[251] K. Samejima, W. Dairman, J. Stone, and S. Udenfriend, *Analyt. Biochem.*, 1971, **42**, 237.
[252] K. T. Kossmann, *Aerztl. Lab.*, 1971, **17**, 375.
[253] Y. Ariyoshi, N. Sato, H. Zenda, and K. Adachi, *Bull. Chem. Soc. Japan*, 1971, **44**, 2558.

described.[254, 255] The chromatography of dinitrophenylamino-acids on thin layers of proteins is claimed to give good results.[256] Conditions for the effective separation of the dinitrophenyl derivatives of arginine and lysine,[257, 258] and for the rapid separation of phenylthiohydantoins,[259] have been reported.

D. Other Methods.—A theoretical study of the paper electrophoretic separation of amino-acids gives results in good agreement with experimental observations.[260] Further details of instrumentation [261] and of the applications of paper electrophoresis to amino-acids [262] have been reported. Many other topics in the analytical chemistry of amino-acids have been discussed, including polarographic determinations in water [263] and in DMSO,[264] partition chromatography,[265, 266] fluorometric measurements,[267] enzyme electrode probes for D-amino-acids,[268] and the detection of sulphur-containing amino-acids.[269]

E. Determination of Specific Amino-acids.—Papers on the determination of the following amino-acids have appeared: glutamic and aspartic acids,[270] tryptophan,[271–273] proline,[274] hydroxyproline,[275] cystine,[276] methionine,[277, 278] lysine,[278, 279] ornithine,[280] histidine,[281] and thyroxine.[282]

[254] J. G. Heathcote, D. M. Davies, C. Haworth, and R. W. A. Oliver, *J. Chromatog.*, 1971, **55**, 377.
[255] A. F. Krivis, *Microchem. J.*, 1971, **16**, 391.
[256] P. R. Brady and R. M. Hoskinson, *J. Chromatog.*, 1971, **54**, 65.
[257] H. M. Jacoby and L. Spero, *Analyt. Biochem.*, 1971, **44**, 299.
[258] T. Skrabka-Blotnicka, *Chem. Analit.* (*Warsaw*), 1971, **16**, 631.
[259] K. D. Kuble and H. Konschewski, *Analyt. Biochem.*, 1971, **44**, 548.
[260] Y. Kiso and E. Falk, *J. Chromatog.*, 1971, **59**, 401.
[261] O. Sova and A. Sullova, *Chem. Listy*, 1971, **65**, 1102.
[262] T. Fukuda and K. Mikami, *Shokuhin Eiseigaku Zasshi*, 1970, **11**, 292.
[263] S. Fujiwara, Y. Umezawa, and H. Ishizuka, *Bull. Chem. Soc. Japan*, 1971, **44**, 1984.
[264] R. T. Koch and W. C. Purdy, *Analyt. Chim. Acta*, 1971, **54**, 271.
[265] Y. Tanimura, *Farumashia*, 1971, **7**, 628.
[266] B. A. Persson, *Acta Pharm. Sueccia*, 1971, **8**, 193.
[267] S. R. Nahorski, *Analyt. Biochem.*, 1971, **42**, 136.
[268] G. G. Guilbault and E. Hrabanova, *Analyt. Chim. Acta*, 1971, **56**, 285.
[269] P. E. Belliveau and R. W. Frei, *Chromatographia*, 1971, **4**, 189.
[270] M. E. Balis, *Methods Biochem. Analysis*, 1971, **20**, 103.
[271] M. Friedman and J. W. Finley, *J. Agric. Food Chem.*, 1971, **19**, 626.
[272] M. A. Marina and Yu. E. Shut, *Nov. Metody Modif. Boikhim. Issed. Zhivotnovod*, 1970, 50.
[273] T. Devenyi, J. Bati, and F. Fabian, *Acta Biochim. Biophys. Acad. Sci. Hung.*, 1971, **6**, 133.
[274] V. Koppel, *Tartu Riikliku Ulikooli Torin*, 1971, **270**, 41.
[275] D. Dabev and H. Struck, *Biochem. Med.*, 1971, **5**, 17.
[276] J. W. Purdie and D. E. Hanafi, *J. Chromatogr.*, 1971, **59**, 181.
[277] G. M. Ellinger and R. H. Smith, *Biochem. J.*, 1971, **124**, 151.
[278] S. Ferenczi, J. Bati, and T. Devenyi, *Acta Biochim. Biophys. Acad. Sci. Hung.*, 1971, **6**, 123.
[279] B. Wheeler, *Israel J. Agric. Res.*, 1971, **10**, 49.
[280] C. W. Moss, F. J. Diaz, and M. A. Lambert, *Analyt. Biochem.*, 1971, **44**, 458.
[281] D. A. Newton and G. K. Summer, *Clinical Chem.*, 1971, **17**, 386.
[282] N. N. Nihei, M. C. Gershengorn, T. Mitsuma, L. R. Stringham, A. Cordy, B. Kuchmy, and C. S. Hollander, *Analyt. Biochem.*, 1971, **43**, 433.

2

Structural Investigations of Peptides and Proteins

BY R. N. PERHAM, J. O. THOMAS, T. L. BLUNDELL, AND R. H. PAIN

PART I: Primary Structure and Chemical Modification *by J. O. Thomas and R. N. Perham*

1 Introduction

Another year, another Report, another attempt to compress the proverbial quart into the pint pot. Gibbon boasted that his English text was chaste, and that all licentious passages were left in the decent obscurity of a learned language.[1] Our readers will also find nothing licentious in these pages: our sins are (we earnestly hope) ones of omission not commission, and we beg the forgiveness of those whose work is inadvertently neglected. Since the amount of material published in this field increases annually and the rate of increase shows no sign of diminishing, such inadvertence regrettably cannot be avoided. Indeed, when confronted with the reference list for this article, your Reporters limply recalled the siren voice of Wordsworth.[2] But sterner counsel prevailed and we, at least, are the wiser for it.

The form of this Report follows that of previous years, with a somewhat arbitrary distinction drawn between methods and results.

2 Methods

The determination of the primary structure and the evolution of proteins has been reviewed [3] and a new method for the comparison of amino-acid sequences has been described.[4] This uses a measure of similarity between every pair of amino-acids based on observed substitutions in homologous proteins, and suggests that the gene duplication predicted for the origin of bacterial ferredoxins has not occurred in plant ferredoxins. Moreover, the sequence repetitions in cytochrome *c* were held not to be significant.

[1] Edward Gibbon, 'Memoirs of my Life and Writings'.

[2] Up! Up! my friend and quit your books
Or surely you'll grow double
Up! Up! my friend and clear your looks
Why all this toil and trouble?

William Wordsworth, 'The Tables Turned'.

[3] P. Jollès and J. Jollès, *Prog. Biophys. Mol. Biol.*, 1971, **22**, 97.

[4] A. D. McLachlan, *J. Mol. Biol.*, 1971, **61**, 409.

The proceedings of a conference devoted to recent developments in the chemical study of proteins have also been published [5] and should provide a useful source book for ideas and references on many aspects of the art.

A. Amino-acid Analysis. (See also Chapter 1, Section 5.)—The amino-acid analyses of 120 proteins drawn from the recent literature have been collated [6] and a statistical method of comparing analyses has been given.[7] It has been shown that if nitroarginine is present in a peptide, tyrosine and phenylalanine suffer chlorination during acid hydrolysis and that methionine, tryptophan, and serine are lost as unidentified products.[8] Presumably the inclusion of a little phenol in the hydrolysis mixture would cure much of this.

It has been reported [9] that the tryptophan fluorescence of eleven reduced proteins in the presence of sodium dodecyl sulphate (SDS) is directly proportional to the tryptophan content of the proteins, thereby making this the basis of an analytical method for tryptophan. Another study [10] has shown that if a protein is hydrolysed in 3N-toluene-*p*-sulphonic acid containing 0.2% 3-(2-aminoethyl)indole in evacuated tubes at 110 °C for up to 72 h, tryptophan is recovered quantitatively, together with the other amino-acids. Since the hydrolysate may be placed directly on to the ion-exchange columns of an amino-acid analyser, without the prior removal of acid that is required when 6N-HCl is used for the hydrolysis, this would seem to represent a neat step forward in technique. The complete hydrolysis of peptides by a mixture of Sepharose-bound peptidases has been described,[11] which also avoids the destruction of tryptophan and, additionally, retains the amide groups of asparagine and glutamine.

Improvements have been reported [12] in the determination of amino-groups with trinitrobenzenesulphonic acid, and a micro-method has been described [13] for detecting reactive carbonyl groups in proteins and peptides, using 2,4-dinitrophenylhydrazine. The colorimetric determination of proline in protein hydrolysates and biological fluids with isatin has also been improved [14] and a colorimetric assay for hydroxylysine has been reported.[15] Proline interferes with this assay and the hydroxylysine and proline must first be separated.

[5] 'Recent Developments in the Chemical Study of Protein Structures', Proceedings of Inserm Meeting, Montpellier, 1971, Institut National de la Santé de la Recherche Médicale, Paris, 1971.
[6] D. M. Kirschenbaum, *Analyt. Biochem.*, 1971, **44**, 159.
[7] H. M. Shapiro, *Biochim. Biophys. Acta*, 1971, **236**, 725.
[8] P. Moritz and R. Wade, *Analyt. Biochem.*, 1971, **41**, 446.
[9] K. R. Shelton and K. S. Rogers, *Analyt. Biochem.*, 1971, **44**, 134.
[10] T.-Y. Lui and Y. H. Chang, *J. Biol. Chem.*, 1971, **246**, 2842.
[11] H. P. J. Bennett, D. F. Elliott, P. J. Lowry, and C. McMarker, *Biochem. J.*, 1971, **125**, 80P.
[12] (*a*) R. Fields, *Biochem. J.*, 1971, **124**, 581; (*b*) B. V. Plapp, S. Moore, and W. H. Stein, *J. Biol. Chem.*, 1971, **246**, 939.
[13] R. Fields and H. B. F. Dixon, *Biochem. J.*, 1971, **121**, 587.
[14] F. N. Boctor, *Analyt. Biochem.*, 1971, **43**, 66.
[15] N. Blumenkrantz and D. J. Prockop, *Analyt. Biochem.*, 1971, **39**, 59.

Many papers on the determination of thiol groups continue to appear. For example, a sensitive polarographic estimation of thiol groups using *N*-ethylmaleimide has been described[16] and 2-vinylquinoline has been suggested[17] as a reagent for the spectrophotometric determination of cysteine residues in proteins. A novel membrane filter assay for protein thiol groups has also been reported:[18] the adduct formed between ^{14}C-labelled mercuribenzoate and a protein that contains thiol groups is retained on a nitrocellulose filter whereas the superfluous label passes through. The minimum size of protein retained by the filter remains to be determined. The disulphide bridges of proteins can be reduced with dithiothreitol in liquid ammonia.[19] After reduction and alkylation the ammonia is removed by evaporation and the excess reagent and the by-products are removed by washing with methanol. The best alkyl halides to use are the chlorides, since iodides and bromides react with ammonia. A strange series of side-reactions between DTNB [5,5′-dithio-bis-(2-nitrobenzoate)] and the disulphide bridges of proteins has also been reported.[20] This involves reduction of the disulphide bridge by the nitrothiophenoxide anion. However, since the nitrothiophenoxide anion is stabilized by resonance, this reaction can best be described as unlikely, and the whole concept has now been refuted by other workers.[21]

Ion-exchange Chromatography. Titanous chloride has been recommended[22] as a substitute for stannous chloride in the reduction of the ninhydrin reagent for analysers since it causes no precipitation in the apparatus, and a simple device for cleaning the Teflon reaction coil in analysers has also been described.[23] It has been reported[24] that the condensation of ninhydrin with aldehydes and primary amines to yield highly fluorescent ternary products can be put to good use in automatic analysis, in which it is 10—100 times as sensitive as the conventional colorimetric procedure.

A detailed study has been made of accelerated amino-acid analysis using lithium citrate buffers on Aminex A5 resin.[25] Particular attention was paid to the effect of resin cross-linking and the avoidance of microbial contamination of the analyser columns. Aminex A5 resin has additionally been used for improving the separation of basic amino-acids and related compounds.[26] The column chromatographic analysis of tryptophan with the basic amino-acids, particularly useful for physiological fluids, has also been described.[27]

[16] P. D. J. Weitzman and H. J. Tyler, *Analyt. Biochem.*, 1971, **43**, 321.
[17] L. H. Krull, D. E. Gibbs, and M. Friedman, *Analyt. Biochem.*, 1971, **40**, 80.
[18] J. S. Krakow and S. P. Goolsby, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 453.
[19] J. Meizenhofer, J. Czombos, and H. Maeda, *J. Amer. Chem. Soc.*, 1971, **93**, 3080.
[20] J. F. Robyt, R. J. Ackerman, and C. G. Chittenden, *Arch. Biochem. Biophys.*, 1971, **147**, 262.
[21] K. Brocklehurst, M. Kierstan, and G. Little, *Biochem. J.*, 1972, **128**, 811.
[22] L. B. James, *J. Chromatog.*, 1971, **59**, 178.
[23] K. D. Hapner, *Analyt. Biochem.*, 1971, **43**, 613.
[24] K. Samejima, W. Dairman, J. Stone, and S. Udenfriend, *Analyt. Biochem.*, 1971, **42**, 237.
[25] G. E. Atkin and W. Ferdinand, *J. Chromatog.*, 1971, **62**, 373.
[26] A. Mondino, G. Bongiovanni, V. Nol, and I. Raffalle, *J. Chromatog.*, 1971, **63**, 411.
[27] B. J. Berridge, W. R. Chao, and J. H. Peters, *Analyt. Biochem.*, 1971, **41**, 256.

Single-column chromatography systems continue to attract attention. Using three lithium citrate buffers and an ammonia filtration system, it has been possible to design a single-column system that will adequately resolve most amino-acids.[28] A modified gradient elution procedure that resolves all the amino-acids from connective tissue proteins has also been described.[29]

Another modification to the analyser involves reading the absorbance of the effluent at 405 nm instead of the more normal 570 nm.[30] This enables proline and hydroxyproline to be quantitated in the same colorimeter channel as the other amino-acids and, moreover, the three colorimeters in conventional analysers can then be used to run three samples simultaneously. It is claimed that there is no loss in sensitivity. Volatile pyridine–formic acid buffers can replace the citrate buffers commonly used, which is particularly useful when one wishes to recover amino-acids from the analyser effluent.[31]

The quantitation of imino-acids can be improved by omitting hydrindantin from the ninhydrin reagent so as to favour the reaction with imino-acids,[32] and new conditions have been given for the rapid analysis of collagen hydrolysates,[33] for the determination of homocitrulline,[12b] and for the estimation of various methylated histidine compounds in muscle extracts.[34] Further studies on the chromatographic determination of D- and L-amino-acids in pneumococcal C-polysaccharide after their coupling with an L-amino-acid *N*-carboxy-anhydride have been reported [35] (see Volume 2 of these Reports) and a detailed analysis of the separation of oligopeptides and amino-acids on Chelex X-100 has been made.[36] The exchanger was used in the Cu^{2+} form and the amino-acids were eluted, free of Cu^{2+}, with ammonia.

High-voltage Electrophoresis and Thin-layer Chromatography. Because of their cheapness and simplicity, these methods continue to be very popular for separating amino-acids. It has been suggested [24] that the fluorescent ninhydrin technique referred to above for automatic analysers might find use in detecting amino-acids on paper and thin-layer chromatograms. A new empirical relation between the chromatographic mobility of a peptide on t.l.c. and its amino-acid composition has been derived [37] that is comparable with the original equation of Pardee.

A series of papers [38] has dealt with the analysis of amino-acids in the form of their dansyl derivatives and with the preparation of Dns-peptide maps.

[28] S. B. Melançon and J. Tayco, *J. Chromatog.*, 1971, **63**, 404.
[29] B. C. Starcher, L. Y. Wenger, and L. D. Johnson, *J. Chromatog.*, 1971, **54**, 425.
[30] J. P. Ellis, jun. and J. B. Garcia, jun., *J. Chromatog.*, 1971, **59**, 321.
[31] M. Redford-Ellis and M. N. Kelson, *J. Chromatog.*, 1971, **59**, 437.
[32] J. P. Ellis, jun. and J. M. Prescott, *J. Chromatog.*, 1971, **61**, 152.
[33] R. M. Osborne, R. W. Longton, and B. L. Lamberts, *Analyt. Biochem.*, 1971, **44**, 317.
[34] A. A. Christman, *Analyt. Biochem.*, 1971, **39**, 181.
[35] J. M. Manning, *J. Biol. Chem.*, 1971, **246**, 2926.
[36] J. Boisseau and P. Jouan, *J. Chromatog.*, 1971, **54**, 231.
[37] C. Haworth and R. W. A. Oliver, *Biochem. J.*, 1971, **124**, 255.
[38] V. A. Spivak, V. V. Shcherbukhin, V. M. Orlov, and J. M. Varshavsky, *Analyt. Biochem.*, 1971, **39**, 271; V. A. Spivak, V. A. Fedoseev, V. A. Orlov, and J. M. Varshavsky, *ibid.*, 1971, **44**, 12; V. A. Spivak, M. I. Levjant, S. P. Katrukha, and J. M. Varshavsky, *ibid.*, p. 503.

Gas–Liquid Chromatography. Further applications of a system for separating nanogram amounts of amino-acids as the *N*-trifluoroacetyl derivatives of their n-butyl esters have been described.[39] The same authors [40] have given details of the conversion of the amino-acids into their derivatives and have looked at the suitability of heptafluorobutyryl as an N-blocking group for higher sensitivity. They have also described the complete separation of the 20 amino-acids found in proteins, using two columns.[41] Other workers have now achieved the same complete separation on a single column.[42] The amino-acids are separated as their *N*-heptafluorobutyryl n-propyl ester derivatives in 43 min.

The conversion of amino-acids into their trimethylsilyl derivatives and their separation in this form have been described,[43] and it is has been reported [44] that methionine can readily be determined by estimating the methyl thiocyanate released by treatment with cyanogen bromide. Since methionine sulphoxide does not react with cyanogen bromide, one can determine the proportion of oxidized methionine in any sample.

B. End-group Analysis and Sequential Degradation.—The methods for identification of N-terminal amino-acids in peptides and proteins have been reviewed.[45]

It has been shown [46] that despite some previous reports to the contrary *Im*-Dnp-histidine does not decompose under conditions of acid hydrolysis. Fluorodinitrobenzene reacts preferentially with the 3′-nitrogen of N^{α}-acetylhistidine. Dnp-arginine and ε-Dnp-lysine have been separated using high-voltage paper electrophoresis at pH 11.2, and it is claimed that the same technique is applicable to the Dns-derivatives.[47] A technique for the determination of dinitrophenylamino-acids using chromatography on columns of nylon powder has also been described.[48] Two-dimensional separation of Dns-amino-acids has been effected by high-voltage paper electrophoresis and chromatography [49] and the identification of Dns-amino-acids in the mass spectrometer has been discussed.[50a] Pivaloyl chloride has been suggested as a reagent for N-terminal analysis of peptides and proteins and the characterization of N^{α}-pivaloyl amino-acid derivatives by mass spectrometry has been described.[50b]

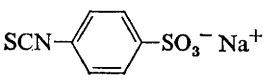
The N-terminal stepwise degradation due to Edman continues to be subject to various attempts to improve it. Thus for small peptides it has

[39] R. W. Zumwalt, K. Kuo, and C. W. Gehrke, *J. Chromatog.*, 1971, **55**, 267.
[40] R. W. Zumwalt, K. Kuo, and C. W. Gehrke, *J. Chromatog.*, 1971, **57**, 193.
[41] C. W. Gehrke, K. Kuo, and R. W. Zumwalt, *J. Chromatog.*, 1971, **57**, 209.
[42] C. W. Moss, M. A. Lambert, and F. J. Diaz, *J. Chromatog.*, 1971, **60**, 137.
[43] C. W. Gehrke and K. Leriner, *J. Chromatog.*, 1971, **57**, 219.
[44] G. M. Ellinger and R. H. Smith, *Biochem. J.*, 1971, **124**, 15P.
[45] J. Rosmus and Z. Deyl, *Chromatog. Rev.*, 1971, **13**, 163.
[46] P. Henkart, *J. Biol. Chem.*, 1971, **246**, 2711.
[47] H. M. Jacoby and L. Spero, *Analyt. Biochem.*, 1971, **44**, 299.
[48] H. Beyer and U. Schenk, *J. Chromatog.*, 1971, **61**, 263.
[49] G. L. Moore and R. S. Antonoff, *Analyt. Biochem.*, 1971, **39**, 260.
[50] (*a*) N. Seiler, H. H. Schneider, and K.-D. Sonnenberg, *Analyt. Biochem.*, 1971, **44**, 451; (*b*) G. Prota, F. Chioccara, and A. Previero, *Biochimie*, 1971, **53**, 51.

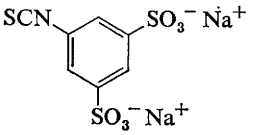
been proposed [51] that methyl isothiocyanate be substituted for phenyl isothiocyanate and that after the first step of degradation the methylthiohydantoin be identified in a sample of the reaction mixture by g.l.c. After the second step, analysis by g.l.c. of another sample of the reaction mixture reveals two methylthiohydantoins, that derived from the second residue in the peptide together with that previously identified as N-terminal. This process can be repeated, but obviously quantitative g.l.c. is required if the peptide contains two or more residues of the same amino-acid. An automatic method for the quantitation of methylthiohydantoins of acidic and neutral amino-acids following their separation on sulphopolystyrene resins has also been described.[52] Conditions have been found for the rapid separation of Pth-amino-acids using t.l.c. on polyamide-coated glass plates,[53] and the use of chemical ionization mass spectrometry for the analysis of Pth-amino-acids has also been described [54] (see also Section 2C). An aberration in the Edman degradation of peptides with N-terminal arginine or histidine residues has been reported [348] (see p. 72).

However, as in previous years, it is the automated Edman degradation procedure that shows the biggest growth in application. Examples of its use will be found throughout this Report, and comment here will be restricted to some reflections on the method. Several useful reviews will be found in an earlier reference.[5] A good typical application of the sequenator is to the determination of the N-terminal sequence of some myeloma light chains [55] (see also Section 7). It was observed that two methods of hydrolysis, with hydriodic acid and with sodium hydroxide–sodium sulphate, allowed direct conversion of the extracted thiazolinones into the corresponding amino-acids without the need for intermediate conversion into other derivatives. One of the problems associated with the use of the sequenator for the analysis of tryptic peptides containing lysine is the high repetitive loss of peptide in the various extractions during the last few steps, which explains why tryptic peptides with C-terminal arginine are favoured.[5] To circumvent this difficulty, it has been proposed that the first step of the degradation should be carried out with a sulphonated phenyl isothiocyanate.[56] With the ε-amino-group of the lysine residue safely modified by this hydrophilic reagent, the extraction losses are much diminished. Of the various reagents (1)—(4) that have been tested, (3) and (4) were reported to be the most effective.[5, 56] It is inevitable, but none the less a pity, that the commercial sequenators are so expensive. Few university laboratories can ever expect to possess one (see last year's Report).

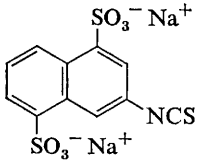
[51] D. E. Vance and D. A. Feingold, *Nature*, 1971, **229**, 121.
[52] V. M. Stepanov, S. P. Katrukha, L. A. Baratova, L. P. Belyanova, and V. P. Korzhenko, *Analyt. Biochem.*, 1971, **43**, 209.
[53] K. D. Kulbe, *Analyt. Biochem.*, 1971, **44**, 548.
[54] H. M. Fales, Y. Nagai, G. W. A. Milne, H. B. Brewer, jun., T. J. Bronzert, and J. J. Pisano, *Analyt. Biochem.*, 1971, **43**, 288.
[55] O. Smithies, D. Gibson, E. M. Fanning, R. M. Goodfliesh, J. G. Gilman, and D. L. Ballantyne, *Biochemistry*, 1971, **10**, 4912.
[56] G. Braunitzer, B. Schrank, A. Ruhfus, S. Petersen, and U. Petersen, *Z. physiol. Chem.*, 1971, **352**, 1730.



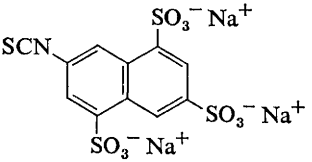
(1)



(2)

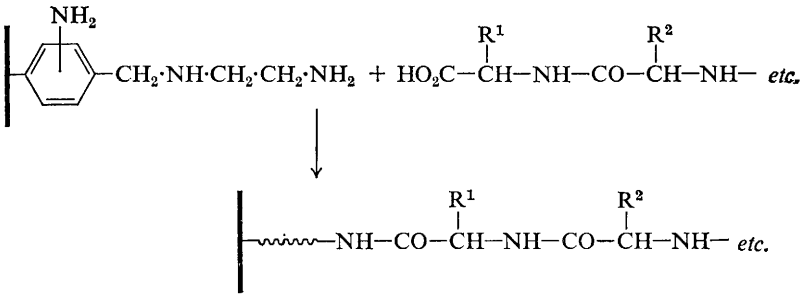


(3)



(4)

In a wholly different approach to the same problem, complete details of a solid-phase Edman degradation have now been given.[5, 57] The peptide is attached by means of its α-carboxy-group to a solid support (Scheme 1) and



Scheme 1

the usual Edman degradation is then carried out using methyl isothiocyanate. The addition of ring amino-groups to the polystyrene resin greatly increases its polarity, thereby enabling it to swell readily in trifluoroacetic acid, which in turn facilitates the cleavage step in the degradation. There are several obvious disadvantages to the method as it stands. For example, side-chain carboxy-groups should also become attached to the resin, which means that the corresponding amino-acid residue remains bound to the support, causing a gap to appear in the sequence. In fact, this does happen with glutamic acid but with aspartic acid residues the degradation actually stops altogether, presumably because a cyclic imide is formed after the activation of the β-carboxy-group in the coupling procedure. Several possible improvements remain to be tested,[57] for example the attachment of lysine-containing peptides through the ε-amino-group to resins containing isothiocyanate groups.[58] Even with the present technique, approximately 20

[57] R. A. Laursen, *European J. Biochem.*, 1971, **20**, 89.
[58] L. M. Dowling and G. Stark, *Biochemistry*, 1969, **8**, 4728.

cycles of degradation have successfully been carried out [57] on the A- and B-chains of insulin, with average yields per step of *ca.* 90%, and the method clearly bears watching.

Further work has been reported [59] on sequential degradation from the C-terminus (see Volume 2 of these Reports). The peptidyl-thiohydantoin is formed in a non-aqueous medium (a mixture of acetic anhydride, acetic acid, and sodium thiocyanate), is cleaved using the acidic form of a cation-exchange resin, and the thiohydantoin is then identified. Four steps of degradation were achieved with ribonuclease A. The C-terminal residue of rat liver lactate dehydrogenase has been identified as phenylalanine by the selective tritiation method [60] (reviewed in Volume 1 of these Reports), the labelled amino-acid being released by treatment with carboxypeptidase, and a stream-splitting device that enables the tritiated amino-acid to be counted in the effluent of the amino-acid analyser has been described.[61]

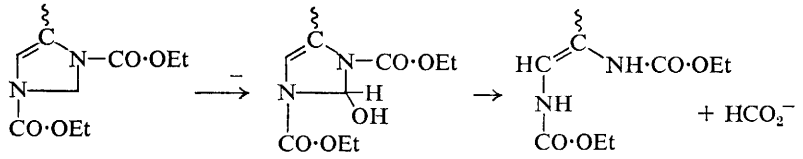
C. Mass Spectrometry.—Perhaps the main hindrance to the establishment of mass spectrometry as a standard tool for the determination of the sequence of purified peptides lies in the relatively large amounts that the technique demands for convenient operation. Whereas these pages last year conveyed the hope that 10—20 nmol would serve to give considerable sequence information, the literature suggests that much larger amounts are routinely being used. The main concrete advance, and one which gives mass spectometry the edge over 'wet' methods in this respect at least, has been in the direct analysis of peptide mixtures by low-resolution mass spectrometry. This also goes a long way to solving the supply problem, since 100 nmol of pure peptide is much harder to produce than mixtures containing comparable amounts. Two laboratories have reported the analysis of peptide mixtures [62, 63] and both have taken advantage of the differential volatility of the permethylated *N*-acetyl-peptide methyl esters. A steady temperature gradient is generated by fractional insertion of the probe into the ion source; when volatilization of a component begins (as indicated on an oscilloscope screen) the temperature gradient is rapidly increased and the spectrum recorded.[63] Mixtures of two or three synthetic peptides up to heptapeptides have been successfully resolved in this way, as well as a mixture of three tryptic peptides from pepsin.[63] Although a knowledge of the amino-acid composition is apparently not essential,[63] it has been suggested [62] that this is highly desirable for unambiguous identification; this would require a further 20 nmol or so. As an indication of growing confidence in mixture analysis, a general strategy has been formulated for investigating oligopeptides and proteins.[63] Briefly, an enzymic

[59] S. Yamashita, *Biochim. Biophys. Acta*, 1971, **229**, 301.
[60] W. T. Hsieh, L. E. Gundersen, and C. S. Vestling, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 69.
[61] G. Cappugi, P. Nassi, C. Treves, and G. Ramponi, *Experientia*, 1971, **27**, 237.
[62] P. Roepstorff, R. K. Spear, and K. Brunfeldt, *F.E.B.S. Letters*, 1971, **15**, 237.
[63] H. R. Morris, D. H. Williams, and R. P. Ambler, *Biochem. J.*, 1971, **125**, 189.

digest of the protein is subjected to gel filtration to obtain a mixture of peptides with molecular weight less than 1000 (*i.e. ca.* ten residues, the limit of the technique at present). The mixture would then be derivatized in the usual way and subjected to direct mass spectral analysis, using differential volatility as above. Larger fragments from the gel column would be digested with a different enzyme and treated similarly. Since peptide purification is the main stumbling block in conventional 'wet' methods of sequence determination, mixture analysis has the enormous advantage of saving labour and time, as well as being a partial solution to the problem of supply.

The other problem, that of 'difficult' amino-acids, is still real but continues to diminish. The hope that permethylation with strictly controlled amounts of methyl iodide might be successful for methionine peptides, as it was for cysteine peptides (see last year's Report), appears to have been borne out; the peptides Met-Met-Gly-Met and Thr-Met gave good electron-impact spectra [64] after permethylation on a microgram scale, using [65] a carbanion (methylsulphinylmethide) : MeI : peptide ratio of 10 : 10 : 1. Good spectra, containing *N*-terminal sequence peaks, have been obtained [66] for small arginine peptides, and for the relatively large bradykinin (a nonapeptide containing two arginines) after derivatization at the 100 nmol level with acetylacetone. This converts arginyl into *N*-2-(4,6-dimethyl)pyrimidylornithine residues; the peptide is then *N*-acetylated and permethylated in the usual way. Desulphurization of sulphur-containing amino-acids is still being explored;[67] however, mass spectra of a series of *S*-carboxymethylated and *S*-carboxamidomethylated cysteine- and (reduced) cystine-peptides have also been described.[68] No further details have been published since last year [69] on the use of diethyl pyrocarbonate to avoid problems of quaternization during the permethylation of histidine-containing peptides. Under the conditions of treatment with diethyl pyrocarbonate (pH 8), the α-amino-group and the side-chains of tyrosine, cysteine, and histidine (but not arginine) are modified. Since the substituted imidazole ring will still quaternize on permethylation, exhaustive treatment with diethyl pyrocarbonate is carried out. This cleaves the ring, eliminating C-2 as formate and converting the remainder into a 1,2-bis(ethoxycarbonamido)ethylene derivative (Scheme 2). This does not form quaternary salts and is extracted satisfactorily into chloroform after permethylation. The method worked successfully on a series of histidine peptides up to a pentapeptide;[69] this contained arginine, which was converted into ornithine by hydrazinolysis

[64] P. A. Leclercq and D. M. Desiderio, jun., *Biochem. Biophys. Res. Comm.*, 1971, **45**, 308.
[65] P. A. Leclercq and D. M. Desiderio, jun., *Analyt. Letters*, 1971, **4**, 305.
[66] P. A. Leclercq, L. C. Smith, and D. M. Desiderio, jun., *Biochem. Biophys. Res. Comm.*, 1971, **45**, 937.
[67] Yu. A. Ovchinnikov, A. A. Kiryushkin, V. A. Gorlenko, and B. V. Rozynov, *Zhur. obshchei Khim.*, 1971, **41**, 660.
[68] Yu. A. Ovchinnikov, A. A. Kiryushkin, V. A. Gorlenko, Ts. E. Agadzhanyan, and B. V. Rozynov, *Zhur. obshchei Khim.*, 1971, **41**, 385 (*Chem. Abs.*, 1971, **75**, 20 992b).
[69] J. F. G. Vliengenthart and L. Dorland, *Biochem. J.*, 1970, **117**, 31P.



Scheme 2

before mass spectrometry. It is possible, however, that derivatization of histidine may be unnecessary if permethylation is carried out with controlled amounts [65] of carbanion and methyl iodide; Glp-Pro-Tyr-His-NH_2 gave a good spectrum.[70] So it appears that the presence of arginine, histidine, methionine, and cysteine residues in peptides does not preclude determination of their sequences by mass spectrometry, but one must, of course, be aware of their presence. Does, then, a routine application of mass spectrometry to the analysis of mixtures, for example, call for a succession of derivatization treatments? A better suggestion,[63] perhaps, is that a preliminary paper-electrophoretic step be included so that cysteine, arginine, histidine, *etc.* could be detected by staining and the sample treated accordingly. In dealing with mixtures (as above) sufficient material would probably be available and the electrophoretic mobilities would also be useful; even dealing with more precious pure peptides, the advantage of being able to circumvent trouble at a later stage would make an amino-acid analysis of the sample highly desirable. Application of the Edman degradation to remove troublesome amino-acids from peptides for mass spectrometry [71] is useful in principle, but with the reservation that it calls for some prior knowledge of the sequence. If this is used on C-terminal lysine peptides a hydrazinolysis step must be included before the mass spectrum is run, so that the ε-PTH derivative (involatile) is liberated as the free amine.[71] Mass spectra of the chromophoric *N*-(azulen-4-yl)acetyl derivatives of amino-acid and peptide methyl esters have been described.[72] These showed strong molecular ions, low fragmentation in the high-mass region, and good sequence peaks for the series of N-terminal proline peptide derivatives studied.

While efforts to deal successfully with mixtures and 'difficult' amino-acids continue, mass spectrometry is being used in peptide chemistry where suitably straightforward problems arise. Thus it was used [73] in the identification of small tryptic peptides from soluble tropoelastin (considered to be the precursor of insoluble elastin). The peptides Ala-Ala-Ala-Lys and Ala-Ala-Lys, identified after permethylation of samples N-acetylated with a mixture of acetic anhydride and its perdeuterio-analogue, apparently occur

[70] P. A. White and D. M. Desiderio, jun., *Analyt. Letters*, 1971, **4**, 141.
[71] N. A. Aldanova, E. I. Vinogradova, S. A. Kazaryan, B. V. Rozynov, and M. M. Shemyakin, *Biochemistry* (*U.S.S.R.*), 1970, **35**, 742.
[72] E. Wünsch and E. Jaeger, *Z. physiol. Chem.*, 1971, **352**, 1584.
[73] L. B. Sandberg, N. Weissman, and W. R. Gray, *Biochemistry*, 1971, **10**, 52.

six times in the tropoelastin polypeptide chain, and are thought to be the areas involved in the formation of the desmosine and isodesmosine cross-links of insoluble elastin (see the section on Structural Proteins in this, and earlier, Reports). Mass spectrometry played an important part [74] in defining the sequence of the hypothalamic thyroid-stimulating-hormone releasing factor as Glp-His-Pro-NH_2 which, with both ends blocked, would present something of a problem for conventional methods of sequence analysis. In another study of this peptide a normal ('peptide') fragmentation was not observed;[75] however, the sample had not been derivatized, and the high temperature used gave some thermal degradation. Whether or not one believes that scotophobin really is a 'specific behaviour-inducing brain peptide', its structure was revealed [76] by mass spectrometry as Ser-Asp-Asn-Asn-Gln-Gln-Gly-Lys-Ser-Ala-Gln-Gln-Gly-Gly-Tyr-NH_2. Spectra of the whole molecule and of the two tryptic peptides arising from cleavage at the conveniently placed lysine residue were obtained after diazomethane treatment, the only derivatization. The advantages of permethylation again become apparent when we read [76] that the sequence was reconstructed from the spectra of di- and tri-peptides resulting from extensive pyrolysis! That (presumably) also made certain amide assignments impossible, whereas this is now recognized as one of the benefits of sequence analysis by mass spectrometry.

In last year's Report, some of the advantages of chemical ionization mass spectrometry over the electron-impact method were outlined, with reference to the spectra of very simple peptides. Chemical ionization spectra of a series of blocked, permethylated peptides up to pentapeptides, and containing most of the amino-acids or suitable derivatives, have now been examined.[77] Two series of ions were recognized and their presence in the spectra was lucidly discussed.[77] When C—N cleavage occurs, retention of the positive charge on the carbon atom gives acyl carbonium ions, *i.e.* the N-terminal sequence-determining ions familiar from electron-impact spectra. Chemical ionization spectra show also the complementary ammonium ions (*i.e.* peaks that determine the C-terminal sequence), and from these data additional sequence information can be deduced and ambiguities can be minimized. In addition, the spectra are cleaner, with less evidence of C—C bond cleavage, and good results were obtained on 20 μg (*ca.* 30 nmol of a hexapeptide). It is suggested [77] that chemical ionization mass spectrometry will complement rather than supersede the electron-impact method for sequence analysis of peptides, despite the attractive bonus of ions determining the C-terminal sequence, because N-terminal sequence ions are often more intense in the electron-impact spectra.

[74] D. M. Desiderio, jun., R. Burgus, T. F. Dunn, W. Vale, R. Guillemin, and D. N. Ward *Org. Mass Spectrometry*, 1971, **5**, 221.
[75] J.-K. Chang, H. Sievertsson, C. Bogentoft, B. Currie, K. Folkers, and G. D. Daves, *J. Medicin. Chem.*, 1971, **14**, 481.
[76] D. M. Desiderio, jun., G. Ungar, and P. A. White, *Chem. Comm.*, 1971, 432.
[77] A. A. Kiryushkin, H. M. Fales, T. Axenrod, E. J. Gilbert, and G. W. A. Milne, *Org. Mass Spectrometry*, 1971, **5**, 19.

Mass spectrometry continues to be recommended for the identification of thiohydantoin derivatives of amino-acids from the Edman degradation of proteins. Chemical ionization spectra of phenylthiohydantoin derivatives [54] gave a clearly defined (M + 1) ion for all amino-acids except arginine, lysine, and *S*-carboxymethylcysteine, but since electron-impact spectra of Pth derivatives also show good parent peaks the main advantage of the chemical ionization method here lies in the basic similarity of the spectra, and in much greater sensitivity. Isotope dilution, with labelled Pth-amino-acids, was used in the identification of the Pth derivative released (*cf.* last year's Report, and the identification of methylthiohydantoins). The method was demonstrated on sperm whale myoglobin and compared with the identification of the thiohydantoins by g.l.c. Automation will truly have arrived when thiazolinones from the sequenator are fed direct into the mass spectrometer![78] The electron-impact spectra of thiazolinones are very similar to those of the corresponding thiohydantoins, presumably as a result of thermal rearrangement in the mass spectrometer; the first five residues of ribonuclease were successfully identified. Direct analysis of thiazolinones holds the advantage of minimal decomposition of those amino-acids (*e.g.* serine) which are sensitive to the acid conditions used for conversion of thiazolinone into thiohydantoin.[78] In another study [79] *p*-bromophenyl isothiocyanate was reported to lead to easier recognition of thiohydantoin peaks within the spectrum using the strong doublets arising from the natural isotope abundance of bromine. The method was illustrated on an uncomplicated octapeptide and the N-terminal sequence of glucagon.

It will be apparent that computer analysis of high-resolution spectra has not featured to any great extent in the determination of peptide sequences during the past twelve months. Those who work at low resolution, and without computers, have outlined [63] what they consider to be unnecessary disadvantages inherent in the former approach, namely decreased sensitivity, increased cost, and the complication of additional black boxes. It does seem that more answers are coming from the low-resolution approach at the moment, but one should perhaps bear in mind certain advantages of computer-aided high-resolution mass spectrometry, *e.g.* objective identification of the sequence ions when these are not the strongest in the spectrum (even for permethylated samples), and a vast amount of information which might be useful in minimizing ambiguities. The differential volatility approach could, of course, be applied to mixtures, as described earlier for low-resolution work. Whether or not one feels that the requirements of sequence analysis have already been met by low-resolution mass spectrometry seems at the moment to be as much a subjective judgement as an objective one. A new computer analysis of high-resolution mass spectra applicable to peptides has been described;[80] termed 'submolecular group

[78] T. Fairwell and R. E. Lovins, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 1280.
[79] F. Weygand and R. Obermeier, *European J. Biochem.*, 1971, **20**, 72.
[80] A. Kunderd, R. B. Spencer, and W. L. Budde, *Analyt. Chem.*, 1971, **43**, 1086.

analysis', it permits identification of ions other than sequence ions. It is likely that aspects of the application of mass spectrometry to peptides will feature in a companion series of Reports, the first of which has already appeared [81] and the second of which is currently being prepared.

D. Cleavage of Protein Chains.—*Enzymic Cleavage.* For many years protein chemists have been bothered by the chymotrypsin-like activity of pure, crystalline trypsin, which cannot be removed even by treatment with active-site-directed inhibitors of chymotrypsin. Several recent papers bear on this problem. It has been reported [82] that gel filtration on Sephadex-G50 can remove the chymotryptic activity. Other experiments [83] have shown α- and β-trypsin purified by chromatography on SE-Sephadex to possess only 'pure' tryptic action. β-Trypsin is the single polypeptide chain form of the enzyme, whereas α-trypsin has a single peptide bond split at position Lys(131)–Ser(132). Moreover, it appears [84] that ψ-trypsin, which differs from α-trypsin in having an additional bond split at Lys(176)–Asp(177), is what confers the chymotrypsin-like activity on unfractionated preparations of crystalline trypsin, although ψ-trypsin clearly differs from chymotrypsin in that it will not cleave some peptides rapidly hydrolysed by chymotrypsin. It is especially interesting that the residue Asp-177 is believed to be of prime importance in defining the specificity of trypsin.

The action of thrombin on fibrinogen is, of course, that of a very restricted trypsin. Various peptides related to fibrinopeptides have been synthesized and tested as substrates for thrombin and trypsin.[85] None is as good as fibrinogen, suggesting that factors other than the primary structure in the immediate vicinity of the susceptible bond contribute to the interaction between fibrinogen and enzyme. Comparable experiments in another laboratory [86] show that denaturation of fibrinogen reduces the susceptibility to thrombin but enhances the susceptibility to trypsin.

Thermolysin continues to arouse interest. A study of the effect of pH on its activity with synthetic substrates indicates that groups of pK 5.15 and 8 are implicated in its action,[87] and Ca^{2+} ions have been shown [88] to affect the kinetic parameters of the enzyme and limit the autodigestion. Since a number of dipeptides containing *S*-alkylcysteine residues have been observed to be substrates for thermolysin, it is possible that useful additional cleavages for sequence work on proteins might occur at such residues.[89]

[81] 'Mass Spectrometry', ed. D. H. Williams (Specialist Periodical Reports), The Chemical Society, London, 1971, Vol. 1.
[82] J. Jentsch, *J. Chromatog.*, 1971, **57**, 450.
[83] V. Keil-Dlouhá, N. Zylber, N.-T. Tong, and B. Keil, *F.E.B.S. Letters*, 1971, **16**, 287.
[84] V. Keil-Dlouhá, N. Zylber, J. M. Imhoff, N.-T. Tong, and B. Keil, *F.E.B.S. Letters*, 1971, **16**, 291.
[85] R. K. H. Lieu, R. H. Andreatta, and H. A. Scheraga, *Arch. Biochem. Biophys.*, 1971, **147**, 201.
[86] Y. Inada, M. Bando, I. Kotoku, A. Matsushima, and J. Hirano, *Biochim. Biophys. Acta*, 1971, **251**, 94.
[87] C. E. Stauffer, *Arch. Biochem. Biophys.*, 1971, **147**, 568.
[88] H. Drucker and S. L. Borchers, *Arch. Biochem. Biophys.*, 1971, **147**, 242.
[89] A. P. Damoglou, H. Lindley, and I. W. Stapleton, *Biochem. J.*, 1971, **123**, 379.

A number of new enzymes have been reported. An aminopeptidase has been isolated from *Aeromonas proteolytica* [90] and a carboxypeptidase C from orange leaves.[91] The carboxypeptidase shows the combined action of carboxypeptidases A and B. A dipeptidase (a metalloenzyme) that will cleave only L-α-dipeptides with free amino- and carboxy-groups has been purified from mouse ascites tumour cells.[92] Two new proteases have also been described. One, a protease from *E. coli*, has a molecular weight of 43 000 and is inhibited by DFP;[93] the other, from *Acremonium kiliense*,[94] has a molecular weight of *ca.* 28 000 and a pH optimum of 10.5. Moreover, it is fully active even after prolonged exposure to 8M-urea. Other enzymes of interest are the peptidoglutaminases from *B. circulans*.[95] These enzymes show no peptidase activity but will selectively hydrolyse the γ-amide of peptide-bound glutamine.

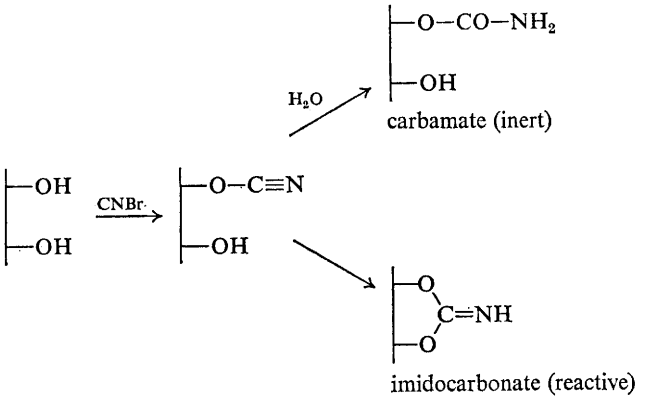
Several studies of neutral proteases from micro-organisms have also been described. The specificity of such enzymes seems to be affected by the sequence of at least five or six residues near the bond actually split in the substrate,[96] a property shared by various serine proteinases of micro-organisms.[97] Peptide bonds involving the N-terminal groups of phenylalanine, leucine, or alanine are rapidly split by the neutral protease of *Micrococcus caseolyticus*, similar to that of *Pseudomonas aeruginosa* and *Streptomyces griseus*,[98] whereas the endopeptidase from the marine bacterium *Vibrio* B-30 shows specificity for the N-terminal side of aromatic amino-acids.[99]

Restriction of Enzymic Cleavage. This has been discussed extensively in previous Reports. Some recent developments in the reversible blocking of protein amino-groups to restrict tryptic cleavage have been reviewed [100] and a comparative study has been made of several reversible reagents for that purpose.[101] 2-Methylmaleic (citraconic) anhydride was held to be the most suitable for most purposes.

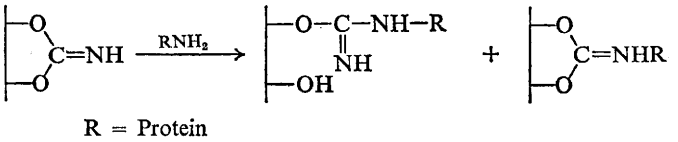
Insolubilized Enzymes. A comprehensive review [102] discusses many aspects of insolubilization of enzymes and its effect on enzymic parameters, and

[90] J. M. Prescott, S. H. Wilkes, F. W. Wagner, and K. J. Wilson, *J. Biol. Chem.*, 1971, **246**, 1756.
[91] B. Sprössler, H.-D. Heilmann, E. Grampp, and H. Uhlig, *Z. physiol. Chem.*, 1971, **352**, 1524.
[92] S. Hayman and E. K. Patterson, *J. Biol. Chem.*, 1971, **246**, 660.
[93] M. Pacaud and J. Uriel, *European J. Biochem.*, 1971, **23**, 435.
[94] S. van Heyningen and D. Secher, *Biochem. J.*, 1971, **125**, 1159.
[95] M. Kikuchi, H. Hayashida, E. Nakano, and K. Sakaguchi, *Biochemistry*, 1971, **10**,. 1222.
[96] K. Morihara and H. Tsweuki, *Arch. Biochem. Biophys.*, 1971, **146**, 291.
[97] K. Morihara, T. Oka, and H. Tsuzuki, *Arch. Biochem. Biophys.*, 1971, **146**, 297.
[98] M. J. Desmazeaud and J. H. Hermier, *European J. Biochem.*, 1971, **19**, 51.
[99] T. Sipos and J. R. Merkel, *Arch. Biochem. Biophys.*, 1971, **145**, 137.
[100] R. N. Perham in Ref. 5, p. 57.
[101] (*a*) A. F. S. A. Habeeb and M. Z. Atassi, *Biochemistry*, 1970, **9**, 4939; (*b*) R. P. Singhal and M. Z. Atassi, *ibid.*, 1971, **10**, 1757.
[102] E. Katchalski, I. Silman, and R. Goldman, *Adv. Enzymology*, 1971, **34**, 445.

another volume [103] contains several articles of interest. A theoretical model describing steady-state catalysis by enzymes immobilized in gel particles was verified by experiment.[104] Many enzymes have previously been attached to cross-linked dextrans activated by treatment with cyanogen bromide; as a further example, tRNA nucleotidyl transferase was insolubilized in this way and shown to be active in the repair of the $_pC_pC_pA$ sequence of tRNA.[105] The chemistry of the attachment has now been described in some detail.[106] The activation of the dextran and the attachment of the protein are thought to take place as shown in Schemes 3 and 4 respectively. A new



Scheme 3



R = Protein

Scheme 4

method [107] for the attachment of enzymes to polymers bearing various functional groups uses isocyanides for the coupling step. The chemistry of the process is described [107] and it is worth noting that it permits coupling through a variety of alternative side-chains on the enzyme (ligand) (*cf.* the use of variously modified agaroses, described in last year's Report, to introduce flexibility in the choice of groups used for the coupling). The isocyanide method was used successfully [108] for the attachment of pepsin to

[103] (*a*) L. Goldstein, *Methods Enzymology*, 1970, **19**, 935; (*b*) E. M. Crook, K. Brocklehurst, and C. W. Wharton, *ibid.*, p. 963; (*c*) B. Alexander and A. M. Engel, *ibid.*, p. 978.
[104] V. Kasche, H. Lundqvist, R. Bergman, and R. Axén, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 615.
[105] S. Litvak, L. Tarrago-Litvak, D. S. Carre, and F. Chapeville, *European J. Biochem.*, 1971, **24**, 249.
[106] R. Axén and S. Ernback, *European J. Biochem.*, 1971, **18**, 351.
[107] R. Axén, P. Vretblad, and J. Porath, *Acta Chem. Scand.*, 1971, **25**, 1129.
[108] P. Vretblad and R. Axén, *F.E.B.S. Letters*, 1971, **18**, 254.

modified agarose. Hyaluronidase attached to agarose by the triazine method [109] could prove useful in structural work on proteoglycans.

Other supports and other methods of insolubilization have also been investigated. Thus, Pronase has been immobilized on porous glass [110] and shown to be useful in the total hydrolysis of proteins when used in conjunction with immobilized leucine aminopeptidase. Acid phosphatase, trypsin, chymotrypsin, and glucose oxidase have been attached to polyacrylamide beads using glutaraldehyde [111] and trypsin, chymotrypsin, and papain have been linked to cross-linked poly(methacrylic acid anhydride),[112] with substantial retention of activity (towards small substrates) in all cases. Papain has also been converted into a giant polymer by cross-linking it with glutaraldehyde:[113] in such a form it will attack only the surface of ox-casein micelles. When chymotrypsin was insolubilized similarly, retention of activity was greater if the glutaraldehyde treatment was carried out in the presence of an inhibitor of the enzyme.[114] Co-insolubilized α-chymotrypsin–papain had both enzymatic activities.[114] In another study [115] glutaraldehyde-insolubilized trypsin and chymotrypsin were used for selective adsorption of the naturally occurring protein inhibitors and *vice versa*; insolubilized concanavalin A was used to isolate glycoproteins and insoluble glycoproteins to purify agglutinins.

The kinetic behaviour of glucose oxidase attached to porous glass has been discussed [116] and the attachment of invertase to bentonite [117] and of amino-acylase to halogenoacetyl-celluloses [118] has been reported.

E. Fractionation Methods.—Discussion is restricted here to fractionation of peptides and proteins; procedures for amino-acids have already been dealt with in Section 2A. Chromatographic and electrophoretic methods are treated in some detail below. It is also worth noting a study of the fractionation of proteins and viruses with poly(ethylene glycol).[119] The effect acts quantitatively like 'salting out', and there is a linear relationship between log solubility and % poly(ethylene glycol). The selectivity of the method is greater for larger proteins and viruses. A useful volume [120] devoted to enzyme purification carries many useful accounts of specialized techniques.

Peptide Separation, Detection, and Identification. Adsorption on to a neutral polystyrene resin (Porapak Q) has been recommended [121] as a

[109] W. H. Stimson and A. Serafini-Fracassini, *F.E.B.S. Letters*, 1971, **17**, 318.
[110] G. P. Royer and G. M. Green, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 426.
[111] P. D. Weston and S. Avrameas, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1574.
[112] A. Conte and K. Lehmann, *Z. physiol. Chem.*, 1971, **352**, 533.
[113] S. H. Ashoor, R. A. Sair, N. F. Olson, and T. Richardson, *Biochim. Biophys. Acta*, 1971, **229**, 423.
[114] E. F. Jansen, V. Tomimatsu, and A. C. Olson, *Arch. Biochem. Biophys.*, 1971, **144**, 394.
[115] S. Avrameas and B. Guilbert, *Biochimie*, 1971, **53**, 603.
[116] M. K. Weibel and H. J. Bright, *Biochem. J.*, 1971, **124**, 801.
[117] P. Monsan and G. Durand, *F.E.B.S. Letters*, 1971, **16**, 39.
[118] T. Sato, T. Mori, T. Tosa, and I. Chibata, *Arch. Biochem. Biophys.*, 1971, **147**, 788.
[119] I. R. M. Jackes, *Biochim. Biophys. Acta*, 1971, **229**, 535.
[120] *Methods Enzymology*, 1971, Vol. 22.
[121] A. Niederwieser, *J. Chromatog.*, 1971, **61**, 81.

method for rapid desalting and fractionation of non-polar amino-acids and peptides. Adsorption increased with increasing ionic strength, demonstrating the hydrophobic nature of the interaction, and elution was performed with water or aqueous-organic mixtures. Such adsorption of amino-acids on to polystyrene is presumably superimposed upon the ion-exchange effect in the separation of amino-acids on sulphonated polystyrene resin in the standard Moore and Stein method, and probably explains why the chromatographic behaviour of several neutral amino-acids does not correlate with their p*K* or pI; hydrophilic and basic amino-acids chromatograph normally. Preliminary work on peptides gave every indication that the effect would be useful for desalting peptides containing non-polar residues.

A theoretical study [122] of the behaviour of amino-acids and peptides during paper electrophoresis suggests that with a knowledge of p*K* values and molecular weights the optimum conditions for separation can be read off from standard graphs. One supposes that visual dye markers are routinely used by the exponents of paper electrophoresis; the mobilities of several dyes at commonly used pH values have been documented [123] and the dyes of choice at each pH indicated. Conditions conducive to the deamidation of side-chains, a recognized hazard in peptide and protein purification, have been reported [124] for two asparaginyl peptides; borate rather than phosphate buffer, low ionic strengths, a pH near 5, and temperatures less than 4 °C should all minimize deamidation.

This year's candidate for the 'diagonal' spot (see Vol. 1, p. 61) is a method for the selective isolation of histidine-containing peptides.[125] Briefly, the procedure is to dinitrophenylate histidine side-chains after the reaction of the amino-groups with citraconic anhydride; the modified protein is then digested with pepsin in 10% formic acid, with concomitant regeneration of the amino-groups. The first electrophoretic step is then carried out at pH 3.5 or 6.5, and the second at right angles after thiolysis of the Dnp residues on the paper by exposure to 2-mercaptoethanol. Histidine peptides will lie off the diagonal in the direction of the cathode. Disulphide bonds must be modified (*e.g.* by performic acid oxidation) at some point before the thiolytic step, preferably at the very beginning. A pH of 3.5 was preferred for the electrophoretic steps since this gave a better resolution of *Im*-Dnp-histidyl peptides in the first dimension and hence purer off-diagonal peptides. Disulphide exchange with cystamine has again been suggested [126] as the basis of a method for the selective isolation of cystinyl peptides; it is also noted that the *S*-(2-aminoethylthio)cysteinyl bond, although similar to the *S*-aminoethylcysteinyl bond, may or may not be cleaved by trypsin.

[122] Y. Kiso and E. Falk, *J. Chromatog.*, 1971, **59**, 401.
[123] K. J. Stevenson, *Analyt. Biochem.*, 1971, **40**, 29.
[124] J. H. McKerrow and A. B. Robinson, *Analyt. Biochem.*, 1971, **42**, 565.
[125] W. H. Cruickshank, T. M. Radhakrishnan, and H. Kaplan, *Canad. J. Biochem.*, 1971, **49**, 1225.
[126] D. J. Cox and J. R. Brown, *Analyt. Biochem.*, 1971, **41**, 139.

A useful trick has been applied to the separation of cysteine and glutathione derivatives;[127] cysteic acid and glutathione, which had the same elution time on Dowex-1, were well resolved on a Dowex-1–Sephadex G-10 mixed-bed column. This is worth bearing in mind in other cases where molecules have similar chromatographic properties but different molecular weights. Satisfactory fingerprints of large polypeptide chains (in this case of m. wt. 80 000—90 000) were obtained by two-dimensional separation of the dansylated digest on silica-gel plates.[128] Even better results were obtained, however, by passing the dansylated mixture through Sephadex G-25 and running a series of plates in parallel on successive cuts from the column. In this way 64—68 of the expected 77 tryptic peptides were resolved, and very little material was required.

It appears that an improvement on the usual starch–iodide method for detecting imino-group-containing compounds on t.l.c. can be achieved by including starch in the silica-gel plates from the beginning;[129] after chromatography only potassium iodide spraying is necessary, and this is done more evenly from 75% aqueous acetone. N-terminal tryptophyl dipeptides (and Trp itself) give intense fluorescence on silica-gel t.l.c. after exposure to gaseous formaldehyde; as little as 0.3—0.1 μg can be detected.[130]

Chromatography. A unified theory proposed earlier for molecular-sieve chromatography and electrophoresis in polyacrylamide gels is borne out in practice.[131] Chromatography on Sephadex ion-exchangers has been discussed[132] and illustrated with fractionation of a tryptic digest of immunoglobulin κ-chains on SE-Sephadex C-25. A simple rule allows the calculation of the optimum column volume and gradient in any particular case. It is noted that resolution is improved by using a shallower gradient only if proteins are eluted from the ion-exchange column at high concentration, and that often a steeper gradient may be better. A word of warning is issued[133] to those who use ribonuclease A to calibrate Sephadex columns for molecular weight determinations; elution of the expected peak from G-100 was followed by further ribonuclease activity several void volumes later, and loading with other proteins displaced still more activity from the gel. This may be ribonuclease itself or an active fragment, and buffers of high ionic strength are recommended for use with commercial calibration kits to minimize interactions between protein and gel. A word of warning of a different sort concerns the degradation of Sephadex (in this case G-150) by 50% formic and acetic acids.[134]

As little as 100 μg of a protein is sufficient for determination of its Stokes'

[127] J. W. Purdie and D. E. Hanagi, *J. Chromatog.*, 1971, **59**, 181.
[128] B. Kremer and J. Ullrich, *Z. physiol. Chem.*, 1971, **352**, 189.
[129] S. D. Killilea and P. O'Carra, *J. Chromatog.*, 1971, **54**, 284.
[130] R. Håkanson and F. Sundler, *J. Chromatog.*, 1971, **59**, 209.
[131] C. J. O. R. Morris and P. Morris, *Biochem. J.*, 1971, **124**, 517.
[132] J. Novotný, *F.E.B.S. Letters*, 1971, **14**, 7.
[133] W. E. Mitch and C. C. Levy, *Biochim. Biophys. Acta*, 1971, **251**, 388.
[134] R. Fairweather and J. H. Jones, *J. Chromatog.*, 1971, **58**, 285.

radius (in only 8 h) if thin-layer gel-filtration is used.[135] For staining, the protein is transferred to a sheet of moist filter paper pressed gently on to the gel layer. The subunit molecular weight also can be determined rapidly by t.l.c. if Sephadex swollen in 6M-guanidine hydrochloride (but not urea) is used.[136] The method requires 10—30 μg of protein and takes about 3 h; cytochrome *c* is run as an internal marker and the usual relationship between mobility and log m. wt. holds. It is difficult to see this method replacing SDS-gels for the determination of subunit molecular weight, but it could well be a more attractive method for determining the chain molecular weight in guanidine hydrochloride than the running of columns in this solvent. The column method is being used successfully for proteins and was recently shown to be applicable also to peptides. Agarose columns run in 6M-guanidine hydrochloride gave[137] the molecular weights of the three cyanogen bromide fragments of apoferritin (in the range 1400—8000) with an accuracy better than 10%. It is worth noting that for molecular weights less than 9000 Sepharose gave much better resolution than the BioGel equivalent, whereas there was very little difference in the range 16 000—80 000. With the increasing use of guanidine hydrochloride as a denaturant in protein chemistry, a statement of the criteria of purity, by Tanford and co-workers, is most welcome.[138] The purity of urea solutions has also received attention,[139a] and the kinetics of cyanate formation were used to plot graphs of the concentration of cyanate in urea solutions as a function of time at different pH values, for periods up to two months, and at 0 and 25 °C. The effect of cyanate on the lens protein α-crystallin in concentrated urea was clearly illustrated by the appearance of additional bands on polyacrylamide gels.[139b] Cyanate formation in urea solutions is, however, slow at low temperatures and therefore not a real problem over short periods; it occurs rapidly at a pH higher than 4. The interaction of urea with proteins has now been studied[140] by means of ultrafiltration; one urea molecule binds per three amino-acid residues, suggesting that the peptide bond is again the major site of interaction (*cf.* the binding of SDS to proteins; see last year's Report).

Electrophoresis. This year again this section deals very largely with electrophoresis in polyacrylamide gels, with or without denaturants. Some consideration is also given to the increasingly used technique of isoelectric focusing.

[135] Z. Wasyl, E. Luchter, and W. Bielański, jun., *Biochim. Biophys. Acta*, 1971, **243**, 11.
[136] F. Heinz and W. Prosch, *Analyt. Biochem.*, 1971, **40**, 327.
[137] C. F. A. Bryce and R. R. Crichton, *J. Chromatog.*, 1971, **63**, 267.
[138] K.-P. Wong, R. Roxby, and C. Tanford, *Analyt. Biochem.*, 1971, **40**, 459.
[139] (*a*) P. Hagel, J. J. T. Gerding, W. Fieggen, and H. Bloemendal, *Biochim. Biophys. Acta*, 1971, **243**, 366; (*b*) J. J. T. Gerding, A. Koppers, P. Hagel, and H. Bloemendal, *ibid.*, p. 374.
[140] J. R. Warren and J. A. Gordon, *Biochim. Biophys. Acta*, 1971, **229**, 216.

A useful survey of the theory and practice of polyacrylamide gel electrophoresis, both analytical and preparative, has appeared [141] with particular attention to the quantitative physical-chemical aspects. A theoretical treatment of chromatography and electrophoresis in polyacrylamide gels has already been mentioned.[131] A further series of papers [142] deals quantitatively with the behaviour of proteins in polyacrylamide gels. The practical aspects of preparative acrylamide gel electrophoresis (both continuous and disc techniques),[143a] and of analytical gel electrophoresis [143b] (the latter with emphasis on problems related to the separation of enzymes) have recently been treated lucidly. A two-dimensional technique has been used to determine the isoelectric points of the proteins of *E. coli* ribosomes without preliminary fractionation.[144] Tryptophan-containing proteins can be located in polyacrylamide gels by fluorimetry;[145] tyrosine and phenylalanine fluorescence alone is not sufficient. The location of dansylated proteins in gels by fluorimetry is mentioned below. The only comments on starch gels are new methods suggested for visualizing proteins. First comes a procedure [146] for the rapid staining (about 5 min) of peptides and proteins by chlorination; thin slices of gel are treated for 1 min with hypochlorite, and the excess is then removed with a hydrazinium sulphate wash for 2—3 min, before staining with aqueous potassium iodide. Other methods that have been suggested [147] involve the immersion of gel slices in copper sulphate solution (photographs should be taken soon afterwards); the inclusion of fluorescein in the gel so that protein-containing regions will quench fluorescence when observed under u.v. light; and finally, dansylation of the gel in acetone to give fluorescent protein bands.

A discontinuous (sulphate–borate) buffer system for SDS-gel electrophoresis has been thoroughly investigated,[148] and the relationship between relative mobility and molecular weight of protein–SDS complexes has been explored by calculating retardation coefficients and free mobilities from gels of different concentrations. It is becoming clear that the intrinsic charge on a protein does determine the amount of SDS it binds, so that migration in SDS-gels may not be determined entirely by size. Maleylation of protein amino-groups caused a disproportionate increase in the apparent molecular weight, an effect that was attributed to decreased binding of SDS,[149] and the abnormal mobilities of two acidic proteins were shown to be normalized by esterification.[150] The effect of charge was also noted in another study;

[141] A. Chrambach and D. Rodbard, *Science*, 1971, **172**, 440.
[142] (*a*) D. Rodbard and A. Chrambach, *Analyt. Biochem.*, 1971, **40**, 95; (*b*) D. Rodbard G. Kapadia, and A. Chrambach, *ibid.*, p. 135; (*c*) J. Lunney, A. Chrambach, and D., Rodbard, *ibid.*, p. 158.
[143] (*a*) L. Shuster in ref. 120, p. 124; (*b*) O. Gabriel in ref. 120, p. 565.
[144] E. Kaltschmidt, *Analyt. Biochem.*, 1971, **43**, 25.
[145] D. M. Easton, H. Lipner, J. Hines, and R. C. Leif, *Analyt. Biochem.*, 1971, **39**, 478.
[146] R. L. Darskus, *J. Chromatog.*, 1971, **55**, 425.
[147] S. J. Tata, *Analyt. Biochem.*, 1971, **42**, 470.
[148] D. M. Neville, jun., *J. Biol. Chem.*, 1971, **246**, 6328.
[149] J.-S. Tung and C. A. Knight, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 1117.
[150] J. G. Williams and W. B. Gratzer, *J. Chromatog.*, 1971, **57**, 121.

glucose oxidase, papain, and pepsin were found to be much more resistant to binding of SDS, and to loss of activity, than were other proteins.[151] It was also found that the maximum amount of SDS bound by proteins could be increased to 2 g (g protein)$^{-1}$ [*cf.* 1.4 g (g protein)$^{-1}$ reported last year] by increasing both the ionic strength (so confirming the hydrophobic nature of the interaction) and the SDS concentration. The effect of protein charge is acknowledged in the determination of the molecular weights of histones by SDS-gel electrophoresis;[152] the standard curve was drawn up using histones of known molecular weight. Histones have a lower mobility than expected in SDS gels, and this is attributed to the reduced net negative charge on the protein–SDS complex. The cationic detergent cetyltrimethylammonium bromide has been suggested[150] as a possible replacement for SDS in an attempt to circumvent the anomalous binding of the anionic detergent to (negatively) charged proteins. A linear log m. wt. *versus* mobility relationship is demonstrated, except for rather steep curvature in the region of high molecular weight. One can presumably expect anomalous binding of this detergent to basic proteins, and the answer might be to run gels in the presence of both anionic and cationic detergents. In the same survey[150] of the limitations of molecular weight determination in detergent–polyacrylamide gels, the point was made that the relation between molecular weight and mobility in SDS gels breaks down for proteins with molecular weight less than *ca.* 15 000, and that below *ca.* 6000 all molecules migrate with the same mobility. The effect is independent of acrylamide concentration, and is interpreted as a loss of asymmetry of the protein–SDS complexes, as expected (*i.e.* assuming a prolate ellipsoid, the major and minor semi-axes become equal, so that the frictional coefficient becomes essentially independent of molecular weight). Despite this, molecular weights of polypeptides in the range 1225—10 000 have been determined in SDS gels;[153] the accuracy claimed was ± 18% in most cases, and intrinsic charge had more effect than for proteins. Better resolution was achieved by including 8M-urea in the gels. When the molecular weights of the cyanogen bromide fragments of collagen (m. wt. *ca.* 5000—67 000) were studied in SDS gels it was found[154] that peptides from the α_1- and α_2-chains lay on two distinct straight lines, the mobilities in each case being lower than for globular proteins of similar molecular weights and being unchanged by various chemical treatments, including succinylation. This behaviour is attributed to a certain rigidity inherent in the structure of collagen (high Pro and Hypro content) which persists even after denaturation. In any event, for peptides from each chain there exists a good linear log m. wt. *versus* mobility relationship.[154] Complications aside, gel electrophoresis continues to make available information that

[151] C. A. Nelson, *J. Biol. Chem.*, 1971, **246**, 3895.
[152] S. Panyim and R. Chalkley, *J. Biol. Chem.*, 1971, **246**, 7557.
[153] R. T. Swank and K. D. Munkres, *Analyt. Biochem.*, 1971, **39**, 462.
[154] H. Furthmayr and R. Timpl, *Analyt. Biochem.*, 1971, **41**, 510.

would be difficult to obtain in other ways. For instance, micro-scale disc-electrophoresis makes possible a study of protein synthesis in single, identified neurons,[155] and a comparison of the sizes and number of copies of proteins in 70S and 80S ribosomes.[156] In a study of the ribosomal proteins of rabbit reticulocytes,[157] electrophoresis in the first dimension was carried out in urea gels in tubes, and in the second dimension in the presence of SDS in polyacrylamide slabs into which the gels were embedded after dialysis to remove urea. Molecular weights of all 63 resolved proteins could be estimated from the second-dimension run. SDS-gel electrophoresis has also given the molecular weights of all the major myofibrillar and sarcoplasmic muscle proteins.[158] There have been many instances in which the proposed quaternary structure of a protein has been checked by studying the SDS-gel pattern after intramolecular cross-linking with dimethyl suberimidate (see last year's Report). The method confirmed that 17β-hydroxysteroid dehydrogenase from human placenta was a dimer,[159] that inosine-5′-phosphate dehydrogenase[160] and glycerol kinase[161] were tetramers, and that carbamyl phosphate synthetase contained one catalytic and one regulatory subunit.[162] In a particularly interesting application[163] the effect of a feedback inhibitor on isopropylmalate synthase, an associating–dissociating system, could be observed directly on SDS gels; when cross-linking was carried out in the presence of the inhibitor, tetramers were shown to have been converted into dimers and monomers.

Electrophoresis in SDS gels is potentially even more powerful now that two reports have appeared of renaturation of proteins from SDS solutions. In a careful study[164] Weber and Kruter describe removal of the detergent with Dowex 1-X2; addition of urea avoids precipitation and irreversible adsorption on to the column. Then follows reactivation as usual from urea, either by dilution into buffer or by dialysis, with good recovery of enzymic activity. The procedure was applied successfully to several oligomeric proteins, even those as demanding as aspartate transcarbamylase, which also regained its allosteric properties, and RNA polymerase, which has the quaternary structure $\beta\beta'\alpha_2\omega_2$. Details are also given for recovering proteins in a fairly active form from (unstained) SDS gels, but it was found necessary to pre-run the gels in order to regain any activity at all. In addition, a procedure adapted for fingerprinting eluted proteins is described.[164] For this purpose the Dowex 1-X2 eluate (6 mol l^{-1} urea) is diluted to a urea

[155] H. Gainer, *Analyt. Biochem.*, 1971, **44**, 589.
[156] T. A. Bickle and R. R. Traut, *J. Biol. Chem.*, 1971, **246**, 6828.
[157] O. H. W. Martini and H. J. Gould, *J. Mol. Biol.*, 1971, **62**, 403.
[158] R. K. Scopes and I. F. Penny, *Biochim. Biophys. Acta*, 1971, **236**, 409.
[159] J. Jarabak and M. A. Street, *Biochemistry*, 1971, **10**, 3831.
[160] T.-W. Wu and K. G. Scrimgeour, *Canad. J. Biochem.*, 1971, **49**, 473.
[161] J. W. Thorner and H. Paulus, *J. Biol. Chem.*, 1971, **246**, 3885.
[162] P. P. Trotter, M. E. Burt, R. H. Haschemeyer, and A. Meister, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2599.
[163] G. Kohlhaw and G. Boatman, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 741.
[164] K. Weber and D. J. Kuter, *J. Biol. Chem.*, 1971, **246**, 4504.

concentration of 2 mol l^{-1}, digested with trypsin, and the tryptic peptides are then retained on Dowex 50-X2 to remove urea, and subsequently stripped with ammonia solution and freeze-dried for fingerprinting. A radioactively labelled membrane protein has recently [165] been eluted from SDS gels and successfully fingerprinted after an ion-exchange step to remove SDS. Another study [166] of the removal of SDS from proteins on an ion-exchange resin gave 100% recovery of activity from ribonuclease. [^{35}S]SDS was used to show that in general 99% of the detergent was removed, and it was noted that the remaining 1% might hinder the refolding of some proteins. It is likely that this method will be widely used, not only to recover small amounts of enzymes from analytical gels run in SDS, but also for membrane preparations solubilized in SDS and fractionated by preparative electrophoresis (see Section 8).

A technique that might find favour is dansylation of proteins as a means of making them visible in SDS gels.[167, 168] The advantages are increased sensitivity (10—50 ng detectable), and direct visualization of the gel through glass in u.v. light. For large samples, electrophoresis of a dansylated portion in parallel with the bulk could prove useful in locating regions of the gel from which proteins are to be eluted.[168] The mobility of proteins is apparently affected negligibly by dansylation.[168] A study [169] of radioactively labelled envelope proteins of *E. coli* used dansylated marker proteins for molecular weight estimations. Dansylation of proteins which have been fixed in polyacrylamide gels is described [170] as a means of identifying their N-terminal residues. Dansylation is carried out in the presence of acrylamide, and dansylated amino-acids are separated and identified on polyamide layers as usual, after hydrolysis and precipitation of the acrylamide. The method was demonstrated successfully on *S*-sulphotrypsinogen. Amino-acid analysis of stained (Amido Schwarz) bands from polyacrylamide gels has been described,[171] a good recovery of amino-acids being achieved by including 2-mercaptoethanol in the hydrolysis mixture. Again, the hydrolysis step was carried out in the presence of the gel. This is perhaps safer than attempted elution of the protein since, for example, dansylated immunoglobulin heavy chains (m. wt. 50 000) were eluted with great difficulty under conditions in which the light chains (m. wt. 25 000) were readily extractable.[168] A membrane protein (m. wt. 105 000) appears, however, to have been eluted successfully from gels.[165]

There have been numerous reports of improved procedures for counting radioactive proteins in polyacrylamide gels. The effectiveness of various commercially available solubilizers has been compared [172] and a procedure

[165] M. S. Bretscher, *J. Mol. Biol.*, 1971, **59**, 351.
[166] J. Lenard, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 662.
[167] D. N. Talbot and D. A. Yphantis, *Analyt. Biochem.*, 1971, **44**, 246.
[168] K. R. Shelton, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 367.
[169] M. Inouye, *J. Biol. Chem.*, 1971, **246**, 4834.
[170] R. Schyne, S. Bricteux-Grégoire, and M. Florkin, *J. Chromatog.*, 1971, **62**, 115.
[171] L. L. Houston, *Analyt. Biochem.*, 1971, **44**, 81.
[172] P. N. Paus, *Analyt. Biochem.*, 1971, **42**, 372.

in which the gel slice is put directly into the scintillation cocktail has been described.[173] A danger in the conventional hydrogen peroxide method of solubilizing gels is the loss of ^{14}C and ^{3}H (as carbon dioxide and water), and conditions have now been described[174] which avoid this. Specific radioactivities of brain proteins in stained gels have been measured:[175] slices of the gel (cross-linked with ethylene diacrylate) were solubilized in alkali and the radioactivity and the stain intensity (related to the amount of protein present) were determined. In a new method of counting,[176] proteins are eluted from the gel slice on to a glass-fibre disc and this is then counted while immersed in scintillation fluid; quenching is said to need no correction. It is possible, however, that there may be difficulties in eluting high-molecular-weight proteins (see above). Another procedure for mechanical fractionation of gels has been described[177] and faster diffusion destaining may be possible if activated charcoal is used for continuous removal of the eluted dye.[178] '. . . and the crooked shall be made straight . . .'[179] if those who photograph their gels in glass tubes adopt a suggestion made recently.[180]

Isoelectric Focusing. This method of using a pH gradient to separate proteins according to their isoelectric points has achieved wide currency both as a preparative method and as a criterion of homogeneity. Of the numerous instances in which it has proved useful only a few representative examples are cited here. The technique has been reviewed[181a] and details of the practical aspects have been clearly presented.[181b] Separation of lipoproteins from serum[182] is a typical example of its use; it is worth noting that a very narrow pH range was achieved in this case by pre-running the column and then collecting the ampholyte in the pH range desired. The tendency for proteins to precipitate at their pI can apparently be overcome by inclusion of 0.5% of a non-ionic detergent (Brij 35) in the pH gradient.[183] Isoelectric focusing in polyacrylamide gels (gel electrofocusing) is also becoming widely used, both preparatively and analytically, and details of the technique have been documented.[184, 185] Its use is illustrated by the preparative separation[186] of six rat haemoglobins, the largest difference in pI being 0.26 and the smallest only 0.04 pH units. A method for the rapid staining of protein bands after gel electrofocusing does not require removal

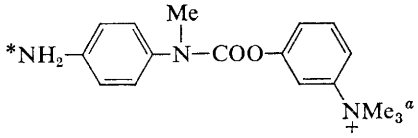
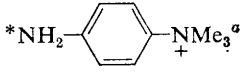
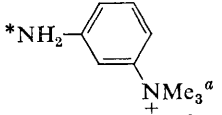
[173] S. A. Leon and A. T. Bohrer, *Analyt. Biochem.*, 1971, **42**, 54.
[174] D. Goodman and H. Matzura, *Analyt. Biochem.*, 1971, **42**, 481.
[175] K. Hemminki, *Acta Chem. Scand.*, 1971, **25**, 3887.
[176] C. W. Helleiner and W. H. Wunner, *Analyt. Biochem.*, 1971, **39**, 333.
[177] B. Bakay, *Analyt. Biochem.*, 1971, **40**, 429.
[178] L. J. Gathercol and L. Klein, *Analyt. Biochem.*, 1971, **44**, 232.
[179] Isaiah, Chap. 40, vs. 4.
[180] D. Oliver and R. Chalkley, *Analyt. Biochem.*, 1971, **44**, 540.
[181] (*a*) N. Catsimpoolas, *Separation Sci.*, 1970, **5**, 523; (*b*) O. Vesterberg in ref. 120, p. 389.
[182] J. J. Albers and A. M. Scanu, *Biochim. Biophys. Acta*, 1971, **236**, 29.
[183] A. D. Friesen, J. C. Jamieson, and F. E. Ashton, *Analyt. Biochem.*, 1971, **41**, 149.
[184] C. W. Wrigley, in Ref. 120, p. 559.
[185] P. Righette and J. W. Drysdale, *Biochim. Biophys. Acta*, 1971, **236**, 17.
[186] S. Stein, M. G. Cherian, and A. Mazur, *J. Biol. Chem.*, 1971, **246**, 5287.

of the ampholyte from the gel.[187] Isoelectric focusing in the absence and presence of urea (whether [188] or not [189] in gels) can yield information about the conformation of a protein, and about the nature of any microheterogeneity. In one study [188] it was found that whereas the pI values of several proteins were not affected by the presence of urea in the gel, that of plasma albumin was increased by *ca.* one pH unit. This was interpreted as the normalization of buried ionizable groups (in this case lysine); compensation of buried acidic and basic groups could account for cases of an unaltered isoelectric point in the presence of urea. In a related case (but no urea) an increase in the pI of TPCK-treated α-chymotrypsin relative to the native enzyme was held [190] to be consistent with the exposure of Ile-16. If multiple bands persist in a purified protein on isoelectric focusing in the presence of urea, this may be taken as evidence of differences in primary structure; it was observed for bovine plasma albumin [188] and it showed that fumarase isoenzymes contained six distinct types of subunit.[191] A procedure for the two-dimensional separation of proteins by a combination of thin-layer gel-filtration and thin-layer isoelectric focusing has been described.[192]

Affinity Chromatography. The power of this elegant method of purifying proteins is reflected in the countless cases in which it is being used to good effect; a selection of these is included in Table 1. The principles, limitations, and execution of the technique have been the subject of several excellent reviews [193–195] and little will be added here. Agarose activated with cyanogen bromide continues to be the support of choice in almost every instance; some exceptions are noted in Table 1. A report [106] of the chemistry of the activation step has already been mentioned; so has a method for coupling biologically active substances to polymers using isocyanides [107] (see Section 2D). Although flexible 'arms' for attachment of ligands to the support were found to be essential in some cases (*cf.* refs. 206, 207, 213, 217) to minimize interactions with the matrix, inspection of Table 1 shows that this is not always so; it is possible, of course, that in many instances 'arms' would have improved the chromatography. It is interesting that in one case at least a flexible 'arm' was found to be undesirable:[219] thymidine kinase was adsorbed better on to a column of Sepharose-5′-amino-5′-deoxythymidine than on to a column of the *N*-(6-aminohexyl) derivative. Affinity chromatography of β-galactosidase [206] did, however, require that the inhibitor ligand be attached *ca.* 21 Å from the matrix (10 Å was much worse). Other observations made in this work [206] may turn out to be generally applicable: 'interacting systems of low affinity' can lead to very effective adsorption,

[187] O. Vesterberg, *Biochim. Biophys. Acta*, 1971, **243**, 345.
[188] M. R. Salaman and A. R. Williamson, *Biochem. J.*, 1971, **122**, 93.
[189] N. Ui, *Biochim. Biophys. Acta*, 1971, **229**, 567.
[190] N. Ui, *Biochim. Biophys. Acta*, 1971, **229**, 582.
[191] P. E. Penner and L. H. Cohen, *J. Biol. Chem.*, 1971, **246**, 4261.
[192] F. Drawert and W. Müller, *Chromatographia*, 1971, **4**, 23.
[193] P. Cuatrecasas and C. B. Anfinsen, *Ann. Rev. Biochem.*, 1971, **40**, 259.
[194] P. Cuatrecasas and C. B. Anfinsen in ref. 120, p. 345.
[195] G. Feinstein, *Naturwiss.*, 1971, **58**, 389.

Table 1 *Some recent applications of affinity chromatography of proteins*

Substance purified	*Ligand (* indicates attachment through side-arm)*	*Elution*	*Ref.*
Acetylcholinesterase: electric eel		'Tensilon' (K_i = 10^{-6} mol l^{-1})	196
		'Tensilon' (K_i = 10^{-6} mol l^{-1})	196
erythrocyte		'Tensilon' (K_i = 10^{-6} mol l^{-1})	196
ara C Protein (*E. coli*)	* 4-Aminophenyl-β-D-6-deoxygalacto-pyranoside	Borate buffer, pH 10	197
Brain-specific antibodies (human)	Brain proteins	Acetic acid	198
Cobalamin-binding proteins	Cobalamin–albumin conjugate[b]	Hydroxycobalmin	199
DNA Polymerase (HeLa cells)	HeLa DNA (with single-stranded end)	0.1M-Phosphate buffer	200
Deoxyribonuclease I inhibitor (calf thymus)	Deoxyribonuclease I (pancreatic)	3M-Guanidine hydrochloride, 1M-acetate, 30% glycerol	201
3-Deoxy-D-*arabino*-heptulosonate-7-phosphate synthetase (tyrosine-sensitive isoenzyme)	Tyrosine[c]	—	202

[196] J. D. Berman and M. Young, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 395.
[197] G. Wilcox, K. J. Clemetson, D. V. Santi, and E. Englesberg, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 2145.
[198] I. Tripatzis, K. Warecka, and M.-C. Wong, *Nature New Biology*, 1971, **230**, 250.
[199] H. Olesen, E. Hippe, and E. Haber, *Biochim. Biophys. Acta*, 1971, **243**, 66.
[200] M. S. Poonian, A. J. Schlabach, and A. Weissbach, *Biochemistry*, 1971, **10**, 424.
[201] U. Lindberg and S. Eriksson, *European J. Biochem.*, 1971, **18**, 474.
[202] M. Takahashi and W. W.-C. Chan, *Canad. J. Biochem.*, 1971, **49**, 1015.

Table 1 (*cont.*)

Substance purified	*Ligand* (* *indicates attachment through side-arm*)	*Elution*	*Ref.*
Dihydrofolate reductase: *Lactobacillus casei*	* Methotrexate ($K_i = 10^{-8}$ mol l^{-1})	Dihydrofolate ($K_m = 5 \times 10^{-6}$ mol l^{-1}) and NADPH	203
chicken liver	* Methotrexate	Dihydrofolate	204
T_4 phage	* N^{10}-Formylaminopterin[a]	Dihydrofolate	205
β-Galactosidase (*E. coli*)	* *p*-Aminophenyl-β-D-thiogalacto-pyranoside	0.1M-Borate	206
Glutamic–oxaloacetic transaminase apoenzyme (pig heart)	* Pyridoxamine 5′-phosphate	Glutamic acid, phosphate	207
Glycogen synthetase	UDP hexitolamine	Glycogen	208
Growth hormone (monkey)	Anti-human-placental lactogen	Not done (contaminant of prolactin)	209
Lactose synthase A protein (bovine)	β-Lactalbumin	Omission of glucose from buffer	210
	UDP hexitolamine	Cacodylate buffer + edta	208
Lipoprotein lipase (bovine milk)	Heparin	NaCl gradient	211
Neuraminidases: micro-organisms	* *N*-(*p*-Aminophenyl)oxamic acid	0.1M-$NaHCO_3$	212
intact influenza virus	* *N*-(*p*-Aminophenyl)oxamic acid	0.1M-$NaHCO_3$	212
Protein kinase (rat parotid, rabbit muscle)	* Cyclic AMP	see text	213
Pyridine-nucleotide-dependent dehydrogenases	* NAD^+ or $NADP^+$[b]	KCl gradient	214
	* NAD^+	NAD^+	215
	* AMP	NAD^+	215
Receptors: acetylcholine (electroplax)	Cobra toxin	—	216
glucagon (liver)	* Glucagon	—	217
Retinol binding protein (human)	Prealbumin	Dec. ionic strength	218
Thymidine kinase (*E. coli*)	5′-Amino-5′-deoxythymidine	0.5M-NaCl	219
Tyrosine hydroxylase (brain)	3-Iodotyrosine	1mM-KOH	220
Proteolytic Enzymes, Zymogens, and Inhibitors			
Carboxypeptidase A	Gly-D-Phe[b]	0.25M-NaCl (pH 7.5)	221
	Gly-D-Phe[d]	Acetic acid	222
Carboxypeptidase B	D-Ala-L-Arg	NaCl gradient	223
Chymotrypsin	4-Phenylbutylamine	pH 3	224
Pepsin	Polylysine (inhibitor)	NaCl gradient	225

Trypsin (bovine) (α- and β-forms separated)	Chicken ovomucoid	Dec. pH gradient	226
	p-(*p'*-Aminophenoxypropoxy)benzamidine	pH 4	227
Thrombin (bovine)	* *p*-Chlorobenzylamine	Benzamidine (1.0 mol l^{-1})	228
Plasminogen (human)	Lysine (inhibitor analogue)	Lysine (0.1 mol. l^{-1})	229
	Lysine[a]	ε-Aminocaproic acid	230
Ovoinhibitor (chicken)	Chymotrypsin (bovine)	0.2M-KCl, pH 2	231
Trypsin inhibitor (peanuts)	Trypsinogen (bovine)	0.01M-HCl	232

[a] Polyacrylamide; [b] Cellulose; [c] Sephadex; [d] Glass.

[203] P. C. H. Newbold and N. G. L. Harding, *Biochem. J.*, 1971, **124,** 1.
[204] B. T. Kaufman and J. V. Pierce, *Biochem. Biophys. Res. Comm.*, 1971, **44,** 608.
[205] J. S. Erickson and C. K. Mathews, *Biochem. Biophys. Res. Comm.*, 1971, **43,** 1164.
[206] E. Steers, jun., P. Cuatrecasas, and H. B. Pollard, *J. Biol. Chem.*, 1971, **246,** 196.
[207] R. Collier and G. Kohlaw, *Analyt. Biochem.*, 1971, **42,** 48.
[208] J. Shaper, R. L. Hill, and R. Barker, *Fed. Proc.*, 1971, **30,** 1265.
[209] H. J. Guyda and H. G. Friesen, *Biochem. Biophys. Res. Comm.*, 1971, **42,** 1068.
[210] I. P. Trayer and R. L. Hill, *J. Biol. Chem.*, 1971, **246,** 6666.
[211] T. Olivecrona, T. Egelrud, P.-H. Iverius, and U. Lindahl, *Biochem. Biophys. Res. Comm.*, 1971, **43,** 524.
[212] P. Cuatrecasas and G. Illiano, *Biochem. Biophys. Res. Comm.*, 1971, **44,** 178.
[213] M. Wilchek, Y. Salomon, M. Lowe, and Z. Selinger, *Biochem. Biophys. Res. Comm.*, 1971, **45,** 1177.
[214] C. R. Lowe, and P. D. G. Dean, *F.E.B.S. Letters*, 1971, **14,** 313.
[215] K. Mosbach, H. Guilford, P.-O. Larsson, R. Ohlsson, and M. Scott, *Biochem. J.*, 1971, **125,** 20P.
[216] J. P. Changeux, J. C. Meunier, and M. Huchet, *Mol. Pharmacol.*, 1971, **7,** 538.
[217] F. Krug, B. Desbuqois, and P. Cuatrecasas, *Nature New Biology*, 1971, **234,** 268.
[218] (*a*) L. Rask, A. Vahlquist, and P. A. Peterson, *J. Biol. Chem.*, 1971, **246,** 6638; (*b*) A. Vahlquist, S. F. Nilsson, and P. A. Peterson, *European J. Biochem.*, 1971, **20,** 160.
[219] W. Rohde and A. G. Lezius, *Z. physiol. Chem.*, 1971, **352,** 1507.
[220] W. N. Poillon, *Biochem. Biophys. Res. Comm.*, 1971, **44,** 64.
[221] J. R. Uren, *Biochim. Biophys. Acta*, 1971, **236,** 67.
[222] P. J. Robinson, P. Dunnill, and M. D. Lilly, *Biochim. Biophys. Acta*, 1971, **242,** 659.
[223] M. Sokolovsky and N. Zisapel, *Biochim. Biophys. Acta*, 1971, **250,** 203.
[224] K. J. Stevenson and A. Landman, *Canad. J. Biochem.*, 1971, **49,** 119.
[225] B. Nevaldine and B. Kassell, *Biochim. Biophys. Acta*, 1971, **250,** 207.
[226] N. C. Robinson, R. W. Tye, H. Neurath, and K. A. Walsh, *Biochemistry*, 1971, **10,** 2743.
[227] G. W. Jameson and D. T. Elmore, *Biochem. J.*, 1971, **124,** 66P.
[228] A. R. Thompson and E. W. Davie, *Biochim. Biophys. Acta*, 1971, **250,** 210.
[229] T. H. Liu and E. T. Mertz, *Canad. J. Biochem.*, 1971, **49,** 1055.
[230] E. E. Rickli and P. A. Cuendet, *Biochim. Biophys. Acta*, 1971, **250,** 447.
[231] G. Feinstein, *Biochim. Biophys. Acta*, 1971, **236,** 74.
[232] K. K. Stewart and R. F. Doherty, *F.E.B.S. Letters*, 1971, **16,** 226.

provided that the insolubilized ligand is present in high concentration, is separated substantially from the matrix backbone, and is readily accessible to the protein in solution. In the particular case of β-galactosidase, agarose beads were good, but polyacrylamide was not porous enough to allow the large enzyme to reach the ligand inside the beads. Inactive monomers of β-galactosidase also bound to the affinity column, suggesting that the method might be applicable to inactive mutant proteins if these can still bind substrate analogues.[206]

Noteworthy are the preliminary reports of isolation of cell receptors; the membrane receptor of the neurotransmitter acetylcholine was separated from acetylcholinesterase,[216] and the cell receptor for glucagon (dissociation constant 10^{-9}—10^{-10}) was isolated.[217] (The glucagon receptor in liver cells has also been studied extensively by other workers.[233]) Of interest too is the use of insolubilized cofactors to resolve groups of enzymes.[214, 215] It seems that the distance of the bound cofactor from the matrix is of less significance in the binding of enzyme than is the conformation of the cofactor.[214] Cofactors insolubilized in Sephadexes of various pore sizes form the basis of a method suggested for determining the molecular weight of dehydrogenases,[234] and NAD^+ bound to porous glass will function in dehydrogenase reactions.[235] Insolubilized UDP could prove useful for purification of a variety of enzymes (*e.g.* glycosyl transferases) which use UDP or its derivatives as substrate.[209] Protein kinases are activated by cyclic AMP, which binds tightly to them. It is perhaps not too surprising, then, that affinity chromatography of a protein kinase preparation on insolubilized AMP gave [213] a fully activated enzyme. This is attributed to the retention by the column of a regulatory unit that binds cAMP (although this has not yet been recovered), and the method is suggested as a good one for isolating the free catalytic unit.[213]

Last year saw an elegant approach to the purification of affinity-labelled active-site peptides by affinity chromatography with the native protein itself as ligand. The same workers describe [236] a method which is generally applicable to the isolation of any modified peptides. This uses a column of an antibody directed against the particular modifying group, so that all modified peptide(s) will be retained; a mixture of adsorbed peptides then calls for separation by conventional methods. In the examples described (for Dnp- and arsanilazo- [236] and nitro- [237] groups) the method is used to identify labelled peptides in a selectively modified protein, but it is worth bearing in mind the suggestion [237] that it might prove useful in sequence work for isolation of all peptides containing a particular modifiable residue (*e.g.* tyrosine, tryptophan, lysine).

[233] M. Rodbell, H. M. J. Krans, S. L. Pohl, and L. Birnbaumer, *J. Biol. Chem.*, 1971, **246**, 1861, and several accompanying papers.
[234] C. R. Lowe and P. D. G. Dean, *F.E.B.S. Letters*, 1971, **18**, 31.
[235] M. K. Weibel, H. H. Weetall, and H. J. Bright, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 347.
[236] M. Wilchek, V. Bocchine, M. Becker, and D. Givol, *Biochemistry*, 1971, **10**, 2828.
[237] M. Helman and D. Givol, *Biochem. J.*, 1971, **125**, 971.

3 Structural Proteins

The term 'structural proteins' is taken here to include the fibrous proteins, proteins connected with motility, and certain other globular proteins that have no identified enzymic activity. Immunoglobulins and membrane proteins are dealt with separately in Sections 7 and 8 respectively. Methylated amino-acids occur in many 'structural' proteins and a useful review of their occurrence, biosynthesis, and biological significance has been published.[238]

A. The Proteins of Motility.—A study has been made of the enzymic methylation of skeletal-muscle contractile proteins [239] which shows that methionine can act as the methyl donor for the methylation of histidine and lysine residues *in vitro*. Actin has been isolated from the soil amoeba *Acanthamoeba*:[240] it resembles muscle actin in many ways but, unlike muscle action, it also contains N^{ϵ}-dimethyl-lysine.

Amino-acid sequences around four of the five cysteine residues of trout actin have been reported.[241] The trout and rabbit muscle actins are clearly highly homologous. The amino-acid sequence around the single residue of 3-methylhistidine in rabbit skeletal muscle actin has also been determined:[242]

-Gly-Gln-Lys-Asp-Ser-Tyr-Val-Gly-Asp-Glu-Ala-Gln-Ser-Lys-Arg-Gly-Ile-Leu-Thr-Leu-Lys-Tyr-Pro-Ile-Glu-3MeHis-Trp-Gly-Ile-Ile-Thr-Asn-Asp-Asp-Met-

During the preparation of myosin, low concentrations of chains intermediate in size between heavy and light chains have been detected using SDS-gel electrophoresis.[243] It is suggested that these proteins may be new proteins of the thick filament assembly. The *S*-β-(4-pyridylethyl)-L-cysteine derivatives of myosin have also been used for isoelectric focusing in polyacrylamide gels.[244] A detailed study of the light chains of myosin has now been published.[245] Three types of light chain were detected, of which two are related (although they differ in molecular weight by up to 9000 daltons) in that they each contain the sequence -Met-Ala-Gly-Gln-Glu-Asp-Ser-Asn-Gly-Cmc-Ile-Asn-Tyr-. Myosin and heavy meromyosin contain two moles of these related light chains, removal of which (at pH 11) causes total loss of ATPase activity. The third type of light chain is chemically different from the other two and can be released from myosin by treatment with DTNB [5,5′-dithiobis-(2-nitrobenzoate)] without loss of ATPase activity, but again there appear to be two moles of this light chain.

[238] W. K. Paik and S. Kim, *Science*, 1971, **174**, 114.
[239] B. Krzysik, J. P. Vergnes, and I. R. McManus, *Arch. Biochem. Biophys.*, 1971, **146**, 34.
[240] R. R. Weihing and E. D. Korn, *Biochemistry*, 1971, **10**, 590.
[241] J. Bridgen, *Biochem. J.*, 1971, **123**, 541.
[242] M. Elzinga, *Biochemistry*, 1971, **10**, 224.
[243] R. Starr and G. Offer, *F.E.B.S. Letters*, 1971, **15**, 40.
[244] P. J. Bechtel, A. M. Pearson, and C. E. Bodwell, *Analyt. Biochem.*, 1971, **43**, 509.
[245] A. G. Weeds and S. Lowey, *J. Mol. Biol.*, 1971, **61**, 701.

The amino-acid sequence around the single residue of 3-methylhistidine in rabbit skeletal muscle myosin has also been established:[246]

-Leu-Leu-Gly-Ser-Ile-Asp-Val-Asp-3MeHis-Gln-Thr-Tyr-Lys-

It is quite unlike the corresponding sequence in actin (see above). Although located uniquely in myosin subfragment-1, its function is obscure, but there is some correlation between the content of 3-methylhistidine and ATP-ase activity. $N^{G}N^{G}$-Dimethylarginine has been detected in the myosin from the developing leg muscle of various animals but not in the actin nor in adult myosin.[247] There was no evidence for the existence of NN'-dimethylarginine or N^{G}-monomethylarginine.

Tropomyosins from a number of invertebrates (crayfish, oyster, abalone, and blowfly) have been shown to have a subunit molecular weight of 34 000, indistinguishable from that of vertebrate tropomyosins.[248] The preparation of parvalbumins, the occurrence of which is restricted to the muscles of animals from the lower phyla of vertebrates, has been described.[249] The major parvalbumin from hake muscle has a molecular weight of 11 500 and glycine as its C-terminal residue.[250]

The tubulin from microtubules of sea urchin sperm flagellae and pig brain [251] and from other mammals [252] has been reported to contain two non-identical but homologous subunits. Similar heterogeneity has been found for the cytoplasmic microtubules of chick embryo brain, the microtubules appearing as heteropolymers.[253] Two distinct proteins have been observed in the microtubules from brain and neuroblastoma cells, and at least five different proteins were recognized in the outer doublet microtubules of flagellae from *Chlamydomonas*.[254] There is also a very interesting report [255] that ATP will induce sliding of tubules in the flagellae of sea-urchin sperm after trypsin treatment.

The partial amino-acid sequence of flagellin from *Salmonella adelaide* has been described.[256] Difficulty was experienced in cleaving a Met–Ser bond with cyanogen bromide, comparable with that reported for Met–hydroxyamino-acid bonds in other proteins (see last year's Report).

B. Collagen.—The chemistry and structure of collagen have been well reviewed in detail.[257] It now appears that collagen is actually synthesized

[246] G. Huszar and M. Elzinga, *Biochemistry*, 1971, **10**, 229.
[247] M. Reporter and J. L. Corbin, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 644.
[248] E. F. Woods and M. J. Pont, *Biochemistry*, 1971, **10**, 270.
[249] J.-F. Pechère, J. Demaille, and J.-P. Capony, *Biochim. Biophys. Acta*, 1971, **236**, 391.
[250] J.-F. Pechère, J.-P. Capony, and L. Ryden, *European J. Biochem.*, 1971, **23**, 421.
[251] H. Feit, L. Slusarek, and M. L. Shelanski, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2028.
[252] R. E. Fine, *Nature New Biology*, 1971, **233**, 283.
[253] J. Bryan and L. Wilson, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1762.
[254] J. B. Olmsted, G.-B. Witman, K. Carlson, and J. L. Rosenbaum, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2273.
[255] K. E. Summers and I. R. Gibbons, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 3092.
[256] B. E. Davidson, *European J. Biochem.*, 1971, **18**, 524.
[257] W. Traub and K. A. Piez, *Adv. Protein Chem.*, 1971, **25**, 243.

as a larger procollagen molecule from which much non-collagen peptide material is subsequently removed enzymically.[258] The presence in normal tissues of an enzyme that trims off such peptide from the N-terminus of procollagen during or after the formation of the typical triple helix has been demonstrated.[259] It is absent in dermatosparaxic animals.

Primary Structure. The enzymic hydrolysis of collagen has been improved by constructing a 'cocktail' of aminopeptidase, prolidase, prolinase, carboxypeptidase, and X-prolyl-aminopeptidase:[260] sequential addition of collagenase, papain, and the 'cocktail' causes 90% hydrolysis of collagen. The zymogen of tadpole collagenase has been isolated.[261]

An unusual collagen, containing three identical α-chains, has been isolated from chick cartilage.[262] The component α-chains are designated α1(II)-chains, since they differ in primary structure from the normal α1(I)-chains. The α1(II)-chain contains much more hydroxylysine than the α1(I)-chain and is apparently restricted to cartilage. Other experiments [263] have shown that the α1- and α2-chains of both rat and bovine collagen differ in molecular weight by 6000 daltons, as estimated by SDS-gel electrophoresis. The significance is unclear as yet but it would have a bearing on collagen cross-linking if substantiated.[263] Differences in the glycosylation of lysine residues in collagens from human skin and bone have also been reported.[264]

Further work has been described on the peptic digestion of collagen from rat-tail tendon.[265] Several fragments were recovered from the digest of denatured collagen and located in tropocollagen by electron microscopy after renaturation. However, for a molecule as large as collagen, cleavage with cyanogen bromide is clearly the method of choice. Thus, studies are in progress on the cyanogen bromide peptides from the α1- and α2-chains of acid-soluble ox collagen,[266] pig-skin collagen,[267] soluble human- and baboon-skin collagens,[268] and the α1-chain of acid-soluble calf-skin collagen.[269] Electron microscopy is again proving useful in establishing the order of the fragments.[267, 269] The amino-acid sequence of peptide CB4 (third from the

[258] G. Bellamy and P. Bornstein, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1138; R. L. Church, S. E. Pfeiffer, and M. L. Tanzer, *ibid.*, p. 2638.
[259] C. M. Lapière, A. Lenaers, and L. D. Kohn, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 3054.
[260] H. B. Bensusan, S. N. Dixit, and S. D. McKnight, *Biochim. Biophys. Acta*, 1971, **251**, 100.
[261] E. Harper, K. J. Bloch, and J. Gross, *Biochemistry*, 1971, **10**, 3035.
[262] E. J. Miller, *Biochemistry*, 1971, **10**, 1652.
[263] B. C. Sykes and A. J. Bailey, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 340.
[264] S. R. Pinnell, R. Fox, and S. M. Krane, *Biochim. Biophys. Acta*, 1971, **229**, 119.
[265] J. Pikkarainen, K. Lampaiho, and K. Kulonea, *Biochim. Biophys. Acta*, 1971, **251**, 141.
[266] D. Volpin and A. Veis, *Biochemistry*, 1971, **10**, 1751.
[267] W. Heinrich, P. M. Lange, T. Stirtz, C. Iancu, and E. Heidemann, *F.E.B.S. Letters*, 1971, **16**, 63.
[268] E. H. Epstein, R. D. Scott, E. J. Miller, and K. A. Piez, *J. Biol. Chem.*, 1971, **246**, 1718.
[269] J. Rauterberg and K. Kühn, *European J. Biochem.*, 1971, **19**, 398.

N-terminus) from the α1-chain of rat-skin collagen has been shown to be:[270]

Gly-Pro-Arg-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Lys-Asn-Gly-Asp-Asp-
1 **10**
Gly-Glu-Ala-Gly-Lys-Pro-Gly-Arg-Hyp-Gly-Gln-Arg-Gly-Pro-Hyp-
20 **30**
Gly-Pro-Gln-Gly-Ala-Arg-Gly-Leu-Hyp-Gly-Thr-Ala-Gly-Leu-Hyp-
40
Gly-Met

The Arg–Hyp bond was found to be slowly cleaved by trypsin, whereas the Lys–Pro bond was completely resistant. In a separate study [271] the peptides CB2 from the α2-chains of chick- and rat-skin collagens have been compared:

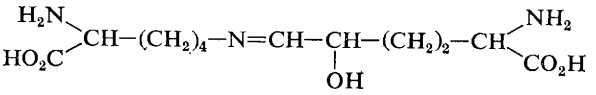
Chick skin **Gly-Pro-Ala -Gly-Asn-Arg-Gly-Ala-Ser-Gly-Pro-Ala-Gly-Val-Lys-**
Rat sin **Gly-Pro-Hyp-Gly-Asn-Arg-Gly-Thr-Ser-Gly-Pro-Ala-Gly-Val-Arg-**
1 **10**
Chick skin **Gly -Pro -Asn -Gly -Asp -Ala -Gly -Arg -Hyp -Gly -Glu -Hyp -Gly-**
Rat skin **(Gly, Pro, Asx, Gly, Asx, Ala, Gly, Arg, Hyp, Gly, Glx, Hyp, Gly,**
20
Chick skin **Leu -Met**
Rat skin **Leu, Met)**
30

The Arg–Hyp bond was again found to be slowly cleaved by trypsin.

The cyanogen bromide peptides from the α1(II)-chain of chick cartilage collagen (see above) have also been isolated and the C-terminal sequence of the chain has been shown to be -Met-Tyr.[272] A non-helical region of *ca.* 20 residues has been detected at the C-terminus of calf-skin collagen.[273] It is rich in aromatic amino-acids and contains a lysine-derived aldehyde, suggestive of cross-links.

Cross-links. The cross-links in collagen from invertebrates, cyclostomes, and elasmobranchs have been compared.[274] Detailed evidence has been presented [275] for the existence of two aldimine bonds as cross-links in the intact collagen from soft tissues: dehydro-hydroxylysino-norleucine (5) occurs as a major component whereas dehydro-lysino-norleucine is only found in traces. In highly insoluble collagens, the proportion of dehydro-lysino-norleucine rises and that of dehydro-hydroxylysino-norleucine falls. Neither aldimine appears to become reduced *in vivo* in collagen, which suggests that factors other than reduction stabilize the cross-links. This is in contrast with elastin, where dehydro-lysino-norleucine does become reduced to give a more stable link.

[270] W. T. Butler and S. L. Ponds, *Biochemistry*, 1971, **10**, 2076.
[271] J. H. Highberger, A. H. Kang, and J. Gross, *Biochemistry*, 1971, **10**, 611.
[272] E. J. Miller, *Biochemistry*, 1971, **10**, 3030.
[273] M. Stark, J. Rauterberg, and K. Kühn, *F.E.B.S. Letters*, 1971, **13**, 101.
[274] A. J. Bailey, *F.E.B.S. Letters*, 1971, **18**, 154.
[275] A. J. Bailey and C. M. Peach, *Biochem. J.*, 1971, **121**, 257.



(5)

Other experiments [276] have shown that selective removal of the non-helical regions from the N-terminus of rat-skin collagen using cyanogen bromide leaves *ca.* 70% of the protein aldehydes still covalently attached to the helical region. These aldehydes participate in cross-links of the Schiff-base type with lysine or hydroxylysine residues of neighbouring molecules, the aldol-type cross-link being largely restricted to the N-terminal region. On the other hand, it has been reported [277] that the long-term gain in tensile strength of scar-tissue collagen in man and guinea-pig does not correlate well with the extent of cross-linking.

Hydroxylated Residues. Collagen contains hydroxylated proline and lysine residues. 2,2′-Bipyridyl blocks the hydroxylation in chick calvaria and permits the isolation of an α1-chain that appears to be non-hydroxylated (a protocollagen).[278] The inhibition of hydroxylation appears also to inhibit the normal extrusion of collagen by fibroblasts in culture.[279] The enzymes involved in the hydroxylation of lysine and proline in protocollagen can be separated chromatographically [280] and the proline hydroxylase from chick embryo has been examined in some detail.[281] An investigation of its specificity [282] shows that the minimum sequence for hydroxylation is an intact -X-Pro-Gly- triplet. The neighbouring amino-acids determine the rate of proline hydroxylation. On the other hand, studies *in vitro* of the α1-CB2 fragments of rat skin and tendon collagen indicate that intracellular factors other than primary structure govern the degree of hydroxylation of susceptible proline residues.[283]

The degree of hydroxylation of lysine residues in the N-terminal telopeptide region of the α1- and α2-chains of various collagens has been investigated.[284] The extent varies with the collagen, skin collagen having no hydroxylysine in these positions except in the chick embryo and newly born rat, the level falling with age.[285]

[276] K. Deshmukh and M. E. Nimmi, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 416.
[277] L. Forrest and D. S. Jackson, *Biochim. Biophys. Acta*, 1971, **229**, 681.
[278] P. K. Müller, E. McGoodwin, and G. R. Martin, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 110.
[279] P. B. Ramaley and J. Rosenbloom, *F.E.B.S. Letters*, 1971, **15**, 59.
[280] R. L. Miller, *Arch. Biochem. Biophys.*, 1971, **147**, 339.
[281] M. Pankalainen and I. K. Kivirikko, *Biochem. Biophys. Acta*, 1971, **229**, 504.
[282] J. O'D. McGee, R. E. Rhoads, and S. Udenfriend, *Arch. Biochem. Biophys.*, 1971, **144**, 343.
[283] R. E. Rhoads, S. Udenfriend, and P. Bornstein, *J. Biol. Chem.*, 1971, **246**, 4138.
[284] M. J. Barnes, B. J. Constable, L. F. Morton, and E. Kodicek, *Biochem. J.*, 1971, **125**, 433.
[285] M. J. Barnes, B. J. Constable, L. F. Morton, and E. Kodicek, *Biochem. J.*, 1971, **125**, 925.

C. Elastin.—Elastin has been the subject of a substantial review.[286] Tropoelastin, the putative precursor of elastin, can now be purified easily from the aorta of copper-deficient pigs.[287] It contains virtually no carbohydrate.[288] Structural studies suggest that the sequences -Ala-Ala-Ala-Lys- and -Ala-Ala-Lys- may each be repeated six times in the peptide chain and may be implicated in the formation of the desmosine and isodesmosine cross-links.[73]

There is mass-spectral evidence [289] for the existence in elastin of cyclic dehydrodesmopiperidines, precursors of desmosine and isodesmosine. The presence of dehydromerodesmosine in ox ligamentum elastin has also been reported.[290] Elastin therefore contains at least two types of Schiff-base cross-links, since the occurrence of dehydro-lysino-norleucine is well established, and both occur naturally in elastin as their reduced derivatives merodesmosine and lysino-norleucine, respectively (*cf.* collagen above).

D. Fibrinogen.—The *S*-carboxymethylated chains of human fibrinogen have been separated by chromatography on CM-cellulose in 8M-urea [291] and a comparative study of human, cow, pig, and sheep fibrinogens has been reported.[292] Peptide maps reveal the close similarity. The C-terminal residues of various mammalian fibrinogens and fibrin have also been established:[293]

Source	*Chain*		
	α-	β-	γ-
Ox, Sheep	Pro	Val	Val
Human, Horse	Val	Val	Val
Pig, Dog	Pro	Val	Ile

The conversion of fibrinogen into fibrin in the lobster has been shown [294] to involve the formation of N^{ε}-(γ-glutamyl)lysine cross-links but, unlike vertebrate fibrinogen, lobster fibrinogen apparently undergoes no prior proteolysis. The existence of cross-links between α-chains in human fibrin has now been reported [295] but no β–β links could be detected, although α–γ and γ–γ links are, of course, already known. At least five or six α-chains

[286] 'Chemistry and Molecular Biology of the Intercellular Matrix', ed. E. A. Balazs, Academic Press, London, 1970, Vol. 1.
[287] L. B. Sandberg, R. D. Zeikus, and I. M. Coltrain, *Biochim. Biophys. Acta*, 1971, **236**, 542.
[288] M. E. Grant, F. S. Steven, D. S. Jackson, and L. B. Sandberg, *Biochem. J.*, 1971, **121**, 197.
[289] M. A. Paz, P. M. Gallop, O. O. Blumenfeld, E. Henson, and S. Seifter, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 289.
[290] M. A. Paz, E. Henson, O. O. Blumenfeld, S. Seifter, and P. M. Gallop, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 1518.
[291] G. Murano, B. Wiman, M. Blombäck, and B. Blombäck, *F.E.B.S. Letters*, 1971, **14**, 37.
[292] T. Cartwright and R. G. O. Kekwick, *Biochim. Biophys. Acta*, 1971, **236**, 550.
[293] M. Okude and S. Iwanaga, *Biochim. Biophys. Acta*, 1971, **251**, 185.
[294] G. M. Fuller and R. F. Doolittle, *Biochemistry*, 1971, **10**, 1311.
[295] R. P. McDonagh, J. McDonagh, M. Blombäck, and B. Blombäck, *F.E.B.S. Letters*, 1971, **14**, 33.

were involved in the cross-linking, indicating it to be an intermolecular event in which several fibrin monomers participate.[295] The occurrence of intermolecular γ–γ cross-links has also been elegantly demonstrated [296] by clotting a mixture of ox and human fibrinogens. Since the human and ox γ-chains differ in primary structure at the cross-linking site, the isolation of a hybrid ox–human γ–γ cross-linked peptide unit can be taken as conclusive evidence of the intermolecular nature of the link.

E. Chromosomal Proteins.—The structure and function of histones and nucleohistones have been the subject of recent reviews.[297, 298] The separation of histones from normal rat liver nuclei has been described [299] and the selective extraction of histones from chicken erythrocyte nuclei using NaCl and urea has been reported.[300] Other experiments [301] have dealt with the separation of histones by fractional displacement with protamine from nucleoproteins *in vitro*.

It has been claimed that histones are better resolved by electrophoresis on cellulose acetate than in polyacrylamide gels.[302] It has also been reported [303, 152] that histones show anomalous electrophoretic behaviour in SDS–polyacrylamide gels. Since they bind as much SDS as other proteins, it is probable that the anomalies are due to the unusually small overall negative charges of histones in this system.[152]

The N-terminal sequence of calf thymus histone III (F_3) has been determined in the sequenator:[304]

Ala-Arg-Thr-Lys-Gln-Thr-Ala-Arg-Lys-Ser-Thr-Gly-Gly-Lys-Ala-
1 **10**
Pro-Arg-Lys-Glu-Leu-Ala-Thr-Lys-Ala-Ala-Arg-.
20

There is some homology with other histones. Treatment of the lysine-rich histones of rabbit thymus with *N*-bromosuccinimide yields a 72-residue fragment from the N-terminus which is rich in basic amino-acids (Figure 1).[305] Comparison with other slightly lysine-rich and with arginine-rich histones reveals that they share this highly asymmetric distribution of basic amino-acids, the C-terminal regions being rich in hydrophobic residues.[305] Further work on the lysine-rich histones of rabbit thymus shows that fraction 3 has a diminished capacity for enzymic phosphorylation whereas

[296] R. F. Doolittle, R. Chen, and F. Lau, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 94.
[297] 'Histones and Nucleohistones', ed. D. M. P. Phillips, Plenum Press, London and New York, 1971.
[298] R. J. DeLange and E. L. Smith, *Ann. Rev. Biochem.*, 1971, **40**, 279.
[299] S. K. Chanda and A. L. Dounce, *Arch. Biochem. Biophys.*, 1971, **146**, 441.
[300] T. Senshu, *Biochim. Biophys. Acta*, 1971, **236**, 349.
[301] D. R. van der Westhuyzen and C. von Holt, *F.E.B.S. Letters*, 1971, **14**, 333.
[302] F. Machicao and J. Sonnenbichler, *Biochem. Biophys. Acta*, 1971, **236**, 360.
[303] L. H. Cohen and B. V. Gotchel, *J. Biol. Chem.*, 1971, **246**, 1841.
[304] M. O. J. Olson, J. Jordan, and H. Busch, *Biochem. Biophys. Res. Comm.*, 1971, **46**, 50.
[305] S. C. Rall and R. D. Cole, *J. Biol. Chem.*, 1971, **246**, 7175.

fraction 4 is readily phosphorylated: this can be explained by the observation that fraction 3 has alanine and fraction 4 has serine at position 37 in the sequence.[306]

Ac-Ser-Glu-Ala-Pro-Ala-Glu-Thr-Ala-Ala-Pro-Ala-Pro-Ala-Glu-Lys-Ser-Pro-
1 10
Ala-Lys-Lys-Lys-Lys-Ala-Ala-Lys-Lys-Pro-Gly-Ala-Gly-Ala-Ala-Lys-Arg-Lys-
20 30
Ala-Ala-Gly-Pro-Pro-Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ala-Ala-Ser-Lys-Glu-
40 50
Arg-Asn-Gly-Leu-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ala-Leu-Ala-Ala-Gly-Gly-Tyr-
60 70

Figure 1 *The N-terminal amino-acid sequence of lysine-rich histone from rabbit thymus*

The histones of chicken erythrocytes and calf thymus are very similar apart from the existence of histone F2c, peculiar to erythrocytes.[307] On the other hand, the histones of sea urchin sperm and calf thymus are readily distinguished.[307] Structural studies are in progress on the cysteine-containing histone F_3 from chicken erythrocytes,[308] and the N-terminal sequence of histone V (F2c) from the same source has been established:[309, 310]

Thr-Glu-Ser-Leu-Val-Leu-Ser-Pro-Ala-Pro-Ala-Lys-Pro-Lys-$\frac{\text{Gln}}{\text{Arg}}$-
1 10
Val-Lys-Ala-Ser-Arg-Arg-Ser-Ala-Ser-His-Pro-Thr-Tyr-Ser-Glu-
20 30
Met-Ile-Ala-Ala-Ala-Ile-Arg-

There is what appears to be an allelic interchange Gln/Arg at position 15.[309] Nonetheless, the relative lack of heterogeneity observed in all these studies of the primary structure of histones reinforces the suspicion that histones have little or no control function in chromatin.

The lysine residues of histones can be converted into homocitrulline in a non-enzymic reaction with carbamyl phosphate [311] but the biological significance remains obscure. A single tyrosine residue is nitrated in histone F1 from calf thymus on treatment with tetranitromethane and the extent of reaction is reduced when the protein is complexed with DNA.[312] Studies with synchronized mammalian cells [313] show that differential methylation of the various histone fractions occurs and that the methylation does not coincide with periods of histone and DNA biosynthesis.

[306] T. A. Langan, S. C. Rall, and R. D. Cole, *J. Biol. Chem.*, 1971, **246**, 1942.
[307] T. Senshu, *Biochim. Biophys. Acta*, 1971, **243**, 323.
[308] W. F. Brandt and C. von Holt, *F.E.B.S. Letters*, 1971, **14**, 338.
[309] P. J. Greenaway and K. Murray, *Nature New Biology*, 1971, **229**, 233.
[310] P. J. Greenaway, *Biochem. J.*, 1971, **124**, 319.
[311] G. Ramponi, J. L. Leaver, and S. Grisolia, *F.E.B.S. Letters*, 1971, **16**, 311.
[312] M. Bustin, *Biochim. Biophys. Acta*, 1971, **251**, 172.
[313] G. R. Shepherd, J. M. Hardin, and B. J. Noland, *Arch. Biochem. Biophys.*, 1971, **143**, 1.

A new histone has been isolated from rainbow trout:[314] it contains no aromatic or sulphur-containing amino-acids and proline is the *N*-terminal residue. During trout spermatogenesis, extensive acetylation of histone occurs and, in histone IV, the acylation is limited to lysine residues at positions 5, 8, 12, and 16.[315] Transformation occurs by replacement of the histones with protamine, and proteolysis appears to play an important role in the removal of histone from the chromatin in the transformation process.[316] Thynnin, the protamine of tuna fish, has been separated into four fractions: all have N-terminal proline and the C-terminal sequence -Arg-Arg-Arg-Arg.[317]

Various papers have been concerned with non-histone proteins of the nucleus. The separation of such proteins from rat liver nuclei by isoelectric focusing in polyacrylamide gels [318] and from mouse chromatin by SDS-gel electrophoresis [319] has been described, and it has been suggested [320] that there may be a single species of chromosomal RNA-binding protein in rat ascites tumour. It is likely that chromosomal RNA is bound to protein through its dihydropyrimidine groups.[320]

F. Ribosomal Proteins.—It is a measure of the growing technical resource of the subject that protein chemists (well, some of them) are prepared to tackle structures as complex as the ribosome. Clearly this problem will not be solved overnight but a growing list of papers documents the early progress. Little strokes fell great oaks.[321]

Polyacrylamide gel electrophoresis in two dimensions has been used to effect an almost complete separation of the proteins from rabbit reticulocyte ribosomes.[157, 322] If SDS is included in the buffer for the second dimension, an estimate of the molecular weight of the protein is also obtainable (see also last year's Report). Differences in the size and number of proteins in 70s and 80s ribosomes have also been demonstrated using SDS-gel electrophoresis:[156] the 80s ribosome contains a greater number of proteins and a greater proportion of these proteins are of high molecular weight.

The proteins from the ribosomes of *E. coli* are receiving the most systematic investigation. Thirty-three pure proteins have been obtained from 50 g of ribosomes using a new method of fractionation that involves treatment with lithium chloride in the presence and absence of urea rather than zonal centrifugation.[323] Details have been given of the purification of all

[314] D. T. Wigle and G. H. Dixon, *J. Biol. Chem.*, 1971, **246**, 5636.
[315] E. P. M. Candido and G. H. Dixon, *J. Biol. Chem.*, 1971, **246**, 3182.
[316] K. Marushige and G. H. Dixon, *J. Biol. Chem.*, 1971, **246**, 5799.
[317] G. Bretzel, *Z. physiol. Chem.*, 1971, **352**, 1025.
[318] M. Gronow and G. Griffiths, *F.E.B.S. Letters*, 1971, **15**, 340.
[319] A. J. MacGillivray, D. Carroll, and J. Paul, *F.E.B.S. Letters*, 1971, **13**, 204.
[320] R. A. Jacobson and J. Bonner, *Arch. Biochem. Biophys.*, 1971, **146**, 557.
[321] Benjamin Franklin, 'Poor Richard's Almanac'.
[322] H. Welfle, J. Stahl, and H. Bielka, *Biochim. Biophys. Acta*, 1971, **243**, 416.
[323] E. Kaltschmidt, V. Rudloff, H. G. Janda, M. Cech, K. Nierhaus, and H. G. Wittmann, *Z. physiol. Chem.*, 1971, **352**, 1545.

twenty-one proteins from the 30s subunit [324] but the mixture of proteins from the 50s subunit is more complex and more difficult to fractionate.[325] Immunological studies show [326] that the twenty-one proteins from the 30s subunit and twenty-four proteins from the 50s subunit are all different in amino-acid sequence but that two have some structural homology. A preliminary study [327] of the tryptic peptides of three proteins from the 30s subunit is in accord with this observation.

Few of the protein thiol groups are reactive in the intact *E. coli* ribosome [328] and a study of the digestibility of the ribosome with trypsin suggests that the 30s subunit may be a less compact structure than the 50s subunit.[329] Digestion of the 30s subunit with ribonuclease T_1 can be used to produce a specific ribonucleoprotein fragment and it was reported that the mild anionic detergent Sarkosyl-L (*N*-laurylsarcosine) is preferable to SDS for gel electrophoresis of the proteins in the fragment.[330] The curious enzyme leucyl,phenylalanyl-tRNA protein transferase of *E. coli* catalyses the transfer of these amino-acids from the corresponding tRNA to the N-terminal arginine residue of a specific protein in the 30s subunit.[331] Photo-oxidation of ribosomes in the presence of Rose Bengal affects the binding sites for tRNA and poly-U, perhaps by modification of histidine in a single protein.[332] The 16s RNA is also modified.

The initiation factor F_1 from *E. coli* ribosomes has been isolated and crystallized:[333] it has a molecular weight of 9400, N-terminal alanine, and C-terminal lysine.

G. Serum and Egg Proteins.—The amino-acid sequence around the pyridoxal-phosphate-binding lysine (*) of bovine serum albumin has been reported and then amended slightly.[334]

The sequence is:

-Ser-Leu-Gln-Phe-Lys-Pro-$\overset{*}{\text{Lys}}$-Lys-

Ten defined fragments, one of which is probably derived from the C-terminus of the protein, have been obtained from peptic digests of bovine serum albumin.[335] Not surprisingly, the protein is more readily digested

[324] I. Hindennach, G. Stöffler, and H. G. Wittmann, *European J. Biochem.*, 1971, **23**, 7.
[325] I. Hindennach, E. Kaltschmidt, and H. G. Wittmann, *European J. Biochem.*, 1971, **23**, 12.
[326] G. Stöffler and H. G. Wittmann, *J. Mol. Biol.*, 1971, **62**, 407; *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2283.
[327] B. Wittmann-Liebold, *Z. physiol. Chem.*, 1971, **352**, 1705.
[328] L. I. Slobin, *J. Mol. Biol.*, 1971, **61**, 281.
[329] F. N. Chang and J. G. Flaks, *J. Mol. Biol.*, 1971, **61**, 387.
[330] R. Brimacombe, J. Morgan, D. G. Oakley, and R. A. Cox, *Nature New Biol.*, 1971, **231**, 209.
[331] M. J. Leibowitz and R. L. Soffer, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1866.
[332] H. F. Noller, C. Chang, G. Thomas, and J. Aldridge, *J. Mol. Biol.*, 1971, **61**, 669.
[333] S. Lee-Huang, M. A. G. Sillero, and S. Ochoa, *European J. Biochem.*, 1971, **18**, 536.
[334] J. A. Anderson, H. W. Chang, and C. J. Grandjean, *Biochemistry*, 1971, **10**, 2408, 3810.
[335] W. G. M. Braam, B. J. M. Harmsen, J. A. L. I. Walters, and G. A. J. van Os, *Internat. J. Protein Res.*, 1971, **3**, 271.

if it is unfolded.[336] In a separate study, seven cyanogen bromide fragments that together account for the whole molecule have been isolated from human serum albumin.[337]

An interchange Arg/Gln has been reported [338] to distinguish the two major genetically transmitted variants of human α-acid glycoprotein (orosomucoid). There is some evidence for duplication in the evolution of the structural genes for transferrin from a variety of vertebrates [339] and for the participation of histidine residues in the binding of iron.[340] It has also been reported [341] that apoferritin (from horse spleen) is in general digested more rapidly and to a greater extent by various proteolytic enzymes than is ferritin, indicating that conformational changes take place when iron is bound.

A study of the cyanogen bromide fragments of ovotransferrin [342] suggests that the molecule (m. wt. 80 000) may contain a duplicated amino-acid sequence, which is not easy to reconcile with earlier work on the thiol sequences.[343] Of the two types of hen egg phosphvitin only one contains methionine (1 residue per chain of m. wt. 34 000), and cyanogen bromide has been used to cleave the molecule into two fragments.[344] A partial structure of the carbohydrate moiety of phosphvitin has been published [345] and the carbohydrate has been shown to be attached to the protein in the sequence:[346]

$$\begin{array}{l}\text{-Ser-Asn-Ser-Gly-[Ser(P)]}_8\text{-Arg-Ser-Val-Ser-His-His-}\\ \qquad\qquad\quad |\\ \text{carbohydrate}\end{array}$$

The occurrence of eight serine phosphate residues in a row is worth remarking.

The primary structure of the 33-residue peptide from plakalbumin has now been established, which enables the sequence of 49 residues at the C-terminus of ovalbumin to be formulated:[347]

-Arg-Gly-Arg-Glu-Val-Val-Gly-Ser-Ala-Glu-Ala-Gly-(Val,Asp)-Ala-Ala-Ser-Val-Ser-Glu-Glu-Phe-Arg-Ala-Asp-His-Pro-Phe-Leu-Phe-Cys-Ile-Lys-His-Ile-Ala-Thr-Asn-Ala-Val-Leu-Phe-Phe-Gly-Arg-Cys-Val-Ser-Pro

[336] W. G. M. Braam, B. J. M. Harmsen, and G. A. J. van Os, *Biochim. Biophys. Acta*, 1971, **236**, 99.
[337] R. H. McMenamy, H. M. Dintzis, and F. Watson, *J. Biol. Chem.*, 1971, **246**, 4744.
[338] R. Nimberg, T. Motoyama, and K. Schmid, *J. Biol. Chem.*, 1971, **246**, 5817.
[339] R. M. Palmour and H. E. Sutton, *Biochemistry*, 1971, **10**, 4026.
[340] A. Bezkorovainy and D. Grohlich, *Biochem. J.*, 1971, **123**, 125.
[341] R. R. Crichton, *Biochim. Biophys. Acta*, 1971, **229**, 75.
[342] J. L. Phillips and P. Azari, *Biochemistry*, 1971, **10**, 1160.
[343] T. C. Elleman and J. Williams, *Biochem. J.*, 1970, **116**, 515.
[344] R. C. Clark and F. J. Joubert, *F.E.B.S. Letters*, 1971, **13**, 225.
[345] R. Shainkin and G. E. Perlmann, *Arch. Biochem. Biophys.*, 1971, **145**, 693.
[346] R. Shainkin and G. E. Perlmann, *J. Biol. Chem.*, 1971, **246**, 2278.
[347] E. O. P. Thompson, R. W. Sleigh, and M. B. Smith, *Austral. J. Biol. Sci.*, 1971, **24**, 525.

The complete amino-acid sequence of the protein subunit of the tetrameric avidin from egg-white has been determined (Figure 2).[348] Carbohydrate is attached at position 17 and half the molecules have isoleucine at position 34 whereas the other half have threonine. There is no obvious similarity to

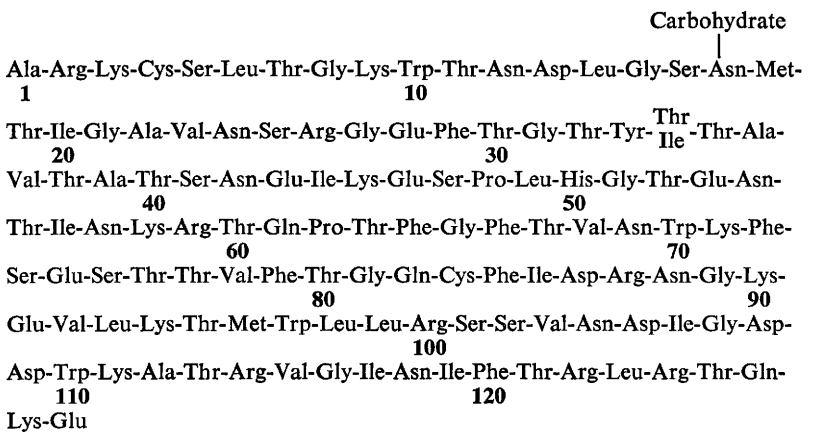


Figure 2 *The amino-acid sequence of avidin from hen egg-white*

lysozyme or α-lactalbumin, making it extremely unlikely that these three proteins had a common genetic ancestor. Two points of technique in the sequence analysis are noteworthy. First, trouble was again found with the cyanogen bromide cleavage of Met–Ser and Met–Thr bonds despite quantitative conversion of the methionine into a homoseryl residue (see ref. 256 above and last year's Report). Secondly, with several peptides containing N-terminal arginine or histidine it was noted that a substantial amount of the penultimate (*i.e.* second) residue is released (presumably as the thiazolinone) in the first step of the Edman degradation, which can lead to erroneous assignment of amino-acid sequence. (Problems such as this lead one to doubt that the automatic sequenator can ever entirely take over as the sole method of sequence analysis in proteins – which is not to deny its usefulness in other circumstances.)

Structural work has also been reported on the eggshell of the oriental garden cricket *Gryllus mitratus*.[349] One major protein was isolated (m. wt. 57 500), rich in serine and *O*-phosphoserine.

H. Miscellaneous.—*Crystallins.* A simple chromatographic method for preparing the polypeptide chains of α-crystallin has been described [350] and the N-terminal amino-acid sequence of both acidic (A1 and A2) peptide

[348] R. J. DeLange and T.-S. Huang, *J. Biol. Chem.*, 1971, **246**, 698.
[349] H. Kawasaki, H. Sato, and M. Suzuke, *Biochem. J.*, 1971, **125**, 495.
[350] G. J. van Kamp, H. J. Hoenders, and H. Bloemendal, *Biochim. Biophys. Acta*, 1971, **243**, 149.

chains of ox α-crystallin has been established:[351]

Ac-Met-Asp-Ile-Ala-Ile-Gln-His-Pro-Trp-Phe-Lys-Arg-

The amino-acid analyses of these chains appear to fit a molecular weight of 12 000 but this value cannot be confirmed by SDS-gel electrophoresis.[351] The subunits from calf lens α-crystallin all have molecular weights of 19 500 or 22 500 and again there is no evidence for smaller chains although three distinct populations of macromolecules were observed, apparently not in equilibrium with one another.[352] One of these subunits has been obtained sufficiently pure for structural studies and its N-terminal sequence has been shown to be:[353]

Ac-Met-Asp-Ile-Ala-Ile-Ser-His-Pro-Trp-Ile-Arg-Pro-Ser-Phe-Phe-Glu-Phe-His-

The molecule was cleaved at the only other methionine residue with cyanogen bromide, and the N-terminal sequence of the second fragment was shown to be:[353]

Ser-Leu-Thr-Lys-Asp-Phe-Asp-Glu-Val-Asn-Ile-Asp-Val-Ser-His-Phe-

Both sequence analyses were carried out in a sequenator. It was also suggested [353] that the acidic and basic polypeptide chains that make up α-crystallin have evolved from a common ancestral form. Other experiments [354] have been concerned with the thiol groups in the acidic chains of calf lens α-crystallin. It is likely that the chains (m. wt. estimated at *ca.* 20 000) contain two methionine residues and a single cysteine residue; the cysteine residue is in the sequence:[354]

-Tyr-Arg-Leu-Pro-Ser-Asn-Val-Asp-Glu-Ser-Ala-Leu-Ser-Cys-Ser-Leu-Ser-Ala-Asp-Gly-Met-Leu-Thr-Phe-Ser-Gly-Pro-Lys-

On the other hand, eight thiol groups can be distinguished in ox γ-crystallin on the basis of the analysis of cysteic-acid-containing peptides after performic acid oxidation,[355] and the C-terminal sequence of the protein has been shown to be -Val-Met-Asp-Phe-Tyr.[356] The results indicate that the γ-crystallin is homogeneous and has a molecular weight of 19 200. In general, therefore, the weight of evidence is now decidedly against the existence of crystallins with molecular weights of only *ca.* 10 000.

Keratins and Wool Proteins. The amino-acid sequence of component C from the high-sulphur proteins of α-keratin (Lincoln wool) has been established

[351] A. E. Leon, J. J. T. Gerding, K. de Groot, H. J. Hoenders, and H. Bloemendal, *Internat. J. Protein Res.*, 1971, **3**, 19.
[352] A. Spector, L.-K. Li, R. C. Augusteyn, A. Schneider, and T. Freund, *Biochem. J.*, 1971, **125**, 337.
[353] R. C. Augusteyn and A. Spector, *Biochem. J.*, 1971, **124**, 345.
[354] P. H. Corran and S. G. Waley, *Biochem. J.*, 1971, **124**, 61.
[355] L. R. Croft and S. G. Waley, *Biochem. J.*, 1971, **121**, 453.
[356] L. R. Croft, *Biochem. J.*, 1971, **121**, 557.

(Figure 3).[357] The protein contains many repetitions of the 'same' sequence of about ten residues, and it was suggested that it may be the product of many gene duplication events. Further, it was pointed out that other structural proteins such as silk fibroin and collagen contain repetitive sequences and

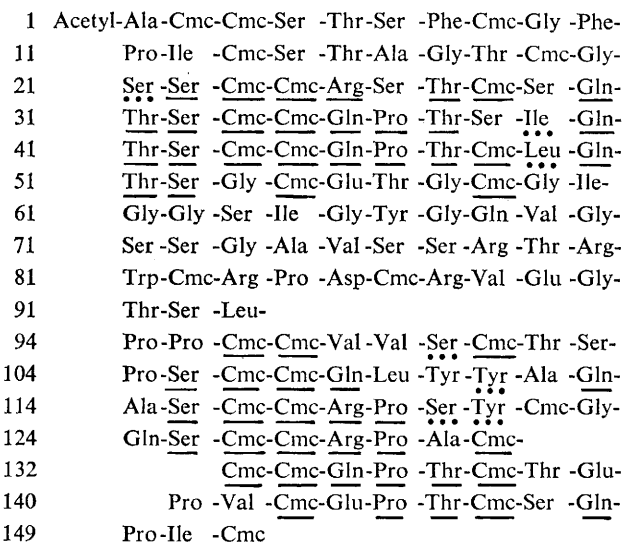


Figure 3 *The complete amino-acid sequence of component* C *of the high-sulphur fraction of Lincoln wool. Homologous residues are underlined, and the sequence has been arranged to show maximum homology*
(Reproduced by permission from *Nature New Biol.*, 1971, **234**, 148)

that this feature may therefore be structurally significant for the wool proteins. The amino-acid sequences of two high-sulphur proteins (98 residues) from reduced Merino wool have also been compared and shown to differ in only four positions.[358]

The presence of N^{ε}-(γ-glutamyl)-lysine cross-links has been demonstrated in the citrulline-containing proteins of hair,[359] and N^{ε}-(β-aspartyl)-lysine cross-links have been reported in native and heated keratin.[360] Dityrosine has been detected in silk fibroin and, in smaller amounts, in keratin, where it is located in the crystalline regions of the fibre rather than the amorphous matrix.[361] The content of aspartic acid, asparagine, glutamic acid, and

[357] T. C. Elleman, *Nature New Biology*, 1971, **234**, 148.
[358] T. Haylett, L. S. Swart, and D. Parris, *Biochem. J.*, 1971, **123**, 191; L. S. Swart and T. Haylett, *ibid.*, 1971, **123**, 201.
[359] H. W. J. Harding and G. E. Rogers, *Biochemistry*, 1971, **10**, 624.
[360] R. S. Asquith, M. S. Otterburn, and K. L. Gardner, *Experientia*, 1971, **27**, 1388.
[361] D. J. Raven, C. Earland, and M. Little, *Biochim. Biophys. Acta*, 1971, **251**, 96.

glutamine in wool proteins has been determined by digesting the *S*-carboxymethylated protein successively with Pronase, Prolidase, and leucine aminopeptidase.[362] The accuracy of the technique was confirmed by carrying out the same estimation for ribonuclease and insulin.

Breed and species differences have been reported for the high-sulphur proteins of wools from various domestic, feral, and wild sheep,[363] and the proteins of normal human hair have been shown to be significantly different from those of mentally retarded siblings.[364] However, since similar differences were found for hair from some individuals not mentally retarded, it is probable that there is only a fortuitous connection between the hair defect and mental retardation in the case described.[364] A blocked N-terminal peptide, Ac-Ser-Cmc-Tyr, has been isolated from chymotryptic digests of keratin from goose feather calamus.[365] It is thought likely that it is the only N-terminal peptide.

Casein. The complete amino-acid sequence of the B variant of ox α_{S1}-casein has now been established and the distinguishing features of the A, C, and D variants have been determined (Figure 4).[366] It has also been reported that the N-terminal sequence of ox β-casein A_2 is:[367]

Arg-Glu-Leu-Glu-Glu-Leu-Asn-Val-Pro-Gly-Glu-Ile-Val-Glu-Ser(P)-Leu-Ser(P)-Ser(P)-Ser(P)-Glu-Glu-Ser-Ile-Thr-Arg-Ile-Asn-Lys-Lys-Ile-Glu-Lys-

and that the *C*-terminal sequence is:[368]

-His-Gln-Pro-His-Gln-Pro-Leu-Pro-Pro-Thr-Val-Met-Phe-Pro-Pro-Gln-Ser-Val-Leu-Ser-Leu-Ser-Gln-Ser-Lys-Val-Leu-Pro-Val-Pro-Glu-Lys-Ala-Val-Pro-Tyr-Pro-Gln-Arg-Asp-Met-Pro-Ile-Gln-Ala-Phe-Leu-Leu-Tyr-Gln-Gln-Pro-Val-Leu-Gly-Pro-Val-Arg-Gly-Pro-Phe-Pro-Ile-Ile-Val

Further work is in progress on the structure [369] of β-casein A_2, and the same N-terminal sequence of 25 residues has been reported for ox β-casein A_1.[370]

Other Proteins. The molecular weight of the coat protein of alfalfa mosaic virus has been redetermined and shown to be 24 500.[371] The primary structure of the coat protein of the RNA bacteriophage Qβ has been

[362] L. A. Holt, B. Milligan, and C. M. Roxburgh, *Austral. J. Biol. Sci.*, 1971, **24**, 509.
[363] R. L. Darskus and J. M. Gillespie, *Austral. J. Biol. Sci.*, 1971, **24**, 515.
[364] R. J. Pollitt and P. D. Stonier, *Biochem. J.*, 1971, **122**, 433.
[365] I. J. O'Donnell, *Austral. J. Biol. Sci.*, 1971, **24**, 179.
[366] J.-C. Mercier, F. Grosclaude, and B. Ribadeau-Dumas, *European J. Biochem.*, 1971, **23**, 41.
[367] B. Ribadeau-Dumas, G. Brignon, F. Grosclaude, and J.-C. Mercier, *European J. Biochem.*, 1971, **20**, 264.
[368] B. Ribadeau-Dumas, G. Brignon, F. Grosclaude, and J.-C. Mercier, *European J. Biochem.*, 1971, **20**, 258.
[369] G. Brignon, B. Ribadeau-Dumas, F. Grosclaude, and J.-C. Mercier, *European J. Biochem.*, 1971, **22**, 179.
[370] W. Manson and W. D. Annan, *Arch. Biochem. Biophys.*, 1971, **145**, 16.
[371] J. Kruseman, B. Kraal, E. M. J. Jaspars, J. F. Bol, F. Th. Brederode, and H. Veldstra, *Biochemistry*, 1971, **10**, 447.

established and compared with that of bacteriophage f2,[372] as explained in last year's Report. It has also been reported [373] that another Qβ-specific protein, *viz.* A_1, of m. wt. 36 000, has an N-terminal sequence which is identical (for eight residues) with that of the Qβ coat protein.[373]

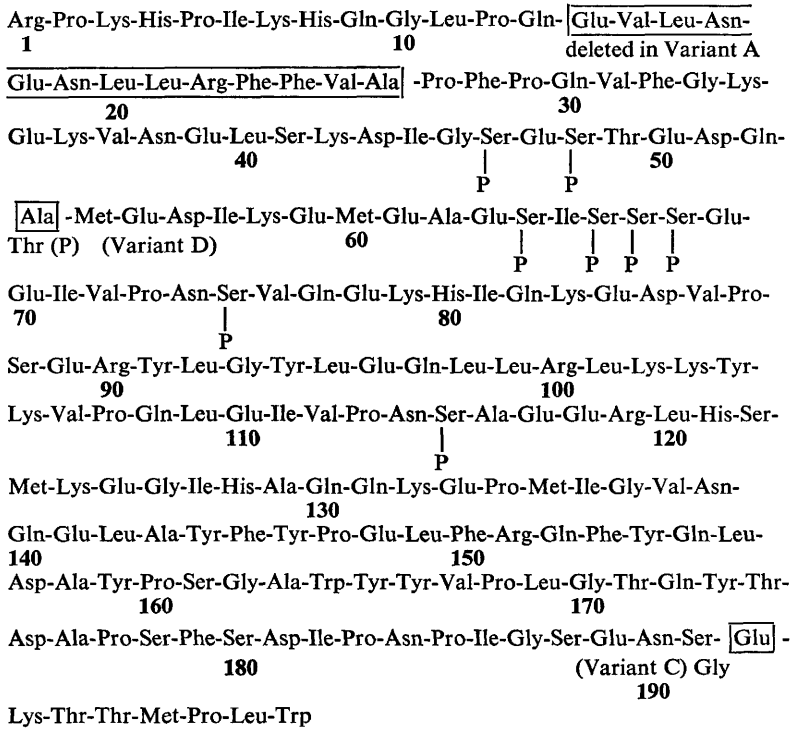


Figure 4 *The amino-acid sequence of bovine α_{S1}-casein. The differences found for the genetic variants A, C, and D are marked in boxes*

Further work has been described [374] on the freezing-point-depressing glycoproteins of the antarctic fish *Trematomus borchgrevenki.* They are shown to contain the tripeptide -Ala-Ala-Thr- in a repeating sequence, the threonine being in glycosidic linkage with the disaccharide galactosyl-*N*-acetylgalactosamine. Even more unusual perhaps are the polymers of arginine and aspartic acid (m. wt. 25 000—100 000) which occur in granules in the blue-green alga *Anabaena cylindrica* and are thought to be storage proteins.[375]

[372] T. Maita and W. Konigsberg, *J. Biol. Chem.*, 1971, **246**, 5003.
[373] A. M. Weiner and K. Weber, *Nature*, 1971, **234**, 206.
[374] A. L. DeVries, J. Vandenheede, and R. E. Feeney, *J. Biol. Chem.*, 1971, **246**, 305.
[375] R. D. Simon, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 265.

4 Peptides and Hormones

A. Pancreatic Hormones.—The isolation of an enzyme that converts proinsulin into insulin has been reported.[376] Details of the structural analysis of ox proinsulin have been given and some forms of the hormone intermediate between proinsulin and insulin were isolated from commercial crystalline insulin.[377] The C-peptide of ox proinsulin has also been purified from pancreas and shown to be identical with that obtained from intact proinsulin.[378] The primary structure of the C-peptide reported in all these papers is slightly different (an inversion of two residues) from that formulated in the first investigation of ox proinsulin. A comparative study [379] of the C-peptides of proinsulin from man, pig, and ox reveals that they show interesting variations in length:

Man	**Glu**-Ala-**Glu**-Asp-Leu-**Gln**-Val-Gly-Gln-Val-**Glu-Leu**-Gly-Gly-Gly-
Pig	**Glu**-Ala-**Glu**-Asn-Pro-**Gln**-Ala-Gly-Ala-Val-**Glu-Leu**-Gly-Gly-Gly-
Ox	**Glu**-Val-**Glu**-Gly-Pro-**Gln**-Val-Gly-Ala-Leu-**Glu-Leu**-Ala-Gly-Gly-

Man	Pro-**Gly**-Ala-Gly-Ser-Leu-Gln-Pro-Leu-Ala-**Leu-Glu-Gly**-Ser-Leu-**Gln**
Pig	Leu-**Gly**-Gly - - Leu-Gln-Ala-Leu-Ala-**Leu-Glu-Gly**-Pro-Pro-**Gln**
Ox	Pro-**Gly**-Ala-Gly - - - - - Gly-**Leu-Glu-Gly**-Pro-Pro-**Gln**

This in turn implies some degree of freedom in the structural requirements of the C-peptide region despite its function in the folding of the proinsulin molecule.[379]

The mouse, like the rat, has been reported to have two different insulins:[380] the molecules are distinguished by a Pro/Ser interchange at position B-9 and a Lys/Met interchange at position B-29. Chemical modification of insulin with phenylglyoxal [381] and a study of semisynthetic insulins [382] both indicate the functional importance of arginine at position B-22. On the other hand, a detailed analysis of the acetylation of insulin [383] showed that acetylation of any of the hormone amino-groups did not affect the biological activity despite changes in the immunological specificity.

The amino-acid sequence of ox glucagon has been shown [384] to be identical with that of pig, first established in 1957.

[376] C. C. Yip, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1312.
[377] C. Nolan, E. Margoliash, J. D. Peterson, and D. F. Steiner, *J. Biol. Chem.*, 1971, **246**, 2780.
[378] (*a*) A. Salokangas, D. G. Smyth, J. Markussen, and F. Sundby, *European J. Biochem.*, 1971, **20**, 183; (*b*) D. F. Steiner, S. Cho, P. E. Oyer, S. Terris, J. D. Peterson, and A. H. Ruben, *J. Biol. Chem.*, 1971, **246**, 1365.
[379] A. S. C. Ko, D. G. Smyth, J. Markussen, and F. Sundby, *European J. Biochem.*, 1971, **20**, 190.
[380] J. Markussen, *Internat. J. Protein Res.*, 1971, **3**, 149.
[381] H. F. Bünzli and H. R. Bosshard, *Z. physiol. Chem.*, 1971, **352**, 1180.
[382] G. Weitzel, U. Weber, J. Martin, and K. Eisele, *Z. physiol. Chem.*, 1971, **352**, 1005.
[383] D. G. Lindsay and S. Shall, *Biochem. J.*, 1971, **121**, 737.
[384] W. W. Bromer, M. E. Boucher, and J. E. Koffenberger, jun., *J. Biol. Chem.*, 1971, **246**, 2822.

B. Pituitary Hormones.—The pituitary hormones have been the subject of furious activity in the past year or so, and many remarkable structural interrelations are coming to light. The purification of human chorionic gonadotropin (CG), has been described [385] and the resolution of the two subunits of luteinizing hormone (LH) from human,[386] pig,[387] and rat [388] pituitaries and of follicle-stimulating hormone (FSH) from human pituitaries [389] has

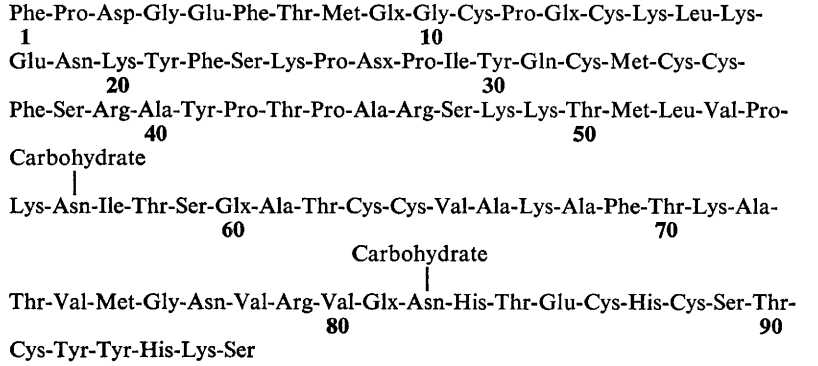


Figure 5 *The amino-acid sequence of the α-subunit of ox TSH*

been reported. Hybridization experiments with CG, LH, FSH, and thyroid-stimulating hormone (TSH) suggest that each molecule contains a hormone-specific β-chain and an α-chain that is similar if not identical in all the hormones.[386, 389, 390] For example, the β-chain of ox TSH together with the α-chain of human CG forms a hormone that is almost as good in thyroid-stimulating activity as TSH itself.[390]

Compelling evidence in support of this idea comes from studies of the primary structure of the hormones. Thus, the amino-acid sequences of the α- and β-chains of ox TSH have been determined (Figures 5 and 6) [391] and it is likely that the α-chains of ox TSH and LH are identical apart from differences in carbohydrate content.[392] The TSH β-chain is found to lack the C-terminal methionine residue in some molecules, and the α-chain is also heterogeneous in that it can lack the Phe-Pro- sequence at the N-terminus.[391] The arrangement of the disulphide bridges remains to be

[385] R. Brossmer, M. Dörner, U. Hilgenfeldt, F. Leindenberger, and E. Trude, *European J. Biochem.*, 1971, **15**, 33.
[386] P. Rathnam and B. B. Saxena, *J. Biol. Chem.*, 1971, **246**, 7087.
[387] G. Hennen, Z. Prusik, and G. C. Maghuin-Rogister, *European J. Biochem.*, 1971, **18**, 376.
[388] D. N. Ward, L. E. Reichart, B. A. Fitzak, H. S. Nahm, C. M. Sweeney, and J. D. Neill, *Biochemistry*, 1971, **10**, 1796.
[389] B. B. Saxena and P. Rathnam, *J. Biol. Chem.*, 1971, **246**, 3549.
[390] J. G. Pierce, O. P. Bahl, J. S. Cornell, and N. Swaminathan, *J. Biol. Chem.*, 1971, **246**, 2321.
[391] T.-H. Liao and J. G. Pierce, *J. Biol. Chem.*, 1971, **246**, 850.
[392] J. G. Pierce, T.-H. Liao, R. B. Carlsen, and T. Reimo, *J. Biol. Chem.*, 1971, **246**, 866.

Phe-Cys-Ile-Pro-Thr-Glu-Tyr-Met-Met-His-Val-Glu-Arg-Lys-Glu-Cys-Ala-Tyr-
1 **10**

Carbohydrate
|
Cys-Leu-Thr-Ile-Asn-Thr-Thr-Val-Cys-Ala-Gly-Tyr-Cys-Met-Thr-Arg-Asx-Val-
20 **30**
Asx-Gly-Lys-Leu-Phe-Leu-Pro-Lys-Tyr-Ala-Leu-Ser-Gln-Asp-Val-Cys-Thr-Tyr-
40 **50**
Arg-Asp-Phe-Met-Tyr-Lys-Thr-Ala-Glu-Ile-Pro-Gly-Cys-Pro-Arg-His-Val-Thr-
60 **70**
Pro-Tyr-Phe-Ser-Tyr-Pro-Val-Ala-Ile-Ser-Cys-Lys-Cys-Gly-Lys-Cys-Asx-Thr-
80 **90**
Asx-Tyr-Ser-Asx-Cys-Ile-His-Glu-Ala-Ile-Lys-Thr-Asn-Tyr-Cys-Thr-Lys-Pro-
100
Gln-Lys-Ser-Tyr-Met
110

Figure 6 *The amino-acid sequence of the β-subunit of ox TSH*

Carbohydrate
|
R-Ser-Arg-Gly-Pro-Leu-Arg-Pro-Leu-Cys-Glx-Pro-Ile-Asn-Ala-Thr-Leu-Ala-
1 **10**
Ala-Glx-Lys-Glx-Ala-Cys-Pro-Val-Cys-Ile-Thr-Phe-Thr-Thr-Ser-Ile-Cys-Ala-Gly-
20 **30**
Tyr-Cys-Pro-Ser-Met-Lys-Arg-Val-Leu-Pro-Val-Ile-Leu-Pro-Pro-Met-Pro-Glx-
40 **50**
Arg-Val-Cys-Thr-Tyr-His-Glu-Leu-Arg-Phe-Ala-Ser-Val-Arg-Leu-Pro-Gly-Cys-
60 **70**
Pro-Pro-Gly-Val-Asx-Pro-Met-Val-Ser-Phe-Pro-Val-Ala-Leu-Ser-Cys-(His,
80
Cys, Gly, Pro)-Cys-Arg-Leu-Ser-Ser-Thr-Asx-Cys-Gly-Pro-Gly-Arg-Thr-Glx-
90 **100**
Pro-Leu-Ala-Cys-Asx-His-Pro-Pro-Leu-Pro-Asp-Ile-Leu
110

Figure 7 *The amino-acid sequence of the β-subunit of ox LH. R is probably acetyl*

determined. These conclusions are further borne out by the analysis of the amino-acid sequence of the α- and β-subunits of sheep LH.[393] The α-subunits of ox TSH [391] and sheep LH [393] are identical apart from an inversion of residues 88 and 89, which is possibly simply the result of an error in one of the two investigations. Unfortunately, the results of independent determinations of the amino-acid sequences of the β-subunit of LH from ox (Figure 7) [394] and sheep [393] do not agree so well, since the differences between the versions are almost certainly more than can be accounted for by the difference in species of origin. The discrepancies are

[393] H. Papkoff, M. R. Sairam, and C. H. Li, *J. Amer. Chem. Soc.*, 1971, **93**, 1531.
[394] G. C. Maghuin-Rogister and A. Dockier, *F.E.B.S. Letters*, 1971, **19**, 209.
[395] G. C. Maghuin-Rogister, J. Closset, and G. P. Hennen, *F.E.B.S. Letters*, 1971, **13**, 301.
[396] G. P. Hennen, I. Klüh, and G. C. Maghuin-Rogister, *F.E.B.S. Letters*, 1971, **19**, 207.
[397] G. C. Maghuin-Rogister and G. P. Hennen, *Eurpean J. Biochem.*, 1971, **21**, 489.

yet to be resolved. It has also been reported that the α-subunits of LH from ox and pig are very similar [395] and that there is heterogeneity in the N-terminal [396] and C-terminal [397] sequences of the ox chain. The N-terminal sequence was determined in a sequenator.

Revisions to the previously published amino-acid sequence of human growth hormone (HGH) have now been agreed [398, 399] (see last year's Report) and the sequence of ox growth hormone has been established.[400] It contains 188 residues, 61% of which are identical with those of HGH, and there is heterogeneity at position 124, 30% of the molecules having valine and 70% having leucine at this position.[400, 401a] Allelic genes appear to be responsible.[401b] The primary structure of human placental lactogen (HPL) has also been established [398, 402, 403] and compared [398, 402–404] with HGH and sheep prolactin (SP). HPL and HGH (each 190 residues) are identical for 85% of their amino-acid sequence and the homology with SP is unmistakable (see last year's Report). It is interesting that growth hormone and prolactin (secreted by the pituitary) and placental lactogen (secreted by trophoblasts during pregnancy) are all potent lactogenic hormones but that, despite the exceedingly close homology, placental lactogen does not promote growth. Further evidence that human pituitary prolactin and HGH are not the same molecule has been presented.[405] In other experiments [406] it has been shown that six out of eight tyrosine residues in HGH can react with tetranitromethane at pH 8. The modified hormone retains full lactogenic activity but only 60% activity in promoting growth. On the other hand, if the modification is carried out in 5M-guanidine hydrochloride, all eight tyrosine residues react, with complete loss of both hormonal activities.

The amino-acid sequences of two neurophysins (pituitary proteins that bind the hormones oxytocin and vasopressin) have been established. That of ox neurophysin-II [407] is shown in Figure 8. The primary structure of pig neurophysin-I [408] is somewhat different: it is found in two forms, one of

[398] H. D. Niall, *Nature New Biol.*, 1971, **230**, 90.
[399] C. H. Li and J. S. Dixon, *Arch. Biochem. Biophys.*, 1971, **146**, 233.
[400] J. A. Santomé, J. M. Dellacha, A. C. Paladini, C. E. M. Wolfenstein, C. Peña, E. Poskus, S. T. Daurat, M. J. Biscoglio, Z. M. M. de Sesé, and A. V. F. de Sangüesa, *F.E.B.S. Letters*, 1971, **16**, 198.
[401] (*a*) H. N. Fernandez, S. T. Daurat, C. Peña, J.-M. Dellacha, J. A. Santomé, and A. C. Paladini, *F.E.B.S. Letters*, 1971, **18**, 53; (*b*) B. K. Seavey, R. N. P. Singh, U. J. Lewis, and I. I. Geschwind, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 189.
[402] C. H. Li, J. S. Dixon, and D. Chung, *Science*, 1971, **173**, 56.
[403] L. M. Sherwood, S. Handwerger, W. D. McLaurin, and M. Lanner, *Nature New Biol.*, 1971, **233**, 59.
[404] T. A. Bewley and C. H. Li, *Experientia*, 1971, **27**, 1368.
[405] U. J. Lewis, R. N. P. Singh, and B. K. Seavey, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 1169.
[406] L. Ma, J. Brovetto-Cruz, and C. H. Li, *Biochim. Biophys. Acta*, 1971, **229**, 444.
[407] (*a*) R. Walter, D. H. Schlesinger, I. L. Schwartz, and J. D. Capra, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 293; (*b*) D. H. Schlesinger, J. D. Capra, I. L. Schwartz, and R. Walter, *Experientia*, 1971, **27**, 213.
[408] T. C. Wuu, S. Crumm, and M. Saffran, *J. Biol. Chem.*, 1971, **246**, 6043.

which has 91 residues whereas the other carries an additional leucine residue at the C-terminus. The disulphide bridges remain to be determined but the ox and pig proteins are clearly homologous. Unlike the ox and the pig, which appear to contain two principal neurophysins (I and II), the rat probably has only one major and one minor component,[409] but the presence of a third form of pig neurophysin has been reported.[410] Since it is likely that oxytocin and vasopressin have separate neurophysins, it has been suggested [411] that there may even be a pro-oxytocin and a provasopressin that are cleaved to yield oxytocin-neurophysin and vasopressin-neurophysin respectively, comparable with the liberation of insulin from proinsulin.

Ala-Met-Ser-Asp-Leu-Glu-Leu-Arg-Gln-Cys-Leu-Pro-Cys-Gly-Pro-Gly-Gly-Lys-
1 **10**
Gly-Arg-Cys-Phe-Gly-Pro-Ser-Ile-Cys-Cys-Gly-Asn-Glu-Leu-Gly-Gln-Phe-Val-
20 **30**
Gly-Thr-Ala-Glu-Ala-Leu-Arg-Cys-Gln-Glu-Glu-Asn-Tyr-Leu-Pro-Ser-Pro-Cys-
40 **50**
Gln-Ser-Gly-Gln-Arg-Pro-Cys-Gly-Ser-Gly-Gly-Arg-Cys-Ala-Ala-Ala-Thr-Ile-
60 **70**
Cys-Cys-Ser-Asp-Glu-Glu-Cys-Val-Pro-Asp-Glu-Gln-Val-Lys-Pro-Gly-Gly-
80
Arg-Gly-Gly-Cys-Phe-Cys-Arg-Val
90

Figure 8 *The amino-acid sequence of ox neurophysin II*

Various papers have dealt with the characterization of a peptide from the hypothalamus that stimulates secretion of LH and FSH from the pituitary. It has been described in pig,[412, 413] sheep,[414] and ox [413] and, following the recognition that the hormone contains tryptophan, [412b, 413] the amino-acid sequence of the pig peptide has been established:[415]

Glp-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH_2

Various chemical modification experiments show [416] that tyrosine, histidine, tryptophan, and arginine are important for the biological activity, and the

[409] A. Norström, J. Sjöstrand, B. G. Livett, L. O. Uttenthal, and D. B. Hope, *Biochem. J.* 1971, **122**, 671.
[410] K. W. Cheng and H. G. Friesen, *J. Biol. Chem.*, 1971, **246**, 7656.
[411] G. D. Burford, C. W. Jones, and B. T. Pickering, *Biochem. J.*, 1971, **124**, 809.
[412] (*a*) A. V. Schally, A. Arimura, Y. Baba, R. M. G. Nair, H. Matsuo, T. W. Redding, L. Debeljuk, and W. F. White, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 393; (*b*) A. V. Schally, R. M. G. Nair, T. W. Redding, and A. Arimura, *J. Biol. Chem.*, 1971, **246**, 7230.
[413] C. Bogentoft, B. L. Currie, H. Sievertsson, J.-K. Chang, K. Folkers, and C. Y. Bowers, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 403.
[414] M. Amoss, R. Burgus, R. Blackwell, W. Vale, R. Fellows, and R. Guillemin, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 205.
[415] Y. Baba, H. Matsuo, and A. V. Schally, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 459.
[416] Y. Baba, A. Arimura, and A. V. Schally, *J. Biol. Chem.*, 1971, **246**, 7581.

synthetic peptide Glp-Tyr-Arg-Trp-NH_2 has been found to possess the power of LH-release.[417] It is interesting that the peptide Glp-Trp-Tyr-Arg-NH_2 is inactive despite the fact that the sequence of residues is the same as that in the native hormone.[417]

It has been observed that, during incubation with human plasma, TSH-releasing hormone, Glp-His-Pro-NH_2, is inactivated as a result of loss of the amide group at the C-terminus.[418] A peptide from pig hypothalamus that shows growth-hormone-releasing activity has also been isolated, and shown to have the amino-acid sequence:[419]

Val-His-Leu-Ser-Ala-Glu-Glu-Lys-Glu-Ala

C. Other Hormones and Peptides.—The preparation of thyroglobulin that contains no sialic acid has been described [420] and heterogeneity in the binding of sialic acid has been demonstrated.[421] A study of the tryptic peptides of non-iodinated pig thyroglobulin suggests [422] that the molecule contains $2n$ peptide chains, where n is an integer awaiting determination. Ox thyroglobulin has been shown [423] to be capable of accepting arginine from Arg-tRNA in an enzymic reaction catalysed by a transferase (see also ref. 331) to give the new N-terminal sequence Arg-Asp-Ile-Phe-. It is suggested [423] that ox thyroglobulin is composed of two chains each with the N-terminal sequence Asp-Ile-Phe-. Further work on the transferase indicates [424] that for acceptor activity a protein needs to have aspartic or glutamic acid as N-terminal residue.

The isolation of three ox parathyroid hormones has been described.[425] All contain 84 residues and the amino-acid sequence of the major form was given in last year's Report. The isolation of pig parathyroid hormone has also been reported [426] and the amino-acid sequence of pig β-lipotropic hormone has been shown to be closely related to that of the sheep hormone determined some years ago.[427]

The curious peptide from ox hypothalamus, substance P, which was first described by von Euler as long ago as 1931, has been resurrected and examined. Its amino-acid sequence turns out to be not unlike that of physalaemin (from the skin of a South American amphibian) and eledoisin

[417] J.-K. Chang, H. Sievertsson, C. Bogentoft, B. L. Currie, K. Folkers, and C. Y. Bowers, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 409.
[418] R. M. G. Nair, T. W. Redding, and A. V. Schally, *Biochemistry*, 1971, **10**, 3621.
[419] A. V. Schally, Y. Baba, R. M. G. Nair, and C. O. Bennett, *J. Biol. Chem.*, 1971, **246**, 6647.
[420] O. Tarutani and S. Shulman, *Biochim. Biophys. Acta*, 1971, **229**, 642.
[421] O. Tarutani and S. Shulman, *Biochim. Biophys. Acta*, 1971, **236**, 384.
[422] T. Hayashi, K. Iwai, and N. Ui, *Biochem. Biophys. Acta*, 1971, **251**, 208.
[423] R. L. Soffer, *J. Biol. Chem.*, 1971, **246**, 1481.
[424] R. L. Soffer, *J. Biol. Chem.*, 1971, **246**, 1602.
[425] H. T. Keutmann, G. D. Aurbach, B. F. Dawson, H. D. Niall, L. J. Deftos, and J. T. Potts, jun., *Biochemistry*, 1971, **10**, 2779.
[426] J. S. Woodhead, J. L. H. O'Riordan, H. T. Keutmann, M. L. Stoltz, B. F. Dawson, H. D. Niall, C. J. Robinson, and J. T. Potts, jun., *Biochemistry*, 1971, **10**, 2787.
[427] L. Graf, E. Barat, G. Cseh, and M. Sajgo, *Biochim. Biophys. Acta*, 1971, **229**, 276.

(from the salivary glands of a cephalopod):[428]

Substance P	Arg-Pro-Lys-Pro-Gln-Gln-**Phe**-Phe-**Gly-Leu-Met-NH$_2$**
Physalaemin	Glp-Ala-Asp-Pro-Asn-Lys-**Phe**-Try-**Gly-Leu-Met-NH$_2$**
Eledoisin	Glp-Pro-Ser-Lys-Asp-Ala-**Phe**-Ile-**Gly-Leu-Met-NH$_2$**

Two other peptides, bombesin and alytesin, from the skin of European amphibians, have also been characterized:[429]

Bombesin	**Glp-Gln-Arg-Leu-Gly-Asn-Gln-Trp-Ala-Val-Gly-His-Leu-Met-NH$_2$**
Alytesin	**Glp-Gly-Arg-Leu-Gly-Thr-Gln-Trp-Ala-Val-Gly-His-Leu-Met-NH$_2$**

Both are hypertensives and uterine stimulants.

Lys-Ser-Cys-Cys-Pro-Asn-Thr-Thr-Gly-Arg-Asn-Ile-Tyr-Asn-Thr-Cys-Arg-
1 **10**
Leu-Gly-Gly-Gly-Ser-Arg-Glu-Arg-Cys-Ala-Ser-Leu-Ser-Gly-Cys-Lys-Ile-Ile-Ser-
20 **30**
Ala-Ser-Thr-Cys-Pro-Ser-Tyr-Pro-Asp-Lys
40

Figure 9 *The amino-acid sequence of viscotoxin B from European mistletoe*

Six peptides from the venom of *Bothrops jararaca* that act as inhibitors of angiotensin-converting enzymes have been shown to have the following amino-acid sequences:[430]

Glp-Gly-Gly-Trp-Pro-Arg-Pro-Gly-Pro-Glu-Ile-Pro-Pro	
Glp - - Trp-Pro-Arg-Pro-Thr-Pro-Gln-Ile-Pro-Pro	
Glp - - Trp-Pro-Arg - - Pro-Gln-Ile-Pro-Pro	
Glp - Asn-Trp-Pro-Arg - - Pro-Gln-Ile-Pro-Pro	
Glp - Asn-Trp-Pro-His - - Pro-Gln-Ile-Pro-Pro	
Glp - Ser-Trp-Pro-Gly - - Pro-Asn-Ile-Pro-Pro	

They are probably the same as the bradykinin-potentiating peptides of the same venom (see last year's Report).

D. Toxins.—The diphtheria toxin from *Corynebacterium diphtheriae* is biosynthesized as a single polypeptide chain of m. wt. 62 000, but most preparations contain a 'nicked' species consisting of two fragments, A (m. wt. 24 000) and B (m. wt. 38 000), linked by a disulphide bridge.[431] The amino-acid sequence of viscotoxin B from the European mistletoe has been determined [432] (Figure 9).

However, most research has been concerned with snake toxins. Thus, the isolation of several neurotoxins with almost identical amino-acid

[428] M. C. Chang, S. E. Leeman, and H. D. Niall, *Nature New Biol.*, 1971, **232**, 86.
[429] A. Anastasi, V. Erspamer, and M. Bucci, *Experientia*, 1971, **27**, 166.
[430] M. A. Ondetti, N. J. Williams, E. F. Sabo, J. Pluščec, E. R. Weaver, and O. Kocy, *Biochemistry*, 1971, **10**, 4033.
[431] D. M. Gill and L. L. Dinius, *J. Biol. Chem.*, 1971, **246**, 1485; D. M. Gill and A. M. Pappenheimer, jun., *ibid.*, p. 1492; R. Drazin, J. Kandel, and R. J. Collier, *ibid.*, p. 1504.
[432] G. Samuelsson and B. M. Pettersson, *European J. Biochem.*, 1971, **21**, 86.

sequences of 71 residues from the Thailand cobra (*Naja naja Siamensis*) and the spectacled Indian cobra (*Naja naja naja*) has been described.[433] The amino-acid sequence of the toxin A from *Naja naja* [434] is shown in Figure 10. On the other hand, the principal neurotoxins of African cobras contain

Ile-Arg-Cys-Phe-Ile-Thr-Pro-Asp-Ile-Thr-Ser-Lys-Asp-Cys-Pro-Asn-Gly-His-
1 10
Val-Cys-Tyr-Thr-Lys-Thr-Trp-Cys-Asp-Gly-Phe-Cys-Ser-Ile-Arg-Gly-Lys-
20 30
Arg-Val-Asp-Leu-Gly-Cys-Ala-Ala-Thr-Cys-Pro-Thr-Val-Arg-Thr-Gly-Val-
40 50
Asp-Ile-Gln-Cys-Cys-Ser-Thr-Asp-Asp-Cys-Asp-Pro-Phe-Pro-Thr-Arg-Lys-
60
Arg-Pro
70

Figure 10 *Amino-acid sequence of toxin A isolated from the venom of the Indian cobra* Naja naja

61 residues, and two such toxins (β and δ) have been isolated from the South African Cape cobra, *Naja nivea*.[435] Toxin δ has the same amino-acid sequence as toxin α of the cobra *Naja haje haje* and toxin β is closely similar. Toxin α from *Naja nivea*, however, has 71 residues [435, 436] and the toxins α and β, though distinct, are homologous.[436] Two further related toxins of 61 residues from the South African Ringhals cobra, *Haemachatus haemachatus*, have also been characterized.[437] The amino-acid sequence of α-bungarotoxin from *Bungarus multicinctus* (Elipidae snake), which contains 74 residues, is very similar to that of toxin α from *Naja nivea*.[438] Crotoxin, a neurotoxin from the Brazilian rattlesnake, has been reported to contain two components.[439]

Toxins have also been isolated from the sea-snake, *Laticauda semifasciata*.[440, 441a] The amino-acid sequences of two forms (62 residues) have been established and compared with those of cobra neurotoxins [441] (Figure 11). The proteins are obviously closely related and have the same arrangement of disulphide bridges (residues 3—24, 17—41, 43—54, and 55—60).[441b] Another toxin (61 residues) has been isolated from Hardwick's sea-snake, *Lapernis hardwickii*.[442]

[433] E. Karlsson, H. Arnberg, and D. Eaker, *European J. Biochem.*, 1971, **21**, 1.
[434] K. Nakai, T. Sasaki, and K. Hayashi, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 893.
[435] D. P. Botes, D. J. Strydom, C. G. Anderson, and P. A. Christensen, *J. Biol. Chem.*, 1971, **246**, 3132.
[436] D. P. Botes, *J. Biol. Chem.*, 1971, **246**, 7383.
[437] A. J. C. Strydom and D. P. Botes, *J. Biol. Chem.*, 1971, **246**, 1341.
[438] D. Mebs, K. Narita, S. Iwanaga, Y. Samejima, and C. Y. Lee, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 711.
[439] R. A. Hendon and H. Fraenkel-Conrat, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1560.
[440] A.-T. Tu, B.-S. Hong, and T. N. Solie, *Biochemistry*, 1971, **10**, 1295.
[441] (*a*) S. Sato and N. Tamiya, *Biochem. J.*, 1971, **122**, 453; (*b*) Y. Endo, S. Sato, S. Ishii, and N. Tamiya, *ibid.*, p. 463.
[442] A.-T. Tu and B.-S. Hong, *J. Biol. Chem.*, 1971, **246**, 2772.

1. *Laticauda semifasciata* Erabutoxin a Arg-Ile -Cys-Phe-Asn-Gln-His-Ser-Ser-Gln-Pro-Gln-Thr- -Lys-Thr-Cys-Pro-Ser-Gly-
2. Erabutoxin b Arg-Ile -Cys-Phe-Asn-Gln-His-Ser-Ser-Gln-Pro-Gln-Thr- -Lys-Thr-Cys-Pro-Ser-Gly-
3. *Naja nigricollis* Toxin α Leu-Glu-Cys-His-Asn-Gln-Gln-Ser-Ser-Gln-Pro-Pro-Thr-Thr-Lys-Thr-Cys-Pro——Gly-
4. *Naja naja atra* Cobrotoxin Leu-Glu-Cys-His-Asn-Gln-Gln-Ser-Ser-Gln-Thr-Pro-Thr- -Thr-Gly-Cys-Ser-Gly-Gly-
5. *Naja haje haje* Toxin α Leu-Gln-Cys-His-Asn-Gln-Gln-Ser-Ser-Gln-Pro-Pro-Thr- -Lys-Thr-Cys-Pro——Gly-

1 10 20

1. Ser-Glu-Ser-Cys-Tyr-Asn-Lys-Gln-Trp-Ser-Asp-Phe-Arg-Gly-Thr-Ile -Ile -Glu-Arg-Gly-Cys-Gly-Cys-Pro-Thr-Val-Lys-Pro-
2. Ser-Glu-Ser-Cys-Tyr-His-Lys-Gln-Trp-Ser-Asp-Phe-Arg-Gly-Thr-Ile -Ile -Glu-Arg-Gly-Cys-Gly-Cys-Pro-Thr-Val-Lys-Pro-
3. Glu-Thr-Asn-Cys-Tyr-Lys-Lys-Val-Trp-Arg-Asp-His-Arg-Gly-Thr-Ile -Ile -Glu-Arg-Gly-Cys-Gly-Cys-Pro-Thr-Val-Lys-Pro-
4. Glu-Thr-Asn-Cys-Tyr-Lys-Lys-Arg-Trp-Arg-Asp-His-Arg-Gly-Tyr-Arg-Thr-Glu-Arg-Gly-Cys-Gly-Cys-Pro-Ser-Val-Lys-Asn-
5. Glu-Thr-Asn-Cys-Tyr-Lys-Lys-Arg-Trp-Arg-Asp-His-Arg-Gly-Ser-Ile -Thr-Glu-Arg-Gly-Cys-Gly-Cys-Pro-Ser-Val-Lys-Lys-

30 40

1. Gly-Ile-Lys-Leu-Ser-Cys-Cys-Glu-Ser-Glu-Val-Cys-Asn-Asn
2. Gly-Ile-Lys-Leu-Ser-Cys-Cys-Glu-Ser-Glu-Val-Cys-Asn-Asn
3. Gly-Ile-Lys-Leu-Asn-Cys-Cys-Thr-Thr-Asp-Lys-Cys-Asn-Asn
4. Gly-Ile-Glu-Ile -Asn-Cys-Cys-Thr-Thr-Asp-Arg-Cys-Asn-Asn
5. Gly-Ile-Glu-Ile -Asn-Cys-Cys-Thr-Thr-Asp-Lys-Cys-Asn-Asn

50 60

Figure 11 *Comparison of the amino-acid sequences of sea-snake and cobra neurotoxins*

Guanidination and nitroguanidination of the amino-groups of staphylococcal enterotoxin B have been reported to have little or no effect on the biological activity.[443] Similarly, modification of the lysine or arginine residues of the toxins from *Laticauda semifasciata* causes no loss of toxicity [440] but Lys-47 appears to be essential for the activity of cobrotoxin.[444] On the other hand, six of the seven free carboxy-groups of cobrotoxin can be modified using a water-soluble carbodi-imide without effect on the biological activity, but modification of the seventh (that of Glu-21) causes a total loss of toxicity.[444, 445] The toxicity of sea-snake venoms is also lost when the single tryptophan residue in each molecule is modified.[440, 442] A study of the reaction of cobrotoxin with tetranitromethane indicates that Tyr-25 is normally 'buried' and essential for activity, whereas Tyr-35 is 'exposed' and inessential.[446] This result is borne out by the fact that Tyr-25 has been conserved in all neurotoxins so far examined [446] (see Figure 11).

E. Proteins of the Nervous System.—The eponymous review [447] gives a good account of the proteins of myelin, in particular encephalitogenic basic protein. The amino-acid sequences of the proteins (170 residues) from ox [448] and human [449] myelin have now been published in detail but there are some discrepancies still between the results. It is worth noting that this is the first structural protein of membranes for which the sequence is known and that trypsin was reported to hydrolyse Lys–Pro and Arg–Pro bonds in the ox protein.[448] As isolated, the protein contains no carbohydrate, but it can be glycosylated enzymically at the threonine residue in position 98.[448] On the other hand, Arg-107 is present in methylated forms, both ω-*N*-monomethyl- and ω-*NN'*-dimethyl-arginine being detected,[448, 450] and an enzyme from guinea-pig brain has been reported [451] to transfer methyl groups from *S*-adenosylmethionine specifically to Arg-107 in the basic protein. The biological significance remains uncertain. Other experiments [452] have also dealt with the enzymic methylation of arginine and lysine residues in brain (and liver) proteins, and the presence of citrulline in a protein fraction from normal human myelin has been reported.[453] Presumably it also arises by modification of arginine *in situ*. A comparison

[443] L. Spero, H. M. Jacoby, J. E. Daldidowicz, and S. J. Silverman, *Biochim. Biophys. Acta*, 1971, **251**, 345.
[444] C. C. Chang, C. C. Yang, K. Nakai, and K. Hayashi, *Biochim. Biophys. Acta*, 1971, **251**, 334.
[445] C. C. Chang, C. C. Yang, M. Kurobe, K. Nakai, and K. Hayashi, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 429.
[446] C. C. Chang, C. C. Yang, K. Hamaguchi, K. Nakai, and K. Hayashi, *Biochim. Biophys. Acta*, 1971, **236**, 164.
[447] E. M. Shooter and E. Roboz Einstein, *Ann. Rev. Biochem.*, 1971, **40**, 635.
[448] E. H. Eylar, S. Brostoff, G. Hashim, J. Caccam, and P. Burnett, *J. Biol. Chem.*, 1971, **246**, 5770.
[449] P. R. Carnegie, *Nature*, 1971, **229**, 25; P. R. Carnegie, *Biochem. J.*, 1971, **123**, 57.
[450] G. S. Baldwin and P. R. Carnegie, *Biochem. J.*, 1971, **123**, 69.
[451] G. S. Baldwin and P. R. Carnegie, *Science*, 1971, **171**, 579.
[452] Y. Kakimoto, *Biochim. Biophys. Acta*, 1971, **243**, 31.
[453] P. R. Finch, D. D. Wood, and M. A. Moscarello, *F.E.B.S. Letters*, 1971, **15**, 145.

with the published amino-acid sequence of ox myelin basic protein suggests [454] that a protein isolated some time ago from pig hypothalamus [455] is in fact the pig myelin basic protein.

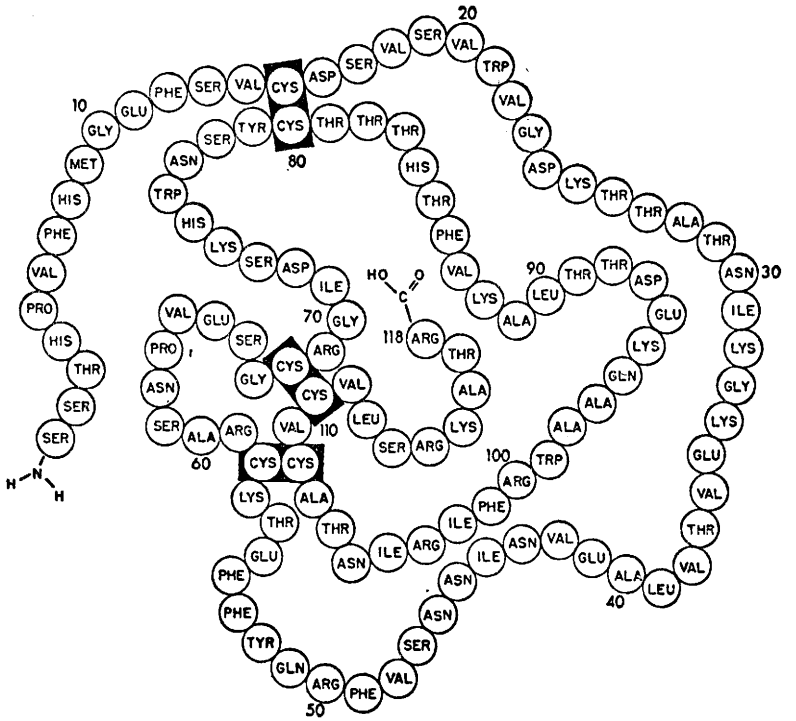


Figure 12 *Schematic representation of the amino-acid sequence of the primary subunit of nerve growth factor from mouse submaxillary glands*
(Reproduced by permission from *Proc. Nat. Sci. U.S.A.*, 1971, **68**, 2417)

Peptic digestion can be used to produce a 46-residue fragment (residues 44—89) from the ox myelin basic protein that retains the ability to induce experimental allergic encephalomyelitis in rabbits.[456, 457] The amino-acid sequences of the corresponding fragments from the guinea-pig, rabbit, monkey, and human proteins are very similar.[457] The single tryptophan residue of the ox and human proteins can be oxidized to the oxindole derivative by treatment with a bromine adduct of 2-(2-nitrophenylsulphenyl)-3-methylindole.[458] No loss of biological activity accompanies the modification. Under stronger conditions, the tryptophyl peptide bond is

[454] R. E. Martenson, G. E. Deibler, and M. W. Kies, *Nature New Biol.*, 1971, **234**, 87.
[455] B. Shome and M. Saffran, *J. Neurochem.*, 1966, **13**, 433.
[456] E. H. Eylar, F. C. Westall, and S. Brostoff, *J. Biol. Chem.*, 1971, **246**, 3418.
[457] R. Shapira, S. S. McKneally, F. Chou, and R. F. Kibler, *J. Biol. Chem.*, 1971, **246**, 4630.

cleaved, and only the N-terminal fragment (116 residues) of the protein retains antigenic determinants.[458]

The amino-acid sequence of nerve growth factor from mouse submaxillary gland has been determined [459] (Figure 12).

5 Enzymes

The Proceedings of the 1971 Cold Spring Harbor Symposium on 'Structure and Function of Proteins at the Three-dimensional Level', [460] is an excellent collection of articles dealing with proteins whose structure is being, or has been, studied by *X*-ray crystallography. A readable review [461] surveys structure and function in several well-characterized proteins and another [462] is concerned with the function of amino-acid side-chains in proteins, particularly the serine side-chain in its various manifestations, *e.g.* as a carrier for phosphate or for phosphopantetheine, or as a reactive nucleophile in the serine proteases, *etc.*

A. Proteolytic Enzymes.—An excellent volume [463] deals with structure, mechanism, formation by activation, and inhibition of all the proteases. Another,[464] equally comprehensive and equally warmly recommended, has a different emphasis and is concerned with preparation, methods of assay, *etc.* of proteases and their naturally occurring inhibitors. A sensitive radioisotopic assay has been suggested [465] as an improvement over the usual haemoglobin and casein methods of assaying proteases; the substrate is haemoglobin completely labelled at carboxy-groups by coupling with radioactive methyl glycinate, and the release of radioactivity is measured.

Carboxypeptidases. A complete picture of carboxypeptidase A as 'a protein and an enzyme' has been presented in a recent review [466] and full details have been published [467] of the determination of the primary structure reported last year. Residue 151, shown as tryptophan in the *X*-ray structure,[468] is confirmed as phenylalanine, and a correction is noted:[469] the amidated glutamic acid residue is Glu-31 not Glu-28. Details have also been given [470] of the stoicheiometric inactivation of carboxypeptidase A

[458] P. R. Burnett and E. H. Eylar, *J. Biol. Chem.*, 1971, **246**, 3425.
[459] R. H. Angeletti and R. A. Bradshaw, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2417.
[460] Cold Spring Harbor Symposium on Quantitative Biology, 1972, vol. 36.
[461] G. P. Hess and J. A. Rupley, *Ann. Rev. Biochem.*, 1971, **40**, 1013.
[462] J. Matheja and E. T. Degens, *Adv. Enzymol.*, 1971, **34**, 1.
[463] 'The Enzymes', ed. P. D. Boyer, Academic Press, N.Y. and London, 3rd edition, 1971, vol. 3.
[464] *Methods Enzymology*, 1970, vol. 19.
[465] H. R. Williams and T.-Y. Lin, *Biochim. Biophys. Acta*, 1971, **250**, 603.
[466] F. A. Quiocho and W. N. Lipscomb, *Adv. Protein Chem.*, 1971, **25**, 1.
[467] R. A. Bradshaw, K. A. Walsh, and H. Neurath, *Biochemistry*, 1971, **10**, 938, 951, 961.
[468] W. N. Lipscomb, J. A. Hartsuck, F. A. Quiocho, and G. N. Reeke, jun., *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **64**, 28.
[469] P. H. Pétra, M. A. Hermodson, K. A. Walsh, and H. Neurath, *Biochemistry*, 1971, **10**, 4023.
[470] (*a*) P. H. Pétra, *Biochemistry*, 1971, **10**, 3163; (*b*) P. H. Pétra and H. Neurath, *ibid.*, p. 3171.

with *N*-ethyl-5-phenylisoxazolium 3′-sulphonate (Woodward's Reagent K) by reaction with Glu-270 at the active site, and carboxy-groups are also implicated [471] at the active site of carboxypeptidase B. Thus affinity labelling with *N*-bromoacetyl-D-[5-^{14}C]arginine at pH 7 was consistent with reaction at a carboxy-group, but not one involved directly in catalysis since treatment of the labelled inactive enzyme with hydroxylamine gave partial reactivation with incorporation of one equivalent of hydroxamate.

Bovine carboxypeptidase A prepared in different ways contains different proportions of three equally active forms (α, β, and γ), and it is pointed out [469] that this may be an important variable in structural and functional studies. Differences in the N-terminal regions of the three forms were conveniently demonstrated using the sequenator:

A_α	**Ala-Arg-Ser-Thr-Asn-Thr-Phe-Asn-Tyr-Ala-**
A_β	**Ser-Thr-Asn-Thr-Phe-Asn-Tyr-Ala-**
A_γ	**Asn-Tyr-Ala-**

Not all laboratories prepare their enzyme in the same way, and it is a little distressing that the amount of β-form in different preparations can vary from 25% to 70%. An improved preparation of the β-form has been described.[472] Although the three forms are equally active, differences clearly exist since, for instance, affinity-labelling of carboxypeptidase A_γ with *N*-bromoacetyl-*N*-methyl-L-phenylalanine caused alkylation of Asn-8 (the N-terminal residue) and His-13 (without loss of activity) in addition to the reaction at Glu-270 which resulted in inactivation.[473] The reaction of 2-hydroxy-5-nitrobenzyl bromide with the amino-terminus of carboxypeptidase A_γ (rather than with tryptophan [474]) may be recalled (see last year's Report). However, other workers, using carboxypeptidase similarly prepared, have reported [475] that tryptophan is modified with a related but water-soluble reagent, dimethyl-(2-hydroxy-5-nitrobenzyl)sulphonium choride, without reaction at the amino-terminus; they attribute the earlier results to unfolding caused by addition of the reagent in solution in acetone.

Activation of bovine procarboxypeptidase has been studied [476] using the arsanilazo chromophore as a probe of environmental changes, and differences between the conformation of arsanilazotyrosine-248 in the crystalline state and in solution have led to the suggestion [477] that 'speculations on the catalytic mechanism based on crystal structure studies alone may require some re-examination', since Tyr-248 has been ascribed the role of proton donor in the catalytic mechanism. The same chromophore has revealed conformational and activity changes when carboxypeptidase is cleaved with

[471] T. H. Plummer, jun., *J. Biol. Chem.*, 1971, **246**, 2930.
[472] G. R. Reeck, K. A. Walsh, and H. Neurath, *Biochemistry*, 1971, **10**, 4690.
[473] G. M. Hass and H. Neurath, *Biochemistry*, 1971, **10**, 3535, 3541.
[474] T. M. Radhakrishnan, R. A. Bradshaw, D. A. Deranleu, and H. Neurath, *F.E.B.S. Letters*, 1970, **7**, 72.
[475] V. R. Naik and H. R. Horton, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 44.
[476] W. D. Behnke and B. L. Vallee, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 760.
[477] J. T. Johansen and B. L. Vallee, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2532.

subtilisin.[478] Substrate binding has been studied[479] by quenching of fluorescence.

The three forms of carboxypeptidase B from activated pancreatic juice have been characterized:[472] the minor form (10%) is a single chain and probably the initial activation product; the two main forms are two-chain proteins which have undergone further splits at the His-95–Met-96 and Arg-92–Glu-93 bonds respectively. Taking advantage of these additional specific cleavages and the availability of a sequenator, comparison of the sequence of carboxypeptidase B with that of carboxypeptidase A was extended by an additional 51 residues,[480] supporting further the hypothesis that they are homologous proteins. Isoleucine at position 255 in carboxypeptidase A is replaced by aspartic acid in the B form, but this may not be the anomaly it seems at first sight since the aspartic acid residue is suitably located in the three-dimensional structure (based on carboxypeptidase A) to provide the negative charge in the binding site of carboxypeptidase B.[480] Purification of bovine carboxypeptidase B by affinity chromatography (Section 2E) with D-Ala-L-Arg as ligand gave a product that was free of carboxypeptidase A but still hydrolysed synthetic substrates of the A form,[224] and an enzyme from orange leaves with the combined specificities of carboxypeptidase A and B has already been mentioned.[91] Four forms of human pancreatic procarboxypeptidase A have recently been demonstrated[481] by isoelectric focusing. The first naturally occurring inhibitor of carboxypeptidase A to be described is the carboxypeptidase B inhibitor (m. wt. *ca.* 3500) from potatoes; combination with the A and B enzymes is mutually exclusive.[482]

Serine Proteases and Their Inhibitors. Two useful source books[463, 464] dealing with all aspects of serine proteases have already been mentioned. Sensitive fluorimetric assays for chymotrypsin using *N*-benzyloxycarbonyl-L-phenylalanine β-naphthyl ester[483] and 4-methylumbelliferyl *p*-trimethylammoniumcinnamate chloride,[484] and a spectrophotometric assay using 4-phenylazobenzyloxycarbonyl-L-Pro-L-Phe-Gly-D-Arg[485] have been described. Trypsin and thrombin can be assayed spectrofluorimetrically with 4-methylumbelliferyl *p*-guanidinobenzoate hydrochloride,[484] an analogue of the *p*-nitrophenyl ester commonly used for spectrophotometric assay of the enzymes.

Certain aspects of the mechanism of action of chymotrypsin and trypsin are being questioned. A kinetic study of the catalytic activity of trypsin

[478] J. F. Riordan and D. M. Livingston, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 695.
[479] N. Lasser and J. Feitelson, *Biochemistry*, 1971, **10**, 307.
[480] G. R. Reeck, K. A. Walsh, M. A. Hermodson, and H. Neurath, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1226.
[481] W. J. Kim and T. T. White, *Biochim. Biophys. Acta*, 1971, **242**, 441.
[482] C. A. Ryan, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 1265.
[483] E. Haas, Y. Elkana, and R. G. Kulka, *Analyt. Biochem.*, 1971, **40**, 218.
[484] D. V. Roberts, R. W. Adams, D. T. Elmore, G. W. Jameson, and W. S. A. Kyle, *Biochem. J.*, 1971, **123**, 41P.

acylated at the active serine with a variety of different carboxy- and dialkylphosphoryl radicals led to the suggestion [486] that the acyl enzyme pathway is not the only one for the hydrolysis of amide and peptide bonds. This, however, is refuted [487] by others who block the active serine of chymotrypsin by carbamylation (*i.e.* with a small polar group, stable to deacylation) and interpret the exactly parallel loss of esterase and amidase activity as evidence that all substrates of serine proteases are hydrolysed *via* an acyl enzyme intermediate. The question of the role of Ile-16 in the catalytic mechanism of α-chymotrypsin is not being settled quite as easily. It will be recalled that the *X*-ray crystallographic work showed that the N-terminal residue Ile-16 forms a salt bridge with Asp-194, thereby maintaining the conformation of the active site, and that the commonly accepted mechanism of α-chymotrypsin action identifies the kinetically important group of p*K* 9 as Ile-16.[488] A study [489] of the reactivity of Ile-16 by the method of competitive labelling with very small amounts of radioactive acetic anhydride (see Section 9) shows that the N-terminus is freely accessible only above pH 9.8, and supports the role of Ile-16 in controlling activity and conformation. On the other hand, retention both of activity and the normal bell-shaped pH-dependence curve when Ile-16 in α-chymotrypsin is hydroxymethylated with formaldehyde,[490a] or when the δ-form is amidinated with either ethyl acetimidate or methyl picolinimidate,[491] as well as retention of activity even after succinylation of the α-amino-group of Ile-16 in δ-chymotrypsin [492] have been taken as evidence that Ile-16 is neither essential for activity nor is it the group of p*K* 9.3 that gives rise to the alkaline limb of the bell-shaped pH-dependence curve. A comparison of α-, α_1-, and δ-chymotrypsins has led to the suggestion [493] that the ionizable group determining the behaviour of α-chymotrypsins at high pH (*i.e.* whose deprotonation leads to disruption of the binding site) is in fact Ala-149, which is the N-terminus of the C-chain in α-chymotrypsin. Ala-149 is bound in peptide linkage in δ-chymotrypsin, in which the C-chain is not cleaved, and in α_1-chymotrypsin the N-terminus of the C-chain is Thr-147. Some support for this comes from the competitive labelling studies [489] which suggest that after deprotonation of Ile-16 (*i.e.* above pH 9.8) a further structural change is dependent on Ala-149. These observations would explain why the pH dependence of the N-terminal group in δ-chymotrypsin does not follow the alkaline

[485] E. Wünsch, A. Högel-Betz, and E. Jaeger, *Z. physiol. Chem.*, 1971, **352**, 1553.
[486] S. E. Bresler, V. M. Krutyakov, and G. P. Vlasov, *European J. Biochem.*, 1971, **18**, 131.
[487] C. E. Stauffer, *F.E.B.S. Letters*, 1971, **16**, 45.
[488] H. L. Oppenheimer, B. Labouesse, and G. P. Hess, *J. Biol. Chem.*, 1966, **241**, 2720.
[489] H. Kaplan, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 1042.
[490] (*a*) M. A. Marini and C. J. Martin, *European J. Biochem.*, 1971, **19**, 162; (*b*) M. A. Marini and C. J. Martin, *ibid.*, p. 153.
[491] S. P. Agarwal, C. J. Martin, T. T. Blair, and M. A. Marini, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 510.
[492] T. T. Blair, M. A. Marini, S. P. Agarwal, and C. J. Martin, *F.E.B.S. Letters*, 1971, **14**, 86.
[493] P. Valenzuela and M. L. Bender, *J. Amer. Chem. Soc.*, 1971, **93**, 3783.

activity curve, and support the idea that the internal split in α-chymotrypsin, releasing Ala-149 as the N-terminal residue of the C-chain, increases the disruptive effect of the deprotonation of Ile-16. Chemical modification of Ile-16 in δ-chymotrypsin [494] is consistent with an appreciable conformational change on binding to the substrate and with the existence of a salt bridge between Ile-16 and Asp-194 in solution, and other studies [495] on α-chymotrypsin demonstrate an ionic-strength-dependent equilibrium between exposed (at low ionic strength) and buried states of Ile-16 at pH 4. A detailed kinetic study [496] of α-chymotrypsin demonstrates two conformational states at neutral pH – the normal active form and as much as 15—20% of an inactive form which is the major form above pH 9. There is a slow interconversion of the two forms, and the authors call for caution in interpreting kinetic results. They give the pK of Ile-16 in the inactive form as 7.94, an indication that the N-terminus is exposed.

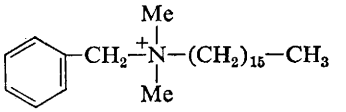
The role of His-57 in the catalytic mechanism has also been explored further. α-Chymotrypsin methylated selectively at His-57 with methyl *p*-nitrobenzenesulphonate (see last year's Report) showed [497] unimpaired substrate binding, but the rate of acylation and of deacylation were much slower, although still dependent on a group of pK_a 7. It is thus suggested that 3-MeHis-57 acts as a general base on the hydroxy-group of Ser-195 in the acylation, and that deacylation occurs by a similar mechanism; the poor hydrogen-bond between N-1 of 3-MeHis-57 and the serine side-chain presumably accounts for the decreased rate. If this is correct then there can be no hydrogen-bond between Asp-102 and 3-MeHis-57 during catalysis, and on this basis the contribution of Asp-102 to the rate enhancement in the native enzyme is estimated [497] as, at most, 5×10^3. A similar explanation might account for the retention of (impaired) activity when the histidine residues are hydroxymethylated with formaldehyde.[490b, 498]

While the details of the actual catalytic mechanism are being sorted out, work continues on elucidation of the nature of the substrate-binding site. Crystallographic studies [499] of bovine chymotrypsin A_γ inhibited with a series of specific active-site-directed peptide chloromethyl ketones confirmed the nature of the specific aromatic binding site (subsite S_1) and identified subsites S_2 and S_3; it was held to be unlikely that any more than three residues on the N-terminal side of the bond cleaved in the substrate would be recognized. Two investigations have compared the dimensions of the binding sites of elastase and chymotrypsin in solution, using in one case [500]

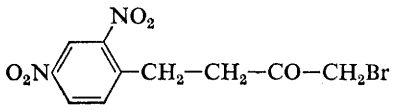
[494] P. Valenzuela and M. L. Bender, *Biochim. Biophys. Acta*, 1971, **235**, 411.
[495] A. Kurosky, J. E. S. Graham, J. W. Dixon, and T. Hofmann, *Canad. J. Biochem.*, 1971, **49**, 529.
[496] A. R. Fersht and Y. Requena, *J. Mol. Biol.*, 1971, **60**, 279.
[497] R. Henderson, *Biochem. J.*, 1971, **124**, 13.
[498] P. Dunlop, M. A. Marini, and C. J. Martin, *Biochim. Biophys. Acta*, 1971, **243**, 320; C. J. Martin, N. B. Oza, and M. A. Marini, *European J. Biochem.*, 1971, **20**, 276.
[499] D. M. Segal, J. C. Powers, G. H. Cohen, D. R. Davies, and P. E. Wilcox, *Biochemistry*, 1971, **10**, 3728.
[500] T. H. Marshall and A. Akgün, *J. Biol. Chem.*, 1971, **246**, 6019.

fatty acid nitrophenyl esters of different chain length, and in the other [501] a series of homologous alkyl isocyanates. The results, which indicate a small binding cavity in elastase, and a larger one in chymotrypsin capable of accommodating bulky side-chains, are in accord with the picture given by *X*-ray crystallography.

Cetyldimethylbenzylammonium chloride (6) inhibition appears to be another test for a serine protease.[502] At acid pH it inhibits chymotrypsin, trypsin, and elastase as well as aspergillopeptidase B and subtilisin, but does not inhibit non-serine proteases. Binding to trypsin is reversible but binding cannot be reversed in the case of chymotrypsin, presumably because the benzene ring binds tightly in the substrate-binding pocket. Inhibition of serine proteinases that differ widely in amino-acid sequence suggests a common element in the environment of the active site; the common charge-relay catalytic mechanism is already accepted as one such feature. A new inhibitor of elastase has been described;[503] the site of the inhibition by 1-bromo-4-(2,4-dinitrophenyl)butan-2-one (7) has been tentatively



(6)



(7)

identified as Glu-6. The inhibitor is not a substrate analogue and there is no glutamic acid either at, or near, the active site as shown by the *X*-ray model. The dinitrophenyl group was found to be essential for inhibition, leading to the speculation that the inhibitor binds to the enzyme at a site involved in binding desmosine or isodesmosine residues of elastin *in vivo*.[503] The reaction between chymotrypsin and phenylalanine chloro(or bromo)methyl ketones, or N-substituted derivatives of these, resulted in alkylation of His-57 in each case indicating that the inhibitors were similarly positioned.[504] Photoaffinity labelling is a powerful method of mapping the active sites of enzymes. Its advantages are specificity in the placement of the label and almost complete non-specificity in the side-chains which the reactive radical generated on activation can attack. A product of photolysis of diazo[^{14}C]acetylchymotrypsin, identified [505] after hydrolysis as *O*-carboxymethyl-Tyr-146 (the C-terminal residue of the B chain), is thought to result from intermolecular reaction within a chymotrypsin dimer, by reaction of the carbene generated at diazoacetyl-Ser-195 in one molecule with Tyr-146 in the other. Photolysis of ethyl diazomalonyltrypsin and chymotrypsin,

[501] W. E. Brown and F. Wold, *Science*, 1971, **174**, 608.
[502] E. Feldbau and C. Schwabe, *Biochemistry*, 1971, **11**, 2131.
[503] L. Visser, D. S. Sigman, and E. R. Blout, *Biochemistry*, 1971, **10**, 735.
[504] E. Shaw and J. Ruscica, *Arch. Biochem. Biophys.*, 1971, **145**, 484.
[505] C. S. Hexter and F. H. Westheimer, *J. Biol. Chem.*, 1971, **246**, 3928.

however, gave *S*-carboxymethylcysteine as the main product after hydrolysis, indicating [506] that the carbene had cleaved a disulphide bond, tentatively identified as the disulphide that maintains the 'histidine loop'.

A new species of chymotrypsin, K-chymotrypsin, has been isolated from activation of bovine chymotrypsinogen A and introduced into the generally accepted activation scheme between δ- and γ-chymotrypsin on the fast activation pathway.[507] Chymotrypsin C from autolysed pig pancreas was shown [508] to differ from the π-form only by loss of the four C-terminal residues of the A chain; unlike the initial cleavage products of chymotrypsinogens A and B it shows no susceptibility to further proteolytic attack. Chymotrypsin-P, generated by papain activation of chymotrypsinogen A, is identical with α-chymotrypsin except that it is three residues shorter at the C-terminus of the A-chain.[509] The second major chymotrypsin (chymotrypsin II), isolated from human pancreas,[510] has only two chains, like bovine δ-chymotrypsin; it is not clear whether chymotrypsins I and II arise from the same or different zymogens. With zymogen granules, rather than the usual pancreatic juice as starting material, chymotrypsinogens, trypsinogens, and procarboxypeptidases have been isolated from chick pancreas.[511]

Lungfish and dogfish, cows and camels, deer and goats . . . not Noah's guest list but some of the species for whose trypsinogens amino-acid sequence information has recently become available. The sequenator has yielded the first 20 amino-acid residues of bovine, lungfish, and dogfish trypsinogens [512] and shown them to be highly homologous (residues in brackets are tenatative assignments):

	1	5 ↓	10
Bovine	Val-Asp-Asp-Asp-	**Asp-Lys-Ile-Val-Gly-Gly-Tyr-Thr-Cys**	-Gly-
Dogfish	Ala-Pro-Asp-Asp-	**Asp-Lys-Ile-Val-Gly-Gly-Tyr-Glu-Cys**	-Pro-
Lungfish	Phe-Pro-Ile	-Glu-Glu- **Asp-Lys-Ile-Val-Gly-Gly-Try-Glu-(Cys)-Pro-**	

	15	20
Bovine	Ala-Asn-Thr	-Val-**Pro**-Tyr-
Dogfish	Lys-His-Ala	-Ala-**Pro**-Trp-
Lungfish	Lys-His-(Thr)-Val-	**Pro**-Trp-

The activation point is the Lys-6–Ile-7 bond: the five N-terminal residues of the resulting trypsins are identical. It is interesting that the usual activation sequence X-$(Asp)_4$-Lys-Ile- is replaced in lungfish trypsinogen by X-Ile-Glu-Glu-Asp-Lys-Ile-; presumably conservation of three negative charges is crucial for the activation. Dromedary [513] and goat [514] trypsinogens have

[506] C. S. Hexter and F. H. Westheimer, *J. Biol. Chem.*, 1971, **246**, 3934.
[507] D. D. Miller, T. A. Horbett, and D. C. Teller, *Biochemistry*, 1971, **10**, 4641.
[508] D. Gratecos and P. Desnuelle, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 857.
[509] M. C. Shaw and T. Viswanatha, *Canad. J. Biochem.*, 1971, **49**, 999.
[510] M. H. Coan, R. C. Roberts, and J. Travis, *Biochemistry*, 1971, **10**, 2711.
[511] R. Zelikson, G. Eilam-Rubin, and R. G. Kulka, *J. Biol. Chem.*, 1971, **246**, 6115.
[512] M. A. Hermodson, R. W. Tye, G. R. Reeck, H. Neurath, and K. A. Walsh, *F.E.B.S. Letters*, 1971, **14**, 222.
[513] S. Bricteux-Grégoire, R. Schyns, and M. Florkin, *Biochim. Biophys. Acta*, 1971, **251**, 79.
[514] S. Bricteux-Grégoire, R. Schyns, and M. Florkin, *Biochim. Biophys. Acta*, 1971, **229**, 123.

the normal activation sequence and the activation peptides (apparently two in the case of the goat zymogen) have the sequences:

Camelus dromedarius	Val-Pro-Ile -**Asp-Asp-Asp-Asp-Lys**-
Goat	Phe-Pro-Val-**Asp-Asp-Asp-Asp-Lys**-
and	Val-**Asp-Asp-Asp-Asp-Lys**-

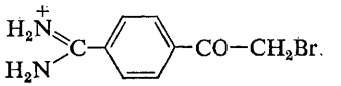
The action of enterokinase on trypsinogen is that of a very restricted trypsin (*cf.* the action of thrombin on fibrinogen, Section 2D). Enterokinase, which has recently been purified [515] and which is specific for native trypsinogen, shows subsite specificity and recognizes the whole of the -$(Asp)_4$-Lys- structure. It is found in the duodenum; although its K_m for trypsinogen is six times larger than that of trypsin, its V_{max} is 2000 times larger, and it may thus be the trigger for activation, producing the initial molecules of trypsin for autocatalysis to take over, conveniently away from the pancreas. Another study [516] of the autoactivation of trypsinogen suggests that the zymogen has an inherent proteolytic activity which allows it to activate itself; since it cannot have an active site maintained by Ile-16, it is suggested [516] that the active site is formed in a dimer using Ser-183 of one molecule and His-46 of the other. The familiar trypsinogen is cationic. Anionic trypsinogens recently purified from pig [517] and bovine [518] pancreas give rise to trypsins with catalytic and inhibition properties similar to those of normal cationic trypsin, and an anionic trypsin-like enzyme has also been purified from *Streptomyces erythreus*.[519] The zymogen of the trypsin-like enzyme cocoonase, from the silk moth, has been characterized;[520] it has a molecular weight of *ca.* 30 000, and *ca.* 20% of the molecule is removed in the activation process, whereas only six residues are lost when trypsinogen is activated. The immunochemical cross-reactivity that exists between trypsin and cocoonase is not present in the zymogens, suggesting perhaps that activation peptides are not subject to the stringent selection pressures that operate on catalytic chains.[520] It will thus be interesting to see whether the -$(Asp)_4$-Lys- sequence is conserved in the acidic fragment removed on activation of prococoonase.

There are many ways of probing the conformational changes that accompany zymogen activation. Two disulphide bonds are readily reduced in trypsinogen but are buried in trypsin,[521] and the tetracarboxymethylated zymogen can be activated to give the expected yield of active sites, showing that the reducible disulphides are not essential structural elements.[521] Selective reduction of only one disulphide (179—203) and carboxymethylation gave on activation carboxymethyltrypsin of impaired catalytic efficiency, probably a consequence of a slightly distorted conformation in the

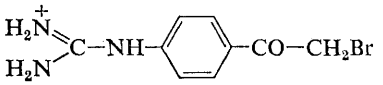
[515] S. Maroux, J. Baratti, and P. Desnuelle, *J. Biol. Chem.*, 1971, **246**, 5031.
[516] J. Kay and B. Kassell, *J. Biol. Chem.*, 1971, **246**, 6661.
[517] P. Voytek and E. C. Gjessing, *J. Biol. Chem.*, 1971, **246**, 508.
[518] A. Puigserver and P. Desnuelle, *Biochim. Biophys. Acta*, 1971, **236**, 499.
[519] N. Yoshida, A. Sasaki, and H. Inoue, *F.E.B.S. Letters*, 1971, **15**, 129.
[520] E. Berger, F. C. Kafatos, R. L. Felsted, and J. H. Law, *J. Biol. Chem.*, 1971, **246**, 4131.
[521] D. L. Sondack and A. Light, *J. Biol. Chem.*, 1971, **246**, 1630.

modified trypsin, since the (carboxymethylated) disulphide bond can no longer become buried in the usual way.[522] The fate of tyrosine and tryptophan residues on activation of trypsinogen has been studied by solvent perturbation methods,[523] and fluorescent probes have also been exploited. Conformational changes on activation have been revealed using non-covalently bound 7-(*p*-methoxybenzylamino)-4-nitrobenzoxadiazole,[524] and the dansyl group has been covalently bound [525] (by reaction at pH 5) to a ring amino-group which had been introduced by specific nitration and reduction of Tyr-137. Trypsin dansylated under normal conditions [526] shows increased activity towards *N*-benzoyl-L-arginine *p*-nitroanilide but not the ethyl ester, suggesting more efficient binding rather than enhanced catalysis.

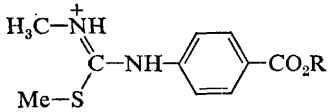
Several investigations have been concerned with the active centre of trypsin. Full details have now been given [527] of the modification of Asp-177 in the active centre using a water-soluble carbodi-imide to couple it to glycinamide. The active-site-directed phenacyl halides *p*-amidinophenacyl bromide (8) and *p*-guanidinophenacyl bromide (9) inactivate trypsin irreversibly,[528] and in the case of (9) this is attributed to formation of an ether link with the active serine, Ser-183. It will be recalled that the active-site-directed tosyl-lysine chloromethyl ketone (TLCK) alkylated His-46 at the active centre, indicating (not surprisingly) small differences in the geometry of binding of the two reagents. From the same laboratory comes the elegant approach [529] of acylating the active site with a group which is only very slowly deacylated and which also carries a functional group potentially able to cross-link to some reactive group in the vicinity. The compound chosen was the isothiuronium derivative (10) of a compound related to the active-site titrant *p*-nitrophenyl *p'*-guanidinobenzoate (11; R = *p*-nitro-



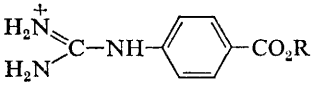
(8)



(9)



(10)

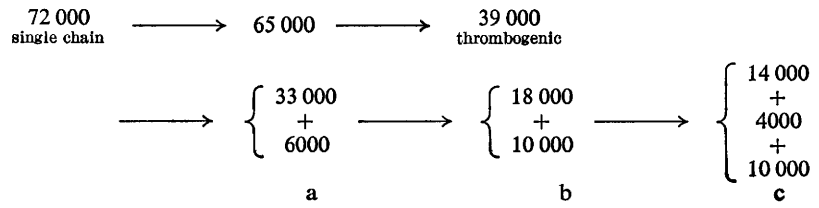


(11)

[522] L. M. Hatfield, S. K. Banerjee, and A. Light, *J. Biol. Chem.*, 1971, **246**, 6303.
[523] G. B. Villanueva and T. T. Herskovits, *Biochemistry*, 1971, **10**, 3358.
[524] R. A. Kenner and A. A. Aboderin, *Biochemistry*, 1971, **10**, 4433.
[525] R. A. Kenner and H. Neurath, *Biochemistry*, 1971, **10**, 551.
[526] J. G. Franklin and J. Leslie, *Canad. J. Biochem.*, 1971, **49**, 516.
[527] A. W. Eyl, jun., and T. Inagami, *J. Biol. Chem.*, 1971, **246**, 738.
[528] D. D. Schroeder and E. Shaw, *Arch. Biochem. Biophys.*, 1971, **142**, 340.
[529] P. Bodlaender and E. Shaw, *Arch. Biochem. Biophys.*, 1971, **147**, 810.

phenyl) (see last year's Report), and it was hoped that the positively charged isothiuronium group would modify a carboxy-group, particularly Asp-177 at the specificity site. Unfortunately this did not occur; covalent bonds were, however, formed when methyl mercaptide was eliminated with mercuric chloride, but the positions of these have not been identified. In another approach [530] to mapping the environment of the active centre it was found that the *trans*-isomer of a photochromic substrate for trypsin, *p*-phenylazobenzoyl-L-Arg-OMe, could be accommodated more easily at the active site than the bulky *cis*-analogue.

The stages involved in the formation of bovine thrombin from prothrombin have been closely examined.[531] The picture which has emerged is as follows:



The chains in all three forms (a, b, and c) are disulphide-linked. All three forms have the same activity towards small substrates but the clotting activities differ (a > b > c). It thus appears fairly certain that the forms having molecular weight 28 000, found in commercial thrombin, are not artefacts but normal prothrombin activation products. The site of the final cleavage to give the three-chain form has been identified [532] as either Arg-73 or Arg-76. Traces of thrombin are apparently required for prothrombin activation in a step that follows a structural change in the zymogen.[533] Inhibition of thrombin by phenylmethylsulphonyl fluoride occurs at, or near, the active site,[534] and resembles inhibition with di-isopropyl fluorophosphate, suggesting once again that thrombin shares the catalytic mechanism of the serine proteases. It is also inhibited by TLCK and by *p*-nitrobenzyloxycarbonylarginine chloromethyl ketone, and the site of alkylation has now been identified [532] as His-43 in the B-chain, which the partial sequence for thrombin [535] shows to be homologous with His-57 of chymotrypsin. The purification of human thrombin has been reported [536] and several active species of the human enzyme with different molecular

[530] M. A. Wainberg and B. F. Erlanger, *Biochemistry*, 1971, **10**, 3816.
[531] K. G. Mann, C. M. Heldebrant, and D. N. Fass, *J. Biol. Chem.*, 1971, **246**, 5994, 6106.
[532] G. Glover and E. Shaw, *J. Biol. Chem.*, 1971, **246**, 4594.
[533] I. P. Baskova, S. M. Strukova, and A. Uriarte, *Biochemistry* (*U.S.S.R.*), 1970, **35**, 670.
[534] R. L. Lundblad, *Biochemistry*, 1971, **10**, 2501.
[535] S. Magnusson, 'Structural Aspects of Thrombin and Prothrombin: Structure–Function Relationships of Proteolytic Enzymes', ed. P. Desnuelle, H. Neurath, and M. Otteson, Munksgaard, Copenhagen, 1970, p. 138.
[536] J. W. Fenton, W. P. Campbell, J. C. Harrington, and K. D. Miller, *Biochim. Biophys. Acta*, 1971, **229**, 26.

weights have been found;[537] how this fits in with the report of glutamic acid as the (sole) C-terminal residue of human thrombin [538] remains to be seen. A thrombin-like enzyme which acts on fibrinogen and which is inhibited by DFP and TLCK has been purified [539] from the venom of the Eastern Diamond rattlesnake *Crotalus adamanteus*; it has a single chain and molecular weight 32 000. The specificity of thrombin action on fibrinogen has already been mentioned in the context of chain cleavage (Section 2D). Thrombin brings about the formation of blood clots and plasmin their solubilization; human plasminogen free from plasmin has been prepared by affinity chromatography [231] and has N-terminal glutamic acid. The *S*-carboxymethylated heavy chains of human plasmin have been characterized [540] (m. wt. 48 000) and have C-terminal arginine. Preliminary work [541] on pig pancreatic kallikrein suggests that it may resemble the other pancreatic serine proteinases in having an -Asp-Ser-Gly- sequence at its active site.

The proceedings [542] of an international meeting on inhibitors of proteolytic enzymes, recently published, are evidence of considerable activity in this field. The other standard works of reference [463, 464] have already been noted. The small size and the single-chain nature of the naturally occurring inhibitors would seem to make them suitable systems on which to study many aspects of protein structure and synthesis.

Bovine pancreatic tissue contains two inhibitors of trypsin: the basic (Kunitz) inhibitor, of known primary and tertiary [543] structure and which has recently been synthesized (58 residues) by the solid-phase method,[544] and the Kazal inhibitor (56 residues), also of known amino-acid sequence. The sequence of the trypsin inhibitor from cow colostrum has recently been determined.[545] It shows considerable homology with the sequence of the bovine basic pancreatic inhibitor (Figure 13) and has -*Lys-Ala- at the active site. The active site appears to be similar in the inhibitors from pig colostrum.[546] When the basic inhibitor combines with trypsin a single easily reducible disulphide bond in both the inhibitor and the enzyme becomes buried,[547] suggesting that 'the mutual binding regions are located near these disulphides'. Conformational changes would also seem to be a possible

[537] T. M. Chulkova and V. N. Orekhovich, *Biochemistry* (*U.S.S.R.*), 1971, **36**, 550.
[538] K.-I. Kasai, Y. Takai, T. Watanabe, and S.-I. Ishi, *Internat. J. Protein Res.*, 1971, **3**, 117.
[539] F. S. Markland and P. S. Damus, *European J. Biochem.*, 1971, **246**, 6460.
[540] L. Summaria, K. C. Robbins, and G. H. Barlow, *J. Biol. Chem.*, 1971, **246**, 2143.
[541] F. Fielder, B. Müller, and E. Werle, *Z. physiol. Chem.*, 1971, **352**, 1463.
[542] 'First Conference on Proteinase Inhibitors', ed. H. Fritz, H. Tschesche, and E. Werle, Walter de Gruyter and Co., Berlin, 1971.
[543] R. Huber, D. Kukla, A. Rühlmann, O. Epp, and H. Formanek, *Naturwiss.*, 1970, **57**, 389.
[544] K. Noda, S. Terada, N. Mitsuyasu, M. Waki, T. Kato, and N. Izumiya, *Naturwiss.*, 1971, **58**, 147.
[545] D. Čechová, V. Jonáková, and F. Šorm in Ref. 542, p. 105.
[546] L. F. Kress, S. R. Martin, and M. Laskowski, *Biochim. Biophys. Acta*, 1971, **229**, 836.
[547] W. Liu, H. Trzeciak, H. Schüssler, and J. Meienhofer, *Biochemistry*, 1971, **10**, 2849.

explanation but another study [548] suggests that these do not occur, and it is estimated, from tritium exchange studies, that between 10 and 25 residues of trypsin are in contact with the inhibitor, a fairly large 'active centre'. Basic inhibitor, in which the easily reducible disulphide bond (14—38) had been reduced and carboxamidomethylated, still inhibited trypsin, but the inhibition was 'temporary' and trypsin activity was gradually released.[549]

CTI Phe-Gln-Thr-Pro-**Pro**-**Asp**-Leu-Cys-Gln-Leu-**Pro**-Gln-Ala-Arg-**Gly**-**Pro**-
1

BPTI Arg-**Pro**-**Asp**-Phe-Cys-Leu-Glu-**Pro**-Pro-Tyr-Thr-**Gly**-**Pro**-
1

CTI **Cys**-**Lys**-**Ala**-Ala-Leu-Leu-**Arg**-**Tyr**-**Phe**-**Tyr**-Asx-Ser-Thr-Ser-Asn-Ala-
*

BPTI **Cys**-**Lys**-**Ala**-Arg-Ile -Ile -**Arg**-**Tyr**-**Phe**-**Tyr**-Asn-Ala-Lys-Ala-Gly-Leu-
*

CTI Cys-Glu-Pro-**Phe**-Thr-**Tyr**-**Gly**-**Gly**-**Cys**-Gln-Gly-Asn-Asn-Asx-**Asn**-**Phe**-

BPTI Cys-Gln-Thr-**Phe**-Val-**Tyr**-**Gly**-**Gly**-**Cys**-Arg-Ala-Lys-Arg-Asn-**Asn**-**Phe**-

CTI Glu-Thr-Thr-**Glu**-Met-Cys-Leu-**Arg**-Ile -Cys-Glu-Pro-Pro-Gln-Gln-Thr-

BPTI Lys-Ser-Ala-**Glu**-Asp-Cys-Met-**Arg**-Thr-Cys-Gly-Gly-Ala
58

CTI Asp-Lys-Ser
67

Figure 13 *Comparison of amino-acid sequences of cow colostrum trypsin inhibitor (CTI) and basic pancreatic trypsin inhibitor (BPTI)*

This was accompanied by the sequential cleavage of three bonds in the inhibitor identified as Lys-15–Ala-16 (the active site), Arg-39–Ala-40, and Arg-17–Ile-18. The *X*-ray model suggests that reduction of the disulphide bond increases the flexibility of the chain segment ending in Lys-15 such that the interaction required at this point for formation of the usual permanent inhibition cannot occur.[549] The basic pancreatic inhibitor can exist in three pH-dependent conformational states;[550] the acid isomerization is dependent on the ionization of a buried carboxylate group which may be the side-chain of Glu-7, and the alkali isomerization on the unmasking of the α-amino-group of the N-terminal arginine (apparent p*K* 9.4), reminiscent of the situation in chymotrypsin. It is well established that the basic pancreatic inhibitor belongs to the class in which a lysyl bond is cleaved in the trypsin–inhibitor complex; a recent investigation [551] of the reaction of the arginine residues with phenylglyoxal showed that five of the six could be modified

[548] M. A. Chepyzheva, G. Ya. Kolomiitseva, and A. G. Tarasenko, *Biochemistry (U.S.S.R.)*, 1971, **36**, 306.
[549] K. A. Wilson and M. Laskowski, *J. Biol. Chem.*, 1971, **246**, 3555.
[550] J.-P. Vincent, R. Chicheportiche, and M. Lazdunski, *European J. Biochem.*, 1971, **23**, 401.
[551] B. Keil, *F.E.B.S. Letters*, 1971, **14**, 181.

without loss of inhibitory activity; the lack of reactivity of the remaining arginine (Arg-53) was difficult to explain in the *X*-ray model, but the fact that it was distinctly removed from the active-site lysine suggested that it was unlikely to be involved in complex formation.[551] A study of the susceptibility of the four lysyl bonds in the Kunitz inhibitor to tryptic digestion

```
Pig I   Thr-Ser-Pro-Gln-Arg-Glu-Ala-Thr-Cys-Thr-Ser -Glu-Val-Ser -Gly-Cys-
Sheep   Asn-Ile -Leu-Gly-Arg-Glu-Ala-Lys-Cys-Thr-Asn-Glu-Val-Asn-Gly-Cys-
                      5                   10                  15

                    *
Pig I   Pro- Lys-Ile-Tyr-Asn-Pro-Val-Cys-Gly-Thr-Asp-Gly-Ile -Thr-Tyr-Ser -Asn-
Sheep   Pro-*Arg-Ile-Tyr-Asn-Pro-Val-Cys-Gly-Thr-Asp-Gly-Val-Thr-Tyr-Ala-Asn-
                    20                  25                  30
Pig I   Glu-Cys-Val-Leu-Cys-Ser-Glu-Asn-Lys-Lys-Arg-Gln-Thr-Pro-Val-Leu-
Sheep   Glu-Cys-Leu-Leu-Cys-Met-Glu-Asn-Lys-Glu-Arg-Gln-Thr-Pro-Val-Leu-
          35                  40                  45
Pig I   Ile-Gln-Lys-Ser-Gly-Pro-Cys
Sheep   Ile-Gln-Lys-Ser-Gly-Pro-Cys
        50                  55
```

Figure 14 *Comparison of the amino-acid sequences of pancreatic secretory trypsin inhibitor from pig (inhibitor I) and sheep*

after guanidination led to the conclusion [552] that homoarginyl bonds are substrates for trypsin although in the sequences -Harg-X-Arg- and -Harg-Arg- they are very poorly split. Care must therefore be taken when retention of inhibitory activity after guanidination is interpreted as evidence against a lysine active site.

The amino-acid sequence of pig pancreatic secretory trypsin inhibitor (Kazal type inhibitor), reported last year, has now been confirmed in another laboratory.[553] This has 56 residues, with 12 differences relative to the bovine sequence determined earlier; the primary structure of the inhibitor from sheep [554] has also been presented and this differs from the inhibitor from pig in 13 positions (Figure 14). The accumulation of substitutions at the N-terminus suggests that this region is not critical for inhibitor activity. All three have identical sequences around the bond cleaved, although this itself may be an arginyl (cow, sheep) or lysyl (pig) bond; the pancreatic secretory inhibitor from the cat also has a lysine residue at the active site.[555] The three disulphide bonds of the bovine inhibitor have recently been identified [556] and probably occur in homologous positions in the inhibitors of cow and pig, which also have three disulphide bonds.

[552] J. Chauvet and R. Acher, *F.E.B.S. Letters*, 1971, **18**, 265.
[553] D. C. Bartelt and L. J. Greene, *J. Biol. Chem.*, 1971, **246**, 2218.
[554] H. Tschesche, E. Wachter, S. Kupfer, R. Obermeier, G. Reidel, G. Haenisch, and M. Schneider, in Ref. 542, p. 207.
[555] H. Tschesche and S. Kupfer, *Z. physiol. Chem.*, 1971, **352**, 764.
[556] O. Guy, R. Shapanka, and L. J. Greene, *J. Biol. Chem.*, 1971, **246**, 7740.

Protease inhibitors from plants are also being extensively investigated, the lima-bean inhibitor, the Bowman–Birk soybean, and the Kunitz soybean inhibitors being the best characterized. The complete sequence of the lima-bean inhibitor (84 residues) is now known [557] but shows no obvious resemblance to the pancreatic inhibitors described above. It is possible that certain regions of the lima-bean sequence may have arisen by gene duplication. For example, (but bearing in mind the Yiddish proverb [558]) the sequence 15—34 is homologous with 43—61, assuming one deletion:

```
 15                  20
-Ser-Cys-Lys-Pro-Cys-Asn-His-Cys-Cys-Cys-Leu-Ser/Ala-Thr-Lys-Ser-
-Ser-Cys-His-Ser-Cys-Ala-Lys-  -  -Cys-Cys-Ile -Ser-Thr-Leu-Ser-
 43                                 50
 30          34
Ile-Pro-Pro-Glx-Cys-
Ile-Pro-Ala-Gln-Cys-
          60  61
```

The anti-chymotryptic site is now identified as the Leu-55–Ser-56 bond, and it is very tempting to speculate that the apparently homologous Lys-28–Ser-29 bond is the anti-tryptic site, and that the double-headed inhibitor has arisen by gene-duplication.[557b] The Lys-28–Ser-29 bond does appear to be readily cleaved by chymotrypsin, or by trypsin after guanidination.[557b] The existence of two functionally independent sites for trypsin and chymotrypsin in lima-bean inhibitor has been confirmed.[559] The Bowman–Birk soybean inhibitor (71 residues) is also double-headed and the amino-acid sequence around the anti-tryptic (lysyl) site is similar to that of the lima-bean inhibitor:

```
                      17  18
Bowman–Birk  -Ala-Thr-Lys-Ser-Asx-Pro-
lima bean    -Ala/Ser-Thr-Lys-Ser-Ile -Pro-
                      28  29
```

There is no homology with the active sites of the basic and secretory pancreatic inhibitors or with the Kunitz soybean inhibitor. Guanidination did not abolish the anti-tryptic activity [560] and again the homoarginyl bond was cleaved by trypsin.[561] In a strange report,[562] 'accidental' chemical modification (when tap-water was used for dialysis) seemed to indicate that certain tyrosine, methionine, and tryptophan residues were not essential for

[557] (*a*) C. G. L. Tan and F. C. Stevens, *European J. Biochem.*, 1971, **18**, 503; (*b*) C. G. L. Tan and F. C. Stevens, *ibid.*, p. 515.
[558] " 'For example' is not proof."
[559] J. Krahn and F. C. Stevens, *F.E.B.S. Letters*, 1971, **13**, 339.
[560] D. S. Seidl and I. E. Liener, *Biochim. Biophys. Acta*, 1971, **251**, 83.
[561] D. S. Seidl and I. E. Liener, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 1101.
[562] S. Odani, T. Koide, and T. Ikenaka, *J. Biochem.* (*Japan*), 1971, **70**, 925.

inhibitory activity in the Kunitz soybean inhibitor. A tetrameric chymotrypsin inhibitor from potatoes [563] appears to dissociate in the presence of chymotrypsin to form complexes of the type I_2E and IE.

Egg-white contains two well-characterized protease inhibitors: ovomucoid and ovoinhibitor (which is about twice as large); the inhibitory properties of ovomucoid from different species differ greatly. In contrast with other inhibitors of chymotrypsin and trypsin, turkey ovomucoid and chicken ovoinhibitor will inhibit elastase;[564] ovomucoid appears to have a common anti-elastase/anti-chymotrypsin site different from the trypsin site, whereas there are three distinct sites in the ovoinhibitor. A study of several avian ovoinhibitors [565] suggests two anti-tryptic and two anti-chymotryptic sites (the latter absent in penguin ovoinhibitor) with an arginyl bond at the anti-tryptic site; in contrast most avian ovomucoids have lysyl active sites. The purification of chicken ovoinhibitor free from ovomucoid by affinity chromatography on a column of agarose–chymotrypsin [232] (see Section 2E) rests on the fact that chicken ovomucoid inhibits only trypsin whereas ovoinhibitor inhibits both trypsin and chymotrypsin, and is therefore selectively retained by the column. Although chicken ovomucoid does not inhibit human trypsin (whereas it does inhibit the bovine enzyme) the two proteins do form a complex, and several bond cleavages occur, one of which is at the active site of the inhibitor;[566] the modified inhibitor will still inhibit bovine trypsin.

Protease inhibitors from human nasal secretion,[567] human sperm plasma,[568] and human tears [569] are being studied, although in some cases the small amounts available are somewhat restrictive.

Neutral Proteases, Acid Proteases, and Thiol Proteases. Some reference to neutral proteases has already been made in the section on chain cleavage (Section 2D). The heterogeneity of commercial Pronase from *S. griseus* has been shown by gel electrophoresis to extend to fourteen bands with various enzymatic activities,[570] and an improved fractionation on carboxymethyl-Sephadex has been described [571] which resolves the major endo- and exopeptidase activities (three of each). The endopeptidases, it will be recalled, have -Asp-Ser-Gly- at their active centres like the mammalian serine proteases. Sequence work on *S. griseus* trypsin was reported last year and

[563] H.-D. Belitz, K.-P. Kaiser, and K. Santarius, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 420.
[564] A. Gertler and G. Feinstein, *European J. Biochem.*, 1971, **20**, 547.
[565] W.-H. Liu, G. E. Means, and R. E. Feeney, *Biochim. Biophys. Acta*, 1971, **229**, 176.
[566] J. Travis, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 793.
[567] K. Hochstrasser, H. Haendle, R. Reichert, E. Werle, and S. Schwarz, *Z. physiol. Chem.*, 1971, **352**, 954.
[568] E. Fink, E. Jaumann, H. Fritz, H. Ingrisch, and E. Werle, *Z. physiol. Chem.*, 1971, **352**, 1591.
[569] F. Kueppers, *Biochim. Biophys. Acta*, 1971, **229**, 845.
[570] B. Löfqvist and L.-B. Sjöberg, *Acta Chem. Scand.*, 1971, **25**, 1663.
[571] L. Jurášek, P. Johnson, R. W. Olafson, and L. B. Smillie, *Canad. J. Biochem.*, 1971, **49**, 1195.

the sequences of the peptic peptides of *S. griseus* protease A, which has wide specificity, have recently been determined.[572] The high homology with other serine proteases, both from mammalian sources and from micro-organisms, can be clearly seen in the N-terminal sequences:[572a]

S. griseus protease A	Ile -Ala-Gly-Gly-Glu-Ala-
Bovine trypsin	Ile -Val-Gly-Gly-Try-Thr-
Bovine chymotrypsin A	Ile -Val-Asn-Gly-Glu-Glu-
Porcine elastase	Val-Val-Gly-Gly-Thr-Glu-
Myxobacter α-lytic protease	Ala-Asn-Ile -Val-Gly-Gly-Ile -Glu-

and in the sequences around the disulphide bridges,[572b] where the homology with *Myxobacter* α-lytic protease and *S. griseus* trypsin is particularly striking. These are important regions of the sequence since one of the bridge peptides contains the active-site serine, and the other the active-site histidine.

Elastase	-Thr-Ala-Ala-His-Cys-,	-Gly-Asp-Ser-Gly-Gly-Pro-Leu-
Chymotrypsin	-Thr-Ala-Ala-His-Cys-,	-Gly-Asp-Ser-Gly-Gly-Pro-Leu-
S. griseus trypsin	-Thr-Ala-Ala-His-Cys-,	-Gly-Asp-Ser-Gly-Gly-Pro-Met-
α-Lytic protease	-Thr-Ala-Gly-His-Cys-,	-Gly-Asp-Ser-Gly-Gly-Ser-Trp-
S. griseus protease A	-Thr-Ala-Gly-His-Cys-,	-Gly-Asp-Ser-Gly-Gly-Ser-Leu-

A comparison [573] of the amino-acid sequences of α-lytic protease and elastase was included in last year's Report (p. 81); based on this homology, and the known three-dimensional structure for elastase, a model was built for α-lytic protease [573] (see last year's Report). Elastase-like enzymes from Pronase which hydrolyse acetyl-L-Ala-L-Ala-L-Ala-OMe are also being investigated.[574] A rapid method for removing metal ions from metallo-neutral-proteases uses gel filtration in columns equilibrated with chelator [575] and has been employed to prepare the apoenzymes of thermolysin and the neutral proteases from *B. subtilis* and *B. cereus*.

The acid proteases constitute another 'family' of enzymes. The best characterized is pepsin, and the extensive studies of its specificity and mechanism of action have been well reviewed,[576] particularly in relation to subsite specificity, *i.e.* the influence on the enzymic cleavage of interactions at sites somewhat removed from the actual catalytic site. Different subsites presumably accounted for the significant differences in the rate of hydrolysis of the Phe–Phe bond by different acid proteases,[577] and penicillopepsin [578] showed the same specificity, with a requirement for

[572] (*a*) P. Johnson and L. B. Smillie, *Canad. J. Biochem.*, 1971, **49**, 1083; (*b*) P. Johnson and L. B. Smillie, *ibid.*, p. 548.
[573] A. D. McLachlan and D. M. Shotton, *Nature New Biol.*, 1971, **229**, 202.
[574] A. Gertler and M. Trop, *European J. Biochem.*, 1971, **19**, 90.
[575] J. Feder and L. R. Garrett, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 943.
[576] J. S. Fruton, *Adv. Enzymol.*, 1970, **33**, 401.
[577] I. M. Voynick and J. S. Fruton, *Proc. Nat. Acad. Sci. U.S.A.* 1971, **68**, 257.
[578] G. Mains, M. Takahashi, J. Šodek, and T. Hofmann, *Canad. J. Biochem.*, 1971, **49**, 1134.

an extended binding site. The *X*-ray studies in progress on pepsin, rennin, and penicillopepsin will show whether an extended binding site is indeed conserved. Preliminary crystallographic studies of the acid protease from *Rhizopus chinensis* have also been reported.[579] The homology between pig pepsin and calf rennin (chymosin) at the N-terminus (reported last year) has now been extended [580] by other workers with sequenator determination of the first 16 residues of leucyl-pepsin (the additional N-terminal leucine residue resulting from activation of pepsinogen with thermolysin, rather than the usual pepsin):

Rennin ↓Gly-Glu-Val-Ala-Ser -Val-**Pro-Leu**-Thr-**Asn-Tyr-Leu-Asn**-Ser-
Leu-pepsin Leu-Ile -Gly-Asp-Glu-**Pro-Leu**-Glu-**Asn-Tyr-Leu-Asn**-Thr-
↑
Rennin Gln-**Tyr-Phe**-
Leu-pepsin Glu-**Tyr-Phe**-

The greatest differences are near the activation points (shown with arrows); the residue preceding Ile-1 in pepsin is now given as Leu and not Glu (see last year's Report), making it easier to explain the different bond cleavages to give pepsin and rennin. Comparison of tryptic peptides obtained after maleylation of prorennin (prochymosin) and pepsinogen showed that the activation peptides were, however, homologous:[581]

Pepsinogen Leu-Val-Lys-Val-**Pro-Leu**-Val-Arg-Lys-Arg-**Ser-Leu-Arg**-
Prorennin Ala-Glu-Ile -Thr-Arg-Ile -**Pro-Leu**-Try-Lys-Gly-Lys-**Ser-Leu-Arg**-

In last year's Report (p. 84) the C-terminal sequences of pig pepsin, human pepsin, and human gastricsin were compared. The 19 C-terminal residues of bovine pepsin have now been determined [582] and appear (not surprisingly) to resemble the C-terminal region of human and pig pepsins more than that of calf rennin:

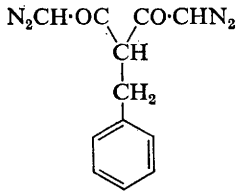
Pig pepsin -Gln-**Tyr**-Tyr-Thr-**Val-Phe-Asp-Arg**-Ala-**Asn-Asn**-Lys-Val-
Bovine pepsin -Gln-**Tyr**-Phe-Thr-**Val-Phe-Asp-Arg**-Gly-**Asn-Asn**-Gln-Ile
Calf rennin [583] -Glu-**Tyr**-Tyr-Ser -**Val-Phe-Asp-Arg**-Ala-**Asn-Asn**-Leu-Val-

Pig pepsin **Gly-Leu-Ala**-Pro-Val-Ala
Bovine pepsin **Gly-Leu-Ala**-Pro-Val-Ala
Calf rennin [583] **Gly-Leu-Ala**-Lys-Ala-Ile

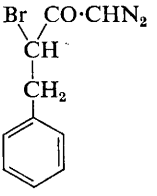
Stoicheiometric labelling of the active site of bovine pepsin with *N*-diazoacetyl[^{14}C]norleucine methyl ester (an inhibitor) [584] led to identification of the sequence around a reactive aspartyl residue as -Ile-Val-Asp-Thr-Gly-Thr-Ser-, identical with the sequence reported earlier for porcine pepsin,

[579] I.D.A. Swan, *J. Mol. Biol.*, 1971, **60**, 405.
[580] V. M. Stepanov, E. A. Timokhina, L. A. Baratova, L. P. Belyanova, V. P. Korzhenko, and I. G. Zhukova, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1482.
[581] B. Foltmann and B. Andersen, *F.E.B.S. Letters*, 1971, **17**, 87.
[582] K. T. Rasmussen and B. Foltmann, *Acta Chem. Scand.*, 1971, **25**, 3873.
[583] B. Foltmann, *Phil. Trans. Roy. Soc.*, 1970, **B257**, 147.
[584] P. A. Meitner, *Biochem. J.*, 1971, **124**, 673.

and homologous with the sequence -Ile-Ala-Asp-Thr-Gly-Thr-Thr- reported for penicillopepsin from *Penicillium janthinellium*;[585] the active site of the minor pepsin from pig (pepsin C) also had -Ile-Val-Asp-Thr- at the active centre.[586] The sequence around the phosphoserine residue in pig pepsin has now been extended to 11 residues,[587] the tryptophan-containing peptides have been re-examined,[588] and the sequences of some chymotryptic peptides from the N-terminal region have been determined in the same laboratory.[589] Two of the three disulphide bonds in pepsin and pepsinogen can be reversibly reduced without loss of activity and have now been identified.[590] The bisdiazoketone (12), designed as a bifunctional inhibitor of pepsin and related acid proteases, reacted stoicheiometrically (1 : 1) and intramolecularly with pepsin;[591] 50% of the label was bound in alkali-stable linkage but the site of reaction was not identified. The bromodiazoketone (13) was also studied [591] and a spectrophotometric assay for pepsin using phenyl sulphate as a specific substrate has been reported.[592]



(12)



(13)

The bovine pepsinogens and pepsins can be separated chromatographically [593] by virtue of their different organic phosphate content, while their amino-acid compositions are very similar. Three bovine pepsinogens have N-terminal Ser-Val- and C-terminal -Val-Ala and contain carbohydrate; the corresponding pepsins all have N-terminal Val, have C-terminal Ala, and are carbohydrate-free.[593a] Bovine pepsinogens and pepsin are also under study in other laboratories.[594, 595] A canine pepsinogen,[596] with one mole of organic phosphate and N-terminal Ala-Ile-, resembles other mammalian

[585] J. Sodek and T. Hofmann, *Canad. J. Biochem.*, 1970, **48**, 1014.
[586] J. Kay and A. P. Ryle, *Biochem. J.*, 1971, **123**, 75.
[587] E. A. Vakhitova, M. M. Amirkhanyan, and V. M. Stepanov, *Biochemistry* (*U.S.S.R.*), 1970, **35**, 1009.
[588] V. I. Vasenev, Yu. S. Kuznetsov, and V. M. Stepanov, *Biochemistry* (*U.S.S.R.*), 1970, **35**, 690.
[589] R. A. Balyulis and V. M. Stepanov, *Biochemistry* (*U.S.S.R*), 1971, **36**, 297.
[590] Y. Nakagawa and G. E. Perlmann, *Arch. Biochem. Biophys.*, 1971, **144**, 59.
[591] S. S. Husain, J. B. Ferguson, and J. S. Fruton, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2765.
[592] T. P. Stein, T. W. Reid, and D. Fahrney, *Analyt. Biochem.*, 1971, **41**, 360.
[593] (*a*) P. A. Meitner and B. Kassell, *Biochem. J.*, 1971, **121**, 249; (*b*) H. M. Lang and B. Kassell, *Biochemistry*, 1971, **10**, 2296.
[594] J. Antonini and B. Ribadeau Dumas, *Biochimie*, 1971, **53**, 321.
[595] A. Z. Vafin, *Biochemistry* (*U.S.S.R.*), 1971, **36**, 35.
[596] J. P. Marciniszyn, jun., and B. Kassell, *J. Biol. Chem.*, 1971, **246**, 6560.

pepsinogens in its content of all amino-acids except methionine, which is high, reminiscent of dogfish and chicken pepsinogens. It has been proposed that the conversion of pepsinogen into pepsin at acid pH involves an intramolecular reaction in which the zymogen cleaves itself.[597]

The structure of papain, the best characterized of the thiol proteases, has recently been reviewed.[598] Clostripain, like trypsin, shows specificity for arginyl and lysyl bonds, but particularly arginyl because of the high affinity of clostripain for the guanidino-group, demonstrated on specific substrate.[599] It will hydrolyse the Arg–Pro bond and *p*-nitrophenyl *p'*-guanidinobenzoate which with trypsin gives a stable acyl enzyme and is therefore useful as a specific active-site titrant. Clostripain is inhibited by TLCK, and the site of alkylation has been tentatively identified as a thiol residue;[600] this might be a useful way of inhibiting clostripain contamination of collagenase.[600]

B. Lysozyme and α-Lactalbumin.—A 6 Å *X*-ray analysis of human urine leukaemic lysozyme [601] confirms the high structural homology with hen egg-white lysozyme predicted from the similarities in primary structure.[602] As expected, the homology of the 'internal' residues (74%) is higher than that of the 'external' residues (52%) although significantly lower than the internal homology in chymotrypsin and elastase, where the overall homology is, in fact, lower. The active site is highly conserved but four of the five residues in the extended binding site are different. The two lysozymes are identical in 77 of the 129 positions (with a deletion and an insertion in the human enzyme) (Figure 15). An almost complete sequence of the lysozyme from human milk has also been presented,[603] and appears to be identical with the human leukaemic lysozyme except for a slight difference in the position of the deletion and an amide assignment.[604] Primary structural studies confirm that multiple duck lysozymes are due to multiple alleles at one locus,[605] unlike the case of the Black Swan which has two non-allelic genes for lysozyme (see Vol. 3 of these Reports). It is curious that the first 30 residues determined (using a sequenator) for goose lysozyme [602]

Arg-Thr-Asp-Cys-Tyr-Gly-Asn-Val-Asn-Arg-Ile-Asp-Thr-Thr-Gly-Ala-Ser-Cys-Lys-Thr-Ala-Lys-Pro-Glu-Gly-Ile-Ser-Tyr-Cys-Gly-

reveal no similarity with other avian lysozymes, suggesting that a different non-allelic gene is being expressed, possibly one of the two expressed in the

[597] M. Bustin and A. Conway-Jacobs, *J. Biol. Chem.*, 1971, **246**, 615.
[598] J. Drenth, J. N. Jansonius, R. Koekoek, and B. G. Wolthers, *Adv. Protein Chem.*, 1971, **25**, 79.
[599] P. W. Cole, K. Murakami, and T. Inagami, *Biochemistry*, 1971, **10**, 4246.
[600] W. H. Porter, L. W. Cunningham, and W. M. Mitchell, *J. Biol. Chem.*, 1971, **246**, 7675.
[601] C. C. F. Blake and I. D. A. Swan, *Nature New Biol.*, 1971, **232**, 12.
[602] R. E. Canfield, S. Kammerman, J. H. Sobel, and F. J. Morgan, *Nature New Biol.*, 1971, **232**, 16.
[603] P. Jollès, *Chimia* (*Switz.*), 1971, **25**, 1.
[604] J. Hermann, J. Jollès, and P. Jollès, *European J. Biochem.*, 1971, **24**, 12.
[605] E. M. Prager and A. C. Wilson, *J. Biol. Chem.*, 1971, **246**, 523.

```
          1                  10                  20
HLL  Lys-Val-Phe-Glu-Arg-Cys-Glu-Leu-Ala-Arg-Thr-Leu-Lys-Arg-Leu-Gly-Met-Asp-Gly-Tyr-Arg-Gly-Ile-Ser-Leu-Ala-Asn-
HEL  Lys-Val-Phe-Gly-Arg-Cys-Glu-Leu-Ala-Ala-Ala-Met-Lys-Arg-His-Gly-Leu-Asp-Asn-Tyr-Arg-Gly-Tyr-Ser-Leu-Gly-Asn-
BAL  Glu-Gln-Leu-Thr-Lys-Cys-Glu-Val-Phe-Arg-Glu-Leu-Lys-  -Asp-Leu-Lys-Gly-Tyr-Gly-Gly-Val-Ser-Leu-Pro-Glu-
               30                  40                  50
HLL  Trp-Met-Cys-Leu-Ala-Lys-Trp-Glu-  Ser-Gly-Tyr-Asn-Thr-Arg-Ala-Thr-Asn-Tyr-Asn-Ala-Gly-Asp-Arg-Ser-Thr-Asp-
HEL  Trp-Val-Cys-Ala-Ala-Lys-Phe-Glu-  Ser-Asn-Phe-Asn-Thr-Gln-Ala-Thr-Asn-Arg-Asn-Thr-  -Asp-Gly-Ser-Thr-Asp-
BAL  Trp-Val-Cys-Thr-Thr-  -Phe-His-Thr-Ser-Sly-Tyr-Asp-Thr-Glu-Ala-Ile-Val-Glu-Asn-  -Asn-Gln-Ser-Thr-Asp-
                    60                  70                  80
HLL  Tyr-Gly-Ile-Phe-Gln-Ile-Asn-Ser-Arg-Tyr-Trp-Cys-Asn-Asp-Gly-Lys-Thr-Pro-Gly-Ala-Val-Asn-Ala-Cys-His-Leu-Ser-
HEL  Tyr-Gly-Ile-Leu-Gln-Ile-Asn-Ser-Arg-Trp-Trp-Cys-Asn-Asp-Gly-Arg-Thr-Pro-Gly-Ser-Arg-Asn-Leu-Cys-Asn-Ile-Pro-
BAL  Tyr-Gly-Leu-Phe-Gln-Ile-Asn-Asn-Lys-Ile-Trp-Cys-Lys-Asn-Asp-Gly-Asp-Pro-His-Ser-Ser-Asn-Ile-Cys-Asn-Ile-Ser-
                         90                  100
HLL  Cys-Ser-Ala-Leu-Leu-Gln-Asp-Asn-Ile-Ala-Asp-Ala-Val-Ala-Cys-Ala-Lys-Arg-Val-Arg-  -Asp-Pro-Gln-Gly-Ile-Arg-
HEL  Cys-Ser-Ala-Leu-Leu-Ser-Ser-Asp-Ile-Thr-Ala-Ser-Val-Asn-Cys-Ala-Lys-Lys-Ile-Val-Ser-Asp-Gly-Asp-Gly-Met-Asn-
BAL  Cys-Asp-Lys-Phe-Leu-Asn-Asn-Asp-Leu-Thr-Asn-Asn-Ile-Met-Cys-Val-Lys-Lys-Ile-Leu-  -Asp-Lys-Val-Gly-Ile-Asn-
          110                 120                 129
HLL  Ala-Trp-Val-Ala-Trp-Arg-Asn-Arg-Cys-Gln-Asn-Arg-Asp-Val-Arg-Gln-Tyr-Val-Gln-Gly-Cys-Gly-  -Val
HEL  Ala-Trp-Val-Ala-Trp-Arg-Asn-Arg-Cys-Lys-Gly-Thr-Asp-Val-Gln-Ala-Trp-Ile-Arg-Gly-Cys-Arg-  -Leu
BAL  Tyr-Trp-Leu-Ala-His-Lys-Ala-Leu-Cys-Ser-Glu-Lys-Leu-Asp-Gln-  -Trp-Leu-  -Cys-Lys-Glu-Leu
```

Figure 15 *Amino-acid sequence of human leukaemia lysozyme (HLL), hen egg-white lysozyme (HEL), and bovine α-lactalbumin (BAL). The numbers correspond to the sequence of HLL.*

Black Swan (unfortunately not the most readily available source of lysozyme for sequence studies, except perhaps to Australians!).

Immunological cross-reactivity is certainly a useful method for investigating similarities between proteins, particularly for comparing different species. However, since antigenic determinants will lie on the surface of a protein, which is also well established as the location of most of the amino-acid changes in different species, immunological cross-reaction can only give information about the most dissimilar regions of the protein. Lack of immunological cross-reactivity, therefore, does not necessarily indicate major differences in tertiary structure (although an antigenic determinant may have several elements brought into proximity by folding) (see also last year's Report). These points are borne out in several studies of lysozyme and α-lactalbumin. Thus in a series of clearly homologous avian lysozymes, sequences which showed less than 40% homology did not cross-react,[606] in keeping with the fact that most evolutionary changes occur at the surfaces of proteins. A better measure of similarity in protein sequences, therefore, is the immunological cross-reactivity of the unfolded molecules; for example, after reduction and carboxymethylation, hen egg-white lysozyme and bovine α-lactalbumin cross-reacted[607] whereas the native molecules did not. Human leukaemia lysozyme and hen egg-white lysozyme similarly cross-reacted only after reduction and carboxymethylation,[608] confirming that immunological cross-reactivity is not a necessary consequence of high sequence homology. The unique antigenic determinant in lysozyme, formed from the loop peptide (residues 60—83), has been further studied; antibodies to the loop peptide have been partially purified[609] and the peptide containing residues 64—82 has been synthesized by the solid-phase method.[610] Nitration of hen lysozyme suggested that Tyr-20 and Tyr-23 belong to the same antigenic determinant,[611] and the selective nitration and acetylation of human lysozyme did not occur at the active site.[612]

The isolation and properties of α-lactalbumin from various sources have been described.[613, 614] The homology between the amino-acid sequences of bovine α-lactalbumin and hen egg-white lysozyme (Figure 15) makes it almost certain that the three-dimensional structures will be basically the same; and it has, of course, already been shown that the α-lactalbumin sequence can be fitted to the lysozyme model. Energy calculations based on this model for the α-lactalbumin structure have, however, been taken as

[606] E. M. Prager and A. C. Wilson, *J. Biol. Chem.*, 1971, **246**, 5978.
[607] R. Arnon and E. Maron, *J. Mol. Biol.*, 1971, **61**, 225.
[608] N. Arnheim, J. Sobel, and R. Canfield, *J. Mol. Biol.*, 1971, **61**, 237.
[609] E. Maron, C. Shiozawa, R. Arnon, and M. Sela, *Biochemistry*, 1971, **10**, 763.
[610] R. Arnon, E. Maron, M. Sela, and C. B. Anfinsen, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1450.
[611] A. D. Strosberg, B. van Hoeck, and L. Kanarek, *European J. Biochem.*, 1971, **19**, 36.
[612] R. L. Fawcett, T. J. Limbird, S. L. Oliver, and C. L. Borders, jun., *Canad. J. Biochem.*, 1971, **49**, 816.
[613] D. V. Schmidt and K. E. Ebner, *Biochim. Biophys. Acta*, 1971, **243**, 273.
[614] N. I. Phillips and R. Jenness, *Biochim. Biophys. Acta*, 1971, **229**, 407.

support [615] for results of an earlier small-angle *X*-ray study in solution [616] which suggested that the gross shapes of lysozyme and α-lactalbumin differed considerably, but these results [616] have now been shown to be invalid [617] and explicable in terms of α-lactalbumin dimer in the solution. A theoretical treatment,[618] which predicted correctly all the helices in lysozyme, predicted helices for the same regions of α-lactalbumin (with a greater helix content in the C-terminal region). On the other hand, marked differences in the susceptibility of the disulphide bonds in the two proteins to reduction and in behaviour in denaturing media [619] were taken as a warning of the dangers of assuming similar conformations from sequence homology. However, slight differences in reactivity can almost certainly be explained in terms of differences in side-chain orientation, and it is likely that the main conformational element, the polypeptide backbone, is indeed very similar in lysozyme and α-lactalbumin. Full details have now been published of the purification of the A protein of lactose synthetase on an agarose–α-lactalbumin affinity column,[620] clear indication that a complex does indeed form between the two proteins. The molecular weight of the A protein, at 112 000, is higher than the generally accepted value.

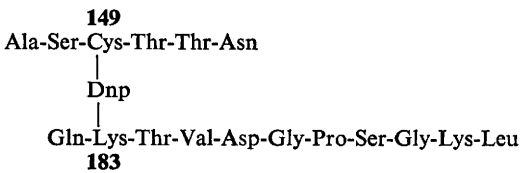
The almost complete amino-acid sequence (157 residues) of λ-phage endolysin has been published.[621] There is no structural similarity between λ-endolysin and T_4 phage or egg-white lysozyme, and it is likely that the enzyme is not a neuraminidase like the other two.[621] Preliminary structural work is under way on the *NO*-diacetylmuramidase (m. wt. 23 400) of a fungus (*Chalaropsis* species) [622] and the characterization of turnip lysozyme (m. wt. 25 000) has been described.[623] This resembles papaya latex lysozyme in having chitinase activity as well as normal lysozyme activity. The papaya enzyme (m. wt. 28 000), with four disulphide bridges, can be reversibly reduced and denatured.[624]

C. Dehydrogenases.—Preliminary sequence studies [625] show that glyceraldehyde 3-phosphate dehydrogenase (GPDH) from the red kangaroo (*Megaleia rufa*) is very similar to the enzyme from the pig and that the sequence of 17 residues around the active-site cysteine is conserved in this species, as in others. An improved preparation has been reported [626] for GPDH apoenzyme from *B. stearothermophilus*, which is much more stable

[615] W. R. Krigbaum and B. H. Rubin, *Biochim. Biophys. Acta*, 1971, **229**, 368.
[616] W. R. Krigbaum and F. R. Kügler, *Biochemistry*, 1970, **9**, 1216.
[617] E. K. Achter and I. D. A. Swan, *Biochemistry*, 1971, **10**, 2976.
[618] P. N. Lewis and H. A. Scheraga, *Arch. Biochem. Biophys.*, 1971, **144**, 584.
[619] A. F. S. A. Habeeb and M. Z. Atassi, *Biochim. Biophys. Acta*, 1971, **236**, 131.
[620] I. P. Trayer and R. L. Hill, *J. Biol. Chem.*, 1971, **246**, 6666.
[621] M. Imada and A. Tsugita, *Nature New Biol.*, 1971, **233**, 230.
[622] J. W.-K. Shih and J. H. Hash, *J. Biol. Chem.*, 1971, **246**, 994.
[623] I. Bernier, E. Van Leemputten, M. Horisberger, D. A. Bush, and P. Jollès, *F.E.B.S. Letters*, 1971, **14**, 100.
[624] A. O. Barel, M. Dolmans, and J. Léonis, *European J. Biochem.*, 1971, **19**, 488.
[625] R. J. Simpson and B. E. Davidson, *Austral. J. Biol. Sci.*, 1971, **24**, 263.
[626] K. Suzuki and J. I. Harris, *F.E.B.S. Letters*, 1971, **13**, 217.

than the apoenzyme from other species, and crystallizes in a form suitable for crystallographic studies. The enzyme from *E. coli* as prepared contains very little NAD^+ but (unlike the yeast enzyme) will crystallize only with bound coenzyme.[627] The amino-acid compositions and peptide maps of GPDH from various insects have been compared,[628] some properties of the enzyme from Ehrlich ascites tumour cells have been recorded,[629] and the isolation of GPDH strongly retained by human erythrocyte ghosts has been described.[630] Cariogenic streptococci possess both an NAD^+-linked and an $NADP^+$-linked GPDH which can be separated;[631] the NADP-specific enzyme is assigned a role in generation of NADPH since the organism lacks alternative pathways of production of the reduced coenzyme, namely the oxidative portion of the pentose-phosphate-shunt pathway and trans-hydrogenase activity.

The lysine residues of rabbit muscle GPDH that reacted most readily with pyridoxal phosphate [632] were residues 191 and 212, while there was no reaction at the 'active' lysine, Lys-183; the sequence around Lys-212, it will be recalled, is homologous with the region containing Lys-97 (the essential pyridoxal-binding lysine) in glutamate dehydrogenase (see last year's Report). For the first time the active-site cysteine, Cys-149, has been cross-linked [633] to the active lysine, Lys-183. Thus the peptide:



was one of four isolated when rabbit muscle apoenzyme was treated with four moles of 1,5-difluoro-2,4-dinitrobenzene per mole of tetramer. The cysteine and the lysine must, therefore, be within 5—6 Å of each other,[633] but it is not possible to distinguish between intra- and inter-monomer reaction (*i.e.* between an active site contributed entirely by one monomer and one situated between monomers). This is of particular interest in view of the apparent pairing of subunits ($\alpha_2\alpha'_2$) suggested by crystallographic studies and, more recently, by chemical studies (see below). Three elegantly conceived active-site-directed spin-labels for GPDH based, respectively, on iodoacetate, *p*-nitrophenyl acetate, and glyceraldehyde 3-phosphate have been described.[634]

[627] G. D'Alessio and J. Josse, *J. Biol. Chem.*, 1971, **246**, 4326.
[628] C. W. Carlson and R. W. Brosemer, *Biochemistry*, 1971, **10**, 2113.
[629] F. P. Mamaril and S. Green, *Arch. Biochem. Biophys.*, 1971, **147**, 583.
[630] M. J. A. Tanner and W. R. Gray, *Biochem. J.*, 1971, **125**, 1109.
[631] A. T. Brown and C. L. Wittenberger, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 217.
[632] B. G. Forcina, G. Ferri, M. C. Zapponi, and S. Ronchi, *European J. Biochem.*, 1971, **20**, 535.
[633] S. Shaltiel and M. Tauber-Finkelstein, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 484.
[634] W. Balthasar, *European J. Biochem.*, 1971, **22**, 158.

Differences in the binding of NAD^+ to GPDH from rabbit muscle and yeast continue to be of considerable interest. Equilibrium, temperature-jump, and stopped-flow kinetic studies,[635] and *X*-ray small-angle scattering [636] all support the concerted mechanism of binding for the yeast enzyme. Temperature-jump studies on the binding of NAD^+ to rabbit muscle GPDH under conditions that permitted direct comparison with the yeast enzyme still supported the sequential mechanism of conformational changes accompanying binding,[637] so the anomaly of substantial differences in coenzyme binding in enzymes that are highly homologous remains. Apparently it cannot be accounted for in terms of differences in the co-enzyme binding-sites [638] and must arise from differences elsewhere in the enzyme molecules.

Further evidence in support of the non-equivalence of the four active sites in GPDH, despite the chemical identity of the four subunits, is based on acylation of only two active sites in rabbit muscle [639] by either furyl-acryloyl phosphate or glyceraldehyde 3-phosphate. However, in another study [640] it was shown that all four sites of the lobster enzyme are acylated by 1,3-diphosphoglycerate, and that NAD^+ dissociates from these sites before acylation. Since hydrolysis of the sturgeon acyl-enzyme occurred fairly rapidly even in the absence of NAD^+, it was suggested [640] that the stability of the rabbit acyl-enzyme should be checked before taking for granted the non-equivalence of the four active sites. However, support for the functional non-identity of subunits and, further, for the existence of enzymically active dimers, comes also from another laboratory.[641]

A series of papers [642] gives full details of the determination of the amino-acid sequence of bovine liver glutamate dehydrogenase (GDH). The complete sequence (506 residues) was reproduced in last year's Report (p. 96); residue 341, shown there as Ile, is now given [642f] as Asn, the difference being attributed to true allelic variation and different starting material. Apart from the variety of lines of attack used for a sequence of this length, and the authors' valid plea for better methods of separating large peptides, it is also worth noting limited tryptic cleavage at many arginyl bonds; this was attributed to neighbouring acidic residues, either in the sequence itself or arising from maleylation of hydroxy-groups.[642c] The reaction of ox-liver

[635] (*a*) K.Kirschner, E. Gallego, I. Schuster, and D. Goodall, *J. Mol. Biol.*, 1971, **58**, 29; (*b*) K. Kirschner, *ibid.*, p. 51.
[636] H. Durchshlag, G. Puchwein, O. Kratky, I. Schuster, and K. Kirschner, *European J. Biochem.*, 1971, **19**, 9.
[637] G. G. Hammes, P. J. Lillford, and J. Simplicio, *Biochemistry*, 1971, **10**, 3686.
[638] D. Eby and M. E. Kirtley, *Biochemistry*, 1971, **10**, 2677.
[639] R. A. MacQuarrie and S. A. Bernhard, *J. Mol. Biol.*, 1971, **55**, 181.
[640] P. J. Harrigan and D. R. Trentham, *Biochem. J.*, 1971, **124**, 573.
[641] J. Ovadi, M. Telegedi, J. Batke, and T. Keleti, *European J. Biochem.*, 1971, **22**, 430.
[642] (*a*) M. Landon, M. D. Melamed, and E. L. Smith, *J. Biol. Chem.*, 1971, **246**, 2360; (*b*) M. Landon, D. Piszkiewicz, and E. L. Smith, *ibid.*, p. 2374; (*c*) W. J. Brattin, jun., and E. L. Smith, *ibid.*, p. 2400; (*d*) T. J. Langley and E. L. Smith, *ibid.*, p. 3789; (*e*) M. Landon, T. J. Langley, and E. L. Smith, *ibid.*, p. 3802; (*f*) M. Landon, T. J. Langley and E. L. Smith, *ibid.*, p. 3807.

GDH with trinitrobenzenesulphonic acid (TNBS) has received further attention,[643, 644] the availability of the sequence making possible the identification of residues involved in previous chemical modification studies. It has been shown [644] that rapid reaction with Lys-428 in three of the chains is followed by slower reaction at Lys-425 in the other three chains; no chain has both lysine residues modified. Despite the chemical identity of the subunits it is as though Lys-425 is unavailable until Lys-428 has reacted; neither residue appears to be involved in catalysis. One practical point of interest in this study [644] is that the use of sulphoethyl-Sephadex at elevated temperatures was found to improve recoveries of peptides modified with hydrophobic groups. An accompanying paper describes the complete characterization of the kinetic and molecular properties of the enzyme after modification.[645] Nitration of GDH, like trinitrophenylation, causes desensitization to allosteric inhibition by GTP, and details have now appeared [646] of the identification of the single tyrosine residue involved as Tyr-412. A kinetic study of the inactivation of GDH by pyridoxal [647] confirms the involvement of the ε-amino-group of Lys-97 in this reaction. The involvement of tyrosine and lysine residues in the reactivity of catalytic and regulatory sites in ox-liver GDH has also been studied by dinitrophenylation,[648] and the subunit structure has been examined using the electron microscope.[649] Preliminary work on GDH from pea roots [650] suggests that it is slightly smaller than the bovine enzyme, with m. wt. 208 000.

An *X*-ray crystallographic study at 5 Å of an abortive ternary complex of lactate dehydrogenase (dogfish M_4) has been published [651] and compared with previous results for the apoenzyme (see this chapter, Part II, Section 4G), and other experiments [652] indicate a covalent bond between the 3-position of the pyridine ring of NAD^+ and pyruvate in the ternary complex with chicken-heart LDH. It is of interest that substituents in this position of the ring in coenzymes influence the rate and equilibrium of hydride transfer.[652] A modifiable arginine residue which becomes inaccessible in the abortive ternary complex is implicated in substrate binding of pig-heart lactate dehydrogenase [653] and a histidine residue may also be involved;[654] other studies suggest a single tyrosine residue to be essential for activity.[655]

[643] C. E. Clark and K. L. Yielding, *Arch. Biochem. Biophys.*, 1971, **143**, 158.
[644] C. J. Coffee, R. A. Bradshaw, B. R. Goldin, and C. Frieden, *Biochemistry*, 1971, **10**, 3516.
[645] B. R. Goldin and C. Frieden, *Biochemistry*, 1971, **10**, 3527.
[646] D. Piszkiewicz, M. Landon, and E. L. Smith, *J. Biol. Chem.*, 1971, **246**, 1324.
[647] D. Piszkiewicz and E. L. Smith, *Biochemistry*, 1971, **10**, 4538.
[648] G. di Prisco, *Biochemistry*, 1971, **10**, 585.
[649] (*a*) R. Josephs, *J. Mol. Biol.*, 1971, **55**, 147; (*b*) A. M. Fiskin, E. F. J. van Bruggen, and H. F. Fisher, *Biochemistry*, 1971, **10**, 2396.
[650] E. Pahlich and K. W. Joy, *Canad. J. Biochem.*, 1971, **49**, 127.
[651] I. E. Smiley, R. Koekoek, M. J. Adams, and M. G. Rossmann, *J. Mol. Biol.*, 1971, **55**, 467.
[652] G. Di Sabato, *Biochemistry*, 1971, **10**, 395.
[653] J. Berghäuser and I. Falderbaum, *Z. physiol. Chem.*, 1971, **352**, 1189.
[654] J. Berghäuser, I. Falderbaum, and Ch. Woenckhaus, *Z. physiol. Chem.*, 1971, **352**, 52.
[655] D. Jeckel, R. Anders, and G. Pfleiderer, *Z. physiol. Chem.*, 1971, **352**, 769.

Beef-heart [656] and chicken-heart [657] LDH have *N*-acetyl-Ala-Thr- at the N-terminus and Leu at the C-terminus, in common with all other H_4 isoenzymes examined;[656] M_4 isoenzymes end with Phe and begin with *N*-acetyl-Ala-Ala- (rat liver and rabbit muscle), although dogfish M_4 has *N*-acyl-Thr-Ala-Leu-.

Carboxamidomethylation of an essential histidine residue in pig mitochondrial malate dehydrogenase (see last year's Report) has been confirmed from another laboratory;[658] there was no reaction at methionine although some confusion had arisen because of the similar chromatographic and electrophoretic properties of *N*-3-carboxamidomethyl-His-Gly-Gly and *S*-carboxamidomethylmethionine.[658] Selective modification of the enzyme with *N*-ethylmaleimide at pH 5 suggested an essential cysteine residue in the enzyme [659] and it will be interesting to see whether this turns out to occur in the cysteine-containing sequence that is conserved in several species.[660] Another study,[661] however, concluded that the involvement of a thiol group in the catalytic mechanism was unlikely, and that tyrosine is implicated in the substrate or coenzyme binding-site. The amino-acid sequences of several tryptic peptides from the supernatant malic dehydrogenase pig heart have been reported.[662]

The covalently linked flavin in beef-liver monoamine oxidase is bound through the 8α-carbon atom of the riboflavin ring to a cysteine residue in the sequence -Ser-Gly-Gly-Cys-Tyr-.[663] Beef-heart succinic dehydrogenase, which has flavin covalently attached to histidine (see last year's Report), has been shown in two laboratories [664, 665] to consist of one large subunit (m. wt. 68 000 or 70 000) and one small (30 000 or 27 000) subunit, both of which contain iron and labile sulphide whereas only the larger subunit contains flavin. Determination of the amino-acid sequence around the active-centre disulphide of the two thioredoxins from yeast (*Saccharomyces cerevisiae*) [666] permits comparison with the sequence from *E. coli* thioredoxin reported last year. The corrected sequence [667] for the same region

[656] L. D. Steginк, B. M. Sanborn, M. C. Brummel, and C. S. Vestling, *Biochim. Biophys. Acta*, 1971, **251**, 31.
[657] M. C. Brummel, B. M. Sanborn, and L. D. Steginк, *Arch. Biochem. Biophys.*, 1971, **143**, 330.
[658] E. M. Gregory, M. S. Rohrbach, and J. H. Harrison, *Biochim. Biophys. Acta*, 1971, **243**, 489.
[659] E. M. Gregory, F. J. Yost, jun., M. S. Rohrbach, and J. H. Harrison, *J. Biol. Chem.*, 1971, **246**, 5491.
[660] T. P. Fondy, G. B. Kitto, and G. A. Driscoll, *Biochemistry*, 1970, **9**, 1001.
[661] L. Siegel and J. S. Ellison, *Biochemistry*, 1971, **10**, 2856.
[662] L. M. Allen, J. Vanecek, and R. G. Wolfe, *Arch. Biochem. Biophys.*, 1971, **143**, 166.
[663] E. B. Kearney, J. I. Salach, W. H. Walker, R. L. Seng, W. Kenney, E. Zeszotek, and T. P. Singer, *European J. Biochem.*, 1971, **24**, 321; W. H. Walker, E. B. Kearney, R. L. Seng, and T. P. Singer, *ibid.*, p. 328.
[664] P. Righetti and P. Cerletti, *F.E.B.S. Letters*, 1971, **13**, 181.
[665] K. A. Davis and Y. Hatefi, *Biochemistry*, 1971, **10**, 2509.
[666] D. E. Hall, A. Baldesten, A. Holmgren, and P. Reichard, *European J. Biochem.*, 1971, **23**, 328.
[667] O. Berglund and A. Holmgren, *J. Biol. Chem.*, 1971, **246**, 5544.

of the thioredoxin induced by phage T_4 is also shown, with a Tyr/Val inversion [667] of the previous [668] sequence:

Thioredoxin II (yeast) -Phe-Ala-Thr-Trp-Cys-Gly-Pro-Cys-Lys-Met-Ile-Ala-Pro-Met-Ile-Glu-Lys-

Thioredoxin I (yeast) -Tyr-Ala-Thr-Trp-Cys-Gly-Pro-Cys-Lys-

E. coli -Trp-Cys-Gly-Pro-Cys-Lys-Met-Ile-

E. coli/T_4 -Cys-Val-Tyr-Cys-Asx-Asx-Ala-Lys-

Apart from the conserved disulphide loop there is no obvious homology between the yeast thioredoxins and the others, and whatever the reason for the two yeast thioredoxins they are clearly the products of different genes.

D. Aldolases.—Preliminary crystallographic studies on a new crystalline form of rabbit-muscle fructose 1,6-diphosphate (FDP) aldolase have been reported,[669] and the catalytic and immunochemical properties of various hybrids of aldolase have been described.[670] The amino-acid sequence around the active-site lysine in rabbit-muscle aldolase has been extended substantially by work in two laboratories [671, 672] but unfortunately the sequences are at variance in many positions. Close homology with aldolase from rabbit muscle is observed in the 28 residues around the active-site lysine of codfish-muscle enzyme,[673] and 12 residues around the lysine in the enzyme from lobster muscle [674] are identical with those in other muscle aldolases:

-Leu-Gln-Gly-Thr-Leu-Leu-Lys-Pro-Asn-Met-Val-Thr-

Aldolases from different species contain different amounts of methionine; the single methionine in the frog-muscle enzyme [675] is near the active-site lysine in the sequence -Lys-Pro-Asn-Met-. This lysine occurs in the sequence -Leu-Leu-Lys-Pro-Ser- in the FDP aldolase from spinach leaf.[676] The unique C-terminal sequence of sturgeon-muscle aldolase,[677] like the rabbit enzyme, is -His-Ala-Tyr, again evidence of homology. Aldolases from liver are also receiving attention. The amino-acid sequence at the

[668] O. Berglund and A. Holmgren, *J. Biol. Chem.*, 1970, **245**, 6036.
[669] E. G. Heidner, B. H. Weber, and D. Eisenberg, *Science*, 1971, **171**, 677.
[670] E. E. Penhoet and W. J. Rutter, *J. Biol. Chem.*, 1971, **246**, 318.
[671] C. Y. Lai and T. Oshima, *Arch. Biochem. Biophys.*, 1971, **144**, 363.
[672] M. Sajgó, *F.E.B.S. Letters*, 1971, **12**, 349.
[673] C. Y. Lai and C. Chen, *Arch. Biochem. Biophys.*, 1971, **144**, 467.
[674] A. Guha, C. Y. Lai, and B. L. Horecker, *Arch. Biochem. Biophys.*, 1971, **147**, 692.
[675] S.-M. Ting, C. L. Sia, C. Y. Lai, and B. L. Horecker, *Arch. Biochem. Biophys.*, 1971, **144**, 485.
[676] G. Ribereau-Gayon, T. Ramasarma, and B. L. Horecker, *Arch. Biochem. Biophys.*, 1971, **147**, 343.
[677] P. J. Anderson, *Canad. J. Biochem.*, 1971, **49**, 372.

active centre of the enzyme from rabbit liver has been extended [678] and the corresponding 36 residues have been established for the enzyme from ox liver [678] (see below); a 27-residue sequence included in this region of the enzyme from ox liver was determined independently.[679] It is worth noting that in a correction [679] of a previous sequence, a proposed serine insertion in the enzyme from liver vanishes, so this is clearly not the reason for the kinetic differences between the muscle and liver enzymes. The active-centre sequences (36 residues) of the enzymes from rabbit and ox liver are identical, but even more remarkable perhaps is the high homology between liver and muscle enzymes (see also last year's Report):[678]

Rabbit liver	-Ala-Leu-Asn-Asp-His-His-Val-Tyr-Leu-Glu-Gly-Thr-Leu-Leu-
Ox liver	-Ala-Leu-Asn-Asp-His-His-Val-Tyr-Leu-Glu-Gly-Thr-Leu-Leu-
Rabbit muscle	-Ala-Leu-Ser -Asp-His-His-Ile -Tyr-Leu-Glu-Gly-Thr-Leu-Leu-
Rabbit liver	Lys-Pro-Asn-Met-Val-Thr-Ala-Gly-His-Ala-Cys-Thr-Lys-Lys-
Ox liver	Lys-Pro-Asn-Met-Val-Thr-Ala-Gly-His-Ala-Cys-Thr-Lys-Lys-
Rabbit muscle	Lys-Pro-Asn-Met-Val-Thr-Pro-Gly-His-Ala-Cys-Thr-Gln-Lys-
Rabbit liver	Tyr-Thr-Pro-Gln-Glu-Val-Ala-Met-
Ox liver	Tyr-Thr-Pro-Gln-Glu-Val-Ala-Met-
Rabbit muscle	Tyr-Ser -His-Glu-Glu-Ile -Ala-Met-

A further study of the thiol groups of rabbit-muscle aldolase has confirmed the presence of an eighth, which is assigned to the N-terminal cyanogen bromide fragment.[680]

Aldolases from various other sources have been described. The enzyme from the muscle of the shark [681] (*Mustelus canis*) is similar to rabbit-muscle aldolase, and the enzyme from the liver of the domestic fowl is also unexceptional in its properties.[682a] Statistical treatment of amino-acid compositions, and immunochemical studies, support the idea that divergences between enzymes from different species are smaller than between the same enzyme from different tissues within a species.[682b] It is interesting that the green alga *Chlamydomonas reinhardii* contains two aldolases whose properties suggest that one is a Class I enzyme whereas the other belongs to Class II.[683]

Unfashionable though trimers may be, sedimentation and chemical evidence in support of a three-subunit structure for 2-keto-3-deoxy-6-phosphogluconate aldolase from *Pseudomonas putida* has been presented,[684]

[678] B. G. Forcina and R. N. Perham, *F.E.B.S. Letters*, 1971, **18**, 59.
[679] S.-M. Ting, C. Y. Lai, and B. L. Horecker, *Arch. Biochem. Biophys.*, 1971, **144**, 476.
[680] C. Y. Lai, C. Chen, J. D. Smith, and B. L. Horecker, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1497.
[681] C. E. Caban and L. F. Hass, *J. Biol. Chem.*, 1971, **246**, 6807.
[682] (*a*) R. R. Marquardt, *Canad. J. Biochem.*, 1971, **49**, 647; (*b*) R. R. Marquardt, *ibid.*, p. 658.
[683] A. M. Guerrini, T. Cremona, and E. C. Preddie, *Arch. Biochem. Biophys.*, 1971, **146**, 249.
[684] R. H. Hammerstedt, H. Möhler, K. A. Decker, and W. A. Wood, *J. Biol. Chem.*, 1971, **246**, 2069.

the strongest evidence perhaps being the generation of four hybrid species from native and maleylated enzyme. It is a Class I aldolase and the sequence around the substrate-binding lysine has been shown [685] to be:

-Phe-*Lys-Leu-Phe-Pro-Ala-Glu-Ile-Ser-Gly-Gly-Val-Ala-Ala-Ile-Lys-

There is clearly no similarity between this and the active site of the Class I fructose 1,6-diphosphate aldolases from muscle (see above). Inactivation of the enzyme with fluorodinitrobenzene has been re-examined [686] and the reaction of chlorodinitrobenzene with the isoenzymes of transaldolase from *Candida utilis* now appears [687] to be with cysteine, and not lysine as reported earlier. The 2-keto-4-hydroxyglutarate aldolase [688] and the 2-keto-4-hydroxybutyrate aldolase [689] from bovine liver are the same enzyme [689] and function *via* a Schiff-base mechanism, *i.e.* they are Class I aldolases.

E. Nucleases.—Staphylococcal nuclease, ribonucleases, and deoxyribonucleases are excellently reviewed in a single volume.[690] A high-resolution (2 Å) structure of an inhibitor complex of the staphylococcal nuclease has now been published,[691] and, with the primary structure of the enzyme from the same strain of *S. aureus* (Foggi) also available,[692] publication of a detailed mechanism of the catalytic action will doubtless soon follow. The enzyme from the Foggi strain has the same sequence as that of the V-8 strain published some years ago, with a change of Leu to His at position 124.

The elegant reconstitution experiments in which inactive fragments from limited tryptic cleavage of the enzyme regain acitivity are now well known, and further experiments continue to contribute to a general picture of protein folding. Two types of active complementing structures were formed simultaneously and in approximately equal amount when the fragments 1—126 and 49—149 were mixed.[693] In one case binding was between residues 1—126 and 111—149 and in the other between 1—48 and 49—149. The redundant portions (49—110 and 49—126 respectively) appeared to be flexible, external to the folded structure, and readily digested with trypsin. Residues 1—126 also bind to the cyanogen bromide fragment 99—149 to give an enzymically active complex from which residues 99—110 could be removed with trypsin. These experiments show that peptide bonds either between residues 48 and 50 (in an exposed loop in the *X*-ray structure) or

[685] D. C. Robertson, W. W. Altekar, and W. A. Wood, *J. Biol. Chem.*, 1971, **246**, 2084.
[686] L. R. Barran and W. A. Wood, *J. Biol. Chem.*, 1971, **246**, 4028.
[687] O. Tsolas, J. de Castro, and B. L. Horecker, *Arch. Biochem. Biophys*, 1971, **143**, 516.
[688] R. D. Kobes and E. E. Dekker, *Biochemistry*, 1971, **10**, 388.
[689] R. S. Lane, A. Shapley, and E. E. Dekker, *Biochemistry*, 1971, **10**, 1353.
[690] 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, and London, 3rd edn., 1971, vol. 4.
[691] A. Arnone, C. J. Bier, F. A. Cotton, V. W. Day, E. E. Hazen, jun., D. C. Richardson, J. S. Richardson, and A. Yonath, *J. Biol. Chem.*, 1971, **246**, 2302.
[692] J. L. Cone, C. L. Cusumano, H. Taniuchi, and C. B. Anfinsen, *J. Biol. Chem.*, 1971, **246**, 3103.
[693] H. Taniuchi and C. B. Anfinsen, *J. Biol. Chem.*, 1971, **246**, 2291.

somewhere between residues 113 and 124 can be cleaved without destroying the information required for formation of a functional structure, and that stable structures resembling the native enzyme can be formed in a number of ways when the minimum information requirement (the entire amino-acid sequence) is fulfilled. Breaks at both sites simultaneously are even permissible and the three fragments will recombine to generate (albeit low) enzymic activity.[694] Solid-phase synthetic studies confirmed (see above) that in the recombination system (1—126) + (99—149), the sequence (99—126) present in both fragments is contributed in the active functional complex by (1—126), since fragment (99—149) may be shortened by several residues without affecting productive complementation with fragment (1—126).[695] Trp-140 may also be replaced by Phe without ill effect.[695] This single tryptophan residue is probably important in stabilizing the structure rather than in catalysis, since modification of tryptophan in the native enzyme with nitrophenylsulphenyl halides disrupted the helical content substantially, with loss of only half the enzymic activity.[696] This is in agreement with the position of the tryptophan residue between two stretches of α-helix in the *X*-ray structure.[691] The importance of Asp-21, Arg-35, Asp-40, and Glu-43 (thought to be involved in binding the inhibitor and Ca^{2+} ions in the *X*-ray structure) has been demonstrated, using analogues of the peptide 6—47 made by solid-phase synthesis,[697] and testing for formation of active complex with 49—149. When charge was preserved (Asp → Glu, Arg → Lys, Glu → Asp) the fragments associated but the complex was inactive; if charge was not preserved there was no binding. The only exception was the substitution of Asp-40 by Asn, which gave rise to active complex. (It is worth noting that this information made available by solid-phase synthesis would not be easily available in any other way.)

Full details have now been published [698] of the solid-phase synthesis of bovine pancreatic ribonuclease A. As a bonus of the solid-phase method, the des-(21—25)-S-protein and S-protein itself were obtained by removing samples from the machine after the appropriate number of couplings; both combined equally well with S-peptide when mixtures were denatured and renatured, showing that the first five residues of S-protein are not required for the folding process.[698] ^{19}F N.m.r. has been used to observe changes in conformation when ribonuclease S, trifluoroacetylated at Lys-1 and Lys-7, associates with S-peptide,[699] and the crystallization of 41-Dnp-S-protein has been reported.[700] Elongation of all four disulphide bonds in ribonuclease A by *ca.* 3 Å by conversion into -S-Hg-S- (see these Reports for

[694] G. Andria, H. Taniuchi, and J. L. Cone, *J. Biol. Chem.*, 1971, **246**, 7421.
[695] I. Parikh, L. Corley, and C. B. Anfinsen, *J. Biol. Chem.*, 1971, **246**, 7392.
[696] I. Parikh and G. S. Omenn, *Biochemistry*, 1971, **10**, 1173.
[697] I. M. Chaiken and C. B. Anfinsen, *J. Biol. Chem.*, 1971, **246**, 2285.
[698] B. Gutte and R. B. Merrifield, *J. Biol. Chem.*, 1971, **246**, 1922.
[699] W. H. Huestis and M. A. Raftery, *Biochemistry*, 1971, **10**, 1181.
[700] D. S. Fung and M. S. Doscher, *Biochemistry*, 1971, **10**, 4099.

1969, 1970) caused conformational changes, but some activity was retained.[701] A further study [702] of the carboxymethylation of histidine residues in the enzyme (see last year's Report) shows that both His-12 and His-119 can be carboxymethylated in the same active site, even though it has generally been assumed that the carboxymethylations are mutually exclusive. The reaction can go to completion in time, but 17% of the species containing 3-Cm His-12 and 1-Cm His-119 is present even after 24 h. His-48 (which the *X*-ray structure shows to be buried) does not react but methionine and lysine do, and the sequence of events has been determined.[702]

Several ribonucleases from micro-organisms are also being studied. A much improved purification of ribonuclease T_1 from Takadiastase (an extract of the mould *Aspergillus oryzae*) has been published,[703] and a series of papers [704] gives full details of the determination of the amino-acid sequence of the enzyme reported in 1965. There are two disulphide bridges [(2—10) and (6—103)] and the sequence is quite unrelated to that of bovine ribonuclease A. Ribonuclease U_1 (from *Ustilago spheroides*) has been purified in two laboratories [705, 706] and ribonuclease N_1 (from *Neurospora crassa*) also described.[706] All three nucleases (T_1, U_1, and N_1) have two disulphide bridges (involving in U_1 the N-terminus [705] and in T_1 the 2-position [704a]) and molecular weights of *ca.* 11 000. A much larger ribonuclease (m. wt. 34 000) which cleaves after purine bases (particularly adenine) has been purified from *Octopus vulgaris*.[707]

Of the two tryptophan residues in bovine pancreatic deoxyribonuclease that can be modified chemically, one is essential for catalysis.[708] Unpublished work from another laboratory quoted in this paper [708] claims that the enzyme has a total of three tryptophan residues, not four.

F. Pyridoxal Phosphate Enzymes.—Studies of the primary structure of aspartate aminotransferase from pig heart [709] place 383 residues in sequence, out of a total of 400—430. Two additional thiol groups per dimer of 90 000 daltons become exposed when the aldimine form (PLP-enzyme)

[701] R. Sperling and I. Z. Steinberg, *J. Biol. Chem.*, 1971, **246**, 715.
[702] J. Bello and E. F. Nowoswiat, *European J. Biochem.*, 1971, **22**, 225.
[703] R. Fields, H. B. F. Dixon, and G. R. Law, *Biochem. J.*, 1971, **121**, 591.
[704] (*a*) K. Takahashi, *J. Biochem.* (*Japan*), 1971, **70**, 477, 603, 617, 803; (*b*) K. Takahashi, *ibid.*, p. 945.
[705] W. C. Kenney and C. A. Dekker, *Biochemistry*, 1971, **10**, 4962.
[706] J. Hashimoto, T. Uchida, and F. Egami, *J. Biochem.* (*Japan*), 1971, **70**, 903.
[707] F. De Lorenzo, G. Molea, and M. Molinaro, *Arch. Biochem. Biophys.*, 1971, **146**, 327.
[708] T. L. Poulos and P. A. Price, *J. Biol. Chem.*, 1971, **246**, 4041.
[709] (*a*) Yu. A. Ovchinnikov, A. A. Kiryushkin, Ts. A. Egorov, N. G. Abdulaev, A. P. Kiselev, N. N. Modyanov, E. V. Grishin, E. I. Vinogradova, M. Yu. Feigina, N. A. Aldanova, and V. M. Lipkin, *F.E.B.S. Letters*, 1971, **12**, 194; (*b*) Yu. A. Ovchinnikov, A. A. Kiryushkin, Ts. A. Egorov, N. G. Abdulaev, A. P. Kiselev, N. N. Modyanov, E. V. Grishin, A. P. Sukhikh, E. I. Vinogradova, M. Yu. Feigina, N. A. Aldanova, V. M. Lipkin, A. E. Braunstein, O. L. Polyanovsky, and V. V. Nosikov, *F.E.B.S. Letters*, 1971, **17**, 133.

is converted into the amine form (PMP + enzyme),[710] and inactivation of the apoenzyme with tetranitromethane can be correlated with nitration of a single tyrosine residue at the coenzyme binding site.[711] Some conclusions about steric relationships at the active site of aspartate aminotransferase can be drawn from a spectral study of the free and bound forms of the apoenzyme inhibitors *N*-(5′-phosphopyridoxyl)glutamic acid and its pyrrolidonecarboxylic acid analogue.[712] These resemble the holoenzyme–glutamate complex. The relation between the quite distinct soluble and mitochondrial forms of glutamic–oxaloacetic transaminase from pig-heart muscle is being investigated at the level of primary structure.[713] A cyanogen bromide fragment from each shows clear homology:[713a]

soluble	**-Ala-Asp-Arg-Ile-Leu-Ser-Met-**
mitochondrial	**-Ala-Asp-Arg-Ile-**Ile **-Ser-Met-**

but the amino-terminal sequences are not obviously related:[713b]

soluble	Ala-Pro-Pro-Ser-Val-Phe-Ala-**Glu**-Val-
mitochondrial	Ser-Ser-Trp-Trp-Ala-His-Val-**Glu**-Met-

and it will be interesting to see what further studies of the sequences hold in store.

Chemical evidence suggests [714] that the histidine decarboxylase from *Micrococcus* sp.n. has six chains, three with *N*-terminal Met and three with *N*-terminal X-Phe, where X awaits identification. It is worth recalling that the enzyme from *Lactobacillus* 30a (last year's Report, p. 103) had ten chains, five of which have N-terminal pyruvoylphenylalanyl residues, and perhaps speculating that a similar situation might obtain for the micrococcal enzyme. There has been some uncertainty about the subunit structure of glutamate decarboxylase; electron microscopy now reveals a hexamer for the enzyme from *E. coli* B.[715] The amino-acid sequence around the pyridoxal-binding lysine residue in glutamate decarboxylase reported last year (these Reports) can now be compared with the sequence determined for the pyridoxal site in arginine decarboxylase:[716]

Arg decarboxylase:

-Ala-Thr-His-Ser-Thr-His-$\overset{*}{\text{Lys}}$-Leu-Leu-Asn-Ala-Leu-Ser-Gln-Ala-Ser-Tyr-

[710] I. Cournil and M. Arrio-Dupont, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 40.
[711] C. Turano, D. Barra, F. Bossa, A. Ferraro, and A. Giartosio, *European J. Biochem.*, 1971, **23**, 349.
[712] R. M. Khomutov, H. B. F. Dixon, L. V. Vdovina, M. P. Kirpichnikov, Y. V. Morozov, E. S. Severin, and E. N. Khurs, *Biochem. J.*, 1971, **124**, 99.
[713] (*a*) T. Watanabe and H. Wada, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 1310; (*b*) H. Wada, T. Watanabe, and A. Miyatake, *ibid.*, p. 1318.
[714] V. N. Prozorovskii, S. R. Mardashev, and A. M. Sokhina, *Biochemistry* (*U.S.S.R.*), 1970, **35**, 788.
[715] C. M. To, *J. Mol. Biol.*, 1971, **59**, 215.
[716] E. A. Boeker, E. H. Fischer, and E. E. Snell, *J. Biol. Chem.*, 1971, **246**, 6776.

Glu decarboxylase:

-Ser-Ile -Ser -Ala-**Ser**-Gly-His-*Lys-Phe-

It remains to be seen whether the -His-*Lys- sequence will be a common feature in pyridoxal-phosphate-dependent decarboxylases.

It is interesting that the same peptide also occurs in the β_2-subunits of tryptophan synthetase from *E. coli* [717] and *Pseudomonas putida* [718] which are clearly homologous in the region containing the pyridoxal-phosphate-binding site:

E. coli	-Arg-Glu-Asp-Leu-Leu-His-Gly-Gly-Ala-His-*Lys-Thr-Asn-Gln-
Pseudomonas putida	-Arg-Glu-Glu-Leu-Asn-His-Thr-Gly-Ala-His-*Lys-Val-Asn-Asn-
E. coli	Val-Leu-Gly-Gln-Ala-Leu-Leu-Ala-Lys-
Pseudomonas putida	Cys-Ile -Gly-Gln-Val-Leu-Leu-Ala-Lys-

This is the first information on the primary structure of the β_2-subunit of the $\alpha_2\beta_2$ complex, whereas the structure of the α-subunit has been known for some time. It has thus been possible [719] to identify the cross-link introduced into the α-subunit by the bifunctional reagent bis(maleimidomethyl) ether (*cf.* modification of haemoglobins, p. 146) as Cys-80/Cys-117, with no reaction at the remaining cysteine, Cys-153; this appears to be 'buried' in native enzyme, but will, however, react with *N*-ethylmaleimide, as a result of a conformational change, in the presence of indole.[720] The cross-linked product from reaction with bis(maleimidomethyl) ether is functionally heterogeneous, with 70% of the modified α-subunit being devoid of independent indole glycerol phosphate activity and unable to form a functional complex with the β_2 subunit.[719] Similar loss of enzymic activity was observed when Cys-80 and Lys-108 were cross-linked with 1,5-difluoro-2,4-dinitrobenzene [721] but not when intramolecular methylene cross-links were introduced into the α-subunit with formaldehyde.[722] The sites of reaction in the latter case [722] were tentatively assigned as Asn-156/Ser-214 and Gln-218/Ser-232. It is worth remarking that studies such as these with cross-linking reagents of different span lengths are a nice and effective way of charting the surface of the protein molecule, with *X*-ray analysis being the final test of their validity. The many mutant forms of tryptophan synthetase which have been well characterized in Yanofsky's laboratory not only illuminate elegantly the effect of defined amino-acid substitution on

[717] R. Fluri, L. E. Jackson, W. E. Lee, and I. P. Crawford, *J. Biol. Chem.*, 1971, **246**, 6620.
[718] R. Maurer and I. P. Crawford, *J. Biol. Chem.*, 1971, **246**, 6625.
[719] W. B. Freedberg and J. K. Hardman, *J. Biol. Chem.*, 1971, **246**, 1439.
[720] W. B. Freedberg and J. K. Hardman, *J. Biol. Chem.*, 1971, **246**, 1449.
[721] J. K. Hardman and D. F. Hardman, *J. Biol. Chem.*, 1971, **246**, 6489.
[722] J. S. Myers and J. K. Hardman, *J. Biol. Chem.*, 1971, **246**, 3863.

protein function but also make this a particularly attractive protein for chemical modification studies.

A sequence of 42 residues containing the pyridoxal-phosphate-binding site of rabbit-muscle phosphorylase is as shown:[723]

-Arg-Val-Ser-Leu-Ala-Glx-Lys-Val-Ile-Pro-Ala-Ala-Asp-Leu-Ser-Glx-Glx-Ile-

Ser-Thr-Ala-Gly-Thr-Gln-Ala-Ser-Gly-Thr-Gly-Asp-Met-*Lys-Phe-Met-Gly-Arg-Thr-Leu(Glx,Asx,Thr)-Met-

Worth noting, perhaps, is the -*Lys-Phe- sequence, also found in the pyridoxal-phosphate-binding site of glutamate decarboxylase (see above). Fluorescence studies [724] support an earlier suggestion that pyridoxal phosphate is bound in a hydrophobic environment in the phosphorylase b dimer, and the particles seen in an electron-microscopic study [725] were taken to be phosphorylase b tetramers. Phosphorylases a and b from yeast have also been described.[726]

G. Other Enzymes.—The composition and C-terminal sequences of erythrocyte carbonic anhydrase from various mammalian species have been determined [727] (asterisked) and compared with existing sequences:

Human	B	-Val-Arg-Ala-Ser-Phe
	C	-Ile -Lys-Ala-Ser-Phe-Lys
Pig	B	-Lys-Ala-Ser-Phe
	C*	-$\frac{\text{Arg}}{\text{Lys}}$-Ser-Phe
Horse	B*	-Val-Arg-Ala-Phe-Phe
	C*	-Ile -Arg-Ala-Ser -Phe-Lys
Ox	B(C)	-Val-Arg-Gly-Phe-Pro-Lys
Sheep	C*	-Val-Arg-Val-Phe-Pro-Lys
Deer	C*	-Pro-Arg

They clearly fall into two groups: those containing C-terminal -Pro-$\frac{\text{Arg}}{\text{Lys}}$, and those that have Phe as the last or penultimate residue. It is suggested that there is a relationship between evolutionary grouping and the predominant carbonic anhydrase, and a scheme for the evolution of B- and C-type carbonic anhydrases by gene duplication from a primordial carbonic anhydrase C is proposed.[727] The enzyme from two species of shark appears to have properties similar to those of the mammalian enzymes.[728] Dinitrophenylation of human carbonic anhydrase B occurred [729] at the 3-position of His-204, also the site of reaction of iodo-

[723] A. W. Forrey, C. L. Sevilla, J. C. Saari, and E. H. Fischer, *Biochemistry*, 1971, **10**, 3132.
[724] D. C. Jones and R. W. Cowgill, *Biochemistry*, 1971, **10**, 4276.
[725] N. A. Kiselev and F. Ya Lerner, *J. Mol. Biol.*, 1971, **62**, 537.
[726] M. Fosset, L. W. Muir, L. D. Nielsen, and E. H. Fischer, *Biochemistry*, 1971, **10**, 4105.
[727] R. B. Ashworth, J. M. Brewer, and R. L. Stanford, jun., *Biochem. Biophys. Res. Comm.*, 1971, **44**, 667.
[728] J. R. Maynard and J. E. Colman, *J. Biol. Chem.*, 1971, **246**, 4455.
[729] P. Henkart and F. Dorner, *J. Biol. Chem.*, 1971, **246**, 2714.

acetate. The residue is not ascribed a role in catalysis although it does appear to be near the active site. Its reactivity towards FDNB is attributed to its high nucleophilicity, thought to be a consequence of a fairly hydrophobic environment.[729]

Additional evidence has been presented [730] that the three main isoenzymes (97%) of triose phosphate isomerase from rabbit muscle are of the form AA, BB, and AB; the nature of the difference between the chains has not yet been revealed in what is known of the amino-acid sequence. Happily the slight uncertainty about the sequence around the glutamic acid residue labelled by halogenoacetol phosphates (see last year's Report) has now disappeared with publication of full details from the two laboratories concerned [731, 732] and some correction in one case.[732] So the sequence is now:

-Trp-Val-Leu-Ala-Tyr-Glu-Pro-Val-Trp-Ala-Ile-Gly-Thr-Gly-Lys-

Amino-acid sequences around the five cysteine residues [733] give an additional 75 residues of the primary structure. Preliminary work on purified triose phosphate isomerase from human erythrocytes [734] shows three main bands on isoelectric focusing.

In the accepted model of the α-keto-acid dehydrogenase multienzyme complexes, a basic structural unit composed of transacylase chains binds the decarboxylase and dihydrolipoyl dehydrogenase enzymes. The subunit molecular weights of the component enzymes of the complexes from *E. coli* (Crookes' strain) have been redetermined [735] and some revision of accepted molecular weights was found necessary. The N-terminus of the pyruvate decarboxylase component (subunit m. wt. 90 000) from the pyruvate dehydrogenase complex of *E. coli* K 12 is Ser-[736, 737] and the C-terminal sequence is -Arg-Leu-Ala.[737] The transacetylase component (subunit m. wt. 80 000) of the same complex has the C-terminal sequence -Arg-Arg-(Val,Leu)-Met and no detectable N-terminus.[738] It has not yet been crystallized but a preliminary *X*-ray analysis of the corresponding enzyme, dihydrolipoyl trans-succinylase, from the α-ketoglutarate dehydrogenase complex has been reported.[739] The organization of the transacetylase and trans-succinylase units as seen in the electron microscope is very similar,

[730] W. K. G. Kietsch, P. G. Pentchev, and H. Klingenburg, *European J. Biochem.*, 1971, **23**, 77.
[731] J. C. Miller and S. G. Waley, *Biochem. J.*, 1971, **123**, 163.
[732] F. C. Hartman, *Biochemistry*, 1971, **10**, 146.
[733] J. C. Miller and S. G. Waley, *Biochem. J.*, 1971, **122**, 209.
[734] E. E. Rozacky, T. H. Sawyer, R. A. Barton, and R. W. Gracy, *Arch. Biochem. Biophys.*, 1971, **146**, 312.
[735] R. N. Perham and J. O. Thomas, *F.E.B.S. Letters*, 1971, **15**, 8.
[736] G. Dennert and D. Eaker, *F.E.B.S. Letters*, 1970, **6**, 257.
[737] O. Vogel and U. Henning, *European J. Biochem.*, 1971, **18**, 103.
[738] O. Vogel, H. Beikirch, H. Müller, and U. Henning, *European J. Biochem.*, 1971, **20**, 169.
[739] D. J. Derosier, R. M. Oliver, and L. J. Reed, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1135.

and the presence of 24 chains in cubic array seems confirmed by the *X*-ray results.[739] The reaction of the decarboxylase components of the pyruvate dehydrogenase complexes of *E. coli* [740] and pigeon breast muscle [741] with mercurials suggests that thiol groups may be involved in binding the cofactor thiamine pyrophosphate.

Ile-Ile-Tyr-Pro-Gly-Thr-Leu-Trp-Cys-Gly-His-Gly-Asn-Lys-Ser-Ser-Gly-Pro-
1 **10**
Asn-Glu-Leu-Gly-Arg-Phe-Lys-His-Thr-Asp-Ala-Cys-Cys-Arg-Thr-His-Asp-
20 **30**
Met-Cys-Pro-Asn-Val-Met-Ser-Ala-Gly-Glu-Ser-Lys-His-Gly-Leu-Thr-Asp-
40 **50**
Thr-Ala-Ser-Arg-Leu-Ser-Cys-Asn-Asp-Asn-Asp-Leu-Phe-Tyr-Lys-Asp-Ser-
60
Ala-Asp-Thr-Ile-Ser-Ser-Tyr-Phe-Val-Gly-Lys-Met-Tyr-Phe-Asn-Leu-Ile-Asn-
70 **80**
Thr-Lys-Cys-Tyr-Lys-Leu-Glu-His-Pro-Val-Thr-Gly-Cys-Gly-Glu-Arg-
90 **100**
Thr-Glu-Gly-Arg-Cys-Leu-His-Tyr-Thr-Val-Asp-Lys-Ser-Lys-Pro-Lys-Val-
110 **120**
Tyr-Gln-Trp-Phe-Asp-Leu-Arg-Lys-Tyr
129

Figure 16 *Amino-acid sequence of phospholipase-A from bee venom*

Pancreatic lipase has been reviewed [742] and the preliminary characterization of the colipase as a small protein (m. wt. 10 000) reported.[743] *Tout c'est la même chose*: histidine may be involved in the active site of pig pancreatic lipase.[744] Phospholipases from bee and snake venoms are being actively studied. The complete amino-acid sequence (129 residues) of phospholipase A_2 from the common European honey bee *Apis mellifica* has been determined (Figure 16);[745] the enzyme is a dimer of molecular weight 40 000 (about 30% of which is carbohydrate) [746] with four disulphide bridges; the pancreatic enzyme (123 residues), whose zymogen sequence was reported last year (p. 102), had six. There is no obvious homology between the two enzymes but it may be worth noting the C-terminal sequence -Lys-Lys-Tyr-Cys of the pancreatic enzyme and the C-terminal sequence -Arg-Lys-Tyr in bee-venom phospholipase. It is now clear that the phospholipases A_2 of the venoms of the snakes *Crotalus adamanteus* [747] and *C. atrox* [748] are also dimers (m. wt. *ca.* 30 000) with a high cystine

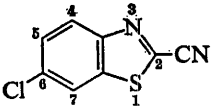
[740] E. R. Schwartz and L. J. Reed, *J. Biol. Chem.*, 1970, **245**, 183.
[741] L. S. Khailova, *Biochemistry* (*U.S.S.R.*), 1971, **36**, 120.
[742] P. Desnuelle, *Biochimie*, 1971, **53**, 841.
[743] M. F. Maylié, M. Charles, C. Gache, and P. Desnuelle, *Biochim. Biophys. Acta*, 1971, **229**, 286.
[744] M. Sémériva, C. Dufour, and P. Desnuelle, *Biochemistry*, 1971, **10**, 2143.
[745] R. A. Shipolini, G. L. Callewaert, R. C. Cottrell, and C. A. Vernon, *F.E.B.S. Letters*, 1971, **17**, 39.
[746] R. A. Shipolini, G. L. Callewaert, R. C. Cottrell, S. Doonan, C. A. Vernon, and B. E.C. Banks, *European J. Biochem.*, 1971, **20**, 459.
[747] M. A. Wells, *Biochemistry*, 1971, **10**, 4074.
[748] Y. Hachimori, M. A. Wells, and D. J. Hanahan, *Biochemistry*, 1971, **10**, 4084.

content, but the C-terminal sequence (*C. adamanteus*) is -Ser-Gly-Cys-Leu. A phospholipase A inhibitor from *Bothrops neuwiedii* venom [749] appears to require thiol groups for its inhibitory activity. A new membrane-bound phospholipase, phospholipase A_1, has been isolated from *E. coli*.[750] It has a different substrate specificity from phospholipase A_2 and appears to have a subunit molecular weight of *ca.* 29 000. The enzyme is stable in 3% SDS and tends to aggregate in its absence, all implying, perhaps, a hydrophobic exterior that may be related to its location at the membrane.

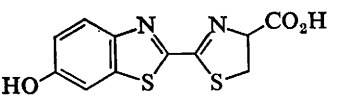
A peptide isolated from the luciferin-binding site of firefly luciferase had the sequence:[751]

Glp-X-Gly-Ala-Val-(Asp)-Ile-Leu-

where X may be tyrosine and attachment is through a thioether linkage. Specific labelling was achieved with the active-site-directed inhibitor 2-cyano-6-chlorobenzothiazole (14), an analogue of firefly luciferin (15). No



(14)



(15)

reaction occurred at the two reactive thiol groups per dimer (m. wt. 100 000) and the importance of precise design of reagent is apparent here because earlier work had shown that these thiol groups are close to the site that binds the carboxy-group of luciferin, *i.e.* removed from the 6-position which reacts in the present instance. A self-associating low molecular weight luciferase with one thiol group per subunit of 12 000 daltons has been isolated [752] from the sea pansy *Renilla reniformis*. The roles of the two non-identical subunits of luciferase (from *Photobacterium fischeri*) in the bioluminescent reaction are clearly different, as shown [753] by the properties of hybrids in which one or other (or both or neither) of the subunits is succinylated. The $\alpha\beta_s$ species had at least 50% of native activity while the $\alpha_s\beta$ enzyme was virtually inactive. The α-subunit appears to participate directly in the catalytic step of the reaction, the function of the β-subunit being unclear.[754] Perhaps some remarks about the 'genetic' approach to the study of quaternary structure are not amiss here. Briefly, in one sort of experiment [755] a bacterium (either wild-type or mutant) carries an episome of another bacterium and produces two types of a given enzyme, one characteristic of each organism. From the number of *in vivo* hybrid

[749] J. C. Vidal and A. O. M. Stoppani, *Arch. Biochem. Biophys.*, 1971, **147**, 66.
[750] C. J. Scandella and A. Kornberg, *Biochemistry*, 1971, **10**, 4447.
[751] R. T. Lee and W. O. McElroy, *Arch. Biochem. Biophys.*, 1971, **146**, 551.
[752] Y. D. Karkhanis and M. J. Cormier, *Biochemistry*, 1971, **10**, 317.
[753] E. A. Meighen, M. Ziegler Nicoli, and J. W. Hastings, *Biochemistry*, 1971, **10**, 4062.
[754] E. A. Meighen, M. Ziegler Nicoli, and J. W. Hastings, *Biochemistry*, 1971, **10**, 4069.
[755] K. K. Lew and J. R. Roth, *Biochemistry*, 1971, **10**, 204.

species shown by electrophoresis the number of subunits in the enzymes can be calculated. The pros and cons of the genetic approach have been considered.[755, 753] Using genetic variants in a rather different, more clearly defined way, hybrids have been constructed from subunits of wild-type and defective β-galactosidase [756] and it could be shown, for example, that a single wild-type subunit in a hybrid tetramer could be active independent of the other subunits. This approach is, of course, limited by the availability of mutants but should nonetheless be a powerful tool in the study of subunit interactions.

The subunit structure of aspartate transcarbamylase from *E. coli* has been thoroughly re-investigated [757] and the structure favoured contains two catalytic trimers intercalated by six regulatory subunits as dimers (Figure 17). There is much evidence for the existence of the catalytic subunit as a trimer, including titration with the transition-state analogue *N*-(phosphonacetyl)-L-aspartate,[758] which combines most of the structural features of the two natural substrates, carbamyl phosphate and L-aspartate, and binds 1000 times more strongly than carbamyl phosphate.

Hybridization of native dimers of creatine kinase with dimers modified with iodoacetamide at the one essential thiol group per subunit gives a stable tetrameric structure, but with only 50% of the native activity.[759] Lombricine kinase, ATP:guanidine phosphotransferase, from *Lumbricus terrestris* (earthworm) muscle, is a dimer (m. wt. *ca.* 80 000) with only one essential thiol group per dimer (perhaps because only one of the subunits is catalytic). Comparison of the sequence around the thiol group [760] with known sequences around the essential thiol (2 per dimer) in creatine kinase and arginine kinase shows that they are clearly homologous:

Lombricine kinase (earthworm)	-Leu-Gly-Tyr-Ile - - -Thr-*Cys-Pro-Gly-Ser-Asn-
Creatine kinase (rabbit muscle)	-Asn-His-Leu-Gly-Tyr-Val-Leu-Thr-Cys-Pro- - -Ser-Asn-
(ox brain)	-Asn-His-Leu-Gly-Tyr-Ile -Leu-Thr-Cys-Pro- - -Ser-Asn-
Arginine kinase (lobster)	-Gln-Thr-Cys-Pro-Thr-Ser-Asn-
Lombricine kinase (earthworm)	Leu-Gly-Thr- - -Leu-Arg-
Creatine kinase (rabbit muscle)	Leu-Gly-Thr-Gly-Leu-Arg-
(ox brain)	Leu-Gly-Thr-Gly-Leu-Arg-
Arginine kinase (lobster)	Leu-Gly-Thr- - -Val-Arg-

The sequence around a single lysine residue in chicken creatine kinase that

[756] F. Melchers and W. Messer, *J. Mol. Biol.*, 1971, **61**, 401.
[757] J. P. Rosenbusch and K. Weber, *J. Biol. Chem.*, 1971, **246**, 1644.
[758] K. O. Collins and G. R. Stark, *J. Biol. Chem.*, 1971, **246**, 6599.
[759] D. Allard and J. C. Dreyfus, *Biochimie*, 1971, **53**, 311.
[760] E. der Terrossian, G. Desvages, L.-A. Pradel, R. Kassab, and N. van Thoai, *European J. Biochem.*, 1971, **22**, 585.

can be labelled with dansyl chloride (after blocking the thiol group with tetrathionate) is:[761]

-Leu-Leu-Val-Pro-Asp-Ser-Lys-Leu-Phe-Ser-Val-Arg-

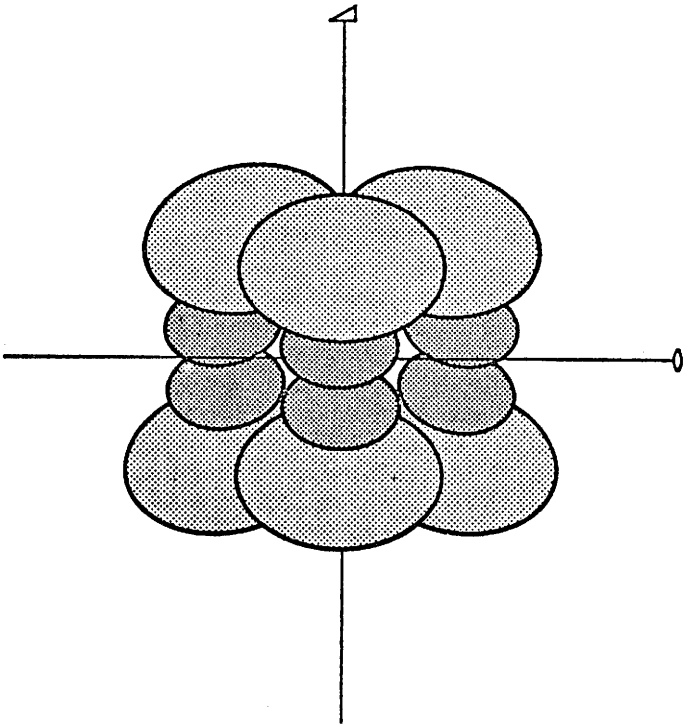


Figure 17 *Proposal for the arrangement of the polypeptide chains in aspartate transcarbamylase. The larger units (light shading) represent the six catalytic chains grouped in two trimers around a 3-fold symmetry axis (vertical line). The two trimers are arranged around three 2-fold symmetry axes (one of which is indicated by a horizontal line). The six regulatory chains (dark shading) are intercalated as dimers between the two catalytic trimers. Placing the subunits in register along the 3-fold axis is used for the purpose of better visualization only*
(Reproduced by permission from *J. Biol. Chem.*, 1971, **246**, 1644)

Numerous other observations on a number of enzymes ought to be mentioned. Phenylalanine ammonia-lyase from yeast has been shown to have dehydroalanine at its active site [762] like histidine ammonia-lyase. The dimeric model (m. wt. 88 000) for yeast enolase has been confirmed [763] and the enzyme shown to be very resistant to exopeptidase digestion, contrary

[761] S. Bose and F. Friedberg, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 271.
[762] D. S. Hodgins, *J. Biol. Chem.*, 1971, **246**, 2977.
[763] P. A. Hargrave and F. Wold, *J. Biol. Chem.*, 1971, **246**, 2904.

to previous reports. Active monomers of enolase have apparently been obtained by frontal elution [764] and the technique is advocated as a possible general method for such purposes.[764] Allosteric inhibition of glutamine synthetase by AMP results from binding of AMP in phosphodiester linkage to a tyrosyl side-chain in each of the chains. The sequence of amino-acid residues around this tyrosine residue, reported last year, has now been extended [765] and a correction of Asp to Asn has been incorporated:

-Ile-His-Pro-Gly-Glu-Ala-Met-Lys-Asp-Asn-Leu-*Tyr-Asp-Leu-Pro-Pro-Glu-Gly-Glu-Ala-Lys-

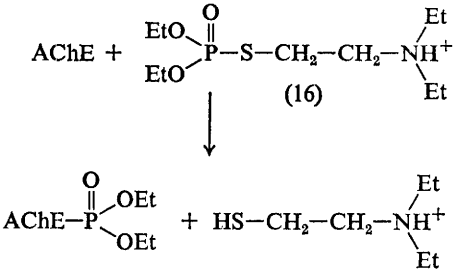
The complete amino-acid sequence (125 residues) has been determined [766] for the Δ^5-3-keto-steroid isomerase of *Pseudomonas testosteroni*, an enzyme with a very high turnover number, a putative trimeric structure, and clusters of hydrophobic residues throughout the sequence, any significance of which is unknown.

When Met tRNA synthetase from *E. coli* is incubated with trypsin at 37 °C the tetramer of 4 × 43 000 molecular weight is converted into an enzymically active dimer of molecular weight 2 × 32 000.[767] Thus proteolysis is obviously the explanation for the dimeric form isolated when an autolytic step is included in the preparation of the enzyme. It is remarkable that loss of 20% of the structure by chain cleavage reduces the activity relative to the tetrameric form by only 20%, and this leads to speculation on the existence of two distinct regions in the subunit, produced by gene fusion. The two regions could be connected by a flexible piece of chain, whose presence intact promotes the formation of tetramers, thereby somehow conferring some selective advantage.[767] In another study, deliberate genetic manipulation gave a fused protein in which histidinol dehydrogenase and imidazolylacetol phosphate:L-glutamate aminotransferase, folded in their usual conformations, were linked, presumably by a length of flexible chain.[768] Mild proteolysis released histidinol dehydrogenase carrying an extra length of polypeptide but the aminotransferase activity was destroyed.[769] In the case of DNA polymerase proteolysis released an active polymerase fragment and an active exonuclease fragment,[770] suggesting again that the two proteins are linked (through a flexible chain) as a result of a gene fusion event.

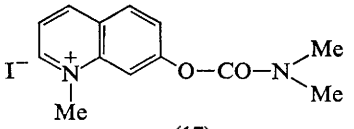
Two new active-site titrants for acetylcholinesterase have been reported: one is an inhibitor and the other a poor substrate. The inhibitor [771] is

[764] S. Keresztes-Nagy and R. Orman, *Biochemistry*, 1971, **10**, 2506.
[765] R. L. Heinrikson and H. S. Kingdon, *J. Biol. Chem.*, 1971, **246**, 1099.
[766] A. M. Benson, R. Jarabak, and P. Talalay, *J. Biol. Chem.*, 1971, **246**, 7514.
[767] D. Cassio and J.-P. Waller, *European J. Biochem.*, 1971, **20**, 283.
[768] M. M. Rechler and C. B. Bruni, *J. Biol. Chem.*, 1971, **246**, 1806.
[769] T. Kohno and J. Yourno, *J. Biol. Chem.*, 1971, **246**, 2203.
[770] H. Klenow, K. Overgaard-Hansen, and S. A. Patkar, *European J. Biochem.*, 1971, **22**, 371.
[771] J. B. Suszkiw, *Analyt. Biochem.*, 1971, **44**, 321.

OO'-diethyl *S*-(2-diethylaminoethyl)phosphorothiolate (16). Liberation of 2-diethylaminoethyl thiolate (Scheme 5) is measured spectrophotometrically after reaction with Ellman's reagent [5,5'-dithiobis-(2-nitrobenzoate)]. The poor substrate is the *N*-methyl-7-hydroxyquinolinium iodide ester of dimethylcarbamic acid (17);[772] decarbamoylation is slow and the zwitter-



Scheme 5



(17)

ionic form of the leaving group is fluorescent. There seems to be some disagreement on the subunit structure of acetylcholinesterase from *Electrophorus electricus* (m. wt. 260 000). In one case [773] an $\alpha_2\beta_2$ tetrameric structure with two active sites is suggested, and in the other [774] a hexamer of subunit molecular weight 42 000.

H. Quaternary Structure.—Some of the newer quaternary structures have already been mentioned in the Report; many others are listed in Table 2.

6 Electron-transport and Oxygen-transport Proteins

A. Electron-transport Proteins.—*Cytochromes.* *X*-Ray crystallography has given the structure of horse-heart ferricytochrome *c* to 2.8 Å resolution;[775] the phases were also used to calculate the structure of the isomorphous ferricytochrome *c* from bonito. A 4 Å resolution study of the reduced form reported from another laboratory [776] indicates some difference in side-chain conformation at the surface in the ferri- and ferro-forms. The primary

[772] T. L. Rosenberry and S. A. Bernhard, *Biochemistry*, 1971, **10**, 4114.
[773] W. Leuzinger, *Biochem. J.*, 1971, **123**, 139.
[774] D. B. Millar and M. A. Grafius, *F.E.B.S. Letters*, 1970, **12**, 61.
[775] R. E. Dickerson, T. Takano, D. Eisenberg, O. B. Kallai, L. Samson, A. Cooper, and E. Margoliash, *J. Biol. Chem.*, 1971, **246**, 1511.
[776] T. Ashida, T. Ueki, T. Tsukihara, A. Sugihara, T. Takano, and M. Kakudo, *J. Biochem. (Japan)*, 1971, **70**, 913.

Table 2 *Quaternary structure of some individual enzymes*

Enzyme	*Source*	*Molecular weight*	*No. of subunits*
Acetolactate-forming enzyme (pH 6)[a]	*Aerobacter aerogenes*	220 000	4
Adenosine triphosphatase[b]	Pig brain	280 000	12
Aldolase[c, d, e, f]	Codfish muscle	160 000	4
	Frog muscle	164 000	4
	Lobster muscle	160 000	4
	Shark muscle	160 000	4
δ-Aminolaevulinate dehydratase[g, h]	Mouse liver	250 000	6
	Rhodopseudomonas spheroides	240 000	6
δ-Aminolaevulinate synthase[i]	*Rhodopseudomonas spheroides*	57 000	1
α-Amylase[j]	Pig pancreas	50 000	2
Anthranilate synthetase[k]	*Serratia marcescens*	141 000	4 ($\alpha_2\beta_2$)
Anthranilate synthetase–anthranilate-5-phosphoribosylpyrophosphate phosphoribosyltransferase[l]	*Salmonella typhimurium*	280 000	4 ($\alpha_2\beta_2$)
Apoferritin[m, n]	Horse spleen	450 000	24
Arylamidase[o]	Human liver	235 000	6
L-Asparaginase[p]	*E. coli*	142 000	4
Aspartase[q]	*E. coli* B	170 000	4
Aspartate transcarbamylase[r]	*E. coli*	300 000	12 ($\alpha_6\beta_6$)
Azoferredoxin[s]	*Clostridium pasteurianum* W5	55 000	2
Butyryl-CoA dehydrogenase[t]	*Peptostreptococcus elsdenii*	150 000	4
Catabolite gene activator protein (*lac*)[u]	*E. coli*	44 000	2
Chorismate mutase–prephenate dehydrogenase[v]	*E. coli* K12	80 000	2
Citrate oxaloacetate lyase[w]	*Klebsiella aerogenes*	560 000	16
Citrate synthase[x]	Rat heart, liver	100 000	2
Concanavalin A[y]	Jack bean	70 000	4
	Jack bean	54 000	2
Cytochrome b_2 (L-lactate) cytochrome *c* oxidoreductase[z]	Baker's yeast	220 000	8 ($\alpha_4\beta_4$)

[*References for Table on pages* 132 *and* 133]

Table 2 (*cont.*)

Enzyme	*Source*	*Molecular weight*	*No. of subunits*
Diacetyl (acetoin) reductase [aa]	*Aerobacter aerogenes*	100 000	4
Dihydrofolate reductase [bb]	*Diplococcus pneumoniae*	20 000	1
DNA-dependent RNA polymerase [cc]	*Anacystis nidulans*	436 000	5 ($\alpha_2\beta\beta'\sigma$)
DNA polymerase [dd]	Phage T7	100 000	1
Enolase [ee]	*E. coli*	90 000	2
F_1-initiation factor [ff]	*E. coli* ribosomes	9 400	1
Follicle-stimulating hormone [gg]	Human	35 000	2
Formyl tetrahydrofolate [hh] synthetase	*Clostridium cylindrosporum*	240 000	4
Galactokinase [ii]	Human red blood cells	55 000	2
β-Glucuronidase [jj]	Rat liver lysozomes	280 000	4
Glutathioneperoxidase I [kk]	Ox blood	84 000	4
Glyceraldehyde 3-phosphate dehydrogenase [ll]	*E. coli*	144 000	4
Glycerate dehydrogenase [mm]	Ox liver	68 000	2
Glycerol kinase [nn]	*E. coli*	215 000	4
Glycerol 3-phosphate dehydrogenase [oo]	Rat liver	63 000	2
Haemoglobin [pp]	*Arenicola cristata*	2.85×10^6	192 ($\alpha_{96}\beta_{96}$)
Hydrogenase [qq]	*Clostridium pasteurianum* W5	60 000	2
Hypoxanthine–guanine phosphoribosyltransferase [rr]	Human erythrocytes	68 000	2
Invertase [ss]	*Neurospora crassa*	210 000	4
2-Keto-3-deoxy-6-phosphogluconate aldolase [tt]	*Pseudomonas putida*	72 000	3
Lactoferrin [uu]	Human	76 000	1
Leucine aminopeptidase [vv]	Ox lens	327 000	6
Malonyl-CoA–acyl-carrier-protein transacylase [ww]	*E. coli*	35 000	1
N-Methylglutamate synthetase [xx]	*Pseudomonas MA*	350 000	12
Molybdoferredoxin [yy]	*Clostridium pasteurianum* W5	170 000	3 ($\alpha_2\beta$)
Nerve growth factor [zz]	Mouse submaxillary gland	29 000	2
Nucleoside diphosphate kinase [aaa]	Pea seed	70 000	4
17β-Oestradiol dehydrogenase [bbb]	Human placenta	68 000	2
Paramyosin [ccc]	*Venus mercenaria*	200 000	2
Phosphoenolpyruvate carboxylase [ddd]	*E. coli*	402 000	4
Phosphoglucose isomerase [eee]	Human erythrocytes	125 000	2

Table 2 (*cont.*)

Enzyme	*Source*	*Molecular weight*	*No. of subunits*
3-Phosphoglycerate kinase [fff]	Yeast	50 000	1
Phosphofructokinase [ggg, hhh]	Yeast	600 000	6
	Human erythrocytes	$(104\,000)_n$	n = 2, 4, 6, 8, or 10
Phospholipase A [iii]	*Apis mellifica*	40 000	2
Prealbumin [jjj]	Human?	56 000	4
Pyruvate decarboxylase [kkk]	*E. coli* K12	200 000	2
Pyruvate kinase [lll, mmm]	Yeast (*Saccharomyces cerevisiae*)	160 000	8
	Saccharomyces carlsbergensis	191 000	4
	Neurospora crassa mitochondria	64 000	1
RNA polymerase [nnn]	Ox brain	21 000	3 ($\alpha_2\beta$)
S-100 protein [ooo]	Ox heart	98 000	2 ($\alpha\beta$)
Succinate dehydrogenase [ppp]	Ox heart	35 000	2
Superoxide dismutase [qqq]	*Rhodospirillum rubrum*	180 000	4
Threonine deaminase [rrr]	*E. coli* K12	68 000	1
Thymidine phosphorylase [sss]	Ehrlich ascites carcinoma	70 000	1
Thymidylate synthetase [ttt, uuu]	*Lactobacillus casei*	70 000	2
	Human	64 000	4
Thyroxine-binding prealbumin [vvv]	Yeast	140 000	2
Transketolase [www]	Houseflies	120 000	4
1,1,1- Trichloro-2,2-bis-(*p*-chlorophenyl)ethane dehydrochlorinase [xxx] [DDT-dehydrochlorinase]	(*Musca domestica* L.)		
tRNA synthetase:	*E. coli*	69 000	1
Glutaminyl [yyy]	Yeast	120 000	2
Seryl [zzz]	*E. coli*	74 000	2
Tryptophanyl [aaaa]	*Pseudomonas putida*	86 000	2
Tryptophan synthetase [bbbb] B component	Human plasma	53 000	1
Vitamin-D-binding protein [cccc]			

[*References for Table on pages* 132 *and* 133]

References for Table 2

[a] N. E. Huseby, T. B. Christensen, B. R. Olsen, and F. C. Størmer, *European J. Biochem.*, 1971, **20**, 209; [b] A. Atkinson, A. D. Gatenby, and A. G. Lowe, *Nature New Biol.*, 1971, **233**, 145; [c] C. Y. Lai and C. Chen, *Arch. Biochem. Biophys.*, 1971, **144**, 467; [d] S.-M. Ting, C. L. Sia, C. Y. Lai, and B. L. Horecker, *Arch. Biochem. Biophys.*, 1971, **144**, 485; [e] A. Guha, C. Y. Lai, and B. L. Horecker, *Arch. Biochem. Biophys.*, 1971, **147**, 692; [f] C. E. Caban and L. F. Hass, *J. Biol. Chem.*, 1971, **246**, 6807; [g] D. Doyle, *J. Biol. Chem.*, 1971, **246**, 4965; [h] S. van Heyningen and D. Shemin, *Biochemistry*, 1971, **10**, 4676; [i] G. R. Warnick and B. F. Burnham, *J. Biol. Chem.*, 1971, **246**, 6880; [j] J. F. Robyt, C. G. Chittenden, and C. T. Lee, *Arch. Biochem. Biophys.*, 1971, **144**, 160; [k] H. Zalkin and L. H. Hwang, *J. Biol. Chem.*, 1971, **246**, 6899; [l] E. J. Henderson and H. Zalkin, *J. Biol. Chem.*, 1971, **246**, 6891; [m] I. Bjork and W. W. Fish, *Biochemistry*, 1971, **10**, 2844; [n] C. F. A. Bryce and R. R. Crichton, *J. Biol. Chem.*, 1971, **246**, 4198; [o] G. H. Little and F. J. Behal, *Biochim. Biophys. Acta*, 1971, **243**, 312; [p] H. Glossmann and W. Bode, *Z. physiol. Chem.*, 1971, **352**, 132; [q] F. B. Rudolph and H. J. Fromm, *Arch. Biochem. Biophys.*, 1971, **147**, 92; [r] J. P. Rosenbusch and K. Weber, *J. Biol. Chem.*, 1971, **246**, 1644; [s] G. Nakos and L. Mortenson, *Biochemistry*, 1971, **10**, 455; [t] P. C. Engel and V. Massey, *Biochem. J.*, 1971, **125**, 879; [u] A. D. Riggs, G. Reiness, and G. Zubay, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 1222; [v] G. L. E. Koch, D. C. Shaw, and F. Gibson, *Biochim. Biophys. Acta*, 1971, **229**, 805; [w] T. J. Bowen and M. G. Mortimer, *European J. Biochem.*, 1971, **23**, 262; [x] T. Moriyama and P. A. Srere, *J. Biol. Chem.*, 1971, **246**, 3217; [y] W. D. McCubbin and C. M. Kay, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 101; J. C. Wang, B. A. Cunningham, and G. M. Edelman, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 1130; [z] F. Lederer and A. M. Simon, *European J. Biochem.*, 1971, **20**, 469; F. Lederer and C. Jacq, *ibid.*, p. 475; [aa] Ø. Hetland, B. R. Olsen, T. B. Christensen, and F. C. Størmer, *European J. Biochem.*, 1971, **20**, 200; [bb] F. M. Sirotnak and J. S. Salser, *Arch. Biochem. Biophys.*, 1971, **145**, 286; [cc] F. Herzfeld and W. Zillig, *European J. Biochem.*, 1971, **24**, 242; [dd] J. L. Oey, W. Strätling, and R. Knippers, *European J. Biochem.*, 1971, **23**, 497; [ee] T. G. Spring and F. Wold, *J. Biol. Chem.*, 1971, **246**, 6507; [ff] P. D. Stahl and O. Touster, *J. Biol. Chem.*, 1971, **246**, 5398; [gg] L. Flohé, B. Eisele, and A. Wendel, *Z. physiol. Chem.*, 1971, **352**, 151; [hh] G. d'Alessio and J. Josse, *J. Biol. Chem.*, 1971, **246**, 4326; [ii] I. Y. Rosenblum, D. H. Autkowiak, H. J. Sallach, L. E. Flanders, and L. A. Fahien, *Arch. Biochem. Biophys.*, 1971, **144**, 375; [jj] J. W. Thorner and H. Paulus, *J. Biol. Chem.*, 1971, **246**, 3885; [kk] T. P. Fondy, K. J. Herwig, S. J. Sollohur, and D. B. Rutherford, *Arch. Biochem. Biophys.*, 1971, **145**, 583; [ll] L. Waxman, *J. Biol. Chem.*, 1971, **246**, 7318; [mm] G. Nakos and L. E. Mortenson, *Biochemistry*, 1971, **10**, 2442; [nn] W. J. Arnold and W. N. Kelley, *J. Biol. Chem.*, 1971, **246**, 7398; [oo] Z. D. Meadhum, jun., H. J. Colvin, jun., and H. D. Braymer, *Biochemistry*, 1971, **10**, 326; [pp] R. H. Hammerstedt, H. Möhler, K. A. Dekker, and W. A. Wood, *J. Biol. Chem.*, 1971, **246**, 2069; [qq] P. Querinjean, P. L. Masson, and J. F. Heremans, *European J. Biochem.*, 1971, **20**, 420; [rr] S. W. Melbye and F. H. Carpenter, *J. Biol. Chem.*, 1971, **246**, 2459; [ss] V. C. Joshi and S. J. Wakil, *Arch. Biochem. Biophys.*, 1971, **143**, 493; [tt] R. J. Pollock and L. B. Hersh, *J. Biol. Chem.*, 1971, **246**, 4737; [uu] G. Nakos and L. Mortenson, *Biochim. Biophys. Acta*, 1971, **229**, 431; [vv] R. H. Angeletti, R. A. Bradshaw, and R. D. Wade, *Biochemistry*, 1971, **10**, 463; [ww] B. Edlund, *F.E.B.S. Letters*, 1971, **13**, 56; [xx] D. J. W. Burns, L. L. Engel, and J. L. Bethune, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 786; [yy] J. Olander, *Biochemistry*, 1971, **10**, 601; [zz] T. E. Smith, *J. Biol. Chem.*, 1971, **246**, 4234; [aaa] K. K. Tsuboi, K. Fukunaga, and C. H. Chervenka, *J. Biol. Chem.*, 1971, **246**, 7586; [bbb] R. K. Scopes, *Biochem. J.*, 1971, **122**, 89; [ccc] H. Wilgus, J. R. Pringle, and E. Stellwagen, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 89; [ddd] K. W. Wenzel, G. Zimmerman, J. Gauer, W. Diezel, S. Liebe, and E. Hofmann, *F.E.B.S. Letters*, 1971, **19**, 285; [eee] R. A. Shipolini, G. L. Callewaert, R. C. Cottrell, S. Doonan, C. A. Vernon, and B. E. C. Banks, *European J. Biochem.*, 1971, **20**, 459; [fff] G. Gonzalez and R. E. Offord, *Biochem. J.*, 1971, **125**, 309; [ggg] O. Vogel and U. Henning, *European J. Biochem.*, 1971, **18**, 103; [hhh] K. Ashton and A. R. Peacocke, *F.E.B.S. Letters*, 1971, **16**, 25; [iii] H. Bischofberger, B. Hess, and P. Röschlau, *Z. physiol. Chem.*, 1971, **352**, 1139; [jjj] H. Küntzel and K. P. Schäfer, *Nature New Biol.*, 1971, **231**, 265; [kkk] P. S. Dannies and L. Levine, *J. Biol. Chem.*, 1971, **246**, 6276; [lll] K. A. Davis and Y. Hatefi, *Biochemistry*, 1971, **10**, 2509; P. Righetti and P. Cerletti, *F.E.B.S. Letters*, 1971, **13**, 181;

[qqq] B. B. Keele, jun., J. M. McCord, and I. Fridovich, *J. Biol. Chem.*, 1971, **246**, 2875; [rrr] R. S. Feldberg and P. Datta, *European J. Biochem.*, 1971, **21**, 438; [sss] M. Schwartz, *European J. Biochem.*, 1971, **21**, 191; [ttt] A. Fridland, R. J. Langenbach, and C. Heidelberger, *J. Biol. Chem.*, 1971, **246**, 7110; [uuu] R. B. Dunlap, N. G. L. Harding, and F. M. Huennekens, *Biochemistry*, 1971, **10**, 88; [vvv] L. Rask, P. A. Peterson, and S. F. Nilsson, *J. Biol. Chem.*, 1971, **246**, 6087; [www] C. P. Heinrich and O. Wiss, *F.E.B.S. Letters*, 1971, **14**, 251; [xxx] M. L. Dinamarea, L. Levenbook, and E. Valdés, *Arch. Biochem. Biophys.*, 1971, **147**, 374; [yyy] W. R. Folk, *Biochemistry*, 1971, **10**, 1728; [zzz] H. Heider, E. Gottschalk, and F. Cramer, *European J. Biochem.*, 1971, **20**, 144; [aaaa] D. R. Joseph and K. H. Muench, *J. Biol. Chem.*, 1971, **246**, 7610; [bbbb] R. Mauerer and I. P. Crawford, *Arch. Biochem. Biophys.*, 1971, **144**, 193; [cccc] P. A. Peterson, *J. Biol. Chem.*, 1971, **246**, 7748.

---

structure of bonito cytochrome *c* has also been determined;[777] it differs from the tuna cytochrome only in the change Asp → Glu at position 61. A three-dimensional structure has been proposed [778] for cytochrome c_{551} from *Pseudomonas aeruginosa* based on the structures of horse and bonito cytochrome *c* and a comparison of amino-acid sequences. Most of the differences could be accounted for by deletion of a single 16-residue loop in the cytochrome from horse heart, and the structure showed how the sequences should be aligned to maximize homology; this was also predicted using a new method for comparing amino-acid sequences.[4] There was no evidence from either the *X*-ray study [775] or the theoretical treatment [4] for repeating polypeptide sequences in cytochrome *c*. Bacterial cytochrome c_{551} and horse-heart cytochrome *c* turn out to be clearly homologous at the level of both primary and tertiary structure. In other cases where homology in primary structures needs some imagination this might turn out not to be so, and a warning of the dangers of detecting otherwise invisible homologies from minimization of minimum mutational distances [778] would be well heeded. Helix probability profiles calculated for 27 cytochrome *c* sequences,[779] again based on the *X*-ray structure of the horse-heart protein, were similar in all but three cases. In these cases (screw-worm fly, fruit fly, and yeast) it appears that either there are errors in the amino-acid sequences, or there are true differences in three-dimensional structure.[779] The former is, perhaps, more likely.

The two fragments obtained by specific cyanogen bromide cleavage at Met-65 in horse-heart cytochrome *c* will reassociate with regeneration of the spectral properties of the parent protein.[780] A method for removing haem [781] is based on selective oxidation of the thioether bridge to the sulphoxide with limited amounts of iodine, followed by cyanogen bromide treatment to remove the haem; the cleavage at Met-65 apparently occurs more slowly and can be avoided. Reversible formylation of Trp-59 in horse-heart cytochrome *c* has also been reported.[782]

[777] T. Nakayama, K. Titani, and K. Narita, *J. Biochem.* (*Japan*), 1971, **70**, 311.
[778] R. E. Dickerson, *J. Mol. Biol.*, 1971, **57**, 1.
[779] P. N. Lewis and H. A. Scheraga, *Arch. Biochem. Biophys.*, 1971, **144**, 576.
[780] G. Corradin and H. A. Harbury, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 3036.
[781] F. Lederer and J. Tarin, *European J. Biochem.*, 1971, **20**, 482.
[782] I. Aviram and A. Schejter, *Biochim. Biophys. Acta*, 1971, **229**, 113.

Many new cytochrome *c* sequences have become available. Peptide mapping and amino-acid composition suggested [783] that the cytochrome *c* from zebra (*Equus guagga Boehmi*) differed from horse cytochrome *c* only in the substitution Thr-47 → Ser. The complete primary structures of cytochromes *c* from a wide range of plants have been reported from one

```
                                                1
S:  Acetyl-Ala-Ser-Phe-Asx-Glx-Ala-Pro-Pro-Gly-Asx-Val-Lys-Ser-Gly-Glu-
C:  Acetyl-Ala-Ser-Phe-Asx-Glx-Ala-Pro-Pro-Gly-Asx-Val-Lys-Ala-Gly-Glu-
M:  Acetyl-Ala-Ser-Phe-Asx-Glx-Ala-Pro-Pro-Gly-Asx-Ser-Lys-Ser-Gly-Glu-
              10                                20
S:  Lys-Ile-Phe-Lys-Thr-Lys-Cys-Ala-Gln-Cys-His-Thr-Val-Asp-Lys-Gly-Ala-
C:  Lys-Ile-Phe-Lys-Thr-Lys-Cys-Ala-Gln-Cys-His-Thr-Val-Glu-Lys-Gly-Ala-
M:  Lys-Ile-Phe-Lys-Thr-Lys-Cys-Ala-Gln-Cys-His-Thr-Val-Asp-Lys-Gly-Ala-
                    30                                40
S:  Gly-His-Lys-Gln-Gly-Pro-Asn-Leu-Asn-Gly-Leu-Phe-Gly-Arg-Gln-Ser-Gly-
C:  Gly-His-Lys-Gln-Gly-Pro-Asn-Leu-Asn-Gly-Leu-Phe-Gly-Arg-Gln-Ser-Gly-
M:  Gly-His-Lys-Gln-Gly-Pro-Asn-Leu-Asn-Gly-Leu-Phe-Gly-Arg-Gln-Ser-Gly-
                          50
S:  Thr-Thr-Pro-Gly-Tyr-Ser-Tyr-Ser-Ala-Ala-Asn-Lys-Asn-Met-Ala-Val-Ile -
C:  Thr-Thr-Ala-Gly-Tyr-Ser-Tyr-Ser-Ala-Ala-Asn-Lys-Asn-Met-Ala-Val-Gln-
M:  Thr-Thr-Ala-Gly-Tyr-Ser-Tyr-Ser-Thr-Ala-Asn-Lys-Asn-Met-Ala-Val-Ile -
          60                                70
S:  Trp-Gly-Glu-Asn-Thr-Leu-Tyr-Asp-Tyr-Leu-Leu-Asn-Pro-TML-Lys-Tyr-Ile-
C:  Trp-Gly-Glu-Ans-Thr-Leu-Tyr-Asp-Tyr-Leu-Leu-Asn-Pro-TML-Lys-Tyr-Ile-
M:  Trp-Glu-Glu-Lys-Thr-Leu-Tyr-Asp-Tyr-Leu-Leu-Asn-Pro-TML-Lys-Tyr-Ile-
                    80                                90
S:  Pro-Gly-Thr-Lys-Met-Val-Phe-Pro-Gly-Leu-TML-Lys-Pro-Gln-Glu-Arg-
C:  Pro-Gly-Thr-Lys-Met-Val-Phe-Pro-Gly-Leu-TML-Lys-Pro-Gln-Asp-Arg-
M:  Pro-Gly-Thr-Lys-Met-Val-Phe-Pro-Gly-Leu-TML-Lys-Pro-Gln-Asp-Arg-
                          100       103
S:  Ala-Asp-Leu-Ile-Ala-Tyr-Leu-Lys-Glu-Ala-Thr-Ala
C:  Ala-Asp-Leu-Ile-Ala-Tyr-Leu-Lys-Glu/Gln-Ala-Thr-Ala
M:  Ala-Asp-Leu-Ile-Ala-Tyr-Leu-Lys-Glu-Ser-Thr-Ala
```

Figure 18 *Complete sequence of sesame (S), castor (C), and mung-bean (M) cytochrome* c. *Numbering is based on that of horse-heart cytochrome* c. *TML is trimethyl-lysine*

laboratory, culminating in a proposed family tree [784] which for plants is of particular interest in the absence of fossil records. All the plant cytochrome *c* sequences have 111 amino-acid residues, with an eight-residue acetylated 'tail' at the N-terminus (compared with horse-heart cytochrome *c*). Other distinctive features are two residues of trimethyl-lysine at positions 72 and 86, a Thr-Thr sequence at positions 42-43, and glutamine at position 47 (numbering is that of horse-heart cytochrome *c*). Figure 18 shows the sesame and castor cytochrome *c* sequences [785] compared with the mung-

[783] L. Gürtler and H. J. Horstmann, *F.E.B.S. Letters*, 1971, **18**, 106.
[784] J. A. M. Ramshaw, M. Richardson, and D. Boulter, *European J. Biochem.*, 1971, **23**, 475.
[785] E. W. Thompson, M. Richardson, and D. Boulter, *Biochem. J.*, 1971, **121**, 439.

bean sequence determined earlier and now amended slightly in the region of residues 5 to 9. Cauliflower [786] and rape [787] (both members of the Brassicaceae) cytochromes *c* are identical, and very similar to that of pumpkin;[788] both pumpkin and rape show some evidence of polymorphism. The proteins from *Abutilon* and *Gossypium* (cotton), from the Malvaceae, are also very similar.[789] In addition to these angiosperms, the complete sequence of cytochrome *c* from a gymnosperm (*Ginkgo biloba* L.) has been determined;[784] it has the usual common features of plant cytochromes *c* and appears to be two residues longer at the C-terminus. The sequence of cytochrome *c* from the yeast *Debaryomyces kloeckeri* (109 residues) [790] showed that it was very similar (*ca.* 80% homology) to cytochromes from *Candida* and *Saccharomyces*. Again there is a 'tail' (5 or 6 residues) at the N-terminus and they all have one residue of trimethyllysine at position 72. A *c*-type cytochrome from *Micrococcus denitrificans* [791] has only one histidine, and with about 135 residues is larger than mammalian cytochrome *c*. A protozoan cytochrome [792] has only one cysteine, in the sequence -Ala-Ala-Gln-Cys-His-(Thr,Gly,Ala)-Lys-, and two histidines.

Other cytochromes have several cysteine–histidine clusters and several haem groups per chain. For instance there are four such clusters per chain of 103 residues in cytochrome c_3 from *Desulfovibrio desulfuricans*;[793] three of these have the standard separation of the two cysteines (-Cys-X-Y-Cys-His-) while the separation is greater in the fourth. Assuming some deletions, this cytochrome is clearly homologous with other *Desulfovibrio* cytochromes. It is also homologous (more so than the other *Desulfovibrio* cytochromes) with cytochrome $c_{551.5}$ (cytochrome c_7) from the green photosynthetic bacterium *Chloropseudomonas ethylica*,[794] thus establishing a connection between these anaerobic bacteria and sulphate-reducing bacteria. The c_7 cytochrome has 68 amino-acid residues and three cysteine–histidine clusters, all with the normal sequence (above) and each presumably carrying one haem. The N-terminal sequence of the haem peptide of cytochrome c_{552} from chloroplasts of *Euglena gracilis* [795] is Ala-Asp-Asp-, and the peptide clearly contains histidine.

Cytochrome b_2 from baker's yeast is found to be an octamer of the type $(\alpha\beta)_4$ where the α-chains have a molecular weight of 36 000 daltons and the

[786] E. W. Thompson, M. Richardson, and D. Boulter, *Biochem. J.*, 1971, **124**, 783.
[787] M. Richardson, J. A. M. Ramshaw, and D. Boulter, *Biochim. Biophys. Acta*, 1971, **251**, 331.
[788] E. W. Thompson, M. Richardson, and D. Boulter, *Biochem. J.*, 1971, **124**, 779.
[789] E. W. Thompson, B. A. Notton, M. Richardson, and D. Boulter, *Biochem. J.*, 1971, **124**, 787.
[790] K. Sugeno, K. Narita, and K. Titani, *J. Biochem. (Japan)*, 1971, **70**, 659.
[791] P. B. Scholes, G. McLain, and L. Smith, *Biochemistry*, 1971, **10**, 2072.
[792] G. Pettigrew, *Biochem. J.*, 1971, **125**, 46P.
[793] R. P. Ambler, M. Bruschi, and J. Le Gall, *F.E.B.S. Letters*, 1971, **18**, 347.
[794] R. P. Ambler, *F.E.B.S. Letters*, 1971, **18**, 351.
[795] M. A. Cusanovich, T. Meyer, S. M. Tedro, and M. D. Kamen, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 629.

β-chains of 21 000 daltons.[796] The 'cytochrome b_2 core' obtained by tryptic digestion derives from the heavy chain and contains the haem-binding site. Spectral evidence suggests that the environment of the haem in the native protein is preserved in the core, and apo-core was thus used for photo-oxidation studies [797] to investigate residues involved in ligand binding. This was lost on destruction of two histidine residues; cysteine was thought not to be involved in ligand bonding, nor was methionine, since this was absent from cytochrome b_2 from another strain. Cytochrome b_5 is found in the endoplasmic reticulum and is strongly bound to the membrane. The molecular weight of cytochrome b_5 isolated from rabbit liver,[798] avoiding proteolysis and lipolysis steps and using only detergent, was higher than usual (16 700 rather than 11 000) and the additional hydrophobic sequence of 40 amino-acid residues may be the clue to its strong association with the membrane. Complete sequences for liver microsomal cytochromes b_5 from man, monkey, pig, and chicken have been proposed [799] from an analysis of the tryptic peptides and homology with the known sequence of rabbit liver cytochrome b_5 (97 residues). Residues 42—72 were totally invariant in all five sequences, and the two histidine residues (positions 43 and 67) make this region a strong candidate for haem binding. The molecules contain no cysteine and, in fact, haem structure and ligand binding in microsomal cytochromes b_5 resemble those of haemoglobin although the spectral and enzymic properties are those of cytochromes. Cytochrome *f*, a *c*-type (c_{555}) from chloroplasts, is also firmly membrane-bound. The protein from spinach grana membranes [800] appears to have one haem-containing polypeptide chain and one non-haem chain, each reported to have molecular weight *ca.* 31 000, somewhat different from existing values for cytochrome *f* from other plant chloroplasts. A re-investigation [801] of the subunit structure of cytochrome oxidase [cytochrome $(a + a_3)$] by SDS-gel electrophoresis suggests the quaternary structure $\alpha\beta\gamma_2\delta_4$ (α = 37 000, β = 19 000, γ = 14 000, δ = 10 000).

Other Electron-transport Proteins. The structure of ferredoxins (non-haem iron proteins) and their function in photosynthesis, nitrogen fixation, and fermentative metabolism have been reviewed.[802] The primary structures of two more bacterial ferredoxins have been established.[803, 804] That of ferredoxin from the thermophile *Clostridium tartarivorum* [803] is compared in Figure 19 with other known bacterial ferredoxin sequences (from

[796] F. Lederer and A.-M. Simon, *European J. Biochem.*, 1971, **20**, 469.
[797] O. Groudinsky, *European J. Biochem.*, 1971, **18**, 480.
[798] L. Spatz and P. Strittmatter, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1042.
[799] F. G. Nóbrega and J. Ozols, *J. Biol. Chem.*, 1971, **246**, 1706.
[800] J. Singh and A. R. Wasserman, *J. Biol. Chem.*, 1971, **246**, 3532.
[801] J. J. Keirns, C. S. Yang, and M. V. Gilmour, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 835.
[802] B. B. Buchanan and D. I. Arnon, *Adv. Enzymol.*, 1970, **33**, 119.
[803] M. Tanaka, M. Haniu, G. Matsueda, K. T. Yasunobu, R. H. Himes, J. M. Akagi, E. M. Barnes, and T. Devanathan, *J. Biol. Chem.*, 1971, **246**, 3953.
[804] J. Travis, D. J. Newman, J. Le Gall, and H. D. Peck, jun., *Biochem. Biophys. Res. Comm.*, 1971, **45**, 452.

Micrococcus aerogenes, C. butyricum, C. pasteurianum, and *C. acidiurici*). Homology is high, particularly around the six iron-binding cysteine residues, and there is no clear indication of the reason for thermostability. The well-established gene-duplication in bacterial ferredoxins is clearly seen in the homology between 1—28 and 29—56. The second new bacterial ferredoxin

M. a. Ala-Tyr-Val-Ile-Asn-Asp-Ser-Cys-Ile-Ala-Cys-Gly-Ala-Cys-Lys-Pro-
C. b. Ala-Phe-Val-Ile-Asn-Asp-Ser-Cys-Val-Ser-Cys-Gly-Ala-Cys-Ala-Gly-
C. p. Ala-Tyr-Lys-Ile-Ala-Asp-Ser-Cys-Val-Ser-Cys-Gly-Ala-Cys-Ala-Ser-
C. a. Ala-Tyr-Val-Ile-Asn-Glu-Ala-Cys-Ile-Ser-Cys-Gly-Ala-Cys-Asp-Pro-
C. t. Ala-His-Ile-Ile-Thr-Asp-Glu-Cys-Ile-Ser-Cys-Gly-Ala-Cys-Ala-Ala-
1 10
M. a. Glu-Cys-Pro-Val-Asn-Ile-Gln-Gln-Gly-Ser-Ile-Tyr-Ala-Ile-
C. b. Glu-Cys-Pro-Val-Ser-Ala-Ile-Thr-Gln-Gly-Asp-Thr-Gln-Phe-Val-Ile-
C. p. Glu-Cys-Pro-Val-Asn-Ala-Ile-Ser-Gln-Gly-Asp-Ser-Ile-Phe-Val-Ile-
C. a. Glu-Cys-Pro-Val-Asp-Ala-Ile-Ser-Gln-Gly-Asp-Ser-Arg-Tyr-Val-Ile-
C. t. Glu-Cys-Pro-Val-Glu-Ala-Ile-His-Glu-Gly-Thr-Gly-Lys-Tyr-Gln-Val-
20 30
M. a. Asp-Ala-Asp-Ser-Cys-Ile-Asp-Cys-Gly-Ser-Cys-Ala-Ser-Val-Cys-Pro-
C. b. Asp-Ala-Asp-Thr-Cys-Ile-Asp-Cys-Gly-Asn-Cys-Ala-Asn-Val-Cys-Pro-
C. p. Asp-Ala-Asp-Thr-Cys-Ile-Asp-Cys-Gly-Asn-Cys-Ala-Asn-Val-Cys-Pro-
C. a. Asp-Ala-Asp-Thr-Cys-Ile-Asp-Cys-Gly-Ala-Cys-Ala-Gly-Val-Cys-Pro-
C. t. Asp-Ala-Asp-Thr-Cys-Ile-Asp-Cys-Gly-Ala-Cys-Gln-Ala-Val-Cys-Pro-
40
M. a. Val-Gly-Ala-Pro-Asn-Pro-Glu-Asp
C. b. Val-Gly-Ala-Pro-Asn-Gln-Glu
C. p. Val-Gly-Ala-Pro-Val-Gln-Glu
C. a. Val-Asp-Ala-Pro-Val-Gln-Ala
C. t. Thr-Gly-Ala-Val-Lys-Ala-Glu
50

Figure 19 *Primary structures of bacterial ferredoxins*

sequence is that from the sulphate-reducing bacterium *Desulfovibrio gigas*,[804] which is unique among bacterial ferredoxins in having only four cysteines. Two are in the first half of the molecule, which is highly homologous with the first half (residues 1—28) of the ferredoxins from other bacteria (see above), and two in the second half, which, in this case, is less homologous either with its own first half or with other ferredoxins. Ferredoxin from *C. pasteurianum* can exist in a dimeric form in which the number of atoms of iron and labile sulphide is reduced by half,[805] and it is suggested that four cysteinyl residues per subunit participate in disulphide bridges in the dimer. This observation will probably call for re-examination of many physicochemical studies on this protein.[805] The role of ferredoxin in reductive nitrogen fixation in *Clostridia* is to accept electrons from hydrogenase and transfer them to nitrogenase; hydrogenase is a non-haem iron protein with four atoms each of iron and labile sulphide, whose distribution within the dimer of molecular weight 60 000 is as yet unknown.[806] The gene

[805] K. Gersonde, E. Trittelvitz, H.-E. Schlaak, and H.-H. Stabel, *European J. Biochem.*, 1971, **22**, 57.
[806] G. Nakos and L. Mortenson, *Biochim. Biophys. Acta*, 1971, **227**, 576.

duplication recognized in bacterial ferredoxins has not occurred in plant ferredoxins,[4] which are also much larger (*ca.* 110 residues) and contain five or six cysteine residues, only four of which are probably essential.[807] The ferredoxin of *E. coli* (m. wt. 12 600) has two atoms of iron and labile sulphide and resembles plant-type ferredoxins rather than those of anaerobic bacteria (*e.g. Desulfovibrio*).[808] The role of ferredoxins in the origin of life and in evolution has been discussed.[809]

When certain anaerobic bacteria are grown on an iron-deficient medium, flavodoxins are produced instead of ferredoxins, and can replace ferredoxins in electron transport. Using a sequenator, the first 41 residues of *Peptostreptococcus elsdenii* flavodoxin and the first 51 of that of *C. pasteurianum* were determined;[810] the sequence of the former (137 residues) was almost completed[811] and the latter was extended. In the regions that can be compared the homology is 45% (Figure 20). If a cysteine is involved in FMN binding this is probably Cys-128 in *P. elsdenii* since the second cysteine at position 55 is replaced by serine in *C. pasteurianum* flavodoxin. The flavodoxin from *E. coli* (m. wt. 14 000) also contains one mole of FMN but appears to resemble the phytoflavins of blue-green algae more than the flavodoxins of anaerobic bacteria.[808]

The rubredoxins are non-haem iron proteins that function as electron carriers in hydroxylations. The sequence of the rubredoxin from the aerobic microbe *Pseudomonas oleovorans* has been almost completely determined (174 residues)[812] and the homology between the regions 1—54 and the C-terminal 119—174 in the sequence is clear evidence of gene duplication. These regions also show homology with the sequences of rubredoxins from anaerobic bacteria, evidence of a common ancestor (Figure 21). The *P. oleovorans* rubredoxin can bind one or two iron atoms (*cf.* transferrin; see last year's Report, p. 67), the iron-binding sites being a group of four cysteine residues near the C-terminus and another group of four near the N-terminus.[813] In the one-Fe form the iron is bound near the C-terminus and the iron-containing peptide from the C-terminus can, in fact, serve as an electron carrier in hydroxylation.[813]

The phycobiliproteins are photosynthetic pigments containing the phycobilin prosthetic group, a linear tetrapyrrole. Several investigations have been concerned with the subunit molecular weights of these proteins. One of the previously accepted molecular weights of 30 000 for the monomer of phycocyanin has been re-offered,[814] but appears to be at variance with

[807] S. J. Aggarwal, K. K. Rao, and H. Matsubara, *J. Biochem. (Japan)*, 1971, **69**, 601.
[808] H. Vetter, jun., and J. Knappe, *Z. physiol. Chem.*, 1971, **352**, 433.
[809] D. O. Hall, R. Cammack, and K. K. Rao, *Nature*, 1971, **233**, 136.
[810] M. Tanaka, M. Haniu, G. Matsueda, K. T. Yasunobu, S. Mayhew, and V. Massey, *Biochemistry*, 1971, **10**, 3041.
[811] M. Tanaka, M. Haniu, K. T. Yasunobu, S. Mayhew, and V. Massey, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 886.
[812] A. Benson, K. Tomoda, J. Chang, G. Matsueda, E. T. Lode, M. J. Coon, and K. T. Yasunobu, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 640.
[813] E. T. Lode and M. J. Coon, *J. Biol. Chem.*, 1971, **246**, 791.
[814] O. Kao, D. S. Berns, and R. MacColl, *European J. Biochem.*, 1971, **19**, 595.

several other studies,[815, 817] which make extensive use of SDS-gel electrophoresis. These suggest that (blue) phycocyanin and (red) phycoerythrin from several strains of blue-green algae all contain two non-identical subunits, each of which is associated with a chromophore; an association

```
          1                                  10
P. e.  Met-     Val-Glu-Ile-Val-Tyr-Trp-Ser-Gly-Thr-Gly-Asn-Thr-Glu-Ala-
C. p.  Met-Lys-Val-Asn-Ile-Ile -Tyr-Trp-Ser-Gly-Thr-Gly-Asn-Thr-Glu-Ala-

                    20                                  30
P. e.  Met-Ala-Asn-Glu-Ile-Glu-Ala-Ala-Val-Lys-Ala-Ala-Gly-Ala-Asp-Val-
C. p.  Met-Ala-Lys-Leu-Ile-Ala-Glu-Gly-Ala-Gln-Glu-Lys-Gly-Ala-Glu-Val-

                              40
P. e.  Glu-Ser -Val-Arg-Phe-Glu-Asp-Thr-Asn-Val-Asp-Asn-Val-Ala-Ser -Lys-
C. p.  Lys-Leu-Leu-Asn-Val-Ser -(Asp,Ala)Lys-Glu-Asp-Asp-Val-Lys-Glu-Ala-

          50                                  60
P. e.  Asp-Val-Ile -Leu-Leu-Gly-Cys-Pro-Ala-Met-Gly-Ser-Glu-Glu-Leu-Glu-
C. p.  Asp-Val-Val-Ala-Phe-Gly-Ser -Pro-Ser -Met-Gly-Ser-Glu-Val (Ser,Gln,

                    70                                  80
P. e.  Asp-Ser -Val-Val -Glu-Pro-Phe-Phe-Thr-Asp-Leu-Ala-Pro-Lys-Gly-Lys-
C. p.  Glu,Glu,Pro,Met) Phe-Leu-Asp-Val-Val-Ser -Ser -Ile -Val-Thr-Gly-Lys-

                              90
P. e.  Lys-Leu-Lys-Val-Gly-Leu-Phe-Gly-Ser-Tyr-Gly-Trp-Ser-Trp(Gly,Gly,
C. p.  Lys-

                    100                                 110
P. e.  Glu)Met-Asp-Ala-Trp-Lys-Gln-Arg-Thr-Glu-Asp-Thr-Gly-Ala-Thr-Val-

                              120
P. e.  Ile-Gly-Thr-Ala-Ile-Val-Asn-Glu-Met-Pro-Asp-Asn-Ala-Pro-Glu-Cys-

          130                       138
P. e.  Lys-Glu-Leu-Gly-Glu-Ala-Ala-Ala-Lys-Ala
```

Figure 20 *Amino-acid sequences of the* P. elsdenii *and* C. pasteurianum *flavodoxins*

of two different chains had also been suggested earlier. The subunit molecular weights [815] of phycoerythrin are 22 000 and 20 000, and of phycocyanin 20 000 and 16 000 (other values [816, 817] are not substantially different). Allophycocyanin has either one subunit [816] of 16 000 daltons or two [815] of 17 500 and 15 000 daltons. An 11-residue peptide from phytochrome of oat seedlings (another chromoprotein with a covalently bound linear tetrapyrrole) has the partial sequence -Leu-Arg-Ala-Pro-His-X-Y-His-Leu-Gln-Tyr-.[818]

[815] A. N. Glazer and G. Cohen-Bazire, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1398.
[816] A. Bennett and L. Bogorad, *Biochemistry*, 1971, **10**, 3625.
[817] P. O'Carra and S. D. Killilea, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1192.
[818] K. T. Fry and F. E. Mumford, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1466.

B. Oxygen-transport and -storage Proteins.—*Haemoglobin and Myoglobin.* With true chauvinism, work on marsupial globins progresses in leaps and bounds. The complete amino-acid sequence of the α-chain of haemoglobin from the grey kangaroo has been determined,[819] the sequence of the β-chain

```
                    1
P. e.       f-Met-Asp-Lys-Tyr-Glu-Cys-Ser-Ile -Cys-Gly-Tyr-Ile -Tyr-
                    1
M. a.       f-Met-Gln-Lys-Phe-Glu-Cys-Thr-Leu-Cys-Gly-Tyr-Ile -Tyr-
P. o. (N-term)      Ala-Ser-Tyr-Lys-Cys-Pro-Asp-Cys-Asn-Tyr-Val-Tyr-
                     119
P. o. (C-term)      Leu-Lys-Trp-Ile -Cys-Ile -Thr-Cys-Gly-His-Ile -Tyr-

                                                          25  26
P. e.       Asp-Glu-Ala-Glu-Gly-Asp-Asp-Gly-Asn-Val-Ala-Ala-Gly-Thr-
                                                          25  26
M. a.       Asp-Pro-Ala-Leu-Val-Gly-Pro-Asp-Thr-Pro-Asp-Gln-Asp-Gly-
                                                      24  25
P. o. (N-term) Asp-Glu-Ser-Ala-Gly-Asn-Val-His-Glu-Gly-Phe-Ser-Pro-Gly-
                                                  143 144
P. o. (C-term) Asp-Glu-Ala-Leu-Gly-Asp-Glu-Ala-Glu-Gly-Phe-Thr-Pro-Gly-
                 Trp

P. e.       Lys-Phe-Ala-Asp-Leu-Pro-Ala-Asp-Trp-Val-Cys-Pro-Thr-Cys-
M. a.       Ala-Phe-Glu-Asp-Val-Ser-Glu-Asn-Trp-Val-Cys-Pro-Leu-Cys-
P. o. (N-term) Thr-Pro-Trp-His-Leu-Pro-Glu-Asp-Trp-Asp-Cys-Pro-Cys-Cys-
                                 Ile
P. o. (C-term) Thr-Phe-Glu-Asn-Ile-Pro-  -Asp-Trp-Asp-Cys-Cys-Trp-Cys-
                 Arg

                                                49          53
P. e.       Gly-Ala-Asp-Lys-Asp-Ala-Phe-     -Val-Lys-Met-Asp
                                                49          53
M. a.       Gly-Ala-Gly-Lys-Glu-Asp-Phe-Glu-Val-Tyr-Glu-Ala
                                                49          53
P. o. (N-term)   -Ala-Val-Arg-Asp-Lys-Leu-Asp-Phe-Met-Leu-Ile-
                                               167
P. o. (C-term) Asx,Pro-Gly-Ala-Thr-Lys-Glu-Asn-Tyr-Val-Leu-Tyr-Glu-

P. o. (C-term) Glu-Lys
```

Figure 21 *Homologous sequences in* P. elsdenii, M. aerogenes, *and* P. oleovorans *rubredoxins*

having been reported earlier. As expected, the contact residues between the subunits are largely conserved and any replacements are conservative. The complete primary structure of myoglobin from the red kangaroo,[820] shows again that amino-acid replacements compared with sperm-whale myoglobin are conservative, the most striking being the change Val → Glu at position 21. The β-chain of red-kangaroo haemoglobin is identical with that of the

[819] J. M. Beard and E. O. P. Thompson, *Austral. J. Biol. Sci.*, 1971, **24**, 765.
[820] G. M. Air and E. O. P. Thompson, *Austral. J. Biol. Sci.*, 1971, **24**, 75.

grey kangaroo except that residue 56 is glycine not alanine.[821] This position was found to be variable in 11 species of the marsupial subfamily Macropodinae, where 55—56 are -Met-X- and X is Ala, Gly, Ser, or Thr. In the potoroo, *Potorous tridactylus*, residue 55 was leucine.[821] The divergence of the Macropodinae and Potoroinae, estimated from comparison of the complete sequence of the potoroo β-chain with that of the grey kangaroo, appears to have started *ca.* 50 million years ago.[821] A wider comparison of globin sequences gave an estimate of the time in evolution when marsupials and eutherians diverged.[822]

	55	60	65
Leghaemoglobin *a*	-Val-**Asn-Pro-Lys**-Leu-Thr-Gly-**His**-Ala-Glu-**Lys**-Leu-Phe-		
Human γ-chain	-Gly-**Asn-Pro-Lys**-Val-Lys-Ala-**His**-Gly-Lys-**Lys**-Val-Leu-		
	70	**75**	
Leghaemoglobin *a*	Ala-Leu-Val-Arg-**Asp**-Ser-Ala-Gly-Gln-**Leu**-Lys-Ala-Ser-		
Human γ-chain	Thr-Ser-Leu-Gly-**Asp**-Ala-Ile-Lys-His-**Leu**-Asp-Asp-Leu-		
	80	**85**	**90**
Leghaemoglobin *a*	Gly-Thr-Val-Val-Ala-Asp-Ala-Ala-**Leu**-Gly-Ser-Val-**His**-		
Human γ-chain	Lys-Gly-Thr-Phe-Ala-Gln-	-**Leu**-Ser-Gly-Leu-**His**-	
	100		
Leghaemoglobin *a*	Ala-Gln-**Lys**-Ala-Val-Thr-Asn-**Pro-Glu**-		
Human γ-chain	Cys-Asp-**Lys**-Leu-His-Val-Asp-**Pro-Glu**-		

Figure 22 *Amino-acid sequences surrounding the haem-linked histidines in soybean leghaemoglobin* a *and in human γ-chain*

The complete sequence of soybean leghaemoglobin *a* (142 residues) is now available.[823] There is 12% identity with human globin in the sequence of 48 residues surrounding the two haem-binding histidines (Figure 22), strong support for evolution from a common ancestor that existed before plants and animals diverged. The histidine residues in leghaemoglobin are 30 residues apart rather than the usual 28, and the -Lys-Lys- sequence usually found near one of the haem-linked histidines is also absent.[823] Haemoglobin II from the edible frog (*Rana esculenta*) has the terminal sequences:[824]

β-chain:	N-terminus	Gly-Ser-Asp-Leu-
	C-terminus	-Lys-Ala-Tyr-His
α-chain:	N-terminus	Acetyl-Ala-Leu-
	C-terminus	-Lys-Tyr

The α-chain of haemoglobin from the viper *Vipera aspis* has N-terminal valine, like mammalian α-chains, rather than the acetylated N-terminus characteristic of amphibians,[825] and the α-chain of haemoglobin from White Leghorn chickens resembles that of man (35 differences) more than that of

[821] E. O. P. Thompson and G. M. Air, *Austral. J. Biol. Sci.*, 1971, **24**, 1199.
[822] G. M. Air, E. O. P. Thompson, B. J. Richardson, and G. B. Sharman, *Nature*, 1971, **229**, 391.
[823] N. Ellfolk and G. Sievers, *Acta Chem. Scand.*, 1971, **25**, 3532.
[824] J.-P. Chauvet and R. Acher, *Internat. J. Protein Res.*, 1971, **3**, 261.
[825] M. Duguet, J.-P. Chauvet, and R. Acher, *F.E.B.S. Letters*, 1971, **18**, 185.

horse (40 differences).[826] The β-chain of haemoglobin from the Japanese monkey (*Macaca fuscata fuscata*) is identical with that known for the β-chain of the rhesus monkey (*M. mulatta*) apart from the change Leu-33 → Met.[827] The chains of the minor adult haemoglobin, $\alpha_2\delta_2$, differ in man and the chimpanzee only at position δ126, which is methionine in man and valine in the chimpanzee.[828] The two species are thus similar in all their globin chains, and even in the mechanism whereby the relative amounts of the two γ-chains (the products of allelic genes) change a few months after birth.[829] The two foetal haemoglobins F and F_1 can be separated by polyacrylamide gel electrophoresis [830] and it now appears [831] that in F_1 both γ-chains are acetylated at the N-terminus. This fractionation method does not require prior removal of haem and will resolve all globin chains. Two allelic genes have been demonstrated for rabbit β-chains [832] and the existence of two α-chain genes has been confirmed.[833] *X*-Ray crystallography of the (monomeric) haemoglobin of the lamprey [834] shows that the haem pocket is roughly the same as that in other globin chains, although the additional 11 residues at the N-terminus cause significant differences elsewhere in the tertiary structure. The results also suggest [834] that two residues should be inserted into the known sequence so that the correct sequences should read:

-Met-Ser-Ser-Met- and -Lys-Leu-Arg-
94 **135**

The large haemoglobins that constitute the respiratory pigments of invertebrates continue to be studied, and an improved *X*-ray analysis of *Chironomus thummi* erythrocruorin at 2.5 Å resolution has been presented.[835] Particularly noteworthy is the presence of seven phenylalanine residues in the haem pocket (in contrast to the three of vertebrates) and the replacement of the distal histidine at E7, invariant in all known vertebrates, by a glutamic acid residue which makes no contact with the haem and protrudes into solution.[835] The structural basis of the pH-induced structural change in the crystalline state has been discussed [836] and the ligands occupying the sixth co-ordination position of the haem in components I and III of erythrocruorin have been identified.[837] The N-terminal 45 residues of haemoglobin

[826] G. Matsuda, H. Takei, K. C. Wu, and T. Schiozawa, *Internat. J. Protein Res.*, 1971, **3**, 173.
[827] G. Matsuda, T. Maita, H. Ota, I. Tachikawa, Y. Yanaka, A. Araya, and Y. Nakashima, *Internat. J. Protein Res.*, 1971, **3**, 53.
[828] W. W. W. De Jong, *Nature New Biol.*, 1971, **234**, 176.
[829] W. W. W. De Jong, *Biochim. Biophys. Acta*, 1971, **251**, 217.
[830] L. D. Stegink, P. D. Meyer, and R. Chalkley, *Analyt. Biochem.*, 1971, **41**, 351.
[831] L. D. Stegink, P. D. Meyer, and M. C. Brummel, *J. Biol. Chem.*, 1971, **246**, 3001.
[832] J. Delaunay, N. Maleknia, and G. Schapira, *Biochim. Biophys. Acta*, 1971, **229**, 712.
[833] U. Flamm, J. S. Best, and G. Braunitzer, *Z. physiol. Chem.*, 1971, **352**, 885.
[834] W. A. Hendrickson and W. E. Love, *Nature New Biol.*, 1971, **232**, 197.
[835] R. Huber, O. Epp, W. Steigemann, and H. Formanek, *European J. Biochem.*, 1971, **19**, 42.
[836] R. Huber, O. Epp, and H. Formanek, *J. Mol. Biol.*, 1971, **57**, 377.
[837] G. Braunitzer, H. Neuwirth, and F. Reinhard, *Z. physiol. Chem.*, 1971, **352**, 757.

from the marine worm *Glycera dibranchiata* (determined in a sequenator) [838] show high homology with sperm-whale myoglobin, a similar tertiary structure having been suggested some years ago. The highly co-operative haemoglobin of *Arenicola cristata* (a polychaete annelid) has two types of chain with m. wt. 13 000 and 14 000 daltons,[839] and it is suggested that one haem is bound by two chains, giving about 96 haems per molecule. The two types of chain in chlorocruorin from the marine worm *Spirographis spallanzani* (both of m. wt. *ca.* 35 000) have N-termini Glu- and Ala-, and C-termini -Arg and -Tyr-His, respectively.[840]

The complete primary structure of human myoglobin has been determined [841] and subsequently amended slightly [842] so that residue 83 is Glu not Gln, and the sequence 19—22 is -Ala-Asp-Ile-Pro-. It agrees with an earlier partial structure [843] in all but a few points. Not unexpectedly perhaps, it resembles other mammalian myoglobins more than it does human haemoglobin, and the 26 differences compared with sperm-whale myoglobin either occur on the surface of the molecule or are highly conservative. Of particular interest is the replacement of valine at B2 by proline, so far found only in the globin of *Chironomus*, a feature which should impart extra stability to the corner.[841] Myoglobin from the gibbon (*Hylobates agilis*) is identical with that of man but for the replacement of Gly at position 23 by Ser.[842] A partial sequence for sheep-heart myoglobin,[844] and the N-terminal sequence of dolphin myoglobin determined on a home-built sequenator,[845] have been reported. This almost completes the determination of the dolphin sequence, with two corrections [845] of earlier results: residue 26 is Glu and residue 15 is Gly. All the sequences are highly homologous with that of sperm-whale myoglobin. Myoglobin from *Aplysia limacina* is unusual in having an acetylated N-terminus and a single histidine; the sequence of 63 residues at the C-terminus is reported.[846] Immunochemical cross-reactivity between peptides from sperm whale and fin-back whale has been used to detect similarities between the proteins [847] (*cf.* immunochemical studies on lysozyme and α-lactalbumin, this volume, p. 10,8 and last year's Report), and a warning has been sounded [848] on proteolytic activity in commercial myoglobin preparations. The activity was purified on an affinity column and resembled trypsin and chymotrypsin.

[838] S. L. Li and A. F. Riggs, *Biochim. Biophys. Acta*, 1971, **236**, 208.
[839] L. Waxman, *J. Biol. Chem.*, 1971, **246**, 7318.
[840] D. Guerritore and R. Zito, *Biochim. Biophys. Acta*, 1971, **229**, 720.
[841] A. E. Romero Herrera and H. Lehmann, *Nature New Biol.* 1971, **232**, 149.
[842] A. E. Romero Herrera and H. Lehmann, *Biochim. Biophys. Acta*, 1971, **251**, 482.
[843] R. L. Hill, C. M. Harris, J. F. Naylor, and W. M. Sams, *J. Biol. Chem.*, 1969, **244**, 2182.
[844] K. Han, D. Tetaert, M. Dautrevaux, V. Moschetto, and G. Biserte, *F.E.B.S. Letters*, 1971, **15**, 116.
[845] I. Kluh and A. Bakardjieva, *F.E.B.S. Letters*, 1971, **17**, 31.
[846] L. Teutori, G. Vivaldi, S. Carta, M. Marinucci, A. Massa, E. Antonini, and M. Brunori, *F.E.B.S. Letters*, 1971, **12**, 181.
[847] M. Z. Atassi and B. J. Saplin, *Biochemistry*, 1971, **10**, 4740.
[848] D. F. Goldspink, D. Holmes, and R. J. Pennington, *Biochem. J.*, 1971, **125**, 865.

Abnormal haemoglobins whose oxygen affinities are atypical have been divided into those in which the amino-acid substitution affects the oxygen-binding site direct and those in which the substitution affects the balance between the tertiary structures of the oxy- and deoxy-forms.[849] Up to the present, abnormal haemoglobins have been identified solely on the basis of abnormal electrophoretic mobility, which means that most of those that have been characterized show a replacement of an amino-acid by one of dissimilar charge. In a new method [850] which permits identification of neutral mutations, the abnormality is revealed by a slight change in chromatographic mobility of the abnormal peptide in a digest. To avoid the uncertainties of comparison with a standard map, the abnormal peptide is identified as the only non-radioactive one when the abnormal haemoglobin and [^{14}C]haemoglobin are co-digested. Thus in Hb-*Christchurch*, which results in haemolytic anaemia, the mutation was shown to be β71 Phe → Ser, important because β71 forms a hydrophobic haem contact. Various other new abnormal haemoglobins have been reported, three of them α-chain variants: Hb-*Fort Worth* (α27 Glu → Gly),[851] Hb-*Atago* (α85 Asp → Tyr),[852] and Hb-*Rampa* (α95 Pro → Ser).[853] Among the β-chain variants characterized are Hb-*Rainier* (β145 Tyr → Cys),[854] Hb-*Bethesda* (β145 Tyr → His),[854] Hb-*Toulouse* (β66 Lys → Glu),[855] Hb-*Bucuresti* (β42 Phe → Leu),[856] Hb-*Ta-Li* (β83 Gly → Cys),[857] and Hb-J *Kaolisiung* (β59 Lys → Thr).[858] A re-examination [859] of Hb-*Hiroshima* shows that it has the substitution β146 His → Asp and not β143 His → Asp as reported earlier, consistent with the proposed role of His-143 in 2,3-diphosphoglycerate (DPG) binding and the apparently normal response of the oxygen affinity of Hb-*Hiroshima* to DPG.[859] Hb-*Toulouse* [855] is noteworthy for the rupture of an ionic bond in the haem pocket, and Hb-*Ta-Li* [857] and -*Rainier* [854] for a substitution by cysteine, resulting in the former case in a tendency to polymerize (*in vitro*) and in the latter case in the formation of an intramolecular disulphide bond, as shown by *X*-ray crystallography (see below). Hb A_2-*Indonesia* [860] has the change δ69 Gly → Arg, and in Hb-*Myiada* [861]

[849] H. Morimoto, H. Lehmann, and M. F. Perutz, *Nature*, 1971, **232**, 408.
[850] R. W. Carrell and M. C. Owen, *Biochim. Biophys. Acta*, 1971, **236**, 507.
[851] R. G. Schneider, B. Brimhall, R. T. Jones, R. Bryant, C. B. Mitchell, and A. I. Goldberg, *Biochim. Biophys. Acta*, 1971, **243**, 164.
[852] N. Fujiwara, T. Maekawa, and G. Matsuda, *Internat. J. Protein Res.*, 1971, **3**, 35.
[853] W. W. W. De Jong, L. F. Bernini, and P. M. Khan, *Biochim. Biophys. Acta*, 1971, **236**, 197.
[854] A. Hayashi, G. Stamatoyannopoulos, H. Yoshida, and J. Adamson, *Nature New Biol.*, 1971, **230**, 264.
[855] D. Labie, J. Rosa, O. Belkhodja, and R. Bierme, *Biochim. Biophys. Acta*, 1971, **236**, 201.
[856] V. Bratu, P. A. Lorkin, H. Lehmann, and C. Predescu, *Biochim. Biophys. Acta*, 1971, **251**, 1.
[857] R. Q. Blackwell, C.-S. Liu, and C.-L. Wang, *Biochim. Biophys. Acta*, 1971, **243**, 467.
[858] R. Q. Blackwell, C.-S. Liu, and T.-B. Shih, *Biochim. Biophys. Acta*, 1971, **229**, 343.
[859] M. F. Perutz, P. Del Pulsinelli, L. Ten Eyck, J. V. Kilmartin, S. Shibata, I. Iuchi, T. Miyaji, and H. B. Hamilton, *Nature New Biol.*, 1971, **232**, 147.
[860] L. I. L. Eng, W. Pribadi, F. W. Boerma, G. D. Efremov, J. B. Wilson, C. A. Reynolds, and T. H. J. Huisman, *Biochim. Biophys. Acta*, 1971, **229**, 335.
[861] Y. Ohta, K. Yamaoka, I. Sumida, and T. Yanase, *Nature New Biol.*, 1971, **234**, 218.

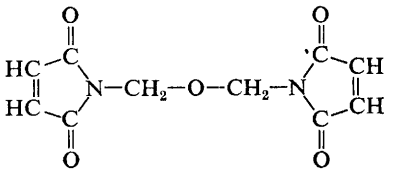
normal β-chains are replaced by chains which are the result of non-homologous cross-over of δ and β genes. It differs from Hb-*Lepore* in that the N-terminus in Hb-*Myiada* derives from the β-chain gene and the C-terminus from the δ gene, rather than *vice versa*. Two other unusual abnormal haemoglobins have elongated α- (Hb-*Constant Spring*)[862] or β-(Hb-*Tak*) chains,[863] and it appears that the explanation is not as simple as a mutation in the normal chain-termination codon.[862] Three further variants of human myoglobin have also been reported,[864] two of them with substitutions at the same position: Arg-139 → Gln,[864a] Arg-139 → Trp,[864b] and Lys-133 → Asn [864c] (numbering of residues as in ref. 841). The contribution of structural studies of abnormal haemoglobins to a precise understanding of the functional consequences of mutation are discussed below. Isoelectric focusing in polyacrylamide gels is likely to be a useful method for characterizing haemoglobin variants,[865] and human Hb A_2 is easily separated from Hb C by chromatography on phosphocellulose,[866] although the variability between different batches of adsorbent is likely to be a nuisance in this method.

The *X*-ray analysis [867] of deoxy Hb-*Rainier* (β145 Tyr → Cys) shows that the cysteine introduced by the mutation forms a disulphide bridge with Cys-93 of the same β-chain, and the abnormal physico-chemical properties of the molecule can be explained.[867] An elegant approach to structure–function relationships is possible through a combination of *X*-ray studies and the availability of more than one variant haemoglobin involving a single position in the chain. Thus crystallographic studies of Hb-*Kansas* (β102 Asn → Thr) and Hb-*Richmond* (β102 Asn → Lys) showed [868] that the highly abnormal properties of the former arose from the distortion caused by the additional methyl group while the properties of the latter were virtually normal; Asn-102 is known to be important since it forms the only hydrogen-bond across the $\alpha_1\beta_1$ interface in oxy-haemoglobin. Further crystallographic studies [869] show that when tyrosine replaces a haem-linked histidine (α87 in Hb M-*Iwate*, or β92 in Hb M-*Hyde Park*) the mutant chains are stabilized in the ferric form and will not combine reversibly with oxygen; the detailed structural changes are described.[869] Arg-92α of normal haemoglobin is replaced by glutamine in Hb J-*Capetown* and by leucine in Hb-*Chesapeake*. The properties of the molecules, resulting from different

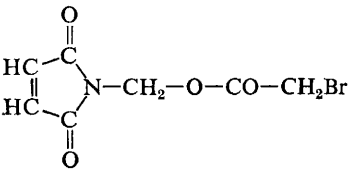
[862] J. B. Clegg, D. J. Weatherall, and P. F. Milner, *Nature*, 1971, **234**, 337.
[863] G. Flatz, J. L. Kinderlerer, J. V. Kilmartin, and H. Lehmann, *Lancet*, 1971, 732.
[864] (*a*) F. E. Boulton, R. G. Huntsman, G. I. Yawson, A. E. Romero Herrera, P. A. Lorkin, and H. Lehmann, *Brit. J. Haematol.*, 1971, **20**, 69; (*b*) F. E. Boulton, R. G. Huntsman, A. E. Romero Herrera, P. A. Lorkin, and H. Lehmann, *Biochim. Biophys. Acta*, 1971, **229**, 716; (*c*) F. E. Boulton, R. G. Huntsman, A. E. Romero Herrera, P. A. Lorkin, and H. Lehmann, *ibid.*, p. 871.
[865] J. W. Drysdale, P. Righetti, and H. F. Bunn, *Biochim. Biophys. Acta*, 1971, **229**, 42.
[866] B. F. Horton and A. I. Chernoff, *J. Chromatog.*, 1971, **63**, 414.
[867] J. Greer and M. F. Perutz, *Nature New Biol.*, 1971, **230**, 261.
[868] J. Greer, *J. Mol. Biol.*, 1971, **59**, 99.
[869] J. Greer, *J. Mol. Biol.*, 1971, **59**, 107.

types of perturbations in each case, have been the subject of a detailed kinetic investigation.[870]

Correlation of specific chemical modification of proteins with an alteration in properties is another way of studying the same problem of structure–function relationships; when this is combined with *X*-ray analysis of the modified molecule the approach becomes doubly powerful. The functional properties of haemoglobin substituted at Cys-93 with a variety of N-substituted maleimides depended not only on the nature of the modifying reagent but also on the ligand state of the protein.[871, 872] In general it was found that reaction with the oxy-form gave a derivative in which co-operativity was eliminated, while it was only impaired if the deoxy-form was used. Two bifunctional reagents reacted at a second site in horse oxy-haemoglobin after initial reaction at Cys-93; bis(maleimidomethyl) ether [871] (18) reacted at His-97β, and *N*-α-(bromoacetoxymethyl)maleimide [873] (19)



(18)



(19)

at Val-1β, although the ester bond in this cross-link was then hydrolysed rapidly. Reaction with the deoxy-form was confined to Cys-93, also the site of reaction of monofunctional N-substituted maleimides with both the oxy- and deoxy-forms.[871] *X*-Ray studies [874] showed that the loss of co-operativity in derivatives alkylated at Cys-93 arose from structural changes that stabilized the oxy-configuration even in the absence of ligands; in particular, displacement of the hydrogen-bonded position of Tyr-145β distorted the region of the β-chain involved in the $\alpha_1\beta_2$ contact. Hence co-operativity, which involves a change of conformation on binding or release of ligands, disappears. The functional consequences of structural alterations have been discussed in detail.[875] The absence of any effect of 2,3-DPG on haemoglobin treated with bis(maleimidomethyl) ether is thus taken to imply that an essential conformational change is involved in DPG function.[876] On the other hand, when Cys-93β reacts with DTNB there is no

[870] R. L. Nagel, Q. H. Gibson, and T. Jenkins, *J. Mol. Biol.*, 1971, **58**, 643.
[871] D. J. Arndt, S. R. Simon, T. Maita, and W. Konigsberg, *J. Biol. Chem.*, 1971, **246**, 2602.
[872] S. R. Simon, D. J. Arndt, and W. Konigsberg, *J. Mol. Biol.*, 1971, **58**, 69.
[873] D. J. Arndt and W. Konigsberg, *J. Biol. Chem.*, 1971, **246**, 2594.
[874] J. K. Moffat, *J. Mol. Biol.*, 1971, **58**, 79.
[875] J. K. Moffat, S. R. Simon, and W. Konigsberg, *J. Mol. Biol.*, 1971, **58**, 89.
[876] J. M. Salhany, *F.E.B.S. Letters*, 1971, **14**, 11.

loss of co-operativity,[877] although the Bohr effect and the oxygen affinity are altered, and if loss of co-operativity is indeed associated with the structural changes revealed by the *X*-ray work [874] then it appears that the large thiolate residue at β-93 does not expel Tyr-145β. Similar retention of co-operativity was observed when Cys-93β reacted with mercurials.[878] A study [879] of the binding of CO_2 by haemoglobin in which either the α-chains or the β-chains, or both, had been modified with cyanate, with retention of co-operativity, confirmed the view that under physiological conditions haemoglobin binds CO_2 through the four terminal amino-groups, and that the amino-termini of the α-chains are those involved in the Bohr effect. The pK_a of the α-amino-group of the free α-chain, determined by reaction with fluorodinitrobenzene,[880] is 7.4. Support for the role of the amino-termini of the β-chains in the control of oxygen binding comes from a study of the two major haemoglobins of the cat;[881] one has the N-terminal sequence Gly-Phe- and is sensitive to 2,3-DPG whereas the other has the blocked N-terminal sequence Acetyl-Ser-Phe- (uncharacteristic of mammalian haemoglobins) and is not sensitive. Further, it has now been suggested [882] that DPG does, in fact, affect the Bohr effect. Thus differences in the Bohr effect of the haemoglobins of mouse, man, and elephant in the presence of physiological concentrations of DPG are attributed to differences in the sensitivity of the oxygen equilibria to DPG arising from small differences in the environment of the DPG-binding site as a result of mutation, reflecting adaptation to the metabolic needs of the animal (greatest in the mouse).[882]

Haemerythrin. The primary structure of the non-haem respiratory pigment, haemerythrin, of the sipunculid worm *Dendrostomum pyroides* has been determined [883] using the sequenator for the first 35 residues, and comparison of tryptic peptides to establish the remainder by homology with the sequence of haemerythrin from *Golfingia gouldii*. The sequences appear to differ in only four positions, and an error in the *G. gouldii* sequence is corrected;[884] residues 10 and 11 are Trp and Asp respecitvely. From two reports concerned with identification of the tyrosine residues implicated in iron binding in haemerythrin [885, 886] Tyr-8 and Tyr-109 are the most likely candidates. It is interesting that side-reactions in the use of tetranitromethane were observed in one case [885] but not the other.[886]

[877] G. Amiconi, E. Antonini, M. Brunori, A. Nason, and J. Wyman, *European J. Biochem.*, 1971, **22**, 321.
[878] B. Giardina, I. Binotti, G. Amiconi, E. Antonini, M. Brunori, and C. H. McMurray, *European J. Biochem.*, 1971, **22**, 327.
[879] J. V. Kilmartin and L. Rossi-Bernardi, *Biochem. J.*, 1971, **124**, 31.
[880] S. H. De Bruin and E. Bucci, *J. Biol. Chem.*, 1971, **246**, 5228.
[881] F. Taketa, A. G. Mauk, and J. L. Lessard, *J. Biol. Chem.*, 1971, **246**, 4471.
[882] S. Tomita and A. Riggs, *J. Biol. Chem.*, 1971, **246**, 547.
[883] R. E. Ferrell and G. B. Kitto, *Biochemistry*, 1971, **10**, 2923.
[884] R. E. Ferrell and G. B. Kitto, *F.E.B.S. Letters*, 1971, **12**, 322.
[885] R. L. Hill and I. M. Klotz, *Arch. Biochem. Biophys.*, 1971, **147**, 226.
[886] J. L. York and C. C. Fan, *Biochemistry*, 1971, **10**, 1659.

7 Immunoglobulins

The interested reader need fear no shortage of reviews in this fast-moving field. Among the good ones this year,[887] that by Williamson can be particularly recommended as a brief introduction to some of the recent developments in the study of cell–cell interactions involved in antibody biosynthesis. A report [888] of the relevant colloquium at the 7th F.E.B.S. Meeting in Varna, Bulgaria, in September 1971 serves to remind one of the strong divergence of opinion about the problem of the genetics of antibody formation, which is also evident in the reviews cited.

The use of an immunoadsorbent (an antibody against myeloma IgE trapped in polyacrylamide gel) for the isolation of IgE has been described [889] and 7s subunits of IgM have been obtained from the lemon shark.[890] Natural 7s IgM has been isolated from human sera and some preliminary structural work has been carried out on its disulphide bridges.[891] Immunoglobulin-like molecules that contain no light chains but only heavy chains of IgM have been reported in the serum of bursectomized, irradiated chicks.[892]

The immune macroglobulin from the gar, *Lepisosteus osseus*,[893] and the paddlefish, *Polyodon spatula*,[894] has a molecular weight of *ca.* 650 000 and appears to have a tetrameric $(H_2L_2)_4$ structure. The same is probably true of the IgM from the giant grouper, *Epinephelus itaira*.[895] No free N-terminal residue could be detected in the light chains of paddlefish IgM, but the heavy chain was shown to have the N-terminal sequence Asp-Val-Val-$\begin{matrix}\text{Leu}\\\text{Val}\end{matrix}$-Thr-.[894] The duck contains curious 5.7s and 7.8s immunoglobulins and it has been reported [896] that the heavy chains in these proteins have molecular weights of 35 000 and 59 000 respectively. Since a study of the evolution of the amino-acid sequence of normal γ-chains shows that the constant region is three times the length of the variable region and suggests that the molecule arose by a series of gene duplication events,[887] it is proposed [896] that the constant region of the heavy chain of duck 5.7s immunoglobulin is only twice as long as the variable region, whereas that

[887] R. R. Porter, 'Harvey Lectures 1969—70' Academic Press, New York, p. 157; C. Milstein and A. J. Munro, *Ann. Rev. Microbiol.*, 1970, **24**, 335; J. R. L. Pink, A. C. Wang, and H. H. Fudenberg, *Ann. Rev. Med.*, 1971, **22**, 145; G. P. Smith, L. Hood, and W. M. Fitch, *Ann. Rev. Biochem.*, 1971, **41**, 969; A. R. Williamson, *Nature*, 1971, **231**, 359.

[888] M. Sela, *F.E.B.S. Letters*, 1971, **19**, 181.

[889] S. Carrel, L. Theilkäs, A. Morell, F. Skvaril, and S. Barandum, *Biochem. J.*, 1971, **122**, 405.

[890] D. G. Klapper, L. W. Clem, and P. A. Small, jun., *Biochemistry*, 1971, **10**, 645.

[891] F. Dolder, *Biochim. Biophys. Acta*, 1971, **236**, 675.

[892] Y. S. Choi and R. A. Good, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2083.

[893] R. T. Acton, P. F. Weinheimer, H. K. Dupree, E. E. Evans, and J. C. Bennett, *Biochemistry*, 1971, **10**, 2028.

[894] R. T. Acton, P. F. Weinheimer, H. K. Dupree, T. R. Russell, M. Wolcott, E. E. Evans, R. E. Schrohenloher, and J. C. Bennett, *J. Biol. Chem.*, 1971, **246**, 6760.

[895] L. W. Clem, *J. Biol. Chem.*, 1971, **246**, 9.

[896] B. Zimmerman, N. Shalatin, and H. M. Grey, *Biochemistry*, 1971, **10**, 482.

of the 7.8s molecule is four times as long. If borne out by further work, this will be a most interesting observation.

A study of the hydrolysis of IgG1 and IgG2 of guinea-pig with papain and pepsin has been described:[897] IgG1 is far more susceptible to proteolysis than IgG2. It has been found [898] that the N-terminal aspartic acid or glutamic acid residues of light chains in intact antibodies are incapable of accepting arginine from Arg-tRNA in an enzyme-catalysed reaction (see also refs. 331, 423) whereas isolated light chains will serve as acceptors, suggesting that their N-terminal residues are folded away in the intact immunoglobulin.

A. Light Chains.—More than 95% of the light chains from chicken and turkey immunoglobulins appear to be homologous with mammalian λ-chains [899] and the light chains of the African lungfish (*Protopterus aethiopicus*) also have N-terminal residues that are predominantly blocked.[900] The unblocked class of chains has the N-terminal sequence Asp-( ) ( )-Leu-Thr-Glx-Asx-Ala-Ser-, in which two deletions are postulated in order to maximize the homology with mammalian light chains,[900] but this sequence is no more like that of human κ-chains than human λ-chains as far as it goes.

However, most recent work has been concerned with human light chains and Bence-Jones proteins. Large deletions, one of which is in the variable region, have been detected in two myeloma light chains,[55, 901] but the relevance to normal antibody variability is obscure. Carbohydrate has been found in the variable region of two myeloma κ-chains [902] and in both cases the amino-acid sequence in the attachment site fits the postulate of sequence specificity for attachment (see last year's Report). The amino-acid sequences of several human κ-chains have also been reported:[903–905] they all belong to subgroup κ_I but indicate that that subgroup may need to be further divided. Similarly, the amino-acid sequences of two human λ-chains of subgroup II [906, 907] and one of subgroup IV [908] have been determined. One of the chains in subgroup II has histidine as N-terminal residue,[906] which is unique in the primary structures of light chains so far reported. Moreover, the existence of genetic polymorphism involving a Ser/Gly interchange at position 154 in the constant region of human λ-chains has been firmly

[897] R. G. Q. Leslie, M. D. Melamed, and S. Cohen, *Biochem. J.*, 1971, **121**, 829.
[898] R. L. Soffer and J. D. Capra, *Nature New Biol.*, 1971, **233**, 44.
[899] J. A. Grant, B. Sanders, and L. Hood, *Biochemistry*, 1971, **10**, 3123.
[900] G. W. Litman, A. C. Wang, H. H. Fudenberg, and R. A. Good, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2321.
[901] O. Smithies, D. Gibson, E. M. Fanning, M. E. Percy, D. M. Parr, and G. E. Connell, *Science*, 1971, **172**, 574.
[902] C. P. Milstein and C. Milstein, *Biochem. J.*, 1971, **121**, 211.
[903] C. P. Milstein and E. V. Deverson, *Biochem. J.*, 1971, **123**, 945.
[904] H. Schiechl and N. Hilschmann, *Z. physiol. Chem.*, 1971, **352**, 111.
[905] M. Braun, W. Leibold, H.-U. Barnikol, and N. Hilschmann, *Z. physiol. Chem.*, 1971, **352**, 647.
[906] H. Ponstingl and N. Hilschmann, *Z. physiol. Chem.*, 1971, **352**, 859.
[907] F. A. Garver and N. Hilschmann, *F.E.B.S. Letters*, 1971, **16**, 128.
[908] H. Ponstingl, M. Hess, and N. Hilschmann, *Z. physiol. Chem.*, 1971, **352**, 247.

established.[909, 910] The interchange is not the result of allelic genes and there appears to be no restriction on the combination of the two possibilities for position 154 with the two possibilities (Lys/Arg) for the Oz interchange at position 191.[909b, 910] In addition, the same variation is found in the constant regions of normal human λ-chains as well as those associated with myelomatosis,[910] and any of the possible λ-chain constant regions seems able to fuse with any of the λ-chain variable-region subgroups.[909b] Formal proof of the non-allelic nature of the variable-region subgroups for human λ-chains has now been produced [911] and the first amino-acid sequence of a mouse myeloma light chain, MOPC 104, has been determined.[912]

During the fractionation of a human Bence-Jones protein, Au, it was observed [913] that the variable region was released (probably by proteolysis in the urine) and that the fragment could be crystallized. In line with the idea that variable and constant regions arose by gene duplication and fusion, the independent folding of variable and constant regions of a λ-chain has been reported.[914] The common folded structure of Bence-Jones proteins has been demonstrated in a study of the reactivity of tyrosine and histidine residues towards iodination.[915]

Amyloidosis is a disease of unknown aetiology characterized by the deposition of a fibrous protein in the extracellular connective tissue of the organ involved. In some cases the deposition of amyloid protein in man is also associated with multiple myeloma, and N-terminal sequence analysis of two such proteins (using a sequenator for 35 and 36 residues respectively) shows [916] that both proteins are derived from homogeneous κ-chains of variable-region subgroup I. On the other hand, human and monkey amyloid proteins of the typical kind (a variety of amyloidosis associated with chronic inflammation) have major components with the following N-terminal sequences:[917]

Human	**Arg-Ser-Phe-Phe-Ser-Phe-Leu-Gly-Glu-Ala-Phe-Asp-Gly-Ala-Arg-**	
Monkey	**Arg-Ser-Trp-Phe-Ser-Phe-Leu-Gly-Glu-Ala-Tyr-Asp-Gly-Ala-Arg-**	
	1	**10**
Human	**Asp-Met-Trp-Arg-Ala-Tyr-Ser-Asp-Met-**	
Monkey	**Asp-Met-Trp-Arg-Ala-Tyr-Ser-Asp-Met-**	
	20	

Again, a sequenator was used for the analysis.

[909] (*a*) M. Hess and N. Hilschmann, *Z. physiol. Chem.*, 1971, **352**, 657; (*b*) M. Hess, N. Hilschmann, L. Rivat, C. Rivat, and C. Ropartz, *Nature New Biol.*, 1971, **234**, 58.
[910] D. Gibson, M. Levanon, and O. Smithies, *Biochemistry*, 1971, **10**, 3114.
[911] F. W. Tischendorf, B. Michelitsch, G. Ledderose, and M. M. Tischendorf, *J. Mol. Biol.*, 1971, **61**, 261.
[912] E. Appella, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 590.
[913] H. J. Schramm, *Z. physiol. Chem.*, 1971, **352**, 1134.
[914] I. Björk, F. A. Karlsson, and I. Berggärd, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1707.
[915] (*a*) B.-K. Seon, O. A. Roholt, and D. Pressman, *J. Biol. Chem.*, 1971, **246**, 887; (*b*) B.-K. Seon, O. A. Roholt, and D. Pressman, *Biochim. Biophys. Acta*, 1971, **229**, 396.
[916] G. G. Glenner, W. Terry, M. Harada, C. Isersky, and D. Page, *Science*, 1971, **172**, 1150.
[917] E. P. Benditt, N. Eriksen, M. A. Hermodson, and L. H. Ericsson, *F.E.B.S. Letters*, 1971, **19**, 169.

B. Heavy Chains.—The amino-acid sequence of the variable region of a human myeloma μ-chain has been determined and the evolutionary origins of μ- and γ-chains have been discussed.[918] Further work has also been described on the VH_{III} subgroup of human heavy chains characterized by an unblocked glutamic acid residue at the N-terminus. Thus, hypervariability has been reported [919, 920] for the region of residues 30—37 (identical results have already been obtained for light chains) and a study of cyanogen bromide fragments [920b] indicates that residues 86—91 and 101—109 are also hypervariable. Further, one IgG and IgM from a single myeloma patient show the same N-terminal sequence for 34 residues, suggesting that the same VH_{III} variable region exists on two different classes of heavy chain,[919] but it is also possible that the subgroup distinctions may not extend throughout the variable region.[920b]

Allotypically related sequences have been described in the variable region of rabbit heavy chains for allotypes A*a*1, A*a*2, and A*a*3 [921] (Figure 23). No such allotypic variation was detected in the constant region of the chains. A total of 16 residues, including six consecutively, show the correlation, which is a remarkably high number for allelic variation (see Section 7E) and the results emphasize the hypervariability of residues 95—115 reported previously from the same laboratory. On the other hand, the A14/A15 allotypes of rabbit IgG have been shown [922] to correlate with a Thr/Ala interchange at position 309 in the Fc region (using the human Eu numbering). Work is in progress on IgG-2 from inbred guinea-pig. Thus, five fragments that together account for 303 residues from the C-terminus of the γ2-chain have been isolated after cyanogen bromide digestion [923] and the amino-acid sequence of a fragment from the variable region (49 residues derived from positions 35—83) has been established.[924]

The C-terminal sequence of the α-chain in human IgA1 and IgA2 has been found to be:[925]

-Met-Ala-Glu-Val-Asp-Gly-Thr-Cys-Tyr

with the penultimate residue probably involved in an intrachain disulphide bridge. Carbohydrate has been shown to be attached at five specific sites in the constant region of the human μ-chain [926] but there was no obvious

[918] A. Shimizu, C. Paul, H. Köhler, T. Shinoda, and F. W. Putnam, *Science*, 1971, **173**, 629.
[919] A. C. Wang, H. H. Fudenberg, and J. R. L. Pink, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1143.
[920] (*a*) J. D. Capra, *Nature New Biol.*, 1971, **230**, 61; (*b*) J. McKehoe and J. D. Capra, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2019.
[921] L. E. Mole, S. A. Jackson, R. R. Porter, and J. M. Wilkinson, *Biochem. J.*, 1971, **124**, 301.
[922] E. Appella, A. Chersi, R. G. Mage, and S. Dubiski, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1341.
[923] B. K. Birshtein, K. J. Turner, and J. J. Cebra, *Biochemistry*, 1971, **10**, 1.
[924] B. K. Birshtein and J. J. Cebra, *Biochemistry*, 1971, **10**, 4930.
[925] J. W. Prahl, C. A. Abel, and H. M. Grey, *Biochemistry*, 1971, **10**, 1808.
[926] A. Shimizu, F. W. Putnam, C. Paul, J. R. Clamp, and I. Johnson, *Nature New Biol.*, 1971, **231**, 73.

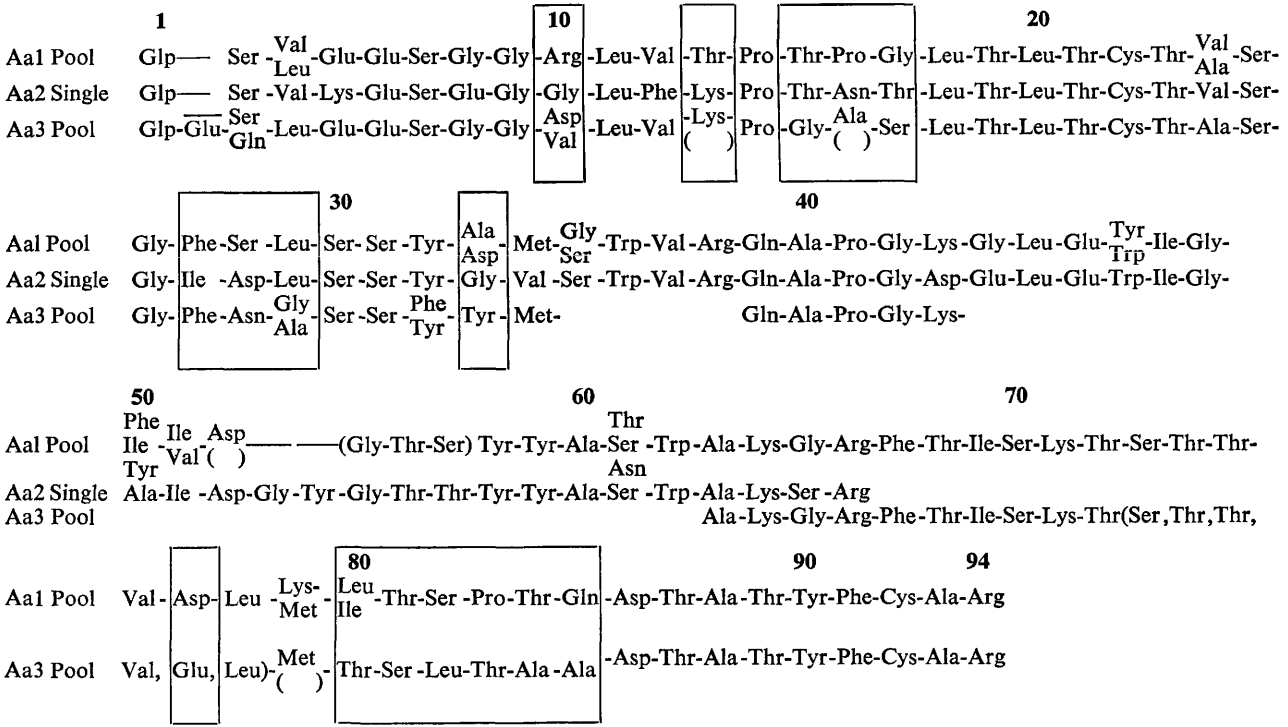
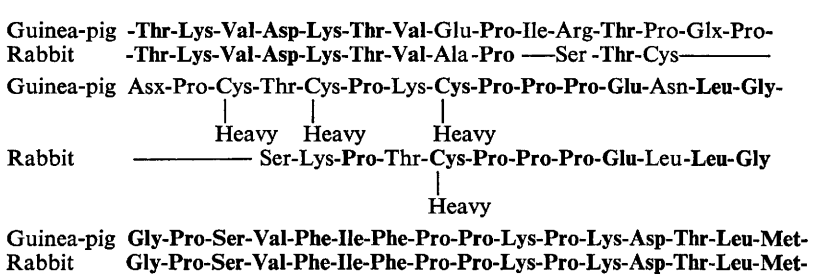


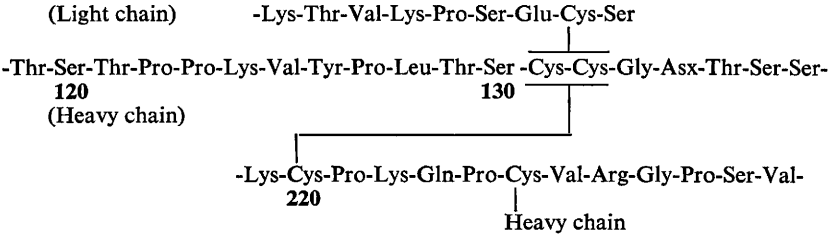
Figure 23 *Comparison of the variable-region sequences of Aa*1, *Aa*2, *and Aa*3 *rabbit heavy chains. Residues in boxes are those that vary between the Aa*1, *Aa*2, *and Aa*3 *sequences, and those in parentheses are minor variants in each sequence. The bars at positions* 2, 53, *and* 54 *indicate a presumed deletion*
(Reproduced by permission from *Biochem. J.*, 1971, **124,** 301)

homology with any glycopeptide from rabbit or human γ-chains. The carbohydrate is of unknown biological function.

C. Disulphide Bridges.—Full details have now been published of the arrangement of the disulphide bridges in the human γ2-chain [927] and five interchain disulphide bridges have been described in the γ-chains of pig immunoglobulin, the results being consistent with the existence of more than one subclass of γ-chain in the pig.[928] The amino-acid sequence in the hinge region of the IgG2 of the inbred strain 13 of guinea-pig has also been determined.[929] Three inter-heavy-chain disulphide bridges were found in a sequence comparable with that for the single inter-heavy-chain bridge of rabbit IgG:[929a]



The light chain in guinea-pig IgG2 is linked to the heavy chain by a disulphide bridge sited approximately one-quarter of the way along the heavy chain from the N-terminus,[930] like that of all other IgG except human IgG1. Other experiments [931] show that the interchain disulphide bridges of goat IgG closely resemble those of the rabbit:



Human IgA exists in two subclasses, IgA1 and IgA2. Subclass IgA2 can be further divided into two forms, Am2(+), the common variant, which has no disulphide bridge linking heavy and light chains and Am2(−), the minor variant, which has. It is reported that the amino-acid sequence in this

[927] C. Milstein and B. Frangione, *Biochem. J.*, 1971, **121**, 217.
[928] J. J. Metzger, J. Novotný, and F. Franek, *F.E.B.S. Letters*, 1971, **14**, 237.
[929] (*a*) K. J. Turner and J. J. Cebra, *Biochemistry*, 1971, **10**, 9; (*b*) B. Oliveira and M. E. Lamm, *ibid.*, p. 26.
[930] B. K. Birshtein, Q. Z. Hussain, and J. J. Cebra, *Biochemistry*, 1971, **10**, 18.
[931] P. H. Strausbach, E. Hurwitz, and D. Givol, *Biochemistry*, 1971, **10**, 2231.

bridge region is the same for γA1- and γA2Am2(−)-chains [932] but that differences occur in the amino-acid sequence of the hinge region of the γA1- and γA2-chains.[933] The primary structure of the hinge region of a human μ-chain has also been determined:[934]

Carbohydrate Heavy chain

-Val-Asp-His-Arg-Gly-Leu-Thr-Phe-Glx-Glx-Asx-Ala-Ser-Ser-Met-Cys-Val-Pro-Asp-Glu-Asp-Thr-Ala-Ile-Arg-Val-Phe-Ala-Ile-Pro-Pro-Ser-Phe-Ala-Ser-Ile-Phe-Leu-Thr-Lys-Ser-Thr-Lys-Leu—

Little sequence homology with γ-chains is apparent in this region. Potentially useful derivatives for *X*-ray crystallography have been produced by inserting mercury into the disulphide bridges of a human IgG1 myeloma protein.[935]

There have been further reports of J-chains in IgM [936, 937] and colostral IgA.[936] It is probable that the J-chain (m. wt. 26 000, but not a light chain) is covalently attached to the immunoglobulin by disulphide bridges and that it is identical in IgM and IgA.[936]

D. Antibody Binding Sites.—The golden fleece of molecular immunology, to coin a phrase, is a chemically homogeneous antibody of defined specificity, preferably one that crystallizes. The role of Jason is hotly contested. Thus, work is in progress on the structure of different antibodies from inbred strains of guinea-pig [924, 929a, 930] and of relatively homogeneous rabbit antibodies against group C streptococcal carbohydrate [938] and type III pneumococci.[939] Homogeneous rabbit antibodies against the *p*-azophenyltrimethylammonium hapten have been reported [940] and the N-terminal sequence of light chains from a homogeneous rabbit anti-*p*-azobenzoate antibody has been established:[941]

Val-Glu-Val-Leu-Thr-Glx-Thr-Pro-Ser-Pro-Val-Ser-Ala-Ala-Val-Gly-Gly-Thr-Val-Thr-Ile-

The N-terminal valine residue is unique in rabbit light chains examined so far.

The next best thing to a natural antibody is a myeloma protein to which an antibody specificity can be assigned. The mouse myeloma IgM, MOPC

[932] E. Mihaesco, M. Seligman, and B. Frangione, *Nature New Biol.*, 1971, **232**, 220; C. Wolfenstein, B. Frangione, E. Mihaesco, and E. C. Franklin, *Biochemistry*, 1971, **10**, 4140.
[933] C. A. Abel and H. M. Grey, *Nature New Biol.*, 1971, **233**, 29.
[934] C. Paul, A. Shimizu, H. Köhler, and F. W. Putman, *Science*, 1971, **172**, 69.
[935] L. A. Steiner and P. M. Blumberg, *Biochemistry*, 1971, **10**, 4725.
[936] J. Mestecky, J. Zikan, and W. T. Butler, *Science*, 1971, **171**, 1163.
[937] B. Frangione, F. Prelli, C. Mihaesco, and E. C. Franklin, *Proc. Nat. Acad. Sci. U.S.A.* 1971, **68**, 1547.
[938] J. B. Fleischman, *Biochemistry*, 1971, **10**, 2753.
[939] J.-C. Jaton, M. D. Waterfield, M. N. Margolies, K. J. Bloch, and E. Haber, *Biochemistry*, 1971, **10**, 1583.
[940] M. H. Freedman and R. H. Painter, *J. Biol. Chem.*, 1971, **246**, 4340.
[941] E. Appella, A. Chersi, O. A. Roholt, and D. Pressman, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2569.

104E, has specificity for carbohydrate haptens [942] and the primary structure of the light chains has been described.[912] However, most work has so far been carried out with the mouse IgA myeloma protein, MOPC 315, which has anti-Dnp activity (see last year's Report). The tentative amino-acid sequence of the light chains has now been established [943] (Figure 24). When

1 **10**
Glp-Ala-Val-Val-Thr-Glx-Glx-Ser-Ala-Leu-Thr-Thr-Ser-Pro-Gly-Gly-Thr-Thr-
20 **30**
Val-Leu-Thr-Cys-Arg-Ser-Ser-Thr-Gly-Ala-Val-Thr-Thr-Ser-Asn-Tyr-Ala-Asn-
40 **50**
Trp-Ile-Glx-Glx-Pro-Asx-Lys-His-Leu-Phe-Thr-Gly-Leu-Ile-Gly-Gly-Thr-Ser-
60 **70**
Asx-Arg-Ala-Pro-Gly-Val-Pro-Val-Arg-Phe-Ser-Gly-Ser-Leu-Ile-Gly-Asx-Lys-
80 **90**
Ala-Ala-Leu-Thr-Ile-Thr-Gly-Ala-Glx-Thr-Glx-Asx-Glx-Ala-Met-Tyr-Phe-Cys-
100
Ala-Leu-Trp-Phe-Arg-Asx-His-Phe-Val-Phe-Gly-Gly-Gly-Thr-Lys-Val-Thr-Val-
110 **120**
Leu-Gly-Gln-Pro-Lys-Ser-Thr-Pro-Thr-Leu-Thr-Val-Phe-Pro-Pro-Ser-Ser-Glx-
130 **140**
Glx-Leu-Lys-Glu-Asn-Lys-Ala-Thr-Leu-Val-Cys-Leu-Ile-Ser-Asx-Phe-Ser-Pro-
150 **160**
Gly-Ser-Val-Thr-Val-Ala-Trp-Lys-Ala-Asx-Gly-Thr-Pro-Ile-Thr-Glx-Gly-Val-
170 **180**
Asx-Thr-Thr-Asx-Pro-Ser-Lys-Gly-Gly-Asx-Lys-Phe-Met-Ala-Ser-Ser-Phe-Leu-
190
His-Leu-Thr-Asx-Ser-Trp-Glx-Arg-Ser-His-Asx-Phe-Ser-Thr-Cys-Gln-Val-Thr-
200 **210**
Asx-Glx-Gly-His-Thr-Val-Glx-Lys-Ser-Leu-Ser-Pro-Ala-Glu-Cys-Leu

Figure 24 *The tentative amino-acid sequence of the light chain from mouse myeloma protein MOPC* 315

compared with the light chain of MOPC 104, multiple differences in the constant region are apparent despite the fact that both are λ-chains. Assuming that these are not allelic differences, the most probable explanation is that there is polymorphism in the genes for the constant region of mouse λ-chains.[943] This has already been established for human λ-chains (see above). Crystallization with hapten of the Fab′ fragment of MOPC 315 (produced by peptic digestion) has also been reported.[944]

Rabbit and mouse antibodies against the *p*-azophenyltrimethylammonium determinant have been allowed to react with the affinity label *p*-trimethylammonium-benzenediazonium difluoroborate:[945] tyrosine residues in both heavy and light chains were specifically modified and further experiments suggested that the same local region in various antibodies is used to form

[942] N. M. Young, I. B. Jocius, and M. A. Leon, *Biochemistry*, 1971, **10**, 3457.
[943] E. P. Schulenberg, E. S. Simms, R. G. Lynch, R. A. Bradshaw, and H. N. Eisen, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2623.
[944] D. Inbar, M. Rotman, and D. Givol, *J. Biol. Chem.*, 1971, **246**, 6272.
[945] J. W. Fenton and S. J. Singer, *Biochemistry*, 1971, **10**, 1429.

active sites of different specificities. A series of bromoacetyl derivatives related to Dnp has been proposed for mapping the binding site of anti-Dnp antibodies [946] and used to study the mouse myeloma MOPC 315.[947] In particular, a bifunctional affinity label was found that cross-linked a tyrosine residue in the light chain and a lysine residue in the heavy chain, tentatively identified as residues 34 and 54 in the respective chains.[947] Another attempt [948] at cross-linking in the active site of MOPC 315 involved labelling Tyr-34 in the light chain using *m*-nitrobenzenediazonium fluoroborate and reducing the resulting azotyrosine to 3-aminotyrosine. By taking advantage of its low p*K*, the new amino-group thus introduced can, in principle, be cross-linked specifically to other nucleophiles using F_2DNB.[949] For some reason this was not possible with MOPC 315 but cross-links to light- and heavy-chain residues were produced. Their identity remains to be determined.[948] Nitration of a nitrophenyl-binding Waldenström macroglobulin using tetranitromethane has been reported [950] to inactivate the molecule by modification of tyrosine residues, the critical residues being located predominantly in the Fd fragment.

A study of the hapten-binding activity in reconstituted mouse myeloma proteins with anti-Dnp specificity suggests that unique pairs of heavy and light chains are required for a particular binding site.[951] If different heavy and light chains – even chains from other anti-Dnp myeloma proteins – were used, no activity could be recovered.

E. Some Problems of Biosynthesis.—This topic is properly beyond the scope of this Report and is covered in depth in various reviews.[887, 888] A few points, however, deserve to be noticed. The subunits of IgM [952] and IgA [953] accumulate intracellularly because the thiol groups that eventually form the disulphide bridges necessary for assembly are blocked. The nature of the blocking group is unknown (perhaps it is cysteine or glutathione) but its removal allows intracellular assembly to occur.[952] Addition of the carbohydrate found in many immunoglobulins appears to take place in several steps at different sites within the myeloma cell but it is not yet clear whether carbohydrate addition is necessary for secretion of the immunoglobulin.[954]

Moreover, the problem of the fusion of the genes for the variable and constant regions of immunoglobulin polypeptide chains is still with us. The

[946] P. H. Strausbauch, Y. Weinstein, M. Wilchek, S. Shaltiel, and D. Givol, *Biochemistry*, 1971, **10**, 4342.
[947] D. Givol, P. H. Strausbauch, E. Hurwitz, M. Wilchek, J. Haimovich, and H. N. Elsen, *Biochemistry*, 1971, **10**, 3461.
[948] N. M. Hadler and H. Metzger, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1421.
[949] P. Cuatrecasas, S. Fuchs, and C. B. Anfinsen, *J. Biol. Chem.*, 1969, **244**, 406.
[950] N. S. Otchin and H. Metzger, *J. Biol. Chem.*, 1971, **246**, 7051.
[951] S. H. Bridges and J. R. Little, *Biochemistry*, 1971, **10**, 2525.
[952] B. A. Askonas and R. M. E. Parkhouse, *Biochem. J.*, 1971, **123**, 629.
[953] R. M. E. Parkhouse, *F.E.B.S. Letters*, 1971, **16**, 71.
[954] F. Melchers, *Biochemistry*, 1971, **10**, 653; Y. S. Choi, P. M. Knopf, and E. S. Lennox, *ibid.*, 1971, **10**, 668.

synthesis of a mouse light chain in a cell-free system from rabbit reticulocytes at the direction of an RNA fraction from a mouse myeloma provides formal proof that the immunoglobulin chain is not formed by a fusion of two distinct polypeptides.[955] More direct evidence that distinct genes control the variable and constant regions comes from a preliminary report [956] that recombination has been detected between markers in the variable region (*a* locus) and constant region (A14/A15 locus) of rabbit γ-chains. Polymorphism for the genes coding for the constant regions of human λ-chains is established.[909, 910] However, opinion remains divided about how many genes are required to code for the variable regions. Germ-line theories have been urged by some [906–908] (see also refs. 887 and 888) but the detailed report of allotypically related sequence changes in the variable region of rabbit heavy chains [921] argues strongly in favour of one or a small number of copies of each allele. It is not unreasonable to hope, *pace* Franklin,[957] that continued study of the primary structure of immunoglobulins and their genetic markers will eventually provide an answer to the riddle.[958]

8 Membrane Proteins

The reader will be aware that this is a rapidly advancing field, and again only the characterization of membrane proteins, not their function, will be the concern of this section. Recent attempts to label and isolate transport systems, drug receptors, and enzymes in membranes have been neatly reviewed,[959] and the distinction has been made (see also last year's Report) between the affinity-labelling approach and characterization of proteins from solubilized whole membrane preparations, the latter being advocated.[959] However, there are ways of avoiding the non-specific labelling held to be a disadvantage of the former method, and its success in certain cases is not denied.[959] A brief survey of the possible modes of organization of protein molecules in membranes comes down in favour of a fluid structure in which protein molecules float in a pool of lipid, rather than the rigidity which the 'structural protein' concept conveys,[960] and chemical modification of whole microsomal membranes has revealed [961] that tryptophan in the proteins is buried and a higher proportion of tyrosine residues and a lower proportion of lysine residues are exposed than in ordinary globular proteins.

A. Solubilization and Fractionation.—It is highly desirable that proteins from membranes should be obtained in their native conformations, and

[955] J. Stavnezer and R. C. C. Huang, *Nature New Biology*, 1971, **230**, 172.
[956] R. G. Mage, G. O. Young-Cooper, and C. Alexander, *Nature New Biology*, 1971, **230**, 63.
[957] 'He that lives upon hope will die fasting', Benjamin Franklin, 'Poor Richard's Almanac'.
[958] 'When you have eliminated the impossible, whatever remains, however improbable, must be the truth', Sir Arthur Conan Doyle, 'The Sign of Four'.
[959] S. I. Chavin, *F.E.B.S. Letters*, 1971, **14**, 269.
[960] N. M. Green, *Biochem. J.*, 1971, **122**, 37P.
[961] A. S. Khandwala and C. B. Kasper, *Biochim. Biophys. Acta*, 1971, **233**, 348.

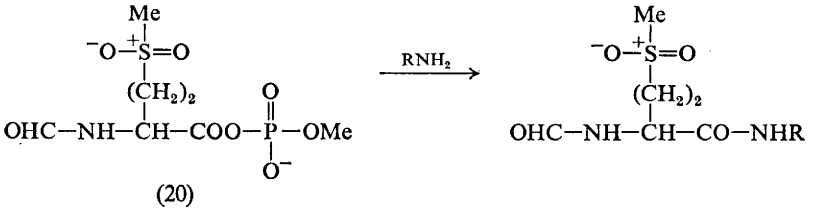
mild methods of extraction are thus needed. Solubilization of most of the proteins of erythrocytes with no evidence of conformational change was achieved with 5 mmol l^{-1} edta alone,[962] suggesting that inorganic cations participate in stabilization of protein–lipid interaction in this system, and 0.1 mol l^{-1} tetramethylammonium bromide (90% solubilization) was accordingly much more effective than 0.1M-NaCl (10% solubilization).[962] Dilute acetic acid is also a useful mild solvent for erythrocyte membrane proteins,[963] but rather less effective than edta. Guanidine hydrochloride is often used to solubilize membranes, although it now appears that, in some cases at least, it may fail since some carbohydrate-rich proteins of erythrocyte membranes remained associated with lipid in 6M guanidine hydrochloride.[964] For the extraction of glycoproteins from membranes, lithium di-iodosalicylate is recommended.[965]

One or two improvements in procedures for fractionating membrane proteins by polyacrylamide-gel electrophoresis have been described. Thus the inclusion of 5% 2-mercaptoethanol in the phenol–acetic acid–urea system gives sharper bands [966] and the three major proteins of rat brain myelin were purified by preparative SDS-gel electrophoresis (10—15 mg of total protein on a gel of dimensions 10.5 × 2.4 cm).[967] This is obviously a method of general application made more attractive by the possibility of renaturation of proteins from solution in SDS,[164, 166] as mentioned earlier (Section 2E). Non-ionic detergents such as Triton X-100 and deoxycholate are frequently used to disrupt membranes and the components are then resolved by gel chromatography in the presence of detergents. In an improved procedure for this step [968] a decreasing (top to bottom) detergent gradient is set up in the column before the sample is applied, and the sample is then eluted with the strongest detergent (necessary initially to keep the sample soluble). The large proteins travel faster than the strong detergent and thus suffer minimal damage. Thus the mitochondrial oligomycin-sensitive ATPase was purified in active form on agarose with a 0—0.3% deoxycholate gradient, whereas irreversible inactivation resulted when 0.3% deoxycholate was used throughout for the chromatography.[968]

B. Red Blood Cell Membranes.—The distribution of proteins on the inside and outside of the erythrocyte membrane has been the subject of many investigations (see also last year's Report). Susceptibility to proteolytic attack or to chemical modification identifies proteins on the outside of the membrane, and the protein bands concerned are easily identified on SDS gels, either by virtue of their radioactivity as a result of modification or

[962] J. A. Reynolds and H. Trayer, *J. Biol. Chem.*, 1971, **246**, 7337.
[963] A. H. Maddy and P. G. Kelly, *Biochim. Biophys. Acta*, 1971, **241**, 290.
[964] A. H. Maddy and P. G. Kelly, *Biochim. Biophys. Acta*, 1971, **241**, 114.
[965] V. T. Marchesi and E. P. Andrews, *Science*, 1971, **174**, 1247.
[966] T. K. Ray and G. V. Marinetti, *Biochim. Biophys. Acta*, 1971, **233**, 787.
[967] T. V. Waehneldt, *Analyt. Biochem.*, 1971, **43**, 306.
[968] P. Swanljung, *Analyt. Biochem.*, 1971, **43**, 382.

through a change of molecular weight as a result of proteolysis. ^{35}S-Labelled formylmethionyl sulphone methyl phosphate (20), which is extremely reactive towards amino-groups (Scheme 6), labels a carbohydrate-free protein (m. wt. 105 000) which is a major component of the



(20)

Scheme 6

external membrane [969] and a glycoprotein. The molecular weight of the latter was initially given as 90 000 from SDS-gels [969] but was later re-estimated [970] as 34 000, since glycoproteins have anomalous mobilities (see below). This glycoprotein contains 70% of the cell-surface sialic acid and is the MN blood-group glycoprotein.[970] Although proteins have not generally been regarded as passing through the lipid bilayer, peptide mapping suggests that both the glycoprotein and the component of molecular weight 105 000 pass right through the membrane.[970, 165] The component of molecular weight 90 000 which can be iodinated specifically on the outside of the membrane with lactoperoxidase (see also last year's Report) is claimed [971] not to be the same as the component of molecular weight 105 000 in the previous study (ref. 969). If this is so then it is worth bearing in mind that the number of external proteins depends to some extent on the method of investigation. In fact, a further study [972] shows that two proteins of molecular weight 90 000 and 105 000 unique to the outer membrane are digested by Pronase, and formation of a stable fragment of molecular weight 65 000 indicates that the proteins are only partly exposed. The anomalous molecular weights of glycoproteins in SDS-gels means that the glycoprotein labelled with [^{35}S]formylmethionyl sulphone methyl phosphate,[970] that labelled with lactoperoxidase [971] (previously [973] not distinguished from the non-glycoprotein on gels), and that with [^{35}S]sulphanilic acid diazonium salt [974] (see also last year's Report) are probably the same protein. In fact, decreased binding of SDS to oligosaccharide side-chains relative to the polypeptide backbone has been conclusively demonstrated [975] and shown not to be corrected by the removal of sialic acid. Glycoproteins

[969] M. S. Bretscher, *J. Mol. Biol.*, 1971, **58**, 775.
[970] M. S. Bretscher, *Nature New Biol.*, 1971, **231**, 229.
[971] D. R. Phillips and M. Morrison, *F.E.B.S. Letters*, 1971, **18**, 95.
[972] D. R. Phillips and M. Morrison, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1103.
[973] D. R. Phillips and M. Morrison, *Biochemistry*, 1971, **10**, 1766.
[974] W. W. Bender, H. Garan, and H. C. Berg, *J. Mol. Biol.*, 1971, **58**, 783.
[975] J. P. Segrest, R. L. Jackson, E. P. Andrews, and V. T. Marchesi, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 390.

thus have anomalously high molecular weights and the most accurate values are obtained in gels of greater than 10% cross-linking, *i.e.* when the sieving effect is greatest. The molecular weight of the major glycoprotein of erythrocytes was estimated as 50 000—55 000 in 12.5% gels (5% gels gave 92 000).[975] An independent study [976] confirmed that the protein labelled from the outside with [^{35}S]sulphanilic acid diazonium salt was the glycoprotein by comparing human and bovine erythrocytes, since the bovine (unlike the human) glycoprotein is clearly resolved from other proteins on gels and its identification as the sole site of radioactive labelling was therefore unambiguous. This comparative approach will probably prove useful in other similar instances.

The effect on various membrane properties when proteins on the outside are digested with Pronase has been studied.[974] All the major proteins and glycoproteins in the erythrocyte membrane have been separated in SDS-gels and the proteins released under various conditions compared.[977] Proteins on the interior of the membrane, neatly studied using inside-out vesicles from erythrocyte ghosts,[978] proved to be decidedly resistant to pronase and only one component was slowly attacked. This had a molecular weight of 89 000 (using reduced proteins as markers in SDS-gels) and had previously been identified [977] as the major component of the erythrocyte membrane (about 30% of the total protein); it was also digested quite rapidly from the outside, again suggesting that there is at least one component which spans the membrane.[978] Other studies with Pronase had also suggested this possibility.[974] Another investigation has also been concerned with the electrophoretic analysis of the proteins of red blood cell membranes [979] and again emphasizes the need to use reduced proteins for calibration in molecular weight determinations; 40% of the proteins were found to have high molecular weights, between 200 000 and 220 000.

It was reported last year that the water-soluble protein spectrin (subunit m. wt. 140 000) represented 20% of the total membrane protein of erythrocytes. Another water-soluble protein, tetkin A, also 20% of the protein, has now been described.[980] It appears to be a dimer with chains of m. wt. *ca.* 220 000 and 240 000 and there is no evidence to suggest that it is a 'precursor' of spectrin. Their relation, if any, to proteins seen as a hollow cylinder and a torus, respectively, in the electron microscope, and released from erythrocyte ghosts at low ionic strength,[981] is unknown.

C. Mitochondrial and Other Membranes.—A lipoprotein in ox-heart mitochondrial membranes that appears to be involved in oxidative

[976] K. L. Carraway, D. Kobylka, and R. B. Triplett, *Biochim. Biophys. Acta*, 1971, **241**, 934.
[977] G. Fairbanks, T. L. Steck, and D. F. H. Wallach, *Biochemistry*, 1971, **10**, 2606.
[978] T. L. Steck, G. Fairbanks, and D. F. H. Wallach, *Biochemistry*, 1971, **10**, 2617.
[979] H. R. Trayer, Y. Nozaki, J. A. Reynolds, and C. Tanford, *J. Biol Chem.*, 1971, **246**, 4485.
[980] M. Clarke, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1063.
[981] J. R. Harris, *Biochem. J.*, 1971, **122**, 38P; *Biochim. Biophys. Acta*, 1971, **229**, 761.

phosphorylation can be specifically labelled *in situ* with [^{14}C]dicyclohexylcarbodi-imide;[982] the molecular weight on gels is as low as 10 000. The ATPase of beef-heart [983] and rat-liver [983, 984] mitochondria has a molecular weight in the region of 360 000—380 000 and a complex subunit structure, suggested as $\alpha_3\beta_3\gamma$ in one case [984] and $\alpha\beta\gamma\delta\varepsilon_6$ in the other,[983b] one of the subunits in this case having a molecular weight of 10 500 daltons and possibly fulfilling the role of the mitochondrial ATPase inhibitor protein.[985] The Na^+–K^+-dependent ATPase from canine renal medulla, obtained in active form by extraction with deoxycholate, probably contains two types of chain,[986] and the ATPase has also been purified in high yield from Lubrol-solubilized brain membranes,[987] the chain molecular weight of the phosphorylated subunit being estimated as 94 000.

Claims that sarcoplasmic reticulum membranes are composed largely of a polypeptide of m. wt. 6000 (see last year's Report) now appear to have been conclusively disproved,[988] a repeat of the 'mini-proteins' story for erythrocyte membranes. Sarcoplasmic reticulum is instead very heterogeneous, the components having molecular weights in the range 30 000—300 000. One of the three major proteins was the calcium-dependent ATPase which had a molecular weight (estimated in SDS-gels) of 106 000 before reduction and two types of chain (m. wt. 20 000 and 60 000) after reduction.[988] In another study [989] the ATPase obtained in active form from membranes solubilized with Triton X-100 had subunit m. wt. 80 000—90 000, and the presence of low-molecular-weight material on the gels was also noted. The active-centre phosphopeptide of the ATPase is being studied [988] and the purified ATPase has been shown to form membranes (possibly one particle thick) with deoxycholate-extracted components, presumably phospholipids.[990] It is thus concluded that the ATPase probably has a structural as well as a functional subunit. Further, it is suggested that the ATPase has all the known properties of the Ca^{2+}-transport system and has a molecular weight of 102 000, with probably one active site per mole.[990] Another study agrees that there are three major protein components of sarcoplasmic reticulum membranes,[991] one of which is the ATPase (m. wt. 100 000) and another of which (m. wt. 62 000) is probably the calcium-binding protein; together

[982] K. J. Cattell, C. R. Lindop, I. G. Knight, and R. B. Beechey, *Biochem. J.*, 1971, **125**, 169.
[983] (*a*) D. O. Lambeth, H. A. Lardy, A. E. Senior, and J. C. Brooks, *F.E.B.S. Letters*, 1971, **17**, 330; (*b*) A. E. Senior and J. C. Brooks, *ibid.*, p. 327.
[984] W. A. Catterall and P. L. Pederson, *J. Biol. Chem.*, 1971, **246**, 4987.
[985] J. C. Brooks and A. E. Senior, *Arch. Biochem. Biophys.*, 1971, **147**, 467.
[986] J. Kyte, *J. Biol. Chem.*, 1971, **246**, 4157.
[987] S. Uesugi, N. C. Dulak, J. F. Dixon, T. D. Hexum, J. L. Dahl, J. F. Perdue, and L. E. Hokin, *J. Biol. Chem.*, 1971, **246**, 531.
[988] A. Martonosi and R. A. Halpin, *Arch. Biochem. Biophys.*, 1971, **144**, 66.
[989] B. H. McFarland and G. Inesi, *Arch. Biochem. Biophys.*, 1971, **145**, 456.
[990] D. H. MacLennan, P. Seeman, G. H. Iles, and C. C. Yip, *J. Biol. Chem.*, 1971, **246**, 2702.
[991] N. Ikemoto, G. M. Bhatnagar, and J. Gergely, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 1510.

these constitute 50% of the total protein. In an independent study an acidic calcium-binding protein isolated by deoxycholate–KCl extraction had m. wt. 44 000 and was monomeric in aqueous solution.[992] Named calsequestrin, it binds 43 mol Ca^{2+} (mol protein)$^{-1}$ and is believed to be hydrophobically bonded to the interior of the sarcoplasmic reticulum, and to be quite distinct from the Ca^{2+}-transport protein.

Proteins from plasma membranes, mitochondrial membranes, and endoplasmic reticulum of rat liver and kidney were compared by gel electrophoresis [993] using the discontinuous sulphate–borate buffer system already mentioned (ref. 148, Section 2E). They were highly heterogeneous and there was evidence of common proteins only where the membranes were functionally related (*e.g.* mitochondrial membranes from liver and kidney). The plasma membranes had three bands in common, one of which (m. wt. 48 000), representing 3—10% of the protein in liver and kidney membranes and also present in erythrocyte membranes, was suggested as being possibly a structural rather than a catalytic component.[993] Plasma membranes from pig lymphocytes retained some biological activity after solubilization of 95% of the membrane proteins with 2% deoxycholate.[994] Studies of the glucagon-sensitive adenyl cyclase in the plasma membranes of rat liver,[233] and the preliminary isolation of the glucagon receptor by affinity chromatography [217] have already been mentioned briefly (Section 2E). The electrophoretic pattern of glycoproteins in cell surface is unique for a particular type of cell although some cells (*e.g.* liver and kidney brush border) appear to have some common subunits.[995] A procedure is given for locating carbohydrate and protein bands in the same gel.[995]

Since the first report [996] of specific labelling of the cholinergic receptor in the electroplax of the electric eel (*Electrophorus electricus*) with α-bungarotoxin (from snake venom, m. wt. 8 000), ^{131}I-labelled bungarotoxin has been used similarly for the receptor in the electric tissue of the torpedo fish (*Torpedo marmorata*);[997] this has one binding site for the toxin per subunit of m. wt. 80 000, and tetramers and other aggregates are formed in 1% Triton. An attempt has been made to purify the α-bungarotoxin-binding component in electric eel electroplax by fractionation on agarose [998] and, rather more effectively, by affinity chromatography on an insolubilized toxin related to α-bungarotoxin [216] as already mentioned (Section 2E). There is every hope that specific binding of α-bungarotoxin will also lead to isolation of the acetylcholine receptor at neuromuscular junctions in higher animals. It has already permitted estimation of the number of cholinergic

[992] D. H. MacLennan and P. T. S. Wong, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1231.
[993] D. M. Neville, jun., and H. Glossmann, *J. Biol. Chem.*, 1971, **246**, 6335.
[994] D. Allan and M. J. Crumpton, *Biochem. J.*, 1971, **123**, 967.
[995] H. Glossmann and D. M. Neville, jun., *J. Biol. Chem.*, 1971, **246**, 6339.
[996] J.-P. Changeux, M. Kasai, and C.-Y. Lee, *Proc. Nat. Acad. Sci. U.S.A.*, 1970, **67**, 1241.
[997] R. Miledi, P. Molinoff, and L. T. Potter, *Nature*, 1971, **229**, 554.
[998] M. A. Raftery, J. Schmidt, D. G. Clark, and R. G. Wolcott, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1622.

receptor sites in frog sartorius and rat diaphragm muscle [999] and in mouse skeletal muscle junctions,[1000] which turn out to be the same as the number of acetylcholine-active centres (estimated with ^{32}P-labelled DFP), a situation that also applies in electric organs of torpedo fish [997] and of electric eels [995] and whose significance is still something of a puzzle. It seems clear, however, that the binding sites are different, *i.e.* the acetylcholine receptor and acetylcholinesterase do not share a common structural element. The acetylcholine receptor in the electroplax has also been affinity labelled *in situ* by reacting its thiol groups with the tritiated quaternary ammonium compound 4-maleimido-α-benzyl-trimethylammonium iodide, an acetylcholine analogue.[1001] There was some difficulty with non-specific labelling (*cf.* ref. 959) which may perhaps be the reason why the number of catalytic sites of acetylcholinesterase in this case turns out to be 4—7-fold greater than the number of cholinergic receptor sites. Another quaternary ammonium compound, 3-bromomethyl-3′-trimethylammonium-methylazobenzene bromide, activates the membrane by binding covalently to the acetylcholine receptor, apparently to thiol groups.[1002] Only the *trans*-isomer is effective. Pure acetylcholinesterase binds covalently two moles of ^{14}C-labelled *NN*-dimethyl-2-phenylaziridinium chloride per 65 000 daltons, one mole of which appears to be bound at the tubocurarine-binding site,[1003] supporting the idea that some part of acetylcholinesterase serves as a natural curare receptor in excitable membranes. It has been known for some time that tubocurarine prevents binding of α-bungarotoxin to the cholinergic receptor.

An enzyme from a bacterial membrane is claimed to be the most highly purified lipoprotein enzyme and the most non-polar protein (58% hydrophobic residues) so far reported, properties which would make it well suited to location in the interior of the membrane. It is the C_{55}-isoprenoid alcohol phosphokinase extracted from the membrane of *Staphylococcus aureus* with butanol and has a single chain (m. wt. 17 000).[1004] A rhodopsin-like protein which contains one mole of retinal per 26 000 daltons has been reported to occur in the purple membrane of *Halobacterium halobium*,[1005] of interest because it has otherwise been found only in the retinae of higher animals. Before it can be said with certainty that some membrane proteins of *Micrococcus lysodeikticus* share common peptide chains,[1006] the possibility that all the proteins studied are contaminated will have to be disproved.

[999] R. Miledi and L. T. Potter, *Nature*, 1971, **233**, 599.
[1000] E. A. Barnard, J. Wieckowski, and T. H. Chiu, *Nature*, 1971, **234**, 207.
[1001] A. Karlin, J. Prives, W. Deal, and M. Winnik, *J. Mol. Biol.*, 1971, **61**, 175.
[1002] E. Bartels, N. H. Wassermann, and B. F. Erlanger, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1820.
[1003] B. Belleau and V. Di Tullio, *Canad. J. Biochem.*, 1971, **49**, 1131.
[1004] H. Sandermann, jun., and J. L. Strominger, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2441.
[1005] D. Oesterhelt and W. Stoeckenius, *Nature New Biol.*, 1971, **233**, 152.
[1006] Y. Fukui, M. S. Nachbar, and M. J. R. Salton, *Biochim. Biophys. Acta*, 1971, **241**, 30.

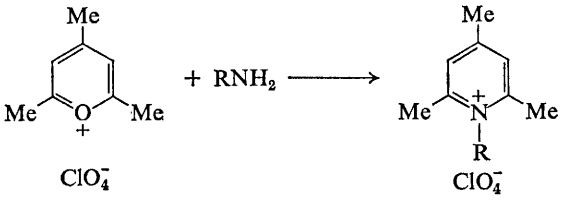
9 Chemical Modification

Numerous examples of chemical modification have already arisen in this chapter where it was more appropriate to discuss them in relation to particular enzymes and proteins. This section will deal mainly with some points of general interest in the use of various reagents for chemical modification. A readable review[1007] surveys the applications of chemical modification to studies of the structure and function of proteins. Modification of the active site is cleanly achieved by affinity-labelling (see Section 9G). Alternatively, the aim may be to map the topography of the molecule by estimating the relative accessibility of various amino-acid side-chains. This is beset with pitfalls, not only because some reagents appear to react faster with buried residues (see below) but because the whole concept of 'buried' and 'exposed' groups in proteins is vague, as illustrated recently by computer analysis of surface residues in the *X*-ray models of some well-characterized proteins.[1008] Further, if residues which are found to be reactive cannot be precisely identified because the amino-acid sequence is not known, the approach is still weaker. In a recent study of carbonic anhydrase a whole battery of group-specific reagents was investigated;[1009] all lysine and arginine residues reacted without inactivation, and there was some evidence of participation of histidine and tryptophan at, or near, the active site and one tyrosine residue that became exposed when zinc was bound.

A. Amino-groups.—In the method of competitive labelling (already mentioned in connection with the pK_a of Ile-16 in chymotrypsin, Section 5A) the competition of particular amino-groups in proteins for small amounts of radioactive acylating agents is measured at various pH values. In this way the ionization constants and reactivities of individual amino-groups can be measured. Thus the amino-terminus of elastase was shown to have pK_a 9.7 and a reactivity much lower than normal.[1010] In this case tritiated (rather than ^{14}C-labelled) acetic anhydride gave the advantage of higher specific activity, and the disadvantage of not being able to detect the peptides on paper by autoradiography was overcome by running markers from a completely [^{14}C]acetylated protein. Pyrylium salts have been suggested for modification of lysine residues[1011] (Scheme 7), the N-substituted pyridinium salt being stable to acid hydrolysis, and quantitative estimation of amino-groups with trinitrobenzenesulphonic acid (TNBS)[12] has already been mentioned (Section 2A). *N*-Hydroxysuccinimide acetate proved to be a highly reactive and selective acylating agent for amino-groups[383] and showed that none of the three amino-groups in insulin was essential for activity. Modification of lysine residues in glutamate dehydrogenase with

[1007] R. B. Freedman, *Quart. Rev.*, 1971, **25**, 431.
[1008] B. Lee and F. M. Richards, *J. Mol. Biol.*, 1971, **55**, 379.
[1009] S. Nees, W. Schmidt, and F. Schneider, *Z. physiol. Chem.*, 1971, **352**, 355.
[1010] H. Kaplan, K. J. Stevenson, and B. S. Hartley, *Biochem. J.*, 1971, **124**, 289.
[1011] M. H. O'Leary and G. A. Samberg, *J. Amer. Chem. Soc.*, 1971, **93**, 3532.

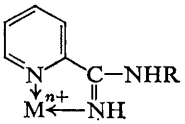
TNBS has been discussed (Section 5C), and the same reagent was used to show that a single lysine residue in pyruvate kinase was involved in ADP binding [1012] and that a single lysine in bovine serum albumin, in the sequence -Leu-Ala-Glu-Lys-Tyr-, was involved in binding fatty-acid anions.[1013]



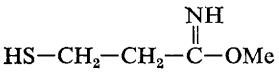
Scheme 7

Reversible blocking of amino-groups by acylation with anhydrides to restrict tryptic cleavage, and an evaluation of various reversible amino-blocking groups for various purposes, has already been mentioned (see Section 2D). The increase in net negative charge of two for each amino-group modified by maleylation caused selective, reversible loss of the elastolytic but not the esterolytic or caseinolytic activity of elastase (total of three lysines);[1014] it was concluded that ε-amino-groups are involved in the secondary binding sites of the enzyme for elastin. After maleylation haemoglobin did not bind haptoglobin,[1015] but this was due to structural changes caused by the negative charges rather than to the importance of amino-groups *per se*, since retention of charge on amidination with ethyl acetimidate was without effect on formation of the haemoglobin–haptoglobin complex. The importance of charge was also shown in another study [12b] in which guanidination or picolinimidation of amino-groups in bovine pancreatic deoxyribonuclease A did not destroy activity, in contrast to introduction of the much smaller carbamyl group which did.

The active picolinimidyl derivative of DNase would seem to be an eminently suitable candidate for the preparation of isomorphous derivatives for *X*-ray crystallography, by chelation of heavy-metal ions, *e.g.* (21). Details have now been published for another reagent (22) designed with the



(21)



(22)

[1012] P. F. Hollenberg, M. Flashner, and M. J. Coon, *J. Biol. Chem.*, 1971, **246**, 946.
[1013] L.-O. Andersson, J. Brandt, and S. Johansson, *Arch. Biochem. Biophys.*, 1971, **146**, 428.
[1014] A. Gertler, *European J. Biochem.*, 1971, **23**, 36.
[1015] W. L. Lockhart and D. B. Smith, *Canad. J. Biochem.*, 1971, **49**, 148.

crystallographer in mind,[1016] the thiol group in this case serving for attachment of heavy atoms. It is hoped that this will prove useful in the *X*-ray diffraction analysis of tobacco mosaic virus, in which there is only one thiol group per protein subunit and into which a single additional heavy-atom site is introduced as a result of the reaction between methyl 3-mercaptopropionimidate (22) and one of the two lysine residues, Lys-68, in the protein subunit of the intact virus. This lysine residue may also be maleylated and the double bond thus introduced should also serve for the attachment of heavy atoms and permit preparation of isomorphous derivatives for *X*-ray crystallography.[1017]

Guanidination of the active-site lysine residue in trypsin inhibitor and cleavage of the homoarginyl bond by trypsin have already been mentioned (Section 5A); so have modification of the amino-groups of cobrotoxin [444] and staphylococcal enterotoxin B [443] (Section 4D). It is worth noting the use of 3,5-dimethyl-1-guanylpyrazole and the nitroguanyl analogue as mild guanidinating agents in the latter case.[443] Cross-linking with dimethyl suberimidate to investigate the quaternary structure of oligomeric proteins has already been discussed (Section 2E), and modification of lysine residues in GPDH with pyridoxal phosphate [632] has been mentioned (Section 5C). Lysozyme modified with one dansyl group can be reversibly denatured, perhaps because the lysine residue to which it is attached is in a hydrophobic environment where a further hydrophobic group does not disrupt refolding.[1018] Reversibility is lost when a further dansyl group is introduced.

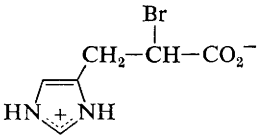
B. Carboxy-groups.—Whereas elastolytic activity was lost selectively when the net negative charge on elastase and subtilisin was increased by maleylation [1014] (see above), it can be conferred upon the slightly alkaline proteinase from *Aspergillus sojae* by modification of the carboxy-groups with a water-soluble carbodi-imide,[1019] thus demonstrating the importance of net positive charge in the adsorption of enzymes on to elastin. However, additional positive charges introduced by coupling with ethylenediamine caused loss of enzymic activity. Similarly, coupling of eight carboxy-groups per subunit of tobacco mosaic virus with ethylenediamine produced monomers, under exceptionally mild conditions,[1020] as a result of the charge change. Modification of an essential carboxy-group in cobrotoxin with a water-soluble carbodi-imide was mentioned earlier (Section 4D), and the role of carboxy-groups in binding thorium ions to thrombin has been investigated.[1021] It is reported that the reactivities of carboxy-groups in proteins to carbodi-imide in 6M guanidine hydrochloride are lower when the proteins have been reduced and carboxymethylated or cyanoethylated.[1022] Formation

[1016] R. N. Perham and J. O. Thomas, *J. Mol. Biol.*, 1971, **62**, 415.
[1017] L. King and R. N. Perham, *Biochemistry*, 1971, **10**, 981.
[1018] N. Okabe and T. Takagi, *Biochim. Biophys. Acta*, 1971, **229**, 484.
[1019] A. Gertler, *F.E.B.S. Letters*, 1971, **19**, 255.
[1020] A. Z. Budzynski and G. E. Means, *Biochim. Biophys. Acta*, 1971, **236**, 767.
[1021] R. W. Colman, *J. Biol. Chem.*, 1971, **246**, 4497.
[1022] R. Frater, *F.E.B.S. Letters*, 1971, **12**, 186.

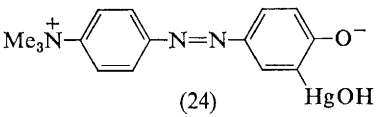
of salt bridges is held to be the cause, and accordingly the effect is less in carboxymethylated proteins, where the additional negative charges help to keep the protein unfolded, than in cyanoethylated samples. Guanidine hydrochloride appears to be not wholly effective in preventing the formation of salt bridges.[1022]

C. Thiol Groups and Disulphide Bridges.—Several reports have dealt with the modification of disulphide bonds in proteins. They can be reduced with dithiothreitol in liquid ammonia and then alkylated with alkyl chlorides to give *S*-methyl derivatives,[19] an alternative to methylation of thiol groups with methyl *p*-nitrobenzenesulphonate in aqueous solution, for which details have now been published [1023] (see also last year's Report). It will be recalled that *S*-methylation was initially suggested as a means of introducing additional points of chemical cleavage into denatured protein chains (see last year's Report); conversion of essential thiols to *S*-methyl derivatives would also be of considerable interest. Analogously, the effect of converting cystine in a disulphide bridge into two alanine residues, by desulphurization of the protein with Raney nickel under mild conditions,[1024] might be useful in elucidating the structural role of the disulphide bridge; desulphurization of peptides for mass spectrometry was described in last year's Report. It occurs much more readily for cysteine and cystine than for methionine, and the extent of desulphurization of proteins depends on the accessibility of the residues concerned.[1024] Conversion of the four disulphide bonds of ribonuclease A into —S—Hg—S— derivatives [701] was mentioned earlier (Section 2E).

DL-α-Bromo-β-(5-imidazolyl)propionic acid (23) acted as a typical thiol reagent with the active thiol of papain, and it is suggested that it might prove useful as an affinity label for the enzymes of histidine metabolism.[1025] 4-(*p*-Trimethylammoniumphenylazo)-2-hydroxymercuriphenol (24) is a new



(23)



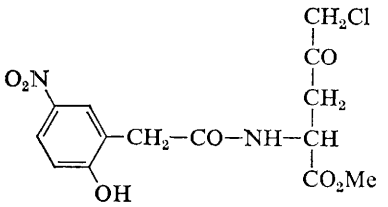
(24)

water-soluble mercurial for the spectrophotometric titration of thiol groups in 8M urea;[1026] addition of edta improves the sensitivity of the method by altering the pK_a of the free reagent. Kinetic analysis of the reaction of *p*-chloromercuribenzoate (PCMB) with the thiol groups of model compounds and of phosphorylase *b* shows [1027] that the reactivity of thiol

[1023] R. L. Heinrikson, *J. Biol. Chem.*, 1971, **246**, 4090.
[1024] M. T. Perlstein, M. Z. Atassi, and S. H. Cheng, *Biochim. Biophys. Acta*, 1971, **236**, 174.
[1025] J. A. Yankeelov, jun., and C. J. Jolley, *Arch. Biochem. Biophys.*, 1971, **147**, 822.
[1026] J. Bloemmen and R. Lontie, *Biochim. Biophys. Acta*, 1971, **236**, 487.
[1027] B. B. Hasinoff, N. B. Madsen, and O. Avramovic-Zikic, *Canad. J. Biochem.*, 1971, **49**, 742.

groups in proteins is determined both by their pK_a values and by their accessibility to PCMB, which is what one might have expected. Other methods for the determination of thiol groups were given in Section 2A.

Modification of one of the three cysteines in NADP-specific isocitrate dehydrogenase from *Azotobacter vinelandii* with a variety of reagents gave almost complete inactivation whereas the thiocyanoalanyl derivative was active, showing that the reactive thiol is not essential for catalysis.[1028] The same observation had already been made for the thiocyanoalanyl derivative of the catalytic subunit of aspartate transcarbamylase.[1029] The thiol groups of formyltetrahydrofolate synthetase[1030] and of carbamyl phosphate synthetase[1031] have been investigated, and the essential thiol groups of creatine kinase can be reversibly protected with DTNB [5,5′-dithiobis-(2-nitrobenzoate)].[1032] Modification of haemoglobin with a series of N-substituted maleimides has already been mentioned (Section 6B) and so has affinity labelling of the acetylcholine receptor with a maleimide derivative of an acetylcholine analogue (Section 8C). Alkylation with *N*-ethylmaleimide was used to explore conformational changes during the catalytic steps of aspartate amino-transferase[1033] (see last year's Report for the use of tetranitromethane for this purpose), leading to the suggestion that two thiol groups are an order of magnitude more reactive in the ketimine complex than in the free enzyme or in any other intermediate. Three halogenomethyl ketones that are useful as reporter groups will alkylate a thiol group at the active site of transglutaminase;[1034] methyl *N*-(2-hydroxy-5-nitrophenylacetyl)-2-amino-4-oxo-5-chloropentanoate (25) is analogous to the natural substrate and is particularly effective.



(25)

D. Tyrosine.—The uncertainty in defining the terms 'buried' and 'exposed' as applied to residues in proteins[1008] mentioned above is borne out in a thorough study of the tyrosine residues in the subtilisins.[1035] There was no clear correlation between the pK values of the tyrosine residues and their

[1028] A. E. Chung, J. S. Franzen, and J. E. Braginski, *Biochemistry*, 1971, **10**, 2872.
[1029] T. C. Vanaman and G. R. Stark, *J. Biol. Chem.*, 1970, **245**, 3565.
[1030] T. Nowak and R. H. Himes, *J. Biol. Chem.*, 1971, **246**, 1285.
[1031] R. Foley, J. Poon, and P. M. Anderson, *Biochemistry*, 1971, **10**, 4562.
[1032] W. J. O'Sullivan, *Internat. J. Protein Res.*, 1971, **3**, 139.
[1033] W. Birchmeier and P. Christen, *F.E.B.S. Letters*, 1971, **18**, 209.
[1034] J. E. Folk and M. Gross, *J. Biol. Chem.*, 1971, **246**, 6683.
[1035] B. Myers, jun., and A. N. Glazer, *J. Biol. Chem.*, 1971, **246**, 412.

reactivity with *N*-acetylimidazole, and it is suggested that care is needed in interpreting results for *N*-acetylimidazole and tetranitromethane (TNM) since both reagents will react preferentially with tyrosine residues in a hydrophobic environment.[1035] This was clearly the case in horse-heart cytochrome *c* when the easily nitratable tyrosines were located in the *X*-ray structure,[775] and caution wisely prevailed in equating reactivity with accessibility in the nitration of carbonic anhydrase B.[1036] The tyrosine nitrated in γ_M macroglobulin also appeared to be in a non-polar environment.[1037] It is worth bearing in mind that spectra of tyrosine derivatives are very sensitive to micro-environment and care is thus needed when extent of reaction is measured spectrophotometrically.[1035] Net loss of tyrosine during nitration, estimated from the final content of (tyrosine + nitrotyrosine), is evidence of side-reactions (see also last year's Report). There have been many indications of cross-linking of proteins, and accordingly biphenyl and triphenyl have now been identified in nitrated proteins and from nitration of tyrosine itself.[1038] It has been pointed out [885] that the free-radical mechanism proposed earlier [1039] for the reaction of TNM with substituted phenoxide ions (Scheme 8) makes cross-linking (in retrospect) seem hardly

$$XPhO^- + C(NO_2)_4 \longrightarrow XPhO^{\bullet} + NO_2^{\bullet} + \bar{C}(NO_2)_3$$

Scheme 8

surprising since the phenoxyl radical will be extremely reactive. Cross-linking does, however, demand the proximity of a second tyrosine residue, and nitration of proteins with TNM can presumably be clear-cut when this is not so. Another investigation using model tyrosine peptides suggested other, non-polymeric, ninhydrin-negative side-products.[1040] Despite side-reactions with TNM, it was possible to correlate loss of α-lactalbumin activity in the lactose synthetase reaction with loss of tyrosine,[1041] and to show that a single tyrosine in pancreatic deoxyribonuclease was essential for activity and was protected from nitration by calcium ions.[1042] In this case extensive side-reaction gave as much as 50% cross-linking at a TNM : DNase ratio of only 5 : 1, and the nitrated monomer was isolated from a polymeric mixture by gel-filtration.[1041] Nitration implicated a tyrosine residue in the NADP-binding site of 6-phosphogluconate dehydrogenase [1043] and in the extended binding site for protein substrates in thrombin.[1044] The tyrosine

[1036] F. Dorner, *J. Biol. Chem.*, 1971, **246**, 5896.
[1037] N. S. Otchin and H. Metzger, *J. Biol. Chem.*, 1971, **246**, 7051.
[1038] J. Williams and J. M. Lowe, *Biochem. J.*, 1971, **121**, 203.
[1039] T. C. Bruice, M. J. Gregory, and S. L. Walters, *J. Amer. Chem. Soc.*, 1968, **90**, 1612.
[1040] N. D. Boyd and D. B. Smith, *Canad. J. Biochem.*, 1971, **49**, 154.
[1041] W. L. Denton and K. E. Ebner, *J. Biol. Chem.*, 1971, **246**, 4053.
[1042] T. E. Hugli and W. H. Stein, *J. Biol. Chem.*, 1971, **246**, 7191.
[1043] M. Rippa, C. Picco, M. Signorini, and S. Pontremoli, *Arch. Biochem. Biophys.*, 1971, **147**, 487.
[1044] R. L. Lunblad and J. H. Harrison, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1344.

residues of cobrotoxin [446] and of human growth hormone [406] were mentioned earlier (Section 4); nitration and acetylation of 40—50% of the tyrosine residues of stem bromelain with loss of activity was consistent with other studies.[1045]

Reduction of nitrotyrosyl to aminotyrosyl residues in proteins is generally achieved with sodium dithionite ($Na_2S_2O_4$) but recently a hint of a complication comes with the detection of small amounts of a derivative tentatively identified as the aminosulphonic derivative of tyrosine when model nitrotyrosine compounds were reduced with dithionite.[1046] Reduction of six nitrated tyrosines in alkaline phosphatase, and acetylation or succinylation of the resultant amino-groups, showed that these tyrosine residues are not involved directly in catalysis, substrate binding, or metal binding,[1047] and in another study aminotyrosine-137 in trypsinogen permitted specific attachment of a dansyl group as a fluorescent probe for monitoring conformational changes on activation [525] (see also Section 5A). The synthesis and spectral properties of two dansylated amines based on 3-aminotyrosine that are models for such fluorescent reporter groups in proteins have been presented.[1048] 3-Aminotyrosine residues can also be generated by reduction of azotyrosine residues with dithionite, and this was done successfully [1049] for Tyr-73 of ribonuclease; the azo-derivative had been formed by affinity labelling with the diazo-derivative of the substrate analogue PUDP [5′-(4-aminophenylphosphoryl)uridine 2′(3′)-phosphate]. Such conversion of tyrosine into aminotyrosine at various sites in enzymes affords an opportunity to study the effects of relatively small changes and introduces possible points of attachment for reporter groups and cross-linking reagents, as done for staphylococcal nuclease some time ago.

Iodination may also occur preferentially at buried tyrosine residues [775] and care is again required. The essential thiol groups in arginine kinase from lobster muscle were reversibly protected with tetrathionate, and then iodination of a single tyrosine was shown to result in inactivation.[1050] The relative reactivities to iodination of the tyrosine residues in the variable region of the Bence-Jones protein Col have now been determined [915a] using the paired-label technique (see below), the constant region having been studied earlier (see last year's Report). The possibility that such differences arise from differences in primary structure around the tyrosines has been eliminated;[1051] common micro-environments have been shown for tyrosine residues in seven Bence-Jones proteins.[915b] (In the paired-label method the 'reactivity' of a tyrosine residue is the ratio of ^{125}I : ^{131}I incorporated when

[1045] K. Goto, N. Takahashi, and T. Murachi, *J. Biochem.* (*Japan*), 1971, **70**, 157.
[1046] J. F. Riordan and M. Sokolovsky, *Biochim. Biophys. Acta*, 1971, **236**, 161.
[1047] P. Christen, B. L. Vallee, and R. T. Simpson, *Biochemistry*, 1971, **10**, 1377.
[1048] R. A. Kenner, *Biochemistry*, 1971, **10**, 545.
[1049] M. Gorecki, M. Wilchek, and A. Patchornik, *Biochim. Biophys. Acta*, 1971, **229**, 590.
[1050] A. Fattoum, R. Kassab, and L.-A. Pradel, *European J. Biochem.*, 1971, **22**, 445.
[1051] Y. Takeda, B.-K. Seon, O. A. Roholt, and D. Pressman, *Biochim. Biophys. Acta*, 1971, **251**, 357.

the protein is iodinated to a low level with $^{125}I_2$ and to a higher level with $^{131}I_2$; the ratio of the labels gives no information about the micro-environment, but this can be deduced from the ratio di-iodotyrosine : mono-iodotyrosine since the introduction of the second iodine atom is favoured by a hydrophobic environment). Other reagents have also been used to investigate the tyrosine and tryptophan residues of Bence-Jones proteins.[1052] Enzymic iodination with lactoperoxidase has been used to label proteins on the outside of erythrocyte membranes (see Section 8 and last year's Report) and more recently to compare the surface proteins of normal and neoplastic lymphocytes,[1053] and to produce labelled hormones of high specific activity for radio-immunoassay.[1054] A simplified preparation of lactoperoxidase from cow's milk has been reported.[1054] Thyroid peroxidase, which also catalyses iodination of proteins, unlike lactoperoxidase exhibits specificity for tyrosine residues in the sequence -Glu-Tyr-, consistent with the amino-acid composition of thyroglobulin, the presumed physiological substrate of thyroid peroxidase.[1055]

E. Tryptophan, Arginine, and Histidine.—Dimethyl (2-hydroxy-5-nitrobenzyl)sulphonium halides, which are water-soluble, represent a promising improvement over 2-hydroxy-5-nitrobenzyl bromide for tryptophan modification [1056] and have already been used on carboxypeptidase [475] (see Section 5A).

There is still no simple reversible chemical modification of arginine residues in proteins, but a method that is potentially useful as a method of reversible restriction of tryptic attack at arginyl bonds has been described.[1057] Tryptic attack is prevented by conversion of arginine into δ-(5-nitro-2-pyrimidyl)ornithine by reaction with nitromalondialdehyde, and susceptibility to tryptic attack is restored by reduction of the ornithyl derivative to the 1,4,5,6-tetrahydropyrimidyl derivative. The sequence has been carried out successfully on model compounds and holds promise for proteins. (Conversion of arginine in peptides into pyrimidylornithine derivatives for mass spectrometry was mentioned in last year's Report.) The reactions of the single arginine residue in insulin [381] and of the arginine residues of the basic pancreatic trypsin inhibitor [551] with phenylglyoxal were mentioned earlier (Sections 4A and 5A); and citraconic anhydride is recommended [1058] for reversible protection of lysine residues in proteins when arginine is being (irreversibly) modified with trimeric butane-2,3-dione. The unfolding caused by citraconylation also ensures that no cross-linking of arginine residues occurs. Modification of arginine with

[1052] T. Azuma, K. Hamaguchi, and S. Migita, *J. Biochem.* (*Japan*), 1971, **69**, 535.
[1053] J. J. Marchalonis, R. E. Cone, and V. Santer, *Biochem. J.*, 1971, **124**, 921.
[1054] J. I. Thorell and B. G. Johansson, *Biochim. Biophys. Acta*, 1971, **251**, 363.
[1055] M. M. Krinsky and J. S. Fruton, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 935.
[1056] W. P. Tucker, J. Wang, and H. R. Horton, *Arch. Biochem. Biophys.*, 1971, **144**, 730.
[1057] A. Signor, G. M. Bonora, L. Biondi, D. Nisata, A. Marzotto, and E. Scoffone, *Biochemistry*, 1971, **10**, 2748.
[1058] J. A. Yankeelov, jun. and D. Acre, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 886.

cyclohexanedione can be carried out in 0.1M triethylamine as solvent,[1059] much milder than the 0.2M-NaOH normally used, and without ill effects on disulphide bonds.

Reaction with diethyl pyrocarbonate showed that there was an essential histidine residue at or near the substrate-binding site in octopine dehydrogenase [1060] and at the active site of arginine oxygenase.[1061] Exhaustive ethoxycarbonylation of histidine peptides for mass spectrometry has already been discussed (Section 2C). 5-Diazonium-1*H*-tetrazole reacts with both tyrosine and histidine residues; in a study of the reaction of the reagent with RNA polymerase [1062] a careful analysis was made of the spectral properties of the azotetrazole derivatives of histidine and tyrosine, and a formula was presented for calculation of the concentration of each present.[1062] The stability of *Im*-Dnp-histidine to acid hydrolysis [46] has been mentioned (Section 2B).

F. Photo-oxidation.—Studies of Rose-Bengal-sensitized photo-oxidation implicated a histidine residue (His-92) at the active centre of ribonuclease T_1 [1063] and at the active centres of neutral subtilopeptidase *amylosaccharitic us* [1064] and thermolysin,[1065] although in the latter case some oxidation of tyrosine and tryptophan also occurred. Photo-oxidation of the 30s ribosomal subunit (*E. coli*) in the presence of Rose Bengal appeared to modify a histidine residue in a single ribosomal protein involved in binding tRNA and poly-U,[332] and loss of activity on photo-oxidation of glucoamylase I in the presence of Methylene Blue appeared to be correlated with loss of tryptophan.[1066]

Further success has been achieved in increasing the specificity of photo-oxidation by anchoring the photosensitizer at a definite point in the molecule (see last year's Report). Thus the catalytic thiol group in papain (Cys-25) was made the point of attachment for a dinitrophenyl group or a fluorescein thiocarbamyl group without conformational changes.[1067] Spectroscopic methods showed that the Dnp group was buried in the active-site cleft, and irradiation with visible light led to specific photo-oxidation of His-159 and Trp-177, consistent with *X*-ray and kinetic evidence. There was no conformational change, suggesting catalytic rather than structural roles for these groups. On the other hand, the fluorescein thiocarbamyl group appeared to be rotating fairly freely around Cys-25, and photo-oxidation in

[1059] A. F. S. A. Habeeb and J. C. Bennett, *Biochim. Biophys. Acta*, 1971, **251**, 181.
[1060] C. Huc, A. Olomucki, Lê-Thi-Lan, D. B. Pho, and N. van Thoai, *European J. Biochem.*, 1971, **21**, 161.
[1061] F. Thomé-Beau, Lê-Thi-Lan, A. Olomucki, and N. van Thoai, *European J. Biochem.*, 1971, **19**, 270.
[1062] F. Zaheer and B. H. Nicholson, *Biochim. Biophys. Acta*, 1971, **251**, 38.
[1063] K. Takahashi, *J. Biochem.* (*Japan*), 1971, **69**, 331.
[1064] D. Tsuru, T. Hirose, and J. Fukumoto, *J. Biochem.* (*Japan*), 1971, **70**, 699.
[1065] T. Abe, K. Takahashi, and T. Ando, *J. Biochem.* (*Japan*), 1971, **69**, 363.
[1066] S. A. Barker, C. J. Gray, and M. E. Jolley, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 654.
[1067] G. Jori, G. Gennari, C. Toniolo, and E. Scoffone, *J. Mol. Biol.*, 1971, **59**, 151.

this case led to oxidation of Trp-26 as well as of the groups mentioned above, and to collapse of the tertiary structure, indicating a critical structural role for Trp-26, which the *X*-ray model shows to be buried. All the modified groups are within 5 Å of Cys-25. The haem group acts as a built-in photosensitizer in the cytochromes, and photo-oxidation of horse-heart ferrocytochrome *c* has now been reported,[1068] the oxidized form having been looked at earlier (see last year's Report). In ferrocytochrome *c*, Tyr-48 and Trp-59 become oxidized as well as His-18 and Met-80, which were oxidized in the ferri-form; presumably the known conformational change on reduction of cytochrome brings the tyrosine and tryptophan residues nearer the haem group. At pH greater than 8.5, photo-oxidation of spinach-leaf or rabbit-muscle (but not rabbit-liver) aldolase leads to loss of the C-terminal tyrosine,[1069] in contrast to loss of histidine which occurs at lower pH (see last year's Report).

Haematoporphyrin binds in the ratio 2 : 1 to lysozyme, and while it is not an active-site-directed photosensitizer it is effective in causing specific photo-oxidation of the protein. Only one of the binding sites for haematoporphyrin appears to contain photo-oxidizable side-chains since Met-12 is the sole residue modified,[1070] and again the fact that this is clearly buried in the *X*-ray model indicates that care is needed with photo-oxidation studies, as with others, in correlating the reactivity of a side-chain with the degree of exposure on the surface of the protein. Some degree of specificity in photo-oxidation can even be achieved with dye photosensitizers free in solution (see above) if paramagnetic ions such as Co^{2+} or Cu^{2+} are used to inhibit the photodynamic action of the dyes.[1071] The protective effect is localized and the method might be a way of achieving fairly specific modification under non-specific conditions. Thus Trp-63 and Trp-123 in lysozyme seemed to be of minor importance for both structure and catalysis.[1071]

G. Affinity Labelling.—Photo-affinity labelling of chymotrypsin has been mentioned in Section 5A, and 'active-site-directed photo-oxidation' in Section 9F. Diazomalonyl derivatives of adenosine 3′,5′-phosphate (cyclic AMP) are potentially useful photo-affinity labels for enzymes and other receptors that bind the nucleotide.[1072] However, the diazoacetyl ester of the 3-hydroxymethyl derivative of NAD^+ bound to yeast alcohol dehydrogenase gave a mixture of products on photoactivation [1073] and it was thought likely that such a non-covalently bound molecule would not

[1068] G. Jori, G. Gennari, M. Folin, and G. Galiazzo, *Biochim. Biophys. Acta*, 1971, **229**, 525.
[1069] L. C. Davis, G. Ribereau-Gayon, and B. L. Horecker, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 416.
[1070] G. Jori, G. Galiazzo, and E. Scoffone, *Experientia*, 1971, **27**, 379.
[1071] G. Jori, G. Gennari, G. Galiazzo, and E. Scoffone, *Biochim. Biophys. Acta*, 1971, **236**, 749.
[1072] D. J. Brunswick and B. S. Cooperman, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1801.
[1073] D. T. Browne, S. S. Hixson, and F. H. Westheimer, *J. Biol. Chem.*, 1971, **246**, 4477.

turn out to be of general use for this reason. Some examples of active-site labelling that have been reported during the past twelve months (a few of which have already been mentioned) are collected in Table 3.

Table 3 *Some recently reported active-site-directed inhibitors*

Protein	*Active-site-directed compound*	*Residue modified*	*Ref.*
Acetylcholine receptor	4-Maleimido-α-benzyltrimethylammonium iodide	Cys	1001
	3-Bromomethyl-3′-trimethylammonium-methylazobenzene bromide	Cys	1002
Adenylosuccinate lyase	Carboxylic–phosphoric mixed anhydride isosteric with AMP (see text)	?	1074
Alcohol dehydrogenase (liver)	Nicotinamide-[5-(bromoacetyl)-4-methylimidazole]-dinucleotide	?	1075
Aldolase (yeast)	Halogenoacetol phosphates	Cys	1076
Aspartate transcarbamylase catalytic subunit (*E. coli*)	Permanganate ion	Cys	1077
Biotin transport system (yeast)	Biotinyl *p*-nitrophenyl ester	?	1078
Carboxypeptidase A_γ	*N*-Bromoacetyl-*N*-methyl-L-Phe	Glu-270 (see text)	473
β-Galactosidase	*N*-Bromoacetyl-β-D-galactopyranosylamine	Met	1079
Glyceraldehyde 3-phosphate dehydrogenase	$I \cdot CH_2 \cdot CO \cdot OR$ $p\text{-}NO_2 \cdot C_6H_4 \cdot O \cdot COR$ $H \cdot CO \cdot R$ } *	Cys-149	634
Glutamate dehydrogenase	Bromoacetyldiethylstilboesterol	?	1080
20β-Hydroxysteroid dehydrogenase	Cortisone 21-iodoacetate	His	1081
Lac repressor	2′-chloro-2′-cyanoethyl-1-thio-β-D-galactopyranoside	?	1082
Luciferase	2-Cyano-6-chlorobenzothiazole	Tyr?	751
Lysine decarboxylase	L-Lysine bromomethyl ketone	?	1083
Pepsin (pig)	1,2-Epoxy-3-(*p*-nitrophenoxy)-propane	Asp or Glu	1084
Pepsin (bovine)	*N*-Diazoacetylnorleucine methyl ester	Asp	584
	1-Diazoacetyl-2-phenylethane	Asp	591
	1,1-Bis(diazoacetyl)-2-phenylethane	Asp + ?	591
	1-Diazoacetyl-1-bromo-2-phenylethane	Asp + ?	591
Ribonuclease (pancreatic)	Pyridoxal phosphate	Lys-7	1085

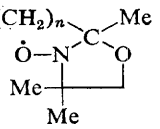
*R = 

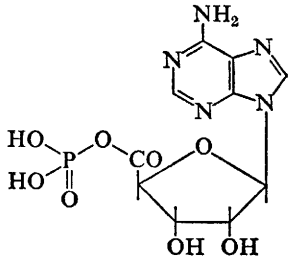
Table 3 (*cont.*)

Protein	*Active-site-directed compound*	*Residue modified*	*Ref.*
Thrombin	*p*-Nitrobenzyloxycarbonyl-arginine chloromethyl ketone	His-43	532
	Tosyl-lysine chloromethyl ketone	His-43	532
Trypsin	*p*-Amidinophenacyl bromide	Ser-183	528
	p-Guanidinophenacyl bromide	Ser-183	528

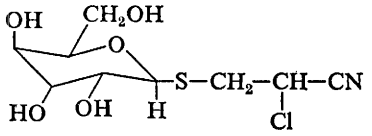
One or two of the examples listed in Table 3 merit further comment. The adenosine 5′-phosphate analogue (26) [1074] and the NAD^+ analogue [1075] could be useful for labelling many enzymes that bind AMP and NAD^+ tightly, and lysine bromomethyl ketone [1088] for the enzymes of lysine metabolism. DL-α-Bromo-β-(5-imidazolyl)propionic acid [1025] as a potential candidate for covalent binding to the enzymes of histidine metabolism has already been mentioned (Section 8C). Permanganate is presumably acting on aspartate transcarbamylase as an analogue of the competitive inhibitor phosphate ion,[1077] and might thus be useful for probing the phosphate binding-sites of other enzymes. It is such a strong oxidizing agent that it shares with carbenes (generated in photo-affinity labelling) the advantage of extreme non-specificity of attack once it has been bound at a specific site. Pyridoxal phosphate is an affinity label for ribonuclease only by virtue of its phosphate group [1085] and it becomes attached to Lys-7 near the active site, not Lys-41 at the active site. The active-site-directed inhibitor for the *lac* repressor,[1082] an analogue of a potent inducer, is a sulphur mustard (27), and the electrophilic carbon atom is substituted with the strongly electron-withdrawing cyanide group as well as with chloride. When pig pepsin was inactivated with the substrate-like epoxide (28),[1084] two moles of reagent became esterified to two carboxy-groups, one of which was protected by substrate and was therefore presumably at, or near, the active site. This was not the aspartyl side-chain that reacts with diazoacetyl-DL-norleucine methyl ester or *p*-bromophenacyl bromide, since two moles of the epoxide became bound even after pretreatment with these inhibitors. It is likely then

[1074] A. Hampton and P. J. Harper, *Arch. Biochem. Biophys.*, 1971, **143**, 340.
[1075] C. Woenckhaus and R. Jeck, *Z. physiol. Chem.*, 1971, **352**, 1417.
[1076] Y. Lin, R. D. Kobes, I. L. Norton, and F. C. Hartman, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 34.
[1077] W. F. Benisek, *J. Biol. Chem.*, 1971, **246**, 3151.
[1078] J. M. Becker, M. Wilchek, and E. Katchalski, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2604.
[1079] J. Yariv, K. J. Wilson, J. Hildesheim, and S. Blumberg, *F.E.B.S. Letters*, 1971, **15**, 24.
[1080] J. Kallos and K. P. Shaw, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 916.
[1081] M. Ganguly and J. C. Warren, *J. Biol. Chem.*, 1971, **246**, 3646.
[1082] R. R. Rando, *Nature New Biol.*, 1971, **234**, 183.
[1083] D. L. Miller and V. W. Rodwell, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 1227.
[1084] J. Tang, *J. Biol. Chem.*, 1971, **246**, 4510.
[1085] G. E. Means and R. E. Feeney, *J. Biol. Chem.*, 1971, **246**, 5532.

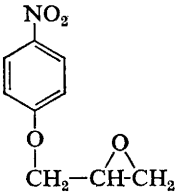
that a second essential carboxy-group reacts with the epoxide, consistent with indications from earlier work of the possible involvement of a second carboxy-group at the active centre of pepsin.



(26)



(27)



(28)

10 Conclusion

The subject grows yearly and it has not been possible to encompass within these pages all that has become known about primary structure and chemical modification of proteins in the year 1971. Much has been omitted, and we would warn our readers that we cannot say of our account what was said of the blessed St. Neot.[1086]

PART II: *X*-Ray Studies *by T. L. Blundell*

1 Introduction

'Protein crystallography' came of age in 1971. To this date there have been forty independent, high-resolution structure analyses (see Table 1), and

[1086] 'That is all, and indeed rather more than all, that is known to men of the blessed St. Neot', J. H. Newman (ed.), 'Lives of the Saints'.

many of these spectacular advances have been chronicled in the report of the Cold Spring Harbor Symposium held in the summer of 1971. However, the workshop on protein crystallography held at Alpbach in March, 1972, showed that this report was soon out of date: four further structures have now been successfully elucidated (see Table 1).

Table 1 *Protein structures solved at* ⩽ 3.5 Å *resolution*

Reported during 1971	*Reported before* 1971
Trypsin [49]	α-Chymotrypsin
Subtilisin Novo [50]	γ-Chymotrypsin
Nuclease [63, 64]	Chymotrypsinogen
Carbonic anhydrase [68]	Elastase
Ferri- and ferro-cytochrome *c* [89, 91]	Subtilisin BPN′
Cytochrome b_5 [93]	Trypsin inhibitor
Flavodoxin [97]	Carboxypeptidase A
High-potential iron protein [95]	Papain
Glycera haemoglobin [109]	Ribonuclease A
Lamprey haemoglobin [109, 110]	Ribonuclease S
Calcium-binding protein [129]	Lysozyme
	Lactate dehydrogenase
Solved but not published at high resolution in 1971	Rubredoxin
	Oxyhaemoglobin
	Deoxyhaemoglobin
Ferredoxin	Sperm-whale myoglobin
Malate dehydrogenase	*Chironomus* haemoglobin (Erythrocruorin)
Thermolysin [59] (recent)	Insulin
Concanavalin	

Independent solution of structures previously solved

α-Chymotrypsin [51]
Ribonuclease A

The publications during 1971 have been of several types. There have been thoughtful reviews of the chemistry and biology in relation to the structures of carboxypeptidase,[1] nuclease,[2] subtilisin,[3] chymotrypsin,[4] elastase,[5] papain,[6] ribonuclease,[7] and insulin.[8] There have also been

[1] J. A. Hartsuck and W. N. Lipscomb, in 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, 1971, vol. III, p. 1; W. N. Lipscomb, *Adv. Protein Chem.*, 1971, **25**, 1.
[2] F. A. Cotton and E. E. Hazen, in 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, 1971, vol. IV, p. 153.
[3] J. Kraut, in 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, 1971, vol. III, p. 547.
[4] D. M. Blow, in 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, 1971, vol. III, p. 185.
[5] J. Drenth, J. N. Jansonius, R. Koekoek, and B. G. Wolthers, in 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, 1971, vol. III, p. 484; *Adv. Protein Chem.*, 1971, **25**, 79.
[6] F. M. Richards and H. W. Wyckoff, in 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, 1971, vol. IV, p. 647.
[7] B. S. Hartley and D. M. Shotton, in 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, 1971, vol. III, p. 323.
[8] T. L. Blundell, E. J. Dodson, G. G. Dodson, D. C. Hodgkin, and M. Vijayan, *Recent Progr. Hormone Res.*, 1971, **26**, 1.

concise accounts of new structures. The emphasis on extracellular proteases appears to have given way to a study of metal-containing proteins. The protein crystallographer has now to be acquainted with inorganic chemistry as well as biochemistry, organic chemistry, physical chemistry, physics, and mathematics! A number of progress reports have been published for problems with subunits: intracellular enzymes of the glycolytic pathway, regulatory enzymes, and virus proteins.

There have been some very exciting reports on peptides of which actinomycin is perhaps the most spectacular, both from the point of view of *X*-ray analytical techniques as well as the interest in the principles of peptide binding to the DNA double helix.

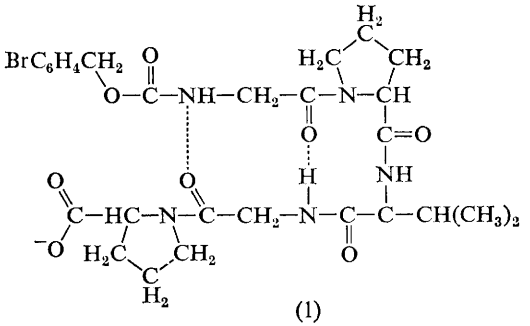
2 Amino-acids and Peptides (See also Chapter 1, Section 3A)

X-Ray studies of amino-acids continue to be reported, but as Hendrickson and Karle [9] comment, they 'hold no surprises'. The crystal and molecular structures of L-cysteic acid monohydrate,[9] L-isoleucine,[10] L-leucine hydriodide,[11] L-tyrosine,[12] an L-histidine complex of molybdenum(v),[13] and bis-(L-prolinato)palladium(II) [14] are described. Preliminary *X*-ray data from powder photographs are given for L-arginine, L-cysteine, DL-lysine, and DL-phenylalanine.[15] The molecular structures of L-tyrosine-*O*-sulphate,[16] a constituent of a number of polypeptide hormones such as gastrin, and of L-arginyl phosphate monohydrate,[17] a model for guanidyl and phosphate interaction such as that found in staphylococcal nuclease, are also reported.

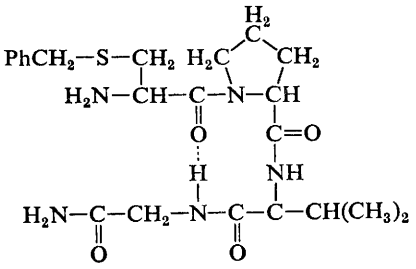
The results of studies on peptides are more interesting. A number of peptides have a hydrogen bond between the carbonyl of a residue and the amide group of the fourth residue further along—giving a ten-membered ring. An example is the structure of bromobenzyloxycarbonylglycyl-L-prolyl-L-leucylglycyl-L-proline ethyl acetate monohydrate,[18] which is a good substrate of the enzyme collagenase. The peptide (1) is folded back at the Pro-2 and Leu residues to give in this case two intramolecular hydrogen bonds at the glycine residues. The conformation has been termed 'U folding' or a 'β turn', and it is characteristic of the 3_{10} helix.

A similar conformation is reported by Rudko, Lovell, and Low in the crystal structure of the C-terminal tetrapeptide of oxytocin, *S*-benzyl-L-cysteinyl-L-prolyl-L-leucylglycinamide (2) [19] and also in the almost isomorphous seleno-analogue. Further, it occurs in a series of new peptide

[9] W. A. Hendrickson and J. Karle, *Acta Cryst.*, 1971, **B27**, 427.
[10] K. Toni and Y. Iitaka, *Acta Cryst.*, 1971, **B27**, 2237.
[11] M. O. Chaney, O. Seely, and L. K. Steinrauf, *Acta Cryst.*, 1971, **B27**, 544.
[12] A. Mostad, H. M. Nissen, and Chr. Rømming, *Tetrahedron Letters*, 1971, 2131.
[13] L. T. J. Delbaere and C. K. Prout, *Chem. Comm.*, 1971, 162.
[14] T. Ito, F. Marumo, and Y. Saito, *Acta Cryst.*, 1971, **B27**, 1062.
[15] B. Khawas, *Acta Cryst.*, 1971, **B27**, 1517.
[16] D. C. Fries and M. Sundaralingam, *Acta Cryst.*, 1971, **B27**, 401.
[17] K. Aoki, K. Nagano, and Y. Iitaka, *Acta Cryst.*, 1971, **B27**, 11.
[18] T. Ueki, S. Bando, T. Ashida, and M. Kakudo, *Acta Cryst.*, 1971, **B27**, 2219.
[19] A. D. Rudko, F. M. Lovell, and B. W. Low, *Nature New Biol.*, 1971, **232**, 18.

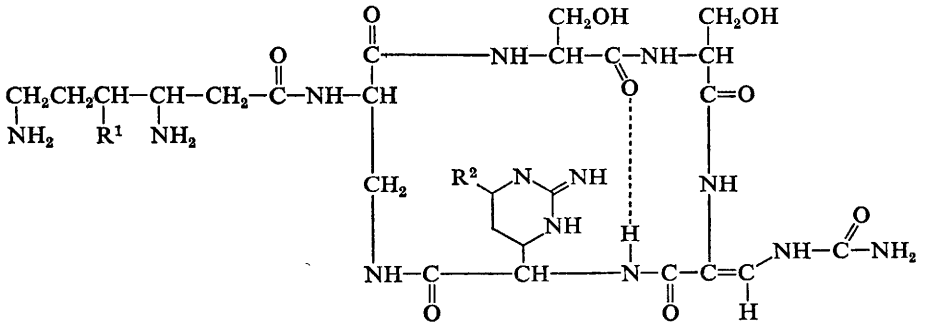


(1)



(2)

antibiotics called tuberactinomycins (TUM) which are effective against tubercular bacilli.[20] The compounds have the structure (3) where R^1 and R^2 can be either H or OH. The 16-membered peptide ring contains the ten-membered ring formed by a hydrogen bond.

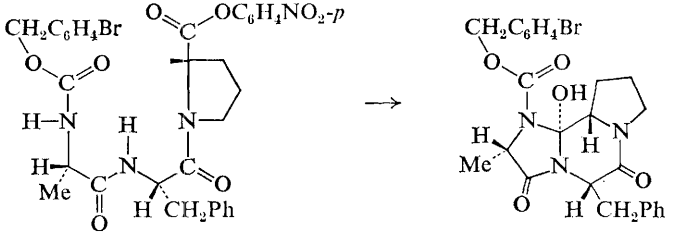


(3)

[20] H. Yoshioka, T. Aoki, H. Goko, K. Nakatsu, T. Noda, H. Sakakibara, T. Take, A. Nagata, J. Abe, T. Wakamiya, T. Shiba, and T. Kaneko, *Tetrahedron Letters*, 1971, 2043.

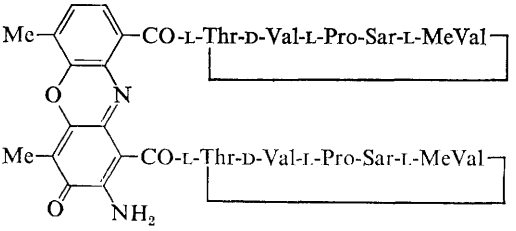
The structures of acetyl-L-proline-*N*-methylamide,[21] LiBr-L-alanyl-glycine,[22] tosyl-L-prolyl-L-hydroxyproline monohydrate,[24] LiBr-glycyl-glycylglycine,[23] and glycyl-L-tryptophanatocopper(II) trihydrate,[25] are reported in detail; the degree of planarity of the peptide groups and the distortion from the regular bond angles are discussed.

The *X*-ray analysis of the product obtained by mild alkaline treatment of the *p*-nitrophenyl ester of *N*-benzyloxycarbonyl-L-alanyl-L-phenylalanyl-L-proline (4) provides an interesting historical footnote.[26] In 1936 Dorothy Wrinch suggested that cyclol structures, formed from intramolecular reactions involving amide groups, may be an important feature of protein architecture. Although the cyclol structure has not been of as general importance as Wrinch expected, evidence is conclusively provided for its existence in structure (5).



(4) (5)

The successful analysis of a 1:2 crystalline complex of actinomycin D and deoxyguanosine has shed light on the stereochemistry of actinomycin binding to DNA.[27, 28] Actinomycin D (6) is a cyclic polypeptide-containing



(6)

[21] T. Matuzaki and Y. Iitaka, *Acta Cryst.*, 1971, **B27**, 507.
[22] J. P. Declercq, R. Meulemans, P. Piret, and M. Van Meerssche, *Acta Cryst.*, 1971, **B27**, 539.
[23] R. Meulemans, P. Piret, and M. Van Meerssche, *Acta Cryst.*, 1971, **B27**, 1187.
[24] M. N. Sabesan and K. Venkatesan, *Acta Cryst.*, 1971, **B27**, 1879.
[25] H. B. Hursthouse, S. A. A. Jayaweera, G. H. W. Milburn, and A. Quick, *Chem. Comm.*, 1971, 207.
[26] S. Cerrini, W. Fedeli, and F. Mazza, *Chem. Comm.*, 1971, 1607.
[27] H. M. Sobell, S. C. Jain, T. D. Sakore, and C. E. Nordman, *Nature*, 1971, **231**, 200.
[28] H. M. Sobell, S. C. Jain, T. D. Sakore, G. Ponticello, C. E. Nordman, *Cold Spring Harbor Symposium*, 1972, **36**, 263.

antibiotic which binds specifically to deoxyguanosine residues of double-helical DNA and so inhibits RNA synthesis. Sobell *et al.* have shown that actinomycin has approximate two-fold symmetry as shown in Figure 1(a).

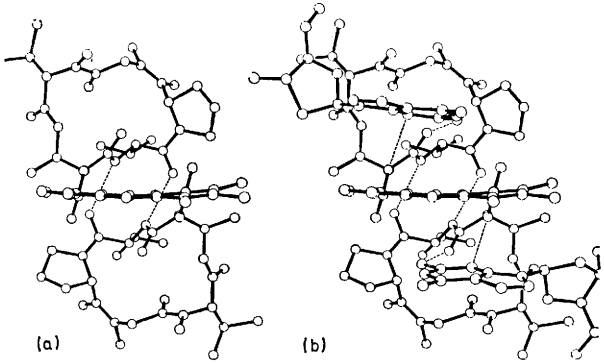


Figure 1 (a) *Computer-drawn illustration of the actinomycin molecule viewed down its approximate two-fold axis;* (b) *same view, but with the deoxyguanosine molecules in place showing two-fold symmetry of the complex*
(Reproduced by permission of Prof. H. Sobell)

The dyad axis lies along a vector connecting the O—N bridging atoms of the phenoxazone ring, and the two polypeptide chains are also approximately related by the dyad. There are strong hydrogen bonds connecting the N—H of one D-valine residue with the C=O of the other D-valine residue. The complex shown in Figure 1(b) also has two-fold symmetry. The two deoxyguanosine residues stack on alternative sides of the phenoxazone ring system and interact with the peptides through two hydrogen bonds. There is a strong bond between the guanosine 2-amino-group and the carbonyl oxygen of the L-threonine residue, and a weaker bond connects the guanine N(3) ring nitrogen with the NH group of the same L-threonine residue. The sugar residues make van der Waals contacts with the isopropyl groups of the L-methylvaline residues.

Sobell *et al.* have shown that these contacts can be incorporated into a model for the binding of actinomycin to DNA so that the phenoxazone ring system intercalates between the base-paired dinucleotide sequence, GpC, while the peptide units lie in the narrow groove of the DNA helix and interact with the deoxyguanosine residues on opposite chains through specific hydrogen bonds. Sobell *et al.* suggest that this mode of binding demonstrates a general principle which several classes of proteins may utilize in recognizing symmetrically arranged sequences on the DNA helix.

3 Methods of Protein Structure Analysis

Although it is often asserted that the techniques of *X*-ray analysis of proteins are worked out, the number of papers published in this area seems as large as ever.

Akervall and Strandberg [29] describe a beautiful technique for growing and mounting protein crystals in a thin-walled capillary with a flattened portion which reduces slippage. Rosenbaum *et al.*[30] report preliminary results using synchrotron radiation as a source particularly for the *X*-ray diffraction studies of fibres. The imidoester methyl 3-mercaptopropionimidate can be used to introduce thiol groups for the attachment of heavy-atom derivatives.[31] A number of interesting theoretical studies and model calculations concerning the method of isomorphous replacement and the use of anomalous scattering are outlined.[32–39]

Now that many protein structures have been determined at high resolution, improving the fit of the protein model to the electron density has become a central problem. Of particular interest is the computer-controlled display system of Barry and North.[40] A number of studies designed to refine protein structures by crystallographic techniques are also discussed. They include real-space refinement,[41] the use of difference Fouriers,[42, 43] the use of the tangent formula,[44] and the inverse Fourier transformation of squared electron density.[45] Perhaps the most spectacular advance has been the use of classical methods of refinement, least squares, and difference Fouriers to improve the structure of the small protein rubredoxin to an agreement value of 0.132. This is the first case in which calculated phases have conclusively been shown to be better than those derived from isomorphous replacement.

In a neutron diffraction study of myoglobin, Schoenborn [46] convincingly shows that this technique provides further structural detail, and in particular can be used to identify the hydrogen-bonding pattern in proteins.

4 Globular Proteins

A. General Structural Principles.—The high-resolution structure analyses reported in 1971 continue to emphasize the variation of architecture

[29] K. Akervall and B. Strandberg, *J. Mol. Biol.*, 1971, **62**, 625.
[30] G. Rosenbaum, K. C. Holmes, and J. Witz, *Nature*, 1971, **230**, 434.
[31] R. N. Perham and J. O. Thomas, *J. Mol. Biol.*, 1971, **62**, 415.
[32] S. Parthasarathy and M. N. Sabesan, *Acta Cryst.*, 1972, **A28**, 51.
[33] S. Parthasarathy, *Acta Cryst.*, 1971, **A27**, 45.
[34] C. W. Bunn, *Acta Cryst.*, 1971, **B27**, 1780.
[35] T. L. Blundell, E. J. Dodson, G. G. Dodson, and M. Vijayan, *Contemp. Phys.*, 1971, **12**, 209.
[36] E. J. Dodson and M. Vijayan, *Acta Cryst.*, 1971, **B27**, 2402.
[37] R. H. Stanford, *Acta Cryst.*, 1971, **B27**, 2036.
[38] W. A. Hendrickson, *Acta Cryst.*, 1971, **B27**, 1474.
[39] W. A. Hendrickson, *Acta Cryst.*, 1971, **B27**, 1472.
[40] C. D. Barry and A. C. T. North, *Cold Spring Harbor Symposium*, 1972, **36**, 577.
[41] R. Diamond, *Acta Cryst.*, 1971, **A27**, 436.
[42] R. Henderson and J. K. Moffat, *Acta Cryst.*, 1971, **B27**, 1414.
[43] K. D. Watenpaugh, L. C. Sieker, J. R. Herriott, and L. H. Jensen, *Cold Spring Harbor Symposium*, 1972, **36**, 359.
[44] C. L. Coulter, *Acta Cryst.*, 1971, **B27**, 1730.
[45] A. N. Barrett and M. Zwick, *Acta Cryst.*, 1971, **A27**, 6.
[46] B. P. Schoenborn, *Cold Spring Harbor Symposium*, 1972, **36**, 569.

between different proteins. Protein structures can contain variable percentages of helix, β-pleated sheet, and irregular non-repeating conformations. The structure of carbonic anhydrase shows that helices may be quite distorted from the geometry of the classical α-helix. The importance of extended β-pleated sheet (which may also be rather distorted) to the structure of protein molecules is further evidenced in carbonic anhydrase, flavodoxin, thermolysin, and nuclease.

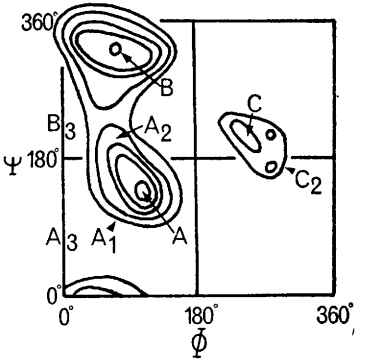


Figure 2 *Empirical peptide energy map for an 'average peptide unit' obtained by calculating the density of experimental values*
(Reproduced by permission from *Nature*, 1971, **234**, 277)

Pohl [47] has suggested using the conformational data of proteins determined at high resolution for the calculation of an empirical protein energy map, which is based on two assumptions: (*a*) that individual influences are averaged out for a large number of different observations, and (*b*) the conformational states of a peptide unit in a collection of proteins correspond to a Boltzmann distribution. The calculated energy maps (Figure 2) show interesting differences from the theoretical maps at A_2 and C_2, where high density indicates that the carbonyls of helices are displaced to a position pointing away from the helix axis, possibly owing to interaction of the π-electrons of a peptide bond with the partial positive charge of the amide proton of the following amide; this also explains low density of A_3 and B_3. Low density at A_1 is due to steric interference with peptides further away in the chain. Conformational data are also presented for glycines and aromatic side-chains.

Although it seems to be generally accepted that hydrophobic forces are important in the structure of proteins, changes of accessibility to solvent on chain folding have not been previously considered quantitatively. Lee and Richards [48] have designed a method of drawing static accessibility contours, and calculating the change of accessibility for different groups of atoms in

[47] F. M. Pohl, *Nature New Biol.*, 1971, **234**, 277.
[48] B. Lee and F. M. Richards, *J. Mol. Biol.*, 1971, **55**, 379.

the protein. They find that the overall accessibility per non-hydrogen atom of different proteins is remarkably constant. The nitrogen and oxygen atoms are 3.5 times as accessible as carbon and sulphur atoms in a native protein. However, the figure for an extended chain is twice. Thus the change in accessibility in going from an extended chain to the folded conformation for these atoms is only about 2. This perhaps sounds a note of caution in discussion on the importance of hydrophobic forces in protein structure.

B. Proteases.—Almost all the *X*-ray studies of proteolytic enzymes reported during 1971 concern the serine proteases. The high-resolution structure of trypsin [49] and of subtilisin Novo,[50] and a second independent study of α-chymotrypsin, are reported.[51] Model-building studies on the α-lytic protease are also described.[52] There have been studies on polypeptide inhibitors with elastase,[53] γ-chymotrypsin,[54] and subtilisin BPN′.[55] Further, the detailed structure of a trypsin inhibitor [56] and a postulated mechanism of binding to trypsin have been described.

Stroud, McKay, and Dickerson [49] report the structure at 2.7 Å resolution of DIP-inhibited bovine trypsin. As yet a detailed description of the electron-density map has not been published; rather, 5 Å models are shown. The analysis is based on four heavy-atom derivatives, only two of which make a significant contribution to the phasing out to 2.7 Å. The average figure of merit is 0.8 at 5 Å but falls to 0.45 at 2.7 Å.

Although only 24% of the residues are invariant between trypsin, chymotrypsin, and elastase, the trypsin analysis shows that this enzyme has similar polypeptide-folding, α-carbon positions, and active-site residues. Changes in the trypsin sequence from that of chymotrypsin are often compensated by further changes which allow the main chain-folding to be preserved. For instance, Met-104 of trypsin replaces a threonine in chymotrypsin, and the methionine side-chain occupies the position of Val-60 side-chain in chymotrypsin. Also, tryptophans at positions 27, 29, and 207 fill a depression in the surface of the enzyme which is occupied by the A-chain of chymotrypsin. The isopropyl group of the inhibitor blocks the active-site binding pocket. Nevertheless, the active site opens on to a

[49] R. M. Stroud, L. M. Kay, and R. E. Dickerson, *Cold Spring Harbor Symposium*, 1972, **36**, 125.
[50] J. Drenth, W. G. J. Hol, J. N. Jansonius, and R. Koekoek, *Cold Spring Harbor Symposium*, 1972, **36**, 107.
[51] R. L. Vandlen and A. Tulinsky, *Fed. Proc. Abs.*, 1971, **30**, 1129.
[52] A. McLachlan and D. M. Shotton, *Nature New Biol.*, 1971, **229**, 202.
[53] D. M. Shotton, A. J. White, and H. C. Watson, *Cold Spring Harbor Symposium*, 1972, **36**, 91.
[54] D. M. Segal, G. H. Cohen, D. R. Davies, J. C. Powers, and P. E. Wilcox, *Cold Spring Harbor Symposium*, 1972, **36**, 85.
[55] J. Kraut, J. D. Robertus, J. J. Birktoft, R. A. Alden, P. E. Wilcox, and J. C. Powers, *Cold Spring Harbor Symposium*, 1972, **36**, 117.
[56] R. Huber, D. Kukla, A. Rühlmann, and W. Steigemann, *Cold Spring Harbor Symposium*, 1972, **36**, 141; A. Rühlmann, H. J. Schramm, D. Kukla, and R. Huber, *Cold Spring Harbor Symposium*, 1972, **36**, 148.

large region of solvent, in the crystal form studied, which makes it potentially useful for the study of substrate analogues.

The active-site binding pocket contains aspartic acid 189, which will be negatively charged at active pH. Model-building studies with arginine and lysine show that this is ideally placed to form a charge interaction with the substrate side-chains, and so explains the specificity of trypsin proteolysis. Asp-189 in the pocket appears to be hydrogen-bonded to Gln-221A, which may help to freeze the conformation of the aspartic acid in an advantageous way.

Stroud *et al.* suggest that the acidic N-terminal hexapeptide Val-$(Asp)_4$-Lys of trypsinogen interacts and is stabilized by a cluster of lysines at 145 or 188A, 15, 222, and 224. The acidic hexapeptide is probably also a calcium-binding site unique to trypsinogen. The close juxtaposition of carboxylates at 71, 77, and 153 on the surface are suggestive of the second calcium-binding site common to both the enzyme and its precursor.

McLachlan and Shotton [52] have proposed a three-dimensional model for α-lytic protease based on the *X*-ray structures of α-chymotrypsin and elastase. They suggest that the differences are chiefly on the surface, where entire loops are missing and a new one is added, as shown in Figure 3. The hydrophobic core thus remains virtually unaltered, so that the two halves of the molecule form an active centre similar to that of elastase.

Drenth and his co-workers [50] have presented a comparison of subtilisin BPN′ and subtilisin Novo. These have the same sequence but have been crystallized under very different conditions:

Subtilisin	*Crystallization conditions*	*Space group*	*Resolution*
BPN′	2.1 M ammonium sulphate, 0.05 M sodium acetate, pH 5.9	*C*2	2.5 Å
Novo	55% (v/v) acetone–water, 0.05 M glycine–NaOH, pH 9.1	$P2_1$	2.8 Å

Although data on the related structures of α-chymotrypsin and chymotrypsinogen, and also ribonuclease A and ribonuclease S, have been published, this represents the first detailed comparison of two such structures in different crystals. Information from this study of the effect of the medium and lattice packing on the conformation is relevant to the question of whether the structure in the crystal is a reflection of the structure of the functioning molecule, either free in solution or attached to a specific surface such as a membrane.

The structures were first related together by comparison of the main-chain and C atoms. The best fit determined by an iterative least-squares method gave a r.m.s. difference of 1.32 Å. By further comparison of the Fouriers and also of each structure with the Fourier of the other, Hol, Drenth, and their co-workers estimate that 0.5 Å differences result from errors in the electron densities, the interpretation, and the model building.

Further errors of 0.8 Å are estimated to result from the determination of the co-ordinates from the model. Thus differences of less than 1.3 Å are not significant.

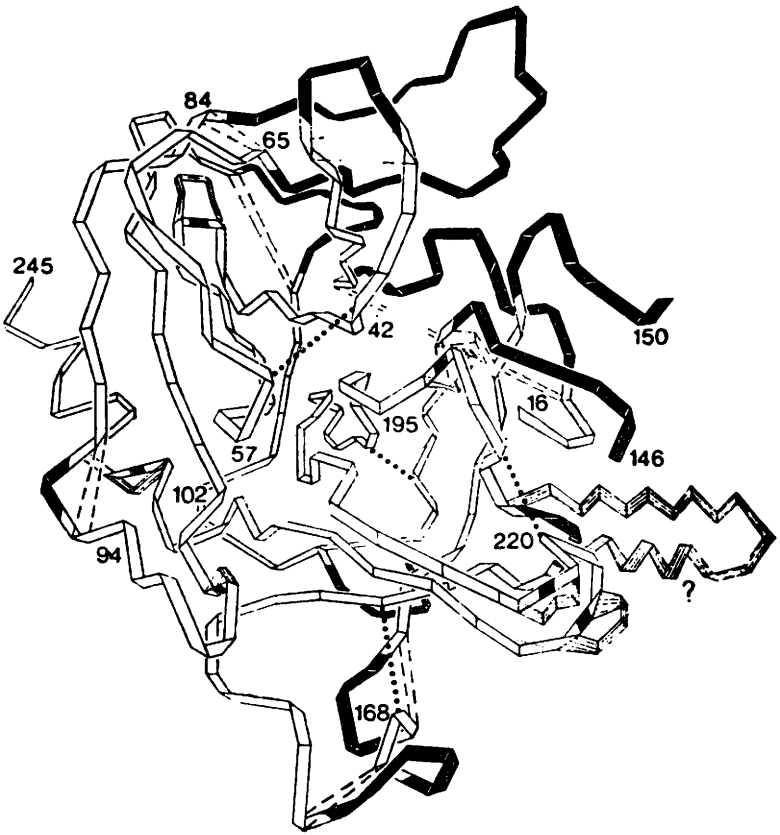


Figure 3 *Ribbon diagram of the polypeptide chain conformation of α-lytic protease drawn on the known skeleton of α-chymotrypsin. Shaded areas are present in chymotrypsin but probably missing in the α-lytic protease; dotted lines are short cuts; striped sections are insertions; beaded lines are disulphide bridges; dark bars are cuts to allow insertions. The active site is in the centre*
(Reproduced by permission from *Nature New Biol.*, 1971, **229**, 202)

The r.m.s. difference in side-group positions—1.92 Å average—is much greater than for backbone positions. Most of the significant differences occur in surface side-chains, many of which correspond to intermolecular contacts. A significant difference of internal side-chains at Val-149 and Val-150 is thought to result from binding of a Tl^+ or K^+ to the nearby Asp-197 in subtilisin Novo, a consequence of the change of the crystallization medium. There are significant differences in only short regions of main chain such as Gly-46—Gly-47 and Pro-129—Ser-130.

The position of the essential Ser-221 is the same in both structures, but the β-C atom of Met-222 is pointing towards solution in Novo and inwards in BPN′ (a difference of 4 Å). However, most active-site residues—for example, Asp-102, Ser-33, Asp-60, or Thr-66—occupy the same positions. In general, Drenth and his co-workers conclude that the difference in medium surrounding subtilisin Novo and subtilisin BPN′ and the different intermolecular contacts in the crystals have not resulted in any major structural change of the subtilisin molecule. Comparative studies of subtilisins BPN′ and Novo are also reported by Kraut *et al.*[57]

A valid criticism of the diffraction results on α-chymotrypsin was that they were obtained at pH 4, well below the p*K* of His-57 involved at the active site. These objections can now be met in the light of the results obtained from studies on the pH dependence of the conformation of the serine proteases.

Vandlen and Tulinsky [51] report a 3.5 Å resolution comparison of the structure of α-chymotrypsin at pH 6.7 with that at pH 3.9, which shows that there has been a conformational change between these pH values. However, a comparison of the structures of native elastase at pH 5 and pH 8.5 shows [53] that in both cases the active-site histidine is hydrogen-bonded to Asp-102 and Ser-195 as found previously in α-chymotrypsin. The only difference appears to be the presence of two bound sulphates at pH 5, one of which lies 6 Å from the side-chains of His-57 and Ser-195. In a similar way, there appears to be no conformational change in the crystals of DIP-inhibited trypsin between pH 5 and pH 9, although there are conformational changes in solution in both trypsin and trypsinogen between these pH values.[49]

Henderson *et al.*[58] have published a concise account of the mechanism of action of α-chymotrypsin as proposed on the basis of *X*-ray studies. They have also studied the role of His-57 in the mechanism of action of α-chymotrypsin. His-57 can be specifically methylated with methyl *p*-nitrobenzenesulphonate. An electron-density difference-map of such a methylated α-chymotrypsin shows that His-57 moves 0.3 Å outwards from its position in native α-chymotrypsin, where it is hydrogen-bonded to the hydroxy-group of Ser-195. The methyl group substituted at $N^{\epsilon 2}$ is clearly the largest feature in the map. The hydrogen bond between $N^{\delta 1}$ and Asp-102 is retained. However, the γ-oxygens of Ser-195 occupy two alternative positions. Although the enzyme still binds substrate analogues, the catalytic rates for substrates are decreased by factors of between 5000 and 20 000. This decrease is thought to result partially from disorientation if the histidine flips so that $N^{\delta 1}$ is used and partially from the absence of the polarizing effect of Asp-102.

[57] J. D. Robertus, R. A. Alden, and J. Kraut, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 334.

[58] R. Henderson, C. S. Wright, G. P. Hess, and D. M. Blow, *Cold Spring Harbor Symposium*, 1972, **36**, 63.

Because of the molecular packing in the crystals of α-chymotrypsin, particularly in the active-site region, attempts to bind larger substrates and substrate analogues than the *N*-formyl-L-tryptophan previously reported have been unsuccessful. However, this is not the case for the crystal forms of the other serine proteases.

Studies of γ-chymotrypsin by Segal *et al.*,[54] and of subtilisin by Kraut *et al.*,[55] with chloromethyl ketone analogues of good phenylalanine polypeptide substrates have indicated further similarities in substrate binding and specificity of these enzymes. It had previously been reported that both enzymes contain a serine at the active site which is acylated by ester substrates, a histidine hydrogen-bonded to this serine which is alkylated by active-site-directed halogenomethyl ketones, and an aspartate which is buried and hydrogen-bonded not only to the active-site histidine but also in each case to a further serine.

The studies now reported show that chloromethyl ketone polypeptide inhibitors bind in an antiparallel β-pleated sheet fashion to a length of extended backbone, Ser-125—Leu-126—Gly-127 in subtilisin, and Ser-214—Trp-215—Gly-216 in γ-chymotrypsin. In each case there are the same geometric relationships of the pleated sheet to the active serine, and glycine residues are involved in β-pleated sheet hydrogen-bonding in both (see Figure 4):

Subtilisin	Ser-125	Leu-126	Gly-127
γ-Chymotrypsin	Ser-214	Trp-215	Gly-216

In the case of γ-chymotrypsin these deductions were made from model building and a comparison of the Ala-Ala-Gly— and Ala-Ala—derivatives of phenylalanine chloromethyl ketone inhibitor complexes, as these involved a rotation of the molecule with respect to the native crystal structure. In both γ-chymotrypsin and subtilisin the subsite S_1 provides a hydrophobic binding pocket for a phenylalanine side-chain; the subsite S_2 is less well defined, while the subsite S_3 can give rise to two specific hydrogen bonds. Stroud *et al.*[49] suggest that trypsin may form a further β-pleated-sheet hydrogen-bond between the polypeptide substrate and the NH of residue 219 (218 is deleted in trypsin). These beautiful experiments show that convergence in evolution between subtilisin and the trypsin family of serine proteases may be quite extensive.

Unfortunately, studies by Shotton *et al.*[53] of elastase with various polypeptide competitive inhibitors indicate a different mode of enzyme substrate interaction. They report difference electron-density maps for complexes of elastase with Ala-Ala-Ala and *N*-acetyl-Ala-Ala at pH 4.2 (at which elastase has negligible activity), with *N*-acetyl-Pro-Ala-Pro-Ala, which is not hydrolysed, at pH 8.5 (the enzyme's optimum pH), and with Lys-Phe, which is the hydrolysis product of the excellent substrate Ala-Ala-Ala-Ala-Lys-Phe, at both pH 5 and pH 8.5. The difference studies of the complexes of Ala-Ala-Ala, *N*-acetyl-Ala-Ala, and *N*-acetyl-Pro-Ala-Pro-Ala all show electron density in the same region, indicating binding of

the C-terminal residue, P_1, beneath Ser-195 with the side-chain methyl extending towards Val-216. P_2 has its side-chain pointing into solution, whereas with P_3 His the side-chain is pointing towards the enzyme surface between the carbonyl of Thr-41 and the β-carbon of Gln-192. However,

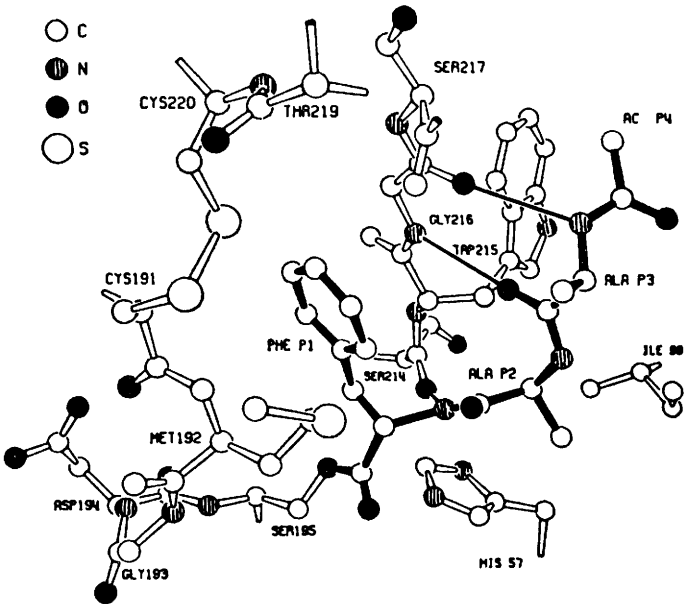


Figure 4 *Atomic model of chymotrypsin A inhibited by acetyl-Ala-Ala-Phe chloromethyl ketone. Only the portion of the enzyme which may interact with the inhibitor has been shown*
(Reproduced by permission from *Cold Spring Harbor Symposium*, 1972, **36**, 85)

the complex of *N*-acetyl-Ala-Ala shows a subsite in the same position as found by Segal *et al.* in γ-chymotrypsin. The difference electron-density maps showed no evidence of binding of the Lys-Phe peptide. Shotton *et al.* concede that the binding found by Segal *et al.* 'seems to be the more favourable', but further studies are clearly required.

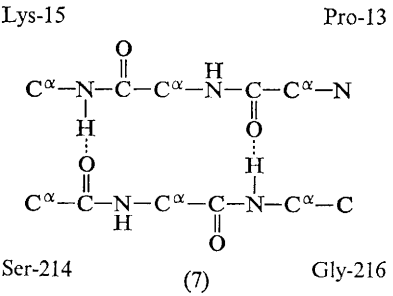
The structure of the basic trypsin inhibitor of bovine pancreas has been studied at 2.5 Å resolution by Huber *et al.*[56] The *X*-ray analysis shows that this small protein of 58 amino-acids is pear-shaped with length of 29 Å and maximum diameter of 19 Å. The major determinant of the structure is a double-stranded anti-parallel β-sheet formed by a loop in the residues between Ala-16 and Gly-36. This β-sheet has a twist of 180°. The remaining polypeptide chain is folded to give one hydrogen bond of a triple-stranded sheet, and three turns of an α-helix at the carboxy-terminus. The structure is stabilized by a hydrophobic core and two further cystine disulphides. There is also a buried asparagine 43 which hydrogen-bonds

to the main chain at positions 5, 7, and 23, so that the segment 38—45 is also tightly held to the rest of the molecule. These interactions lead to a molecule that is stable towards denaturing agents and heat, as well as to almost all proteolytic enzymes.

The molecule has a distinct dipolar character; the binding region is on the positively charged side and involves a major site at Lys-15 and a minor site at Arg-39. The two sites are close together and mutually exclusive. They lie in a symmetrical way about the disulphide bridge, cystine 14—38. If cystine 14—38 is cleaved, the modified inhibitor binds to trypsin in a substrate-like manner and the peptide bonds at Lys-15—Ala-16 and Arg-39—Ala-40 are split. The trypsin inhibitor of soya bean also binds in a similar substrate-like fashion. The lack of peptide cleavage in the native pancreatic trypsin inhibitor is presumably a consequence of stabilization by the β-sheet hydrogen-bonds, the various hydrophobic interactions, and the disulphide 14—38. These would together hold the new C- and N-terminus to the body of the structure, and reduce the potential entropy gain on cleavage.

An homologous inhibitor from bovine colostran has 50% amino-acid changes. Nevertheless, the disulphides, the non-polar core, and the buried asparagine are retained so that a close structural similarity may be assumed. Furthermore, the surface region around the major binding site 12—16 and 35—38 is also retained, although the subsidiary site is different.

The pancreatic trypsin inhibitor binds to trypsin, chymotrypsin, plasmin, and kallikrein, but does not inhibit elastase and subtilisin. Model-building studies of the inhibitor chymotrypsin complex [56] show that the enzyme and inhibitor have highly complementary structures. If Lys-15 (in a non-protonated form) is placed in the specificity pocket with the C and NH of Lys-15 in similar positions to those of tryptophan in the formyl-L-tryptophan–chymotrypsin complex, the residues on the N-terminal side of lysine then form an antiparallel β-structure with the enzyme similar to that proposed for γ-chymotrypsin (7), and there appear to be a number of



(7)

favourable hydrophobic interactions. As a result, the association rate is as high as that of substrates. If peptide cleavage occurs, dissociation—the deacylation steps—would be hindered owing to lack of access of water.

A similar mechanism in soya-bean inhibitor may explain the abnormal deacylation which is often observed. Preliminary *X*-ray data for the crystalline complexes of the pancreatic trypsin inhibitor with chymotrypsin and trypsin have been reported.[58]

Matthews *et al.*[59] have determined the structure of the extracellular proteolytic enzyme thermolysin to 2.3 Å resolution. This enzyme of molecular weight 37 500 contains one zinc and four calcium ions, and is interesting because of its unusual heat stability. The active site contains a zinc atom tetrahedrally co-ordinated to histidine-142, histidine-146, glutamate-166, and a water molecule, and in this way resembles carboxypeptidase A. The precise details of the co-ordination of the calcium ions are not reported. However, it is of interest that two have a centre-to-centre distance of 3.8 Å and are located within an interior region of the protein, surrounded by carbonyl and carboxylate groups. Loss of calcium does not hinder proteolysis at room temperature, but the enzyme is no longer heat stable.

Preliminary *X*-ray data are also reported for the acidic protease from the fungus *Rhizopus chinensis*.[60]

C. Nucleases.—Carlisle and his colleagues (unpublished work) have completed an independent analysis of ribonuclease A.

The data in the literature on the specificity and mechanism of action of ribonuclease have been reviewed by Richards and Wyckoff. The two-step reaction catalysed by ribonuclease [61] is shown in Figure 5, along with a schematic view of the active centre of ribonuclease as deduced from *X*-ray data using the enzyme complex with the substrate analogue, dinucleoside phosphonate, UpcA. This substrate analogue is identical with the substrate, UpA, except that the 5′-oxygen of the adenosine residue has been converted into a methylene carbon atom, so that the P—C bond is not labile in the presence of RNA. Although the *X*-ray work cannot prove conclusively the mechanism of step 2, the same geometry is observed in the simultaneous binding of 3′-CMP and of the activator 3′-AMP, suggesting that His-119 plays a similar role in the steps of a push–pull mechanism.

A crystallographic study of the binding of cupric ion with ribonuclease S has been carried out [62] in order to investigate its inhibitory effects with this enzyme. Allewell and Wyckoff have studied at 6 Å resolution the binding of Cu^{II} to RNase S, RNase E, and 41-Dnp-RNase S in the presence of 3M-$(NH_4)_2SO_4$ and acetate buffer. At pH 5.5 with 0.1M acetate buffer, there are four intramolecular binding sites which are close to His-119, His-105, and Glu-86, and a sulphate ion bound at the active centre on both RNase S and RNase E. There are a further three intermolecular sites. In the 41-Dnp-RNase there are minor changes and the

[59] B. W. Matthews, J. N. Jansonius, P. M. Colman, B. P. Schoenborn, and D. Dupourque, *Nature New Biol.*, 1972, **238**, 37.
[60] I. D. A. Swan, *J. Mol. Biol.*, 1971, **60**, 405.
[61] F. M. Richards, H. W. Wyckoff, W. D. Carlson, N. M. Allewell, B. Lee, and Y. Mitsui, *Cold Spring Harbor Symposium*, 1972, **36**, 35.
[62] N. M. Allewell and H. W. Wyckoff, *J. Biol. Chem.*, 1971, **246**, 4657.

site close to the sulphate is absent. At pH 7 in the presence of 0.67M acetate and 0.3M ethanolamine the binding at all sites is decreased, although the occupancies at the nitrogen ligands decrease less than those with oxygen ligands.

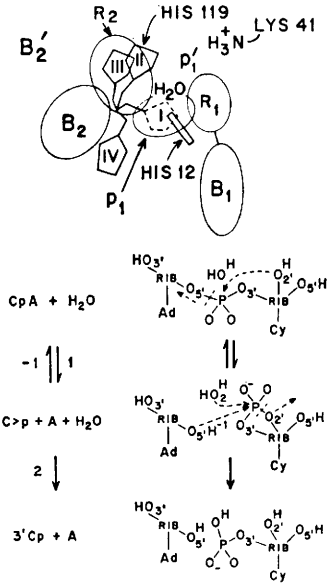


Figure 5 *The diagram at the top is a schematic view of the active centre of ribonuclease as deduced from* X*-ray data from the protein and some substrate complexes.* B_1, R_1, R_2, *and* B_2 *represent the positions of the bases and riboses of a dinucleotide found for UpcA, and in certain other analogues.* B'_2 *is the probable position of the second pyrimidine in dinucleotides such as CpU;* p_1 *is a phosphate position occupied by* SO_4^{2-} *in the crystals. The reaction diagrams are meant to parallel the active-site diagram*

(Reproduced by permission from 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, 1971, vol. IV, p. 647)

Cotton *et al.*[63] report the high-resolution structure of an inhibitor complex of the extracellular nuclease of *Staphylococcus aureus*. The 2 Å resolution electron-density map was based on the analysis of two heavy-atom derivatives: one involves the replacement of the calcium ion for a barium, whereas the other uses 5-iododeoxyuridine 3′,5′-diphosphate in place of thymidine 3′,5′-diphosphate, a replacement of CH_3 by I. The report contains an interesting and useful discussion of the use of anomalous scattering in different stages of the *X*-ray analysis.

The electron-density map allows a complete tracing of the polypeptide chain except for a few residues at the chain termini, which are thought to

[63] A. Arnone, C. J. Bier, F. A. Cotton, V. W. Day, E. E. Hazen, jun., D. C. Richardson, J. S. Richardson, and A. Yonath, *J. Biol. Chem.*, 1971, **246**, 2302.

be disordered as they project into solvent. About 30 of the 149 residues form three separate sections of helix, while 24 residues form a rather irregular, antiparallel pleated sheet. The inhibitor is bound in a large pocket which is predominantly hydrophobic with the exception of the calcium and inhibitor ligands.

The calcium ion is co-ordinated by an approximately square array of carboxylate groups (Glu-43, Asp-21, Asp-40, and Asp-19) with the distance to Asp-19 being somewhat longer than the others.[64] There is also an outer co-ordination sphere comprising a complex array of hydrogen-bonded threonines, water molecules, and other structural elements. The 5′-phosphate of the inhibitor is bound to Arg-35 and Arg-87 and one oxygen lies *ca.* 4 Å from the calcium ion and is probably β-bridged by a hydroxy-group. The 3′-phosphate interacts with Tyr-85 and probably with Lys-84.

Simple esters of thymidine 5′-phosphate or pdTp are hydrolysed with cleavage of the 5′-C—O—P bond between phosphorus and oxygen. It is thought that a second calcium ion binds between the two phosphate groups. This calcium and the two arginines may then serve as general acid catalysts and facilitate the nucleophilic attack of the hydroxy-group (bound to the other calcium) on the 5′-phosphate groups, with expulsion of the 5′-oxygen atom.

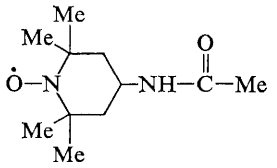
D. Glycoside Hydrolases.—Although comparative biochemical studies have previously cast doubt on the expected homology of human and hen egg-white lysozymes, a 6 Å resolution Fourier of human lysozyme indicates that their tertiary structures are in fact very similar.[65]

Human lysozyme, like hen egg-white lysozyme, has 129 amino-acid residues in its polypeptide chain, but 41% of the amino-acids are different. There is strong evidence that there is in fact one additional residue located between residues 47 and 48 in the hen egg-white chain, in a position on a hair-pin bend in the β-pleat structure which can easily be made to accommodate an extra residue. A compensating deletion occurs at Ser-100, which is also on the surface of the enzyme. As expected, most of the amino-acid changes occur on the surface of the enzyme where the side-chains protrude into solvent. A number of changes do, however, occur in the hydrophobic core, such as the replacement of the methionine residues, but these appear to be accommodated in the tertiary structure. Of the internal residues, 74% are invariant, a smaller percentage than for chymotrypsin and elastase.

The active site including residues Glu-35 and Asp-52 remains substantially unchanged, but there appear to be sequence changes in the residues comprising subsites A, B, and C. The major structural changes which may affect substrate binding appear in subsite A, which is thought to be the weakest in hen egg-white lysozyme.

[64] F. A. Cotton, C. J. Bier, V. W. Day, E. E. Hazen, jun., S. Larsen, *Cold Spring Harbor Symposium*, 1972, **36**, 243.
[65] C. C. F. Blake and I. D. A. Swan, *Nature New Biol.*, 1971, **232**, 12.

The binding of the spin label 4*N*-acetamido-2,2,6,6-tetramethylpiperidine *N*-oxide (8) with tetragonal lysozyme has been studied at 6 Å resolution. The spin label is rather different from the inhibitor, *N*-acetylglucosamine.



(8)

However, difference electron-density maps indicate two binding sites in the cleft close to sugar sites A and C, with interactions through the acetamido-group analogous to the binding of *N*-acetylglucosamine. There is also a stronger third binding site in a hydrophobic pocket near the surface, which does not appear to have biological significance.

Preliminary *X*-ray studies are reported for sweet-potato β-amylase.[67] These indicate a space group of $P4_122$ and a molecular weight of 206 000, which corresponds to a tetramer. At low resolution a pseudocell of the same space group indicates that the tetramer contains an approximate two-fold axis.

E. Carbonic Anhydrase.—Strandberg and his colleagues [68] have calculated a 2 Å resolution electron-density map of human erythrocyte carbonic anhydrase C. A polypeptide chain of 258 amino-acid residues is indicated and two sequenced fragments have been fitted in positions 1—88 and 224—258. The structure has a gross shape of 41 × 41 × 47 Å and is built essentially of three layers. Seven distorted sections of right-handed α-helix are in the surface layers. In contrast, the middle layer is an extensive mainly antiparallel pleated sheet structure with a total twist of 220° and comprising 37% of the total residues. There are aromatic regions between the central sheet and the surface layers.

The active site is a 15 Å deep conical cavity between the pleated sheet and the upper surface layer. The zinc atom is bound to $N^{\varepsilon 2}$ of His-93 and His-95, and $N^{\delta 1}$ of His-117 (part of the pleated sheet). The distorted tetrahedral co-ordination is completed by a water molecule which is also bound to the O^{γ} of Thr-197. The active-site region also indicates two histidines, His-63 and His-128, which lie on opposite sides of the entrance to the cavity. There are eight solvent peaks, not all directly bound to protein functional groups, in the cavity.

[66] L. J. Berliner, *J. Mol. Biol.*, 1971, **61**, 189.
[67] P. M. Colman and B. W. Matthews, *J. Mol. Biol.*, 1971, **60**, 163.
[68] K. K. Kannan, A. Liljas, I. Waara, P.-C. Bergsten, S. Lovgren, B. Strandberg, U. Bengtsson, U. Carlbam, K. Fridborg, L. Jarup, and M. Pelef, *Cold Spring Harbor Symposium*, 1972, **36**, 221.

A series of difference Fouriers have been used to study the replacement of zinc with other metal ions and the nature of interaction of sulphonamides and anions with carbonic anhydrase C. Copper, cobalt, and manganese bivalent ions bind at the zinc site, but mercury is displaced by 0.6 Å from this position. Iodide, cyanide, and cyanate bind to the zinc whereas bromide and chloride bind *via* the zinc-bound water at *ca.* pH 6. 3-Acetoxymercuri-4-aminobenzenesulphonamide binds through a sulphonamide nitrogen or oxygen with a Zn—S distance of 3 Å. The strength of the inhibitor binding appears to result from hydrogen-bonding and van der Waals interactions, as well as from entropic contributions due to displacement of solvent. As HCO_3^- competes with anions it is deduced that this also binds to zinc. Little about the mechanism of action of this enzyme can be deduced at present from these structural studies.

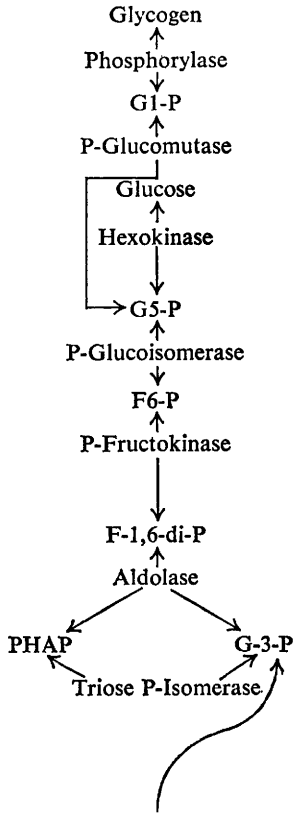
F. Enzymes of the Glycolytic Pathway.—Work on the enzymes of the glycolytic pathway is being carried out in laboratories at Oxford, Bristol, Cambridge, and Yale; several of the studies have led to low-resolution electron-density maps and should give more detailed results in the very near future. These studies have been reviewed by Campbell *et al.*[69] and are summarized in Table 2. Certain of the studies demand more detailed discussion.

A preliminary study [72] of the Patterson projections of hexokinase-B suggests that the molecules are identical and there is a molecular two-fold screw axis parallel to the crystallographic *c*-axis involving translation of 3.6 Å along this local axis.

A 6 Å resolution electron-density map of chicken triose phosphate isomerase (TIM) has been calculated by Banner *et al.*[75] This isomerase catalyses the reaction (*a*) in Scheme 1 and has an interesting advantage over most enzymes because the substrate dihydroxyacetone phosphate (DHAP) can be studied when bound to the crystals. The 6 Å map of the native enzyme shows that the dimer has a local two-fold axis and has dimensions 45 Å parallel to the local axis and 70 × 40 Å in the plane perpendicular to this. The DHAP–TIM complex has a slightly different crystal unit cell as shown in Table 2, but a low-resolution (8 Å) difference map indicates two substrate binding sites related by the local diad axis, although there appear to be conformational changes. Crystals of TIM with a bound transition state intermediate analogue phosphoglycollate (9)

[69] J. W. Campbell, E. Duee, G. Hodgson, W. D Mercer, D. K. Stammers, P. L. Wendell, H. Muirhead, and H. C. Watson, *Cold Spring Harbor Symposium*, 1972, **36**, 165.
[70] F. S. Matthews, *Fed. Proc.*, 1967, **26**, 831.
[71] R. K. Scopes and I. F. Penny, *Biochim. Biophys. Acta*, 1971, **236**, 409.
[72] T. A. Steitz, *J. Mol. Biol.*, 1971, **61**, 695.
[73] P. A. M. Eagles, L. N. Johnson, M. A. Joynson, C. H. McMurray, and H. Gutfreund, *J. Mol. Biol.*, 1969, **43**, 533.
[74] A. I. Gobyunov, N. S. Andreyeva, and V. L. Shpitsberg, *Biofizika*, 1969, **14**, 1116.
[75] D. W. Banner, A. C. Bloomer, G. A. Petsko, D. C. Phillips, C. I. Pogson, *Cold Spring Harbor Symposium*, 1972, **36**, 151.

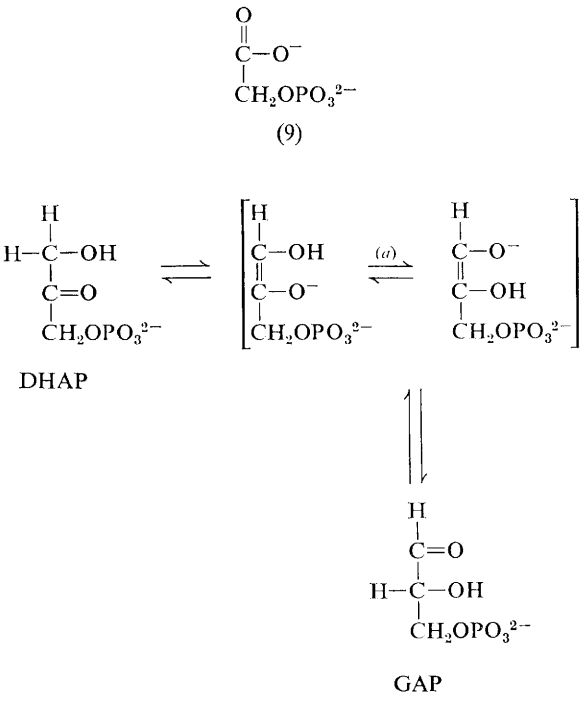
Table 2 *X-Ray studies of enzymes in the glycolytic pathway*



	Form	*Molecular weight*	*Space groups*			*Ref.*
Phosphorylase			$P2_1$ [$P4_12_12_4$ or $P4_32_12$*]			70
P-Glucomutase	Rabbit	63 000	Not suitable for *X*-ray studies			71
Hexokinase	Yeast	102 000	Dimer in asymmetric unit $P2_12_12_1$ a = 215.2 Å, b = 58.5 Å, c = 90.7 Å			72
P-Glucoisomerase	Pig	108 000 Dimer	$P4_12_12$ or $P4_32_12$ $a = b$ = 95 Å, c = 138 Å			69
P-Fructokinase	No mammalian form crystallized					
Aldolase	Rabbit		Monoclinic			73
			Hexagonal			74
Triose P-Isomerase	Chicken	49 000	$P2_12_12_1$			75
		Per dimer	*a*	*b*	*c*	75
	Native		105.9	74.4	61.8 Å	
	+DHAP		102.1	74.4	62.2 Å	
	+phosphoglycollate		99.4	74.4	62.5 Å	

Glyceraldehyde 3-P-dehydrogenase	Lobster	150 000	$P2_12_12_1$	76
	Human	Tetramer	$C2$	77
	Crayfish		$C2$ $a = 128$ Å, $b = 100$ Å, $c = 80$ Å, $\beta = 114.3°$, 2 subunits/asymmetric unit	69
↓				
1.3-di-P-G				
↑				
P-Glycerokinase	Yeast	45 000	$C2$ $a = 126.6$ Å, $b = 54.4$ Å, $c = 98.0$ Å, $\beta = 133.9°$	78
	Horse muscle	48 000 Monomer	$P2_1$ $a = 50.8$ Å, $b = 106.9$ Å, $c = 86.3$ Å, $\beta = 98.6°$	79
↓				
3-P-G				
↑				
P-Glyceromutase	Yeast	110 700 Tetramer	$C2$ $a = 96$ Å, $b = 86$ Å, $c = 81$ Å, $\beta = 120°$, 2 subunits/ asymmetric unit	69
↓				
2-P-G				
↑				
Enolase	Yeast	88 000 Dimer	$P1$ $a = 64$ Å, $b = 74$ Å, $c = 109$ Å, $\alpha = 100°$, $\beta = 92°$, $\gamma = 107°$	69
↓				
P-E-P				
↑				
Pyruvate kinase	Human muscle	240 000	$I222$	69
		Tetramer	$a = 87$ Å, $b = 117$ Å, $c = 129$ Å	
↓				
Pyruvate				

* Eagles *et al.*, unpublished observations.



(9)

DHAP

GAP

Scheme 1

show further changes in cell dimensions as shown in Table 2, which may be a result of a further conformational change.

Crystallographic studies of both yeast and horse-muscle phosphoglycerate kinase (PGK) are progressing well. Watson *et al.*[78] have produced a good mercuric acetate derivative of the yeast enzyme and have recently calculated a 6 Å electron-density map. Blake *et al.*[79] have computed a 6 Å electron-density map of the horse-muscle enzyme which shows that the long polypeptide chain of *ca.* 420 residues is folded into two distinct globular units, which are clearly different in chain conformation. There is no suggestion that the molecule is composed of two polypeptide chains. A single active site has been located on one of the subunits by calculating difference electron-density maps on the isomorphous binary complex PGK—Mg—ADP.

G. Dehydrogenases.—A new map at 2.5 Å resolution of the apoenzyme of dogfish-muscle lactate dehydrogenase is reported by Rossmann and his

[76] H. C. Watson and L. J. Banaszak, *Nature*, 1964, **204**, 918.
[77] A. I. Gobyunov and N. S. Andreyeva, *Molekulyarnaya Biologiya*, 1967, **1**, 261.
[78] H. C. Watson, P. L. Wendell, and R. K. Scopes, *J. Mol. Biol.*, 1971, **57**, 623.
[79] C. C. F. Blake, P. R. Evans, and R. K. Scopes, *Nature New Biol.*, 1972, **235**, 195.

colleagues.[80] Although the gross conformation of the polypeptide chain derived from the 2.8 Å map is confirmed, a number of minor modifications have been made to the interpretation. More secondary structure is now included and the number of amino-acid residues is increased from 311 to 331; this secondary structure is illustrated in Figure 6. The unusual arrangement of the first 20 residues extending away from the rest of the subunit is confirmed. Tetraiodofluorescein binding to LDH has been the subject of a 4 Å resolution crystallographic study.[81] The crystals are isomorphous with the apoenzyme and have one binding site for dye per monomer of the enzyme. The position of the dye is coincident with the adenosine portion of bound NAD^+.

The structure at 5 Å resolution of the abortive ternary complex of lactate dehydrogenase, NAD, and pyruvate resembles that of the apoenzyme in that it has 222 symmetry, although the packing of subunits in the tetramer is different.[82] The main-chain folding is similar in the structures except in one part close to the coenzyme binding site, where the chain stands out in the apoenzyme but moves *ca.* 12 Å to fold over the active site in the ternary complex. An improved electron-density map at 3 Å resolution with phases between 5 Å and 3 Å based on isomorphous and anomalous scattering differences from one mercury derivative confirms these conclusions. The conformation of the coenzyme is also described on the basis of the 3 Å electron-density map. The L-lactate molecule is thought to be oriented under the nicotinamide, with one carboxy-group interacting with the carboxamide of the nicotinamide. His-195 and Asp-168 are close to the essential thiol at Cys-165 and these three groups, lying inside the subunit, may form a charge relay system which enhances the negativity of the histidine, thought to be the 'source and sink' for the proton in the reaction.

The structure at 5 Å resolution of horse-liver alcohol dehydrogenase is described by Branden *et al.*[83] A two-fold crystallographic axis relates to two identical subunits of the dimeric enzyme. The subunit is *ca.* 45 × 55 × 110 Å. The binding site of one zinc atom per subunit is clearly identified in a cleft close to the surface of the enzyme. The position of the second zinc atom is not evident. The active-site cleft is clear but lies about 30 Å from the position of zinc atom found from the electron-density map, and so this zinc atom cannot be involved in biological activity.

The structure of pig heart-muscle malate dehydrogenase determined at 5 Å resolution, and more recently at 3 Å resolution by Tsernoglou *et al.*,

[80] M. G. Rossmann, M. J. Adams, M. Buehner, G. C. Ford, M. L. Hackert, P. J. Lentz, jun., A. McPherson, jun., R. W. Schevitz, and I. E. Smiley, *Cold Spring Harbor Symposium*, 1972, **36**, 179.

[81] P. M. Wassarman and P. J. Lentz, jun., *J. Mol. Biol.*, 1971, **60**, 509.

[82] I. E. Smiley, R. Koekoek, M. J. Adams, and M. G. Rossmann, *J. Mol. Biol.*, 1971, **55**, 467.

[83] C.-I. Bränden, E. Zeppezauer, B.-O. Söderberg, R. Boiwe, B. Nordström, G. Söderlund, M. Zeppezauer, P.-E. Warner, and A. Akeson, in 'Wenner-Gren Symposium on Structure and Function of Oxidation–Reduction Enzymes', ed. A. Akeson and A. Ehrenborg, in the press.

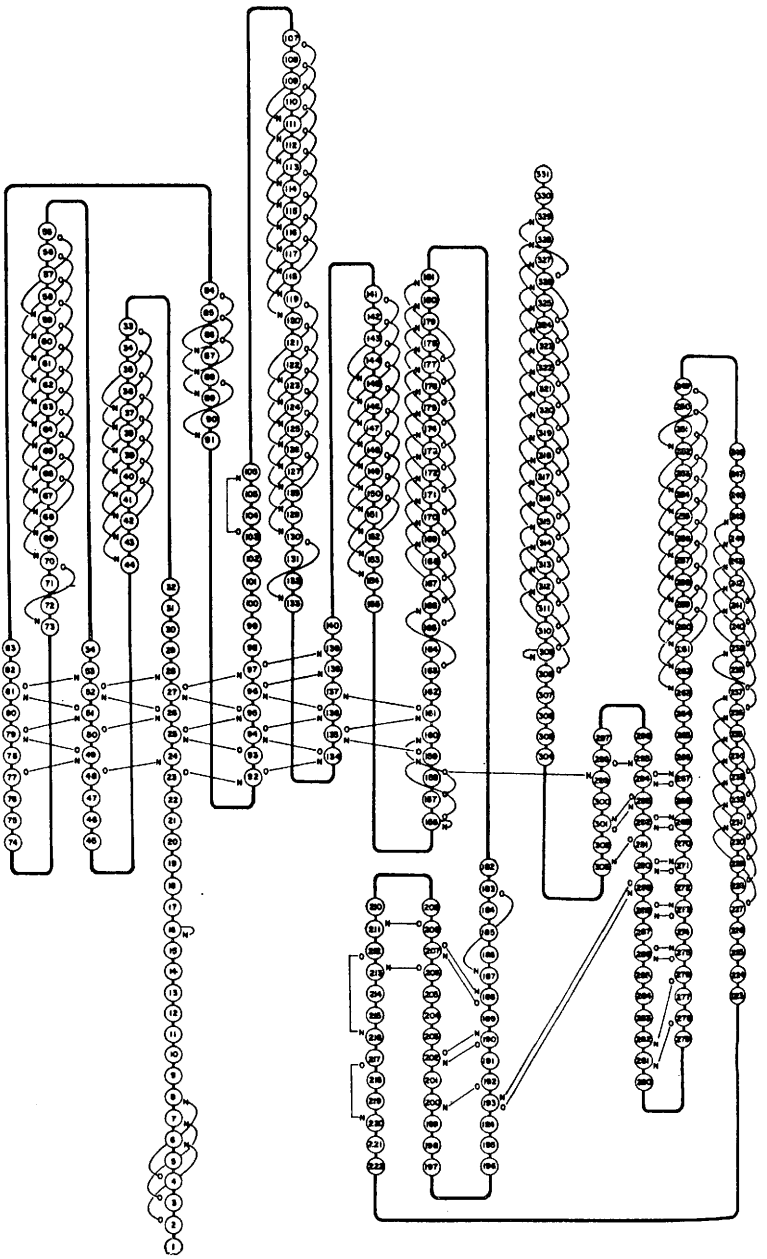


Figure 6 *Schematic diagram of hydrogen bonds between main-chain atoms within one subunit of LDH*
(Reproduced by permission from *Cold Spring Harbor Symposium*, 1972, **36**, 179)

shows remarkable resemblances to lactate dehydrogenase.[84] The enzyme is a dimer, and this forms the crystallographic asymmetric unit. The two molecules of the dimer appear to have slightly different conformations, but are related by an approximate dyad axis at an angle of *ca.* 10—25° to the *b* axis. Only one of the two molecules has bound NAD^+. The enzyme has the same chain-folding as lactate dehydrogenase, but lacks the N-terminal arm, which assists the formation of tetramers from dimers in the latter case. The molecular weight of the malate dehydrogenase dimer is 70 000 and this represents the largest unit determined at high resolution by *X*-ray analysis to date.

However, an even more ambitious project is the study of lipoyl transsuccinylase, the core of the α-ketoglutarate dehydrogenase complex from *E. coli*; preliminary *X*-ray studies are reported by DeRosier *et al.*[85, 86] This core of the complex contains 24 identical subunits and can be crystallized in a face-centred cubic lattice, *F*432. Optical diffraction studies of the electron micrographs of these crystals give structural detail to 40 Å resolution, and indicate that the core complex has octahedral symmetry with the 24 globular units arranged near the vertices of a truncated cube. The information from the low-resolution electron-microscope images and from the subunit composition have been combined with the *X*-ray diffraction data to 18 Å resolution, in order to estimate phases and calculate an electron-density map. This shows 24 peaks at a radius of *ca.* 64 Å situated at the vertices of the truncated cube, and an indication of trimer clustering at a radius of *ca.* 42 Å. Studies with heavy-atom derivatives are now in progress.

H. Regulatory Enzymes.—Studies of two multi-subunit regulatory enzymes, glutamine synthetase [87] and aspartate transcarbamylase [88] are reported.

Glutamine synthetase catalyses the condensation of glutamate with ammonia to form glutamine, which serves as a nitrogen donor to various metabolites such as histidine, tryptophan, AMP, CTP, carbonyl phosphate, and glucosamine 6-phosphate. The enzyme is also regulated by changes in Mg^{2+} and Mn^{2+} concentrations. The enzyme contains 12 identical subunits of molecular weight 48 500, each of which has binding sites for the substrates and effectors. Crystals of 0.3 × 0.2 × 0.1 mm have been grown using the modified micro-diffusion-cell method of Weber and Goodkin in Eisenberg's laboratory.[87] *X*-Ray patterns to 7 Å resolution,

[84] D. Tsernoglou, E. Hill, and L. J. Banaszak, *Cold Spring Harbor Symposium*, 1972, **36**, 171.
[85] D. J. DeRosier, R. M. Oliver, and L. J. Reed, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1135.
[86] D. J. DeRosier and R. M. Oliver, *Cold Spring Harbor Symposium*, 1972, **36**, 199.
[87] D. Eisenberg, E. G. Heidner, P. Goodkin, M. N. Dastoor, B. H. Weber, F. Wedler, and J. D. Bell, *Cold Spring Harbor Symposium*, 1972, **36**, 291.
[88] D. C. Wiley, D. R. Evans, S. G. Warren, C. H. McMurray, B. F. P. Edwards, W. A. Franks, and W. N. Lipscomb, *Cold Spring Harbor Symposium*, 1972, **36**, 285.
[89] R. E. Dickerson, T. Takano, D. Eisenberg, O. B. Kallai, L. Samson, A. Cooper, and E. Margoliash, *J. Biol. Chem.*, 1971, **246**, 1511.

combined with knowledge of the arrangement of the subunits found from electron microscopy, indicate that the space group is probably *P*42. The results are consistent with a model based on a double hexagonal ring pattern 'like two benzene rings with their faces together', which may have either symmetry 6 or 622, but exclude a model based on 32 symmetry (see Figure 7).

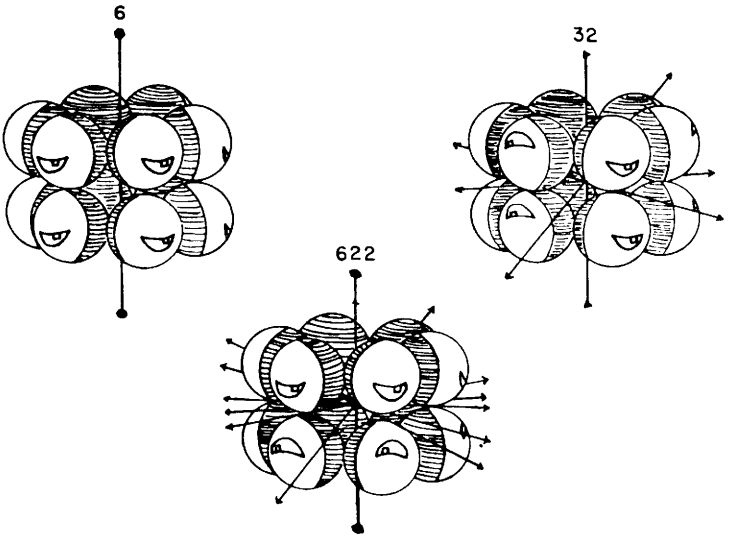


Figure 7 *Symmetries of the glutamine synthetase molecule consistent with molecular shapes seen in the electron micrographs. Symmetry* 32 *is inconsistent with the* X*-ray results*
(Reproduced by permission from *Cold Spring Harbor Symposium*, 1972, **36**, 285)

Studies of the regulatory enzyme aspartate transcarbamylase are further advanced.[88] This enzyme catalyses the first unique reaction in the pyrimidine biosynthetic pathway. The *X*-ray studies of Wiley *et al.*[88] show that this enzyme of molecular weight 310 000 containing six catalytic and six regulatory chains crystallizes in two forms with space groups *P*321 and *R*32. A 5.5 Å resolution electron-density map based on two derivatives has been computed for the *R*32 form; the molecule has 32 symmetry. A clear molecular boundary shows that the molecule has a triangular shape of side 105 ± 10 Å and is 92 ± 10 Å long. The centre of the molecule is empty and is ringed with areas of electron density. The dense areas connect across two-fold axes to form dimers, probably involving contacts between regulatory chains. Further from the centre the contacts, probably between catalytic chains, are trimeric. The active-site region located by a mercury-containing inhibitor appears to be in the plane of maximum contact within the trimer.

I. Redox Proteins.—The study of redox proteins by *X*-ray analysis has been very productive during the year. High-resolution structures of ferri- and ferro-cytochrome *c*, cytochrome b_5, ferredoxin, rubredoxin, high-potential iron protein, and two flavodoxins have now been worked out, although not all were published in 1971.

A series of exciting papers on cytochrome *c* by Dickerson *et al.*[89–92] includes a report on the structures at 2.8 Å resolution of horse and bonito ferricytochromes,[89] a speculative discussion on chain flexibility in ferricytochrome,[90] a report of tuna ferrocytochrome at 2.45 Å resolution [91] and a discussion of sequence and structure homologies in bacterial and mammalian-type cytochromes.[92]

The *X*-ray analysis of ferricytochrome [89] shows that the haem sits in a crevice with thioester bonds to Cys-14 and Cys-17. The iron is bound to His-18 on the right wall and to Met-80 on the left wall as viewed in Figure 8. One of the propionic side-chains is buried at the bottom of the crevice and is extensively hydrogen-bonded, while the other is on the surface. The molecule is constructed in two halves with residues 1—47 on the right side and residues 48—91 on the left. The C-terminal residues 92—104 are wrapped across the right half like a 'strap of a suitcase'.

The peptide is mainly an extended chain which is wrapped round the haem—one layer thick—with hydrophobic residues on the inside and hydrophilic on the outside. The N- and C-terminal residues form short helices. There are two channels leading to the haem; the left channel is lined with residues 52—74 and includes Tyr-74, Trp-59, and Tyr-67, which lie parallel to each other. The right channel is lined with residues 6—20 which are also mainly hydrophobic. The basic groups—including 19 lysines—are segregated into two groups on the surface around the two channels, and there is an acidic surface region between these.

X-Ray analysis of ferrocytochrome [91] shows that this has a more closely packed structure, consistent with its increased resistance to proteolytic digestion, heat denaturation, denaturation at air–water interfaces, and the difficulty of replacement of Met-80 with cyanide or azide. The short C- and N-terminal helices and the peptide attached to the haem (12—18) have the same structure as in the oxidized form (see Figure 8). Nevertheless, there is a dramatic change in residues 77 through 83 on the left side; Phe-82 swings into the channel to block it while Ile-81 swings more to the surface. There is a smaller motion in residues 19 through 25, which results in the blocking of the right channel. Possible mechanisms of electron transfer have been considered: they involve the binding of the cytochrome oxidase to the right-hand side and of the reductase to the left-hand side—each with basic residues. It is postulated that binding of the reductase

[90] R. E. Dickerson, T. Takano, O. N. Kallai, and L. Samson, in ref. 83.
[91] T. Takano, R. Swanson, O. B. Kallai, and R. E. Dickerson, *Cold Spring Harbor Symposium*', 1972, **36**, 397.
[92] R. E. Dickerson, *J. Mol. Biol.*, 1971, **57**, 1.

constrains Phe-82 and Ile-81 to the oxidized structure. An electron is then transferred by way of overlap of aromatic π-electron clouds of Tyr-74 and Tyr-67 to Met-80 and thence to the Fe^{III}. This electron transfer may break a hydrogen bond of Tyr-67 with Thr-78, so that the conformational change to the reduced form is triggered.

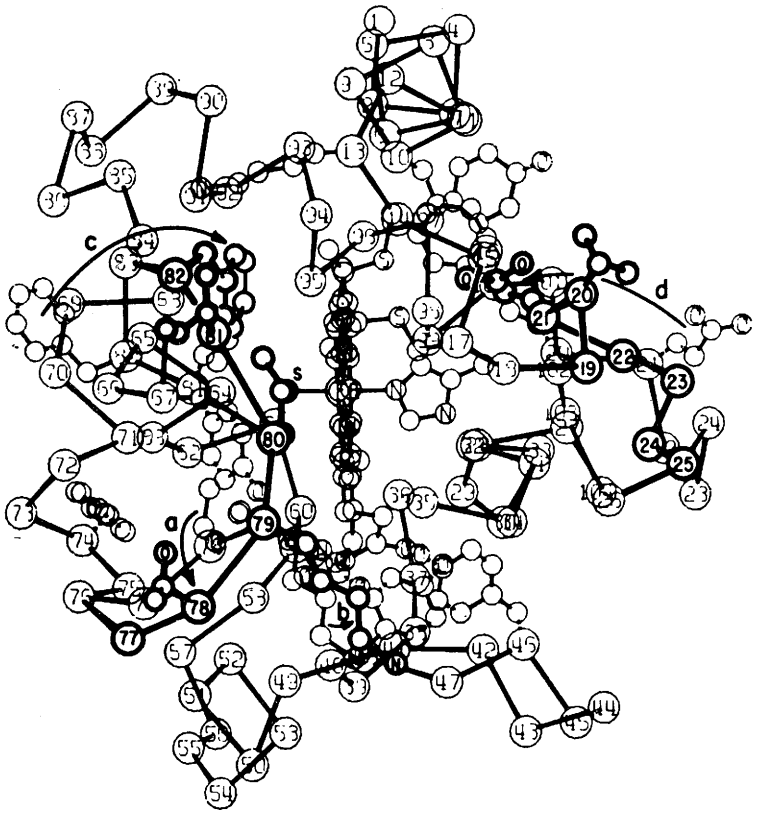


Figure 8 *The altered conformation of residues* 19—25 *and* 77—82 *in reduced cytochrome* c, *superimposed on the oxidized structure*
(Reproduced by permission from the *Cold Spring Harbor Symposium*, 1972, **36**, 397)

Studies of the sequence changes in cytochrome *c* found in 30 species ranging from man through moth to castor bean show that the hydrophobic nature of the haem box is conserved. A number of residues on the left side including the aromatics Trp-59, Tyr-67, and Tyr-74 are conserved, which is consistent with the importance of this part of the molecule to electron transfer. The character of the basic region is also conserved, which explains the observation that cytochrome *c* from one species will react *in vitro* with cytochrome–oxidase preparations from distantly related species.

Model-building studies on bacterial cytochromes [92] show that the sequence of cytochrome c_2 from *Rhodospirillum rubrum*, with 114 amino-acids, can be easily accommodated in a similar tertiary structure. The cytochrome c_{551} from *Pseudomonas aeruginosa* containing 82 amino-acids can be fitted only with some difficulty, involving deletions in positions 60—80 as postulated previously. This results in the exposure of one side of the haem. However, a model with deletion of positions 30 through 47 is possible, with much smaller structural modifications.

The structure of calf-liver cytochrome b_5 has been determined [93] at 2.8 Å, and more recently at 2 Å resolution.[94] This protein has a molecular weight of 11 000 and contains 93 amino-acids. It interacts specifically with a flavoprotein, cytochrome b_5 reductase, which catalyses the transfer of electrons from NADH to the haem iron of the cytochrome.

The conformation of the polypeptide chain is shown in Figure 9. There are five short helical regions and four short segments of extended chain which form a pleated sheet structure. The haem group lies in a hydrophobic crevice, which comprises two antiparallel helical segments forming the walls and the pleated sheet structure forming the floor. The haem iron is co-ordinated by histidines 39 and 63. The haem group is oriented so that the vinyl groups are deeply buried while the propionic acid groups are in an aqueous environment, one being bound to the surface of the molecule while the other is free. The arrangement of the haem relative to the polypeptide thus more closely resembles that in myoglobin rather than in cytochrome *c*, where one of the propionic groups is buried. The surface of the molecule is polar and the protein is soluble. It is now thought that cytochrome b_5 has a further 40 amino-acid residues which are cleaved off when it is solubilized; these residues are involved in binding to the membrane.

A preliminary account of the structure, at 2.25 Å resolution, of the high-potential iron–sulphur protein (HiPIP) of *Chromatium vinosum* is reported by Kraut and co-workers.[95] As a result of oxidation by the *X*-ray beam during data collection, the reported structure is a hybrid of unknown proportions of the reduced and oxidized states. The molecule, which comprises 86 amino-acid residues, is a prolate ellipsoid of 35 × 20 Å built in two halves comprising residues 1—42 and 47—86. The inorganic cluster of four iron and four sulphur atoms lies in a hydrophobic environment between these halves and is bound to cysteines 43, 46, 63, and 77. One turn of 3_{10} helix connects Cys-43 and Cys-46.

The geometry of the iron–sulphur cluster derived from the electron-density map is confirmed by comparison with a $(F_o - F_c)$ difference Fourier (where F_c is calculated from the atomic positions excluding the inorganic and cysteine sulphur atoms) and a Bijvoet difference Fourier. The iron

[93] F. S. Mathews, M. Levine, and P. Argos, *Nature New Biol.*, 1971, **233**, 16.
[94] F. S. Mathews, P. Argos, and M. Levine, *Cold Spring Harbor Symposium*, 1972, **36**, 387.
[95] C. W. Carter, S. T. Freer, Ng. H. Xuong, R. A. Alden, and J. Kraut, *Cold Spring Harbor Symposium*, 1972, **36**, 381.

atoms are arranged in a tetrahedron with an inorganic sulphur atom bound above each triangular face to three iron atoms; the bond lengths given in Table 3 indicate the presence of covalent iron–iron interactions. At this stage there is no evidence for a distortion from tetrahedral symmetry, as

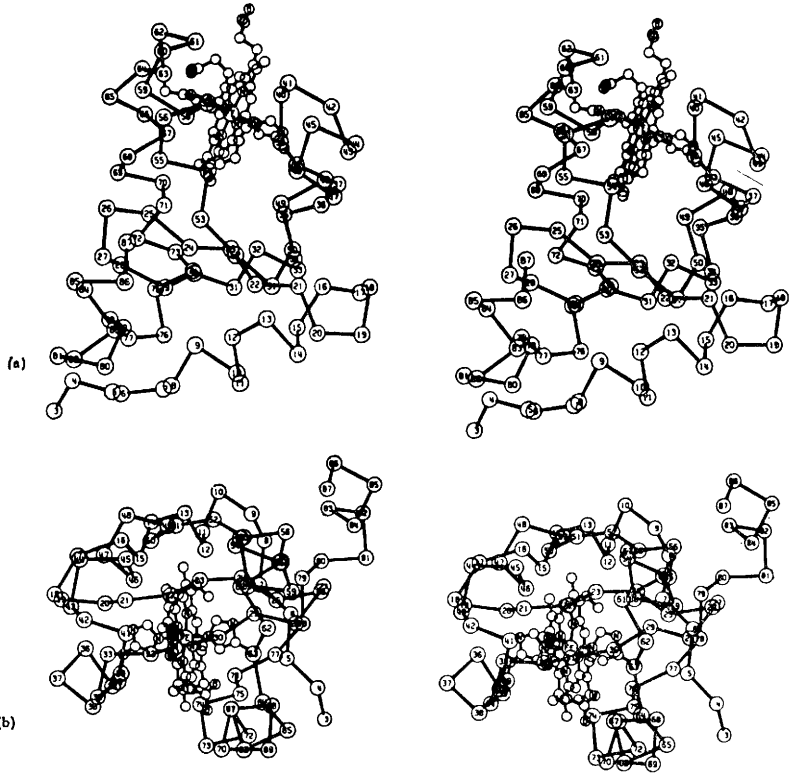


Figure 9 *Stereo drawings of the α-carbons of cytochrome* b_5. *The haem group and side-chain atoms of His-39 and -63 are included. Figures* (a) *and* (b) *show views parallel and perpendicular to the crystallographic axis* c, *respectively*
(Reproduced by permission from the *Cold Spring Harbor Symposium*, 1972, **36**, 387)

indicated by electron spin resonance and Mössbauer spectroscopy of the protein and actually found in the model compound $(C_5H_5FeS)_4$ (see Table 3).

The structure of rubredoxin, previously reported, has been refined at 1.5 Å resolution by Watenpaugh *et al.*,[43] and the sequence is now known. The iron co-ordination appears to have highly significant deviations from tetrahedral geometry. Jensen *et al.*[96] have also determined the structure of a ferredoxin, which contains two tetrahedral clusters of four iron atoms

[96] L. Jensen, personal communication.

which resemble the cluster found in the high-potential iron protein. Each iron cluster is co-ordinated to four cysteines and the relation of peptide to cluster in the two cases is very similar.

Table 3 *Average bond lengths in the HiPIP cluster and its analogue* [95]

Chromatium *HiPIP Cluster*				$(C_5H_5FeS)_4$		
Bond type	*Mean** /Å	*Range* /Å	*Precision* /Å	*Bond type*	*Mean** /Å	*Accuracy* /Å
Fe—Fe	3.06	3.0–3.1	±0.14	Fe—Fe	2.650	±0.006
					3.365	±0.006
Fe—S_{inorg}	2.35	2.1–2.9	±0.24	Fe—S_{inorg}	2.206	±0.002
					2.256	±0.003
Fe—S_{cys}	2.01	1.8–2.1	±0.26			
Mean Fe—S distance = 2.28 Å						

*Mean values represent average distances for similar bonds. Precisions given for the HiPIP cluster were calculated from the standard deviations of repeated measurements.

The structure of a clostridial flavodoxin at 3.25 Å is described by Ludwig *et al.*[97] The flavodoxins are small proteins of molecular weight *ca.* 15 000 which can replace ferrodoxins in certain electron-transferring reactions. The proteins contain no iron; rather they have a non-covalently bound flavin mononucleotide prosthetic group. The clostridial flavodoxin crystallizes in space group $P3_121$ in all oxidation states, but there are changes of cell dimensions and diffracted intensities between them.

A 3.25 Å resolution electron-density map of the semiquinone form shows that flavodoxin is a 'smorgasbord of secondary structures'. The chain folding shown in Figure 10 contains 35% helix, a central region of pleated sheet (one parallel pair and two pairs anti-parallel), and a number of 3_{10} bends. The exact orientation of the FMN prosthetic group appears to be in some doubt. However, Jensen *et al.*[96] have recently computed a 2.5 Å electron-density map based on one samarium derivative for a closely related flavodoxin, which shows the same general chain folding and indicates clearly the orientation of the FMN group.

J. Haemoglobins.—Perutz and his colleagues have described a number of experiments designed to test their stereochemical interpretation of the co-operative effects in haemoglobin, presented in 1970. The experiments include *X*-ray studies of chemically modified and mutant haemoglobins.

The constraints which clearly distinguish the deoxy-form from the oxy-structure include the salt bridges formed by the C-terminal residues shown in Scheme 2. Perutz and Ten Eyck have studied crystallographically haemoglobins modified at these groups. 3.5 Å difference electron-density maps on human haemoglobins have shown that removal of histidine

[97] M. L. Ludwig, R. D. Anderson, P. A. Apgar, R. M. Burnett, M. E. Le Quesne, and S. G. Mayhew, *Cold Spring Harbor Symposium*, 1972, **36**, 369.

HC3(146)β either enzymatically or sterically by attachment of *N*-ethylsuccinimide to the nearby cysteine F9(93)β leaves the rest of the oxy- or deoxy-structure undisturbed. Removal of arginine HC3(141)α produces structural changes in both the oxy- and the deoxy-forms, leading to new

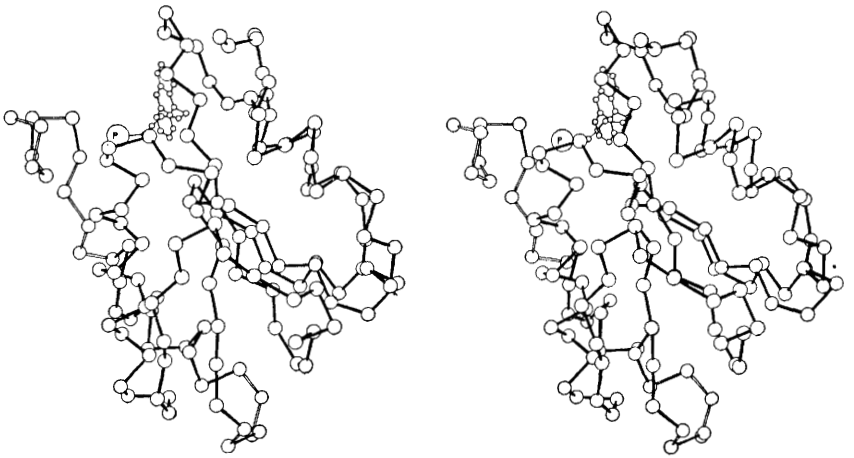
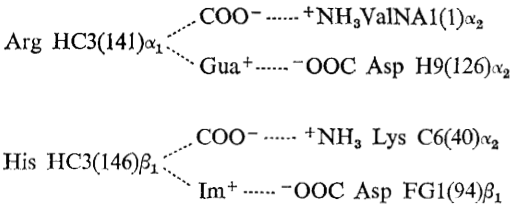


Figure 10 *A stereo representation of the α-carbon positions of the polypeptide chain of flavodoxin. Where the interpretation is tentative the bonds connecting atoms have been drawn in open form. The FMN is seen in the upper centre*
(Reproduced by permission of Prof. M. Ludwig)



Scheme 2

salt bridges. Neither of these modifications, if carried out singly, inhibits the two alternative quaternary structures. On the other hand, removal of both arginine and cysteine or removal of arginine and attachment of *N*-ethylsuccinimide to cysteine (which prevents the formation of the histidine salt bridges) inhibits crystallization of the normal deoxy-form. Instead, the deoxy-derivatives crystallize in a lattice which is closely related to that of normal human oxyhaemoglobin. Perutz and Ten Eyck conclude that oxygen affinity and Hill's constant are directly dependent on the strength and number of salt bridges constraining the tetramer in the deoxy-form. Removal of a sufficient number of the salt bridges prevents

formation of the quaternary deoxy-structure even in the total absence of haem ligand. This shows that the subunit contacts in the deoxy-form are stressed even in the absence of oxygen, and that to remain in that conformation the tetramer must be clamped by the salt bridges.

A further series of chemical and *X*-ray studies lead to similar conclusions. Moffat and his colleagues have shown that in bis-(*N*-maleimidomethyl) ether–oxyhaemoglobin, and related derivatives, reaction at cysteine F9(93)β leads to a displacement of tyrosine HC2(145)β. The resulting disruption of the hydrogen bond formed by the tyrosyl hydoxy-group to the main-chain carbonyl group of valine FG5(98)β is thought to lead to the observed reduction of co-operativity.

Morimoto *et al.*[99] have interpreted in stereochemical terms the amino-acid substitutions in abnormal haemoglobins, which produce clinical symptoms by altering the oxygen affinity. They find examples of changed affinity due to residues at the haem pocket which either restrict and/or increase oxygen binding (Zurich, Bristol, Hammersmith, and Bucaresti). They also find changes in residues which stabilize the tertiary structure of either the oxy- or deoxy-form and therefore change the oxygen affinity (Shepherd's Bush, Peterborough, and Agenoni). Finally, they find examples of changes which alter the balance of interactions between subunits (Hirose, Malmo, Yakima, Kempsey, Kansas, and Yoshizuka).

Perutz, Greer, and their colleagues have obtained more-detailed crystallographic information for certain abnormal haemoglobins. Table 4 summarizes the chemical and physiological properties of the mutants so far studied by difference Fourier techniques. Only one of these, haemoglobin M Iwate, crystallizes in a different space group and required a completely new *X*-ray analysis.

The changes in deoxyhaemoglobin Rainier[100,101] are represented schematically in Figure 11. The random mutation at the penultimate residues places a cysteine in a position to form a disulphide with perfect geometry and very little distortion of the rest of the molecule. The presence of the disulphide explains the high resistance of haemoglobin Rainier, particularly of the β-chain, to denaturation by alkali and to digestion by carboxypeptidase.

The two salt bridges made by the C-terminal histidine are disrupted as the terminal residue rotates to an α-helical conformation, and removal of the salt bridges contributes to lowering of the Bohr effect. The α-helical arrangement appears to allow a new hydrogen-bond between the carboxylate at the terminus with the α-amino-group of the opposite β-chain so that the oxy-form is relatively stabilized. This explains the increased oxygen affinity.

[98] M. F. Perutz and L. F. Ten Eyck, *Cold Spring Harbor Symposium*, 1972, **36**, 295.
[99] H. Morimoto, H. Lehmann, and M. F. Perutz, *Nature*, 1971, **232**, 408.
[100] J. Greer, *Cold Spring Harbor Symposium*, 1972, **36**, 315.
[101] J. Greer and M. F. Perutz, *Nature New Biol.*, 1971, **230**, 261.

Table 4 *Physiological properties of abnormal haemoglobins, the structures of which have been investigated by X-ray diffraction*

Mutant	*Position*	*From*	*To*	*n*[c]	*Oxygen affinity*	*Bohr effect*[d]	*Clinical symptom*	*X-Ray study of deoxy-haemoglobins (reference number)*
Rainier	βHC2 (145)	Tyr	Cys	1.5	Very high	Halved	Polycythemia	3.5 Å difference Fourier (104)
Kansas	βG4 (102)	Asn	Thr	1.3	Very low	Large	Cyanosis[a]	5.5 Å difference Fourier (101)
Richmond	βG4 (102)	Asn	Lys	2.5	Normal	Normal	Normal	4 Å difference Fourier (101)
Chesapeake	αFG4 (92)	Arg	Leu	1.4	High	Normal	Polycythemia[b]	5.5 Å difference Fourier (103)
J Capetown	αFG4 (92)	Arg	Gln	2.3	Slightly high	Normal	Normal	5.5 Å difference Fourier (103)
M Iwate	αF8 (87)	His	Tyr	1.1	Low	None	Cyanosis	5.5 Å *X*-ray analysis, different space group (102)
M Hyde Park	βF8 (92)	His	Tyr	1.3	Normal	Normal	Methemoglobinemia	3.5 Å difference Fourier (102)
Hiroshima	βH24 (146)	His	Asp		High	Halved		3.5 Å difference Fourier (105)

[a] Cyanosis: blue colour of skin owing to deoxyhaemoglobin in capillaries. [b] Polycythemia: excess of red cells. [c] *n*: Hills constant, a measure of the interaction between the four haem groups. [d] Bohr effect: dependence of oxygen affinity on pH.

In haemoglobins Kansas and Richmond,[102] an asparagine which is close to the haem group is changed. The lysine in this position in haemoglobin Richmond causes only local changes. On the other hand, the replacement of asparagine by threonine in Kansas would cause unfavourable steric

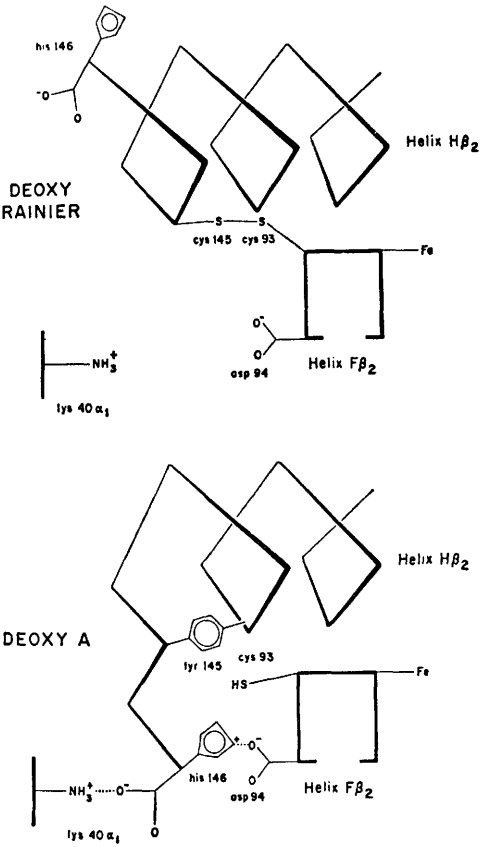


Figure 11 *A schematic representation of the structures of deoxyhaemoglobin Rainier and normal deoxyhaemoglobin*
(Reproduced by permission from *Cold Spring Harbor Symposium*, 1972, **36**, 315)

interactions between the γ-methyl of the threonine and the methyl and vinyl side-chains of the haem group. The difference Fourier indicates that this is avoided by movement not only of the haem but also by helices Fβ, Eβ, and Bβ. These substantial changes give rise to a decrease of co-operativity, oxygen affinity, and Bohr effect.

Haemoglobins M Hyde Park and M Iwate[103] have a haem-linked histidine replaced by a tyrosine in the β- and α-chains, respectively. In

[102] J. Greer, *J. Mol. Biol.*, 1971, **59**, 99.
[103] J. Greer, *J. Mol. Biol.*, 1971, **59**, 107.

deoxyhaemoglobin Hyde Park, 20—30% of β haem groups appear to be lost, but otherwise there are only local changes in structure. However, in haemoglobin M Iwate both the proximal tyrosine (induced by mutation) and the distal histidine appear to bind haem iron. Both the met- and the deoxy-forms have the deoxy-quaternary structure, and this explains the low oxygen affinity.

Crystallographic studies on haemoglobins J Capetown and Chesapeake [104] show that there are no structural distortions in the deoxy-forms of either mutant. On the other hand, replacement of the invariant arginine by leucine at αFG4(92) in Chesapeake appears to lead to impermissible van der Waals contacts in the oxy-form, which give rise to a distorted structure. Possibly in going from deoxy- to oxy-haemoglobin Chesapeake the β_2-chain moves too far past FGα_1, thus strengthening the oxy-form and increasing the oxygen affinity.

A crystallographic study of haemoglobin Hiroshima [105] has shown that the previously proposed amino-acid substitution histidine 143(H21)β to aspartic acid is incorrect. Although this change might account for a diminished Bohr effect, the normal response of its oxygen affinity to 2,3-DPG was inconsistent with the proposed role of histidine 143 in 2,3-DPG binding by haemoglobin. Difference electron-density maps of deoxyhaemoglobin Hiroshima have now revealed that the replacement occurs not in position 143 but in 146, thus supporting the role of histidine 146β in the alkaline Bohr effect. These results have been confirmed by chemical methods.

The structure of horse carboxyhaemoglobin labelled with the spin label *N*-(1-oxyl-2,2,6,6-tetramethyl-4-piperidinyl)iodoacetamide has been determined.[106] The label has two orientations, which correspond to the weakly immobilized and strongly immobilized components of the e.s.r. spectra found in solution and in crystals. One is free and the other displaces the C-terminus of the β-chain, causing perturbations in the β-chains, and to a lesser extent in the α-chains.

The small-angle scattering functions of haemoglobin for *X*-rays and neutrons have been calculated from the atomic co-ordinates of haemoglobin.[107] The theoretical scattering functions were compared with the experimental, and the good agreement gives confidence to the statement that the three-dimensional structures of haemoglobin in single crystals and in aqueous solution are the same.

The three-dimensional structures of a number of monomeric myoglobins and haemoglobins have been reported. As expected, all have the myoglobin

[104] J. Greer, *J. Mol. Biol.*, 1971, **62**, 241.
[105] M. F. Perutz, P. del Pulsinelli, L. Ten Eyck, J. V. Kilmartin, S. Shubata, I. Fuchi, T. Miyaji, and H. B. Hamilton. *Nature New Biol.*, 1971, **232**, 149.
[106] J. K. Moffat, *J. Mol. Biol.*, 1971, **55**, 135.
[107] R. Schneider, A. Mayer, W. Schmatz, J. Schelten, R. Franzel, and H. Eicher, *European J. Biochem.*, 1971, **20**, 179.

fold. The course of evolution of these globins is shown in Figure 12, and the percentage mismatch between pairs is illustrated in Table 5.

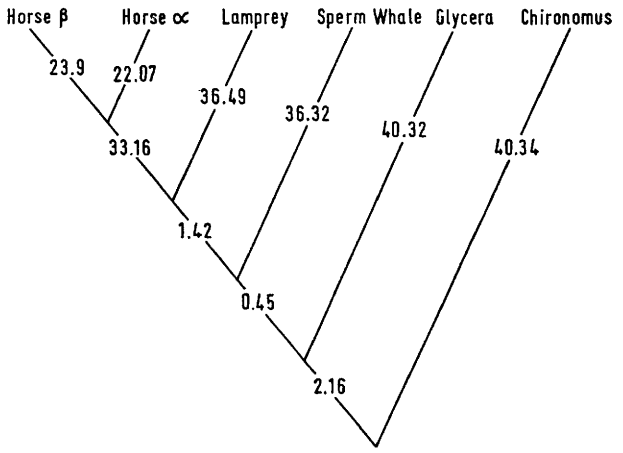


Figure 12 *Haemoglobin evolution. The line running from lower right to upper left represents the main course of evolution*

Table 5 *Percentage mismatch between pairs of amino-acid sequences from some globins*

Horse α Hb	54				
Lamprey Hb	75	70			
Sperm whale Mb	72	74	75		
Glycera Hb	77	77	80	76	
Chironomus Hb	82	79	79	79	85
	Horse β	Horse α	Lamprey	Sperm whale	*Glycera*

Huber *et al.*[108] compare the structure of the insect-larval haemoglobin from *Chironomus* (earlier mistakenly named erythrocruorin) with the structure of sperm-whale myoglobin. They show that although only 21% of the amino-acids are identical in both proteins, the structures are closely similar. Of particular note is the absence of the distal histidine E7. This is invariant in all the known vertebrate haemoglobins, but in *Chironomus* is replaced by a glutamate, which protrudes into the surrounding solution. This indicates that a histidine is not essential for the function of haemoglobin as a reversible oxygen carrier.

[108] R. Huber, O. Epp, W. Steigemann, and H. Formanek, *European J. Biochem.*, 1971, **19**, 142.

Love *et al.* have determined the structures of two other evolutionary remote haemoglobins.[109] The structure of lamprey haemoglobin determined[109, 110] at 2 Å resolution shows that the extra C-terminal nonapeptide runs in an extended configuration along the H helix, making two hydrogen bonds and covering a methionine and an isoleucine which would otherwise be exposed to the solvent. A deletion of nine residues, G18 to H2, is accommodated by a shortening of the G and H helices. The structure of the annelid worm *Glycera* haemoglobin has been determined[109] at 2.5 Å resolution. The structure is similar to that of sperm-whale myoglobin. As in the *Chironomus* haemoglobin, the distal residue is not histidine, but in this case leucine. Lattman *et al.*[111] have shown that, at least at low resolution, yellow tuna-fin myoglobin has a similar 'myoglobin fold'.

At the present time the relation of structure to physiology has not been fully worked out for lamprey and *Glycera* haemoglobins, and further studies are under way. The study of the insect larval *Chironomus* haemoglobin is more advanced. This haemoglobin has been shown to undergo a conformational change in the crystalline state between pH 7 and pH 5.5, which is thought to be related to the dissociation of haem and globin at acidic pH, and may also give rise to the large changes in oxygen affinity observed in solution.[112]

K. Concanavalin A.—The structure of concanavalin A from jack bean has been studied in a number of different laboratories. This plant protein is capable of agglutinating various types of cells and inducing cellular transformation and mitosis. Each subunit of molecular weight 27 000 contains three mutually interacting binding sites for polysaccharides, Mn^{2+}, and Ca^{2+}. The protein aggregates to dimers below pH 6 and to tetramers above pH 7, and the subunit structure giving rise to more than one site may be responsible for the agglutination.

X-Ray studies on the same crystal form have been carried out in three different laboratories and are reported by Quiocho *et al.*[113, 114] and by Hardman *et al.*[115] Patterson syntheses have confirmed the space group as *I*222 with one subunit per asymmetric unit. The two electron-density maps, computed independently at 4 Å, are of very good quality with figures of merit 0.79 and 0.78. Electron-density maps at higher than 3 Å resolution

[109] W. E. Love, P. A. Klock, E. E. Lattman, E. A. Padlan, K. B. Ward, and W. A. Hendrickson, *Cold Spring Harbor Symposium*, 1972, **36**, 349.
[110] W. A. Hendrickson and W. E. Love, *Nature New Biol.*, 1971, **232**, 197.
[111] E. E. Lattman, C. E. Nockolds, R. H. Kretsinger, and W. E. Love, *J. Mol. Biol.*, 1971, **60**, 271.
[112] R. Huber, O. Epp, and H. Formanek, *J. Mol. Biol.*, 1971, **57**, 377.
[113] F. A. Quiocho, G. N. Reeke, J. W. Becker, W. N. Lipscomb, and G. M. Edelman, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1853.
[114] G. N. Reeke, J. W. Becker, and F. A. Quiocho, *Cold Spring Harbor Symposium*, 1972, **36**, 277.
[115] K. D. Hardman, M. K. Wood, M. Schiffer, A. B. Edmundson, and C. F. Ainsworth, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1393; *Cold Spring Harbor Symposium*, 1972, **36**, 271.

have also been computed, but their interpretation has not been published. The sequence is at present unknown.

Each subunit has a half ellipsoidal shape with base 46 × 26 Å and height 42 Å. The dimers are elongated ellipsoids formed by subunits related by the crystallographic two-fold axis parallel to the *c* axis. The dimers are paired at greater distances across points of 222 symmetry to form tetrahedral tetramers. Although large portions of the polypeptide chain can be clearly traced, an unambiguous interpretation has not been possible at this stage. However, a large depression on each subunit may be the binding site for saccharides, and this has been positively identified by difference Fourier studies at 2.8 Å resolution.[116] The metal-free protein has been crystallized in space group $P2_122_1$ with unit cell dimensions nearly equal to those of the native crystals,[117] but with the asymmetric unit containing a dimer. The crystals revert to a form close to the native structure on soaking in Mn^{2+} and Ca^{2+} ions. Studies at 5.5 Å resolution using the rotation and translation functions indicate the presence of a two-fold local axis at 7° to the *c* axis with a screw component involving a translation of 6 ± 2 Å relative to this axis. These results are thought to indicate a large conformational change on demetallization.

L. Toxins.—Crystallographic studies on the small neurotoxic proteins of snake venoms are an important and exciting area of study which surely must attract much attention in the future. Low and her colleagues [118] have now described their preliminary *X*-ray studies on two such toxins, erabutoxin *a* and erabutoxin *b* from the sea snake *Laticauda serrifasciata*. These toxins each have 62 amino-acids and differ only at one position in the sequence. They form isomorphous crystals of space group $P2_12_12_1$ and all dimensions are given in Table 6. Histidine 26 of erabutoxin *b* has been

Table 6 *Cell parameters and estimated standard deviations for orthorhombic crystals of erabutoxins* a *and* b *and di-iodoerabutoxin* b

Cell parameters	*Erabutoxin* a	*Erabutoxin* b	*Di-iodoerabutoxin* b
a/Å	50.0 ± 0.1	49.76 ± 0.05	51.5 ± 0.5
b/Å	46.5 ± 0.1	46.77 ± 0.04	47.4 ± 0.1
c/Å	21.2 ± 0.1	21.58 ± 0.02	21.8 ± 0.1

specifically di-iodinated and this gives crystals which have the same space group and closely related cell dimensions to the native protein as shown in Table 6. All crystal forms appear to give good quality diffraction patterns, and data to 1.8 Å resolution have been collected on the erabutoxin *b*.

M. Hormones.—Further details of the structure of zinc insulin hexamers based on a 2.8 Å resolution electron-density map (see Figures 13 and 14)[119]

[116] J. W. Becker, G. N. Reeke, and G. M. Edelman, *J. Biol. Chem.*, 1971, **247**, 6123.
[117] A. Jack, J. Weinzierl, and A. J. Kalb, *J. Mol. Biol.*, 1971, **58**, 389.
[118] B. W. Low, *J. Biol. Chem.*, 1971, **246**, 4366.
[119] T. L. Blundell, J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. C. Hodgkin, D. A. Mercola, and M. Vijayan, *Nature*, 1971, **231**, 506.

and the relation of the structure to the chemistry, sequence variation, and biological role of insulin have been discussed.[120]

There is now a considerable body of evidence for the view that in solutions approximating physiological conditions, insulin retains an ordered and stable structure and resembles crystalline insulin.[120] The

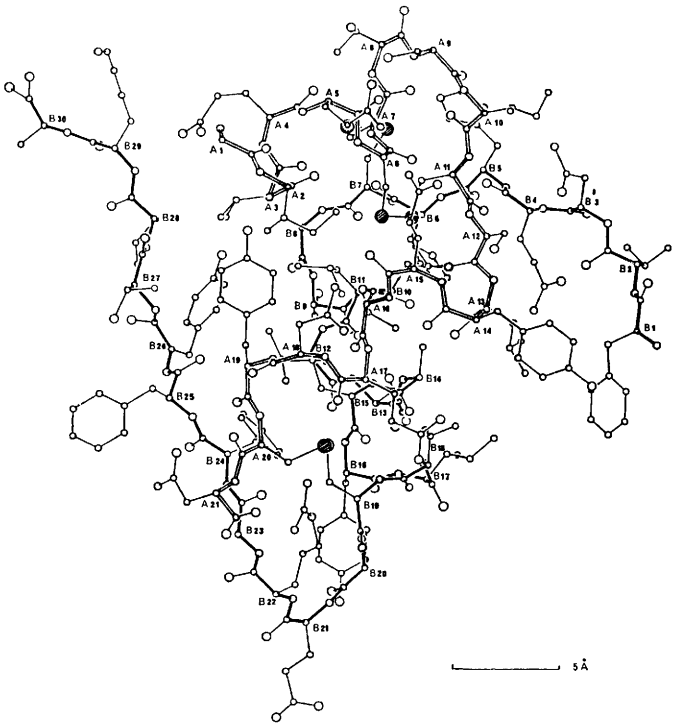


Figure 13 *The atomic positions in the insulin monomer*

residues which form the compact non-polar core of the molecule are conserved in the twenty sequenced insulins. Chemical studies designed to examine the availability of different functional groups—whether they are buried or exposed—and also to study the geometric relations between parts of the molecule are in general agreement with the *X*-ray model. The general tertiary structure appears to be common to insulins of fishes and mammals, and is conserved in solution. These are features which are common to most enzyme structures, but may not be true of smaller polypeptide hormones. Evidence is presented concerning the three-dimensional structure and the role of the connecting peptide of proinsulin, the mode of

[120] T. L. Blundell, E. J. Dodson, G. G. Dodson, D. C. Hodgkin, and M. Vijayan, *Recent Progr. Hormone Res.*, 1971, **20**, 1.

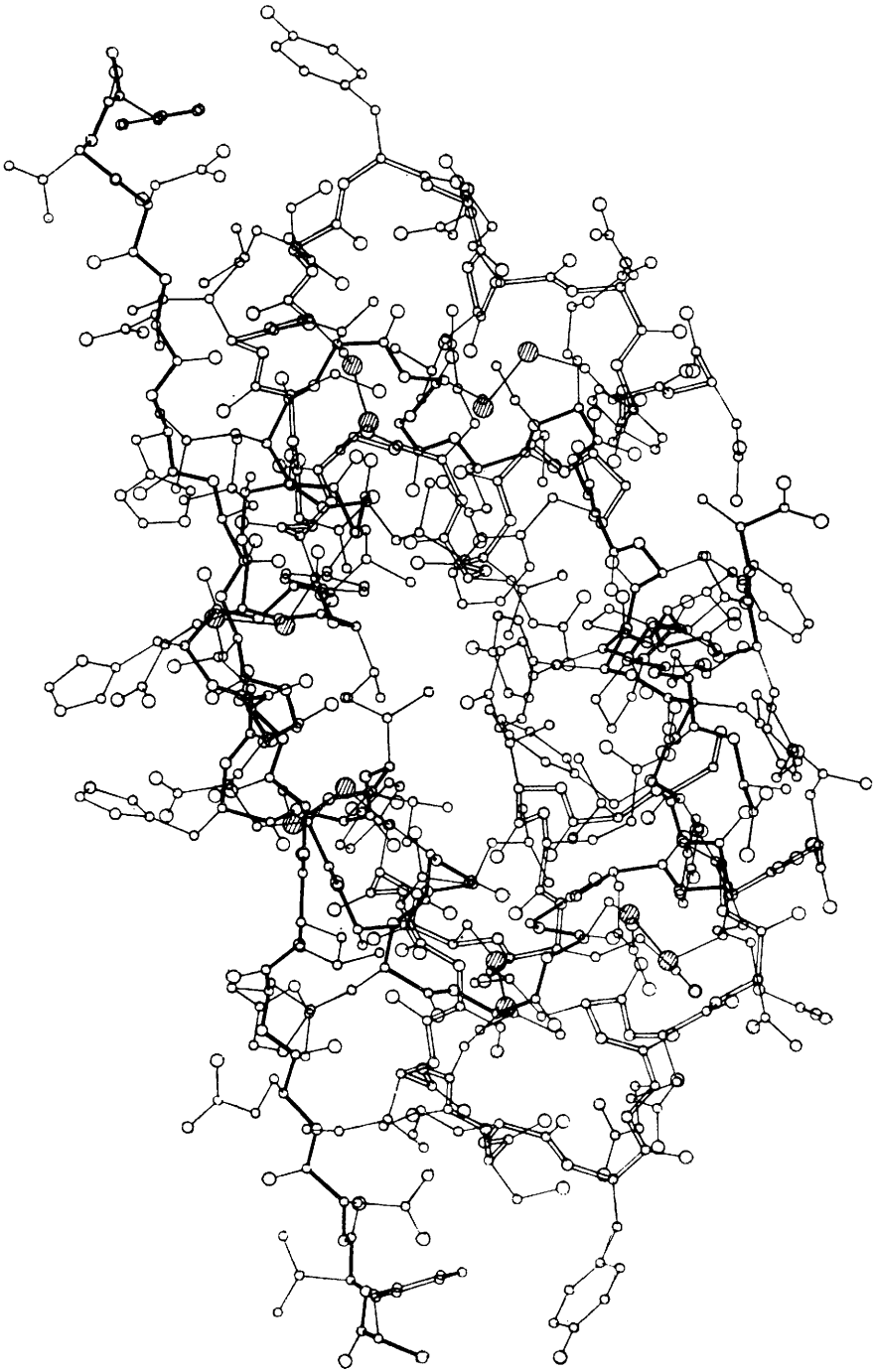


Figure 14 *The atomic positions in the insulin dimer as viewed along the local approximate two-fold axis*

storage of insulin in β granules, the aggregation of insulins of different species, and the active region of the molecule.

Preliminary *X*-ray and diffraction studies are reported for small crystals (1—5) of angiotensinamide II.[121]

N. Plasma Proteins.—Preliminary *X*-ray studies of the plasma protein prealbumin are reported.[122] Prealbumin has a strong interaction with retinol (vitamin A alcohol) binding protein, and its principal function appears to relate to plasma transport of vitamin A, rather than L-thyroxine, which it also binds. The protein crystallizes in the orthorhombic system, but the molecular symmetry of the prealbumin molecule gives rise to pseudo body-centring. This has confused attempts to fix the space group, which now appears to be $P22_12_1$. The molecule is a tetramer of molecular weight 54 000 ± 5000, composed of subunits with identical, or possibly two slightly different, tertiary structures arranged tetrahedrally.

O. Immunoglobulins.—Crystallographic studies of immunoglobulins are under way in a number of laboratories. These include studies of complete immunoglobulin molecules, of Bence-Jones proteins, and of fragments.

Davies and his colleagues have succeeded in obtaining a 6 Å resolution electron-density map of a human γG1 immunoglobulin molecule.[123, 124] This myeloma protein contains two light polypeptide chains of molecular weight 22 500 each and two heavy polypeptide chains of molecular weight 53 000, held together by disulphide bonds and by non-covalent forces. The crystals of space group *C*2 contain two immunoglobulin molecules of molecular weight 150 000. There is one half-molecule per asymmetric unit and the crystallographic two-fold axis coincides with the molecular two-fold axis relating the two halves of the molecule.

The *X*-ray data extend to *ca.* 4.5 Å resolution, but their quality is low outside the 6 Å sphere. The extreme sensitivity of the crystals of immunoglobulin to *X*-radiation presented considerable difficulties and *ca.* eight crystals were required to collect the 2000 reflections in the 6 Å sphere. The crystals were also very sensitive to many heavy-atom salts, but nevertheless three good derivatives have been prepared by soaking the crystals in (*a*) 10^{-3}M *p*-chloromercuribenzenesulphonate, (*b*) 10^{-2}M mercuric cyanide, (*c*) 10^{-3}M chloroplatinate ion. An electron-density map has been calculated at 6 Å with an average figure of merit of 0.73. The authors have not tried to interpret the chain folding at this resolution, rather they have attempted to define the boundary and the shape of the immunoglobulin molecule. The density in the region of the dyad axis at $z = 0$ has indicated

[121] S. Fermandjian, J. L. Morgat, P. Fromageot, C. Legressus, and P. Maire, *F.E.B.S. Letters*, 1971, **16**, 192.
[122] C. C. F. Blake, I. D. A. Swan, C. Rerat, J. Berthou, A. Laurent, and B. Rerat, *J. Mol. Biol.*, 1971, **61**, 217.
[123] V. R. Sarma, E. W. Silverton, D. R. Davies, and W. D. Terry, *J. Biol. Chem.*, 1971, **247**, 3753.
[124] L. W. Labaw and D. R. Davies, *J. Biol. Chem.*, 1971, **247**, 3760.

that this is the molecular axis. The density appears to be located in three globular regions. One of these regions around the dyad axis most probably corresponds to the F_c (the complement) fragment while the other two are related by the dyad axis and may be the F_{ab} (the antibody) fragments. The most difficult task has been to join the F_{ab} regions to the F_c region. There are four possible arrangements of which the T model is favoured.

These same immunoglobulin crystals have also been the subject of a study in the electron microscope.[125] The molecule in projection appears to have a shape varying between a Y and a T, in agreement with previous electron microscope studies and the *X*-ray diffraction investigation.

Edmundson *et al.*[126] also report crystallization and *X*-ray studies of a serum IgG (in this case a euglobulin) from a patient with multiple myeloma. Prismatic crystals as large as 0.6 × 0.6 × 3.0 mm are obtained by a dialysis against deionized water at 4 °C using a microdiffusion method. Unfortunately—because they are grown at low ionic strength—the crystals dissolve when placed in certain solutions of heavy atoms. However, attempts to introduce ammonium or sodium sulphate into the crystals to which heavy atoms could be added without affecting the solubility were unsuccessful. The prismatic crystals have space group $C222_1$, with the asymmetric unit continuing one half-molecule, or one light and one heavy chain. The two halves are related by a two-fold rotation axis as in the cryoglobulin of Sarma *et al.*[123] The euglobulin crystals are less sensitive to radiation damage than those of the cryoglobulin, and reflections extend to 3.5 Å resolution.

Edmundson *et al.*[126] have also studied crystals of the λ-type Bence-Jones protein (the light chain of the immunoglobulin) from the same myeloma patient. The protein crystallizes in orthorhombic form from deionized water and as a trigonal form from ammonium sulphate. The orthorhombic form of space group $P2_12_12$ contains the M—S—S—M dimer. As with the euglobulin, the salt-free crystals dissolved in most heavy-atom compounds, but $PtCl_4^{2-}$ made the crystals less soluble, after which they could be treated by dilute solutions of heavy-atom salts; the rotation function has demonstrated the existence of a local two-fold axis, and a preliminary 6 Å map has been calculated based on the chloroplatinite derivative. The trigonal form also contains the M—S—S—M dimer. Unlike the orthorhombic form, the preparation of heavy-atom derivatives was not hampered by solubility problems in the trigonal form. Heavy-atom positions of three derivatives are being refined with a view to computation of a 6 Å map.

X-Ray studies of a second Bence-Jones protein of antigenic type (κ) are being carried out in Münich by Epp *et al.*[127]

[125] V. R. Sarma, D. R. Davies, L. W. Labaw, E. W. Silverton, and W. D. Terry, *Cold Spring Harbor Symposium*, 1972, **36**, 413.
[126] A. B. Edmundson, M. Schiffer, M. K. Wood, K. D. Hardman, K. R. Ely, and C. F. Ainsworth, *Cold Spring Harbor Symposium*, 1972, **36**, 427.
[127] O. Epp, R. Huber, and W. Stergemann, unpublished results.

X-Ray studies on the F_{ab} fragments prepared by papain and pepsin fragmentation of human myeloma IgG's are being carried out by Poljak *et al.*[128] The preparation of heavy-atom derivatives has caused considerable difficulties. A 6 Å electron-density map has been computed, which shows two clear domains (unpublished work) and is clearly of good quality. The data have been collected to 3.5 Å resolution, and refinement of the heavy-atom positions of the derivatives has been completed and an electron-density map computed.

P. Muscle Proteins.—Kretsinger *et al.* have grown crystals and reported preliminary crystal data of three of the lower molecular weight albumins of carp.[129] They have succeeded in solving the structure at 2 Å resolution of one, a calcium-binding protein, which may be analogous to troponin-A of mammalian and avian muscle. It has a high percentage of phenylalanine (10%) and alanine (20%), little or no tryptophan, tyrosine, methionine, histidine, cystine, or arginine. The interpretation of the electron-density map was achieved by fitting the sequenced tryptic and thermolytic peptides with the aid of the homologous hake albumin sequence.

The polypeptide chain comprises 107 residues. There are three helices involving residues 42—50, 77—87, and 98—106. The molecule is a prolate spheroid with the amino-terminal tucked in so that the ten phenylalanines, which are invariant between hake and carp, are in internal positions. Two spherical peaks in the electron density are interpreted as calcium ions (Kretsinger, unpublished results).

X-Ray diffraction and electron microscopy studies of tropomyosin crystals by Cohen and co-workers[130] indicate that the crystal dynamics may be analogous to structural changes regulating contractile activity of muscle. Tropomyosin is a two-stranded α-helical coiled-coil which forms macroscopic crystals. As the crystals contain 95% water in the lattice they are quite disordered and show a mosaic spread of not less than 1° in their *X*-ray pictures. Nevertheless, by using electron micrographs to phase the *X*-ray projection data, Fouriers have been computed for the [100] and [001] views (see Figure 15), from which a model has been derived. The tropomyosin molecules, each 400 Å long, are bonded head-to-tail and cross-connected. The diagonally oriented strands observed in micrographs consist of a pair of parallel molecular filaments each 20 Å in diameter and separated by *ca.* the same distance. Separation between cross-over points in the lattice is alternately short (165—175 Å) and long (225—235 Å) and these dimensions are found in all polymorphic forms. The alternation of distances gives rise to a bending of the filaments. There is a continuous

[128] R. J. Poljak, L. M. Amzel, H. P. Avey, L. N. Becka, D. J. Goldstein, and R. L. Humphrey, *Cold Spring Harbor Symposium*, 1972, **36**, 421.
[129] R. H. Kretsinger, C. E. Nockolds, C. J. Coffee, and R. A. Bradshaw, *Cold Spring Harbor Symposium*, 1972, **36**, 217.
[130] C. Cohen, D. L. D. Caspar, D. A. D. Parry, and R. M. Lucas, *Cold Spring Harbor Symposium*, 1972, **36**, 205.

range of lattice parameters of the tropomyosin crystals, which appears to be dependent on tryponin binding in the crystals. The changes in the crystals correspond to a scissor-like motion of the molecules relative to each other. An analogous structural change may occur when the filaments are bound in the grooves of the actin helix, and this may be the mode of communication between the calcium-binding troponin complex and actin.

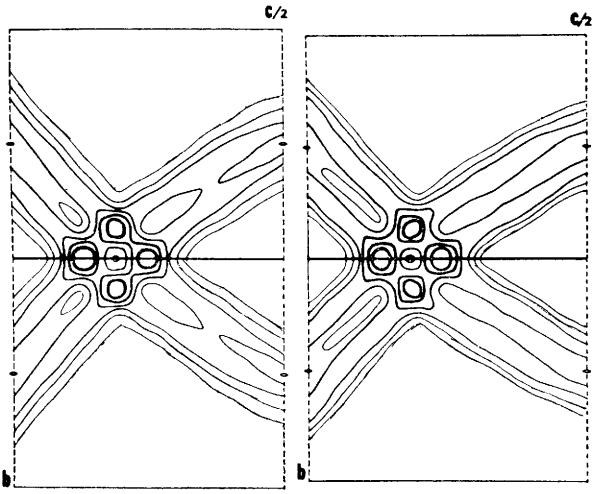


Figure 15 *Electron-density maps of* [100] *projection of two crystal forms of tropomyosin*
(Reproduced by permission from *Cold Spring Harbor Symposium*, 1972, **36**, 205)

Harrison *et al.*[131] discuss the assembly of myosin. They have determined by *X*-ray analysis the helical arrangement of bridges on the thick filament. Placing rods of length 1450 Å on this surface lattice generates the molecular overlaps found in *in vitro* 'segment' aggregates.

Q. Viruses.—The Cold Spring Harbor Symposium has brought together a number of *X*-ray studies of viruses, and has shown how these studies relate to investigations on the mechanism and control of virus assembly.

Barrett *et al.*[132] review the advances in *X*-ray techniques which have led to the present 10 Å resolution electron-density map of tobacco mosaic virus. Four heavy-atom derivatives of this helical virus have been used to calculate phase angles. Comparison of *X*-ray data for the virus with that for the repolymerized helical protein without RNA indicates that the RNA forms a helix having a radius of *ca.* 38 Å. A chain can be followed through

[131] R. G. Harrison, S. Lowey, and C. Cohen, *J. Mol, Biol.*, 1971, **59**, 531.
[132] A. N. Barrett, J. Barrington Leigh, K. C. Holmes, R. Leberman, E. Mandlekow, P. von Sengbusch, and A. Klug, *Cold Spring Harbor Symposium*, 1972, **36**, 433.

the electron density and this polypeptide forms a herring-bone pattern at the intersubunit contact. Three-dimensional image reconstruction of electron micrographs by Finch and Klug[133] are broadly consistent with this model. Studies with an electron microscope and ultracentrifuge[134–138] indicate the way in which the tobacco mosaic virus aggregates to form this structure. The aggregation involves the intermediate formation of discs, built of rings of 17 subunits, which have also been studied by *X*-ray diffraction[139] and electron microscopy.[140] The initiation of the growth of the nucleoprotein helix is thought to be the binding of RNA to two protein discs which then dislocate into 'lock-washers' before further discs are stacked and dislocated to give the helix.

The use of the rotation function to study the arrangement of subunits in spherical viruses has led to some interesting discussion.[141] A preliminary analysis of the *X*-ray data on the satellite tobacco necrosis virus crystals indicated that these might have octahedral rather than icosahedral symmetry. However, all solutions to the rotation function originally considered involved a particle two-fold axis parallel to the monoclinic two-fold axis. It has been demonstrated that if this restriction is relaxed, octahedral symmetry is generated in the Patterson by the packing of two virus particles of icosahedral symmetry.[142] This special icosahedral solution appears to be definitely preferable to the octahedral solution.

Icosahedral structures with triangulation numbers of $T = 7$ and 3 are indicated by three-dimensional image reconstruction of human wart virus and tomato bushy stunt virus,[143] and a three-dimensional *X*-ray analysis at 30 Å resolution of the bushy stunt virus is consistent with this interpretation.[144]

Although a complete virus is rather large to handle by presently available techniques, a study of the six-co-ordinated morphological units (hexons) from the surface of the adenovirus seems more feasible. The adenovirus type-2 hexon, which has a molecular weight of *ca.* 333 000, forms bipyramidal crystals of length up to 0.6 mm.[145] These have four hexons in a cubic unit cell with $a = 149.9$ Å, space group $P2_13$. Tetrahedral crystals of type-5 adenovirus hexon seem to be isomorphous with the type-2

[133] J. T. Finch and A. Klug, *Phil. Trans.*, 1971, **B261**, 211.
[134] P. J. G. Butler and A. Klug, *Nature New Biol.*, 1971, **229**, 47.
[135] A. C. H. Durham and A. Klug, *Nature New Biol.*, 1971, **229**, 42.
[136] A. C. H. Durham, J. T. Finch, and A. Klug, *Nature New Biol.*, 1971, **229**, 37.
[137] A. Klug and A. C. H. Durham, *Cold Spring Harbor Symposium*, 1972, **36**, 449.
[138] P. J. G. Butler, *Cold Spring Harbor Symposium*, 1972, **36**, 461.
[139] P. F. C. Gilbert, Ph.D. Thesis, University of Cambridge, 1970.
[140] R. A. Crowther and L. A. Amos, *J. Mol. Biol.*, 1971, **60**, 123.
[141] K. Akervall, B. Strandberg, M. G. Rossmann, U. Bengtsson, K. Fridborg, H. Johannisen, K. K. Kanan, S. Lövgren, G. Petef, B. Öberg, D. Eaker, S. Hjerlén, L. Rydén, and I. Moring, *Cold Spring Harbor Symposium*, 1972, **36**, 469.
[142] A. Klug, *Cold Spring Harbor Symposium*, 1972, **36**, 483.
[143] R. A. Crowther and L. A. Amos, *Cold Spring Harbor Symposium*, 1972, **36**, 489.
[144] S. C. Harrison, *Cold Spring Harbor Symposium*, 1972, **36**, 495.
[145] R. M. Franklin, U. Pettersson, K. Akervall, B. Strandberg, and L. Philipson, *J. Mol. Biol.*, 1972, **57**, 383.

crystals.[146] The diffraction patterns are of good quality to 2.8 Å resolution. The hexon contains three dimers (the asymmetric unit) related by a crystallographic three-fold axis. The present work indicates that the hexon does not have six-fold rotational symmetry, but that there may be an approximate dyad perpendicular to the three-fold axis. However, studies on hexon groups of nine suggest the hexon must be a polar oligimer, emphasizing that the perpendicular dyad must be interpreted with caution.[147]

5 Fibrous Proteins

A. Keratins.—It is known that keratins giving the 'α-pattern' contain α helices coiled in pairs as two-stranded ropes, and it has been thought previously that these 'ropes' have an axial period of 198 Å. Fraser and MacRae[148] have used a high-resolution focusing *X*-ray camera to obtain diffraction patterns from several native keratins, and find that calculations based on the 198 Å period give discrepancies considerably in excess of experimental error between calculated and observed values of reflexions in their fibre *X*-ray patterns. Instead they suggest a 470 Å axial period, in which the microfibrils have an initial screw axis with 19 units in two turns. The pseudo-period of 198 Å is thought to be a consequence of the fact that units separated by this distance have almost identical environments, rather than an indication of molecular dimensions.

Fraser *et al.*[149] also present a model for feather keratin in which the two β-sheet units are symmetrically disposed about the fibre axis. These are built into a helical structure with four pairs of units per turn.

B. Silks.—Several investigations are relevant to the structure of silks, which may contain between 20 and 60% glycine. The structures of small crystals of poly-(L-Ala-Gly-Gly-Gly) I and II (AGGG I and II), poly-(L-Ala-Gly-Gly) II (AGG II), and poly-(L-Ala-Gly) II (AG II) have been studied by *X*-ray powder diffraction, selected-area electron diffraction, and electron microscopy.[150–152] AGGG II has a structure similar to that of polyglycine II. AGG II has a structure also based upon a 3_1-helical conformation, but is composed of double layers of helices probably bridged by hydrogen-bonded water molecules. The unit cell of AG II is orthorhombic with cell dimensions $a = 4.72$ Å, $b = 14.4$ Å, and $c = 9.6$ Å. The structure proposed has dihedral angles alternately in the α- and β-conformations, and the resulting chains form hydrogen-bonded sheets. They have a screw-axis repeat of 9.6 Å resembling a highly contracted β-form. The powder diffraction pattern resembles that of silk I which

[146] G. Cormick, P. B. Sigler, and H. S. Ginsberg, *J. Mol. Biol.*, 1971, **57**, 397.
[147] R. M. Franklin, S. C. Harrison, U. Pettersson, L. Philipson, C. I. Brandén, and P.-E. Wemer, *Cold Spring Harbor Symposium*, 1972, **36**, 503.
[148] R. D. B. Fraser and T. P. MacRae, *Nature*, 1971, **233**, 138.
[149] R. D. B. Fraser, T. P. MacRae, D. A. D. Parry, and E. Suzuki, *Polymer*, 1971, **12**, 35.
[150] B. Lotz and H. D. Keith, *J. Mol. Biol.*, 1971, **61**, 195.
[151] B. Lotz and H. D. Keith, *J. Mol. Biol.*, 1971, **61**, 201.
[152] J. C. Andries, J. M. Anderson, and A. G. Walton, *Biopolymers*, 1971, **10**, 1049.

suggests that this protein has a similar structure. AGG I possesses a supercoiled cross-β structure, long fibres of which show some twisting and intertwining.

C. Collagen.—Miller and Wray [153] describe the equatorial *X*-ray pattern of collagen fibrils and discuss the implications as to the lateral packing of the tropocollagen molecules, which are 2900 Å long, 15 Å wide, and have a triple-helical conformation. They propose a model involving a five-stranded rope.

The role of the tripeptide sequence Ala-Pro-Gly in the structure of collagen has been studied by *X*-ray analyses of oligomers and of a polymer containing this sequence.[154] The polymer and oligomers longer than the hexapeptide can adopt three different types of structure, two of which are composed of polyproline II-like helices, and the other is a triple helix. Solution studies show the polymer to be unordered in aqueous solution. Thus there is a striking difference of (Ala-Pro-Gly) from the previously studied (Pro-Ala-Gly), which is triple helical both in solid and in solution. The two sequences appear to have different roles in collagen structure.

D. Flagella.—*X*-Ray and optical diffraction patterns of flagella of different species of bacteria are closely similar [155] and show a 52 Å periodicity and width of 130—115 Å. The data are consistent with a helical arrangement of subunits having an 8_4 screw-axis.

PART III: Conformation and Interaction of Peptides and Proteins in Solution

edited by R. H. Pain, with contributions by P. M. Bayley, J. R. Brocklehurst, C. E. Johnson, G. Kellett, N. C. Price, H. W. E. Rattle, B. Robson, and R. M. Stephens

1 Theoretical Aspects of Protein Structure

contributed by B. Robson

There has been continuing interest in the possibility of predicting the conformation of a protein by starting simply with its amino-acid sequence. This section considers relevant work published subsequent to (or not covered by) the discussion of this topic in Volume 1 [1] and other recent reviews [2–5] on related subjects.

A. The General Approach using Energy Functions.—The method of predicting protein conformation using energy functions follows from experimental

[153] A. Miller and J. S. Wray, *Nature*, 1971, **230**, 437.
[154] B. B. Doyle, W. Traub, G. P. Lorenzi, and E. R. Blout, *Biochemistry*, 1971, **10**, 3052.
[155] J. N. Champress, *J. Mol. Biol.*, 1971, **56**, 295.
[1] P. M. Hardy, in 'Amino-acids, Peptides, and Proteins', ed. G. T. Young (Specialist Periodical Reports), The Chemical Society, London, 1969, **1**, 127.
[2] G. N. Ramachandran and V. Sasisekharan, *Adv. Protein Chem.*, 1968, **23**, 283.
[3] H. A. Scheraga, *Chem. Rev.*, 1971, **71**, 195.
[4] J. F. Brandts, in 'Structure and Stability of Biological Macromolecules' ed. S. N. Timasheff and G. D. Fasman, Marcel Dekker, New York, 1969, **2**, 213.
[5] C. Tanford, *Adv. Protein Chem.*, 1970, **24**, 1.

evidence [6, 7] that the native conformation of a protein is the conformation of lowest free energy which is available to it. This may exclude some cases in which there have been changes in covalent structure since biosynthesis.

The magnitude of the task requires a computer-aided approach. Firstly, a suitable representation of the covalent structure of the molecule is provided as input or is generated by the computer from the specified amino-acid sequence using, for example, a dictionary of data for the residue types and the tree concept of structure representation.[8] Secondly, the conformational variables may be split into two data sets, one containing those bond-rotation angles which may be usefully varied, and the other containing the bond lengths and the angles between the bonds. This division is arbitrary but useful, since the former set represents both those variables that are most usually adjusted or preset by operator intervention, and the 'soft variables' in the sense used by Scheraga.[3] Thirdly, the conformational energy is calculated as the sum of intrinsic bond-rotation potentials and of van der Waals, electrostatic, and hydrogen-bonding energies, estimated for every pair of interacting atoms. Because of theoretical and computational difficulties, interactions are treated as pair interactions despite the possibility that more-exact calculations may involve a many-body potential, as has been suggested for the van der Waals potential.[9] A complete calculation should also take account of interactions with the solvent. Finally, the energy calculation is repeated again and again under the control of a procedure for minimizing the free energy of the molecule with respect to the conformational variables.

The principal problem in the above procedure is the complexity of the energy surface, which contains many minima. The search for better procedures for locating the global minimum continues, and a recent method [10] is applicable to a discontinuous energy surface such as might in practice be encountered when taking the solvent conformation into account. The rapid increase in the complexity of the energy surface with molecular size accounts for the continued popularity of the small cyclic peptide gramicidin S as a testing ground for different computational methods.[11]

It is necessary to note that a great deal of the work which has been done is based on an essentially static view of polypeptide conformation, whereas there are also internal vibrations and possibly also bond rotations which should be included in the free energy.[3, 12] The reader is also referred to recent work [13, 14] on the phonons of α-helices for a more dynamic picture.

[6] C. B. Anfinsen, *Brookhaven Symp. Biol.*, 1962, **15**, 184.
[7] C. B. Anfinsen, *Harvey Lectures*, 1967, **61**, 95.
[8] J. Hermans and D. Ferro, *Biopolymers*, 1971, **10**, 1121.
[9] I. R. McDonald and K. Singer, *Ann. Reports (A)*, 1970, **67**, 45.
[10] S. N. Ghani, *Computer Aided Design*, 1972, **4**, 71.
[11] P. De Santis and A. M. Liquori, *Biopolymers*, 1971, **10**, 699.
[12] D. A. Brand, W. G. Miller, and P. J. Flory, *J. Mol. Biol.*, 1966, **23**, 47.
[13] K. Itoh and T. Shimanouchi, *Biopolymers*, 1970, **9**, 383.
[14] B. Fanconi, E. W. Small, and W. L. Peticolas, *Biopolymers*, 1971, **10**, 1277.

B. Solvent Effects.—There is currently no detailed understanding of the conformation of water around specific groups in specific conformations, and it remains to be seen just how good this understanding needs to be before satisfactory predictions of polypeptide conformation can be made. The most useful information at present in the prediction of the conformation of globular proteins is that the interior of a globular protein tends to consist of well-packed non-polar groups.[15] Such a tendency is usually interpreted as the formation of a large number of 'hydrophobic interactions' between non-polar groups as a necessary prerequisite for a stable globular structure. Studies on the nature of this bond continue.[16,17] For the purpose of estimating the decrease in free energy when a group is transferred to the non-polar interior of a protein, an empirically derived scale of hydrophobicities has recently been constructed.[18] The restriction which a distribution of polar and non-polar residues places on the geometry of the stably folded protein molecule, and the exposure to the solvent of the two classes, has been considered.[19] These considerations suggest that introducing an artificial force of attraction between non-polar groups might be a reasonable choice for a first approximation. However, in using such an approximation it should not be forgotten that the van der Waals and hydrogen-bonding interactions with the solvent, as well as the high dielectric constant of the water environment, are all more or less independent factors which should be considered in their own right.

A great deal of work has been carried out neglecting solvent effects, *i.e. in vacuo*. However, although the system whose free energy is to be minimized consists both of the polypeptide and its surrounding solvent, the predominating importance of the van der Waals repulsive forces between the atoms of the polypeptide,[2] coupled with the fact that many parts of the polypeptide chain for which alternative possible conformations are to be compared are within the interior of the protein and away from the solvent, means that useful information can be obtained even when the solvent is neglected.

C. Studies on the Conformation of Single Residues using Energy Functions.—Because of the problem of multiple minima, there has been considerable interest in studying the conformational energies of individual residues; these continue to be *in vacuo* calculations. At the very least, such studies point to high-energy conformations for each residue which should be excluded from the search for the global energy minimum except as a last resort. At best, they may point to a small number of well-defined minima for the conformation of each residue, of which one might hopefully be the conformation of that residue in the stably folded protein. In the latter case,

[15] B. Lee and F. M. Richards, *J. Mol. Biol.*, 1971, **55**, 379.
[16] S. Lewin, *Nature*, 1971, **231**, 80.
[17] A. Ray, *Nature*, 1971, **231**, 313.
[18] Y. Nozaki and C. Tanford, *J. Biol. Chem.*, 1971, **246**, 2211.
[19] R. E. Gates and H. F. Fisher, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2928.

we have the notion of a stereochemical code.[20] Conformational plots for single residues may be represented as contours of equal energy drawn on a plot of the dihedral angle ϕ around the bond N—C^{α} *vs.* the dihedral angle ψ around the bond C^{α}—CO.

The progress in developing ϕ–ψ plots over the past two years has been principally in extending the study to all types of side-chains which occur in proteins and to the effects on the ϕ–ψ plots of side-chain conformation.[21–23] Similar plots based on quantum-mechanical calculations [24–28] have also been developed in the same direction.[29–31] Generally speaking, the amide bond continues to be treated as fixed, planar, and *trans*, despite the possibility that *X*-ray crystallographic data on lysozyme is not incompatible with deviations from this [32] and despite experimental observations on small peptide analogues.[33] An empirical ϕ–ψ diagram, based upon the observed distribution of residue conformations in proteins of known conformation, has been devised for a generalized side-chain.[34]

D. Predictions of Local Conformations in Polypeptides.—It is possible that the global minimum of the conformational energy of a polypeptide does not need to be found before the conformation of a particular local region of the polypeptide chain can be predicted. This would be the case if the conformation of the local region were determined entirely by interactions within that region. Moreover, such a region may have a nucleating function, guiding the folding of the rest of the polypeptide chain around it. Intuitive reasoning continues to support the concept of such nucleation sites as necessary requirements for the attainment of the native conformation in a sufficiently short period of time, for the protein in nature as well as for the protein in the computer.[35]

For this reason there has been great interest in making predictions of local conformations based on statistical analysis of the available data from proteins of known sequence and conformation. Until recently, such studies were mainly confined to the conformation which occurred most frequently in the proteins of known conformation, the right-hand α-helix.

[20] A. M. Liquori, *Quart. Rev. Biophys.*, 1969, **2**, 65.
[21] V. Sasisekharan and P. K. Ponnuswamy, *Biopolymers*, 1970, **9**, 1249.
[22] P. K. Ponnuswamy and V. Sasisekharan, *Biopolymers*, 1971, **10**, 565.
[23] V. Sasisekharan and P. K. Ponnuswamy, *Biopolymers*, 1971, **10**, 583.
[24] R. Hoffman and A. Imamura, *Biopolymers*, 1969, **7**, 207.
[25] B. Pullman, 'Aspects de la Chimie Quantique Contemporaire', ed. R. Daudel and A. Pullman, Colloque International du CNRS, 1971, p. 263.
[26] B. Maigret, D. Perahia, and B. Pullman, *J. Theor. Biol.*, 1970, **29**, 275.
[27] B. Pullman, B. Maigret and D. Perahia, *Compt. rend.*, 1970, **270**, *D*, 1395.
[28] B. Pullman, B. Maigret, and D. Perahia, *Theor. Chim. Acta*, 1970, **18**, 99.
[29] B. Maigret, B. Pullman, and D. Perahia, *Biopolymers*, 1971, **10**, 107.
[30] B. Maigret, D. Perahia, and B. Pullman, *Biopolymers*, 1971, **10**, 491.
[31] B. Maigret, D. Perahia, and B. Pullman, *Biopolymers*, 1971, **10**, 1649.
[32] A. C. T. North, *Biochem. J.*, 1971, **125**, 86P.
[33] R. Lumley Jones, *Trans. Faraday Soc.*, 1970, **66**, 2491.
[34] F. M. Pohl, *Nature New Biology*, 1971, **234**, 279.
[35] Molecular Biology Correspondent, *Nature*, 1971, **233**, 523.

It may be noted that they were also restricted to identifying as helical those residues in the available data which occurred within at least one recognizable turn of α-helix, not only because the publication of such data has preceded the availability of good ϕ,ψ data by a considerable time, but because the property of interest to most workers seems to be the ability to 'help make' a right-hand α-helix in a particular situation. In addition, related studies have been made using values for the percentage helix in proteins obtained by o.r.d. measurements rather than from *X*-ray crystallographic data.[36] Thus they are really referring to a particular *helical state* of a residue which is something more than a particular pair of ϕ,ψ angles. More recently, studies have been made on other local conformations, namely the pleated sheet [37] and the β-fold.[38]

The studies on statistical methods of predicting α-helical regions can be more generally viewed as an attempt to break the biological code relating amino-acid sequence to the α-helical conformation, and two main conclusions have been drawn concerning the way information is carried by this code. Information has been shown to reside in single residues on the one hand,[39–42] and in the arrangement of residues along the backbone on the other;[41, 43–45] both conclusions lead to reasonable predictions of observed helical regions. By applying an information theory approach, it has been possible to measure just how much information each residue provides for its own tendency to be helical or non-helical, and how much it receives from its near neighbours.[46, 47] Whereas the information in single residues can be identified with the relative magnitude of the energy of the right-hand α-helical region on the ϕ–ψ diagram,[48] the nature of the information provided by neighbours is less well understood. It must, however, by definition reside in residue–residue interactions in which either one or both side-chains is important. Some work emphasizes the nature of both side-chains,[43, 44, 49, 50] while observations [41, 51] that residues are not symmetrically distributed between the N- and C-termini of helices has led to

[36] D. E. Goldsack, *Biopolymers*, 1969, **7**, 299.
[37] A. U. Finkelstein and O. B. Ptitsyn, *J. Mol. Biol.*, 1971, **62**, 613.
[38] P. N. Lewis, H. Momany, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2293.
[39] A. V. Guzzo, *Biophys. J.*, 1965, **5**, 809.
[40] J. W. Prothero, *Biophys. J.*, 1966, **6**, 367.
[41] O. B. Ptitsyn, *J. Mol. Biol.*, 1969, **42**, 501.
[42] D. Kotelchuck and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 14.
[43] P. F. Periti, G. Quagliatori, and A. M. Liquori, *J. Mol. Biol.*, 1967, **24**, 313.
[44] M. Schiffer and A. B. Edmunson, *Biophys. J.*, 1967, **7**, 121.
[45] B. W. Low, F. M. Lovell, and A. D. Rudko, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 1519.
[46] R. H. Pain and B. Robson, *J. Mol. Biol.*, 1971, **58**, 237.
[47] R. H. Pain and B. Robson, *Proceedings of the 1st European Biophysics Congress*, 1971, **1**, 33.
[48] B. Robson and R. H. Pain, *Nature*, 1970, **227**, 62.
[49] T. Iio, *Biopolymers*, 1971, **10**, 1583.
[50] D. Kotelchuck, M. Dygert, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **63**, 615.
[51] D. A. Cook, *J. Mol. Biol.*, 1967, **26**, 143.

the suggestion [37, 41] that side-chain backbone interactions may be the important factor.

Another approach to the coding of the α-helix lies in obtaining empirical data from experimental studies on artificial polypeptides in solution.[52, 53] Studies to obtain the appropriate parameters for predicting helical regions in globular proteins continue;[54–57] note that these parameters are the analogue of single-residue information. These solution studies emphasize that a locally determined conformation prior to formation of the globular protein may be the most stable conformation other than the time-random coil, but that the time-random coil may be more stable. Tertiary interactions must therefore have a stabilizing influence. These factors have been experimentally investigated.[58]

E. Predictions of the Stability of Globular Proteins.—The problem of predicting the stability of a globular protein in its native state is clearly related to finding the conformation of lowest energy. It does not necessarily follow, however, that the stereochemistry must be predicted in detail before reasonable predictions can be made. This is the case if the protein is treated as a monomolecular micelle with non-polar groups tending to be inside and with polar groups outside.

If the protein is to be treated as a micelle of this type with otherwise indeterminate internal stereochemistry, predicting the stability of the protein would seem to be a relatively simple task. However, this would only be true if exact measures of the factors involved were available, for the stability of a globular protein depends upon a small difference between countering forces of large magnitude. More specifically, the principal factors which contribute to the free energy of the native state with respect to the fully denatured, time-random coil are the difference in configurational entropy, and the change in free energy in going from polypeptide–water to intramolecular polypeptide contacts. More precise studies should take into account changes in electrostatic effects. Both Brandts [4] and Tanford [59] have attempted to obtain estimates of the stability of globular proteins and the principal points of disagreement are in the change in configurational entropy of the backbone and the relative stabilities of polypeptide–water and intramolecular polypeptide hydrogen-bonds. The latter discrepancy is mainly due to the apparent inconsistency between measurements from hydrogen-bonded low-molecular-weight compounds and studies on the helix–coil transitions of artificial polypeptides.[59]

[52] D. Poland and H. A. Scheraga, 'Theory of Helix–Coil Transitions in Biopolymers', Academic Press, New York, 1970.
[53] P. N. Lewis, N. Gō, M. Gō, D. Kotelchuck, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1970, **65**, 810.
[54] A. Roig and M. Cortijo, *Biopolymers*, 1971, **10**, 321.
[55] A. Roig, F. G. Blanco, and M. Cortijo, *Biopolymers*, 1970, **10**, 329.
[56] D. Puett and A. Ciferri, *Biopolymers*, 1971, **10**, 547.
[57] V. E. Bychova, O. B. Ptitsyn, and T. F. Barskaya, *Biopolymers*, 1971, **10**, 2181.
[58] J. Hermans and D. Puett, *Biopolymers*, 1971, **10**, 895.
[59] C. Tanford, *Adv. Protein Chem.*, 1970, **24**, 1.

2 Spin Labels

contributed by N. C. Price

Spin-labelling has continued to be a widely used approach to examine various aspects of protein structure and function. Several excellent reviews of the technique are available.[60–62] This section deals with the literature which appeared in 1971.

A. Haemoglobin.—Haemoglobin (Hb) has been perhaps the most intensively studied protein. Since the *X*-ray diffraction data on both oxy- and deoxy-Hb are available,[63] many of the spin-labelling studies can be interpreted and discussed in considerable detail. McConnell has published a review article on the application of the spin-labelling technique to the study of co-operative oxygen binding to Hb.[64] In this he discusses the earlier experiments in which the β93-Cys residues of the protein were allowed to react with spin-label derivatives of iodoacetamide.[65, 66] As mentioned in last year's Report,[67] the interpretation of some of these experiments has been complicated by the finding (from *X*-ray studies) that the introduction of the bulky spin labels at the β93 residues can cause considerable distortion of the Hb structure.[68] Two orientations of the spin label with respect to the protein molecule were found in the *X*-ray studies, corresponding to the weakly and strongly immobilized components of the e.s.r. spectrum of the label.[66] In one of the orientations the label is essentially free in solution on the surface of the protein; in the other orientation it displaces the amino-acid residues at the C-terminus of the β-chain. McConnell also describes [64] some hitherto-unpublished experiments in which a spin-label triphosphate derivative was shown to bind to ferrous Hb with a stoicheiometry of one label per tetramer and a dissociation constant very similar to that for diphosphoglyceric acid, a well-known effector of Hb. By analogy with other organic phosphate derivatives, the spin label was assumed to bind close to the dyad symmetry axis of Hb. The label could detect a conformational change caused by the addition of carbon monoxide to the α_2(β-ferric cyanide)$_2$Hb molecule.

A new approach to the study of the ligand-induced structural changes in Hb, and the relationship of these to the co-operativity between ligand sites, has been reported by Asakura *et al.*[69] This involved the use of a spin-labelled protohaem, in which either the 6- or 7-positions, or both, of the

[60] C. L. Hamilton and H. M. McConnell, in 'Structural Chemistry and Molecular Biology', ed. A. Rich and N. Davidson, Freeman, San Francisco, 1968, p. 115.
[61] O. H. Griffith and A. S. Waggoner, *Accounts Chem. Res.*, 1969, **2**, 17.
[62] H. M. McConnell and B. G. McFarland, *Quart. Rev. Biophys.*, 1970, **3**, 91.
[63] M. F. Perutz, *Nature*, 1970, **228**, 726.
[64] H. M. McConnell, *Ann. Rev. Biochem.*, 1971, **40**, 227.
[65] S. Ogawa and H. M. McConnell, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 19.
[66] H. M. McConnell, S. Ogawa, and A. Horwitz, *Nature*, 1968, **220**, 787.
[67] N. C. Price, in 'Amino-acids, Peptides, and Proteins', ed. G. T. Young (Specialist Periodical Reports), The Chemical Society, London, 1971, **3**, 205.
[68] J. K. Moffat, *J. Mol. Biol.*, 1971, **55**, 135.
[69] T. Asakura, J. S. Leigh, jun., H. R. Drott, T. Yonetani, and B. Chance, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 861.

haem were labelled. These haem derivatives were then combined with the apoproteins of myoglobin, Hb, cytochrome *c* peroxidase, and horse-radish peroxidase. The e.s.r. spectra of the spin-labelled myoglobin and Hb were similar (the labels being weakly immobilized), indicating the similarity of the haem environments in these two proteins. In cytochrome *c* peroxidase, the label was quite strongly immobilized. Horse-radish peroxidase showed only one very broad line in the e.s.r. spectrum, which did not appear to result from simple immobilization of the label by the protein, since no similar spectrum could be observed by varying the viscosity of the medium of the free spin-label. Asakura *et al.* concluded that the spectrum in horse-radish peroxidase arose from interactions between the two spin labels on the protohaem group. In the other haemoproteins this interaction does not arise, presumably because the haem pockets in these proteins are sufficiently large to keep the two labels apart. Spin–spin interactions between the iron of the haem and the spin labels (of the type discussed by Leigh [70]) could be detected in these haemoproteins after the addition of ligands (fluoride, azide, or cyanide). Using the Leigh theory,[70] the distances between the iron atom and the labels could be calculated. In the case of haemoglobin, this calculated distance agrees well with that estimated from the *X*-ray structure.

Asakura and Drott have also used these spin-labelled haem derivatives to obtain evidence for haem–haem interactions in Hb.[71] Two kinds of valency hybrid haemoglobins containing spin-labelled ferric haem in one type of subunit were prepared. The e.s.r. spectra of the (α-spin label Fe^{3+})$_2$ (β Fe^{2+}-O_2)$_2$ and (α Fe^{2+}-O_2)$_2$(β-spin label Fe^{3+})$_2$ haemoglobins were different, suggesting that the two chains have non-equivalent conformational properties in the Hb molecule. Deoxygenation of the α-labelled Hb altered the e.s.r. amplitude and lineshape reversibly, whereas that of the β-labelled Hb altered the lineshape. This change in the spectrum of the α-labelled Hb was explained by a change either in the haem–label distance or in the relaxation time of the haem iron in the α subunit on oxygenation of the β subunit.

Haptoglobin (an α-globulin) is known to bind to haemoglobin, forming a 1 : 1 complex and resulting in a marked alteration of the haem function (notably an increase in peroxidatic activity and the ability to halogenate certain organic molecules).[72] The spin-labelling method was used together with optical spectra, circular dichroism, and other e.s.r. measurements to assess the extent of alteration of the Hb structure on complex formation.[72] When Hb spin-labelled at the β93-Cys residues was added to haptoglobin, the e.s.r. spectrum of the label showed no indication of the equilibrium between strongly and weakly immobilized states of the spin label which exists in oxy- (but not deoxy-) Hb.[66] It was concluded that the F helix of

[70] J. S. Leigh, jun., *J. Chem. Phys.*, 1970, **52**, 2608.
[71] T. Asakura and H. R. Drott, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 1199.
[72] M. W. Makinen and H. Kon, *Biochemistry*, 1971, **10**, 43.

the β-chain of Hb assumes a deoxy-like conformation on complex formation, independent of the state of oxidation or ligation of the haem iron.

B. Enzyme Binding Sites.—*Creatine Kinase.* The studies on spin-labelled creatine kinase first reported in 1969 [73] have now been considerably extended.[74, 75] An iodoacetamide-type spin label was allowed to react with the 'essential' cysteine residue of the enzyme. The e.s.r. spectrum of this labelled (inactive) enzyme suggested that the label was highly immobilized.[73] Addition of adenosine 5′-diphosphate (ADP) or adenosine 5′-triphosphate (ATP) to the labelled enzyme (designated E·) resulted in small changes in the e.s.r. spectrum; no change was observed on the addition of creatine or creatine phosphate. In the presence of the obligatory bivalent metal ion, the nucleotides caused larger changes. Thus MgADP decreased the amplitude of the centre peak of the e.s.r. spectrum by about 20% and shifted the outer peaks. It was found, however, that the enhancement of the relaxation rate of water protons by the spin label attached to the enzyme was a more sensitive parameter with which to study the effects of ligands. Dissociation constants for the ternary complexes (*e.g.* E·–Mg^{2+}–ADP) were determined by this method and differences between the ternary complexes with ADP and ATP could be demonstrated. These differences in the properties of the various complexes are consistent with the conformational differences previously postulated on the basis of the reactivity of the essential cysteine residue.[76] Titrations of E· with the Mn–nucleotide complexes were also performed.[75] In these cases, the binding of Mn^{2+} leads to a decrease in the amplitude of the e.s.r. spectrum of the labelled enzyme, with no change in line shape. Dissociation constants for MnADP and MnATP from the respective ternary complexes were evaluated as 0.064 mmol l^{-1} and 0.13 mmol l^{-1}; these values are very close to the corresponding dissociation constants for the unmodified enzyme. Using the theory developed by Leigh to describe the interaction of two spins in a rigid matrix,[70] the distances between Mn^{2+} and the spin label in the ternary complexes were evaluated.[75]

Glyceraldehyde 3-Phosphate Dehydrogenase. Some of the complexities which can arise in interpreting spin-labelling experiments are illustrated by the work with glyceraldehyde 3-phosphate dehydrogenase.[77] This enzyme has been under intensive investigation, particularly since the observation that negative co-operativity is displayed between the coenzyme binding sites.[78, 79] Several types of spin labels were used in this study, some of varying chain length. They included analogues of iodoacetic acid (an

[73] J. S. Taylor, J. S. Leigh, jun., and M. Cohn, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **64**, 219.
[74] J. S. Taylor, A. McLaughlin, and M. Cohn, *J. Biol. Chem.*, 1971, **246**, 6029.
[75] M. Cohn, H. Diefenbach, and J. S. Taylor, *J. Biol. Chem.*, 1971, **246**, 6037.
[76] W. J. O'Sullivan and M. Cohn, *J. Biol. Chem.*, 1966, **241**, 3116.
[77] W. Balthasar, *European J. Biochem.*, 1971, **22**, 158.
[78] A. Conway and D. E. Koshland, jun., *Biochemistry*, 1968, **7**, 4011.
[79] J. J. M. de Vijlder and E. C. Slater, *Biochim. Biophys. Acta*, 1968, **23**, 167.

irreversible inhibitor of the enzyme) and two types of substrate analogue; all of these should react at the 'essential' Cys-149 residue. In addition, a non-covalently bound spin-label analogue of the coenzyme β-nicotinamide adenine dinucleotide (NAD^+), previously used in studies on alcohol dehydrogenase,[80, 81] was employed. The inhibitor and substrate analogues reacted with the enzyme, but the resulting e.s.r. spectra were exceedingly complex. They appeared to represent the superposition of at least three different types of spectra (corresponding to weak, strong, and intermediate immobilization of the labels). The various spin-labelled enzyme derivatives differed only in the relative proportions of these three types of spectra. Addition of NAD^+, which bound weakly to the modified enzymes, altered only the relative amounts of the three types. It was concluded that the active site geometry of the enzyme is complex and cannot be described as a simple cleft, but rather as a system of narrow and wider spaces which is altered by the binding of coenzyme.

Glycogen Phosphorylase b. Bennick *et al.* used spin-labelling in conjunction with n.m.r. and fluorescence measurements to determine the spatial relationship between the active and allosteric sites of glycogen phosphorylase *b*.[82] A single cysteine residue per subunit (molecular weight 92 500) was spin-labelled; the resulting e.s.r. spectrum was characteristic of a fairly mobile radical, whose rotational correlation time (3×10^{-9} s) was 30 times shorter than that for the whole protein molecule. The labelled enzyme was fully active and had an adenosine 5'-monophosphate (AMP) activation curve identical to that of the unmodified enzyme, indicating that both the active sites and the subunit interactions in the dimeric molecule were unaffected by the labelling. Addition of AMP changed the e.s.r. spectrum of the labelled enzyme; the extent of this change correlated well with the extent of activation of the enzyme. The unpaired electron of the label also caused broadening of the n.m.r. lines of substrate (glucose 1-phosphate) and activator (AMP) protons. From the magnitude of these effects, the distances between the label and these various protons were calculated, using the Solomon–Bloembergen equations.[83, 84] The resulting map of the various sites of the enzyme is of interest in that the active and allosteric sites are relatively close, compared with the overall dimensions of the protein molecule.[82]

Monomeric Enzymes. Hydrolytic enzymes, especially the serine esterases, have been quite extensively studied by the spin-label method,[85, 86] partly because the high reactivity of the 'essential' residues offers the possibility of

[80] A. S. Mildvan and H. Weiner, *Biochemistry*, 1969, **8**, 552.
[81] A. S. Mildvan and H. Weiner, *J. Biol. Chem.*, 1969, **244**, 2465.
[82] A. Bennick, I. D. Campbell, R. A. Dwek, N. C. Price, G. K. Radda, and A. G. Salmon, *Nature New Biology*, 1971, **234**, 140.
[83] I. Solomon, *Phys. Rev.*, 1955, **99**, 559.
[84] I. Solomon and N. Bloembergen, *J. Chem. Phys.*, 1956, **25**, 261.
[85] J. C. Hsia, D. J. Kosman, and L. H. Piette, *Biochem. Biophys. Res. Comm.*, 1969, **36**, 75.
[86] J. D. Morrisett, C. A. Broomfield, and B. E. Hackley, jun., *J. Biol. Chem.*, 1969, **244**, 5758.

selective attachment of spin labels to these enzymes. A note of caution, however, has been sounded recently by the results of Hoff *et al.*,[87] who studied the reaction of a substrate-like fluorophosphate spin label with the three enzymes atropinesterase, subtilisin, and α-chymotrypsin at pH values of *ca.* 8. In each case, a small amount of non-specific labelling occurred (*i.e.* two types of e.s.r. signal were observed). This side-reaction (which represented 1% of the total e.s.r. signal for atropinesterase and about 10% for the other two enzymes) could be minimized by using less than stoicheiometric amounts of spin label in the reaction mixture. For each enzyme, the major component in the e.s.r. spectrum arose from a strongly immobilized label, whose rotational correlation time was of the same order as that for the whole enzyme molecule. These results are in contrast to those reported by Morrisett *et al.*,[86] who found that the same label reacted specifically with acetylcholinesterase and α-chymotrypsin (at pH 4), to yield derivatives in which the label was only partially immobilized relative to the protein. Clearly, a thorough study of these labelling reactions under various conditions is required.

In one of the first applications of the spin-label method, it was shown that a spin-label derivative of *p*-nitrophenyl acetate acts as a substrate for α-chymotrypsin, and that at low pH the acyl enzyme intermediate could be isolated.[88] This acyl enzyme has now been crystallized, and the e.s.r. spectrum has been studied as a function of the orientation of the crystal to the applied magnetic field.[89] The α-chymotrypsin unit cell consists of four molecules, essentially arranged as two dimers. When the applied magnetic field was perpendicular to the two-fold screw axis relating the two dimers, and also either parallel or perpendicular to the two-fold rotation axis relating the units of the dimer, only a single molecular orientation was revealed in the e.s.r. spectrum of the crystalline labelled enzyme. This shows that the spin group takes on a single orientation relative to the host protein, a situation which contrasts with that in spin-labelled Hb, where the label takes on two isomeric orientations relative to the protein molecule.[66, 68]

A crystallographic study of the complex between lysozyme and a spin-label derivative of acetamide has been reported.[90] The label is an inhibitor of the enzyme, 50% inhibition occurring at a label concentration of 60 mmol l^{-1}. The *X*-ray study showed that the label binds at three sites on the enzyme. Two of these sites are relatively weak and occur at sub-sites A and C of the active-site cleft.[91] A third, stronger site was found as a hydrophobic pocket near the surface of the molecule around the four-fold screw axis. The biological significance of this third binding site is not clear; however, it would appear not to be an artifact of crystallization, since preliminary n.m.r. measurements of lysozyme in the presence of the spin label suggest that this hydrophobic pocket also exists in solution.

[87] A. J. Hoff, R. A. Oosterbaan, and R. Deen, *F.E.B.S. Letters*, 1971, **14**, 17.
[88] L. J. Berliner and H. M. McConnell, *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **55**, 708.
[89] L. J. Berliner and H. M. McConnell, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 651.
[90] L. J. Berliner, *J. Mol. Biol.*, 1971, **61**, 189.
[91] C. C. F. Blake, L. N. Johnson, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *Proc. Roy. Soc.*, 1967, **B 167**, 378.

C. New Spin Labels.—Cooke and Duke, in a short communication,[92] report the synthesis and some properties of a spin-label analogue of ATP (in which the amino-group at the 6-position is converted into thiol, which is subsequently caused to react with an iodoacetamide-type spin label). The spin-labelled nucleotide (designated NSL) could bind to actin and also cause superprecipitation of actomyosin, though under more forcing conditions than for ATP. In the presence of Mg^{2+}, NSL bound to creatine kinase, but subsequent addition of creatine caused no further changes in the e.s.r. spectrum. Detailed studies with this analogue are awaited, so that a more complete assessment of its ability to replace ATP in enzyme reactions can be made.

Elek and Keleti have described some studies with a new type of spin label.[93] This is an iron–nitric oxide complex (of average *g*-value 2.03) which can be bound to various amino-acids, particularly cysteine and histidine.[94] The cysteine and histidine complexes can be readily distinguished on the basis of their e.s.r. spectra. In the reaction of the complex with glyceraldehyde 3-phosphate dehydrogenase, it was shown that both cysteine and histidine residues of the enzyme took part in the reaction. Prior reaction of the cysteine groups (with *p*-chloromercuribenzoate or iodoacetic acid) hindered the development of the e.s.r. signal. Reaction with the complex inactivated the enzyme, the inhibition being competitive with respect to glyceraldehyde 3-phosphate but of a mixed type with respect to NAD^+ and phosphate. This new type of spin label may be of importance in that the unpaired electron is much closer to the amino-acid side-chain of the enzyme than is the case with the normal spin labels described in the rest of this section.

Other papers on spin labels appeared in 1971 in the following fields: theoretical aspects of correlation times,[95] studies with model compounds as hydrophobic probes,[96] polypeptide mobility,[97] protein microstructure,[98] muscle components,[99, 100] nucleic acids,[101] and membranes.[102–110]

[92] R. Cooke and J. Duke, *J. Biol. Chem.*, 1971, **246**, 6360.
[93] G. Elek and T. Keleti, *Acta Biochim. Biophys. Acad. Sci. Hung.*, 1971, **6**, 107.
[94] J. C. Woolum, E. Tiezzi, and B. Commoner, *Biochim. Biophys. Acta*, 1968, **160**, 311.
[95] G. Poggi and C. S. Johnson, jun., *J. Magn. Resonance*, 1970, **3**, 436.
[96] C. Jolicoeur and H. L. Friedman, *J. Phys. Chem.*, 1971, **75**, 165.
[97] A. Sanson and M. Ptak, *Compt. rend.*, 1970, **271**, *D*, 1319.
[98] G. I. Likhtenshtein, Yu. B. Grebenshchikov, P. Kh. Bobodzhanov, and Yu. V. Kokhanov, *Mol. Biol.* (*U.S.S.R.*), 1970, **4**, 682.
[99] J. C. Seidel, M. Chopek, and J. Gergely, *Arch. Biochem. Biophys.*, 1971, **142**, 223.
[100] R. W. Burley, J. C. Seidel, and J. Gergely, *Arch. Biochem. Biophys.*, 1971, **146**, 597.
[101] A. S. Girshovich, M. A. Grachev, D. G. Knorre, V. P. Kumarev, and V. I. Levinthal, *F.E.B.S. Letters*, 1971, **14**, 199.
[102] B. L. Bales and M. E. Baur, *Chem. Phys. Letters*, 1970, **7**, 341.
[103] J. C. Hsia, H. Schneider, and I. C. P. Smith, *Canad. J. Biochem.*, 1971, **49**, 614.
[104] J. C. Hsia, P. T. S. Wong, and D. H. MacLennan, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 88.
[105] J. Seelig and W. Hasselbach, *European J. Biochem.*, 1971, **21**, 17.
[106] J. K. Raison, J. M. Lyons, R. J. Mehlhorn, and A. D. Keith, *J. Biol. Chem.*, 1971, **246**, 4036.
[107] W. L. Hubbell and H. M. McConnell, *J. Amer. Chem. Soc.*, 1971, **93**, 314.
[108] J. F. W. Keana and R. J. Dinerstein, *J. Amer. Chem. Soc.*, 1971, **93**, 2808.
[109] P. Jost, L. J. Libertini, V. C. Herbert, and O. H. Griffith, *J. Mol. Biol.*, 1971, **59**, 77.
[110] O. H. Griffith, L. J. Libertini, and G. B. Birrell, *J. Phys. Chem.*, 1971, **75**, 3417.

3 Fluorescence

contributed by J. R. Brocklehurst

The importance of the various fluorescence techniques for studying structural problems in proteins and peptides is reflected in the increasing number of papers dealing with the techniques. Several reviews have appeared this year covering all the aspects to be dealt with below. A general review of fluorescence and its applications to proteins and membranes, which was mentioned last year, appeared during 1971.[111] Stryer has reviewed the uses of fluorescent probes in studying proteins and membranes,[112] while Radda has reviewed the use of several probe techniques as applied to studies of conformational changes in proteins.[113] Chen has reviewed the uses of depolarization and fluorescence lifetime studies in macromolecular chemistry.[114, 115] A book on the fluorescence and phosphorescence of proteins and nucleic acids has also appeared.[116] The proceedings of the 5th Johnson Foundation Colloquium were eventually published this year and contain a number of papers about the application of fluorescent probes, including a discussion of the interpretation of probe data [117, 118] and their applications to horse-liver alcohol dehydrogenase [119] and glutamate dehydrogenase.[120]

A. Interpretation of Fluorescence.—The problems of interpretation of fluorescence (and especially fluorescent probe) data have received a good deal of attention. Brand and his co-workers in particular have been active in examining the physical chemistry of various types of probe. The solvent-dependent changes in fluorescence of the aminonaphthalenesulphonate-type of probe are governed by solvent reorientation energy, and the quantum yield of fluorescence is inversely proportional to the singlet–triplet energy separation.[121] Intramolecular charge transfer is also an important factor in the fluorescence properties of these molecules.[122] By resolving the fluorescence emission spectrum of 1-anilinonaphthalene 8-sulphonate (ANS) on a nanosecond time-scale it is possible to see shifts in the emission peak

[111] G. K. Radda, in 'Current Topics in Bioenergetics', ed. D. R. Sanadi, Academic Press, New York, vol. IV, 1971, p. 81.
[112] L. Stryer, in 'Molecular Properties of Drug Receptors', CIBA Foundation Symposium, ed. R. Porter, Churchill, London, 1970, p. 133.
[113] G. K. Radda, *Biochem. J.*, 1971, **122**, 385.
[114] R. F. Chen, *Fluorescence News*, 1970, **5**, 1.
[115] R. F. Chen, *Analyt. Letters*, 1971, **4**, 459.
[116] S. V. Konev, 'Fluorescence and Phosphorescence of Proteins and Nucleic Acids', Plenum, London, 1971.
[117] B. Chance and G. K. Radda, in 'Probes of the Structure and Function of Macromolecules and Membranes', ed. B. Chance, C. P. Lee, and J. K. Blasie, Academic Press, New York, vol. 1, 1971, p. 11.
[118] L. Brand, C. J. Seliskar, and D. C. Turner, in ref. 117, p. 17.
[119] J. R. Gohlke, L. Brand, J. R. Heitz, and R. H. Conrad, in 'Probes of the Structure and Function of Macromolecules and Membranes', ed. B. Chance, T. Yonetani, and A. S. Mildvan, Academic Press, New York, vol. II, 1971, p. 37.
[120] J. R. Brocklehurst and G. K. Radda, in ref. 117, p. 59.
[121] C. J. Seliskar and L. Brand, *Science*, 1971, **171**, 799.
[122] C. J. Seliskar and L. Brand, *J. Amer. Chem. Soc.*, 1971, **93**, 5405.

owing to solvent reorientation during the excited-state lifetime of the fluorescing species. However, the shifts are not seen in very viscous media, which suggests that some ANS emission spectra in protein systems may give incorrect indications of local polarity.[123] Similar work has been done with 2-toluidinonaphthalene 6-sulphonate (TNS) bound to bovine serum albumin (BSA), showing how relaxation processes can be studied on a nanosecond time-scale.[124]

B. Fluorescent Probes.—*Non-covalently Bound.* ANS binding to a number of proteins has been investigated. In yeast alcohol dehydrogenase there are two binding sites which are distinct from the coenzyme site (unlike the liver enzyme), and ANS appears to report a conformational change induced by coenzyme binding.[125] ANS binds to butyryl and acetyl cholinesterases with a blue shift and enhancement of its fluorescence, and competitively inhibits both enzymes. There are two and three binding sites, respectively, and binding constants have been derived from the fluorescence data.[126] Rabbit-liver fructose 1,6-diphosphatase enhances ANS fluorescence on binding. The enhancement is quenched reversibly when the substrate is bound. This quenching is not due to a change in the ANS dissociation constant, so it is postulated that a conformational change is induced by the substrate.[127] In a similar way substrates, substrate analogues, and metal ions enhance the fluorescence of ANS bound to an Na^+- and K^+-stimulated ATPase.[128]

Various fluorescent probes have been used to look at structural changes in BSA following irradiation,[129] and to follow the kinetics of denaturation of horse-liver alcohol dehydrogenase.[130] Binding data suggest that the TNS response to the pepsin–pepsinogen conversion is due to a change in probe environment. The probe fluorescence is enhanced by the inhibitor peptide, which is released during activation.[131] A new class of polarity-sensitive probes has been synthesized from 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) and has also been shown to be useful for studying the activation of pepsinogen.[132] Other probes include 3-benzyl-4-methyl-7-diethylamino-coumarin, which binds to an 'aromatic' site in α-chymotrypsin uncovered by activation of chymotrypsinogen,[133] and chlorpromazine, which quenches aldolase fluorescence.[134]

One of the problems in the use of fluorescent probes has always been a lack of specificity in their binding, and the related difficulty of locating the

[123] S. K. Chakrabarti and W. R. Ware, *J. Chem. Phys.*, 1971, **55**, 5494.
[124] L. Brand and J. R. Gohlke, *J. Biol. Chem.*, 1971, **246**, 2317.
[125] F. M. Dickinson, *F.E.B.S. Letters*, 1971, **15**, 17.
[126] S. T. Christian and R. Janetzko, *Arch. Biochem. Biophys.*, 1971, **145**, 169.
[127] H. Aoe, M. G. Sarngadharan, and B. M. Pogell, *J. Biol. Chem.*, 1970, **245**, 6383.
[128] M. Mayer and Y. Avi-dor, *Israel J. Med. Sci.*, 1970, **6**, 726.
[129] C.-G. Rosen, *Internat. J. Radiation Biol.*, 1971, **19**, 587.
[130] J. R. Heitz and L. Brand, *Arch. Biochem. Biophys.*, 1971, **144**, 286.
[131] J. L. Wang and G. M. Edelman, *J. Biol. Chem.*, 1971, **246**, 1185.
[132] R. A. Kenner and A. A. Aboderin, *Biochemistry*, 1971, **10**, 4433.
[133] Y. Nishimura, O. Takenaka, and K. Shibata, *J. Biochem.* (*Japan*), 1971, **70**, 293.
[134] M. J. Seghatchian, *Biochem. Pharmacol.*, 1971, **20**, 683.

binding sites. (Two of the above papers are, however, exceptions.[135, 136]) Various ways have been found to overcome this problem. One is the use of a more-specific probe. Thus lanthanide ions have been used as fluorescent probes of transferrin, the number of binding sites being one or two depending on the ionic radius. These sites are at least 43 Å from the ferric ion site since no energy transfer is observed.[135] An analogue of ATP [Dns-$NH \cdot CH_2 \cdot CH_2 \cdot O(PO_3)_2PO_3$] acts as a normal substrate for heavy meromyosin ATPase. Its fluorescence is enhanced and blue-shifted on binding. Energy transfer from the protein to the substrate analogue is observed for the triphosphate derivative but not for the mono- or di-phosphates.[136]

Two different methods of locating fluorescein binding sites have been used. In studies of the binding of tetraiodofluorescein to dogfish lactate dehydrogenase (by absorption spectroscopy, *not* by fluorescence), the existence of a unique binding site was confirmed by *X*-ray studies.[137] The binding of fluorescein to BSA was investigated by equilibrium dialysis. The binding sites were then located by covalently attaching fluorescein isothiocyanate to the protein. The two sites are very close and the amino-acid sequence around the sites was obtained after isolation of the labelled peptides.[138]

The latter method has also been used to locate the reaction sites of dansyl chloride on bovine fibrinogen. The various sites involved the labelling of different amino-acids,[139] and on the basis of the fluorescence properties of the various dansyl amino-acids the heterogeneity in the fluorescence of the various conjugates could be explained.[140] Dansylated 3-aminotyrosine has been studied as a model for proteins modified by amination and subsequent dansylation.[141] Dansylation of an amino-tyrosine in trypsin and trypsinogen causes 50% inhibition. The activation process causes a red shift and quenching of the dansyl fluorescence (though there is no concomitant change in tryptophan fluorescence), suggesting a localized conformational change.[142]

Covalently Bound. Dansyl-cysteine covalently bound to actin acts as a probe of polymerization. The fluorescence of the probe is blue-shifted and enhanced on binding, and these changes are intensified by polymerization. From polarization studies a rotational relaxation time was calculated which was similar to the value expected for a spherical molecule of molecular weight 47 000.[143] The difficulties involved in this sort of calculation (particularly uncertainty in the measurement of fluorescence lifetimes)

[135] C. K. Luk, *Biochemistry*, 1971, **10**, 2838.
[136] M. Onodera and K. Yagi, *Biochim. Biophys. Acta*, 1971, **253**, 254.
[137] P. M. Wasserman and P. J. Lentz, *J. Mol. Biol.*, 1971, **60**, 509.
[138] L.-O. Andersson, A. Rehnström, and D. L. Eaker, *European J. Biochem.*, 1971, **20**, 371.
[139] E. Mihalyi and A. Albert, *Biochemistry*, 1971, **10**, 237.
[140] E. Mihalyi and A. Albert, *Biochemistry*, 1971, **10**, 243.
[141] R. A. Kenner, *Biochemistry*, 1971, **10**, 545.
[142] R. A. Kenner and H. Neurath, *Biochemistry*, 1971, **10**, 551.
[143] H. C. Cheung, R. Cooke, and L. Smith, *Arch. Biochem. Biophys.*, 1971, **142**, 333.

have been reviewed by Squire, who also suggested some ways of getting around the difficulties.[144] Fluorescence-polarization studies have also been used to calculate equilibrium and thermodynamic constants for the dansyl-α_{S1}-casein–κ-casein interaction. The authors were able to conclude that the interaction was largely hydrophobic in character.[145] Polarization of fluorescence measurements of a fluorescein mercuric acetate label at the active site have been used to measure isothermal uncoiling of lactate dehydrogenase.[146]

Another type of covalently attached probe is pyridoxal (or pyridoxal 5′-phosphate). Model studies suggested that this probe would report different states of ionization around its binding site.[147] In fact it seems to measure the proton-donating properties of its environment and distinguishes between L-aspartate aminotransferase and phosphorylase *b*.[148] Polarization studies of pyridoxal 5′-phosphate-labelled lysozyme indicate two rotational relaxation times, both of which are shorter than those obtained with the fluorescein-labelled enzyme, and which do not correlate with molecular weight data.[142] Model studies suggested that the shorter of the two was due to rotation of the Schiff base relative to the protein. An inhibitor-induced conformational change is suggested by the increase in polarization and lifetime of the bound pyridoxal on binding the inhibitor to lysozyme.[149]

C. Protein Fluorescence.—The use of the native fluorescence of the biological system is in many ways preferable to the above approaches, since it does not involve introducing a perturbant into the system. The native fluorescence of proteins arises from tryptophan, and to lesser extents tyrosine and phenylalanine. Studies of the isolated amino-acids can yield information useful in studying them in a protein environment. Thus 3-aminotyrosine and simpler aminophenols have been studied as models for 3-aminotyrosine in proteins.[150] The mechanisms of the effect of pH on the quantum yield and lifetime of tyrosine fluorescence have been elucidated, the rate of protonation–deprotonation in the excited state being very important.[151]

The effect of solvent on the fluorescence decay of indole has been studied to help to interpret the fluorescence of tryptophan in proteins.[152] The fluorescence of phosphorylase *b* is higher than (and blue-shifted relative to) that of a mixture of tryptophan and tyrosine of the same composition as the enzyme. Dissociation of the enzyme lowers its fluorescence, but only

[144] P. G. Squire, *Biochim. Biophys. Acta*, 1970, **221**, 425.
[145] R. Clarke and S. Nakai, *Biochemistry*, 1971, **10**, 3353.
[146] D. S. Markovich, N. V. Umrikhima, R. T. Fin, G. B. Krapivinskii, and M. V. Vol'kenshtein, *Doklady Akad. Nauk S.S.S.R.*, 1970, **194**, 1365.
[147] M. Arrio-Dupont, *Photochem. and Photobiol.*, 1970, **12**, 297.
[148] M. Arrio-Dupont, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 653.
[149] R. Irwin and J. E. Churchich, *Biochemistry*, 1971, **10**, 5329.
[150] R. W. Cowgill, *Photochem. and Photobiol.*, 1971, **13**, 183.
[151] M. Fayet and Ph. Wahl, *Biochim. Biophys. Acta*, 1971, **229**, 102.
[152] W. B. de Lauder and Ph. Wahl, *Biochim. Biophys. Acta*, 1971, **243**, 153.

urea causes a red shift.[153] A study of the tryptophan fluorescence of seventeen proteins has shown that there is no simple relation between the quantum yield of fluorescence and the degree of exposure to solvent, though after denaturation the quantum yields were spread over a much narrower range of values. This suggests that either the amino-acid sequence around tryptophan is important in determining its quantum yield, or that even after denaturation the tertiary structure around tryptophan is still intact.[154] U.v. inactivation of ribonuclease leads to quenching of tryptophan fluorescence, possibly by an S^{2-} group. Model studies showed that glutathione, dithioglycollic acid, *etc.* were indeed very efficient quenchers.[155]

Binding of Small Molecules. One very important use of protein fluorescence is for studying substrate and coenzyme binding. In the Shethra flavoprotein there is tyrosine-to-tryptophan energy transfer which is interrupted on binding the flavin. The subsequent quenching of tryptophan fluorescence is used to determine the kinetics of flavin binding.[156] In D-amino-acid oxidase the tryptophan fluorescence is progressively blue-shifted as the two flavin molecules bind to the apoenzyme, suggesting that the tryptophan is close to the active site and is shielded by the two flavins.[157] The binding of NAD^+ or its analogues to pig lipoyl dehydrogenase leads to quenching of bound flavin, though there is no direct correlation between this and the effect of NAD^+ on the diaphorase activity of the enzyme.[158] In a model flavinyl-aromatic amino-acid peptide the fluorophores are mutually quenched, the quenching becoming more efficient as the chain length decreases. The efficiency of flavin quenching is, in the order tryptophan > tyrosine > phenylalanine.[159]

Only one of the two tryptophans in each subunit of liver alcohol dehydrogenase is quenched by NADH or NAD^+, and titration of this quenching has been used to determine coenzyme binding constants. The quenching could be due to a specific interaction between coenzyme and an active-site tryptophan, or could be caused by a conformational change.[160] More-recent work on this system has shown that the quenching is a result of energy transfer to the bound coenzyme. This energy transfer is more efficient to the first NADH molecule which binds. Excitation energy appears to be transferred between different subunits. In this study the quenching was also used to measure velocity constants.[161]

[153] S. Damjanovich, S. Toth, and J. Sümegi, *Acta Biochim. Biophys. Acad. Sci. Hung.*, 1970, **5**, 279.
[154] M. J. Kronman and L. G. Holmes, *Photochem. and Photobiol.*, 1971, **14**, 113.
[155] S. Arian, M. Benjamini, J. Feitelson, and G. Stein, *Photochem. and Photobiol.*, 1970, **12**, 481.
[156] J. A. D'Anna and J. A. Tollin, *Biochemistry*, 1971, **10**, 57.
[157] F. Y.-H. Wu, S.-C. Tu, C.-W. Wu, and D. B. McCormick, *Biochem. Biophys. Res. Comm.*, 1970, **41**, 381.
[158] G. Su and J. E. Wilson, *Arch. Biochem. Biophys.*, 1970, **143**, 253.
[159] F. Y.-H. Wu and D. B. McCormick, *Biochim. Biophys. Acta*, 1971, **229**, 440.
[160] P. L. Luisi and R. Favilla, *European J. Biochem.*, 1970, **17**, 91.
[161] H. Theorell and K. Tatemoto, *Arch. Biochem. Biophys.*, 1971, **142**, 69.

The quenching of glyceraldehyde 3-phosphate dehydrogenase fluorescence by bound NAD^+ has been used to measure binding constants and the number of binding sites for NAD^+.[162, 163] The latter study showed that at a high enough NAD^+ concentration the fourth binding site can be detected,[163] in contrast to earlier work.[162] NAD^+ also quenches UDP-galactose-4-epimerase fluorescence. Reduction of the coenzyme causes no increase in quenching, but a 50% efficient energy transfer to NADH is detected.[164] The fluorescence of pyruvate kinase (yeast) is quenched by K^+, Mg^{2+}, and phosphoenol pyruvate, and even more so by fructose diphosphate. However, the fact that fructose diphosphate binding depends upon the presence of Mg^{2+}, *etc.*, suggests that FDP induces a conformational transition.[165] The fluorescence of Phe *t*RNA synthetase is specifically quenched by Phe *t*RNA but not by other *t*RNA's.[166] Similarly, valyl-*t*RNA synthetase fluorescence is quenched by valyl *t*RNA, valine, ATP, and Mg^{2+}. The effect of pH on the fluorescence and the binding constants suggested a possible interaction between the catalytic and *t*RNA binding sites.[167]

In contrast, isoleucine-*t*RNA synthetase fluorescence is not quenched by substrates; substrate binding was measured using the substrate inhibition of urea denaturation.[168] However, TNS binds to this enzyme without affecting its activity. The enhanced TNS fluorescence is partially quenched when substrates bind to the enzyme, thus providing a much more sensitive method for measuring binding parameters.[169]

Protein–Protein Interaction. The technique of fluorescence quenching has also been used to study protein–protein interactions. The tryptophan fluorescence of pepsin is quenched when suitable acceptors bind to the protein. The interaction between pepsin and haemoproteins was studied by using the haem group as acceptor. The influence of redox state, pH, *etc.*, suggest that the interaction is largely hydrophobic in nature, and that the haem group is remote from the area of contact between the proteins.[170]

Protein Structure. Protein fluorescence can also be used as a tool for studying protein structure. Human and bovine serum albumins have one and two tryptophans, respectively, and the fluorescence of these residues is one example of the difference between the two proteins. Binding of alkyl derivatives enhances HSA fluorescence but quenches that of BSA.[171] The one tryptophan in HSA displays two fluorescence lifetimes, but this could be

[162] S. F. Velick, J. P. Baggott, and J. M. Sturtevant, in 'Pyridine Nucleotide Dependent Dehydrogenases', ed. H. Sund, Springer, Berlin, 1970, p. 229.
[163] N. C. Price and G. K. Radda, *Biochim. Biophys. Acta.*, 1971, **235**, 27.
[164] A. U. Bertland, *Biochemistry*, 1970, **9**, 4649.
[165] R. T. Kuczenski and C. H. Suelter, *Biochemistry*, 1971, **10**, 2862.
[166] J. G. Farelly, J. W. Longworth, and M. P. Stulberg, *J. Biol. Chem.*, 1971, **246**, 1266.
[167] C. Helene, F. Brun, and M. Yaniv, *J. Mol. Biol.*, 1971, **58**, 349.
[168] G. R. Penzer, E. L. Bennett, and M. Calvin, *European J. Biochem.*, 1971, **20**, 1.
[169] E. Holler, E. L. Bennett, and M. Calvin, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 409.
[170] R. A. Badley and F. W. J. Teale, *J. Mol. Biol.*, 1971, **58**, 567.
[171] J. Steinhardt, J. Krijn, and J. G. Leidy, *Biochemistry*, 1971, **10**, 4005.

due to heterogeneity in the sample.[172] BSA binds almost 100 molecules of sodium dodecyl sulphate (SDS); the first twelve bind very rapidly, and only these twelve affect the protein fluorescence. The fluorescence is first quenched, then blue-shifted and slightly enhanced, as the twelve molecules of SDS are titrated on.[173] Studies of the effect of pH and iodide ion suggest that the quenching is due to a neighbouring amino-group, while the enhancement and blue shift reflect a slightly more hydrophobic environment as SDS binds.[174] Polarization of protein fluorescence (and hence rotational relaxation times) has been used to study the reversible dicarboxylic anhydride modification of BSA. Conditions can be chosen such that the protein is in its expanded form (involving exposure of lysine groups) and equivalent to the native protein at pH 2 (where it is irreversibly expanded).[175] Fluorescence studies of the tryptophans in BSA and HSA during conformational changes (induced by urea, SDS, dioxan, *etc.*) indicate that these residues are located in a conformationally labile hydrophobic fold of the protein with a positive group at the entrance to the fold.[176]

Tryptophan fluorescence was one of the properties used to compare a semi-synthetic staphylococcal nuclease with the native enzyme.[177] This enzyme undergoes an acid–base induced unfolding and refolding. The kinetics of refolding have been followed by measuring the tryptophan fluorescence in a stopped-flow apparatus.[178] The results obtained by this method correlated well with n.m.r. studies of the histidine residues of the enzyme.[179] Changes in fluorescence have been used to follow pH-induced conformational changes in native [180] and insolubilized [181] chymotrypsin, and to examine changes in tryptophan environment following heat denaturation of BSA and trypsin.[182] Ion quenching of tryptophan has been used in structural studies of glucagon,[183] lysozyme,[184] and several albumins.[185]

An example of the way in which several fluorescence techniques may be used to tackle the same problem is a study of the apomyoglobins from *Aplysia* and sperm whale. Both proteins have two tryptophans, but fluorescence polarization and spectral data show that the two in the

[172] W. B. de Lauder and Ph. Wahl, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 398.
[173] C. J. Halfman and T. Nishida, *Biochim. Biophys. Acta*, 1971, **243**, 294.
[174] C. J. Halfman and T. Nishida, *Biochim. Biophys. Acta*, 1971, **243**, 284.
[175] A. Jonas and G. Weber, *Biochemistry*, 1970, **9**, 4729.
[176] M. N. Ivkova, N. S. Vedenkina, and E. A. Burshtein, *Mol. Biol.* (*U.S.S.R.*), 1971, **5**, 214.
[177] I. M. Chaiken, *J. Biol. Chem.*, 1971, **246**, 2948.
[178] H. F. Epstein, A. N. Schechter, R. F. Chen, and C. B. Anfinsen, *J. Mol. Biol.*, 1971, **60**, 499.
[179] H. F. Epstein, A. N. Schechter, and J. S. Cohen, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2042.
[180] J. R. Garel and B. Labouesse, *Biochimie*, 1971, **53**, 9.
[181] D. Gabel, I. Z. Steinberg, and E. Katchalski, *Biochemistry*, 1971, **10**, 4461.
[182] K. K. Turoverov and B. V. Shchelchkov, *Biofizika*, 1970, **15**, 965.
[183] H. Bornet and H. Edelhoch, *J. Biol. Chem.*, 1971, **246**, 1785.
[184] S. S. Lehrer, *Biochemistry*, 1971, **10**, 3254.
[185] E. A. Chernitskii and N. M. Kozlova, *Vestei Akad. Nauk Beloruss. S.S.R., Ser. bujal. Navuk*, 1971, **1**, 47.

Aplysia protein are in different environments, and that one is mobile and in contact with an aqueous environment. The two in sperm-whale apomyoglobin are both in a rigid, non-polar environment. ANS binds in the haem crevice of both, but its environment is more polar in the *Aplysia* protein. Despite these differences, rotational relaxation measurements with dansyl-labelled proteins indicate that both are approximately spherical with a rotational relaxation time of *ca.* 30 ns.[186]

In surveying the literature for 1971, it has been apparent (to the Reporter at least) that there is an increasing awareness of the limitations of fluorescence techniques and a desire to use them in a more meaningful way. The great interest in protein fluorescence is particularly refreshing since, despite the fact that it is technically more difficult to study than are fluorescent probes, the data are usually easier to interpret.

4 Mössbauer Spectroscopy

contributed by C. E. Johnson

The study of proteins by Mössbauer spectroscopy, while still not a large field, nevertheless showed a continuing progress of published papers in 1971. Virtually all the work done was confined to the isotope ^{57}Fe, and the studies were in the main directed towards obtaining information on the electronic state and the environment of iron at the active centres of biological molecules. The samples were usually in the form of a frozen aqueous solution of the molecules. An advantage of the Mössbauer technique in this kind of investigation is that it gives data local to the Mössbauer nucleus (*i.e.* ^{57}Fe) and does not depend upon a knowledge of the complete molecular structure. As in previous years, most of the work was concentrated on two kinds of molecule (*i*) proteins containing the haem group and (*ii*) the iron–sulphur proteins.

A complete and convenient source of information on this field is 'The Index of Publications in Mössbauer Spectroscopy of Biological Materials', compiled by Dr. Leopold May, Department of Chemistry, The Catholic University of America, Washington, D.C., 20017, U.S.A. The Index is kept up to date by additions three or four times a year, and has been the main source of references used in writing this section of this Report.

Review papers were given by Lang[187] at the International Conference on the Application of the Mössbauer Effect (Tihany, Hungary, 1969) and by Johnson[188] at the XVIth Conference on Magnetism and Magnetic Materials (Miami, U.S.A., 1970). A review article by May[189] on biological applications was included in a book on Mössbauer spectroscopy.

[186] S. R. Anderson, M. Brunori, and G. Weber, *Biochemistry*, 1970, **9**, 4723.
[187] G. Lang, 'Proceedings of Conference on the Application of the Mössbauer Effect, 1969', ed. I. Dezsi, Akad. Kiado, Budapest, Hungary, 1971.
[188] C. E. Johnson, *J. Appl. Phys.*, 1971, **42**, 1325; *Physics Today*, 1971, **24**, 35.
[189] L. May, in 'An Introduction to Mössbauer Spectroscopy', ed. L. May, Plenum Press, New York, 1971, p. 180.

A. Haem Proteins.—A comprehensive review of work up to 1970 has been given by Lang,[190] which includes data and discussions on compounds of haemoglobin, myoglobin, and cytochromes. In these proteins the immediate environment of the iron (the haem group) is well known and has been extensively studied, but the state of the iron atoms may be Fe^{3+} or Fe^{2+} and either high or low spin, depending upon the other ligands. This state is often not easily determined in a large molecule by conventional chemical methods.

New measurements have been carried out to study the equilibrium between the high- and low-spin states in the undecapeptide of ferricytochrome *c*.[191] Measurements on metmyoglobin and its derivatives have been made by Maeda *et al.*[192] An investigation at high pressures[193] shows that changes in the state of the iron are a function of pressure, and may be used to study the equilibrium of the various charge and spin states. In all of this work, very small changes in the electronic state of the iron associated with the biochemical activity of the proteins may be detected using the Mössbauer effect.

B. Iron–Sulphur Proteins.—The nature of the iron atoms in these proteins has been of great interest since the magnetic properties were first studied by e.s.r. measurements. In contrast to the haem proteins, the structure is not completely known (though the amino-acid sequences have been determined), but the state of the iron atoms seems to be similar in all of them. The proteins appear to be magnetic when they are in the reduced state, and the e.s.r. signal shows a spin $S = \frac{1}{2}$ and unusual g-values which are slightly anisotropic, centred about $g = 1.94$. No e.s.r. signal is detected when they are oxidized. A review of their properties, including Mössbauer spectroscopy, has been given by Tsibris and Woody.[194]

Mössbauer studies have centred on the observations and interpretation of the hyperfine interaction in the reduced (*i.e.* paramagnetic) proteins, the measurements being made at low temperatures and in large magnetic fields. Work has been centred on the simpler proteins which contain two iron atoms and two sulphur atoms per molecule, and which accept one electron per molecule on reduction. New data have been obtained on *Scenedesmus* and spinach ferredoxins,[195] on the two-iron ferredoxins from spinach, parsley, pig adrenal cortex, *Azotobacter vinelandii*, and *Clostridium*

[190] G. Lang, *Quart. Rev. Biophys.*, 1970, **3**, 1.
[191] R. Nassif, M. Sellers, and L. May, *Appl. Spectroscopy*, 1971, **25**, 150; L. May, R. Nassif, and M. Sellers, in 'Applications of Low Energy X- and Gamma-rays', ed. C. A. Ziegler, Gordon and Breach, London, 1971, p. 257.
[192] Y. Maeda, Y. Morita, and C. Yoshida: *J. Biochem.* (*Japan*), 1971, **70**, 509.
[193] D. C. Grenoble, C. W. Frank, C. R. Bargeron, and H. G. Drickamer, *J. Chem. Phys.*, 1971, **55**, 1633.
[194] J. C. M. Tsibris and R. W. Woody, *Co-ordination Chem. Rev.*, 1970, **5**, 417.
[195] K. K. Rao, R. Cammack, D. O. Hall, and C. E. Johnson, *Biochem. J.*, 1971, **122**, 257; C. E. Johnson, R. Cammack, K. K. Rao, and D. O. Hall, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 564.

pasteurianum,[196] and on adrenodoxin [197] and putidaredoxin.[198] The results on all of these proteins are similar (though not identical) and show that the molecules in the reduced state contain one Fe^{3+} atom with spin $S_1 = \frac{5}{2}$ and one Fe^{2+} atom with spin $S_2 = 2$; the two iron atoms in the molecule are coupled together antiferromagnetically to give a resultant spin $S = S_1 + S_2$ of $\frac{1}{2}$. This is in accord with observations of e.s.r. and magnetic susceptibility and confirms the model proposed by Gibson *et al.*[199] to explain the *g*-values. In the oxidized state, one electron is lost by the molecule, and it is assumed that the strong antiferromagnetic coupling between two ferric atoms each with $S = \frac{5}{2}$ results in a non-magnetic molecule. Thus the model of a two-iron centre which explains the spectroscopic properties of these proteins now seems to be well established.

Interest in these proteins will be heightened by the recent *X*-ray determination [200] of the molecular structure of one of them, the ferredoxin from the bacteria *Clostridium pasteurianum*, which contains eight iron atoms per molecule.

C. Other Proteins.—Further measurements have been made on haemerythrin, in its deoxy, oxy, and several oxidized (met) states.[201] This protein is a non-haem iron protein responsible for oxygen transport in the sipuncilid worm *Golfingia gouldii*. The molecule consists of eight subunits, each containing two iron atoms and, on oxygenation, each subunit binds one molecule of oxygen. Cyclic oxygenation and deoxygenation experiments were performed to distinguish between oxidized and oxygenated forms. Oxyhaemerythrin gave a Mössbauer spectrum consisting of two doublets of equal intensity. The iron atoms in the oxygenated protein were shown to have different environments and are antiferromagnetically coupled together so that each subunit is non-magnetic.

The methaemerythrin complexes were also shown to be non-magnetic, from the absence of hyperfine interaction in their Mössbauer spectra measured at low temperatures (4.2 K) and in large magnetic fields (30 kG). These compounds appear to contain pairs of antiferromagnetically coupled Fe^{3+} atoms in closely similar environments. The sign of the electric field gradient at the iron nuclei for the complex metisothiocyanatohaemerythrin was found to be negative. By comparing the postulated structures of

[196] W. R. Dunham, A. J. Bearden, I. T. Salmeen, G. Palmer, R. H. Sands, W. H. Orme-Johnson, and H. Beinert, *Biochim. Biophys. Acta*, 1971, **253**, 134; W. R. Dunham, G. Palmer, R. H. Sands, A. J. Bearden, H. Beinert, and W. H. Orme-Johnson, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1119.

[197] R. Cammack, K. K. Rao, D. O. Hall, and C. E. Johnson, *Biochem. J.*, 1971, **125**, 849.

[198] E. Münck, P. G. Debrunner, J. C. M. Tsibris, and I. C. Gunsalus, *Biochemistry*, 1972, **11**, 855.

[199] J. F. Gibson, D. O. Hall, J. H. M. Thornley, and F. R. Whatley, *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **56**, 987.

[200] L. C. Sieker, E. Adman, and L. H. Jensen, *Nature*, 1972, **235**, 42.

[201] K. Garbett, C. E. Johnson, I. M. Klotz, M. Y. Okamura, and R. J. P. Williams, *Arch. Biochem. Biophys.*, 1971, **142**, 574.

oxyhaemerythrin and deoxyhaemerythrin, a possible mechanism and kinetic behaviour for the oxygenation equilibrium has been suggested.

Measurements have also been reported on the iron storage proteins ferritin [202] and haemosiderin.[188]

5 Nuclear Magnetic Resonance

contributed by H. W. E. Rattle

Slowly but surely, n.m.r. is beginning to fulfil some of its early promise as a tool for the investigation of polypeptides and proteins. Notable among the 1971 papers is an increasing tendency to forsake 'straight' analysis of proton spectra in favour of techniques which sidestep lack of resolution, for example by considering resonances whose chemical shifts or relaxation times are perturbed by paramagnetic ions, by changing to the observation of other nuclei, notably ^{19}F and ^{13}C, or by selective deuteriation. These more subtle approaches show greater promise of producing meaningful biological information, in that they may be more readily tailored to the investigation of specific problems. The era of recording n.m.r. data simply because it was readily available, without regard for biological realities, appears at last to have ended.

A. Peptides and Polypeptides.—Two papers have appeared which may make the analysis of proton spectra easier in some cases; one, from Giessner-Prettre and Pullman,[203] presents curves of calculated ring-current shifts, whereas Ramachandran and Chandrasekaran [204, 205] have employed model compounds to compile tables of the α-C*H*–N*H* coupling constants and their variation with dihedral angle, which may well prove useful in the conformational analysis of smaller peptide molecules. Workers who are synthesizing peptides by solid-phase methods may be interested in a method developed by Bayer *et al.*,[206] in which the ^{19}F chemical shifts of trifluoroacetyl groups bonded to the end of the peptide at any stage of the synthesis provide the basis for a sensitive end-group analysis, enabling the purity of the product to be estimated. Small peptides and oligopeptides provide a rich field for structural analysis by n.m.r., as evidenced by papers such as that of Tonelli *et al.*,[207] whose n.m.r. data on the structure of the cyclic decapeptide antamanide are consistent with the predictions from energy calculations. Data useful to workers in this and other sections of the field have been produced by Cohen,[208] who gives shift data for some

[202] J. F. Boas and G. J. Troup, *Biochim. Biophys. Acta*, 1971, **229**, 68.
[203] C. Giessner-Prettre and B. Pullman, *J. Theor. Biol.*, 1971, **31**, 287.
[204] G. N. Ramachandran, R. Chandrasekaran, and K. D. Kopple, *Biopolymers*, 1971, **10**, 2113.
[205] G. N. Ramachandran and R. Chandrasekaran, *Biopolymers*, 1971, **10**, 935.
[206] E. Bayer, P. Hunziker, M. Mutter, R. E. Sievers, and R. Ullman, *J. Amer. Chem. Soc.*, 1972, **94**, 265.
[207] A. E. Tonelli, D. J. Patel, M. Goodman, F. Naider, H. Faulstich, and T. Wieland, *Biochemistry*, 1971, **10**, 3211.
[208] J. S. Cohen, *Biochim. Biophys. Acta*, 1971, **229**, 603.

oligopeptides containing aromatic residues, including some secondary structure effects on the aromatic resonances.

In the polypeptide field, the interesting left-handed aspartate helices have been the subject of several studies, notably by Bradbury *et al.*,[209] Temussi and Goodman,[210] Paolillo *et al.*,[211] and Silverman *et al.*[212] Chemical shifts of the α-C*H* and N*H* appear to differ between left- and right-handed versions of the same polymer, but there is no evidence of ring interactions in the left-handed helical forms of poly-β-benzyl-L-aspartate. Meanwhile, the old controversy about the origin of the separate peaks of the helical and random forms of polypeptides in trifluoroacetic acid–deuteriochloroform solutions has been revived by Tam and Klotz [213] and their contentions have been answered by Bradbury *et al.*[214] Two papers on poly-L-proline by Torchia and Bovey [215] and Torchia [216] cast some new light on this interesting polymer, while fresh evidence for the rapid internal motions of poly-β-alanine is produced by Applequist and Glickson.[217]

Small biological polypeptides, notably oxytocin and vasopressin, have been the subject of work by several groups.[218–220] Spectral assignments, information on hydrogen-bonding or the lack of it, and other conformational data are presented. Chemical-shift changes and conformational effects in the ^{13}C spectrum of valinomycin on complexing with potassium ions are reported.[221]

B. Proteins.—Moving on to proteins, evidence of the increasing availability of ^{13}C spectrometers is found in the presentation of the ^{13}C spectrum of lysozyme [222] and of the first fifteen residues of RNase, prepared synthetically.[223] Analysis of ^{13}C spectra may be assisted by the data presented by Voelter *et al.*[224] on chemical shift values for amino-acids, and by some

[209] E. M. Bradbury, B. G. Carpenter, C. Crane-Robinson, and H. Goldman, *Macromolecules*, 1971, **4**, 557.
[210] P. A. Temussi and M. Goodman, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1767.
[211] L. Paolillo, P. A. Temussi, E. Trivellone, E. M. Bradbury, and C. Crane-Robinson, *Biopolymers*, 1971, **10**, 2555.
[212] D. N. Silverman, G. T. Taylor, and H. A. Scheraga, *Arch. Biochem. Biophys.*, 1971, **146**, 587.
[213] I. M. Klotz and J. W. O. Tam, *J. Amer. Chem. Soc.*, 1971, **93**, 1313.
[214] E. M. Bradbury, P. Cary, C. Crane-Robinson, L. Paolillo, T. Tancredi, and P. A. Temussi, *J. Amer. Chem. Soc.*, 1971, **93**, 5916.
[215] D. A. Torchia and F. A. Bovey, *Macromolecules*, 1971, **4**, 246.
[216] D. A. Torchia, *Macromolecules*, 1971, **4**, 440.
[217] J. Applequist and J. D. Glickson, *J. Amer. Chem. Soc.*, 1971, **93**, 3276.
[218] V. J. Hruby, A. I. Brewster, and J. A. Glasel, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 450.
[219] P. H. Von Dreele, A. I. Brewster, H. A. Scheraga, M. F. Ferger, and V. Du Vigneaud, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1028.
[220] J. Feeney, G. C. K. Roberts, J. H. Rockey, and A. S. V. Burgen, *Nature New Biol.*, 1971, **232**, 108.
[221] M. Ohnishi, M. C. Fedarke, J. D. Baldeschweiler, and L. F. Johnson, *Biochem. Biophys. Res. Comm.*, 1972, **46**, 312.
[222] J. C. W. Chien and J. F. Brandts, *Nature New Biol.*, 1971, **230**, 209.
[223] M. H. Freedman, J. S. Cohen, and I. M. Chaiken, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 1148.
[224] W. Voelter, G. Jung, E. Breitmaier, and E. Bayer, *Z. Naturforsch.*, 1971, **26b**, 213.

measurements [225] of longitudinal relaxation time (T_1) which show a ten-fold shortening of T_1 in the rigid side-chain of lysine compared with that of the ε-carbon. The longer-established technique of using n.m.r. to follow the titration of histidine resonances in enzymes is this year represented by work on ribonuclease,[226] staphylococcal nuclease,[179] and human carbonic anhydrase.[227] The latter study is followed by one [228] which utilizes ^1H and ^{19}F resonance to establish association and dissociation constants for the binding of various carboxylate ligands to human carbonic anhydrase. Other inhibitor or substrate binding studies have been carried out on lysozyme [229–231] and on liver alcohol dehydrogenase,[232] chymotrypsin,[233] acetylcholinesterase,[234] and ribonuclease.[235] These all use the fairly well-established technique of observing differential broadening and shift of those enzyme and substrate/inhibitor resonances which are visible. A fairly new development is the use of a ^{19}F-containing substrate or inhibitor to provide an uncluttered spectrum in which to seek changes. An interesting development in this direction is reported by Raftery and Huestis,[236] who bound covalently a small ^{19}F-containing 'reporter' molecule near the active site of ribonuclease S to obtain information on conformational changes. Another approach to the study of enzyme binding is to use a substrate or inhibitor containing a paramagnetic ion which perturbs the spectrum of the protein in regions representing the binding site. Binding of metal ions to bovine serum albumin [237, 238] has been reported, and Krugh [239] describes the use of a paramagnetic analogue of ATP as a spin label in the study of the active site of DNA-polymerase.

Of course, many proteins contain one or more metal ions at the active site, and the changes induced in the protein spectrum by oxidation, reduction, or removal of the metal can be most informative. Studies of the binding of carbon monoxide [240] and n-butyl isocyanide [241] to haemoglobin, and work on the linkage between the haem groups of the different subunits

[225] A. Allerhand, D. Doddrell, V. Glushko, D. W. Cochran, E. Wenkert, P. J. Lawson, and F. R. N. Gurd, *J. Amer. Chem. Soc.*, 1971, **93**, 544.
[226] N. L. R. King and J. H. Bradbury, *Nature*, 1971, **229**, 404.
[227] R. W. King and G. C. K. Roberts, *Biochemistry*, 1971, **10**, 558.
[228] P. W. Taylor, J. Feeney, and A. S. V. Burgen, *Biochemistry*, 1971, **10**, 3866.
[229] P. W. Kent and R. A. Dwek, *Biochem. J.*, 1971, **121**, 11P.
[230] B. Capon, H. Ashton, and R. L. Foster, *Chem. Comm.*, 1971, 512.
[231] J. F. Studebaker, B. D. Sykes, and R. Wien, *J. Amer. Chem. Soc.*, 1971, **93**, 4579.
[232] R. H. Sarma and C. L. Woronick, *Res. Comm. Chem. Pathol. Pharmacol.*, 1971, **2**, 177.
[233] B. Capon and H. Ashton, *Chem. Comm.*, 1971, 513.
[234] G. Kato and J. Yung, *Mol. Pharmacol.*, 1971, **7**, 33.
[235] H. Ruterjans and O. Pongs, *European J. Biochem.*, 1971, **18**, 313.
[236] M. A. Raftery and W. H. Heustis, *Biochemistry*, 1971, **10**, 1181.
[237] J. L. Sudmeier and J. J. Pesek, *Inorg. Chem.*, 1971, **10**, 860.
[238] J. L. Sudmeier and J. J. Pesek, *Analyt. Biochem.*, 1971, **41**, 39.
[239] T. R. Krugh, *Biochemistry*, 1971, **10**, 2594.
[240] R. Cassoly, Q. H. Gibson, S. Ogawa, and R. G. Shulman, *Biochem. Biophys. Res. Comm*, 1971, **44**, 1015.
[241] T. R. Lindstrom, J. S. Olson, N. H. Mock, Q. H. Gibson, and C. Ho, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 22.

of this molecule [242] use this technique, while a great deal of information on the precise state of the iron in ferredoxins has emanated from the Du Pont laboratories.[243–245] Other studies on metal-containing proteins which may be noted concern liver alcohol dehydrogenase,[246] cytochrome *c*,[247, 248] and an interesting proton relaxation enhancement study on the contribution of manganese to the activity of arginase.[249]

After the excitement of papers containing results from selectively deuteriated proteins, very little further has appeared. However, Cohen *et al.*[250] have used partially deuteriated tyrosines in staphylococcal nuclease to simplify the spectrum for inhibition studies. Another contribution which stands out as a little unusual is that of Phillips *et al.*,[251] in which the indole NH resonances of the tryptophan in lysozyme are identified by various means, including deuterium exchange and chemical modification.

Among proteins, the histones stand out as unique in their relatively simple n.m.r. spectra, dominated as they are by a few types of residue, and by the asymmetric distribution of the residues along the molecule. Reports on the binding of histone fractions to DNA [252] and their interaction with each other [253] hold the promise of greater things to come.

Finally, a look back to one of the earliest applications of n.m.r., the binding of water to large molecules, shows that the method is still alive and well. Collagen,[254, 255] keratin,[256] and poly-α-amino-acids [257] have been studied.

All in all, 1971 will probably be looked on as a year of consolidation rather than of dramatic advance in the application of n.m.r. to biological molecules.

[242] D. G. Davis, T. R. Lindstrom, N. H. Mock, J. J. Baldassare, S. Charache, R. T. Jones, and C. Ho, *J. Mol. Biol.*, 1971, **60**, 101.
[243] M. Poe, W. D. Phillips, J. D. Glickson, and C. C. McDonald, *Proc. Nat. Acad. Sct. U.S.A.*, 1971, **68**, 68.
[244] J. D. Glickson, W. D. Phillips, C. C. McDonald, and M. Poe, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 271.
[245] M. Poe, W. D. Phillips, C. C. McDonald, and W. H. Orme-Johnson, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 705.
[246] R. L. Ward and J. A. Happe, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1444.
[247] G. E. Krejcarek, L. Turner, and K. Dus, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 983.
[248] R. K. Gupta and S. H. Koenig, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1134.
[249] H. Hirsch-Kolb, J. Helmut, and D. J. Greenberg. *J. Biol. Chem.*, 1971, **246**, 395.
[250] J. S. Cohen, M. Feil, and I. M. Chaiken, *Biochem. Biophys. Acta*, 1971, **236**, 468.
[251] W. D. Phillips, J. D. Glickson, and J. A. Rupley, *J. Amer. Chem. Soc.*, 1971, **93**, 4031.
[252] M. Boublik, E. M. Bradbury, C. Crane-Robinson, and H. W. E. Rattle, *Nature New Biol.*, 1971, **229**, 129.
[253] E. M. Bradbury, P. D. Cary, C. Crane-Robinson, P. L. Riches, and E. W. Johns, *Nature New Biol.*, 1971, **233**, 265.
[254] B.-M. Fung and P. Trautmann, *Biopolymers*, 1971, **10**, 391.
[255] A. A. Khanagov, *Biopolymers*, 1971, **10**, 789.
[256] L. J. Lynch and K. H. Marsden, *Search*, 1971, **2**, 95.
[257] I. D. Kuntz, *J. Amer. Chem. Soc.*, 1971, **93**, 514.

6 Infrared Spectroscopy

contributed by R. M. Stephens

A. Model Compounds, Amino-acids, and Oligopeptides.—*N*-Methylacetamide and its derivatives have for a long time been used as model compounds for interpreting the i.r. spectra of proteins and polypeptides, and the i.r. spectra of *N*-acetylamino-acid methylamides, *N*-isopropylacetamide, *N*-methylisobutyramide and their *N*-deuteriated homologues have been recorded down to 300 cm^{-1}.[258] The C=O in-plane and out-of-plane bending vibration bands of the $CH_3 \cdot CO \cdot NH \cdot C^{\alpha}$ group (amides IVa and VIa) and those of the $C^{\alpha} \cdot CO \cdot NH \cdot CH_3$ group (amides IVb and VIb) have been assigned. Two crystalline modifications, form I and form II, were found from the compounds prepared from L-alanine, DL-leucine, L-aspartic acid, and DL-phenylalanine. The two forms show quite different skeletal vibrations, which suggests rotational isomerism. The two amide bands were found near 630 cm^{-1} and 600 cm^{-1} in form II; the *X*-ray structure analyses suggest that these two forms have different hydrogen-bond structures.

Normal-co-ordinate calculations have been made on dimethylformamide and various isotopic substituted forms using Urey–Bradley, modified Urey–Bradley, and symmetrized valence force fields.[259] The frequency shifts on isotopic substitution have been explained and the results of the calculations are discussed in comparison with other related models. Calculations using adiabatic approximation methods have shown that the splitting of the amide I and II bands was due to coupling between CO and NH groups, resulting from vibrational excitation between these groups.[260] Low-temperature spectra of a number of amino-acids have been studied between 400 and 33 cm^{-1}, enabling additional analytical information to be obtained when the amino-acids are in the zwitterion form.[261]

Conformational analyses using i.r. and n.m.r. spectroscopy have been made on *N*-benzyloxycarbonylglycyl-L- and -D-leucine methyl esters and several other related dipeptides, showing that 7-membered rings were present in polypeptide esters which had *cis*-rotamers of the N(6)—C(5) bond.[262] The i.r. spectra of the oligoglycines $Me(NH \cdot CH_2 \cdot CO)_n NH \cdot Et$ ($1 \leqslant n \leqslant 4$) and their *N*-deuteriated homologues in the solid state, have been analysed between 3500 and 200 cm^{-1}.[263] The similarity of these spectra to those of polyglycine II shows that all the oligomers have an analogous structure. The effect of solvent on the conformation of films of low-molecular-weight poly-(γ-benzyl-L-glutamate) has shown that mainly

[258] K. Yasushi, T. Shimanouchi, M. Sato, and T. Tatsuno, *Biopolymers*, 1971, **10**, 1059.
[259] G. Durgaprasad, D. N. Sathyanarayano, and C. C. Patel, *Bull. Chem. Soc. Japan*, 1971, **44**, 316.
[260] J. M. Konarski, *Chem. Phys. Letters*, 1971, **9**, 54.
[261] W. R. Feairheller and J. T. Millo, *Appl. Spectroscopy*, 1971, **25**, 175.
[262] S. V. Zenin, N. Ya. Krasnobrizhii, V. E. Minaev, N. A. Poddubnaya, and C. B. Sergeev, *Zhur. obshchei Khim.*, 1971, **41**, 665.
[263] M. Avignon and C. Garrigou-Lagrange, *Spectrochim. Acta*, 1971, **27A**, 297.

helices are formed from chloroform and that the amount of β structure present increases with increasing amounts of either benzene, dioxan, *m*-cresol, or dimethylformamide.[264]

The interaction of hydroxocobalamine (vitamin B_{12b}) with α-amino-acids and peptides has been studied using i.r. analysis.[265] The results show that charge donation occurs from the N of the amino-acid, in the un-ionized form. It was seen that vitamin B_{12b} forms a complex more readily with glycine than with other amino-acids, and also to peptides with N-terminal glycine.

B. Synthetic Polypeptides.—Hexafluoroisopropanol is a solvent for most proteins and non-ionic polypeptides, and the conformations of several polypeptides have been determined in this solvent using i.r. and u.v. spectroscopy, circular dichroism, and optical rotatory dispersion.[266] Most polymers studied, including poly(methyl-L-glutamate), poly-(L-methionine), poly-(*N*-benzyloxycarbonyl-L-lysine), and poly-(L-homoserine), had an α-helical conformation, whereas poly-[(γ-morpholinylethyl)-L-glutamamide] had a random-coil conformation. Spectroscopic studies of alternating polypeptides, *e.g.* poly-(γ-benzyl-L-glutamate : γ-benzyl-D-glutamate), have indicated that, in helicogenic solvents, conformations are present which are different from the normal α-helix.[267] The properties of such polypeptides in which polymerization was accompanied by appreciable racemization were similar to those of poly-(L-aspartic acid esters) and a distorted type of α-helical conformation was suggested. When racemization was negligible a new conformation, the nature of which was uncertain, was apparently formed. Conformational analyses [268] of ethyl-L-glutamate oligopeptides and of co-oligopeptides containing glutamate residues have been made for both solution and solid state by observing the amide I and II absorption frequencies. Trimethyl phosphate supports the folded structure of oligomers above the heptamer, whereas in chloroform solutions these oligomers assume the extended β structure above the pentamer. Conformations of polypeptides and proteins in aqueous solutions can be studied using i.r. spectroscopy only if D_2O solutions are used. Poly-(*S*-carboxyethyl-L-cysteine), a higher side-chain homologue of poly-(*S*-carboxymethyl-L-cysteine), is soluble in D_2O, and its conformation is dependent upon the pD of the solution.[269] At high ionization the polymer is randomly coiled, but below pH 5.5 it has a pleated sheet structure. Poly-(L-tyrosine) in water–ethanol solvents has a conformation depending on the concentration of ethanol.[270] Between 20 and 40% ethanol concentration the polymer can

[264] T. Iio, *Bull. Chem. Soc. Japan*, 1971, **44**, 859.
[265] J. G. Heathcote, G. H. Moxon, and M. A. Slifkin, *Spectrochim. Acta*, 1971, **27A**, 1391.
[266] J. R. Parrish and E. R. Blout, *Biopolymers*, 1971, **10**, 1491.
[267] H. N. Rydon, P. M. Hardy, J. G. Haylock. D. I. Marlborough, H. T. Storey, and R. C. Thompson, *Macromolecules*, 1971, **4**, 435.
[268] M. Goodman, Y. Masuda, and S. Verdini, *Biopolymers*, 1971, **10**, 1031.
[269] H. Maeda and S. Ikeda, *Biopolymers*, 1971, **10**, 1635.
[270] G. Conio, E. Patrone, and F. Salario, *Macromolecules*, 1971, **4**, 283.

exist as either the α-helical or β-conformation depending upon kinetic factors, but at concentrations of ethanol of 45% and above, only the α-helical form was observed.

I.r. and *X*-ray diffraction studies on atoms of polypeptides and oligopeptides with repeating sequences of L-alanyl-L-prolylglycine have shown that polymers and oligomers longer than the hexapeptide could adopt three different forms, depending upon the solvent to which they have been exposed:[271] (i) hydrogen-bonded sheets of 'polyproline II-like' helices, when dried from aqueous solution, (ii) a more compact hydrogen-bonded 'polyproline II-type' sheet structure, when precipitated from organic solvents or aqueous solution, and (iii) a triple helical structure, when dried from trifluoroethanol. The solid-state i.r. spectra of the polymer and higher oligomers of the three forms showed a clear correlation between the amide frequencies and their structures. I.r. and c.d. solution studies indicated that $(Ala\text{-}Pro\text{-}Gly)_n$ and the oligomers were unordered in aqueous solution, partly ordered in trifluoroethanol solution, and highly structured in ethylene glycol–hexafluoroisopropyl alcohol solution. These conformations were different, however, from that of $(Pro\text{-}Ala\text{-}Gly)_n$, which was found to be triple helical in the solid state and in solution.

Poly-(L-histidine) films at different degrees of protonation have been produced and subjected to an i.r. study between 4000 and 650 cm^{-1} using *N*-deuteriated films.[272] The amide II and III bands showed that the peptide group was present in the *trans*-form. The amide I and II bands were present at between 0 and 50% protonation, indicating that the poly-(L-histidine) was α-helical; at 100% protonation it had a random-coil conformation with a little β-material present. Between 0 and 50% protonation, no hydration water was bound to the backbone. At 0% protonation all NH groups were linked to each other or to water molecules, whereas at 50% protonation, $N^+H \cdots H$ bonds formed between imidazole rings, and the chloride ions were bonded to NH groups. At 100% protonation, water of hydration was also bonded to the C=O group of the backbone. The NH groups of the backbone, like those of the rings, endeavour to bond to the Cl^- ions, which leads to a strong steric constraint of the random coil.

C. Solvation and Structure.—The effect of water on the structure of proteins can be observed by recording their i.r. spectra at different relative humidities (RH). I.r. spectra of undenatured bovine tendon collagen [273] have been recorded at 25 °C and at relative humidities between 0 and 95%. The samples were examined by electron microscopy before and after they had been exposed to i.r. radiation, to see if denaturation had taken place. The frequencies and intensities of the amide bands changed gradually over the RH range 0—75%. Pronounced changes were seen in the amide II vibration as water molecules gradually attached themselves to peptide NH

[271] B. B. Doyle, W. Taub, G. P. Lorenzi, and E. R. Blout, *Biochemistry*, 1971, **10**, 3052.
[272] J. Muehlinghaus and G. Zundel, *Biopolymers*, 1971, **10**, 711.
[273] H. Susi, J. S. Ard, and R. J. Carroll, *Biopolymers*, 1971, **10**, 1597.

bonds within the triple helix over a wide range of relative humidities. Changes in the CH deformation bands suggest that hydrogen-bonding between CH and O also occurs, and that it is stronger at high humidity. The effect of concentrated aqueous solutions of LiBr on randomly coiled sodium poly-(L-glutamate) have been followed using i.r. and n.m.r. techniques.[274] Between 0 and 4 mol l^{-1} LiBr, water is removed from the polymer by an electrostriction effect. However, above 4 mol l^{-1} LiBr a stronger interaction occurs in which competition for the available water in the system forces the electrolyte to form an association with the peptide group either as a fully or partially hydrated species. It was proposed that the anion is associated with the nitrogen atoms of the peptide group, and that the cation is situated within the vicinity of the carbonyl oxygen. The i.r. spectra of several peptides, and of serum albumin and its hydrolysate, have been recorded between 1580 and 1800 cm^{-1} in D_2O solution. The absorption band observed at 1620—1685 cm^{-1} was attributed to the free carbonyl group in the peptide chain.[275]

By observing changes in the amide frequencies, attempts have been made to examine the nature of the interaction of trifluoroacetic acid with poly-(L-alanine).[276] It has been shown that this polypeptide is not protonated by the acid, and that hydrogen-bonding occurs between the acid and the peptide group. Interactions between proteins and polysaccharides have been studied using attenuated total reflectance spectroscopy.[277] Considerable frequency shifts were observed for the amide I and II bonds of the protein and for the OH deformation of the polysaccharide, suggesting that hydrogen-bonding was involved.

7 Circular Dichroism and Optical Rotatory Dispersion*

contributed by P. M. Bayley

The volume of published work continues to increase. Circular dichroism is becoming the predominant technique for peptide structure determination. Commercial instruments are generally available, and the results are more generally interpretable. Unless explicitly stated otherwise, reference is to c.d. work.

The major fields of activity and the corresponding parts of this section of the Report are: theoretical, in A; structural studies on small model systems and synthetic polypeptides, in B; criteria of folding and refolding in proteins, in C; perturbation of protein structure by non-chromophoric ligands and the induced optical activity of chromophores bound to proteins, in D; and membrane systems, in H.

[274] C. B. Baddiel, D. Chaudhuri, and B. C. Stace, *Biopolymers*, 1971, **10**, 1169.
[275] M. A. Sem'onov and V. Ya. Maleev, *Dopovidi Akad. Nauk Ukrain. R.S.R. Ser. B*, 1971, **33**, 136.
[276] P. Combelas and C. Garrigou-Lagrange, *Compt. rend.*, 1971, **272**, *C*, 153.
[277] G. F. Trott and E. E. Woodside, *J. Colloid Interface Sci.*, 1971, **36**, 40.

* See also Chapter 1, Section 3C.

A. General.—*Reviews.* A general guide to the application of o.r.d. and c.d. to the study of biopolymers presents the physical basis of the methods without excessive reliance on theoretical development.[278] The relationship between optical activity and molecular structure is illustrated with reference to simple model systems in the peptide and nucleotide field, and to polynucleotides, proteins, and aggregates (membranes, ribosomes, and viruses). The analysis of peptide c.d. and the use of aromatic c.d. as a probe for limited conformational changes have been discussed for some intensively studied systems.[279] O.r.d. and c.d. of biological macromolecules have been reviewed: relative merits of the two techniques, applications of m.c.d., u.v., and i.r., and the study of oriented specimens are briefly discussed.[280] Specialized reviews deal with the study of drug interaction with biological systems,[281] the binding of small molecules by macromolecules,[282] and the molecular size and conformation of immunoglobulins.[283]

Theory. A non-perturbation method for computing the rotational strength of transitions in helical poly-(L-alanine) using Bogoliubov exciton theory includes the effects of the high-energy π^+–π^- transition (148 nm) on properties associated with the two principal transitions of the peptide link, n–π^- (210 nm) and π^0–π^- (191 nm). The method is suitable for structures of helical symmetry and avoids the problem of solving large matrices.[284] Whereas first-order perturbation theory adequately represents the absorption properties, being relatively insensitive to doubling the effective oscillator strength at 148 nm, the rotational strength calculations were more sensitive to the assumed properties in this region, indicating a preference for non-perturbation methods in this case.

Monte Carlo methods have been applied to the generation of a set of random polypeptide conformations, in conformity with calculated conformational energies, and the rotational properties associated with various subsets of 5 or 10 peptides have been computed,[285] using methods which have previously been used for computations on α and β structures.[286] For a simulated random polypeptide, short-wavelength π–π^* rotational strength is negative below 190 nm, whereas longer wavelength π–π^* components generate positive rotational strength above 200 nm, in opposition to the negative rotational strength associated with the n–π^* transition. This result is at variance with the usual assignment of the characteristic

[278] I. Tinoco, jun. and C. R. Cantor, *Methods Biochem. Analysis*, 1970, **18**, 81.
[279] S. N. Timasheff, in 'The Enzymes', ed. P. D. Boyer, Academic Press, New York, 3rd edn., 1970, vol. 2, p. 398.
[280] C. A. Bush, in 'Physical Techniques in Biological Research', ed. G. Oster, Academic Press, New York, 2nd edn., vol. 1, 1971, p. 347.
[281] C. F. Chignell, *Adv. Drug Res.*, 1970, **5**, 55.
[282] J. H. Perrin and P. A. Hart, *J. Pharm. Sci.*, 1970, **59**, 431.
[283] K. J. Dorrington and C. Tanford, *Adv. Immunology*, 1970, **12**, 333.
[284] F. M. Loxsom, L. Tterlikkis, and W. Rhodes, *Biopolymers*, 1971, **10**, 2405.
[285] V. A. Zubkov, T. M. Birshtein, I. S. Milevskaya, and M. V. Volkenstein, *Biopolymers*, 1971, **10**, 2051.
[286] V. A. Zubkov and M. V. Volkenstein, *Mol. Biol.*, 1970, **4**, 599.

positive ellipticity in random polypeptides at 220 nm as deriving from n–π^*, but agrees with other recent calculations.[287]

In computations using the Kirkwood formulation of natural optical activity, the local origins of groups must be carefully chosen to eliminate local transition dipoles where these are computed from wave-functions based upon atomic co-ordinates.[288]

Theories of the differential scattering of right- and left-handed circularly polarized light and for the distortion of absorption bands by flattening, phenomena which are of importance in the observation of o.r.d. and c.d. of suspensions of optically active particles, have generally treated the two effects as mutually separable. A new treatment of Mie scattering considers the absorption and refraction as the imaginary and real parts of a single complex analytic function [289] (see also ref. 290), and results are presented for solid spheres and for shells of dimensions equivalent to biological membranes.

The theoretical treatment of the conformational dependence of optical activity achieves a new level of refinement, in treating the origins of optical activity in chromophores extrinsic to the polypeptide backbone. Using methods as previously described,[291, 292] calculation of the rotational strengths of the transitions in four possible structures of poly-(L-tyrosine) produced agreement with experimental data only for the structure with right-handed α-helical peptide conformation.[293] The orientation of phenolic side-chains allows some degree of flexibility. In a detailed study of the origin of the haem Cotton effects in myoglobin and haemoglobin,[294, 295] the only mechanism able to account for the rotational strength of the Soret transitions derives from the coupling of the porphyrin transitions with the π–π^* transitions of nearby aromatic residues. Contributions from coupling with the peptide transitions are negligible. Principal contributors are within 4—6 Å (centre-to-centre distance), though tyrosine residues up to 12 Å make significant contributions. In haemoglobin, coupling of haem transitions with the aromatic residues of adjacent subunits is significant. The characteristic asymmetric shape of the Soret c.d. is shown to derive from oppositely signed strong effects of different magnitudes in the near-degenerate transitions; orientations have been proposed for these transitions. These results establish the theoretical rationale for the use of chromophoric probes for conformation, and indicate the complexity and multiplicity of effects which can contribute to the rotational strength of an extrinsic chromophore.

[287] D. Aebersold and E. S. Pysh, *J. Chem. Phys.*, 1970, **53**, 2156.
[288] P. J. Stiles, *Mol. Phys.*, 1971, **22**, 731.
[289] D. J. Gordon, *Biochemistry*, 1972, **11**, 413.
[290] A. S. Schneider, *Chem. Phys. Letters*, 1971, **8**, 604.
[291] I. Tinoco, jun., *Adv. Chem. Phys.*, 1962, **4**, 113.
[292] R. W. Woody, *J. Chem. Phys.*, 1968, **49**, 4797.
[293] A. K. Chen, and R. W. Woody, *J. Amer. Chem. Soc.*, 1971, **93**, 29.
[294] M. C. Hsu and R. W. Woody, *J. Amer. Chem. Soc.*, 1971, **93**, 3515.
[295] M. C. Hsu and R. W. Woody, *J. Amer. Chem. Soc.*, 1969, **91**, 3679.

The theory has been presented for the temperature-dependent optical activity due to quantized molecular rotation, which is likely to be implicated in low-temperature i.r. work.[296] Expressions are also given for the optical rotation of oriented molecules, for asymmetric molecules belonging to a few simple groups.

Analysis. Following extremely careful calibration of an instrument for o.r.d. and c.d., essentially perfect agreement has been found between the o.r.d. spectrum and the Kronig–Kramers transform of the observed c.d. spectrum both for helical poly-(L-lysine) (in 0.5 mol l^{-1} sodium perchlorate, pH 6.30 and 25.0 °C) and helical poly(methyl-L-glutamate) (in trifluoroethanol at 25.0 °C);[297] solvent corrections were omitted. The absence of divergence between the o.r.d. and c.d. (from 185 nm upwards) is used to question the theoretical predictions of exciton theory. Although this is unlikely to gain general acceptance, the absence of dispersive effects originating from higher-energy transitions is surprising, and requires examination in other systems.

The properties of Standard α, β-I, β-II, and random conformations have been reviewed.[279] The analysis of protein c.d. spectra as a linear combination of contributions from α-helical, β-structured, and irregular regions is claimed to correlate better with the fractional contents of these forms determined by *X*-ray crystallography if the spectrum of poly-(L-serine) in 8 mol l^{-1} lithium chloride is used as the reference component for random coil.[298] An alternative to the multicomponent analysis can be expressed as:

$$[\theta] = \sum x_i[\theta]_i = x_\alpha[\theta]_\alpha + x_\beta[\theta]_\beta + x_R[\theta]_R$$

assuming that the observed c.d. spectrum $[\theta]$ can be expressed as the sum of fractional contributions x_i of three types of structure correlating with α-helix, β-structure, and the remaining structure, R. Using the spectral data for three proteins (ribonuclease, lysozyme, and myoglobin) and the values of x_i for each (from *X*-ray data), the reference spectra $[\theta]_\alpha$, $[\theta]_\beta$, and $[\theta]_R$ are found, respectively, to correlate highly with α-helix, but less so with β-structure or random polypeptide. These three spectra gave an analysis of carboxypeptidase A in good agreement with *X*-ray data, but were less successful for chymotrypsin and chymotrypsinogen.[299]

A similar approach using the values of x_i in conjunction with a spectroscopic parameter *P*, where *P* could be b_0, $[m]_{233}$, or $[\theta]_{222}$, may be formulated:

$$P_{obs} = \sum x_i P_i$$

Based upon the *X*-ray data for the same three reference proteins, plus papain and lactate dehydrogenase, good parameters are found for the α-helix, but less satisfactory results for the other identifiable structures.[300]

[296] Y. N. Chiu, *J. Chem. Phys.*, 1970, **52**, 1042.
[297] J. Y. Cassim and J. T. Yang, *Biopolymers*, 1970, **9**, 1475.
[298] H. Rosenkranz and W. Scholtan, *Z. physiol. Chem.*, 1971, **352**, 896.
[299] V. P. Saxena and D. B. Wetlaufer, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 969.
[300] Y.-H. Chen and J. T. Yang, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 1285.

Variability of helix parameters for different models in the same solvent have been noted in ref. 352. These methods suffer the disadvantage of treating the structure of globular proteins as composed of a limited number of structures of invariant properties and with additivity between the components.

Instrumental. C.d. measurements into the vacuum-u.v. to 135 nm at 1.6 nm spectral bandwidth and with resolution of the order of 10^{-5} O.D. units at 10 s time-constant have been reported,[301] allowing work on model compounds in the vapour state, and hopefully, as for absorption,[302] on films of polypeptides.

Cleavage of the peptide backbone has been observed after u.v. irradiation during spectropolarimetric measurements in aqueous solution; the n–π^* region was found to be more sensitive than the π–π^* region.[303]

Modifications of commercial instruments have been made to allow simultaneous recording of absorption and c.d. data, facilitating work on turbid preparations,[304] to provide automatic scanning of temperature,[305] and to cover the range 1.1—1.8 μm using quarter-wave plates of mica in studying the highly dichroic spectra of metalloproteins.[306] O.r.d. studies have also been reported in the range 0.7—2.0 μm for synthetic polypeptides, from which a new property, specifically characteristic of β-structure, has been inferred.[307, 308]

Low-temperature studies at 77 K in transparent water–glycerol glasses continue to improve the resolution of fine structure in aromatic transitions,[309] and to identify conformational states which have only a transient lifetime at normal temperatures.[310] Computer summation of repeated scans is necessary. Individual stages of conformational transitions have also been observed at − 40 °C in fluid aqueous ethylene glycol mixtures.[311]

Magnetic circular dichroism (m.c.d.) measurements on model aromatic compounds show a dominant effect from tryptophan, readily resolved from a smaller effect due to tyrosine and a negligible effect of phenylalanine.[312] The quantitative determination of tryptophan content of a range of proteins was in good agreement with analytical data.[313]

[301] W. C. Johnson, jun., *Rev. Sci. Instr.*, 1971, **42**, 1283.
[302] J. Brahms, J. Pilet, H. Damany, and V. Chandrasekharan, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 1130.
[303] W. D. Wilson and J. F. Foster, *Biochem. Biophys. Res. Comm.*, 1970, **38**, 552.
[304] J. R. Krivacic, D. E. Wisnosky, and D. W. Urry, *Analyt. Biochem.*, 1971, **43**, 547.
[305] W. W. Martz and S. Aktipis, *Analyt. Biochem.*, 1971, **39**, 327.
[306] W. A. Eaton and W. Lovenberg, *J. Amer. Chem. Soc.*, 1970, **92**, 7195.
[307] Yu. N. Chirgadze, S. Yu. Venyaminov, and V. M. Lobachev, *Biopolymers*, 1971, **10**, 809.
[308] S. Yu. Venyaminov and Yu. N. Chirgadze, *Doklady Akad. Nauk S.S.S.R.*, 1970, **195**, 722.
[309] J. Horwitz and E. H. Strickland, *J. Biol. Chem.*, 1971, **246**, 3749.
[310] J. Horwitz and J. Heller, *Biochemistry*, 1971, **10**, 1402.
[311] P. Douzou, R. Sireix, and F. Travers, *Proc. Nat. Acad. Sci. U.S.A.*, 1970, **66**, 787.
[312] G. Barth, R. Records, E. Bunnenberg, C. Djerassi, and W. Voelter, *J. Amer. Chem. Soc.*, 1971, **93**, 2545.
[313] G. Barth, W. Voelter, E. Bunnenberg, and C. Djerassi, *J. Amer. Chem. Soc.*, 1972, **94**, 1293.

Optical rotation has been used to follow the relaxation of poly(benzyl-L-glutamate) over the millisecond range following electric field orientation.[314] Electric dichroism studies themselves on this polymer show highly resolved vibrational structure in the dichroic spectra.[315, 316] Optical rotation has also been used to follow sedimentation during analytical ultracentrifugation.[317]

Determination of the optical purity of a sample by means of the dissymmetry of the luminescent emission from an asymmetric sample excited with circularly polarized light of either sense has been presented in theory,[318] following the principles involved in measuring the optical activity in racemic mixtures.[319, 320]

B. Small Molecules, Model Compounds, and Synthetic Polymers.—*Amino-acids and Derivatives.* Careful examination of the c.d. of α-amino-acids and their esters [321–323] and of α-hydroxy-acids [322, 323] indicates a weak negative band at 230—245 nm approximately two orders of magnitude less than the well-resolved positive effect at 206—209 nm which has generally been assigned to the n–π^* carboxyl transition. The band is absent for ester hydrochlorides in 95% ethanol,[323] but present in aqueous solution.[321] At elevated temperatures and in solvents of low polarity the band increases, whereas at 206—209 nm it decreases, suggesting a conformational equilibrium.[322, 324] The n–π^* origin of the effect is generally accepted, but the relevant n-electron may derive from the heteroatom at C_α,[322] or from the carboxylic function;[324] in the latter case the heteroatom is considered to be nodal to the chromophore.

Distinctions between carboxamide and carboxyl function are drawn in a study of amino-sugars and sialic acids and polymers derived from them. The stereochemistry of gangliosides, which contain both components, may be partially assigned.[325, 326]

Chromophoric derivatives of α-amino-acids are sought for identification of amino-acid chirality: methyl isothiocyanate adducts show faster reaction, enhanced water solubility, and simpler spectra compared with the phenyl isomers.[327] The sign of the effect at 260—270 nm was independent of solvent for water, methanol, chloroform, or dioxan. The derivative with

[314] B. R. Jennings and E. D. Baily, *Nature*, 1970, **228**, 1309.
[315] J. B. Milstien and E. Charney, *Biopolymers*, 1970, **9**, 991.
[316] E. Charney, J. B. Milstien, and K. Yamaoka, *J. Amer. Chem. Soc.*, 1970, **92**, 2657.
[317] A. J. Sophianopoulos, *Analyt. Biochem.*, 1970, **34**, 278.
[318] S. S. Eaton, *Chem. Phys. Letters*, 1971, **8**, 251.
[319] H. P. J. M. Dekkers, C. A. Emeis, and L. J. Oosterhoff, *J. Amer. Chem. Soc.*, 1969, **91**, 4589.
[320] C. A. Emeis and L. J. Oosterhoff, *Chem. Phys. Letters*, 1967, **1**, 129.
[321] C. Toniolo, *J. Phys. Chem.*, 1970, **74**, 1390.
[322] J. C. Craig and W. E. Pereira, jun., *Tetrahedron*, 1970, **26**, 3457.
[323] J. C. Craig and W. E. Pereira, jun., *Tetrahedron Letters*, 1970, 1563.
[324] I. Listowsky, G. Avigad, and S. Englard, *J. Org. Chem.*, 1970, **35**, 1080.
[325] A. L. Stone and E. H. Kolodny, *Chem. and Phys. Lipids*, 1971, **6**, 274.
[326] A. L. Stone, *Biopolymers*, 1971, **10**, 739.
[327] C. Toniolo, *Tetrahedron*, 1970, **26**, 5479.

cysteine, though reacting slowly, may have some analytical value.[328] *N*-Acetoacetylamino-acids exhibit a solvent-dependent c.d. and a pH dependence corresponding to carboxyl ionization and enolization; tertiary amides (*e.g.* proline) show inverted effects.[329] The β-carbonylamides give rise to azlactones, the geometric isomers of which exhibit distinctive c.d. effects.[330] *N*-Dithioethoxycarbonyl-α-methyl-amino-acids exhibit similarly signed c.d. independent of solvents (methanol, dioxan, chloroform, or benzene) but aromatic amino-acid (*e.g.* α-methylphenylalanine) derivatives, though solvent-insensitive, are oppositely signed. Of the hydantoin derivatives, only the α-methyl-α-phenylglycine derivative was exceptional.[331] A new and rapid resolution of DL-α-methylphenylalanine is reported.[332]

The spectroscopic properties of the disulphide component of L-cystine have been examined in the crystalline state in isotropic preparations in KBr discs.[333] The anomalous, split c.d. of L-cystine is attributed to the proximity of adjacent disulphides in the crystal, whereas in L-cystine hydrochloride no such effects occur, owing to the interposition of the chloride ion. The latter compound, exhibiting a broad negative effect from 250 to beyond 300 nm,[334] correlates with symmetry assignments made on the basis of solution properties of restricted molecules.[335]

Metal complexes of the chiral amino-acids are optically active in the visible spectrum: an empirical rule is derived from the observed additive properties of asymmetric ligands bound to copper(II), in which the metal and chelating atoms are on nodal planes, and only the amino-acid side-group acts as perturbant.[336] Studies of the mixed complexes of copper(II) with glycine and histidine derivatives suggest that histidine complexes resemble histamine rather than glycine.[337] Alcoholic solutions of amino-acid esters form dipeptides in the presence of copper(II), a synthetic activity which complements the hydrolysis properties of copper(II)–peptide complexes.[338] Individual components of the complex ligand-field transition in carboxylatopenta-aminecobalt(III) complexes involving L-amino-acids have been resolved by the effects in c.d. spectra of adding polarizable oxyanions: the component of A_{2g} origin is negative; those of E_g origin are of opposite sign, being positive at lower energy.[339]

[328] C. Toniolo and G. Jori, *Biochim. Biophys. Acta*, 1970, **214**, 368.
[329] C. Toniolo, F. Filira, and C. DiBello, *Biopolymers*, 1971, **10**, 2275.
[330] C. DiBello, F. Filira, and C. Toniolo, *Biopolymers*, 1971, **10**, 2283.
[331] K. Achiwa, S. Terashima, M. Hatsuhiko, N. Takamura, T. Kitagawa, K. Ishikawa, and S. Yamada, *Chem. and Pharm. Bull. (Japan)*, 1970, **18**, 61.
[332] F. W. Bollinger, *J. Medicin. Chem.*, 1971, **14**, 373.
[333] N. Ito and T. Takagi, *Biochim. Biophys. Acta*, 1970, **221**, 430.
[334] P. C. Kahn and S. Beychok, *J. Amer. Chem. Soc.*, 1968, **90**, 4168.
[335] M. Carmack and L. A. Neubert, *J. Amer. Chem. Soc.*, 1967, **89**, 7134.
[336] J. M. Tsangaris and R. B. Martin, *J. Amer. Chem. Soc.*, 1970, **92**, 4255.
[337] H. Sigel, R. E. Mackenzie, and D. B. McCormick, *Biochim. Biophys. Acta*, 1970, **200**, 411.
[338] S. I. Yamada, S. Terashima, and M. Wagatsuma, *Tetrahedron Letters*, 1970, 1501.
[339] C. J. Hawkins and P. J. Lawson, *Inorg. Chem.*, 1970, **9**, 6.

Dipeptides and Oligopeptides. Optical rotatory properties of molecules containing two peptide groups have been obtained for *N*-acetyl-L-prolinamide, L-3-acetamidopyrrolidin-2-one, L-pyroglutamide, and L-alanine-dioxopiperazine, which have their conformation partially restricted by cyclic structures.[340] Results for water and dioxan are in keeping with theoretical calculations for a four-transition dipeptide model.[341]

The 3-cephem chromophore, present in cephalosporin antibiotics, contains a homoconjugated π-system composed of a tertiary amide and an enamine held dissymmetrically in a rigid conformation. Intense and oppositely signed effects at 260 nm (positive) and 230 nm (negative) are assigned respectively to a π–π^* and, from solvent effects, to an n–π^* transition, respectively.[342] The latter assignment appears dubious because of the intensity and molecular structure.

Cyclic hexapeptides with one and two side-chains (for which n.m.r. indicates the presence of two transannular hydrogen bonds) appear to exist in two conformations; *cyclo*-(Gly-Gly-Leu-Gly-Gly-Leu-) is assigned a C_2 symmetry. The monoleucyl derivative exhibits c.d. apparently similar to that of random poly-(L-glutamate); in the cyclohexapeptide, however, both the positive 214 nm and the weaker negative 222 nm effects are attributed to n–π^* transitions, the difference in energies deriving from the intramolecular hydrogen-bonding.[343]

The synthetic cyclononapeptide cyclolinopeptide A, [*cyclo*-(Leu-Ile-Ile-Leu-Val-Pro-Pro-Phe-Phe-)] exhibits solvent dependence in hexafluoroisopropanol, trimethyl phosphate, or trifluoroethanol, under conditions where the rigid model compound camphorolactone showed minor solvent shifts. Considerable conformational flexibility evidently exists for this cyclic nonapeptide.[344]

A bicyclic peptide, *SS'*-bis-*cyclo*-(Gly-Cys-Gly-Gly-Pro-), containing two homodetic peptide rings, has been shown to undergo a progressive conformational change in the presence of 0—4.0M-KCl. The dichroism at 260 nm indicates a right-handed (P) chirality for the disulphide, with dihedral angle 75—90°.[345] In the cyclic synthetic decapeptide [2,7-cystine]-gramicidin S, containing an intramolecular disulphide, P chirality and the unique feature of a dihedral angle greater than 90° are inferred from the chemical shift of valine protons perturbed by the sulphur *p*-orbitals.[346] The observation of a well-defined, broad, negative c.d. band with peak at 271 nm correlates with the predicted quadrant rule for organic disulphides.[347]

[340] E. B. Nielsen and J. A. Schellman, *Biopolymers*, 1971, **10**, 1559.
[341] P. M. Bayley, E. B. Nielsen, and J. A. Schellman, *J. Phys. Chem.*, 1969, **73**, 228.
[342] R. Nagarajan and D. O. Spry, *J. Amer. Chem. Soc.*, 1971, **93**, 2310.
[343] S. M. Ziegler and C. A. Bush, *Biochemistry*, 1971, **10**, 1330.
[344] F. Naider, E. Benedetti, and M. Goodman, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1195.
[345] R. Schwyzer, A. Tun-kyi, P. Moser, and M. Caviezel, *Helv. Chim. Acta*, 1970, **53**, 15.
[346] U. Ludescher and R. Schwyzer, *Helv. Chim. Acta*, 1971, **54**, 1637.
[347] J. Linderberg and J. Michl, *J. Amer. Chem. Soc.*, 1970, **92**, 2619.

For linear oligopeptides of isoleucine in trifluoroethanol, u.v. absorption and c.d. indicate the onset of ordered structure at the heptamer and octamer level, with spectra typical for β-conformations. These structures appear to be more stable than the oligomers of alanine, and an intermolecular association is inferred from the disappearance of characteristic spectra on dilution.[348, 349] For oligomers of lysine in water at pH 11.9 and in methanolic solutions, the appearance of an ordered structure is only partial even at 22 residues. Helix–coil data can be obtained from the dependence of helicity on chain length.[350] It is notable that at pH 3.8 the intensity of c.d. bands of the random form continues to increase up to DP = 156.[351]

Polypeptides. The use of hexafluoroisopropanol as a solvent for proteins and synthetic polypeptides allows observation of u.v., visible, and i.r. spectroscopic properties.[352] The solvent is less polarizable than water, with $n_{589} = 1.2752$. The homopolymers of γ-methyl L-glutamate, L-methionine, ε-*N*-benzyloxycarbonyl-L-lysine, and L-homoserine exhibit α-helical o.r.d. and c.d. spectra with small but significant differences; the parameters $[m]_{199}$ and $[\theta]_{222}$ showed least variability.

The stabilities of α-helical homopolymers of L-lysine, L-ornithine, and L-diaminobutyric acid, investigated as a function of pH at 25 °C and as a function of temperature at pH 11.75, indicate that the side-chains make an important contribution to the stability, with a progressive decrease as the number of methylenes is decreased.[353] Chemical cross-linking of poly-(L-lysine) stabilizes the helical structure against thermal transitions on heating in aqueous solution at pH 11.5.[354] The helix content of poly-(L-arginine) and poly-(L-lysine) undergoes a sharp increase as the proportion of alcohol in aqueous alcohol mixtures exceeds 75—80% by volume for trifluoroethanol, 2-chloroethanol, or methanol. Calf thymus histone, fractionated or unfractionated, showed a steady increase, attributed to the presence of a range of regions of varying helix-forming capability.[355]

Further effects of side-chain composition on conformation in aqueous solution are shown by poly-(L-lysine) substituted at the ε-amino-group with amino-acids. The glycyl derivative appeared random under conditions where the phenylalanyl and leucyl derivatives were helical.[356] Introduction of L-alanyl residues into the sequentially random copolymer with L-lysine results in a helical conformation of lower thermodynamic stability.[357]

[348] C. Toniolo, *Biopolymers*, 1971, **10**, 1707.
[349] M. Goodman, F. Naider, and C. Toniolo, *Biopolymers*, 1971, **10**, 1719.
[350] A. Yaron, E. Katchalski, A. Berger, G. D. Fasman, and H. A. Sober, *Biopolymers*, 1971, **10**, 1107.
[351] G. E. Perlmann and K. Grizzuti, *Biochemistry*, 1971, **10**, 4168.
[352] J. R. Parrish, jun. and E. R. Blout, *Biopolymers*, 1971, **10**, 1491.
[353] M. J. Grourke and J. H. Gibbs, *Biopolymers*, 1971, **10**, 795.
[354] I. M. Klotz and J. U. Harris, *Biochemistry*, 1971, **10**, 923.
[355] M. Boublik, E. M. Bradbury, C. Crane-Robinson, and H. W. E. Rattle, *European J. Biochem.*, 1970, **12**, 258.
[356] N. Anand, N. S. R. K. Murthy, F. Naider, and M. Goodman, *Macromolecules*, 1971, **4**, 564.
[357] H. Sugiyama and H. Noda, *Biopolymers*, 1970, **9**, 459.

Poly-(*S*-carboxyethylcysteine) changes from a random conformation to β-structure as the pH is reduced below 5.5; intermolecular association is involved.[358] This behaviour parallels that of the methyl analogue in forming β-structure in aqueous solution, and high molecular weight favours the conversion.[359, 360] Spectra qualitatively resembling β-structure are shown by poly-(L-histidine) in aqueous solution at pH 5.2. This conformation is reached by a two-stage process from the random conformation as the pH is increased from 4.2.[361] A specific interaction between perchlorate and the side-chains of poly-(1-benzyl-L-histidine) in trifluoroethanol is implicated in the formation of a characteristic ordered structure when perchloric acid is added to protonate the polymer stoicheiometrically. The c.d. shows a strong positive band at 225 nm and a stronger negative band at 198 nm; other acids give randomization.[362]

The origins of rotational strength in poly-(L-tyrosine) have been examined theoretically, and interactions between peptide and aromatic transitions are important.[293] Simple additivity of helical and aromatic effects is therefore unreliable. Tyrosine oligopeptides up to $n = 12$ are random in dimethyl sulphoxide, 1,2-propanediol, or dimethylformamide.[363] The helical conformation of the polymer in the latter two solvents is destabilized sharply when dimethyl sulphoxide is added to 80% or 50% by volume, respectively. For poly-(L-tyrosine) in aqueous solution, changes in c.d. spectra between pH 10.6 and 11.2 are attributed to the partial ionization of phenolic chromophores, with the conformation of the helical backbone remaining essentially unchanged.[364]

The polytripeptide poly Tyr-Ala-Glu shows a helix–coil transition above pH 6 in low, and above pH 11 in high salt concentrations. At intermediate pH values the transition can be induced by decreasing the salt concentration.[365] Stabilization is due to hydrogen-bonding between the *i*th glutamate and the $(i + 4)$th tyrosine side-chain, attached to an α-helical backbone. The resulting array of phenolic chromophores contributes characteristically to the o.r.d. at 220—240 nm and shows 8% hyperchromicity at 278 nm.

Poly-(β-benzyl-L-aspartate) is known to adopt the left-handed α-helix in chloroform, whereas substituted benzyl derivatives adopt the right-handed α-helix in the same solvent. This critical dependence of conformation upon interactions of side-chains with solvent is further shown by the appearance of a mirror-image spectrum for the polymer itself in trimethyl

[358] H. Maeda and S. Ikeda, *Biopolymers*, 1971, **10**, 1635.
[359] S. Makino and S. Sugai, *Biopolymers*, 1970, **9**, 1049.
[360] L. Stevens, R. Townend, S. N. Timasheff, G. D. Fasman, and J. Potter, *Biochemistry*, 1968, **7**, 3717.
[361] Y. P. Myer and E. A. Barnard, *Arch. Biochem. Biophys.*, 1971, **143**, 116.
[362] A. Cosani, E. Peggion, M. Terbojevich, and M. Acampora, *Macromolecules*, 1971, **4**, 390.
[363] J. Engel, E. Liehl, and C. Sorg, *European J. Biochem.*, 1971, **21**, 22.
[364] S. Friedman and P. O. Ts'o, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 510.
[365] J. Ramachandran, A. Berger, and E. Katchalski, *Biopolymers*, 1971, **10**, 1829.

phosphate, indicating that it too can have the right-handed helical conformation.[366]

In continuation of studies of synthetic analogues of the proline-rich regions of collagen, sequential copolypeptides (alternating) of the form $(\text{Pro-Ala})_n$, $(\text{Pro-Gly})_n$, and $(\text{4-Hyp-Gly})_n$ at -45 °C in ethylene glycol–water, show c.d. resembling poly-proline-II and native collagen with the characteristic strong positive band at 220 nm.[367] At higher temperatures the spectra resemble that of heat-denatured collagen. The presence of proline (or hydroxyproline) at every third residue does not seem a necessary condition for adopting the collagen conformation. The stability of the ordered structure in the hydroxylated derivative is enhanced by intramolecular hydrogen-bonding:[368] ordered structures are achieved at -25 °C in trifluoroethanol. Collagen-like structures are formed by poly Pro-Ser-Gly and poly Pro-Ala-Gly in propanediol at normal temperatures, but not in aqueous solution.[369] A three-fold increase in molecular weight for either polymer indicates formation of a triple helix. Trimerization occurs with poly $(\text{Pro-Pro-Gly})_n$ for $n = 10$, 15, or 20, and dissociation at high temperature is accompanied by a change in optical rotation resembling the thermal transition of collagen.[370]

Model polytripeptides for the non-proline regions of collagen of the form poly Ala-Glu(OEt)-Gly and poly Ala-Gly-Gly are generally insoluble in aqueous solution.[371–373] Studied as a suspension or film, the former shows c.d. resembling antiparallel β-structure, whereas the latter resembles a random conformation, or the poly-proline-II spectrum shifted to the blue. I.r. measurements support the first assignment; *X*-ray diffraction suggests a poly-glycine-II conformation in the second case.

Asymmetric polyamides deriving from optically active dicarboxylic acids of cyclo-propane, -butane, and -pentane and the diamide derivatives of piperazine and ethylenediamine exhibit c.d. similar to the polypeptides. Piperazine derivatives give rigid polymers with strong c.d. bands at 220 nm. Both simple bands and oppositely-signed bands are observed, depending upon solvent.[374]

C. Denaturation.—*Detergents and Neutral Salts.* Sodium dodecyl sulphate (SDS), at low concentrations close to the critical micelle concentration, changes the c.d. spectrum of many globular proteins to a characteristic

[366] V. Gianconti, F. Quadrifoglio, and V. Crescenti, *J. Amer. Chem. Soc.*, 1972, **94**, 297.
[367] W. L. Mattice and L. Mandelkern, *J. Amer. Chem. Soc.*, 1970, **92**, 5285.
[368] W. L. Mattice and L. Mandelkern, *Biochemistry*, 1971, **10**, 1926.
[369] F. R. Brown, tert., A. DiCorato, G. P. Lorenzi, and E. R. Blout, *J. Mol. Biol.*, 1972, **63**, 85.
[370] Y. Kobayashi, R. Sakai, K. Kakiuchi, and T. Isemura, *Biopolymers*, 1970, **9**, 415.
[371] J. M. Anderson, W. B. Rippon, and A. G. Walton, *Biochem. Biophys. Res. Comm.*, 1970, **39**, 802.
[372] W. B. Rippon and A. G. Walton, *Biopolymers*, 1971, **10**, 1207.
[373] W. B. Rippon, J. M. Anderson, and A. G. Walton, *J. Mol. Biol.*, 1971, **56**, 507.
[374] C. G. Overberger and Y. Shimokawa, *Macromolecules*, 1971, **4**, 718.

form, which has some features in common with that of the α-helix.[375] Ribonuclease is affected only after reduction and carboxymethylation. With immunoglobulin fragment Fc, alkyl sulphates of chain length up to C_{14} were progressively more effective, though the relative magnitudes of the resulting spectral bands differed in detail.[376] Elastase, chymotrypsin, and pepsin are transformed by SDS; trypsin and lysozyme less so. After dialysis, all were inactive and had spectra characteristic of a high content of β-structure, for which confirmation was obtained from i.r.[377] The exact nature of the conformations adopted in SDS is unclear: from the effect of chain length, hydrophobic factors are evidently involved, together with the micellization of the protein–detergent complex. Remarkable constancy for widely different proteins is observed in the effect of SDS, and apparently for transformations which follow upon dialysis. Aromatic c.d. is in general reduced in intensity in SDS [375–377] but rarely eliminated.

Binding of alkane derivatives in the range C_8—C_{14} by bovine serum albumin causes limited unfolding which occurs in several steps; again, C_{12} and C_{14} derivatives are more effective.[378] In the interaction of alkylpyridinium chloride with oxymyoglobin, derivatives with alkanes C_{12} and C_{14} were ineffective, while C_{16}—C_{20} effected a drop in Soret absorbance which paralleled a 20% decrease in $[m']_{235}$.[379]

The protein–bacteriochlorophyll complex from a green photosynthetic bacterium undergoes dissociation in SDS, with transformation of the protein conformation.[380] Cationic and non-ionic detergents effect dissociation only. In the case of the bacteriochlorophyll complex from purple bacteria, the mild detergent action of octanoic acid produces a soluble complex with apparently monomeric c.d. properties at 860 nm (compared with the dimeric properties of chromatophore suspensions). Organic solvents effect solution, but with spectra shifted to 770 nm.[381] Various degrees of potency are thus available in the detergents, which may be utilized to simulate essentially hydrophobic interactions between proteins and ligands, and for inter- and intra-molecular effects in protein structure.

Denaturation and unfolding is generally accomplished with concentrated guanidine hydrochloride (GHCl); the thiocyanate salt is apparently even more potent.[382] Pancreatic trypsin-inhibitor is resistant to a concentration of 6 M-GHCl,[383] and combination of this and analogous inhibitors to trypsin stabilizes the latter against unfolding by 8 mol l^{-1} urea.[384] Of

[375] B. Jirgensons and S. Capetillo, *Biochim. Biophys. Acta*, 1970, **214**, 1.
[376] P. K. Lee and B. Jirgensons, *Biochim. Biophys. Acta*, 1971, **229**, 631.
[377] L. Visser and E. R. Blout, *Biochemistry*, 1971, **10**, 743.
[378] J. Steinhardt, J. Krijn, and J. G. Leidy, *Biochemistry*, 1971, **10**, 4005.
[379] K. Nakaya, K. Yamada, M. Onozawa, and Y. Nakamura, *Biochim. Biophys. Acta*, 1971, **251**, 7.
[380] Y. D. Kim, *Arch. Biochem. Biophys.*, 1970, **140**, 354.
[381] H. Steffen and M. Calvin, *Biochem. Biophys. Res. Comm.*, 1970, **41**, 282.
[382] S. Lapanje, *Biochim. Biophys. Acta*, 1971, **243**, 349.
[383] J. P. Vincent, R. Chicheportiche, and M. Lazdunski, *European J. Biochem.*, 1971, **23**, 401.
[384] N. Levilliers, M. Peron, B. Arrio, and J. Pudles, *Arch. Biochem. Biophys.*, 1970, **140**, 474.

considerable potency are the concentrated solutions of lithium salts which apparently randomize synthetic polypeptides,[385, 386] but are less effective than GHCl in unfolding bovine serum albumin.[387] The denaturing efficiency of the amides with myoglobin, cytochrome *c*, and chymotrypsinogen, like that of the alcohols, glycols, and ureas, increases with increasing hydrocarbon content and alkyl substitution, suggesting a predominantly hydrophobic mechanism for these agents.[388, 389]

Unfolding and Refolding. The zymogen of Streptococcal proteinase (M = 44 000) unfolds to a random coil in 6 M-GHCl; following dialysis, the protein can be reactivated by 90%, indicating effectively full restoration of native conformation. The proteolysed zymogen, formed by removal of the first 100 residues, is irreversibly denatured by GHCl.[390] The regeneration of activity from unfolded reduced lysozyme is accelerated by the presence of a mixture of thiol and disulphide reagents; rapid refolding of the chain (relative to regeneration of disulphides) occurs on dilution of the denaturant, since the kinetics of recovery of activity are constant for refolding from different states.[391] Refolding of bacteriophage MS2 coat protein to a capsid of normal size and antigenicity occurs on removal of GHCl.[392] The native and refolded conformations are high in β-structure, demonstrating, as with the immunoglobulins (section F), that refolding is realizable with a wide range of conformations.

D. Proteins.—This section is divided into aromatic (and disulphide) chromophores whose c.d. properties reflect local environmental effects; non-chromophoric proteins, whose peptide chromophores indicate backbone conformation; proteins with natural additional chromophores, whose (extrinsic) properties are determined by binding modes which are relevant to biological function; and extrinsic properties of added unnatural ligands which are dependent on binding-site conformation. Changes in any of these properties with added (non-chromophoric) ligands may indicate conformational changes.

Aromatic and Disulphide Chromophores. These chromophores present useful probes for protein structures, *e.g.* in chymotrypsin ribonuclease, carbonic anhydrase, or pyrocatechase.[279] Theoretical interpretation of the spectrum of poly-L-tyrosine shows the mechanism for generating rotational strength at 277 nm and 227 nm (both positive) in the phenol chromophore.[293] The π–π^* transitions of the peptide and the 192.5 nm phenol transitions are not resolvable and cannot be characterized simply from the extrema

[385] M. L. Tiffany and S. Krimm, *Biopolymers*, 1969, **8**, 347.
[386] D. G. Dearborn and D. B. Wetlaufer, *Biochem. Biophys. Res. Comm.*, 1970, **39**, 314.
[387] M. E. Noelken, *Biochemistry*, 1970, **9**, 4117.
[388] T. T. Herskovits, H. Jaillet, and T. Desena, *J. Biol. Chem.*, 1970, **245**, 6511.
[389] T. T. Herskovits, H. Jaillet, and B. Gadegbeku, *J. Biol. Chem.*, 1970, **245**, 4544.
[390] M. C. Lin and M. Bustin, *J. Biol. Chem.*, 1970, **245**, 3384.
[391] V. P. Saxena and D. B. Wetlaufer, *Biochemistry*, 1970, **9**, 5015.
[392] P. J. Oriel, P. Lindsey, and C. Schueneman, *Biopolymers*, 1971, **10**, 1661.

of the total c.d. envelope. Synthetic systems involving tyrosine [363–365] and histidine [361] have been mentioned; as with disulphides,[333–335, 345, 346] direct conformational evidence is derivable mainly from the longest-wavelength transitions.

The enhanced resolution available from low-temperature studies indicates that a single tyrosine residue with 0–0 band at 288.5 nm in ribonuclease-A is shifted 2.5 nm to the blue in ribonuclease-S.[309] *X*-Ray evidence suggests that Tyr-25 is partially exposed in the latter, though its phenolic OH is still buried and hydrogen-bonded. Low-temperature results so far obtained correlate well with those at normal temperature, and the extra degree of resolution represents the most powerful method of analysing the overlapping effects in this region.

Assignments and resolution of individual aromatic amino-acids have been given for the coat protein of tobacco mosaic virus: from model compounds and pH dependence of spectra, contributions from phenylalanine (248, 252, and 257 nm), tyrosine (265, 274, and 281 nm), and tryptophan (291 and 296 nm) have been detected.[393] Spectra of the proteins from other strains and mutants differing in amino-acid composition show that replacements of aliphatics do not affect aromatic c.d., whereas replacements of the aromatics cause significant changes, offering a means of resolving the total spectrum. If this method could be coupled with information about the degree of burying of replaced residues of known sequence, *e.g.* by low-temperature studies,[309] a direct relationship between primary sequence and conformation in solution would be available. Mutation provides selective alteration of primary sequence which has previously been achieved by chemical modification for resolution of aromatic c.d. for lysozyme.[394]

Pancreatic ribonuclease-A, containing no tryptophan residues, is frequently used as a standard system, though a range of environments for tyrosine and cysteine complicates the interpretation of the origins of the c.d. effects. The effect of acid is to alter the aromatic c.d. preferentially;[395] at constant pH and ionic strength, increasing the concentration of acetic acid in the buffer gives progressive diminution of c.d. by a reversible and co-operative effect. Changes in o.r.d. at 228 nm between pH 6.5 and 1.2 are not paralleled by changes in c.d. below 230 nm, suggesting that these effects originate with the aromatic residues rather than with peptides.[396] The reduction in c.d. (276 nm, less negative; 241 nm, less positive) follows $pK = 1.5$. Photo-oxidation with haematoporphyrin at different stages of acid transition allows formation of sulphoxide at Met-29 alone, or with Met-13, or at all four methionine residues together.[397] Whereas the first

[393] A. Z. Budzynski, *Biochim. Biophys. Acta*, 1971, **251**, 292.
[394] V. I. Teichberg, C. Kay, and M. Sharon, *European J. Biochem.*, 1970, **16**, 55.
[395] J. R. Cann, *Biochemistry*, 1971, **10**, 3713.
[396] P. McPhie, *J. Biol. Chem.*, 1971, **246**, 5537.
[397] G. Jori, G. Galiazzo, A. M. Tamburro, and E. Scoffone, *J. Biol. Chem.*, 1970, **245**, 3375.

product retains the major portion of the aromatic c.d., after modification of Met-13 most of this c.d. is lost and the peptide region appears significantly randomized.

By contrast, elongation of the disulphides by formation of $(Cys)_2Hg$ bridges gives a derivative with the gross conformation and a small amount of the activity of the native enzyme. Peptide c.d. is retained, but the aromatic region is reduced to approximately 25%.[398] Elongation of cystine-II—VII would force Asp-38 and (buried) Tyr-92 apart. These results suggest that the major contribution to the c.d. at 275 nm comes from tyrosine residues, and probably from Tyr-92. Significant contributions from disulphide transitions would presumably correlate with retention of peptide c.d. Their absence from the enzyme after conformational transition in acid and the similarity in c.d. of this form with the mercurated derivative suggests that their contribution in the native form is minor; they may be more significant at 77 K.[399]

Studies on the deoxyribonucleases from different sources (the acid, pancreatic, and *S. aureus* enzymes) show that considerable structural differences exist between these functionally related enzymes.[400] Thus the aromatic c.d. is positive, weakly positive and negative, and strongly negative in the three cases, and helical content is low in each. The pancreatic enzyme has a partially reversible thermal transition between two states differing in tyrosine environments,[401] but otherwise the deoxyribonucleases show little resemblance to the pancreatic ribonuclease.

The effects of solvents on the aromatic c.d. of lysozyme show two stages: at concentrations up to 40% aqueous ethanol, the positive c.d. (mainly tryptophan) is intensified, and fine structure is better resolved. At 90% ethanol, a dramatic reversal of sign occurs in the 280—290 nm region, the negative region at 250—260 nm is less intense, and the system exhibits an isodichroic point at *ca.* 275 nm. Simultaneously the 210—225 nm region is intensified and indicates adoption of a more helical conformation.[402] Similar effects are observed with a number of alcohols, only ethylene glycol being relatively ineffective. Inversion has been observed for *N*-acetyl-L-tyrosine derivatives on changing from dioxan to methane;[399] the dependence of tyrosine c.d. upon side-chain orientation has been treated theoretically.[403]

Conformational transitions induced by change of the external medium include: for asparaginase, aromatic c.d. which is constant from pH 4.5 to 7.5, but showing intensification of the 250 nm region at pH 11—12, although the far-u.v. indicates extensive randomization;[404] for insulin, some intensity retained at 70 °C, reduction of intensity with retention of fine-structure in

[398] R. Sperling and I. Z. Steinberg, *J. Biol. Chem.*, 1971, **246**, 715.
[399] J. Horwitz, E. H. Strickland, and C. Billups, *J. Amer. Chem. Soc.*, 1970, **92**, 2119.
[400] S. N. Timasheff and G. Bernardi, *Arch. Biochem. Biophys.*, 1970, **141**, 53.
[401] S. B. Zimmerman and N. F. Coleman, *J. Biol. Chem.*, 1971, **246**, 309.
[402] K. Ikeda and K. Hamaguchi, *J. Biochem. (Japan)*, 1970, **68**, 785.
[403] T. M. Hooker, jun. and J. A. Schellman, *Biopolymers*, 1970, **9**, 1319.
[404] H. Rosenkranz and W. Scholtan, *Z. physiol. Chem.*, 1971, **352**, 1081.

80% acidic methanol or 70% 2-chloroethanol, related to changes in side-chain conformation rather than non-conformational solvent effects;[405, 406] for pepsinogen, reversible conversion by alcohols into a specific conformation from which pepsin activity cannot be obtained by acidic hydrolysis and which is characterized by low c.d.;[407] for human and bovine growth hormone, distinctive aromatic c.d. which also allows differentiation of the three components of the human hormone (in spite of similar amino-acid compositions), and distinctive far-u.v. spectra, with higher helical structure for the bovine hormone, though hormonal activity is also associated with its peptide fragment, which is richer in β-structure;[408] and for the human chorionic somatomammotropin, stable conformation from pH 12—4, small changes in aromatic conformation at pH 3.6, and total loss of order in 50% acetic acid (both reversible), the unfolding behaviour resembling bovine growth hormone and ovine prolactin.[409, 410]

With N-terminal fragments of angiotensin-II, ordered structures resembling β-structure appear at the heptapeptide level (in trifluoroethanol, or in aqueous solution at 60 °C), and at the octapeptide level the addition of the functionally important phenylalanine residue (adjacent to proline) has striking spectroscopic consequences.[411]

More-specific effects observable by the aromatic c.d. are: for concanavalin A, involvement of tyrosine and tryptophan in pH-dependent conformational changes, predominant involvement of tyrosine in changes induced by Ca^{2+} and α-methyl D-glucopyranoside,[412] and by α-methyl D-mannoside, the ligand not conferring stability against the dramatic loss of structure at pH 9;[413] for thyroxine-binding prealbumin, stable conformation from pH 12—3.5, expansion over pH 3.5—2.5 with loss of tryptophan-like spectrum;[414] for pyruvate kinase, a large increase in intensity in going from 37 to 5 °C, conformational transitions inducible by 0.1 mol l^{-1} K^+ (activating) or 0.1 mol l^{-1} Li^+ (non-activating), independent of Mg^{2+} and phosphoenol pyruvate;[415] for fructose 1,6-diphosphatase (rabbit liver), intensification of the aromatic c.d. occurs in the presence of fructose 1,6-diphosphate and AMP together, but not individually, with fine-structure more resolved at pH 9.1 than 7.5, and the enzyme generally being little affected in peptide conformation by pH or SDS, although the latter

[405] M. J. Ettinger and S. N. Timasheff, *Biochemistry*, 1971, **10**, 824.
[406] M. J. Ettinger and S. N. Timasheff, *Biochemistry*, 1971, **10**, 831.
[407] H. Neumann and M. Shinitzky, *Biochemistry*, 1971, **10**, 4335.
[408] M. Sonenberg and S. Beychok, *Biochim. Biophys. Acta*, 1971, **229**, 88.
[409] T. A. Bewley and C. H. Li, *Arch. Biochem. Biophys.*, 1971, **144**, 589.
[410] S. Aloj and H. Edelhoch, *J. Biol. Chem.*, 1971, **246**, 5047.
[411] S. Fermandjian, J.-L. Morgat, and P. Fromageot, *European J. Biochem.*, 1971, **24** 252.
[412] W. D. McCubbin, K. Oikawa, and C. M. Kay, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 666.
[413] M. N. Pflumm, J. L. Wang, and G. M. Edelman, *J. Biol. Chem.*, 1971, **246**, 4369.
[414] W. T. Branch, J. Robbins, and H. Edelhoch, *J. Biol. Chem.*, 1971, **246**, 6011.
[415] R. A. Wildes, H. J. Evans, and R. R. Becker, *Biochim. Biophys. Acta*, 1971, **229**, 850.

eliminates the aromatic effects;[416] for carbonic anhydrase (parsley), less-intense aromatics than the mammalian enzyme, with no positive dichroism, correlating with the lower tryptophan content;[417] for human lysozyme, which has more-intense tryptophan c.d. than the hen egg-white enzyme, and which shows a band at 313 nm due possibly to a tryptophan–tyrosine interaction whose intensity increases with pH, perturbation of the whole region, including the new band formed on binding *N*-acetyl-D-glucosamine;[418] for phospholipase A_2 from *Crotalus adamanteus* venom, intense tryptophan spectrum, associated with a dimeric protein of high α-helix content;[419] for allantoicase, two-fold reduction of aromatic c.d. on binding the competitive inhibitor *N*-carbamoyl-D-asparagine (plus a lesser reduction of an extrinsic effect at 418 nm attributed to asymmetrically bound Mn^{2+}), with effectively identical behaviour for both the 0.9 S and 10.8 S forms of the enzyme.[420]

Aromatic c.d. is also a useful indicator of the degree of conformational perturbation which occurs on association of two protein components as in the haemoglobin–haptoglobin complex,[421–423] immunoglobulin structure,[283] and the troponin–tropomyosin association which forms the relaxing protein.[424] In the latter case, the helix content of the components is 42% and 98%, respectively, and that of the complex 72%, representing additivity. The aromatic region is non-additive, with the characteristic positive dichroism of troponin at 295 nm disappearing, and the complex being altogether more negative from 260—275 nm than either component, indicating that conformational rearrangements of at least the aromatic side-chains have occurred.

Non-chromophoric Proteins. A significant change in the peptide c.d. of α-elastin solubilized in trifluoroethanol is effected by Ca^{2+}.[425] Since the effect is also found in the presence of 4% trifluoroacetic acid, metal binding at neutral sites is implicated, effects due to direct perturbation of the n–π^* by the metal ion being considered of secondary importance. More generally, the absence of significant perturbation of peptide c.d. is taken as evidence of the maintenance of identical overall conformation, though subtler conformational effects may be involved with specific ligands. Thus while histidine, as end-product inhibitor, affects a difference

[416] A. M. Tamburro, A. Scatturin, E. Grazi, and S. Pontremoli, *J. Biol. Chem.*, 1970, **245**, 6624.
[417] A. J. Tobin, *J. Biol. Chem.*, 1970, **245**, 2656.
[418] J. P. Halper, N. Latovitzki, H. Bernstein, and S. Beychok, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 517.
[419] M. A. Wells, *Biochemistry*, 1971, **10**, 4078.
[420] E. J. S. Gravenmade, C. Van der Drift, and G. D. Vogels, *Biochim. Biophys. Acta*, 1971, **251**, 393.
[421] M. W. Makinen and H. Kon, *Biochemistry*, 1971, **10**, 43.
[422] H. Hamaguchi, A. Isomoto, Y. Miyake, and H. Nakajima, *Biochemistry*, 1971, **10**, 1741.
[423] M. Waks, P. C. Kahn, and S. Beychok, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1232.
[424] I. Staprans and S. Watanabe, *J. Biol. Chem.*, 1970, **245**, 5962.
[425] D. W. Urry, J. R. Krivacic, and J. Haider, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 6.

spectrum in phosphoribosyl transferase equivalent to the burying of several tyrosine residues, only minor effects are observed in the 210—225 nm region in c.d.[426] Indole increases the reactivity of sulphydryl groups in the alpha subunit of *E. coli* tryptophan synthetase, but indole glycerol phosphate has the opposite effect; neither affects the far-u.v. c.d.[427] The activation of α-amylase from *B. subtilis* by Ca^{2+} can also be effected by a range of lanthanide ions, with the activation decreasing with increasing ionic radius; no large structural changes are observed on removal of the activating ions.[428] Characterization of these subtler effects requires more-sensitive probe techniques.

Conformational effects associated with the binding of substrates and cofactors, observed by changes in far-u.v. properties, are found with the dissociation of ATP from bovine cardiac globular actin by treatment with Dowex 1 and obligatory dialysis against H_4edta,[429] and with the binding of aspartate to the allosteric phosphoenol pyruvate carboxylase from *E. coli*.[430]

The α-crystallin from ocular lens has a very high proportion of β-structure (as well as highly resolved fine-structure in the aromatic region). γ-Crystallin has similar properties, although the far-u.v. c.d. includes an additional negative band at 203 nm, with predominant β-structure, and a functional significance for this form in producing the transparent macrostructure of lens is suggested.[431] The phosphoprotein phosvitin is known to have a random conformation at pH 3—5, and to adopt a β-structure at pH 2.[432] This form is also induced reversibly at pH 3—5 by addition of alcohols, glycols, or dioxan.[433] At pH 2, the organic solvents intensify the c.d. at 195 nm but do not affect the magnitude at 222 nm. Also, at pH values where phosvitin itself would be random, addition of poly-L-lysine to give a molar equivalence between charged lysine and phosphate residues induces β-structure. The process is more efficient with high-molecular-weight polymer (M = 24 000 and above); similar results are obtained with protamine.[351] This result indicates that two random polymers can interact in aqueous solution to form a structure of high order.

The tropomyosins of different species (crayfish, oyster, abalone, and blowfly) have been compared in their o.r.d. properties; the high proportion of α-helix found in the mammalian form is also found for these invertebrates. Molecular weight and electron-microscopic appearance indicate close similarities in subunit structure.[434] Far-u.v. o.r.d. properties have been used to characterize the ageing process of bovine mercaptalbumin.[435]

[426] F. Blasi, S. M. Aloj, and R. F. Goldberger, *Biochemistry*, 1971, **10**, 1409.
[427] W. B. Freedberg and J. K. Hardman, *J. Biol. Chem.*, 1971, **246**, 1449.
[428] G. E. Smolka, E. R. Birnbaum, and D. W. Darnall, *Biochemistry*, 1971, **10**, 4556.
[429] W. D. McCubbin and C. M. Kay, *Biochim. Biophys. Acta*, 1970, **214**, 272.
[430] T. Yoshinaga, K. Izui, and H. Katsuki, *J. Biochem.* (*Japan*), 1970, **68**, 747.
[431] H. A. Jones and S. Lerman, *Canad. J. Biochem.*, 1971, **49**, 426.
[432] G. Taborsky, *J. Biol. Chem.*, 1970, **245**, 1054.
[433] G. E. Perlmann and K. Grizzuti, *Biochemistry*, 1971, **10**, 258.
[434] E. F. Woods and M. J. Pont, *Biochemistry*, 1971, **10**, 270.
[435] H. J. Nikkel and J. F. Foster, *Biochemistry*, 1971, **10**, 4479.

A discrete product is formed, for which the *N–F* transition and the acid expansion of the native protein are continuous. This product is formed by disulphide interchange. Dissociation of multimers of urease to give the species with $M = 240\,000$ is effected by 90% propane-1,2-diol without conformational effects in $[m]_{235}$, indicating the mildness but effectiveness of the glycol as a dissociating agent.[436]

Chromophoric Proteins. NAD(P)-dehydrogenases. L-Glutamate dehydrogenase binds one molecule NADH per protomer in the presence of the substrate L-glutamate, with production of a positive c.d. band at 340 nm. In the presence of the allosteric effector guanosine triphosphate or Zn^{2+}, this band doubles in intensity;[437] assuming identical c.d. effect for all bound species, the unmasking of a second site for NADH by the effector is suggested. The positive band of the complex between enzyme and NADPH undergoes a remarkable inversion in the presence of L-glutamate, independently of the presence of GTP. The binding of NADPH to dihydrofolate reductase is accompanied by dichroism at 340 nm, plus a small perturbation of the whole aromatic region.[438] Conformational mobility, induction of increased c.d. in NADPH, and the possibility of inversion of the symmetry are characteristics of these NADH binding-sites.

Flavoproteins. A detailed study of the flavin chromophore in model compounds and flavoproteins shows that at least six vibronic bands corresponding to three π–π^* transitions occur within the total absorption and c.d. envelope; no n–π^* transitions were located.[439] Resolution of these bands is performed for model compounds in organic solvents, and provides the basis for assignments in aqueous solution. The magnitude and sign of the rotational strength of the long-wavelength vibronic components are sensitive to the interaction of the ribityl hydroxy-group with the isoalloxazine ring, and allows similarities within the flavin oxidases and dehydrogenases to be established. Resolution and reconstitution of the Shethna flavoprotein are reported; similarities with two other flavoproteins extend into the far-u.v.[440] No major differences are found in the protein when the flavin is reduced to the semiquinone or hydroquinone forms. Similarly, no strong perturbation of the tryptophan c.d. is found when apoenzyme is combined with a range of flavin analogues, suggesting that the tryptophan–flavin interaction inferred from other spectral data has no direct counterpart in c.d.[441] A similar extensive resolution of the flavin bands in reduced and oxidized mammalian lipoamide dehydrogenase shows that all features are accounted for in terms of bands present in FAD, whether the enzyme itself is reduced or not.[442] These gaussian resolutions bring a further dimension into the

[436] C. C. Contaxis and F. J. Reithel, *J. Biol. Chem.*, 1971, **246**, 677.
[437] J.-M. Jallon and M. Iwatsubo, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 964.
[438] J. H. Freisheim and L. D'Souza, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 803.
[439] D. E. Edmondson and G. Tollin, *Biochemistry*, 1971, **10**, 113.
[440] D. E. Edmondson and G. Tollin, *Biochemistry*, 1971, **10**, 124.
[441] D. E. Edmondson and G. Tollin, *Biochemistry*, 1971, **10**, 133.
[442] A. H. Brady and S. Beychok, *J. Biol. Chem.*, 1971, **246**, 5498.

use of the natural chromophoric probes, and the methods are of general applicability to all attached chromophores.

A similar distinction between the flavin c.d. properties of oxidases and dehydrogenases has been made on the basis of the properties of the pyridine nucleotide transhydrogenase from *Azotobacter vinelandii*.[443] The holo-enzyme (containing FAD) resembles FMN, with signs of the bands distinct from those of the Shethna protein. The importance of asymmetry and the ability of different but related bands to act independently are shown by the elimination of c.d. from the 430 nm band of D-amino-acid oxidase on binding *p*-aminobenzoate. The 380 nm band retains its c.d. and both bands are present in absorption.[444] Small perturbations of the flavin c.d. in *p*-hydroxybenzoate hydroxylase from *Pseudomonas putida* are effected by the substrate analogues benzoate and *p*-fluoro- (but not *p*-amino-)benzoate.[445] Also, changes in the oxidized enzyme c.d. with NADPH (anaerobically) indicate complex formation.

Pyridoxal enzymes. A similar effect to the above is observed with L-glutamate decarboxylase on binding cyclic analogues of L-glutamate. The characteristic pyridoxal phosphate band at 420 nm is slowly converted into one at 380 nm with loss of c.d., but addition of substrate restores the band and the activity.[446] Similar effects have been reported for enzyme–substrate complexes of (supernatant) aspartate aminotransferase.[447] The mitochondrial and supernatant enzymes [448] are closely similar in their peptide c.d., and, although the former enzyme shows more pronounced aromatic effects, no significant differences were observed between the subforms (A, B, C and α, β, γ) of the two enzymes. Similar behaviour in enzyme–substrate and –inhibitor complexes is found; the slow substrate *erythro*-β-hydroxyaspartate produces an intense negative band at 492 nm, characteristic of a quinone-like intermediate of the reaction scheme; the more-rapid substrate α-methyl aspartate produces absorption at 430 and 360 nm, with only the latter active in c.d.[448] Where these bands are linked by protonic equilibria, either a difference in binding of the two forms or a symmetrical environment (determined by rigidity or by a flexible, time-averaged effect) seems to be indicated for one form.

Conformational changes accompanying photo-oxidation of the supernatant enzyme are indicated by the virtual disappearance of dichroism when photo-oxidized apo-enzyme binds pyridoxamine phosphate, whereas the dichroism in the amino-form of the enzyme derived from photo-oxidized holo-enzyme is well-defined. Only histidine is destroyed in the

[443] H. W. J. Van den Broek, J. S. Santema, and C. Veeger, *European J. Biochem.*, 1971, **24**, 55.
[444] K. Yagi, N. Nodi, M. Nishikimi, and A. Kotaki, *J. Biochem. (Japan)*, 1970, **68**, 293.
[445] N. Teng, G. Kotowycz, M. Calvin, and K. Hosokawa, *J. Biol. Chem.*, 1971, **246**, 5448.
[446] L. P. Sastchenko, E. S. Severin, D. E. Metzler, and R. M. Khomotov, *Biochemistry*, 1971, **10**, 4888.
[447] V. I. Ivanov and M. Y. Karpeisky, *Adv. Enzymol.*, 1969, **32**, 21.
[448] M. Martinez-Carrion, D. C. Tiemeier, and D. L. Peterson, *Biochemistry*, 1970, **9**, 2574.

process.[449] A further example of the highly absorbing complex at 492 nm is given by cystathionine gamma-synthetase from *Salmonella* in the presence of the slow substrate *O*-succinylserine.[450] If the extended conjugation of the absorbing intermediate is twisted away from planarity, strong effects (*e.g.* see ref. 448) of either sign are to be expected.

Haem-proteins. The origins of rotational strength in the Soret transitions of haem proteins have been treated theoretically.[294, 295] The striking features of the system are (*i*) the wide range of aromatic groups with which the haem transitions interact (the distal His-64, His-97, Phe-33, Tyr-103, and Tyr-146) though, since asymmetry is the prime consideration, proximity as such, *e.g.* of the His-93 residue which acts as the fifth ligand, does not confer strong perturbing power, and (*ii*) the resultant spectra derive from two near-degenerate transitions of high rotational strength, whose relative magnitudes are determined by orientation of the haem relative to its perturbing environment and the degree and nature of the chemical substituents in the porphyrin ring. Variations in these factors will affect the resultant c.d. profile. The multiplicity of Soret Cotton effects, most clearly seen in c.d., can be explained for many haem proteins in terms of the overlap of these oppositely signed and unequal bands located at different wavelengths. Even the apparently symmetrical, oppositely signed couplet can be produced in this way and, in the absence of other evidence of intermolecular aggregation, the assignment of this type of spectrum to asymmetric stacking of chromophores is inconclusive when the chromophore itself contains near-degenerate transitions.

Examination of eight mammalian myoglobins shows that human and monkey myoglobins differ from the remainder in having lower helical content and a lower ratio for the c.d. values at 208 nm and 220 nm.[451] However, these species were still able to cross-react immunologically with several others; only sperm-whale myoglobin was distinct in its cross-reactivity. There is therefore little correlation between the cross-reactivity and the gross conformation as judged by helical content.

Comparison of a variety of haemoglobins shows the close similarity in conformation of mammalian haemoglobins; those of the lamprey and toad are noticeably different. In general, similarities between γ and β are seen, and these are distinct from α. Reduction of the 430 nm c.d. of deoxy-Hb by dialysis (and the smaller effect with oxy-Hb) may be reversed by adding back 2,3-diphosphoglycerate. Carboxypeptidase treatment also diminishes the Soret c.d. of deoxy-Hb. The protoporphyrin complex with apo-Hb has distinctive positive and slightly asymmetric Soret c.d., resembling Hb, but shows four distinctive bands at 510, 545, 575, and 630 nm (all positive), compared with the 535 and 575 nm bands of the haem, correlating with the

[449] M. Martinez-Carrion, R. Kuczenski, D. C. Tiemeier, and D. L. Peterson, *J. Biol. Chem.*, 1970, **245**, 799.
[450] S. Guggenheim and M. Flavin, *J. Biol. Chem.*, 1971, **246**, 3562.
[451] M. Z. Atassi, *Biochim. Biophys. Acta*, 1970, **221**, 612.

extra transitions observed in the absorption spectra.[452] A number of unnatural haems complexed with apo-Hb have been found to have oxygen affinities which correlate inversely with the negative inductive effects of substituents in side-chains 2 and 4.[453]

The spectroscopic coupling between the haem associated with one chain and the aromatic residues of another, predicted by theory,[294] is found to be more significant in semi-haemoglobins lacking the haem from the α-chain,[454] since this species differs considerably from that of isolated β. By contrast, the species lacking haem from the β-chain closely resembles isolated α-chain in its Soret c.d.

In studying the effect of chloroform-saturated buffers on haem proteins, myoglobin is effectively inert, whereas oxy-Hb shows small effects in $[m']_{233}$. The largest effect is given by met-Hb with dichloromethane, when a 25% reduction is observed.[455] Similarly, the anaesthetic halothane, 2-bromo-2-chloro-1,1,1,-trifluoroethane, causes a reversible effect on $[m']_{233}$ of oxy-, deoxy-, and met-Hb, with the predominant effect, in isolated chains, being on the β-chain.[456]

In the 1:1 complex between haptoglobin and haemoglobin, the haem function is altered and is more susceptible to peroxidatic activity.[421] Structural alterations in the region of haem are observed by loss of the characteristic negative portion of the Soret c.d. of oxy-, deoxy-, and met-Hb, and by the red-shift of the maxima.[422] Even so, only small deviations from additivity of the two components are observed in the peptide and aromatic c.d. on complex formation, though the haem transition at 251 nm does indicate a change in environment.[423] The haptoglobin itself contains significant amounts of β-structure. In the serum β-glycoprotein haemopexin, when complexed with haem (for which it is transporter), the Soret effect (positive) is relatively symmetrical: like apo-Hb, there is evidence for a conformational change on binding haem, though this is much smaller.[457] Soret effects of both sign are found with cytochrome b_5 (microsomal, pig liver) where, for both ferrous and ferric forms, the predominant effect at 430 nm is negative.[458] Again, significant differences are found at 200—220 nm between the holo- and apo-protein, the latter having lost the aromatic fine-structure, and the holo-protein can be reconstituted.

Cytochrome *c* contains meso-haem as prosthetic group; the oxidized state of this cytochrome from a wide variety of sources is found to have very similar dichroic features, in the Soret region, though some differences exist in the α and β bands.[459] In the far-u.v., a significant amount of α-helix is indicated, and a minimum at 190 nm following the 194 nm

[452] Y. Sugita, M. Nagai, and Y. Yoneyama, *J. Biol. Chem.*, 1971, **246**, 383.
[453] Y. Sugita and Y. Yoneyama, *J. Biol. Chem.*, 1971, **246**, 389.
[454] R. Cassoly and R. Banerjee, *European J. Biochem.*, 1971, **19**, 514.
[455] D. J. Bucher and W. D. Brown, *Biochemistry*, 1971, **10**, 4239.
[456] L. H. Laasberg and J. Hedley-Whyte, *J. Biol. Chem.*, 1971, **246**, 4886.
[457] U. Muller-Eberhard and K. Grizzuti, *Biochemistry*, 1971, **10**, 2062.
[458] E. Schnellbacker and L. Lumper, *Z. physiol. Chem.*, 1971, **352**, 615.
[459] Y. P. Myer, *Biochim. Biophys. Acta*, 1970, **214**, 94.

maximum is a characteristic feature. The c.d. properties of the weak band at 650 nm associated with the reducibility of ferricytochrome *c* by ascorbate is found not to correlate with the behaviour of α and β bands as a function of pH.[460] The band can be eliminated without a major conformational change: its simulation with the undecapeptide of cytochrome *c* on addition of *N*-acetylmethionine suggests a specific liganding mechanism.[460] Cardiac cytochrome c_1 contains the same prosthetic group as cytochrome *c*, but is distinguished by a major positive band in the Soret region for the oxidized and reduced form and, in the latter, by the high degree of fine structure in the α and β bands at 500—570 nm.[461] These differences can be accounted for by relatively small differences in haem environment, following the derivation above, which would still allow considerable structural homology between the two systems.

Resolution of o.r.d. effects in the Soret region of cytochrome oxidase in oxidized and reduced forms indicates that the rotational strength of 0.5 DM is associated with an asymmetric c.d. profile, which may be represented by two relatively narrow, positive, gaussian bands.[462] The recent theoretical advances show this assignment to have empirical value only.[294] However, they do allow a fuller interpretation of the difference in rotational strength of the far-u.v. peptide transitions between oxidized and reduced forms. Since the haem chromophore does not derive rotational strength from interaction with the helical transitions (by analogy with myoglobin), these differences enforce the argument of a different helical content with different oxidation states. These differences have been directly observed in c.d.,[463] as have the asymmetric (and therefore multi-component) Soret c.d. effects in the oxidized, reduced, and CO-reduced complexes. The ability of deoxycholate to render the Soret peak of the reduced complex a symmetrical band signifies a change in environment of the chromophore, but not necessarily an environment of higher symmetry.

Small Cotton effects persist with alkali-denatured cytochrome oxidase, and the reduced form shows a considerable change in bandshape.[464] These effects are considerably smaller than those seen in haem complexed with α-helical poly-lysine of D- or L-configuration. These effects are complicated by the tendency of the components to aggregate. A simpler picture is presented by complexes involving protohaem, where the effects appear generally to be accounted for by oppositely signed bands, whose intensities and relative intensities change as a function of the titration of the polymer. The theoretical considerations [294] also help to enforce the inference of different conformation in the oxidized and reduced forms of cytochrome P-450; the Soret Cotton effect is present in oxidized and CO-reduced

[460] C. Greenwood and M. T. Wilson, *European J. Biochem.*, 1971, **22**, 5.
[461] C. A. Tu, F. C. Yong, L. Yu, and T. E. King, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 508.
[462] T. E. King, P. M. Bayley, and F. C. Yong, *European J. Biochem.*, 1971, **20**, 103.
[463] Y. P. Myer, *J. Biol. Chem.*, 1971, **246**, 1241.
[464] F. C. Yong and T. E. King, *European J. Biochem.*, 1971, **20**, 111.

P-450, but lacking in reduced P-450.[465] Since the oxidation state appears not to influence the orientation of the porphyrin Soret transitions,[294] this difference is attributable to a change in the environment of the haem. Likewise, the appearance of a negative Soret c.d. in lactoperoxidase, and its cyanide, azide, and fluoride complexes, indicates its characteristic environment.[466] Negative Soret c.d. is found for both oxidized (418 nm) and reduced (429 nm and 406 nm) cytochrome *b*-563 from larvae of the housefly.[467] Changes in aromatic and far-u.v. c.d. are observed with change of oxidation state. Cytochrome P-450 (the camphor methylene hydroxylase from *Pseudomonas putida*) can be studied in the presence of the substrate camphor, being inactive in the absence of the electron-donor putidaredoxin and its reductase, when significant changes in the negative Soret transition (from 418 to 392 nm) and in the α and β bands occur on binding camphor.[468]

The appearance of split, oppositely signed c.d. effects in the bacteriochlorophyll–protein complex of the green photosynthetic bacterium *Chloropseudomonas ethylicum* is at least partially due to molecular aggregation in an asymmetric stack; the split c.d. at 809 and 400 nm is lost at acidic and alkaline pH, and considerably modified by methanol. The bacteriochlorophyll chromophore itself shows marked asymmetric aggregation.[380, 469] The complex (M = 150 000) contains 20 bacteriochlorophyll molecules in four subunits, and associates further in the chloroplast. The chromophore–chromophore interaction here is of potential functional significance. The characteristic extrinsic effect (in o.r.d.) when hydrogen peroxide is added to horse-radish peroxidase is not observed at − 60 °C but appears on warming to 10 °C. The process is reversible, and interconvertible conformational states are suggested.[311]

Metalloproteins. A review of the state and function of copper in biological systems discusses studies of c.d. of the copper transitions in stellacyanin, plastocyanin, azurin, and *Polyporus* laccase.[470] The cupro-zinc protein from bovine erythrocytes is unique in having positive bands at 610, 440, and 350 nm; the presence of the metal stabilizes against urea denaturation.[471] Denaturation of spinach ferredoxin by urea (5 mol l^{-1}) is promoted by low ionic strength,[472] and the system can be reconstituted with restoration of the 400 nm extrinsic effects.[473] Perturbation of these transitions occurs when the protein combines with ferredoxin-NADP reductase, although there is no e.s.r. change in the reduced ferredoxin when complexed.[474]

[465] F. C. Yong, T. E. King, S. Oldham, M. R. Waterman, and H. S. Mason, *Arch. Biochem. Biophys.*, 1970, **138**, 96.
[466] R. J. Maguire and H. B. Dunford, *Canad. J. Biochem.*, 1971, **49**, 666.
[467] Y. Okada and K. Okunuki, *J. Biochem. (Japan)*, 1970, **67**, 603.
[468] J. A. Peterson, *Arch. Biochem. Biophys.*, 1971, **144**, 678.
[469] Y. D. Kim and B. Ke, *Arch. Biochem. Biophys.*, 1970, **140**, 341.
[470] R. Malkin and B. G. Malstrom, *Adv. Enzymol.*, 1970, **33**, 177.
[471] E. Wood, D. G. Dalgleish, and W. Bannister, *European J. Biochem.*, 1971, **18**, 187.
[472] D. H. Petering and G. Palmer, *Arch. Biochem. Biophys.*, 1970, **141**, 456.
[473] D. Petering, J. A. Fee, and G. Palmer, *J. Biol. Chem.*, 1971, **246**, 643.
[474] R. Cammack, J. Neumann, N. Nelson, and D. O. Hall, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 292.

The removal of metal from ferritin increases the intensity of 200—300 nm c.d., suggesting an increase in order, and a partial disordering at pH 2.8—2.5 is observed in the aromatic c.d.[475]

Visual pigments. An exciting experiment on bovine visual pigment 500 at 77 K shows circular dichroism effects associated with the photochemical processes following illumination, thus extending the concepts of asymmetry to the investigation of different absorbing species involved in the visual process.[310] Illumination at 500 nm generates large c.d. effects (state-I) with 560 nm (positive) and 480 nm (negative), which persist in the absence of illumination; a second stable species (state-II) is obtained by further illumination at greater than 560 nm, and this is characterized by effects of sign opposite to state-I. Interconversion of the two is possible by appropriate illumination, with an isodichroic point at 512—518 nm. The native pigment is not isodichroic with this system, nor can it be regenerated from state-I or state-II. While the exact nature of these species awaits elucidation, a new means of studying these complex interconversions has now been established.

C.d. criteria have been used for establishing the regeneration of native rhodopsin following digitonin extraction, when it is found that phospholipid is necessary for optimal regeneration, possibly through stabilization of the opsin conformation.[476] Extrinsic effects are observable in the 230 nm region in the complex of vitamin A with retinol-binding protein and prealbumin.[477]

Added Extrinsic Chromophores. Naphthalenesulphonate derivatives. A number of dyes related to anilinonaphthalenesulphonate (ANS) give characteristically different extrinsic c.d. when bound to apo-haemoglobin. Thus whereas ANS shows a negative band at 370 nm and 330 nm (− 0.08 DM, *i.e.* weak compared with haem),[294] 1-benzylaminonaphthalene-8-sulphonate shows no c.d., and *N*-iodoacetyl-*N'*-(8-sulphonyl-1-naphthoyl)-ethylenediamine shows a positive band (370 nm) and a negative (330 nm).[478] From the enhanced fluorescence yield on binding, it is inferred that all three bind at the same site, though the orientation of the naphthalene component is evidently different for each. Different c.d. properties are found for the bis-(ANS) derivative on binding to the H and M isozymes of lactic dehydrogenase; the hybrids show intermediate properties, but are not a proportionate combination of H and M, although only the two spectroscopic species are present.[479] The primary sites for covalent attachment of dimethylaminonaphthalenesulphonate (DNS) are the exterior NH_2 groups. By contrast with non-covalently bound ANS, DNS retains considerable mobility, and no extrinsic effects are found at low levels of

[475] G. C. Wood and R. R. Crichton, *Biochim. Biophys. Acta*, 1971, **229**, 83.
[476] H. Shichi, *J. Biol. Chem.*, 1971, **246**, 6178.
[477] L. Rask, P. A. Peterson, and I. Bjork, *Biochemistry*, 1972, **11**, 264.
[478] M. C. Hsu and R. W. Woody, *Biopolymers*, 1971, **10**, 1421.
[479] S. R. Anderson, *Biochemistry*, 1971, **10**, 4162.

labelling with fibrinogen [480] or lysozyme.[481] Higher degrees of labelling lead to irreversible conformational changes.

Purine and pyrimidine nucleotides. 5′-GMP, at concentrations below the level at which it induces aggregation to specific helical form, binds to poly-lysine at pH 7, when the polypeptide is random, to form a characteristic complex with a high degree of order.[482] The same complex is formed from the L- or D-polypeptide, indicating that the nucleotide provides the dominant stereochemistry in these diastereoisomeric complexes. The band is blue-shifted relative to that of the gel, possibly indicating a different geometry or a lesser size. The polypeptide evidently changes conformation, and only the guanosine nucleotide is effective.

Binding of the adenine cofactors and analogues ADP, dADP, or ATP to creatine phosphokinase produces a positive extrinsic effect at 260 nm, which is unaffected by metals or substrates.[483] The intensification of the cofactor c.d. (20-fold for ADP, plus a reversal of sign) is attributed to interaction with far-u.v. transitions of tryptophan, but the effect of immobilization must also be considered (see refs. 486 and 487). 6-Mercaptopurine ribonucleotide triphosphate (an ATP analogue) binds to G-actin without any change in its c.d., but the consequent polymerization to F-actin causes a three-fold increase.[484] The aromatic spectrum is intensified in the F-actin, and tyrosine contributions are evidently reversed; such conformational effects could alter the mobility of the purine cofactor also.

A detailed study of the interaction of ribonuclease-T-1 with 2′-GMP and 3′-GMP shows that the c.d. spectra of the enzyme and inhibitor are non-additive for the complex.[485] The enzyme shows tyrosine bands at 280 nm, with high intensity at 241 nm, and a single minimum at 210 nm, indicating little α-helix or β-structure. Following a sharp thermal transition at 40 °C, both tyrosine regions are lost. The possibility of coupling between the purine and aromatic residues is discussed in accounting for the generation of dichroism (280 nm negative, 250 nm positive) in the complex. An alternative interpretation, deriving from binding of 3′-GMP to RNase-T-1, is that binding freezes the nucleotide in the *syn*-conformation, the biphasic difference spectrum resembling an intensified version of the acidic spectrum of 3′-GMP.[486] The same immobilization mechanism is invoked to account for the enhanced c.d. of 2′-CMP bound to pancreatic ribonuclease A, where the bound form is designated the *anti*-conformation.[487]

Ribonuclease-St (from *Streptomyces erythreus*), differing from T-1 in having no tryptophan, will bind both guanosine and cytosine nucleotides.[488]

[480] E. Mihalyi and A. Albert, *Biochemistry*, 1971, **10**, 237.
[481] N. Okabe and T. Takagi, *Biochim. Biophys. Acta*, 1971, **229**, 484.
[482] K. Wulff, H. Wolf, and K. G. Wagner, *Biochem. Biophys. Res. Comm.*, 1970, **39**, 870.
[483] J. H. R. Kagi, T. K. Li, and B. L. Vallee, *Biochemistry*, 1971, **10**, 1007.
[484] A. J. Murphy, *Biochemistry*, 1971, **10**, 3723.
[485] C. Sander and P. O. Ts'o, *Biochemistry*, 1971, **10**, 1953.
[486] T. Oshima and K. Imahori, *J. Biochem.* (*Japan*), 1971, **70**, 197.
[487] T. Oshima and K. Imahori, *J. Biochem.* (*Japan*), 1971, **70**, 193.
[488] N. Yoshida, K. Kuriyama, T. Iwata, and H. Otsuka, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 954.

The free enzyme also shows a strong 234 nm tyrosine effect. 3′-GMP binds to produce additional effects as with T-1; a three-fold enhancement of 2′-(3′-)CMP c.d. on binding to RNase-A and a similar enhancement for this ligand with RNase-St are observed. Coupling with aromatics is again suggested. (A buried tyrosine is exposed when 3′-CMP binds to RNase-A.[489])

Discrepancies exist in the absolute magnitudes of the enhancements for these nucleotides. The intensifications observed in fully rigid model compounds suggest that immobilization could account for an enhancement of several fold over the value in free solution.[490]

Metals. The replacement of zinc in enzyme systems by cobalt(II) provides absorption and c.d. bands in the visible region whose properties reflect the conformational state of the active site.[491–493] In alcohol dehydrogenases, the metal and NADH (or phenanthroline) interact mutually;[491] in yeast aldolase, binding of fructose 1,6-diphosphate affects the cobalt bands (while the activator, K^+, has no effect), and the loss of activity below pH 8 is paralleled by a loss of asymmetry in the cobalt transitions;[492] in carboxypeptidase, the optical activity of the cobalt transitions is affected variously by a range of inhibitors—those with complex kinetics show multiple spectral changes, the most striking effect being with glycyl-L-tyrosine, which inverts the transitions, intensifying them several-fold, while still maintaining the same basic co-ordination to the metal.[493]

Other extrinsic effects. The reaction of 1.7 moles of tyrosine per mole of procarboxypeptidase with diazotized arsanilic acid produces an arsanilazo-procarboxypeptidase with extrinsic effects at 305, 400, and 500 nm associated with a broad, featureless absorption band.[494] Activation of the complex (M = 87 000) to carboxypeptidase (M = 34 600) doubles the intensity at 438 nm, and shifts the lower band to 525 nm. The kinetics of activation can be followed by these changes, which reflect conformational transitions associated with the complex activation process. The inhibitor β-phenylpropionate alters the c.d. of the labelled enzyme, but not the labelled precursor, suggesting that the labelled residues are close to the active site.

Induced optical activity at 320 and 280 nm is reported for thiamine pyrophosphate binding to transketolase from baker's yeast,[495] and at 245 and 280 nm for α- and β-naphthylamidines, which are competitive inhibitors

[489] E. R. Simons, *Biochim. Biophys. Acta*, 1971, **251**, 126.
[490] D. W. Miles, R. K. Robins, and H. Eyring, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **57**, 1138.
[491] D. E. Drum and B. L. Vallee, *Biochem. Biophys. Res. Comm.*, 1970, **41**, 33.
[492] R. T. Simpson, R. D. Kobes, R. W. Erbe, W. J. Rutter, and B. L. Vallee, *Biochemistry*, 1971, **10**, 2466.
[493] S. A. Latt and B. L. Vallee, *Biochemistry*, 1971, **10**, 4263.
[494] W. D. Behnke and B. L. Vallee, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 760.
[495] G. A. Kochetov, R. A. Usmanov, and V. P. Merzlov, *Biochemistry* (*U.S.S.R.*), 1970, **35**, 1085.

for trypsin.[496] The binding of the drugs chlorothiazide [497] and phenylbutazone, flufenamic acid, and dicoumarol [498] to serum proteins causes extrinsic effects which may be used to assess the stoicheiometry and affinity of the binding processes. Heterogeneity of binding is most likely for aromatic ligands with serum proteins, and the induction of c.d. may be different at different sites.

Differences in the binding of substrate analogues and non-metabolizable competitive inhibitors are indicated from the inability of the latter to generate extrinsic effects with *meta*-pyrocatechase, and their ability to reduce the effect at 317 nm which is characteristic of the anaerobic complex with catechol.[499]

The induction of c.d. in acridine orange on binding to a polypeptide with β-structure produces effects of opposite sign to those found on binding to α-helix;[500] with lower molecular weight polypeptides, different effects are observed, including inversion. The self-association of the ligand complicates the interpretation; such association may change the polymer conformation.

Binding of 4-(8-hydroxy-5-quinolylazo)-1-naphthalenesulphonate to bovine carbonic anhydrase generates extrinsic effects in the ligand. The presence of zinc, to which the ligand chelates, is obligatory.[501] Relatively non-specific, but still optically active, zinc complexes are formed in the presence of excess zinc, but these may be removed selectively by H_4edta. Binding of aquocobalamin to bovine serum albumin results in typical c.d. which can be simulated with corrinoid complexes with histidine, suggesting that this is the ligand for the sixth position in complexes with albumin.[502]

Chemical Modification and Peptide Cleavage. Chemical modification of tyrosine to the amino-derivative (*via* nitro-tyrosine) gives no net change of c.d. spectrum of the alkaline phosphatase of *E. coli*.[503] Formation of the di-isopropyl phosphofluoridate derivative of trypsin causes little change in c.d. spectra of the aromatics, while solvent perturbation spectra indicate fractional extra burying of both tyrosine and tryptophan.[504] Attachment of one 5-mercapto-2-nitrobenzoic acid residue to the sulphydryl of yeast inorganic pyrophosphatase, with retention of full activity, does not generate extrinsic dichroism, presumably because of effective symmetry of the chromophore through mobilization.[505]

[496] K. Tanizawa, S. Ishii, K. Hamaguchi, and Y. Kanaoka, *J. Biochem.* (*Japan*), 1971, **69**, 893.
[497] A. Breckenridge and A. Rosen, *Biochim. Biophys. Acta*, 1971, **229**, 610.
[498] C. F. Chignell and D. K. Starkweather, *Mol. Pharmacol.*, 1971, **7**, 229.
[499] F. Hirata, A. Nakazawa, M. Nozaki, and O. Hayaishi, *J. Biol. Chem.*, 1971, **246**, 5882.
[500] S. Ikeda and T. Imae, *Biopolymers*, 1971, **10**, 1743.
[501] R. W. Henkens and J. M. Sturtevant, *Biochemistry*, 1972, **11**, 206.
[502] R. T. Taylor and M. L. Hanna, *Arch. Biochem. Biophys.*, 1970, **141**, 247.
[503] P. Christen, B. L. Vallee, and R. T. Simpson, *Biochemistry*, 1971, **10**, 1377.
[504] G. B. Villanueva and T. T. Herskovits, *Biochemistry*, 1971, **10**, 4589.
[505] T. Negi, T. Samejima, and M. Irie, *J. Biochem.* (*Japan*), 1971, **70**, 359.

Non-chromophoric chemical modifications with conformational consequences are: reaction of Trp-140 in Staphylococcal nuclease with nitrophenylsulphenyl halides, when 40% reduction in helix content is observed (the Trp-140 residue being located between two helical regions) although immunological effectiveness is unimpaired;[506] and formation of the non-aggregated species maleyl-, glutaryl-, and succinyl-β-casein (a protein of low helical content), which can still show a temperature transition similar to that of the unmodified protein.[507]

Peptide cleavage allows examination of the conformation of a portion of primary structure in isolation: fragments of angiotensin-II,[411] ribonuclease,[508] lysozyme,[509] and myoglobin [510] can generally be induced to take on conformations different from those of the same sequence in the total structure by suitable choice of solvents. With sufficiently large fragments, *e.g.* the overlapping fragments 1—126 and 99—149 of Staphylococcal nuclease, the original conformations can be established, in this case even in the presence of 28 redundant residues, though the regenerated activity is low.[511] By contrast, lysozyme structure collapses on cleavage at two methionine residues, although reactivation may be possible.[512]

Fragments D and E of fibrinogen, obtained by plasmin treatment, are found to contain higher helical contents than the parent protein. D (M = 66 000) resembles fibrinogen in the near-u.v., whereas E (M = 22 000) is markedly different.[513] Such large fragments would be expected to retain most of their original conformation.

E. Nucleic Acid–Protein Complexes.—Two aspects are of interest with respect to protein conformation, namely the ability of synthetic polypeptides to form complexes and conformational effects in the nucleotide and protein components of the complex.

Model Systems. In forming the complex between poly-L-lysine and double-stranded polyadenylic acid, the polynucleotide c.d. is effectively halved when the ratio of lysine residue to phosphate is unity.[514] These changes may be due to base orientation or to large aggregate formation. However, they are in marked contrast to the behaviour of the complexes of the same polypeptide with DNA,[515] or with polyuridylic acid,[516] when a strikingly enhanced c.d. is observed (at lysine: phosphate = 0.7). Poly-lysine of

[506] I. Parikh and G. S. Omenn, *Biochemistry*, 1971, **10**, 1173.
[507] M. T. Evans, L. Irons, and M. Jones, *Biochim. Biophys. Acta*, 1971, **229**, 411.
[508] J. E. Brown and W. A. Klee, *Biochemistry*, 1971, **10**, 470.
[509] C. A. Benassi, R. Ferroni, M. Guarneri, A. Guggi, A. M. Tamburro, R. Tomatis, and R. Rocchi, *F.E.B.S. Letters*, 1971, **14**, 346.
[510] R. P. Singhal and M. Z. Atassi, *Biochemistry*, 1971, **10**, 1756.
[511] H. Taniuchi and C. B. Anfinsen, *J. Biol. Chem.*, 1971, **246**, 2291.
[512] Y. Ota, Y. Hibino, K. Asaba, K. Sugiura, and T. Samejima, *Biochim. Biophys. Acta*, 1971, **236**, 802.
[513] A. Z. Budzynski, *Biochim. Biophys. Acta*, 1971, **229**, 663.
[514] B. Davidson and G. D. Fasman, *Arch. Biochem. Biophys.*, 1971, **144**, 650.
[515] D. Carroll, *Biochemistry*, 1972, **11**, 421.
[516] D. Carroll, *Biochemistry*, 1972, **11**, 426.

very high molecular weight is relatively ineffective at producing these effects, as are poly-L-arginine and protamine. Similar enhanced c.d. effects have been observed with native DNA in the presence of f1 and f2a1 histones, and characteristic aggregated and extended structures have been observed by electron microscopy.[517] Formation of these structures depends upon the initial state of the protein component and is most effective for histone, which is largely unfolded.

Natural Systems. Three calf thymus histones have different conformations, which change in the presence of salt and organic solvents;[518] conformational changes in the binding of histone f2a1 to poly(vinylphosphate) and DNA are observed at a ratio of basic residues to phosphate of 0.3, and the peptide conformation is qualitatively helix-like, although the c.d. intensity is attenuated.[519] Conversely, total histone affects the polynucleotide conformation.[520] Using purified fractions, the lysine-rich histone f1 causes a change in polynucleotide c.d. on binding to DNA; phosphorylated f1 is less effective, suggesting a mechanism for controlling DNA conformation *via* phosphorylation.[521] The C-terminal fragment (M = 15 000), with a higher lysine content, is even more potent than f1, whereas the N-terminal fragment alone (M = 6 000) is ineffective, and hence in DNA complexes with f1 the N-terminal portion evidently modulates the action of the C-terminal portion.[522]

Conditions of ionic strength are clearly critical in determining interactions of histones and nucleic acids. The absence of conformational effects in forming nucleohistone IV (DNA plus histone f2a1) is reported, the complex being formed by dialysis from urea into 5 mM-tris.[523] However, annealing of the components in the presence of 5 M urea, followed by dialysis with 0.15 M-NaCl, allows formation of the characteristic complex, differing in c.d. properties from DNA.[524] The free histone is partially, helical in 0.14 M-NaF, but largely unfolded in 0.01 M-NaF. These properties suggest that although randomized histones are more effective at forming the complexes, as with the extended structures, the complexes once formed are stabilized by conditions under which they are more structured. Even so, the c.d. of bound f1, unlike f2a1, appears to be predominantly an extended form; the two histones act antagonistically, f1 being more effective at blocking changes due to f2a1 than *vice versa*.[525] The general picture is one of a subtle balance between conformational effects in both polynucleotide and polypeptide components.

[517] D. E. Olins and A. L. Olins, *J. Mol. Biol.*, 1971, **57**, 437.
[518] Y. H. Oh, *J. Biol. Chem.*, 1970, **245**, 6404.
[519] T. E. Wagner, *Nature*, 1970, **227**, 65.
[520] I. Fric and J. Sponar, *Biopolymers*, 1971, **10**, 1525.
[521] A. J. Adler, B. Schaffhausen, T. A. Langan, and G. D. Fasman, *Biochemistry*, 1971, **10**, 909.
[522] G. D. Fasman, M. S. Valenzuela, and A. J. Adler, *Biochemistry*, 1971, **10**, 3795.
[523] H. J. Li, I. Isenberg, and W. C. Johnson, jun., *Biochemistry*, 1971, **10**, 2587.
[524] T. Y. Shih and G. D. Fasman, *Biochemistry*, 1971, **10**, 1675.
[525] T. Y. Shih and G. D. Fasman, *Biochemistry*, 1972, **11**, 398.

F. Immunological Systems—*Immunoglobulins.* In a review of the molecular size and conformation of immunoglobulins, particular reference is made to the dissociation, unfolding, and reconstitution of the complexes, and the specificity involved in interactions between heavy and light chains.[283] The non-additivity of the two components in the complex, and the prevalence of β-structure, are particularly notable. Reconstitution of the heavy and light chains of IgG (myeloma and pooled) at pH 5.4, when both species are dimerized non-covalently, leads to an IgG product which is effective immunologically, but which resembles the original only when homologous pairing is involved.[526] Restoration of the original conformation is inferred from c.d. for the recombination of heavy and light chains from rabbit IgG, removing the requirement for unique complementarity between the two components which had previously been proposed.[527–529]

Studies on the variable and constant halves of κ-type Bence-Jones protein show the latter to be richer in β-structure, and the two halves to be additive at 230 nm, although non-additivity in the aromatic region suggests that some orientation of side-chains occurs between the two moieties.[530] Tryptic digestion of the human IgG and IgM allows comparison of their Fab and Fc fragments, showing the former to be similar, but the latter to differ in their β-structure content and their pH-dependent conformational changes.[531] The papain fragments of the atypical myeloma IgG (Sackfield) show that the Fab portion is unusually short and has a more intense $[m]_{225}$ trough.[532]

Comparison of IgG of widely different phylogenetic background shows the remarkable constancy of conformation in this class of proteins.[533] Studies on turtle,[534] bowfin,[535] and eel [536] are reported, and wider comparisons show only the lamprey IgG to contain any significant amount of α-structure. Otherwise the domination of the c.d. profiles by the 217 nm trough indicates the prevalent β-structure. Possible evolutionary significance of this structure is discussed; being an open structure, relatively few substitution sites would be necessary to achieve a variety of conformational changes suitable for promoting multi-domain interactions.[537]

[526] G. T. Stevenson and K. J. Dorrington, *Biochem. J.*, 1970, **118**, 703.
[527] I. Bjork and C. Tanford, *Biochemistry*, 1971, **10**, 1271.
[528] I. Bjork and C. Tanford, *Biochemistry*, 1971, **10**, 1280.
[529] I. Bjork and C. Tanford, *Biochemistry*, 1971, **10**, 1289.
[530] A. C. Ghose and B. Jirgensons, *Biochim. Biophys. Acta*, 1971, **251**, 14.
[531] A. C. Ghose, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1144.
[532] G. E. Connell, K. J. Dorrington, A. F. Lewis, and D. M. Parr, *Canad. J. Biochem.*, 1970, **48**, 784.
[533] G. W. Litman, R. A. Good, D. Frommel, and A. Rosenberg, *Proc. Nat. Acad. Sci. U.S.A.*, 1970, **67**, 1085.
[534] S. L. Chartrand, G. W. Litman, N. Lapointe, R. A. Good, and D. Frommel, *J. Immunol.*, 1971, **107**, 1.
[535] G. W. Litman, D. Frommel, J. Finstad, and R. A. Good, *J. Immunol.*, 1971, **106**, 747.
[536] B. Jirgensons, G. F. Springer, and P. R. Desai, *Comparative Biochem. Physiol.*, 1970, **34**, 721.
[537] G. W. Litman, D. Frommel, A. Rosenberg, and R. A. Good, *Biochim. Biophys. Acta*, 1971, **236**, 647.

Antibody–Hapten Interactions. The ability of the hapten ε-DNP-lysine to stabilize the Fab fragment of anti-DNP antibody against unfolding by 4 mol l^{-1} guanidine hydrochloride is indicated by the retardation of loss of 233 nm c.d. and the persistence of aromatic c.d.[538] The hapten *p*-azobenzenearsonate attached to the helical poly Tyr-Ala-Glu shows positive c.d. at 420 nm, and a smaller negative c.d. when attached to the random copoly Tyr,Ala,Glu, which also has random conformation.[539] On binding univalent antibody for the hapten, the c.d. of the random system increases two-fold, whereas that of the helical system inverts and shifts to 465 nm. Different antibody–hapten conformations are indicated; the conformational sensitivity of this extrinsic chromophore is also recognized in enzyme systems.[494]

Similar subtle effects are shown in the interaction of the haptens 2,4-DNP-lysine and 2,4,6-TNP-lysine with the mouse myeloma protein MOPC-315, and with pooled mouse anti-DNP and anti-TNP.[540] For either ligand, the c.d. spectrum for each antibody complex is distinct, and for any antibody, the c.d. spectrum with either ligand is different. The MOPC-315 complexes show more varied spectra, with bands of both sign. The development of weak c.d. in the DNP transition of the complex of DNP-lysine with MOPC-315 was used to establish 1.7 binding sites per mole of $M = 153\,000$.[541]

G. Antibiotics and Hormones.—Peptides of these types tend to be of intermediate size, often with cyclic structures, and with interesting conformational possibilities. The most striking development has been the proposal of the π (L–D) helix as the conformation for the membrane-active gramicidin-A.[542] The alternating sequence of D- and L-amino-acids can be fitted into a helix with intramolecular hydrogen-bonding running axially, and the diameter of the resulting cylindrical structure, determined by the number of residues per turn, could accommodate the cations whose transport is promoted by gramicidin-A. A head-to-head arrangement of two molecules would span the lipid bilayer.[543] Spectroscopic and chemical evidence in favour of the hypothesis has been presented.[544, 545] C.d. spectra are markedly solvent-sensitive, the spectrum in trifluoroethanol bearing some resemblance to a left-handed helix, and this is proposed to be the characteristic spectrum for the π (L–D) helix.[545]

The conformation of the membrane-active cyclic peptide alamethicin has been shown to be strongly solvent-sensitive, and to take on a highly

[538] R. E. Cathou and T. C. Werner, *Biochemistry*, 1970, **9**, 3149.
[539] A. Conway-Jacobs, B. Schechter, and M. Sela, *Biochemistry*, 1970, **9**, 4870.
[540] M. Glaser and S. J. Singer, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2477.
[541] J. M. Rockey, K. J. Dorrington, and P. C. Montgomery, *Nature*, 1971, **232**, 192.
[542] D. W. Urry, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 672.
[543] D. W. Urry, M. C. Goodall, J. D. Glickson, and D. F. Mayers, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1907.
[544] J. D. Glickson, D. F. Mayers, J. M. Settine, and D. W. Urry, *Biochemistry*, 1972, **11**, 477.
[545] D. W. Urry, J. D. Glickson, D. F. Mayers, and J. Haider, *Biochemistry*, 1972, **11**, 487.

ordered structure in apolar and lipid-like environments, when it is also known to be monomeric.[546]

The correlation between conformation and biological activity of gramicidin-S analogues shows that the inactive 5,5′-alanine derivative has an inverted o.r.d. profile, in contrast to the parent molecule. However, other replacements at this position, *e.g.* the disarcosyl derivative, are active, suggesting that the presence of proline at positions 5 and 5′ is not obligatory.[547] The 2,2′-diaminobutyric acid derivative (active) exhibits an even more intense o.r.d. spectrum than gramicidin-S itself.

Development of ordered structure in small peptides is not solely a function of chain length; bradykinin (9 residues) is random, secretin (27 residues) resembles lysozyme, and there is evidence of intramolecular interaction between the terminal regions, whereas cholecystokinin-pancreozymin (33 residues) is apparently random.[548] For these small and potentially mobile molecules, interaction with the receptor must also be important in defining biological specificity and activity. The sequence requirements for the interaction of the neurophysins with the hormones oxytocin and vasopressin have been studied by means of the effects of synthetic tripeptides analogous to the terminal regions of the hormone, on the 260—300 nm c.d. (mainly disulphide) of the neurophysins. It is found that the α-NH_2 and the nature of residues 2 and 3 are critical.[549]

H. Membrane Systems.—*Artefacts.* The distortion of spectroscopic properties of particulate systems by absorption flattening and light scattering has been recognized as a source of artefacts in c.d. and o.r.d. measurements on membrane systems. In treating these distortions quantitatively, expressions have been derived for the simultaneous occurrence of the two phenomena, by considering the real and imaginary parts of the particle refractive index and its intrinsic optical activity.[289, 290] The neglect of scattering (because Rayleigh scattering overestimates the scattering intensity from large particles) is not justified.[290] The flattening effect has been evaluated for three idealized shapes [550] and applied explicitly to the artefacts encountered with erythrocyte ghosts.[551] Explicit treatments of the two interrelated phenomena are given for spheres and ghost-like particles.[289]

Other treatments include scattering functions for left- and right-circularly polarized light,[552] and the experimental confirmation of distortions occurring (to different extents) in measurement of c.d. and (unpolarized) absorption

[546] A. I. McMullen, D. I. Marlborough, and P. M. Bayley, *F.E.B.S. Letters*, 1971, **16**, 278.
[547] T. Kato, M. Waki, S. Matsura, and N. Izumiya, *J. Biochem. (Japan)*, 1970, **68**, 751.
[548] A. Bodanszky, M. Bodanszky, E. J. Jorpes, V. Mutt, and M. A. Ondetti, *Experientia*, 1970, **26**, 948.
[549] E. Breslow, H. L. Aanning, L. Abrash, and L. Schmir, *J. Biol. Chem.*, 1971, **246**, 5179.
[550] D. J. Gordon and G. Holzwarth, *Arch. Biochem. Biophys.*, 1971, **142**, 481.
[551] D. J. Gordon and G. Holzwarth, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2365.
[552] D. W. Urry and J. Krivacic, *Proc. Nat. Acad. Sci. U.S.A.*, 1970, **65**, 845.

on membranes.[553] Thus, the distortion of c.d. of erythrocyte ghosts at 222 nm was found to be insignificant, no effects of fragmentation being observed, while at other wavelengths the red-shift and attenuation effects were found to vary with molecular aggregation as predicted by theory.[553] This subject has recently been reviewed.[554]

Membrane Proteins. Extraction of erythrocyte membranes with 6 mol l^{-1} guanidine hydrochloride and β-mercaptoethanol resolves a protein of $M = 200\,000$, so far studied only as a random coil in the presence of the denaturant.[555] The encephalitic protein extractable from normal human myelin retained a random conformation in the presence of a variety of detergents, and only lysolecithin at pH 5.0 affected the c.d., suggesting that lipid–protein interactions are important in determining conformation.[556] The proteins from human and bovine brain and spinal cord are closely similar in conformational properties.[557, 558] A lipoprotein extracted from brain by chloroform–methanol shows high helical content in this solvent, and even higher in trifluoroethanol.[559] The high proportion of apolar amino-acids and the relative absence of polar side-chains accounts for its solubility properties. A basic protein extracted from pig brain shows random conformation in water, and increasing helical content in 50% aqueous n-propanol and 90% trifluoroethanol.[560] The apo-lipoproteins from very-low-density lipoproteins of human plasma, distinguishable by their C-terminal residues, also exhibit characteristic conformational differences in aqueous solution.[561]

Phospholipid Micelles. Liposomal preparations of phosphatidylserine interact with random poly-L-lysine, converting it into a form with intense c.d. at 225 nm, closely resembling the α-helix.[562] The interaction, which can be reduced by increased salt concentration, converts the liposomes into vesicles and immobilizes the hydrocarbon chains. Inclusion of phosphatidylcholine in the liposomes reduced the effectiveness of generating the helical form.

An interesting micellar effect is found in which optical activity is induced in a symmetrical dye by inclusion in micelles made from an optically active surfactant. Sulphethidole with L- or D-*N*-decyl-*NN*-dimethylalanine hydrobromide, when peaks of opposite sign at 288 and 255 nm are found. The L-isomer gives the same signs as the dye on bovine serum albumin.[563]

[553] M. Glaser and S. J. Singer, *Biochemistry*, 1971, **10**, 1780.
[554] D. W. Urry, *Biochim. Biophys. Acta*, 1972, **265**, 115.
[555] J. T. Gwynne and C. Tanford, *J. Biol. Chem.*, 1970, **245**, 3269.
[556] J. S. Anthony and M. A. Moscarello, *Biochim. Biophys. Acta*, 1971, **243**, 429.
[557] L. P. Chao and E. R. Einstein, *J. Neurochem.*, 1970, **17**, 1121.
[558] Y. Oshiro and E. H. Eylar, *Arch. Biochem. Biophys.*, 1970, **138**, 606.
[559] G. Sherman and J. Folch-Pi, *J. Neurochem.*, 1970, **17**, 597.
[560] S. E. Kornguth and J. H. Perrin, *J. Neurochem.*, 1971, **18**, 983.
[561] W. V. Brown, R. I. Levy, and D. S. Fredrickson, *J. Biol. Chem.*, 1970, **245**, 6588.
[562] G. G. Hammes and S. E. Schullery, *Biochemistry*, 1970, **9**, 2555.
[563] J. H. Perrin and P. Idsvoog, *J. Pharm. Sci.*, 1970, **59**, 1525.

8 Dissociation and Association of Proteins

contributed by G. L. Kellett

In 1963, approximately thirty proteins were noted by Reithel [564] as likely to consist of subunits. It is a measure of how far the study of subunit systems has progressed that in 1971 Frieden [565] was able to draw a useful distinction between those enzyme systems which do or do not undergo reversible polymerization. Since the forces involved in either intra- or inter-enzyme interactions are very similar, the distinction is made on functional grounds, and Frieden has listed no less than thirty-one enzymes in which reversible association–dissociation processes may be of significance in controlling the rate of catalysis *in vivo*. We will first consider recent changes in some of the techniques which have made this advance possible.

A. Analytical Ultracentrifugation Techniques.—Most users of the analytical ultracentrifuge have adopted the sedimentation–diffusion equilibrium technique [566, 567] for quantitative work,[568–571] rather than velocity sedimentation. The latter still serves a powerful role in the detection of associating–dissociating systems through the effect of interaction on the shape of the boundary profile,[572–574] but its apparent relegation to a diagnostic role is unfortunate in view of the relative ease with which velocity sedimentation data may be quantitatively analysed.[575] The balance in the application of the two techniques is likely to be redressed, however, for whereas progress in sedimentation–diffusion equilibrium methods has been confined primarily to a consolidation of computer routines for data analysis,[576] two potentially invaluable velocity techniques, those of difference [577] and active-enzyme [578] sedimentation, have come to fruition.

Three recent sedimentation–diffusion equilibrium data analyses all avoid conversion of the c *vs.* r^2 distribution into one of M *vs.* c, as has previously been customary.[579] In contrast, the new methods analyse c *vs.* r^2 data directly by expressing the equilibrium concentration distribution as a sum

[564] F. J. Reithel, *Adv. Protein Chem.*, 1963, **18**, 124.
[565] C. Frieden, *Ann. Rev. Biochem.*, 1971, **40**, 653.
[566] D. A. Yphantis, *Biochemistry*, 1964, **3**, 297.
[567] E. G. Richards, D. C. Teller, and H. K. Schachman, *Biochemistry*, 1968, **7**, 1054.
[568] E. T. Adams, jun. and H. Fujita, in 'Ultracentrifugal Analysis in Theory and Experiment', ed. J. W. Williams, Academic Press, New York, 1962, p. 119.
[569] E. T. Adams, jun. and J. W. Williams, *J. Amer. Chem. Soc.*, 1964, **86**, 3454.
[570] E. T. Adams, jun., *Biochemistry*, 1965, **4**, 1646.
[571] E. T. Adams, jun. and D. L. Filmer, *Biochemistry*, 1966, **5**, 2971.
[572] G. A. Gilbert, *Proc. Roy. Soc.*, 1959, **A250**, 377.
[573] G. A. Gilbert and R. C. Ll. Jenkins, *Proc. Roy. Soc.*, 1959, **A253**, 420.
[574] G. A. Gilbert, *Proc. Roy. Soc.*, 1963, **A276**, 354.
[575] E. Chiancone, L. M. Gilbert, G. A. Gilbert, and G. L. Kellett, *J. Biol. Chem.*, 1968, **243**, 1212.
[576] 'Advances in Ultracentrifugal Analysis', ed. D. A. Yphantis, *Ann. New York Acad. Sci.*, 1969, vol. 164.
[577] M. W. Kirschner and H. K. Schachman, *Biochemistry*, 1971, **10**, 1900, 1919.
[578] R. Cohen and M. Mire, *European J. Biochem.*, 1971, **23**, 267, 276.
[579] P. D. Jeffrey and J. H. Coates, *Biochemistry*, 1966, **5**, 489.

of the exponential distributions of the individual species present. Haschemeyer and Bowers [580] have tested a least-squares extension of Newton's method with simulated data. Kellett [581] has extended a method previously applied with success to gel chromatography and velocity sedimentation data,[575] and has analysed experimental data on the haemoglobin tetramer–dimer dissociation–association equilibrium. Dyson and Isenberg [582] have used an extension of Bay's work on moments and introduced a smoothing routine. They tested their analysis with mixtures of serum albumin and ovalbumin, while Kelly and Reithel have applied it to fit the β-lactoglobulin monomer–dimer system at isoelectric pH.[583] The self-association of chymotrypsinogen A [584] has been analysed using the routine of Teller *et al.*[585] Roark and Yphantis [586] have theoretically analysed the influence of the Donnan effect in sedimentation–diffusion equilibrium.

Difference Sedimentation. Richards and Schachman [587, 588] showed some years ago that the differences in two sedimentation coefficients could in principle be measured directly with great precision by utilizing the Rayleigh interference system to subtract the concentration distribution curves for two solutions contained in a double-sector ultracentrifuge cell. Kirschner and Schachman [577] have now identified the systematic errors which prevented the rapid development of the original study. These errors, which resulted from differences between double-sector compartments, alignment of the Rayleigh mask, and fringe displacements across the difference boundary owing to unequal dilution effects, may be overcome by a combination of devices, including the use of unfilled epoxy-cells, the design of a new Rayleigh mask and improvement in alignment procedures,[589, 590] and a single baseline correction. Fringe bowing, which occurs in the region of large concentration gradients, may be eliminated by measurement of the zero-order fringe selected by 'white light' exposure.

When applied to proteins in the molecular weight range 30 000—300 000 dalton, changes in sedimentation coefficient of 0.01 *s* can be measured readily to an accuracy of better than 5% (0.0005 *s*). The percentage accuracy is independent of the size of the difference measured. It is tempting to ascribe small changes in sedimentation coefficient upon binding ligand

[580] R. H. Haschemeyer and W. F. Bowers, *Biochemistry*, 1970, **9**, 435.
[581] G. L. Kellett, *J. Mol. Biol.*, 1971, **59**, 401.
[582] R. D. Dyson and I. Isenberg, *Biochemistry*, 1971, **10**, 3233.
[583] M. J. Kelly and F. J. Reithel, *Biochemistry*, 1971, **10**, 2639.
[584] D. D. Miller, T. A. Horbett, and D. C. Teller, *Biochemistry*, 1971, **10**, 4641.
[585] D. C. Teller, T. A. Horbett, E. G. Richards, and H. K. Schachman, *Ann. New York Acad. Sci.*, 1969, **164**, 66.
[586] D. E. Roark and D. A. Yphantis, *Biochemistry*, 1971, **10**, 3241.
[587] E. G. Richards and H. K. Schachman, *J. Amer. Chem. Soc.*, 1957, **79**, 5324.
[588] E. G. Richards and H. K. Schachman, *J. Phys. Chem.*, 1959, **63**, 1578.
[589] M. W. Kirschner, Ph.D. Thesis, Univ. of Calif., Berkeley, Calif., 1971.
[590] E. G. Richards, D. C. Teller, V. D. Hoagland, jun., R. H. Haschemeyer, and H. K. Schachman, *Analyt. Biochem.*, 1971, **41**, 215.

to induced conformational change, as is the case with aspartate transcarbamylase. However, a 3% increase in the sedimentation coefficient of glyceraldehyde 3-phosphate dehydrogenase upon binding of nucleotide to all four sites [591] can be accounted for within experimental error solely by increase in molecular weight and decrease in apparent specific volume.[577] Changes in sedimentation coefficient resulting from association–dissociation processes can in principle be resolved by studies of concentration dependence.

The wedge-window technique for the comparison of sedimentation coefficients, introduced by Gerhart and Schachman [592] and extended by Schumaker and co-workers,[593] has been applied to several proteins. Poillon and Fiegelson [594] have observed a 6.6% increase in $s_{20,w}$ upon the binding of the allosteric effector tryptophan to tryptophan oxygenase. Charlwood [595] has found that $s_{20,w}$ of apoferritin increases by 1.8% upon the binding of two atoms of iron per mole, while Atkinson and associates [596] have used the method for the accurate determination of the sedimentation coefficient of isocitrate dehydrogenase.

Terminology for Sedimentation and Gel-filtration Experiments. There is a need for an established terminology for the different kinds of experiments and boundaries which may be used in the formally analogous techniques [597, 598] of ultracentrifugation and gel chromatography, especially for measurements designed to increase sensitivity to changes in sedimentation coefficient, s, and elution volume, V. Kirschner and Schachman [577] have applied the term 'difference sedimentation' to the Rayleigh subtraction technique described above, and have preferred to reserve the original term of 'differential sedimentation' [588] for layering experiments with concentration boundaries.[599] A number of these experiments have also been performed by gel chromatography, both with polymerizing [575] and non-polymerizing systems,[600] by Gilbert and associates, who have defined several boundary types. Moreover, Gilbert [601] has also introduced the term 'differential gel-chromatography' independently to describe a technique for the measurement of a small difference, ΔV, in elution volume, in which a boundary is formed by layering.

The fluxes J_1 and J_2, across any given plane in the plateau region of each of two solutions forming a boundary, are given by $v_1 c_1$ and $v_2 c_2$, respectively; c is concentration and v is the velocity of molecules at the given concentration; v thus represents either sedimentation coefficient or elution volume

[591] R. Jaenicke and W. B. Gratzer, *European J. Biochem.*, 1969, **10**, 158.
[592] J. C. Gerhart and H. K. Schachman, *Biochemistry*, 1968, **7**, 538.
[593] V. Schumaker and P. Adams, *Biochemistry*, 1968, **7**, 3422.
[594] W. N. Poillon and P. Fiegelson, *Biochemistry*, 1971, **10**, 753.
[595] P. A. Charlwood, *Biochem. J.*, 1971, **125**, 1019.
[596] L. D. Barnes, G. D. Kuehn, and D. E. Atkinson, *Biochemistry*, 1971, **10**, 3939.
[597] G. K. Ackers and T. E. Thompson, *Proc. Nat. Acad. Sci. U.S.A.*, 1965, **53**, 342.
[598] G. A. Gilbert, *Nature*, 1966, **210**, 299.
[599] R. T. Hersh and H. K. Schachman, *J. Amer. Chem. Soc.*, 1955, **77**, 5258.
[600] G. A. Gilbert and G. L. Kellett, *J. Biol. Chem.*, 1971, **246**, 6079.
[601] G. A. Gilbert, L. M. Gilbert, C. E. Owens, and N. A. Shawky, *Nature New Biol.* 1972, **235**, 110.

since these are analogues.[597, 598] The difference in flux, ΔJ, is given by

$$\Delta J = J_1 - J_2 = v_1 c_1 - v_2 c_2 \qquad (1)$$

The distinction in the use of terms is then based on whether what is measured is the difference in flux, ΔJ, or the differential with respect to concentration, $\mathrm{d}J/\mathrm{d}c$. We may therefore distinguish between four boundary types: differential, finite-difference, difference, and integral. Their velocity parameters will be denoted by $\mathscr{V}$, V_Δ, ΔV, and V respectively.

By definition, a 'differential boundary' is one formed between two solutions differing by an infinitely small increment in concentration. The velocity of such a boundary is given by

$$\mathscr{V} = \mathrm{d}J/\mathrm{d}c \qquad (2)$$

This definition necessarily implies that no constituent disappears across the boundary. In practice, of course, a differential boundary cannot be studied directly. Use is therefore made of the 'finite-difference' boundary that is formed between two solutions which differ by a finite percentage, say 10—20%, of their average concentration, $(c_1 + c_2)/2$. The value of V_Δ is measured experimentally and extrapolation of V_Δ *vs.* c, for a series of boundaries formed at constant $(c_1 + c_2)/2$, to $c = 0$ gives the value of $\mathscr{V}$ for a differential boundary at $(c_1 + c_2)/2$. The velocity of a finite-difference boundary is given by the integral form of equation (2) as $V_\Delta = \Delta J/\Delta c$. When at least one, but not all, of the constituents disappears across the boundary, the difference is no longer finite. The boundary is therefore simply called a 'difference boundary'. This may be formed between two solutions of the same protein at identical concentration, but differing in the presence or absence of effector, for example, or between two solutions of closely related but distinct proteins, such as different haemoglobins. This forms the basis of the technique of 'difference gel-chromatography' (see below), in which ΔJ is measured directly. Since the experiment is designed such that $c_1 = c_2 = c$, equation (1) shows that

$$\Delta V = \Delta J/c = v_1 - v_2 \qquad (3)$$

The extreme case of a difference boundary occurs when all components disappear across it, that is $c_2 = 0$, and the boundary is between solution (c_1) and solvent. Combination of equations (1) and (3) gives the velocity of the boundary as

$$V = J_1/c_1 = v_1 \qquad (4)$$

Since v is the integral velocity corresponding to the velocity of molecules at the plateau concentration c, the boundary between solution and solvent is termed the 'integral boundary'. The 'differential boundary' has a special significance in boundary theory. Thus an integral boundary may be considered to be constructed from a series of differential boundaries, and

the application of equation (2) at a series of concentrations from the plateau to solvent will describe the integral profile shape.[602, 603]

We see then that 'differential sedimentation' measures $\mathscr{V}$ or the nearest accessible experimental quantity V_Δ, which represents the way in which J changes with c, by the use of a differential or finite-difference boundary, respectively. 'Difference sedimentation' directly measures the ΔJ between two independent integral boundaries by Rayleigh subtraction. The equivalent subtraction in 'difference gel-chromatography' is most easily achieved by layering to form a difference boundary. Since equation (4) identifies V with v, the resulting ΔV is given by equation (3) as $V_1 - V_2$. Thus each half of the 'difference boundary' may be regarded as an integral boundary. The wedge-window technique [592] measures directly neither a difference nor a differential. It is simply a technique which allows a more accurate comparison of the progress of two integral boundaries without the technical demands of difference sedimentation. The wedge-window technique might well be called 'comparative sedimentation'.

Active-enzyme Sedimentation. Cohen and Mire have now described [578] in full the technique of analytical active-enzyme centrifugation, in which the band sedimentation of enzyme through a substrate-containing solution is followed by the changes in absorbance caused by appearance of product or disappearance of substrate. The considerations to be observed include the correct enzyme amount in relationship to rotor speed so that the reaction can be monitored over the desired period, the physical conditions to afford convection-free band sedimentation, the relative concentrations of reactants to ensure a uniform rate of catalysis throughout the band, and the relative sedimentation rates of other enzymes in coupled assay systems. The utilization of the assay system for observation confers not only the obvious advantage of permitting the investigation of an enzyme in its active state, but also both selectivity, so that impure preparations can be studied, and sensitivity, which can be modulated by choice of observation wavelength. In a representative experiment, 10—20 μl of enzyme are applied at an initial concentration of 10 μg ml^{-1} to result in a concentration at band centre as low as 0.1 μg ml^{-1}. Clearly, this enormous sensitivity lends the method readily to the determination of the minimum size of the active unit in reversibly polymerizing systems under the influence of substrates or effectors. Thus, at the concentrations used for kinetic studies, the active form of glucose 6-phosphate dehydrogenase is found to be the dimer,[578] even though NADP promotes association to the tetramer at high concentrations.[604] Similarly, the 13*S*-oligomer of beef-liver glutamate dehydrogenase displays both glutamate and alanine dehydrogenase activities at concentrations

[602] L. M. Gilbert and G. A. Gilbert, *Biochem. J.*, 1965, **97**, 90.
[603] L. M. Gilbert and G. A. Gilbert, in 'Regulation of Enzyme Activity and Allosteric Interactions', Fed. European Biochem. Soc. Symp., ed. E. Kvamme and A. Pihl, Academic Press, London, vol. 73, 1968.
[604] R. Yue, E. Noltmann, and S. Kirby, *J. Biol. Chem.*, 1969, **244**, 1353.

which are so low that the effectors ADP and GTP have no influence on the state of polymerization. At higher concentrations, earlier work has shown that GTP promotes polymerization and glutamate dehydrogenase activity, while ADP promotes dissociation and alanine dehydrogenase activity.[605]

The optical absorption photoelectric scanning system [606] possesses clear attributes for following the sedimentation of active enzymes and other proteins selectively, although few applications have been reported.[607] This facility is particularly powerful when more than one wavelength is used, and has recently been applied twice in this way; first,[581] to ensure the integrity of ligand-free haemoglobin at very low concentrations, where it may readily be oxygenated or oxidized to methaemoglobin, by monitoring at the Soret peaks for each species in turn; secondly,[608] to demonstrate all-or-none dissociation of glyceraldehyde 3-phosphate dehydrogenase by mercurial, by monitoring the mercaptide absorption at 254 nm and the protein absorption at 280 nm.

B. Gel Chromatography.—The attributes of selectivity and sensitivity, possessed in common by all methods based upon absorption optics, are combined in the technique of gel chromatography with the virtue of relatively inexpensive apparatus. Moreover, gel chromatography has unique advantages for the study of interacting systems, namely: the use of frontal analysis provides both a leading and trailing profile, so that non-enantiographic effects resulting from interaction may be assessed; column parameters may be altered by use of gels of different porosity, so as to modify profile shapes; layering experiments for the formation of difference boundaries may be performed with the minimum of technical difficulty. Gilbert and Kellett [600] have shown that elution volume data on the interaction of ovalbumin and myoglobin are consistent with the formation of a 1 : 1 complex at pH 6.8, with a 0.01 mol l^{-1} phosphate buffer at 20.0 °C. By monitoring the column effluent at 280 and 410 nm, they were able to follow the constituent concentrations independently and to show that the centroids of the $\bar{V}_A^{\alpha}$ and $\bar{V}_B^{\alpha\beta}$ boundaries (Longsworth notation) [609] are not coincident. Although the separation of centroids in cases of this kind has been recognized for some while,[609] it has not previously been taken into account because it has not been possible to follow constituents independently. Neglect of this separation for ovalbumin and myoglobin is shown to lead to large errors in the calculated equilibrium constant. The ovalbumin–myoglobin elution profiles are markedly non-enantiographic, and, to the Reporter's knowledge, represent the only published example of profiles for a system which follows the general case (A + B $\leftrightharpoons$ C) predicted

[605] C. Frieden, *J. Biol. Chem.*, 1959, **234**, 809.
[606] H. K. Schachman and S. J. Edelstein, *Biochemistry*, 1966, **5**, 2681.
[607] I. Z. Steinberg and H. K. Schachman, *Biochemistry*, 1966, **5**, 3728.
[608] G. D. Smith and H. K. Schachman, *Biochemistry*, 1971, **10**, 4576.
[609] L. G. Longsworth, in 'Electrophoresis', ed. M. Bier, Academic Press, New York, 1959.

by Gilbert and Jenkins [573] in which $V_C < V_B < V_A$. Finite-difference boundary experiments were used to establish the dependence of the elution volumes of the free species upon concentration.

Difference Chromatography. The technique of difference gel chromatography has been developed by Gilbert [601, 610] for the precise measurement of a small difference, ΔV, in the elution volumes of two closely related proteins. This difference can then be related to the differences in hydrodynamic parameters, and also to differences in the subunit interaction energies of associating–dissociating systems. In the difference mode of gel chromatography, sufficient solution of the first protein is applied to the column to achieve a plateau in the effluent. The second solution, at identical protein concentration, is then layered onto the column directly. When the effluent is scanned at an isosbestic point for the two proteins, a net deviation from the plateau concentration, either a hump or a dip, is seen only if they differ in elution volume. The area of the net deviation gives a direct measure of the difference between the masses occluded by the column [equation (3)] in the steady states corresponding to saturation with each solution independently. The power of the technique is illustrated by its use in determining the dimer–tetramer association constant for the hybrid formed between met- and oxy-haemoglobins in dilute salt at neutral pH.

The value of the association constant for pure oxyhaemoglobin has been established by absolute methods to be 250 dl g^{-1}.[575] The exact value of the constant for methaemoglobin, relative to oxyhaemoglobin, was then established by difference experiments between oxy- and met-Hb. Further experiments in which oxy–met mixtures were layered over pure oxy provided the information to calculate the hybrid constant. The ratio of the association constant for met to that of oxy was found to be 0.62 ± 0.01, corresponding to a difference in dimer–dimer interaction energy of 1.11 kJ mol^{-1}. The ratio for the hybrid to oxy was 1.6, a value indicative of perfect dimer–dimer hybridization (when statistical factors are taken into account).

Zimmerman and Ackers [611, 612] have theoretically simulated the effects of column parameters such as axial dispersion, gel bead size, gel porosity, and flow rate on the changes in profile shapes as the protein moves through the gel. They emphasize that gel columns are always non-uniformly packed, so that the column partition coefficient for a species in the plateau region is not constant. As a consequence, the direct scanning approach for the study of developing profile shapes [613] is less efficient than its analogue, velocity sedimentation, but the necessary corrections may be made. Kerestes-Nagy and Orman [614] have reported that yeast enolase exists as an

[610] G. A. Gilbert, *Nature*, 1966, **212**, 296.
[611] J. K. Zimmerman and G. K. Ackers, *J. Biol. Chem.*, 1971, **246**, 1078.
[612] J. K. Zimmerman, D. J. Cox, and G. K. Ackers, *J. Biol. Chem.*, 1971, **246**, 4242.
[613] E. E. Brumbaugh and G. K. Ackers, *J. Biol. Chem.*, 1968, **243**, 6315.
[614] S. Kerestes-Nagy and R. Orman, *Biochemistry*, 1971, **10**, 2506.

active monomer at concentrations below 0.7 μg ml^{-1} at temperatures above 40 °C at pH 7.4. This is so even in the presence of Mg^{2+} and substrate, 2-phosphoglycerate, both of which are known to enhance the formation of dimer.[615] Monomers produced previously by high concentrations of salt were inactive.[615]

C. Light Scattering.—Light scattering has traditionally suffered from a lack of sensitivity both towards low concentrations and low molecular weights, primarily because of the lack of a sufficiently powerful light source. This situation has altered quite dramatically with the advent of the laser. The York group [616] have described the construction of a photometer employing a 50 mW argon-ion laser and have measured the excess scattering of lysozyme solutions over solvent with a range of only 5.3% at protein concentrations ($\sim$ 25 μg ml^{-1}) over one hundred-fold lower than those previously reported [617] to yield a molecular weight of (14 000 $\pm$ 240) dalton. The extension of the concentration scale facilitated extrapolation to zero concentration and a small negative dependence of $[Kc/R(90)]$ was observed, consistent with the existence of the monomer–dimer equilibrium previously reported.[618] The molecular weight distribution of Tamm–Horsfall urinary glycoprotein has been studied directly at concentrations as low as 2 μg ml^{-1}.[619] An alternative approach combined independent measurements of sedimentation and diffusion coefficients.[620] It is here that the use of the laser really comes into its own, for diffusion coefficients may be measured very precisely in as little as two minutes by intensity-fluctuation spectroscopy.[621] The two alternative studies of Tamm–Horsfall glycoprotein [619, 622] at the microgram level are in good agreement and yield a minimum molecular weight of about 14×10^6 dalton. The equilibrium constant of the myosin monomer–dimer system has been determined from measurements of the spectral broadening and intensity of scattered laser light.[623]

D. Transport Studies.—The interpretation of transport data on interacting protein systems depends greatly on the relative rates of attainment of equilibrium between species and their separation due to transport.[624] Only when the separation rate is much greater than that of equilibration will Schlieren peaks in the ultracentrifuge correspond to individual species. These peaks will be fully or partially resolved depending upon the relative

[615] T. H. Gawronski and E. W. Westhead, *Biochemistry*, 1969, **8**, 4261.
[616] J. A. Finnigan, D. J. Jacobs, and J. C. Marsden, *J. Colloid Interface Sci.*, 1971, **37**, 102.
[617] M. Halwer, G. C. Nutting, and B. A. Brice, *J. Amer. Chem. Soc.*, 1951, **73**, 2786.
[618] M. R. Bruzzesi, E. Chiancone, and E. Antonini, *Biochemistry*, 1965, **4**, 1796.
[619] J. A. Finnigan, D. J. Jacobs, and J. C. Marsden, *Biochim. Biophys. Acta*, 1971, **236**, 52.
[620] C. J. Oliver, E. R. Pike, A. J. Cleave, and A. R. Peacocke, *Biopolymers*, 1971, **10**, 173.
[621] R. Foord, E. Jakeman, C. J. Oliver, E. R. Pike, R. J. Blagrove, E. Wood, and A. R. Peacocke, *Nature*, 1970, **227**, 242.
[622] G. L. Kellett and H. Gutfreund, *Nature*, 1971, **227**, 921.
[623] T. J. Herbert and F. D. Carlson, *Biopolymers*, 1971, **10**, 2231.
[624] L. G. Longsworth and D. A. MacInnes, *J. Gen. Physiol.*, 1942, **25**, 507.

sedimentation coefficient of each species. On the other hand, when the equilibration rate is greater than that of separation, fully resolved peaks are never seen and only partial resolution is achieved for systems more complex than monomer–dimer. The latter shows only a single asymmetric peak.[572] However, if protein–solvent interactions are present, then, when the rate of equilibration is greater than that of separation, even a single protein which does not undergo self-association or dissociation may exhibit partial resolution.[624–626]

We see then that partial resolution may be a consequence of at least three fundamentally different alternatives and it thus becomes extremely important to distinguish between them. A case in point is the ultracentrifuge study by Morimoto and Kegeles [627] of the haemocyanin (17*S*)-monomer–(25*S*)-dimer equilibrium. At pH 9.4—9.7, instead of the expected single asymmetric peak, they observed partial resolution of the Schlieren pattern. They were able to show from pressure-dependence effects described below that equilibrium was achieved rapidly compared with separation. Since the presence of no other protein components could be detected, they attributed the partial resolution to the dependence of the equilibrium on Ca^{2+} concentration in the supporting solvent. Further studies on Ca^{2+} and pH dependence showed that the dimerization could be described by the equation:

$$2 \text{ monomer} + 5Ca^{2+} + 3H^{+} \rightleftharpoons \text{dimer}$$

The dimerization is accompanied by a large molar volume of reaction and is therefore pressure-dependent. A rapid increase in rotor speed for solutions overlayered with mineral oil caused an increase in size of the leading peak. The original Schlieren pattern was restored immediately upon return to the lower speed.

E. Kinetic Studies.—There is now a body of known associating–dissociating systems for which equilibrium between species is likely to be attained rapidly compared with their separation due to transport in the ultracentrifuge.[565] The partially resolved Schlieren peaks displayed by such systems do not correspond to individual species, for the dynamic equilibrium exists throughout the whole boundary. If it were possible by some means to increase the rate of separation to be greater than that of equilibration then, as noted earlier, the peaks would correspond to individual species and the areas would provide a direct measure of their equilibrium concentrations. One way to speed up 'separation' is to make use of an alternative technique, such as the use of a kinetic property which can be measured so quickly that equilibrium is 'frozen' during that period. Thus Gibson and co-workers [628] have found that above pH 10 in dilute salt,

[625] J. R. Cann, 'Interacting Macromolecules', Academic Press, New York, 1970.
[626] J. R. Cann and W. B. Goad, *J. Biol. Chem.*, 1965, **240**, 148.
[627] K. Morimoto and G. Kegeles, *Arch. Biochem. Biophys.*, 1971, **142**, 247.
[628] M. E. Anderson, J. K. Moffat, and Q. H. Gibson, *J. Biol. Chem.*, 1971, **246**, 2796.

deoxyhaemoglobin dissociates to dimers, while Kellett [629] has observed similar dissociation in 2 M-NaI at pH 7.0. The combination curve of haemoglobin with carbon monoxide in these conditions is biphasic. The fast phase is due to dimer, which combines at a rate some twenty-fold greater than that of the slow phase due to tetramer. (The fast rate is in fact the same as that for isolated chains, showing that the dimer is non-co-operative in ligand-binding reactions.)

At the concentrations studied, the rate of attainment of equilibrium between tetramer and dimer is slow ($t_{\frac{1}{2}}$ is several seconds) compared with the rate of carbon monoxide combination ($t_{\frac{1}{2}}$ = 50 ms) so that the amplitudes of the fast and slow phases directly represent the concentrations of dimer and tetramer in the deoxyhaemoglobin. Similar considerations are responsible for the observation of rapidly reacting haemoglobin in flash photolysis experiments.[630]

Ligand-induced Association–Dissociation. The addition or removal of ligand also induces further dissociation of tetramer or association of dimer, respectively.[581] Since the latter processes are usually slow compared with the former, they can be separated. Kellett and Gutfreund [622] have made use of this fact to measure the rate constant of dimer self-association (4.4×10^5 l mol^{-1} s^{-1}) from a change in extinction accompanying dimer to tetramer association induced by removal of oxygen. For some enzymes, the rate of ligand-induced polymerization may be comparable to the rate of catalysis and the latter will be dependent upon the former if the polymerization is between species of different catalytic activities. This seems to be so for phosphofructokinase,[631, 632] which shows a concentration-dependent lag in the initiation of catalysis, presumably owing to substrate-induced association of inactive subunits. This association may be dependent upon a prior conformational change. Recovery of activity is at present perhaps the most general of the techniques available for following reassociation on short time-scales, *i.e.* of the order of minutes. β-Galactosidase tetramer dissociates to monomer in 90% glycerol at pH 9.2. [633] Direct titration to pH 7 is accompanied by full restoration of activity within 2—3 minutes, which proceeds without the electrophoretically detectable intermediates. The reassociation of *N*(10)-formyltetrahydrofolate synthetase is cation-dependent and proceeds in two steps, the first having low cation specificity and the second having an absolute requirement for monovalent cations, especially NH_4^+.[634] Two methods which may offer more general physical means of following association–dissociation reactions are the use of a

[629] G. L. Kellett, *Nature New Biol.*, 1971, **234**, 189.
[630] S. J. Edelstein, M. J. Rehmar, J. S. Olson, and Q. H. Gibson, *J. Biol. Chem.*, 1970, **245**, 4372.
[631] E. C. Hulme and K. F. Tipton, *F.E.B.S. Letters*, 1971, **12**, 197.
[632] H. W. Hofer, *Z. physiol. Chem.*, 1971, **352**, 997.
[633] C. C. Contaxis and F. J. Reithel, *Biochem. J.*, 1971, **124**, 623.
[634] R. E. MacKenzie and J. C. Rabinowitz, *J. Biol. Chem.*, 1971, **246**, 3731.

fluorescent probe [635] and light scattering.[636] Radda and co-workers [635] have studied the AMP-induced association of phosphorylase *b* from the enhancement in fluorescence of bound 2-methylanilinonaphthalene-6-sulphonate. Tai and Kegeles [636] have investigated the temperature-dependent relaxation of haemocyanin association–dissociation by observation of scattered light at the 546 nm mercury line. There seems to be no reason why the replacement of the mercury arc with a powerful laser should not allow this scattering to be observed by rapid-reaction techniques for much smaller molecules than haemocyanin and at much lower concentrations.

F. Subunit Structure of Proteins.—It is now a routine procedure to determine the subunit structure of newly isolated proteins. Determination of subunit molecular weight is perhaps the most direct method. The table lists the molecular weights for native protein and subunits, together with the dissociating reagent used, for some of the proteins investigated in 1971.

Table Subunit structure of some proteins

Protein	*Native*	*Subunits*	*Number*	*Dissociating reagent*
Aequorin[a]	31 000	31 000	1	SDS
Lactose synthetase A protein[b]	42 900	44 000	1	SDS, 6 M-GHCl
Cyclic AMP receptor protein[c] (mediating cyclic-AMP-dependent gene transcription in *E. coli*)	45 500	25 500	2	6 M-GHCl
Haemagglutinin[d] (*Lens culinaris* A and B)	49 000	24 500	2	SDS, 6 M-GHCl
α-Amylase[e] (porcine pancreas)	50 000	25 000	2	SDS
ω-Amidase[f] (rat liver)	58 000	27 000	2	7 M-GHCl
Hydrogenase[g] (*Clostridium pasteurianum*)	60 000	30 000	2	SDS
Prealbumin[h] (human)	62 500	15 500	4	6 M-GHCl SDS
17β-Oestradiol dehydrogenase[i] (human placenta)	67 700	33 500	2	SDS
Thymidylate synthetase[j] (amethopterin-resistant *Lactobacillus casei*)	70 000	35 000	2	SDS

[635] D. J. Birkett, R. A. Dwek, G. K. Radda, R. E. Richards, and A. G. Salmon, *European J. Biochem.*, 1971, **20**, 494.
[636] M. S. Tai and G. Kegeles, *Arch. Biochem. Biophys.*, 1971, **142**, 258.

Table (*cont.*)

Protein	*Native*	*Subunits*	*Number*	*Dissociating reagent*
D-Glycerate dehydrogenase[k]	70 000	34 000	2	SDS
2-Keto-3-deoxy-6-phosphogluconate aldolase[l] (*Pseudomonas putida*)	73 000	24 000	3	SDS
2-Decylcitrate synthase[m] (*Penicillium epiculisporum lehman*)	90 000	45 000	2	SDS
Succinic dehydrogenase[n]	105 000	70 000 +27 000	1 1	SDS
Transketolase[o] (baker's yeast)	140 000	70 000	2	SDS
Aldolase[p] (*Mustelus canis*)	160 000	40 000	4	6 M-GHCl
Pyruvate kinase[q] (*Saccharomyces cerevisiae*)	161 000	19 600	8	SDS 6 M-GHCl
Invertase[r] (*Neurospora crassa*)	210 000	51 000	4	6 M-GHCl
Glycerol kinase[s] (*E. coli*)	217 000	56 000	4	SDS 6 M-GHCl
δ-Aminolevulinate dehydratase[t] (*Rhodopseudomonas spheroides*)	250 000	40 000	6	SDS
δ-Aminolevulinate dehydratase[u] (mouse liver)	250 000	39 500	6	SDS 6 M-GHCl
AMP deaminase[v] (rabbit muscle)	278 000	73 000	4	SDS 6 M-GHCl
Phenylalanyl t-RNA synthetase[w] (yeast)	280 000	75 000 +63 000	2 2	SDS
Anthranilate synthetase-anthranilate-5-phosphoribosyl pyrophosphate phosphoribosyltransferase[x] (*S. typhimurium*)	280 000	62 000	2	SDS 8 M-urea
Isocitrate dehydrogenase (DPN)[y] (yeast)	300 000	39 000	8	SDS
Aspartate transcarbamylase[z] (*E. coli*)	310 000	33 500 +17 000	3 3	SDS
Leucine aminopeptidase[aa]	327 000	57 000	6	SDS

Table (*cont.*)

Protein	*Native*	*Subunits*	*Number*	*Dissociating reagent*
N-Methylglutamate synthetase[bb] (*Pseudomonas* M.A.)	350 000	30 000	12	Succinylation 7 M-GHCl
Phosphoenolpyruvate decarboxylase[cc] (*E. coli*)	402 000	102 000	4	SDS 3.5 M-GHCl
Apoferritin[dd, ee] (horse spleen)	440 000—465 000	18 000—18 500	23—25	6 M-GHCl
		18 300—18 800	23—25	6 M-GHCl SDS

[a] Y. Kohama, O. Shimomura, and F. H. Johnson, *Biochemistry*, 1971, **10**, 4149; [b] I. P. Trayer and R. L. Hill, *J. Biol. Chem.*, 1971, **246**, 6666; [c] W. B. Anderson, A. B. Schneider, M. Emmer, R. L. Perlman, and I. Pastan, *J. Biol. Chem.*, 1971, **246**, 5929; [d] I. K. Howard, H. J. Sage, M. D. Stein, N. M. Young, M. A. Leon, and D. F. Dyckes, *J. Biol. Chem.*, 1971, **246**, 1590; [e] J. F. Robyt, C. G. Chittenden, and C. T. Lee, *Arch. Biochem. Biophys.*, 1971, **144**, 160; [f] L. B. Hersh, *Biochemistry*, 1971, **10**, 2884; [g] G. N. Nakos and L. E. Mortenson, *Biochemistry*, 1971, **10**, 2442; [h] L. Rask, P. A. Peterson, and S. F. Nillson, *J. Biol. Chem.*, 1971, **246**, 6087; G. Gonzalez and R. E. Offord, *Biochem. J.*, 1971, **125**, 309; C. C. F. Blake, I. D. A. Swan, C. Rerat, J. Berthon, A. Laurent, and B. Rerat, *J. Mol. Biol.*, 1971, **217**, 61; [i] D. J. W. Burns, L. L. Engel, and J. L. Bethune, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 786; [j] R. B. Dunlop, N. G. L. Harding, and F. M. Huennekens, *Biochemistry*, 1971, **10**, 88; [k] I. Y. Rosenblum, D. H. Antkowiak, H. J. Sallach, L. E. Flanders, and L. A. Fahren, *Arch. Biochem. Biophys.*, 1971, **144**, 375; [l] R. H. Hammerstedt, H. Mohler, K. A. Dekker, and W. A. Wood, *J. Biol. Chem.*, 1971, **246**, 2069; [m] A. Mahlen, *European J. Biochem.*, 1971, **22**, 104; [n] K. A. Davis and Y. Hatefi, *Biochemistry*, 1971, **10**, 2509; [o] C. P. Heinrich and O. Wiss, *F.E.B.S. Letters*, 1971, **14**, 251; [p] C. E. Caban and L. F. Hass, *J. Biol. Chem.*, 1971, **246**, 6807; [q] K. Ashton and A. R. Peacocke, *F.E.B.S. Letters*, 1971, **246**, 6807; [r] Z. D. Meachum, H. J. Colvin, and H. D. Braymer, *Biochemistry*, 1971, **10**, 326; [s] J. W. Thorner and H. Paulus, *J. Biol. Chem.*, 1971, **246**, 3885; [t] S. V. Heyningen and D. Shemin, *Biochemistry*, 1971, **10**, 4676; [u] D. Doyle, *J. Biol. Chem.*, 1971, **246**, 4965; [v] A. Boosman, D. Sammons, and O. Chilson, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1025; [w] J. Schmidt, R. Wang, S. Stanfield, and B. R. Reid, *Biochemistry*, 1971, **10**, 3264; [x] E. J. Henderson and H. Zalkin, *J. Biol. Chem.*, 1971, **246**, 6891; [y] L. D. Barnes, G. D. Kuehn, and D. E. Atkinson, *Biochemistry*, 1971, **10**, 3939; [z] V. Pigiet, E. A. Meighen, H. K. Schachman, J. P. Rosenbuch, and K. Weber, *J. Biol. Chem.*, 1971, **246**, 144; [aa] S. W. Mellbye and F. H. Carpenter, *J. Biol. Chem.*, 1971, **246**, 2459; [bb] R. J. Pollock and L. B. Hersh, *J. Biol. Chem.*, 1971, **246**, 4737; [cc] T. E. Smith, *J. Biol. Chem.*, 1971, **246**, 4234; [dd] I. Bjork and W. W. Fish, *Biochemistry*, 1971, **10**, 2844; [ee] C. F. A. Bryce and R. R. Crichton, *J. Biol. Chem.*, 1971, **246**, 4198.

SDS Gel Electrophoresis. SDS gel electrophoresis has become by far the most widely applied method. Originally introduced as an empirical procedure by Shapiro and co-workers,[637] and confirmed in its reliability with some forty proteins by Weber and Osborn,[638] the method has now acquired a theoretical basis for its success. Reynolds and Tanford have shown that at SDS monomer concentrations in excess of 8×10^{-4} mol l^{-1},

[637] A. L. Shapiro, E. Vanuela, and J. V. Maizel, *Biochem. Biophys. Res. Comm.*, 1967, **28**, 815.

[638] K. Weber and M. Osborn, *J. Biol. Chem.*, 1969, **244**, 4406.

all proteins bind 1.4 g SDS per g protein [639] and that the complex is a rod-like particle whose length is a unique function of molecular weight.[640] Charge effects resulting from the protein are apparently swamped by those from SDS, so that the charge per unit mass of the complex is constant. However, charge effects from the protein may be important for those which have undergone prior extensive chemical modification, *e.g.* by maleylation.[641] Recent variations of the technique include the covalent labelling of proteins with *NN*-dimethylanilinonaphthalene sulphonate, so that bands may be detected visually by fluorescence,[642, 643] and the introduction of a discontinuous buffer system to improve resolution.[644] The polypeptide chains within an oligomeric protein may be cross-linked by selective chemical modification of lysyl residues with dimethyl suberimidate.[645] Subsequent electrophoresis in SDS gels resolves a series of bands representing integral multiples from the protomer molecular weight up to the oligomer. Weber and Kuter [646] have demonstrated reversible denaturation of proteins by SDS and have achieved partial recovery of enzyme activity upon reconstitution of chains separated by SDS gel electrophoresis.

A unified theory of gel chromatography and electrophoresis has been presented in different forms. Morris and Morris [647] have confirmed the simple linear relationship between reduced electrophoretic mobility (ratio of the mobility at the measured gel concentration to that at zero gel concentration) and molecular-sieve partition coefficient proposed by them earlier.[648] Rodbard and Chrambach have provided an alternative formulation valid for non-spherical particles.[649, 650]

Measurements in Guanidine Hydrochloride. Gel chromatography on agarose in either SDS [651, 652] or 6 mol l^{-1} guanidine hydrochloride [653] represent alternative methods of determining subunit molecular weight,[653–655] though these suffer in comparison to SDS gel electrophoresis from a lack of resolution. Guanidine hydrochloride at 6 mol l^{-1} is, of course, a popular solvent for sedimentation equilibrium studies in which the slope of the

[639] J. A. Reynolds and C. Tanford, *Proc. Nat. Acad. Sci. U.S.A.*, 1970, **66**, 1002.
[640] J. A. Reynolds and C. Tanford, *J. Biol. Chem.*, 1971, **245**, 5161.
[641] J.-S. Tung and C. A. Knight, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 1117.
[642] K. R. Shelton, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 367.
[643] D. N. Talbot and D. A. Yphantis, *Analyt. Biochem.*, 1971, **44**, 246.
[644] D. M. Neville, *J. Biol. Chem.*, 1971, **246**, 6328.
[645] G. E. Davis and G. R. Stark, *Proc. Nat. Acad. Sci. U.S.A.*, 1970, **66**, 651.
[646] K. Weber and D. J. Kuter, *J. Biol. Chem.*, 1971, **246**, 4504.
[647] C. J. O. R. Morris and P. Morris, *Biochemistry*, 1971, **124**, 517.
[648] C. J. O. R. Morris, in 'Protides of the Biological Fluids', ed. H. Peeters, Elsevier, Amsterdam, 1967, vol. 14, p. 453.
[649] D. Rodbard and A. Chrambach, *Proc. Nat. Acad. Sci., U.S.A.*, 1970, **65**, 970.
[650] D. Rodbard and A. Chrambach, *Analyt. Biochem.*, 1971, **40**, 95.
[651] J. P. Rosenbuch and K. Weber, *J. Biol. Chem.*, 1971, **246**, 1644.
[652] W. W. Fish, J. A. Reynolds, and C. Tanford, *J. Biol. Chem.*, 1970, **245**, 5166.
[653] I. P. Trayer and R. L. Hill, *J. Biol. Chem.*, 1971, **246**, 6666.
[654] I. Bjork and W. W. Fish, *Biochemistry*, 1971, **10**, 2844.
[655] L. Rask, P. A. Peterson, and S. J. Nillson, *J. Biol. Chem.*, 1971, **246**, 6087.

log c *vs.* r^2 plot is given by $M_2(\partial\rho/\partial c_2)_\mu$. M_2 is the molecular weight of anhydrous protein and $(\partial\rho/\partial c_2)_\mu$ the reduced density increment at dialysis equilibrium with supporting solvent. This formulation is valid for the three-component system which exists at high guanidine hydrochloride concentrations.[656] $(\partial\rho/\partial c_2)_\mu$ is also equal to $(1 - \phi'\rho_0)$, where ϕ' is the apparent specific volume of the protein and ρ_0 the solvent density. Since ρ_0 is high for guanidine hydrochloride solutions, any error in assuming a value for ϕ', which is subject to the effects of preferential solvation, is greatly magnified. The only sure way to determine M_2 is therefore by the independent determination of $(\partial\rho/\partial c_2)_\mu$.[657] This is demanding experimentally if the high level of precision necessary in the determination of subunit molecular weights is to be achieved. Small changes in the partial specific volume of proteins as a result of exposure to denaturing solvents are often attributed to the effects of preferential interaction. There is, however, always a small volume change ($\Delta\bar{V} \approx -0.015$ ml g^{-1}) accompanying unfolding,[658] which should be taken into account.[659] Indeed, the change is sufficiently large that dilatometric studies may be used to follow denaturation.[660] The use of SDS in ultracentrifuge experiments has been reinvestigated.[661]

Hybridization Techniques. Isozymes. The hybridization of isozymes has provided much useful information about the subunit structure of those enzymes for which naturally occurring electrophoretic variants are available.[662] In general, for a protein containing n subunits, $(n + 1)$ bands will be observed upon electrophoresis of the hybridization mixture. Meighen and Schachman have recently extended this approach to those proteins for which variants are unavailable naturally, by their production through chemical modification.[663] Following earlier observations on haemerythrin,[664] they selectively succinylated lysyl residues and, from the resulting mixture of dissociated and undissociated protein, isolated the latter for use as a homogeneous electrophoretic variant.

Hybridization of native and succinylated aldolase resulted in the appearance of five bands, consistent with a tetrameric structure, as did glyceraldehyde 3-phosphate dehydrogenase.[665] Experiments utilizing both inter- and intra-subunit hybridization showed that aspartate transcarbamylase contains two catalytic subunits, each consisting of three

[656] E. F. Casassa and H. Eisenberg, *Adv. Protein Chem.*, 1964, **19**, 287.
[657] E. Reisler and H. Eisenberg, *Biochemistry*, 1969, **8**, 4572.
[658] M. Cassman and H. K. Schachman, *Biochemistry*, 1971, **10**, 1015.
[659] T. J. Bowen and M. G. Mortimer, *European J. Biochem.*, 1971, **23**, 262.
[660] S. Katz and J. E. Miller, *Biochemistry*, 1971, **10**, 3569.
[661] W. Barnett and S. P. Spragg, *Nature*, 1971, **234**, 191.
[662] C. L. Markert, *Ann. New York Acad. Sci.*, 1968, **151**, 14.
[663] E. A. Meighen and H. K. Schachman, *Biochemistry*, 1970, **9**, 1163.
[664] S. Kerestes-Nagy, L. Lazer, M. H. Klapper, and I. M. Klotz, *Science*, 1965, **150**, 357.
[665] E. A. Meighen and H. K. Schachman, *Biochemistry*, 1970, **9**, 1177.

chains.[666] Maleylation [667] provides an alternative to succinylation and has the advantage of being readily reversible.[668] The construction of electrophoretic variants by genetic manipulation has been used to show that both histidinol dehydrogenase and 6-phosphogluconate dehydrogenase are dimers.[669]

Haemoglobins. The formation of hybrids containing active and inactive subunits provides a means whereby subunit interactions and the contributions made by individual subunits to oligomeric structures may be assessed. For example, artificially prepared haemoglobin intermediates of the type $\alpha_2\beta_2^{CN+}$ or $\alpha_2^{CN+}\beta_2$ formed between normal and inactive cyanomet (CN+) subunits do not bind oxygen co-operatively but have an undiminished Bohr effect.[670] The production of succinylated variants allows other proteins to be studied in this way, for the succinylated subunits are inactive, not only in the pure species but also in the hybrids. The normal subunits are active within the oligomer and, in the case of aldolase, the ratios of the observed activities for each of the three hybrids to that of native aldolase are 1 : 2 : 3 : 4, respectively. Thus each subunit contributes equally and independently to enzyme activity, even though isolated unmodified subunits are inactive. Clearly, some subunit interaction responsible for activity exists within the oligomer. Similar observations have been made using naturally occurring aldolase variants by Penhoet and Rutter,[671] as well as for glyceraldehyde-3-phosphate dehydrogenase.[665]

Bacterial luciferase. The value of the approach is well illustrated by studies on bacterial luciferase, which consists of dissimilar subunits, α and β, of unknown function. Meighen and co-workers [672] found that hybridization of native and succinylated luciferase gave three bands upon electrophoresis, confirming the dimeric structure. Quantitative complementation of isolated α and β subunits with the inactive succinylated derivative, $\alpha_s\beta_s$, was used to prepare the hybrids $\alpha_s\beta$ and $\alpha\beta_s$. The former had at least half the specific activity of the native molecule, whereas the latter was inactive, suggesting that the subunits have different functions. Further studies showed that several catalytic centre properties were altered only in molecules possessing an altered α subunit. It was therefore concluded that the α subunit was involved in the catalytic steps of the bioluminescence reaction. The role of the β subunit is as yet unknown.

The bacterial luciferase consists of non-identical subunits and, as such, is distinct from the earlier class of enzymes which consist of apparently identical subunits, each contributing independently and equivalently to the

[666] E. A. Meighen, V. Pigiet, and H. K. Schachman, *Proc. Nat. Acad. Sci. U.S.A.*, 1970, **65**, 234.
[667] R. H. Hammerstedt, H. Mohler, K. A. Decker, and W. A. Wood, *J. Biol. Chem.*, 1971, **246**, 2069.
[668] P. J. G. Butler, J. I. Harris, and R. Lebermann, *Biochem. J.*, 1967, **103**, 78.
[669] K. K. Lew and J. R. Roth, *Biochemistry*, 1971, **10**, 204.
[670] M. Brunori, G. Amiconi, E. Antonini, and J. Wyman, *J. Mol. Biol.*, 1970, **49**, 461.
[671] E. E. Penhoet and W. J. Rutter, *J. Biol. Chem.*, 1971, **318**, 246.
[672] E. A. Meighen, M. Z. Nicoli, and J. W. Hastings, *Biochemistry*, 1971, **10**, 4062.

activity of the oligomer. However, even with these enzymes conditions can be chosen such that only half-of-the-sites activity is displayed. Koshland and co-workers have recently added cytidine triphosphate synthetase to the growing list of such enzymes.[673] In this enzyme, only half the glutamine sites bind the affinity label 6-diazo-5-oxonorleucine, whereas total glutamine activity is lost. The labelled enzyme, however, retains its activity towards ammonia and nucleotides. The phenomenon of negative co-operativity, in which ligand binding to the first site induces a conformational change which renders the second site incapable of reaction, may afford a general mechanism for half-of-the-sites activity. The principal condition for this behaviour appears to be the use of a relatively poor substrate, although alkaline phosphatase is in this respect an exception.[674]

Cold Inactivation. Reversible dissociation of active, oligomeric enzyme into inactive subunits provides one mechanism for the phenomenon of cold inactivation.[675] Several more enzymes have been added this year to the list reviewed by Utter and associates.[675] Cold inactivation of chicken-liver phosphofructokinase is accompanied by dissociation of a 14*S*-species to 5*S* fragments.[676] The rate of inactivation increases with decreasing temperature and decreasing concentration, while the reaction product, fructose 1,6-diphosphate, which is known to promote association,[677] protects against inactivation. Similarly, the allosteric modifier L-leucine protects against cold inactivation at low concentrations of L-threonine deaminase from *Rhodospirillum rubrum*.[678] Inactivation is accompanied by dissociation of tetramer to dimer. Fatty-acid synthetase from the mammary gland of lactating rat dissociates in the cold with loss of activity from a 13*S*- to a 9*S*-species.[679] Bovine-kidney arginosuccinase [680] behaves similarly in respect of cold inactivation to the enzyme from bovine liver.[681]

G. Protein–Small Molecule Equilibria.—The presentation by Klotz of the theory of equilibria between proteins and small molecules has been predominant in this field since it was first written.[682] This theory has been extended with a study by Klotz and Hunston [683] of the properties of graphical representations of multiple classes of binding sites. Thompson and Klotz [684] have also re-analysed the problem in terms of probabilistic considerations rather than equilibrium ones. A log–log plot of moles of

[673] A. Levitzki, W. B. Stallcup, and D. E. Koshland, jun., *Biochemistry*, 1971, **10**, 3371.
[674] D. R. Trentham and H. Gutfreund, *Biochem. J.*, 1968, **106**, 455.
[675] J. J. Irias, M. R. Olmsted, and M. F. Utter, *Biochemistry*, 1969, **8**, 5136.
[676] N. Kono and K. Uyeda, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 1095.
[677] A. Parmeganni, J. H. Luft, D. S. Love, and E. G. Krebs, *J. Biol. Chem.*, 1966, **241**, 4625.
[678] R. S. Feldberg and P. Datta, *European J. Biochem.*, 1971, **21**, 447.
[679] S. Smith and S. Abraham, *J. Biol. Chem.*, 1971, **246**, 6428.
[680] R. C. Bray and S. Ratner, *Arch. Biochem. Biophys.*, 1971, **146**, 531.
[681] E. A. Havir, H. Tamir, S. Rather, and R. C. Warner, *J. Biol. Chem.*, 1965, **240**, 3079.
[682] I. M. Klotz, in 'The Proteins', ed. H. Neurath, vol. 1, 1953, p. 727.
[683] I. M. Klotz and D. L. Hunston, *Biochemistry*, 1971, **10**, 3065.
[684] C. J. Thompson and I. M. Klotz, *Arch. Biochem. Biophys.*, 1971, **147**, 178.

ligand bound *vs.* free ligand is suggested as the most convenient way of presenting the binding data. Laiken and Nemethy [685] have adopted a new approach to the question of multiple binding in considering the binding of flexible ligands to proteins. The flexible ligand may adopt a number of configurations and as a consequence cannot be regarded as binding either independently or even in identical configurations, in contrast to the Klotz theory. The ΔpH method of Scatchard and Black for the determination of the binding of ionic ligands to proteins has been reinvestigated.[686] Cann has continued his outstanding series of studies on the interaction of aliphatic acids with proteins.[687]

H. Examples of Association–Dissociation Equilibria.—*Enzymes.* In addition to those systems already mentioned which display concentration-dependent association–dissociation phenomena, others investigated in 1971 include chymotrypsin,[688] phycocyanin,[689, 690] phosphoribosyl pyrophosphate synthetase,[691] fumarase,[692] follicle-stimulating hormone,[693] and D-amino-acid oxidase.[694] Systems which display concentration-dependent effects only in the presence of dissociating reagents include prealbumin [695] and aldolase A,[696] while lactate dehydrogenase in the presence of bis-(1-anilino-8-naphthalenesulphonate) exhibits association of the tetramer which can be reversed by the addition of NADH.[697] The reported dissociation of bovine fibrinogen [698] has been refuted.[699] The molecular weight of beef-heart mitochondrial ATPase has been determined to be 360 000 dalton,[700] the earlier and lower values of *ca.* 280 000 being attributed to inaccuracies in the assessment of dissociation phenomena. The rat-liver enzyme has been reported to have a molecular weight of 384 000 dalton,[701] raising the possibility that the enzyme species obtained from different sources are similar. Use has been made of the weak interaction between the A protein of lactose synthetase with α-lactalbumin to purify the A protein by affinity chromatography using an α-lactalbumin–sepharose column. The A protein binds in the presence of glucose and may be eluted by glucose omission.[653]

[685] N. Laiken and G. Nemethy, *Biochemistry*, 1971, **10**, 2101.
[686] C. J. Halfman and J. Steinhardt, *Biochemistry*, 1971, **10**, 3564.
[687] J. R. Cann, *Biochemistry*, 1971, **10**, 3707, 3713.
[688] K. C. Aune and S. N. Timasheff, *Biochemistry*, 1971, **10**, 1609.
[689] R. MacColl, J. J. Lee, and D. S. Berns, *Biochem. J.*, 1971, **122**, 421.
[690] O. Kao, D. S. Berns, and R. MacColl, *European J. Biochem.*, 1971, **19**, 595.
[691] I. H. Fox and W. N. Kelley, *J. Biol. Chem.*, 1971, **246**, 5739.
[692] J. W. Tiepel and R. L. Hill, *J. Biol. Chem.*, 1971, **246**, 859.
[693] B. B. Saxena and P. Rathnam, *J. Biol. Chem.*, 1971, **246**, 3549.
[694] K. Shiga and T. Shiga, *Arch. Biochem. Biophys.*, 1971, **145**, 701.
[695] L. Rask, P. A. Peterson, and S. F. Nillson, *J. Biol. Chem.*, 1971, **246**, 6087.
[696] C. J. Masters and D. J. Winzor, *Biochem. J.*, 1971, **121**, 735.
[697] S. R. Anderson, *Biochemistry*, 1971, **10**, 4163.
[698] F. C. Capet-Antonini and S. Guinand, *Biochim. Biophys. Acta*, 1970, **200**, 486.
[699] G. F. Endres and H. A. Scheraga, *Arch. Biochem. Biophys.*, 1971, **144**, 519.
[700] D. O. Lambeth, H. A. Lardy, A. E. Senior, and J. C. Brooks, *F.E.B.S. Letters*, 1971, **17**, 330.
[701] W. A. Catterall and P. L. Pedersen, *J. Biol. Chem.*, 1971, **246**, 4987.

Haemoglobin. Antonini and Brunori have reviewed the field of haemoglobin association–dissociation equilibria not long ago.[702] The comments here will therefore be limited to the most recent data available in so far as the latter are directly relevant to studies on the functional properties of haemoglobin. Detectable dissociation of the tetramer into $\alpha_1\beta_1$ dimers (Perutz notation) does not occur at erythrocyte concentrations, but does so within the technically accessible range. Although further dissociation of dimers into monomers at neutral pH has been reported in the past, a recent ultracentrifuge study of oxyhaemoglobin using photoelectric scanning optics has shown that at concentrations down to 1 μg ml^{-1} monomer formation does indeed occur in the absence of H_4edta, but that it is irreversible.[703] Monomer formation at pH 7 from oxy- or met-haemoglobin is inhibited by 10^{-3} mol l^{-1} H_4edta.

Bucci [704] was unable to find any monomer formation at the microgram level using complement fixation as a means of detecting hybrid formation which proceeds through monomer formation.[705] The presence of monomers need not normally be considered therefore in the interpretation of solution studies at pH 7.

An important problem in the elucidation of haemoglobin mechanism, therefore, is whether the dimer possesses the functional interactions responsible for co-operativity within the tetramer. One approach to this problem, developed by the Rome group, is to establish criteria of co-operativity based upon a comparison of the kinetic, spectral, and equilibrium properties of the co-operative tetramer with those of the non-co-operative, isolated chains. The 'dimer model' for haemoglobin is based, then, upon the finding that haemoglobin maintains its tetrameric characteristics even in high salt conditions where the dissociation of tetramer to dimer of both oxy- and deoxy-haemoglobin has been reported to be almost complete.[706–709] The 'dimer model' attributes a very high degree of co-operativity to the free dimer. However, since the maximum value of the Hill coefficient, n, is two for a dimer and the observed value in high salt is three, a thermodynamic paradox exists.[710] The excess is attributed to weak inter-dimer interactions, *i.e.* within the tetramer, in contrast to strong intra-dimer interactions.

An alternative explanation for the 'salt-paradox' has been recently advanced, based in part upon ultracentrifuge investigations of haemoglobin at microgram concentrations in high salt concentrations.[581] After

[702] E. Antonini and M. Brunori, *Ann. Rev. Biochem.*, 1970, **39**, 977.
[703] G. L. Kellett and H. K. Schachman, *J. Mol. Biol.*, 1971, **59**, 387.
[704] E. Bucci, *J. Mol. Biol.*, 1971, **55**, 401.
[705] J. Vinograd and W. D. Hutchinson, *Nature*, 1960, **187**, 216.
[706] A. Rossi-Fanelli, E. Antonini, and A. Caputo, *Adv. Protein Chem.*, 1964, **19**, 73.
[707] G. Guidotti, *J. Biol. Chem.*, 1967, **242**, 3673.
[708] N. M. Anderson, E. Antonini, M. Brunori, and J. Wyman, *J. Mol. Biol.*, 1970, **47**, 205.
[709] G. Amiconi, E. Antonini, M. Brunori, E. Chiancone, and P. Vecchini, *J. Biol. Chem.*, 1971, **246**, 1883.
[710] A. Rossi-Fanelli, E. Antonini, and A. Caputo, *J. Biol. Chem.*, 1961, **236**, 391, 397.

taking several precautions to prevent oxygenation or oxidation of haemoglobin during sedimentation velocity and equilibrium experiments, and making appropriate corrections for three-component systems,[656] no dissociation of deoxyhaemoglobin at neutral pH could be detected. Failure to observe the necessary precautions always resulted in low molecular weights and apparent dissociation, indicated by curvature of ln c *vs.* r^2 plots. Failure to correct for preferential solvation effects also resulted in low molecular weights. Dissociation of deoxyhaemoglobin was not detectable even at 10 μg ml^{-1} in either 2 M-NaCl or 1 M-NaI solution at pH 7. The latter electrolyte is a particularly powerful dissociating agent for oxyhaemoglobin and a comparison of the two led to the conclusion that the degree of dissociation of oxyhaemoglobin was *not less* than 10^5 times that of deoxyhaemoglobin. This finding provides an explanation for the salt paradox, since co-operative effects may take place in the tetrameric deoxy-form, followed by dissociation into dimers upon oxygenation. Furthermore, this independent study is in agreement with the published *X*-ray high-resolution models of haemoglobin which show that six inter-dimer hydrogen-bonds exist in the deoxy-form that are absent in oxyhaemoglobin.[711, 712]

Perutz has argued strongly that the tetramer is the functional co-operative unit of haemoglobin and has proposed a detailed mechanism in which the interaction energy of co-operativity is derived from the stepwise release of the constraining hydrogen-bonds in the deoxy quaternary structure as each chain is oxygenated.[713] This mechanism implies that the $\alpha_1\beta_1$ dimer is non-co-operative in ligand-binding reactions.

Several studies have been made of the properties of transient deoxyhaemoglobin dimers using rapid-reaction techniques. Some of the factors in their design have been outlined earlier in a consideration of the rates of association–dissociation reactions.

The reactions of deoxy-dimers have been studied using a difference in extinction between deoxy-dimer and tetramer in the Soret region,[622] by means of the change in tyrosine ionization above pH 10 accompanying tetramer dissociation,[714] and in flash photolysis studies which make use of differences in reactivity towards ligand.[630] These studies have shown that the kinetic and spectral properties of transient deoxy-dimers are the same as those of non-co-operative chains.

Conditions have recently been devised in which stable deoxyhaemoglobin dimers can be titrated with ligand. Kellett has shown that deoxyhaemoglobin exists in a tetramer–dimer dissociation–association equilibrium in 2 M-NaI at pH 7.0.[629] The value of n is concentration-dependent

[711] M. F. Perutz, H. Muirhead, J. M. Cox, L. C. G. Goaman, F. S. Matthews, E. L. McGandy, and L. E. Webb, *Nature*, 1968, **219**, 29.
[712] H. Muirhead and J. Green, *Nature*, 1970, **228**, 516.
[713] M. F. Perutz, *Nature*, 1970, **228**, 726.
[714] M. E. Anderson, J. K. Moffat, and Q. H. Gibson, *J. Biol. Chem.*, 1971, **246**, 2796.

and reversible, being about 2.0 at 84 μmol l^{-1} (10% dimer) and decreasing to 1.0 at 0.5 μmol l^{-1} (75% dimer). Perutz and associates have also come to the conclusion that the dimer is non-co-operative, from an independent and simultaneous study on des-Arg$^{141\alpha}$-haemoglobin.[715] Removal of Arg-141α by digestion with carboxypeptidase B results in the elimination of four of the six inter-dimer hydrogen-bonds, but leaves unchanged the structure of the $\alpha_1\beta_1$ dimer. Sedimentation velocity studies at pH 7 show that des-Arg$^{141\alpha}$-haemoglobin is dissociated into dimer in 0.9 M-$MgCl_2$ when the value of n is 1.0. It exists as a tetramer in 0.1 mol l^{-1} phosphate when the value of n is 2.0. This behaviour is reversible.

A thorough study of the role of association–dissociation equilibria in the co-operativity of lamprey haemoglobin has been presented by Anderson and Gibson.[716, 717] The co-operative characteristics of haemoglobin (containing about 192 haems!) from earthworm have also been studied.[718]

Cyclic-AMP-dependent Protein Kinases. The specific site of action of cyclic AMP in a physiological process has been identified for the first time, following a proposal by Krebs and co-workers concerning the mechanism of cyclic-AMP-dependent protein kinases in the regulation of phosphorylase *a* formation in skeletal muscle.[719] The occurrence of cyclic-AMP-dependent protein kinases is widespread and they possess a broad specificity. Thus the enzyme from muscle, in addition to histone and casein, will phosphorylate phosphorylase kinase and glycogen synthetase.

A series of recent publications from Krebs' laboratory has now dealt with the mechanism of the protein kinase in some detail.[719–723] The inactive form of the kinase consists of a regulatory and a catalytic subunit. Activation occurs as a result of dissociation of the subunits induced by binding of cyclic AMP to the regulatory subunit. The separation of these subunits of the protein kinases from adrenal cortical tissue,[724, 725] erythrocyte,[726] and skeletal muscle [727] has recently been achieved. The dissociation is fully reversible upon removal of cyclic AMP. The system

[715] J. A. Hewitt, J. V. Kilmartin, L. F. Ten Eyck, and M. F. Perutz, *Proc. Nat. Acad. Sci. U.S.A.*, 1972, **69**, 203.
[716] M. E. Andersen, *J. Biol. Chem.*, 1971, **246**, 4800.
[717] M. E. Andersen and Q. H. Gibson, *J. Biol. Chem.*, 1971, **246**, 4790.
[718] K. J. Boelts and L. J. Parkhurst, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 637.
[719] M. A. Brostrom, E. M. Reimann, D. A. Walsh, and E. G. Krebs, *Adv. Enzyme Regulat.*, 1970, **8**, 191.
[720] D. A. Walsh, J. P. Perkins, C. O. Brostrom, E. S. Ho, and E. G. Krebs, *J. Biol. Chem.*, 1971, **246**, 1968.
[721] D. A. Walsh, C. D. Ashby, C. Gonzalez, D. Calkins, E. H. Fincher, and E. G. Krebs, *J. Biol. Chem.*, 1971, **246**, 1977.
[722] E. M. Reimann, D. A. Walsh, and E. G. Krebs, *J. Biol. Chem.*, 1971, **246**, 1986.
[723] C. O. Brostrom, J. D. Corbin, C. A. King, and E. G. Krebs, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2444.
[724] G. N. Gill and L. D. Garren, *Biochem. Biophys. Res. Comm.*, 1970, **39**, 335.
[725] G. N. Gill and L. D. Garren, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 786.
[726] M. Tao, M. L. Salas, and F. Lipmann, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **67**, 408.
[727] E. M. Reimann, C. D. Brostrom, J. D. Corbin, C. A. King, and E. G. Krebs, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 187.

behaves as a '*V*' system as defined by Monod, Wyman, and Changeux, since cyclic AMP increases the V_{max} but not K_m for substrates. As a result, co-operativity is observed only for cyclic AMP, but not for substrates such as casein and ATP. The classic allosteric enzyme, aspartate transcarbamylase, had previously been unique in containing distinct subunits for catalytic and regulatory properties. ATCase does not dissociate except under the influence of group-specific reagents such as mercurial.[728]

I. Comment.—The particular examples of haemoglobin and cyclic-AMP-dependent protein kinases illustrate the study of associating–dissociating systems in two distinct ways.

The association and dissociation of the catalytic and regulatory subunits in protein kinases is important for *in vivo* regulation of activity. The association–dissociation reactions therefore have an intrinsic interest *per se* as a means of control. On the other hand, detectable dissociation of haemoglobin does not occur at erythrocyte concentrations. The association–dissociation reactions therefore have no physiological significance. However, since the equilibria established reflect directly the strength of the subunit interactions and quaternary constraints which exist within the oligomer, the association–dissociation reactions provide an important probe for the investigation of the tetramer structure–function relationships.

[728] J. C. Gerhart and H. K. Schachman, *Biochemistry*, 1965, **4**, 1054.

3
Peptide Synthesis

BY J. H. JONES AND B. RIDGE

1 Introduction

Those who provide the material for this chapter continue to be as busy as bees. Indeed the activities of the world's peptide chemists are in many respects comparable to the behaviour of the honey-bee. For example, when a foraging worker bee discovers a new source of food she returns to the hive and there performs a frantic dance, infecting the other workers with her enthusiasm so that they join in and then rush out to find the new food.[1] In the past year or so the new sources of sustenance have been provided, *inter alia*, by the elucidation of the structures of several synthetically accessible potent oligopeptides with important pharmacological effects (see Section 3). The dancing bees have quite rightly reported their discoveries with alacrity, and there are already signs that many workers are in hot pursuit—so much that in some cases different workers appear to be sucking at the same flower. The analogy can be developed further, but breaks down in one embarrassing particular: bees are able to produce a range of pharmacologically active peptides in relatively large amount with effortless superiority.[2]

Retrospective literature searching in peptide chemistry is greatly facilitated by the availability of such comprehensive sources as Greenstein and Winitz,[3] and Schröder and Lübke,[4] which cover up to the early 1960's. We hope that this series performs a similar function starting from the late 1960's. It can be seen, however, that there is a gap between these periods which has not been comprehensively dealt with, and news of a compilation by Pettit [5] of the peptide-synthesis literature published during this period was therefore very welcome. The book will serve a useful function but will no doubt provoke considerable criticism among its users. It is stated in the Introduction and Glossary that 'In general, nomenclature follows current

[1] K. von Frisch, 'The Dancing Bees', translated by D. Isle and N. Walker, Methuen, London, 2nd edn., 1966.
[2] C. R. Diniz and A. P. Corrado, in 'Pharmacology and Toxicology of Naturally Occurring Toxins', ed. H. Raskova, Pergamon Press, Oxford, 1971, vol. II, part IV.
[3] J. P. Greenstein and M. Winitz, 'Chemistry of the Amino Acids', Wiley, New York, 1961, vol. 2, ch. 10.
[4] E. Schröder and K. Lübke, 'The Peptides', vols. 1 and 2, Academic Press, New York, 1965 and 1966.
[5] G. R. Pettit, 'Synthetic Peptides', Van Nostrand Reinhold, New York, 1970, vol. 1.

IUPAC–IUB commission recommendations', but in fact the mode of use of abbreviated nomenclature deviates grossly from that which has been recommended: this is to be deplored in such a publication since it is obviously liable to perpetuate abusage. In addition many users would have found it more useful to have peptide derivatives with sequences related to any particular natural peptide listed in one place than to have entries for the most part classified according to the arbitrary criterion of chain length.

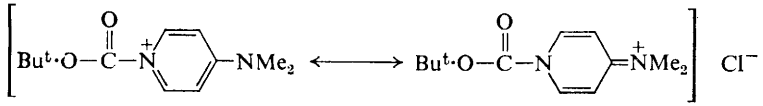
Wünsch [6] has written an authoritative, concise summary of the arguments for and against different strategies of peptide synthesis which is an admirable survey of current thinking and which will no doubt influence future development considerably, since many suggestions for improvements are made. A number of other general reviews of the subject have appeared, and the proceedings [7, 8] of two important symposia of general relevance have been published in full: the proceedings of two others have appeared in abstracted form.[9, 10] There have been a large number of new books, reviews, and conference proceedings concerned with biological applications of synthetic peptides—these are listed in Section 3.

2 Methods

A number of reviews have appeared.[10a–10g]

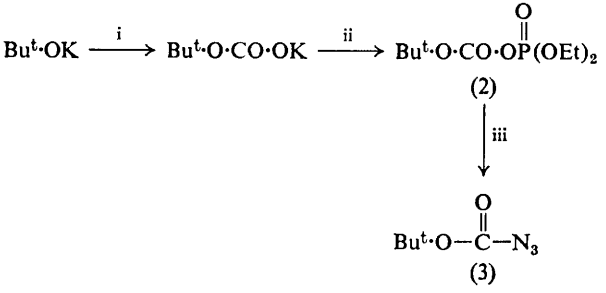
A. Protective Groups.—*Established Methods of Amino-group Protection.* A preliminary report [11] has appeared concerning a new water-soluble reagent which can be used for the t-butoxycarbonylation of amines. The reagent, 4-dimethylamino-1-t-butoxycarbonylpyridinium chloride (1), which is prepared by adding 4-dimethylaminopyridine to an excess of unstable t-butoxycarbonyl chloride, is claimed to be stable at 0 °C in the dry state,

[6] E. Wünsch, *Angew. Chem. Internat. Edn.*, 1971, **10**, 786.
[7] Proceedings of the Tenth European Peptide Symposium held at Abano, Italy, 1969, 'Peptides 1969', ed. E. Scoffone, North-Holland Publishing Co., Amsterdam 1971.
[8] Proceedings of a Symposium on the Chemistry of Peptides held at Santa Monica, California, 1970: *Intra-Sci. Chem. Reports*, 1971, **5**, issues 3 and 4 (not available at the time of writing: see vol. 5 of this Report for reference to individual papers).
[9] Abstracts of a Belgian–German Joint Biochemical Meeting on Peptides and Proteins held at Liège, Belgium, 1971: *Z. physiol. Chem.*, 1971, **352**, 1.
[10] Abstracts of papers presented at a Colloquium on Polypeptide Hormones—Structure and Function: *Biochem. J.*, 1971, **125**, 49P.
[10a] M. Oya and R. Katakai, *Yuki Gosei Kagaku Kyokai Shi*, 1971, **29**, 751 (*Chem. Abs.*, 1971, **75**, 141 129j).
[10b] B. Dugonjic and D. Kolbah, *Tehnika (Belgrade)*, 1970, **25**, 2373 (*Chem. Abs.*, 1971, **74**, 142 311d).
[10c] H. Yajima, *Yuki Gosei Kagaku Kyokai Shi*, 1971, **29**, 27 (*Chem. Abs.*, 1971, **75**, 6260e).
[10d] T. Wieland, ref. 7, p. 3.
[10e] B. J. Johnson, *Ann. Reports Medicin. Chem.*, 1969 (publ. 1970), 307.
[10f] R. Geiger, *Angew. Chem. Internat. Edn.*, 1971, **10**, 152.
[10g] T. Kato, H. Aoyagi, M. Waki, N. Mitsuyasu, and N. Izumiya, *Tampakushitsu Kakusan Koso*, 1971, **16**, 56, 139, 221, 293 (*Chem. Abs.*, 1971, **74**, 142 316j, **75**, 88 911e, 88 912f, 88 913g).
[11] E. Guibé-Jampel and M. Wakselman, *Chem. Comm.*, 1971, 267.



(1)

although it hydrolyses rapidly in neutral aqueous solution. Its chief advantage is that it reacts rapidly with the sodium salts of amino-acids (in aqueous solution at 25 °C at pH 10) to give good yields of the protected derivatives, when the reagent is used in equivalent molar quantity. The reagent may prove useful in semisynthetic studies for the introduction of t-butoxycarbonyl groups on to the free amino-groups of peptides in aqueous media. Another reagent for introducing the t-butoxycarbonyl protecting group is t-butylcarbonic diethylphosphoric anhydride (2),[12] which has recently been used as a precursor for the preparation of t-butoxycarbonyl azide (3). The procedure (Scheme 1),[13] although con-

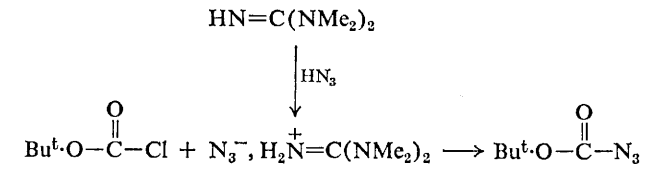


Conditions: i, CO_2; ii, $Cl-P(OEt)_2$; iii, KN_3

Scheme 1

venient, gives only a moderate overall yield of the azide. The direct preparation of this azide by the action of hydrazoic acid on the unstable chloroformate proceeds in only fair yield.[14] However, the action of tetramethylguanidinium azide (4)—an ionic azide which is soluble in organic solvents—on the chloroformate gives a nearly quantitative yield of crude (3), the purity of which is better than 98% (Scheme 2).[15] This method has also been applied to the preparation of phenyl and t-amyl azidoformates and, on account of its convenience, may become the method of choice for preparing (3).

[12] D. S. Tarbell and M. A. Insalaco, *Proc. Nat. Acad. Sci., U.S.A.*, 1967, **57**, 233.
[13] M. A. Insalaco and D. S. Tarbell, *Org. Synth.*, 1970, **50**, 9.
[14] H. Yajima and H. Kawatani, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 182.
[15] K. Sakai and J.-P. Anselme, *J. Org. Chem.*, 1971, **36**, 2387.



(4)

Scheme 2

The selective removal of *N*-t-butoxycarbonyl protecting groups in the presence of *N*-benzyloxycarbonyl groups can lead to partial cleavage of the *N*-benzyloxycarbonyl groups when conventional deblocking agents such as trifluoroacetic acid or hydrogen chloride in organic solvents are used (see ref. 16 for literature documentation). A comparative study of the efficacy of selective removal of such groups using a series of reagents[16] has revealed that the required selective cleavage is best achieved with either 70% aqueous trifluoroacetic acid or boron trifluoride diethyl etherate in the presence of glacial acetic acid.[17] In the case of the latter reagent at least equivalent molar quantities of the boron trifluoride reagent are required to remove *N*-t-butoxycarbonyl groups, whereas only catalytic amounts of the reagent are required to remove t-butyl ester protecting groups. Both the reagents mentioned cause slow cleavage of benzyl ester groups (particularly of γ-benzylglutamic acid derivatives). The boron trifluoride method is particularly suited to the removal of *N*-t-butoxycarbonyl groups from cysteine peptides which possess acid-labile *S*-protecting groups such as *S*-trityl[16,17] or *S*-diphenylmethyl,[17] both of which remain intact during the treatment. Hydroxylic side-chains require protection if the boron trifluoride method is to be applied, otherwise partial acetylation is likely to occur.[16] Mercaptoethanesulphonic acid (5) has been recommended as a

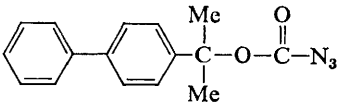
$$HS{\cdot}CH_2{\cdot}CH_2{\cdot}SO_3H$$

(5)

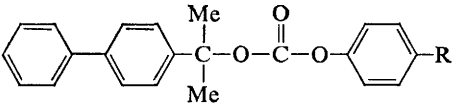
suitable reagent for the selective removal of *N*-t-butoxycarbonyl groups in the presence of *N*-benzyloxycarbonyl groups.[18] A moderate molar excess of the reagent in glacial acetic acid causes the complete cleavage of *N*-t-butoxycarbonyl groups within a few minutes at room temperature, without any detectable cleavage of the side-chain protecting groups of ε-benzyloxycarbonyl-lysine or of γ-benzyl glutamate. A paper has appeared on the selective removal of an *N*-t-butoxycarbonyl group in the presence of a sterically hindered t-butyl ester.[19]

[16] E. Schnabel, H. Klostermeyer, and H. Berndt, *Annalen*, 1971, **749**, 90.
[17] R. G. Hiskey, L. M. Beacham, tert., V. G. Matl, J. N. Smith, E. B. Williams, jun., A. M. Thomas, and E. T. Wolters, *J. Org. Chem.*, 1971, **36**, 488.
[18] A. Loffet and C. Dremier, *Experientia*, 1971, **27**, 1003.
[19] A. Michalik and B. Liberek, *Zeszyty Nauk., Mat., Fiz., Chem.*, 1970, **10**, 157 (*Chem. Abs.*, 1971, **74**, 134 155).

The 2-(*p*-biphenylyl)isopropoxycarbonyl amino-protecting group is usually introduced by means of the action of 2-(*p*-biphenylyl)isopropoxycarbonyl azide (6) or 2-(*p*-biphenylyl)isopropyl phenyl carbonate (7; R = H) on the benzyltrimethylammonium salt of an amino-acid in absolute



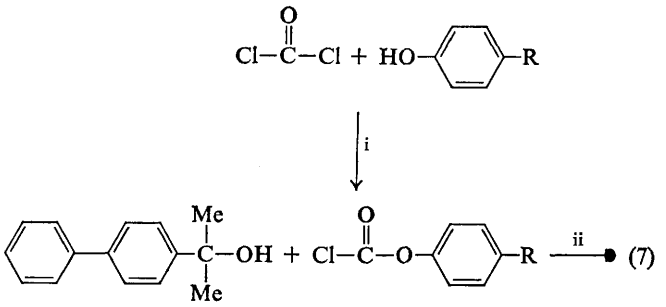
(6)



(7)

R = H, CO_2Me, COMe, or Ph

DMF. Three modified 2-(*p*-biphenylyl)isopropyl phenyl carbonates have been prepared [20] in which the phenyl moiety is substituted in the *p*-position by the electron-withdrawing substituents methoxycarbonyl (7; R = CO_2Me), acetyl (7; R = COMe), or phenyl (7; R = Ph). These compounds were synthesized by a method analogous to that used for the preparation of the parent phenyl mixed carbonate (Scheme 3). They are

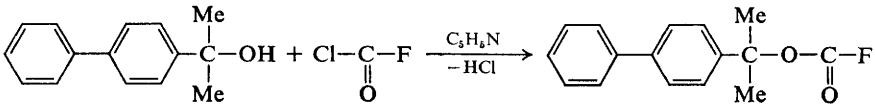


Conditions: i, $C_6H_5 \cdot NMe_2$; ii, C_5H_5N

Scheme 3

crystalline compounds, recrystallize well, and can be used to prepare 2-(*p*-biphenylyl)isopropoxycarbonyl amino-acids, under the anhydrous conditions mentioned above, in good yields. They are also more thermally stable than the parent compound and can be kept at room temperature for several months without appreciable decomposition. 2-(*p*-Biphenylyl)-isopropoxycarbonyl fluoride (8) has been prepared (Scheme 4), and although

[20] E. Schnabel, G. Schmidt, and E. Klauke, *Annalen*, 1971, **743**, 69.



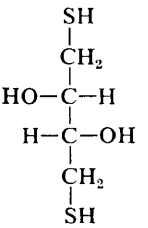
(8)

Scheme 4

it is only stable for a few weeks at − 70 °C, it can be used to acylate amino-acids in aqueous organic solvents between 0 and − 10 °C. A wide range of amino-acid derivatives was prepared with good yields.[20]

Citric acid, the reagent usually employed for acidifying aqueous solutions containing acid-labile amino-acid and peptide derivatives prior to extraction, can cause difficulties because of its solubility in the organic phase. This difficulty can be overcome by replacing the citric acid by an inorganic acid of comparable acidity, namely potassium hydrogen sulphate.[21]

The *N-o*-nitrophenylsulphenyl amino-protecting group is usually removed by treatment of the protected peptide with hydrogen chloride in ethyl acetate, or with thiols. Another reagent has been added to the latter category: Cleland's reagent [*threo*-2,3-dihydroxybutane-1,4-dithiol (9)]

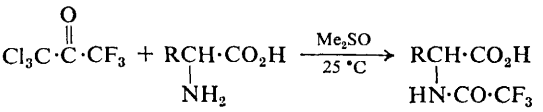


(9)

removes the *o*-nitrophenylsulphenyl group quantitatively from *o*-nitrophenylsulphenylalanine in two hours at pH 8—9 when two equivalents of the reagent are used.[22]

A method has been developed for the trifluoroacetylation of amino-acids and peptides under neutral conditions.[23] Trifluoroacetyl derivatives are of importance in peptide chemistry, since they form the basis of two methods of monitoring racemization tests, *viz.* gas–liquid chromatography and ^{19}F n.m.r. spectroscopy. The acyl group can be introduced by reaction of an excess of trichlorotrifluoroacetone in dimethyl sulphoxide with the amino-component (Scheme 5). The yields of the acylamino-acids obtained were moderate, but poor yields of two acyldipeptides were obtained. Two optically active amino-acids gave derivatives whose optical rotation values

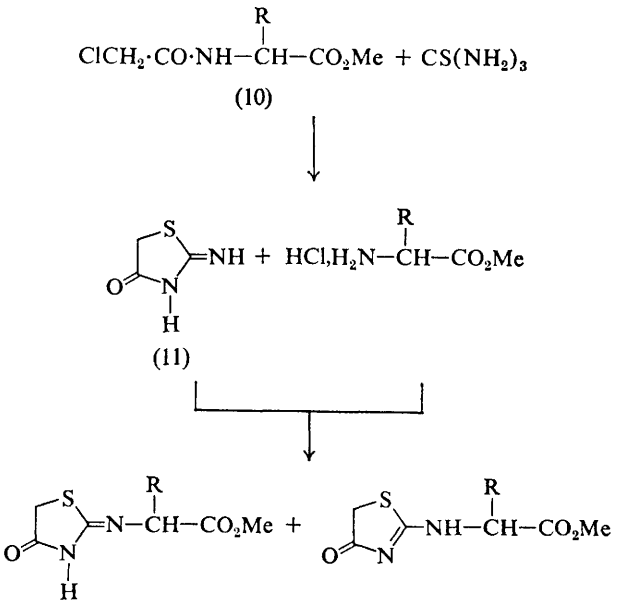
[21] R. Spangenberg, P. Thamm, and E. Wünsch, *Z. physiol. Chem.*, 1971, **352**, 655.
[22] K. P. Polzhofer and K. H. Ney, *Tetrahedron*, 1971, **27**, 1997.
[23] C. A. Panetta and T. G. Casanova, *J. Org. Chem.*, 1970, **35**, 4275.



Scheme 5

were comparable to the literature values. Clearly this study is preliminary in nature and the usefulness of this method will be evaluated later when attempts have been made to optimize yields.

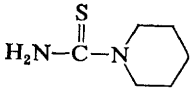
The removal of *N*-chloroacetyl protecting groups from amino-acids and peptides using thiourea often proceeds in unsatisfactory yield owing to secondary reactions. The nature of these side-reactions has been studied in the case of methyl *N*-chloroacetyl-L-valinate (10). It was shown that the 2-iminothiazolidinone (11), formed in the primary reaction, undergoes further reaction with the liberated amino-ester (Scheme 6). Steglich and



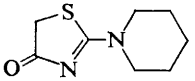
Scheme 6

Batz [24] therefore advocate the use of the *NN*-disubstituted thiourea 1-piperidinethiocarboxamide (12). The *N*-chloroacetyl residue can be removed quantitatively after boiling for several hours in ethanol with (12), the amine and 2-piperidinothiazolin-4-one (13) being produced without interference from side-reactions. A number of optically pure materials

[24] W. Steglich and H. G. Batz, *Angew. Chem. Internat. Edn.*, 1971, **10**, 75.



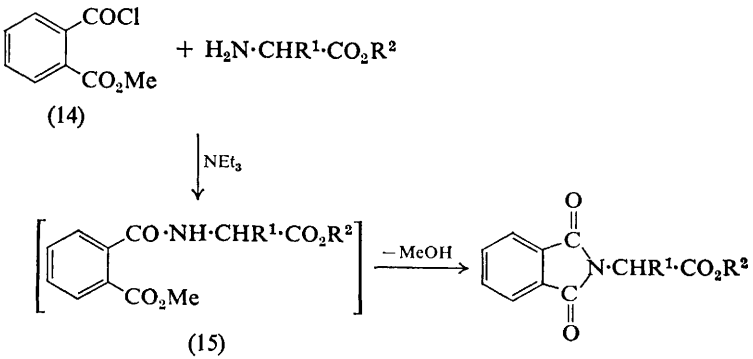
(12)



(13)

were prepared in good yield. The utility of the *N*-chloroacetyl group in peptide synthesis has been extended by showing that the dipeptide derivative obtained from *N*-chloroacetyl-L-valine and L-valine t-butyl ester by the dicyclohexylcarbodi-imide–*N*-hydroxysuccinimide method showed no racemization in the Weygand test after removal of the *N*-chloroacetyl group with (12) and conversion into the trifluoroacetyldipeptide methyl ester. Furthermore, the *N*-chloroacetyl group has been used to protect the ω-amino-function of ornithine in a synthesis which utilized *N*-trifluoroacetyl for α-protection. The former group can be removed selectively in the presence of the latter by using reagent (12), whereas the latter group can be removed selectively by treatment with sodium borohydride in ethanol.

An alternative to the use of *N*-carbethoxyphthalimide for the introduction of *N*-phthaloyl groups has been devised. *o*-Methoxycarbonylbenzoyl chloride (14) reacts readily with amino-esters in an anhydrous medium giving excellent yields of optically pure phthaloylated derivatives, presumably *via* an intermediate *o*-methoxycarbonylbenzoyl amino-acid ester (15) (Scheme 7).[25]



(14)

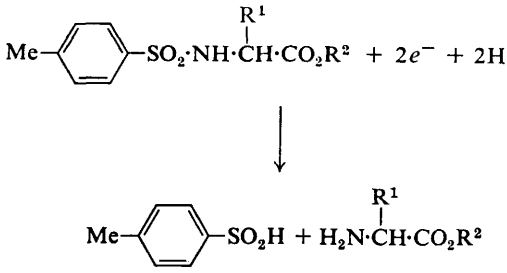
(15)

Scheme 7

Cathodic reduction has been applied for the cleavage of *N*-tosyl groups [26] from amino-acids and peptides. The *N*-tosyl substrate is electrolysed in methanol or aqueous methanol containing two equivalents of sodium hydroxide in a cathode compartment with a lead (or porous zinc) electrode.

[25] D. N. Reinhoudt, D. A. Hoogwater, and H. C. Beyerman, ref. 7, p. 7.
[26] K. Okamura, T. Iwasaki, M. Matsuoka, and K. Matsumoto, *Chem. and Ind.*, 1971, 929.

The anode compartment of the cell, which is separated from the cathode compartment by a membrane, contains hydrochloric acid with a carbon electrode. A number of *N*-tosyl amino-acids and peptides gave very good yields of cleaved material which were optically pure, as judged by optical rotation. In addition it was shown that the selective cleavage of *N*-tosyl groups in the presence of *N*-benzyloxycarbonyl, t-amyloxycarbonyl, t-butoxycarbonyl, and *S*-benzyl groups is possible by this method. Since toluene-*p*-sulphinic acid was isolated after electrolysis of an *N*-tosyl derivate, it was assumed that the cathodic reduction proceeds *via* a two-electron change resulting in S—N bond cleavage (Scheme 8).[26]



Scheme 8

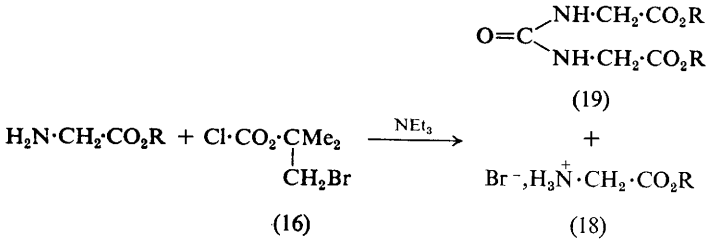
The application of the *N*-trityl protecting group for industrial peptide synthesis has recently been examined.[27] The direct preparation of *N*-trityl amino-acids is claimed to proceed in good yield when amino-acid derivatives are treated with two equivalents of trityl chloride in an aprotic solvent in the presence of triethylamine.[27]

New Methods of Amino-group Protection. t-Butoxycarbonyl groups are outstandingly useful for amino-group protection. In recent years the trend has been to produce modified t-butylurethanes, in which one of the methyl substituents has been replaced by a group capable of stabilizing the incipient carbonium ion still further, thus resulting in increased acid-lability of the protecting group. Recent investigations [28–30] have been concerned with the effect of replacing one of the methyl groups by substituents which are highly electronegative. Such groups will destabilize the incipient carbonium ion but by the same token render the acylating agents (*e.g.* the chloroformates) more stable.

Carpino and his colleagues studied [28] several α-halogeno-t-butoxycarbonyl derivatives with a view to their eventual conversion into t-butoxycarbonyl

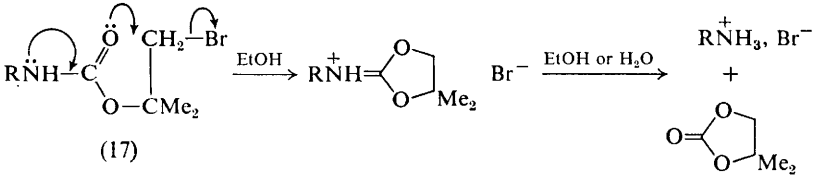
[27] T. Tamaki and S. Kudo, *Yuki Gosei Kagaku Kyokai Shi*, 1971, **29**, 599 (*Chem. Abs.*, 1971, **75**, 98 808r).
[28] L. A. Carpino, K. N. Parameswaran, R. K. Kirkley, J. W. Spiewak, and E. S. Schmitz, *J. Org. Chem.*, 1970, **35**, 3291.
[29] G. L. Southard, B. R. Zaborowsky, and J. M. Pettee, *J. Amer. Chem. Soc.*, 1971, **93**, 3302.
[30] E. Wünsch and R. Spangenberg, *Chem. Ber.*, 1971, **104**, 2427.

derivatives. The stable α-bromo-t-butyl chloroformate (16) was prepared, and although simple amines reacted with it to give high yields of the corresponding urethanes (17), amino-acids or amino-esters failed to react normally. Instead of the expected urethane the amine hydrobromide (18) and/or the urea derivative (19) were obtained (Scheme 9). Subsequent



Scheme 9

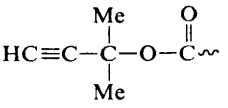
experiments showed that the amine hydrobromide probably arose by 'self-cleavage' of the expected urethane (Scheme 10). Hydrogenation of the simple model α-bromo-t-butylurethanes over a palladium–carbon catalyst in methanol solution in the presence of ammonium acetate gave only moderate yields of the corresponding t-butoxycarbonyl derivatives. Thus

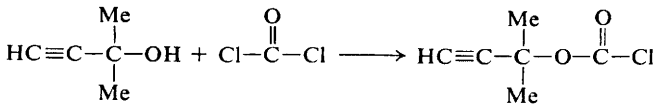


Scheme 10

the 'self-cleavage' reaction and the moderate yields of simple t-butoxycarbonyl derivatives obtained by reduction make this interesting method impractical for the introduction of the t-butoxycarbonyl group.

The use of the 1,1-dimethyl-2-propynyloxycarbonyl group (20) for amino-group protection has been investigated in a preliminary fashion.[28, 29] 1,1-Dimethyl-2-propynyloxycarbonyl chloride (21) was prepared from 2-methyl-3-butyn-2-ol (22) and phosgene (Scheme 11) in low yield, but although it is distillable at water-pump pressure it darkens on storage.[28]

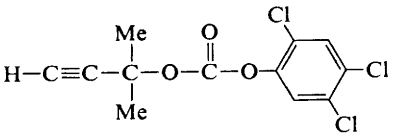




(22) (21)

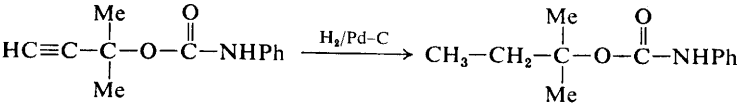
Scheme 11

Preparation of 1,1-dimethyl-2-propynyl 2,4,5-trichlorophenyl carbonate (23) has been mentioned [29] but without any experimental description. The carbanilate (24) was prepared [28] from the chloroformate (21) and aniline: catalytic reduction gave the corresponding t-amyloxycarbonyl derivative [31] (25) (Scheme 12). In a preliminary report [29] it has been shown, however,



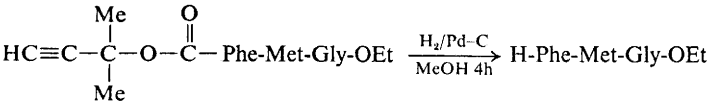
(23)

that the catalytic hydrogenolysis of 1,1-dimethyl-2-propynyloxycarbonyl-amino-protected peptides [*e.g.* (26)] which contain sulphur proceeds with cleavage of the *N*-protecting group (Scheme 13). Presumably the partially poisoned catalyst only allows reduction to the tertiary allylic derivative,



(24) (25)

Scheme 12



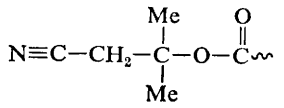
(26)

Scheme 13

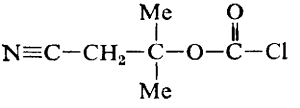
which becomes susceptible to allylic hydrogenolysis. The full paper of this study is awaited with interest so that the potential of this new protecting group can be evaluated.

The *N*-(cyano-t-butoxycarbonyl) group (27) has been introduced for the protection of amino-functions. Amino-acids protected in this way are

[31] N. Schachat and J. J. Bagnell, jun., *J. Org. Chem.*, 1963, **28**, 991.



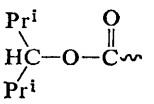
(27)



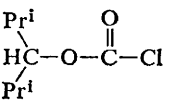
(28)

obtained by the action of crude cyano-t-butoxycarbonyl chloride (28) under Schotten–Baumann conditions.[30] It was demonstrated that a cyano-t-butoxycarbonyl amino-acid could be coupled with a suitable amino-component using the dicyclohexylcarbodi-imide–*N*-hydroxysuccinimide procedure. The new protecting group can be cleaved with weakly basic reagents (aqueous potassium carbonate or triethylamine at pH 10) *via* β-elimination, but is relatively stable in the presence of anhydrous trifluoroacetic acid. Glycyl-L-tryptophan (free peptide) prepared using this method of protection possessed the same optical rotation as a specimen prepared by a standard route.[30]

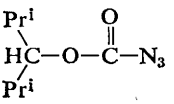
Sakakibara and his co-workers have carried out a preliminary study of the efficacy of the di-isopropylmethoxycarbonyl group (29) for protecting the ε-amino-function of lysine.[32] Both di-isopropylmethyl chloroformate



(29)



(30)



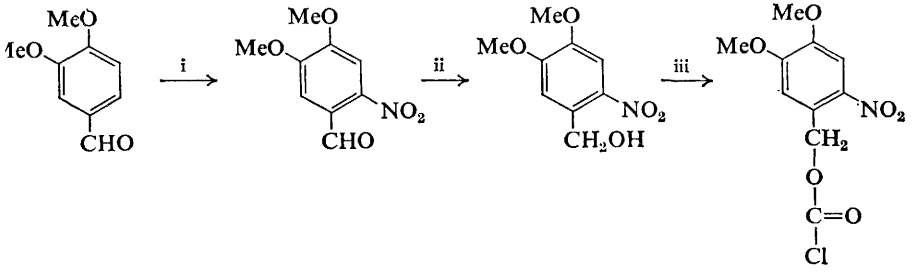
(31)

(30) and azidoformate (31) were prepared and utilized to form ε-di-isopropylmethoxycarbonyl-lysine *via* its copper complex, and the resulting protected amino-acid was used in peptide synthesis. The protecting group was shown to be considerably more stable to hydrogen chloride in acetic acid than is an ε-benzyloxycarbonyl group, but it could be removed completely with anhydrous hydrogen fluoride in the presence of anisole at 20 °C for 1 h. Its use is therefore recommended in cases where there would otherwise be a risk of partial cleavage of an *N*-ε-benzyloxycarbonyl group.[32]

A further preliminary report has appeared on the use of the photochemically labile 6-nitroveratryloxycarbonyl group [33] for amino-group protection. This group can be introduced with 6-nitroveratryl chloroformate (32), which is prepared according to Scheme 14. Irradiation of 6-nitroveratryloxycarbonyl derivatives at 350 nm brings about an intramolecular redox reaction, which results in the liberation of carbon dioxide, 6-nitrosoveratraldehyde, and the free amino-group (Scheme 15). The

[32] S. Sakakibara, T. Fukuda, Y. Kishida, and I. Honda, *Bull. Chem. Soc. Japan*, 1970, **43**, 3322.

[33] A. Patchornik, B. Amit, and R. B. Woodward, ref. 7, p. 12.

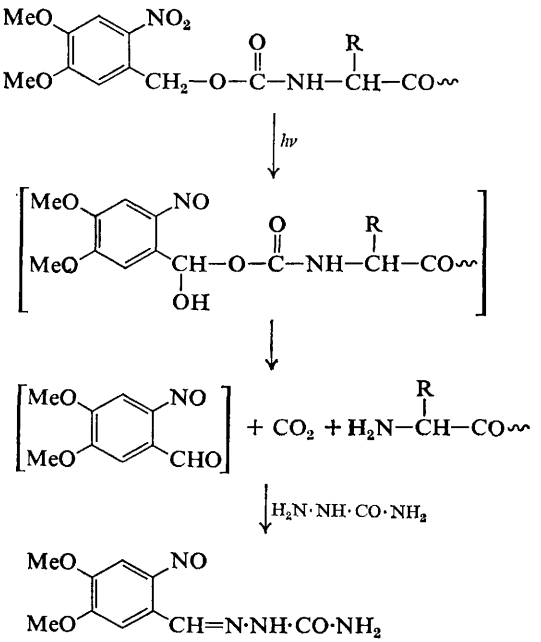


(32)

Conditions: i, HNO_3; ii, $NaBH_4$; iii, $COCl_2$

Scheme 14

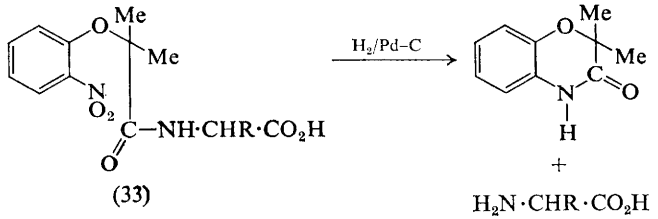
yields of free amino-acids and peptides can be made practically quantitative by carrying out the photoreaction in the presence of strong acids or aldehyde reagents such as semicarbazide hydrochloride. The 6-nitroveratryloxycarbonyl group can be removed readily from photolabile derivatives of tryptophan providing that wavelengths below 320 nm are filtered out. This method of removal is stated to be compatible with protection by benzyloxycarbonyl, t-butoxycarbonyl, trifluoroacetyl, tosyl, and benzyl groups,



Scheme 15

while *o*-nitrophenylsulphenyl and triphenylmethyl groups are sensitive to irradiation. No racemization of optically active derivatives was observed.

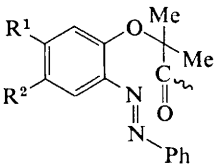
A further preliminary study of 2-methyl-2-*o*-nitrophenoxypropionyl-protected amino-acids (see vol. 2 of these Reports, p. 150) revealed that the group is easily removed by hydrogenolysis (Scheme 16).[34] Peptide-bond



(33)

Scheme 16

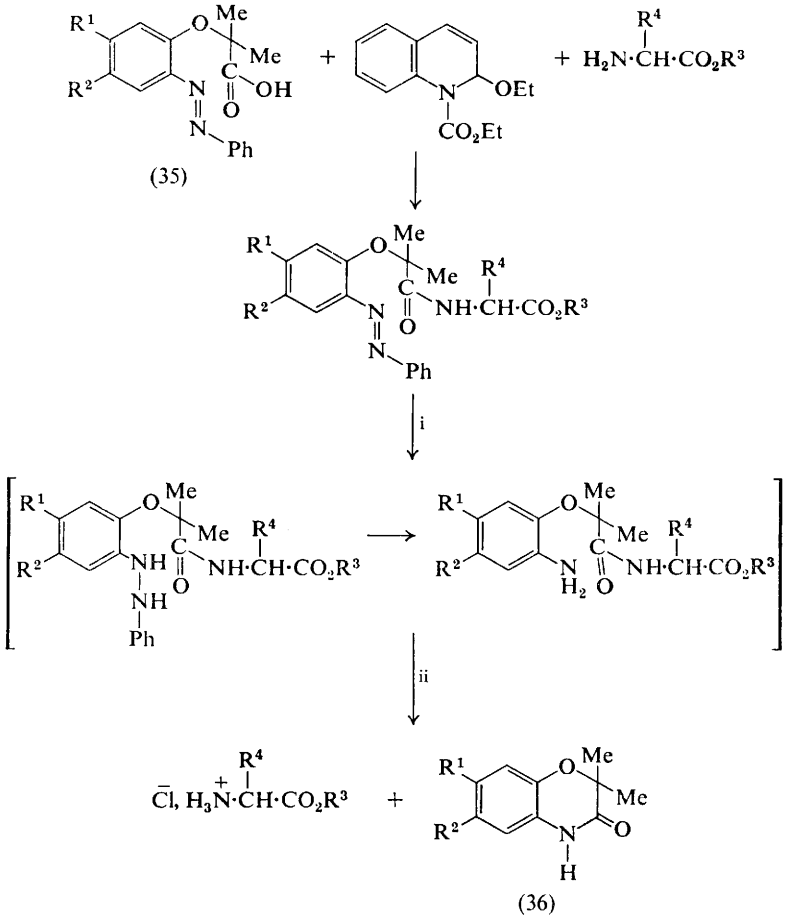
formation with (33) using the acid chloride method proceeds with complete racemization, whereas coupling mediated by mixed anhydrides or dicyclohexylcarbodi-imide is claimed to lead to optically pure products. A related pair of protecting groups which are removable by neighbouring-group participation are α-methyl-α-(4,5-dimethyl-2-phenazophenoxy)propionyl (34; $R^1 = R^2 = Me$) and α-methyl-α-(4-methyl-2-phenazophenoxy)-propionyl (34; $R^1 = H$, $R^2 = Me$). These groups are introduced by



(34)

coupling the corresponding acids (35) with amino-esters using 'EEDQ'—*i.e.* 2-ethoxy-*N*-ethoxycarbonyl-1,2-dihydroquinoline (Scheme 17). The group is removed by converting the phenylazo-moiety into a nucleophilic reduction product using potassium borohydride and palladium on carbon. Acidification of the reaction mixture causes cleavage of the amide bond, with concomitant cyclization of the protecting moiety to a 3,4-dihydro-2,2-dimethyl-2*H*-1,4-benzoxazin-3-one (36). The optical stability of the protected derivatives has not been investigated, or their usefulness in peptide synthesis evaluated.[35] Although this kind of protecting group is chemically ingenious, the multiple possibilities for side-reactions and the predictable optical lability of the acids (33) during coupling combine to make any practical application in peptide chemistry most unlikely.

[34] F. Cuiban, *Tetrahedron Letters*, 1971, 2471.
[35] C. A. Panetta and Altaf-Ur-Rahman, *J. Org. Chem.*, 1971, **36**, 2250.



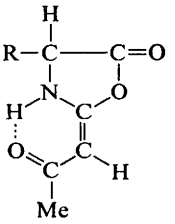
Conditions: i, KBH_4–Pd/C; ii, HCl

Scheme 17

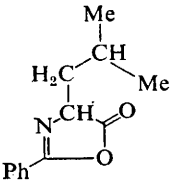
The observation that peptide synthesis with *N*-acetoacetylamino-acid derivatives proceeds, in certain circumstances (*e.g.* using the carbodi-imide method), to give optically pure products has been rationalized by assuming that the carboxy-component is converted into a 2-acetonylidenoxazolidin-5-one (37),[36, 37] which couples more rapidly with nucleophiles than does the corresponding oxazolinone. This tentative conclusion was reached on the basis of the spectroscopic properties of the compounds resulting from the treatment of *N*-acetoacetylamino-acids with dicyclohexylcarbodi-imide.

[36] C. Di Bello, F. Filira, and F. D'Angeli, ref. 7, p. 35.
[37] C. Di Bello, F. Filira, and F. D'Angeli, *J. Org. Chem.*, 1971, **36**, 1818.

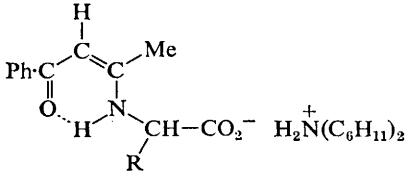
Derivatives such as (37) undergo reaction with nucleophiles to yield optically pure products, and comparative qualitative rate measurements showed that 2-acetonyliden-4-isobutyloxazolidin-5-one (37; R = Bu^{l}) couples with amines more rapidly than does 4-isobutyl-2-phenyl-2-oxazolin-5-one (38).



(37)



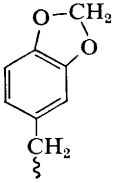
(38)



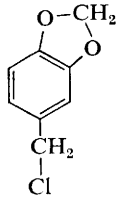
(39)

An improved method for the synthesis of amino-acids protected with β-dicarbonyl compounds has appeared.[38, 39] The amino-acid, in aqueous alcohol containing dicyclohexylamine, is treated with benzoylacetone to yield the crystalline *N*-(2-benzoyl-1-methylvinyl)amino-acid dicyclohexylammonium salt (39). The preparation of a number of active esters of β-dicarbonyl *N*-protected amino-acids has been described.[40]

Protection of Carboxy-groups. The utility of the piperonyl group (40) for carboxy-group protection has been investigated.[41] A limited number of



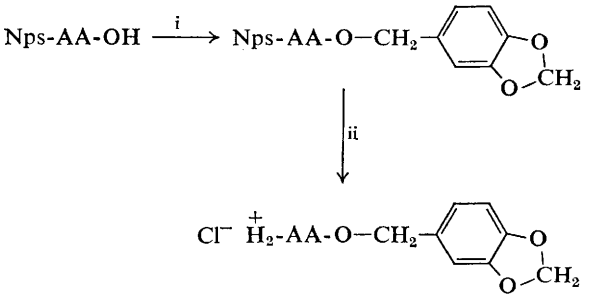
(40)



(41)

[38] G. L. Southard, G. S. Brooke, and J. M. Pettee, *Tetrahedron*, 1971, **27**, 1359.
[39] A. Balog, D. Breazu, E. Vargha, F. Gonczy, and L. Beu, *Rev. Roumaine Chim.*, 1970, **15**, 1375 (*Chem. Abs.*, 1971, **75**, 6264j).
[40] A. Balog, E. Vargha, D. Breazu, L. Beu, and G. Gonczy, *Rev. Roumaine Chim.*, 1970, **15**, 1391 (*Chem. Abs.*, 1971, **75**, 6286t).
[41] F. H. C. Stewart, *Austral. J. Chem.*, 1971, **24**, 2193.

piperonyl esters were prepared by treating the *N*-protected amino-acid with piperonyl chloride (41) in the presence of triethylamine and subsequently removing the *N*-protecting group (Scheme 18). A number of



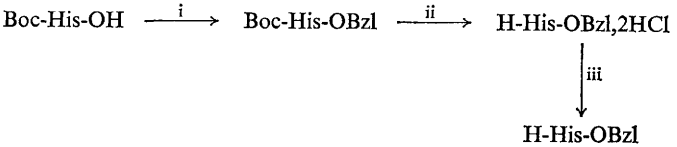
Conditions: i, (41)–NEt_3; ii, HCl in MeOH or dioxan

Scheme 18

benzyloxycarbonyldipeptide piperonyl esters were prepared by the *p*-nitrophenyl ester method, and their conversion to known free dipeptides enabled the absence of racemization during these synthetic steps to be inferred. The new ester group is removable in the presence of benzyloxycarbonyl by brief treatment with trifluoroacetic acid, or removable together with benzyloxycarbonyl by treatment with 2M hydrogen bromide in acetic acid.

The preparation of further amino-acid *p*-methoxybenzyl ester hydrochlorides, and their use in peptide synthesis, has been reported.[42, 43] The derivatives of glycine, L-glutamine, L-asparagine, and *S*-benzyl-L-cysteine were prepared *via* the corresponding *o*-nitrophenylsulphenyl derivatives,[42] whereas the esters of the imino-acids sarcosine, L-proline, and L-4-hydroxyproline were synthesized using the corresponding *N*-nitroso-derivatives.[43] However, in the sarcosine series some cleavage of the ester moiety was observed when the nitroso-group was removed by acidolysis.

An improved synthesis of L-histidine benzyl ester from t-butoxycarbonylhistidine in 66% yield (Scheme 19) has been reported.[44] Another diazo-

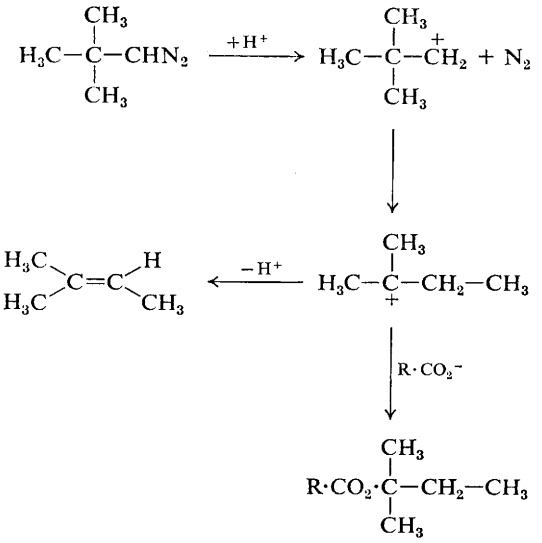


Conditions: i, $Ph{\cdot}CHN_2$; ii, 4M-HCl–dioxan; iii, K_2CO_3

Scheme 19

[42] J. A. Maclaren, *Austral. J. Chem.*, 1971, **24**, 1695.
[43] F. H. C. Stewart, *Austral. J. Chem.*, 1971, **24**, 1749.
[44] A. M. Felix and D. P. Winter, *Org. Prep. Proced.*, 1970, **2**, 255.

alkane, namely 1-diazo-2,2-dimethylpropane (42) has been used as a reagent to prepare t-amyl esters of peptide acids.[45] As shown in Scheme 20, the diazo-alkane, on protonation, gives a neopentyl carbonium ion, which rearranges to the t-amyl carbonium ion, thus giving the ester mentioned.



Scheme 20

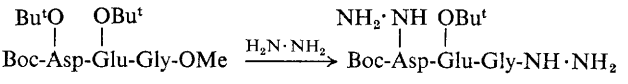
The resulting esters are stable to hydrazinolysis but are cleaved by short exposure to hydrogen bromide in acetic acid. *C*-Terminal arginine peptides protected as t-amyl esters are susceptible to tryptic hydrolysis.

In the picolyl ester method of peptide synthesis the crude product was originally separated from co-products and by-products, either by extraction into aqueous citric acid, or by adsorption on sulphoethyl-Sephadex (see previous volumes of these Reports). Full details of a synthesis of bradykinin (in which the citric acid method was used: see ref. 297) and of [5-valine]-angiotensin II (exemplifying the sulphoethyl-Sephadex technique: see ref. 423) have now been published. The use of the macroreticular sulphonic acid resin Amberlyst-15 is now recommended for the isolation stage,[46, 47] since it can be used in anhydrous organic solvents. In a typical procedure the coupling solution is exhaustively cycled through the Amberlyst column, and the product is subsequently eluted with DMF containing pyridine. In order to overcome the reluctance of some protected peptide picolyl esters, with lipophilic side-chains, to extract into an acidic aqueous phase,

[45] C. F. Hayward and R. E. Offord, ref. 7, p. 116.
[46] J. Burton and G. T. Young, *Israel J. Chem.*, 1971, **9**, 201.
[47] J. Burton, G. A. Fletcher, and G. T. Young, *Chem. Comm.*, 1971, 1057.

the additional use of the picolyl ester to protect carboxy-side-chains has been recommended.

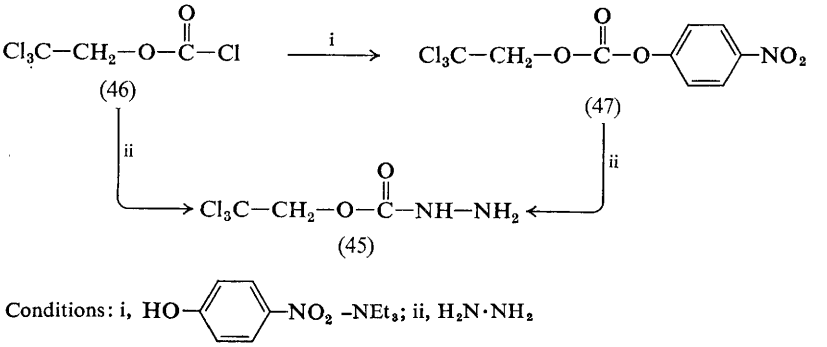
The use of *N*-protected hydrazine derivatives to protect carboxy-groups enables the ready preparation of peptide hydrazides which might otherwise be difficult to obtain by direct hydrazinolysis of the corresponding peptide alkyl esters. For instance, the hydrazinolysis of peptide (43) leads to a product (44) which contains two hydrazide groupings (Scheme 21). A new



(43) (44)

Scheme 21

protected hydrazine has been described, namely trichloroethoxycarbonyl-hydrazine (45),[48] which is prepared either directly from the corresponding trichloroethoxycarbonyl chloride (46) and hydrazine or indirectly *via* hydrazinolysis of trichloroethyl *p*-nitrophenyl carbonate (47) (see Scheme 22). Trichloroethoxycarbonyl hydrazides are prepared by coupling

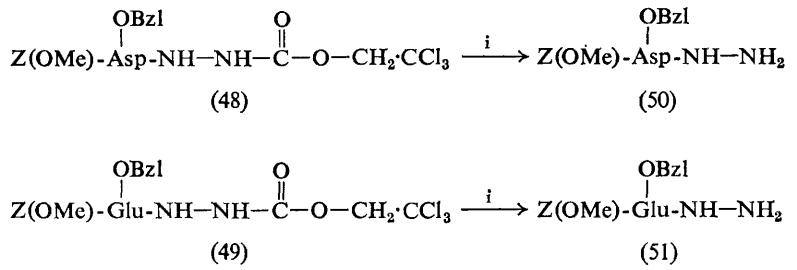


Scheme 22

acylamino-acids with (45) using dicyclohexylcarbodi-imide or the mixed-anhydride method. Conversion to the corresponding hydrazides is by means of zinc dust in acetic acid, which does not disturb other commonly used protecting groups [*e.g.* (48)—(50) and (49)—(51) in Scheme 23]. No optical rotations are quoted in the work just reported.

A paper has appeared dealing with the conversion of amino-acids into the corresponding ethyl esters using azeotropic distillation.[49]

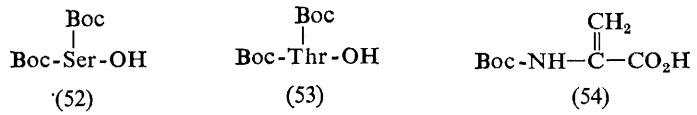
[48] H. Yajima and Y. Kiso, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 420.
[49] M. Dymocky, E. F. Mellon, and J. Naghski, *Analyt. Biochem.*, 1971, **41**, 487.



Conditions: i, Zn dust–$MeCO_2H$

Scheme 23

Protection of Hydroxy-groups. Side-chain hydroxy-functions are amenable to protection with the t-butoxycarbonyl group,[50] although the conditions for the removal of such a group have not been defined adequately. *N*-Mono-t-butoxycarbonyl derivatives of serine and threonine can be obtained in good yield by treating the amino-acid with the stoicheiometric amount of t-butoxycarbonyl fluoride at pH 8.0, whereas the use of excess acylating agent, particularly at higher pH, yields in addition some *NO*-bis-t-butoxycarbonyl derivatives [(52), (53)] which can be obtained in a pure state by



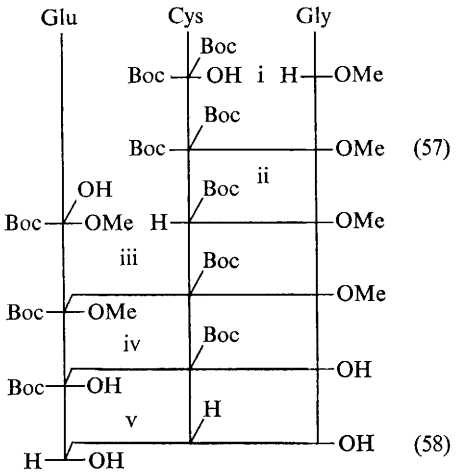
counter-current distribution. When this acylation is carried out with serine in triethylamine the dehydro-alanine derivative (54) is the major product.

The partially resolved mixture of enantiomers obtained from the treatment of *N*-formyl-*O*-benzyl-DL-serine with brucine can be further separated by selective solubilization in ether and t-butyl alcohol.[51] This provides an economic route to *O*-benzyl-L-serine.

Protection of Thiol Groups and Synthesis of Cystine Peptides. The t-butoxycarbonyl group can be used for masking the side-chain thiol function of cysteine. Treatment of cysteine with t-butoxycarbonyl fluoride at pH 8,[50] or with t-butoxycarbonyl chloride in the presence of triethylamine,[52] yields *NS*-bis-t-butoxycarbonyl-L-cysteine (55). Acylation of cysteine with

[50] E. Schnabel, J. Stoltefuss, H. A. Offe, and E. Klauke, *Annalen*, 1971, **743**, 57.
[51] A. S. Dutta and J. S. Morley, *Chem. Comm.*, 1971, 883.
[52] M. Muraki and T. Mizoguchi, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 1708.

t-butoxycarbonyl azide, on the other hand, is reported to yield mainly *NN'*-bis-t-butoxycarbonyl-L-cystine (56)[52] with a little (55). The *N*-protecting group of (55) could be removed by the brief action of 2.8M hydrogen chloride in ethyl acetate,[52] and the *S*-protecting group by treatment with iodine, thiocyanate, or heavy metal ions,[50] or by prolonged treatment with 2M sodium hydroxide solution.[50, 52] Investigations of the differential acid-lability of the *N*- and *S*-protecting groups of (57) showed that the selectivity of acidolytic *N*-deprotection was not complete, although it was sufficient for use in a synthesis of glutathione (Scheme 24).[52] Con-



Conditions: i, DCCI; ii, 2.8M-HCl–$MeCO_2Et$; iii, mixed anhydride; iv, 0.33M-NaOH; v, 1.4M-HCl–$MeCO_2H$

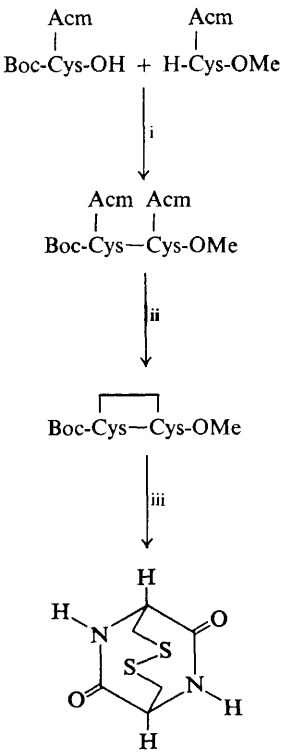
Scheme 24

version of (55) into (56)[52] by alkaline *S*-deprotection followed by iodine oxidation gave a product with an optical rotation some 14% lower than previously recorded figures, which seems ominous in view of the known optical lability of cysteine derivatives.

The *S*-acetamidomethyl protecting group is removable by means of mercury(II) salts at pH 4. It is also possible to form cystine peptides directly from *S*-acetamidomethylcysteine derivatives by a simple procedure involving oxidation with iodine.[53] This technique has been illustrated with the synthesis of *cyclo*-L-cystine (Scheme 25).

The t-butylthio-group has been introduced for masking thiol functions.[54] Some derivatives were prepared by treating cystine or *NN'*-bis-(t-butoxycarbonyl)cystine with t-butylmercaptan in basic solution in the presence

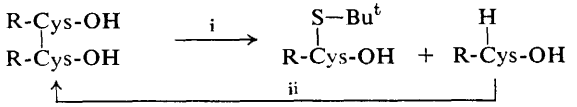
[53] B. Kamber, *Helv. Chim. Acta*, 1971, **54**, 927.
[54] E. Wünsch and R. Spangenberg, ref. 7, p. 30.



Conditions: i, DCCI; ii, I_2–MeOH; iii, $H{\cdot}CO_2H$ and cyclization by heating at 45 °C

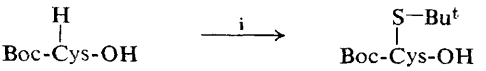
Scheme 25

of oxygen (Scheme 26), or by treating *N*-protected cysteine with t-butylsulphenylthiocyanate (59) in the presence of a tertiary base (Scheme 27). The *S*-t-butylthio protecting group can be removed by oxidative sulphitolysis followed by reduction; alternatively Cleland's reagent (9) can be used



Conditions: i, Bu^t-SH–base; ii, O_2.

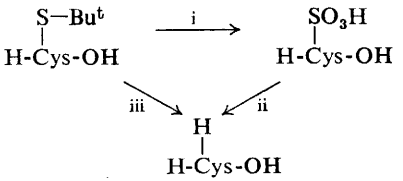
Scheme 26



Conditions: i, Bu^t-S-SCN (59)–*N*-methylmorpholine

Scheme 27

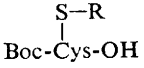
directly (Scheme 28). A number of derivatives (60) with various *S*-alkyl-type protecting groups have been evaluated in solid-phase peptide synthesis.[55] The *S*-protecting groups shown in (60) are removable under



Conditions: i, HSO_3^-; ii, H_2S or $NaBH_4$; iii, Cleland's reagent

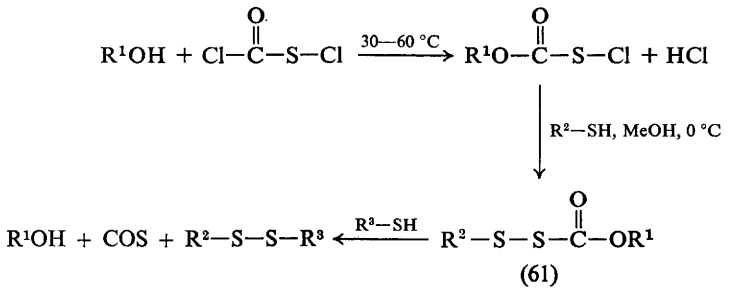
Scheme 28

the mild conditions of oxidative sulphitolysis or by mercaptoethanol treatment. Thiol–disulphide exchange reactions of *S*-alkylthiocysteine derivatives with reduced glutathione (58) have also been studied,[56] and a number of solid-phase syntheses of insulin A-chain analogues using these *S*-alkylthio protecting groups have been described.[57]



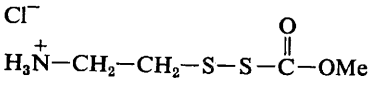
(60) R = Me, Et, Pr^n, Pr^i, Bu^t, or Bzl

A new method has been described for the formation of simple unsymmetrical disulphides, based on the reaction of sulphenyl thiolcarbonates (61) with thiols (Scheme 29).[58] The main advantages of this method are that the reagents (61) are stable and easily prepared and do not appear to undergo side-reactions with other functional groups—*e.g.*, the amino-group



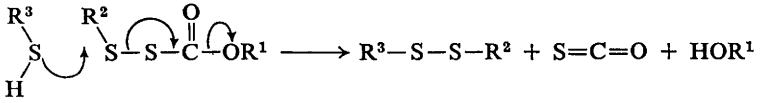
Scheme 29

[55] U. Weber and P. Hartter, *Z. physiol. Chem.*, 1970, **351**, 1384.
[56] U. Weber, P. Hartter, and L. Flohe, *Z. physiol. Chem.*, 1970, **351**, 1389.
[57] U. Weber, K. H. Herzog, H. Grossmann, P. Hartter, and G. Weitzel, *Z. physiol. Chem.*, 1971, **352**, 419.
[58] S. J. Brois, J. F. Pilot, and H. W. Barnum, *J. Amer. Chem. Soc.*, 1970, **92**, 7629.



(62)

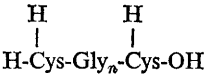
in (62). Nevertheless, they are highly reactive towards thiols, and highly selective in that reaction. In some cases formation of disulphide is catalysed by trace amounts of tertiary amine. The process can be written as shown in Scheme 30 or, alternatively, the synchronous collapse of a six-membered



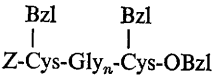
Scheme 30

cyclic transition state may be involved. The application of this procedure to peptides is awaited with interest, particularly on account of the inertness of sulphenyl thiolcarbonates towards amino-groups.

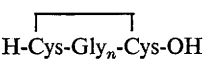
A series of L-cysteinyl-polyglycyl-L-cysteine peptides (63), generated by the action of sodium in liquid ammonia on the fully blocked peptides (64),



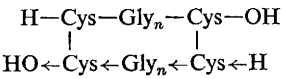
(63) n = 0, 1—9, 12, and 15



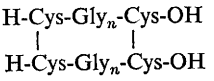
(64)



(65)



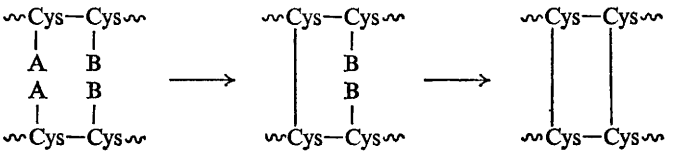
(66)



(67)

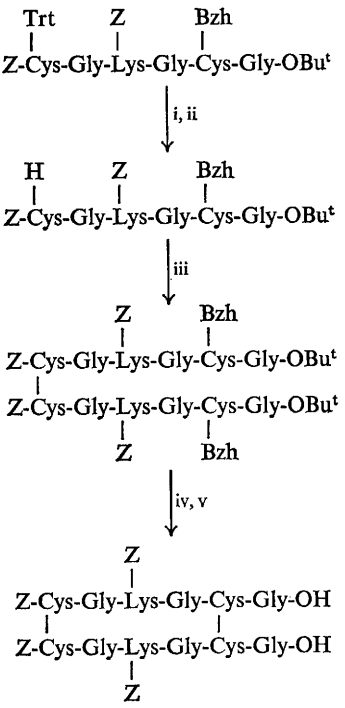
were autoxidized in dilute aqueous solution at pH 8.5 and provided a series of disulphides of varying complexity.[59] The lower members of the series gave rise to mixtures containing varying proportions of cyclic monomer (65) and antiparallel dimer (66). The hexapeptide (63; n = 4) and higher members, in contrast, gave predominantly the monomeric cyclic disulphides. The amounts of these monomers formed agree quite well with those predicted by a simple statistical theory,[59] indicating that the nature of the

[59] P. M. Hardy, B. Ridge, H. N. Rydon, and F. O. dos S. P. Serrão, *J. Chem. Soc.* (*C*), 1971, 1722.

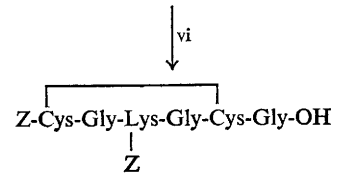


(68) (69) (70)

Scheme 31



(71)



(72)

Conditions: i, $AgNO_3$–C_5H_5N–DMF; ii, H_2S; iii, I_2; iv, $(SCN)_2$–$CF_3 \cdot CO_2H$–$CH_3 \cdot CO_2H$; v, BF_3–$CH_3 \cdot CO_2H$; vi, Δ, DMF

Scheme 32

oxidation products is controlled mainly by the probability of encounter of the two thiol groups when the two sulphur atoms are separated by three or more glycine residues. No evidence for the formation of parallel cyclic dimers (67) was obtained in this work. A general route to such compounds involves the synthesis of bis-cysteine peptides with their thiol groups differentially protected (Scheme 31). Selective removal of A from (68) enables the symmetrical disulphide (69) to be formed; this is followed by selective removal of B and subsequent ring closure to the cyclic bis-cystine peptide (70). A number of such peptides have been prepared using *S*-trityl or *S*-benzoyl groups for A-protection (Scheme 31) and *S*-benzhydryl for B-protection.[60] A representative example is shown in Scheme 32.[60, 61] The striking observations was made that cyclic bis-cystine peptides [*e.g.* (71)] revert to the monomeric form (72) on heating in DMF, the rearrangement being base-catalysed.[60]

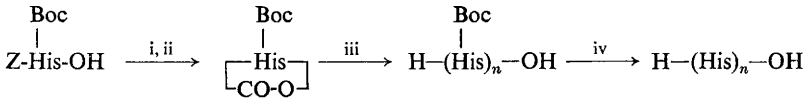
Protection of Histidine Side-chains. The t-butoxycarbonyl group has also been used for imidazole side-chain protection. The introduction of the group into histidine by acylation with t-butoxycarbonyl azide [62] or fluoride [50] to yield (73) must be carried out at carefully controlled pH: if the pH is too high concomitant hydrolysis of the newly formed acyl-imidazole occurs.[63] Earlier studies of the use of the N^{im}-t-butoxycarbonyl

(73) (74)

group indicated that although the reduction in the basicity of the imidazole ring was satisfactory, some derivatives such as (74) lacked stability,[64] in addition to being labile to reagents such as hydrazine and methanolic ammonia.[50] A re-evaluation of the situation showed that although hydrogen bromide in trifluoroacetic or acetic acid, or hydrogen fluoride, cleaves (73) to give free histidine, selective removal of the N^{α}-protecting group is possible using 2M hydrogen chloride in dioxan.[62] The protecting group has been further tested in a synthesis of poly-L-histidine (Scheme 33).[62]

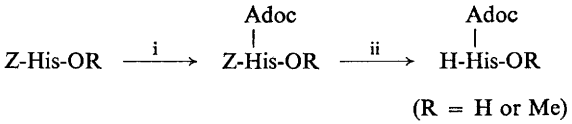
A convenient preparation of N^{im}-adamantyloxycarbonylhistidine has been reported (Scheme 34).[65] An application of the N^{im}-tosyl protecting group in a solid-phase synthesis of [5-isoleucine]-angiotensin II has been described.[66] The synthesis employed the standard Merrifield procedure in

[60] R. G. Hiskey, G. W. Davies, M. E. Safdy, T. Inui, R. A. Upham, and W. C. Jones, jun., *J. Org. Chem.*, 1970, **35**, 4148.
[61] R. G. Hiskey and J. B. Adams, jun., *J. Org. Chem.*, 1966, **31**, 2178.
[62] M. Fridkin and H. J. Goren, *Canad. J. Chem.*, 1971, **49**, 1578.
[63] G. Losse and U. Krychowski, *J. prakt. Chem.*, 1970, **312**, 1097.
[64] B. O. Handford, T. A. Hylton, K.-T. Wang, and B. Weinstein, *J. Org. Chem.*, 1968, **33**, 4251.
[65] M. A. Tilak, R. Russell, and M. L. Hendricks, *Org. Prep. Proced.*, 1971, **3**, 17.
[66] T. Fujii and S. Sakakibara, *Bull. Chem. Soc. Japan*, 1970, **43**, 3954.



Conditions: i, $SOCl_2$; ii, Ag_2O; iii, NEt_3; iv, $HBr–CH_3·CO_2H$

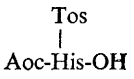
Scheme 33



Conditions: i, Adoc-Cl; ii, H_2–Pd/C

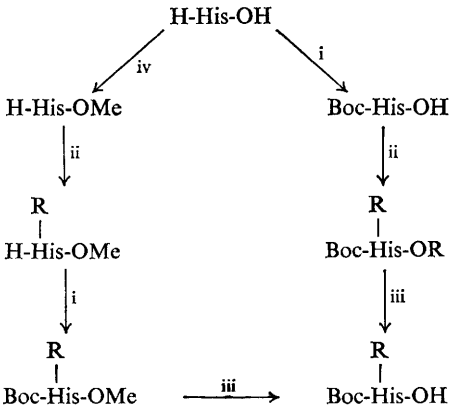
Scheme 34

which the histidine residue was introduced as (75). The naked peptide was liberated from the resin by the hydrogen fluoride procedure and was obtained pure, possessing full hypertensive activity, after Dowex-1 chromatography.



(75)

Some new histidine derivatives protected in the side-chain by aralkyl groups have been prepared (Scheme 35),[63] and a comparative study of the lability of these and other N^{im}-substituted compounds (N^{im}-2,4-dinitrophenyl, N^{im}-picryl, N^{im}-benzyloxycarbonyl, and N^{im}-piperidinocarbonyl)

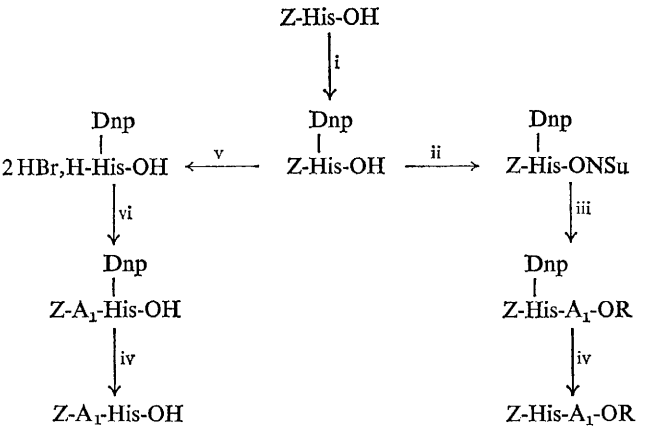


Conditions: i, Boc-N_3,pH 10; ii, Trt-Cl or Bzh-Br; iii, NaOH; iv, HCl–MeOH

Scheme 35

to various conditions has been made, with a view to using the compounds in solid-phase work.[67]

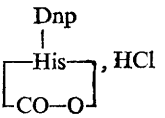
The thiolysis [68] of 2,4-dinitrophenylimidazoles (see vol. 2 of these Reports, p. 153) has been applied in classical preparations of some simple histidine-containing dipeptides (Scheme 36).[69] The rate of thiolysis of



Conditions: i, Dnp-F; ii, HONSu–DCCI; iii, H-A_1-OR; iv, HS·CH_2·CH_2·OH; v, HBr–CH_3·CO_2H; vi, Z-A_1-ONSu

Scheme 36

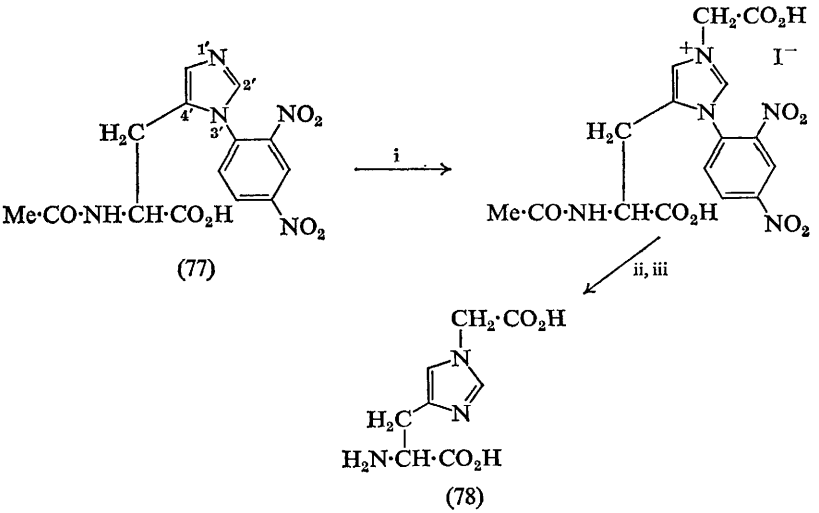
N^{im}-dinitrophenyl derivatives is sensitive to pH, to the amount of thiol used in excess, and to the solvent. Accordingly it is desirable to establish the optimum conditions for removal in each particular case. This is simplified by the development [69] of a spectrophotometric method for following the course of the reaction. In order to ensure a rapid release of histidine from its protecting moiety, and consequent diminution of side-reactions, it is recommended [69] to use a thousand-fold excess of mercaptoethanol or mercaptoacetic acid at pH 7—8. The use of Cleland's reagent [22] for the same purpose has been described. The *N*-carboxy-anhydride of N^{im}-dinitrophenylhistidine (76) has been prepared, and used in a synthesis of



(76)

[67] G. Losse and U. Krychowski, *Tetrahedron Letters*, 1971, 4121.
[68] S. Shaltiel, *Biochem. Biophys. Res. Comm.*, 1967, **29**, 178.
[69] S. Shaltiel and M. Fridkin, *Biochemistry*, 1970, **9**, 5122.

poly-L-histidine[70] (see Section 2E below). It has been shown that in N^{α}-acetyl-N^{im}-2,4-dinitrophenylhistidine (77) nitrogen 3′ of the imidazole ring bears the dinitrophenyl moiety.[71] This result was obtained by carboxymethylating compound (77); after thiolysis and acid hydrolysis the resulting carboxymethyl derivative was shown to be identical in its chromatographic mobility with $N^{1'}$-carboxymethylhistidine (78) (see Scheme 37). N^{im}-Di-



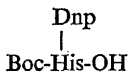
(77) (78)

Conditions: i, $I \cdot CH_2 \cdot CO_2^-$, pH 5; ii, $HS \cdot CH_2 \cdot CH_2 \cdot OH$; iii, hydrolysis

Scheme 37

nitrophenylhistidine does not decompose under the standard conditions of protein hydrolysis: a paper [71] on its quantitative determination on the amino-acid analyser has appeared.

Further reports on the use of (79) in solid-phase peptide synthesis comprise those concerning the tricosapeptide 124—146 of human haemoglobin β-chain,[72] thyrotropin-releasing hormone,[73] and luteinizing-hormone releasing factor.[74]



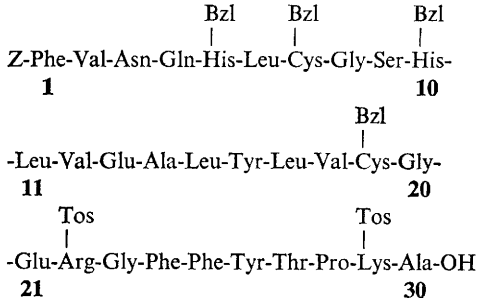
(79)

[70] M. Fridkin and S. Shaltiel, *Arch. Biochem. Biophys.*, 1971, **147**, 767.
[71] P. Henkart, *J. Biol. Chem.*, 1971, **246**, 2711.
[72] F. Chillemi, ref. 7, p. 84.
[73] P. Rivaille and G. Milhaud, *Helv. Chim. Acta*, 1971, **54**, 355.
[74] P. Rivaille, A. Robinson, M. Kamen, and G. Milhaud, *Helv. Chim. Acta*, 1971, **54**, 2772.

Final Deprotection. Methods for stripping a finished synthetic peptide of its protective groups after their work is done still leave much to be desired. Even syntheses which have been planned so that the mildest possible final deprotection is called for can run into difficulty: this was so in the case of calcitonin (see vol. 3 of these Reports, p. 253), where side-reactions were observed when the protecting groups (all of which were derived from t-butyl alcohol) were removed by brief treatment with cold concentrated hydrochloric acid. In more typical examples reagents of greater vigour are required for the last step. The use of anhydrous hydrogen fluoride for this purpose is the subject of a recent review by Sakakibara.[75] Experience with hydrogen fluoride has been very varied (see examples in Section 3), and it does not seem possible to generalize about its value: neither is it clear precisely what all the side-reactions induced by it are. Of older vintage is the use of sodium in liquid ammonia for final deprotection. Although this reagent has sometimes had quite disastrous results, in the past there was often no alternative for many syntheses, and the best that could be done was to minimize the side-reactions after extensive experimentation to locate the optimum conditions for each case. One of the main side-reactions is the cleavage of acyl–proline bonds (see ref. 76 for leading references to previous observations on this point). A recent study using model peptides suggests that Thr—Pro (insulin B27—B28) and Ser—Pro bonds are especially labile.[77] In a bovine (ovine, porcine) insulin B-chain synthesis reported during the year by Katsoyannis *et al.*,[76] cleavage of the Thr^{B27}—Pro^{B28} bond by sodium in ammonia was essentially quantitative, and in fact the reagent has been used [78] on a preparative scale to make de-(28—30)-B-chain.* It appears that this cleavage can be suppressed by carrying out the sodium–ammonia treatment in the presence of a large excess of sodamide,[76] and Katsoyannis and his collaborators were able to convert the human [79] and bovine (80) [76] protected B-chain derivatives to the corresponding deprotected B-chains (isolated as bis-*S*-sulphonates) in this way without fission of the threonylproline bond. This is clearly a finding of considerable potential usefulness, but the addition of sodamide may not prove a panacea for all the failings of sodium–ammonia since although no threonylproline cleavage was detected, both crude deprotected

[75] S. Sakakibara, in 'Chemistry and Biochemistry of Amino-acids, Peptides and Proteins', ed. B. Weinstein, vol. 1, p. 51.
[76] P. G. Katsoyannis, C. Zalut, A. Tometsko, M. Tilak, S. Johnson, and A. C. Trakatellis, *J. Amer. Chem. Soc.*, 1971, **93**, 5871.
[77] A. Marglin, *Internat. J. Protein Res.*, 1972, **4**, 47.
[78] P. G. Katsoyannis, C. Zalut, A. Harris, and R. J. Meyer, *Biochemistry*, 1971, **10**, 3884.
[79] P. G. Katsoyannis, J. Ginos, C. Zalut, M. Tilak, S. Johnson, and A. C. Trakatellis, *J. Amer. Chem. Soc.*, 1971, **93**, 5877.

* The des-(B28—B30)-bovine insulin (which also has a modified B-*C*-terminal group) derived from this degraded B chain was found to be fully active, and Katsoyannis *et al.* make the point that many earlier syntheses which used sodium–ammonia may have to be re-evaluated.



(80)

B-chain bis-*S*-sulphonates contained substantial amounts of at least two unidentified contaminants.

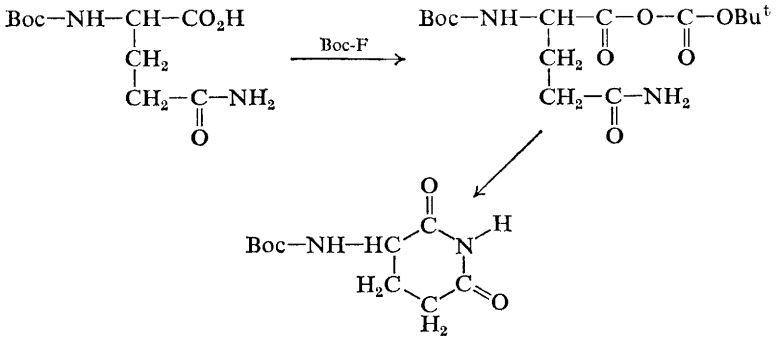
Katsoyannis *et al.* give full details [79] of the special glassware devised and employed by them for the convenient and controlled addition of sodium–ammonia solutions to their reaction mixtures.

Final deprotection of a protected (N^{α}-benzyloxycarbonyl, *O*-benzyl, ω-nitro-, and 4-picolyl ester groups) bradykinin by catalytic hydrogenolysis was accompanied by some reduction of the phenylalanine rings: this reduction could be made quantitative by using extended reaction times (see ref. 297).

Miscellaneous Matters relating to Protective Groups. A new simple quantitative method of reducing sulphoxides with sodium borohydride–cobalt chloride [80] has not as yet been applied to methionine sulphoxide derivatives. Methionine sulphoxide is quantitatively reduced by Cleland's reagent,[22] which has been recommended as an antioxidant to protect methionine peptides during chromatography. A neutral side-product isolated from the acylation of glutamine with t-butoxycarbonyl fluoride has been identified as *N*-t-butoxycarbonyl-α-amino-glutarimide (Scheme 38).[50] When peptides containing both nitroarginine and phenylalanine are hydrolysed with hydrochloric acid, chlorinated phenylalanines are formed.[81]

It has been known for many years that exposure to alkali prolongs and potentiates the melanotropic activity of pituitary peptides (ref. 82 has introductory references on this topic). This effect has been attributed to racemization, which results in delayed proteolytic inactivation. A recently published [82] detailed study provides further substantiation of this interpretation and concludes with the warning that 'The very significant levels of D-amino-acids formed after only 10 min of exposure of melanotropin to alkali should serve as a warning to peptide-synthetic chemists'. In fact the conditions used were 0.1M sodium hydroxide at 60 °C—hot alkali is

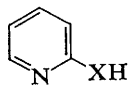
[80] D. W. Chasar, *J. Org. Chem.*, 1971, **36**, 613.
[81] P. Moritz and R. Wade, *Analyt. Biochem.*, 1971, **41**, 446.
[82] S. Lande and A. B. Lerner, *Biochim. Biophys. Acta*, 1971, **251**, 246.



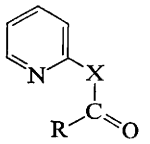
Scheme 38

very rarely used in synthetic operations—but five minutes of such treatment were sufficient for 50% prolongation of melanotropic activity and 10% racemization of some residues. The caveat is clear, although vigorous alkaline deprotection steps are in any case already anathema to many of us because of the side-reactions which often result with complex peptide derivatives.

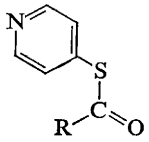
B. Formation of the Peptide Bond.—*Activated Esters.* In continuation of the study of amide bond formation assisted by intramolecular general base catalysis, which results in rapid racemization-free couplings, active esters



(81) X = O
(82) X = S



(83) X = O
(84) X = S

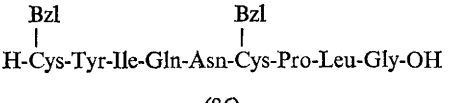


(85)

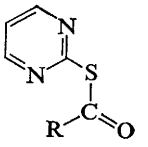
of 2-hydroxypyridine (81) and 2-mercaptopyridine (82) have been investigated.[83] Whereas the esters (83) were prepared using dicyclohexylcarbodi-imide in pyridine solution or in solvents containing two equivalents of pyridine [84] (in solvents such as acetonitrile or dichloromethane, *N*-acylureas were formed), the thiolesters (84) could be prepared in ethyl acetate.[83] A wide range of t-butoxycarbonyl- and benzyloxycarbonyl-amino-acid pyridyl esters [84] and thiolesters [83] were prepared (see Appendix B). Both types of ester are considerably more reactive towards nucleophiles than are the corresponding *p*-nitrophenyl esters or 4-pyridyl thiolesters (85).[83, 84] They react rapidly and exothermically even with hindered amino-esters to give very high yields of protected dipeptides. Both types

[83] K. Lloyd and G. T. Young, *J. Chem. Soc.* (*C*), 1971, 2890.
[84] A. S. Dutta and J. S. Morley, *J. Chem. Soc.* (*C*), 1971, 2896.

of ester react with alcohols, albeit rather slowly, and can be used for forming depsipeptides and *O*-acyl derivatives of serine, threonine, and tyrosine. In contrast to the usual solvent dependence for the aminolysis of esters such as those of *p*-nitrophenol, the rate of coupling for 2-pyridyl esters and 2-pyridyl thiolesters increases [83, 84] when the polarity of the medium is decreased. It appears that this type of solvent dependence may be characteristic of aminolyses which proceed by intramolecular general base catalysis. On account of this higher reactivity of the 2-pyridyl esters in non-polar solvents such as dichloromethane, their use in solid-phase peptide synthesis has been recommended, and this application has been evaluated by a synthesis of *SS'*-dibenzyl-oxytocinoic acid (86).[84] One 2-pyrimidinyl thiolester (87) was also prepared: it behaved similarly to the corresponding 2-pyridyl ester.



(86)



(87)

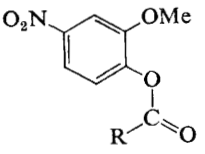
A note has appeared on the marked catalysis of active ester couplings by the anhydrous sodium salt of 2-hydroxypyridine.[85] In various model systems the rate enhancement of aminolysis due to the salt, relative to the hydroxypyridine, is of the order of several hundreds. It is possible that these reactions proceed *via* nucleophilic catalysis in which either a pyridyl ester (83) or an *N*-acylpyridone is formed as a reactive intermediate.

Monoacyl derivatives of catechol also undergo aminolysis by an intramolecularly assisted mechanism: the use of these in sequential polypeptide synthesis is covered elsewhere in this chapter (Section 2E).

The preparation of 4-nitroguaiacyl esters of *N*-protected amino-acids (88) and their use in peptide synthesis has been described recently.[86]

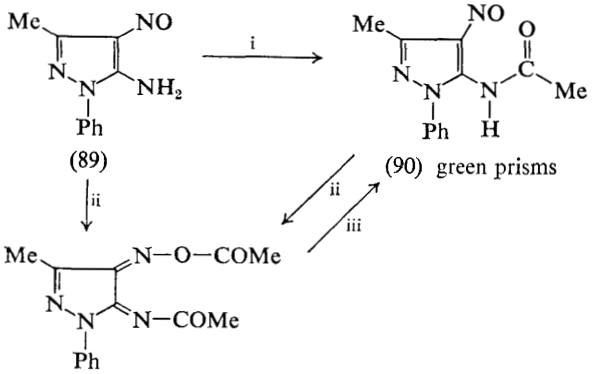
It was observed that 5-amino-3-methyl-4-nitroso-1-phenylpyrazole (89) reacts with acetic anhydride and yields, depending on the conditions used, either (90) or (91). The observation that ammonia converts (91) into (90)[87] (Scheme 39) formed the basis of an investigation of compounds related to (91) as new active esters. In the presence of dicyclohexylcarbodi-imide, (89) reacts with one or two moles of an *N*-protected amino-acid to yield (92) or

[85] N. Nakamizo, *Bull. Chem. Soc. Japan*, 1971, **44**, 2006.
[86] K. Bankowski and S. Drabarek, *Roczniki Chem.*, 1971, **45**, 1205.
[87] M. Guarneri and P. Giori, *Gazzetta*, 1969, **99**, 463.



(88)

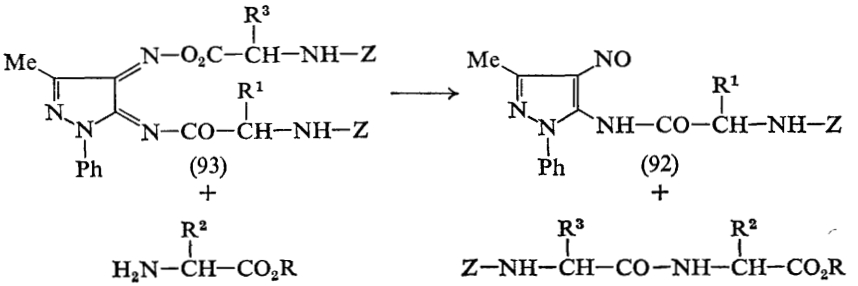
(93) respectively.[88] Compounds such as (93) are stable, crystalline acylating agents which react rapidly at room temperature with amino-esters (Scheme 40): the yields of peptides were high and no racemization was detected in sensitive tests. The co-product (92) is readily removed from the crude peptide by extraction with aqueous sodium carbonate, and can be readily recovered by acidification. The rapidity of the coupling reaction and its



(89) (90) green prisms (91) red crystals

Conditions: i, $(MeCO)_2O$, 45 °C; ii, $(MeCO)_2O$, 60 °C; iii, NH_3

Scheme 39

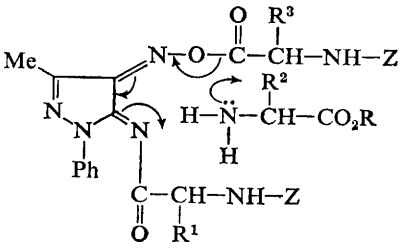


Scheme 40

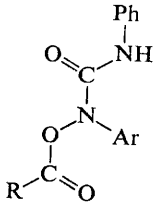
[88] M. Guarneri, P. Giori, and C. A. Benassi, *Tetrahedron Letters*, 1971, 665.

freedom from racemization suggest that the nitrogen atom of the 5-acylimino-group of (93) may be effecting intramolecular general base catalysis of the aminolysis reaction (94).

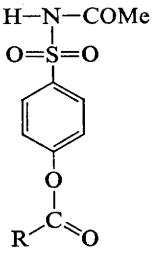
Other relevant papers have been concerned with applications of *O*-acyl-*N*-hydroxyureas (95) [89] and of *p*-(acetylsulphamoyl)phenyl esters (96) [90] in



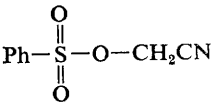
(94)



(95)



(96)



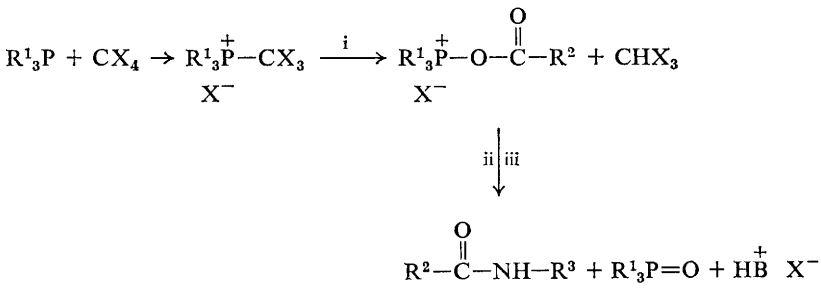
(97)

peptide synthesis and with the preparation of cyanomethyl esters by treatment of *N*-protected amino-acid salts with (97).[91]

Coupling Methods involving Acyloxyphosphonium Salts. Simple amides can be prepared in high yield [92] using the adduct of triphenylphosphine and carbon tetrachloride, which is transformed into a triphenylacyloxyphosphonium salt by treatment with the carboxyl component and then is treated

[89] D. Sarantakis, W. W. Light, A. R. Craig, and B. Weinstein, *Synthesis*, 1971, 328 (*Chem. Abs.*, 1971, **75**, 88 915j).
[90] G. Kupryszewski and F. Muzalewski, *Roczniki Chem.*, 1970, **44**, 2341 (*Chem. Abs.*, 1971, **75**, 49 555x).
[91] M. Leplawy and J. Zabrocki, *Z. Chem.*, 1971, **11**, 16 (*Chem. Abs.*, 1971, **75**, 49 546v).
[92] L. E. Barstow and V. J. Hruby, *J. Org. Chem.*, 1971, **36**, 1305.

with two equivalents of the amino-component (Scheme 41; :B = $R^3 \cdot NH_2$). Racemization can be minimized (to as little as 1—2% in the Young test) [9] by using phosphines or amides of phosphorous or phosphoric acids of higher basicity than triphenylphosphine, namely hexamethylphosphorotriamide (98) or (99) (Scheme 41; :B = *N*-methylmorpholine).



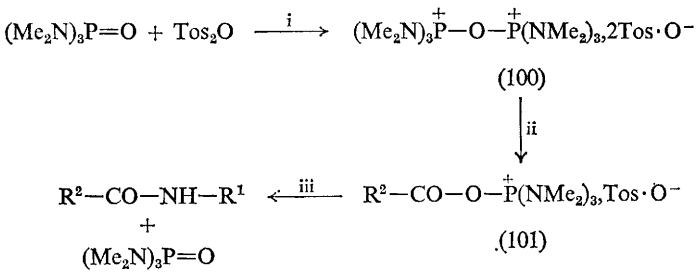
Conditions: i, $R^2 \cdot CO_2H$; ii, :B; iii, $R^3 \cdot NH_2$

Scheme 41

The above methods are reminiscent of the racemization-free condensation involving activated derivatives of hexamethylphosphorotriamide (see also vol. 2 of these Reports, p. 156). Dissolution of toluene-*p*-sulphonic

(98) (99)

anhydride in an excess of hexamethylphosphorotriamide leads to the formation of an activated derivative (100).[94] Addition to this solution of a carboxylate salt generates (101), which readily undergoes aminolysis to form a peptide bond (Scheme 42). In the absence of chloride ion this

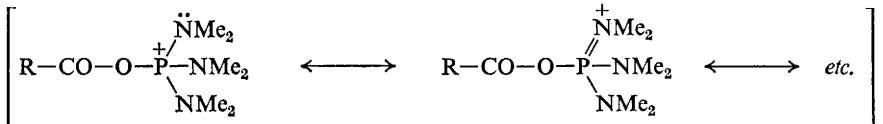


Conditions: i, Dry, room temperature, 15 min; ii, $R^2 \cdot CO_2^-$, 0 °C, 5—10 min; iii, $R^1 \cdot NH_2$

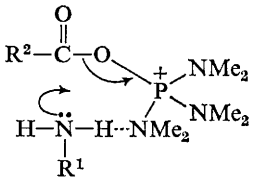
Scheme 42

[93] S. Yamada and Y. Takeuchi, *Tetrahedron Letters*, 1971, 3595.
[94] G. Gawne, G. W. Kenner, and R. C. Sheppard, ref. 7, p. 23.

method of peptide synthesis is largely racemization-free, which is ascribed to the low intrinsic activity of the resonance-stabilized intermediate (101). Aminolysis (but not oxazolinone formation) may be catalysed by anchimeric assistance (102) of the dimethylamino-groups or by initial attack at the

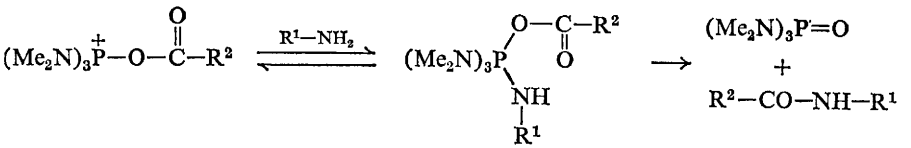


(101)



(102)

phosphorus atom (Scheme 43). Some tentative support for the latter possibility arises from the observation that coupling involving proline or sarcosine as amino-component leads to the production of some dimethylamide of the carboxyl component.

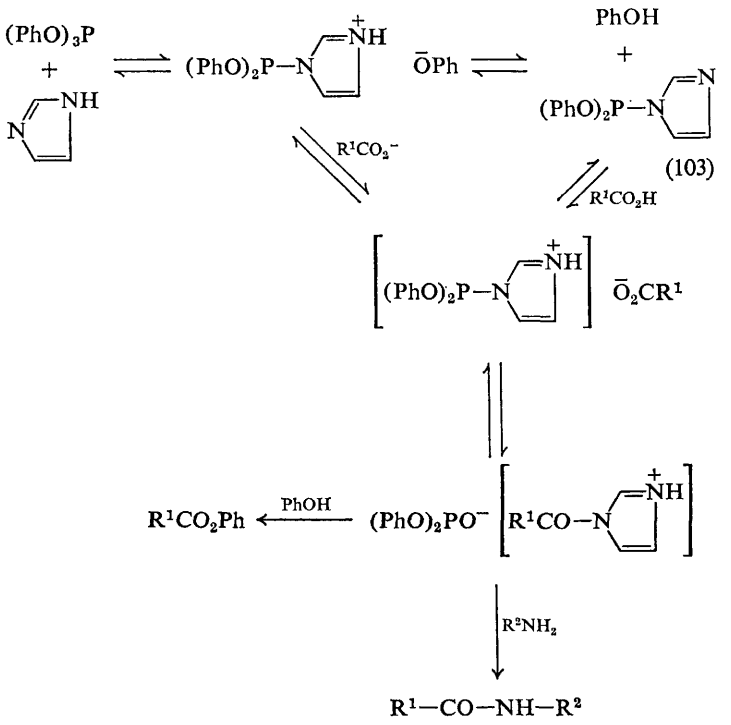


Scheme 43

The method of peptide synthesis using triphenyl phosphite and imidazole has been studied from a mechanistic point of view.[95] Conductivity studies are interpreted to indicate that triphenyl phosphite and imidazole interact with the formation of a compound containing a phosphorus–nitrogen bond. The postulated intermediate (103), prepared from diphenylphosphorochloridite and imidazole (Scheme 44), was shown to be reactive to aminolysis, although it was not characterized on account of its lack of stability. It is probable that the coupling proceeds *via* an acylimidazole, since in the absence of amino-component a phenyl ester results.

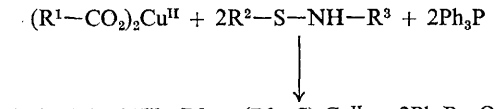
An acyloxyphosphonium salt is probably involved in the 'oxidation–reduction condensation' method of amide-bond synthesis (see vol. 1 of these Reports, p. 190). The observation that sulphenamides are attacked

[95] Yu. V. Mitin and O. V. Glinskaya, *J. Gen. Chem.* (*U.S.S.R.*), 1971, **41**, 1152.



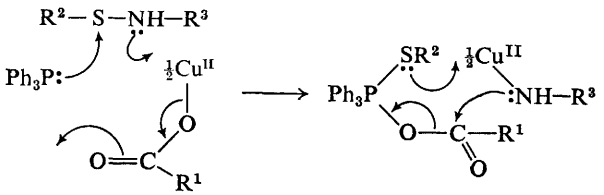
Scheme 44

by protic soft nucleophiles led to the finding that a carboxamide is produced from triphenylphosphine and a carboxylic acid in the presence of a sulphenamide. This in turn led to the development of a one-step procedure for peptide synthesis in which the copper(II) salt of an *N*-protected amino-acid (or an *N*-protected amino-acid plus a metal chloride and triethylamine) couples with an *o*-nitrophenylsulphenyl amino-acid ester in the presence of triphenylphosphine (oxygen acceptor) to give the *N*-protected dipeptide ester, triphenylphosphine oxide, and a copper(II) complex (Scheme 45).[96] The copper ions are present not only to scavenge the liberated mercaptan



Scheme 45

[96] M. Ueki, H. Maruyama, and T. Mukaiyama, *Bull. Chem. Soc. Japan*, 1971, **44**, 1108.



$$R^1{-}CO{-}NH{-}R^3 + Ph_3P{=}O + R^2S\tfrac{1}{2}Cu^{II}$$

Scheme 46

but also to activate the sulphur–nitrogen bond of the sulphenamide to nucleophilic attack (Scheme 46). The coupling proceeds in high yield but is a victim of racemization which, however, can be reduced in the presence of acidic additives (such as *N*-hydroxysuccinimide) which take over the role of activating the sulphur–nitrogen bond by protonation. The most favourable conditions elaborated (Scheme 47) gave benzoyl-L-leucylglycine ethyl ester with 74% optical purity.[96]

$$2R^1{-}CO_2H + 2R^2{-}S{-}NH{-}R^3 + (p\text{-}MeO\cdot C_6H_4)_2Hg^{II} + 2Ph_3P$$

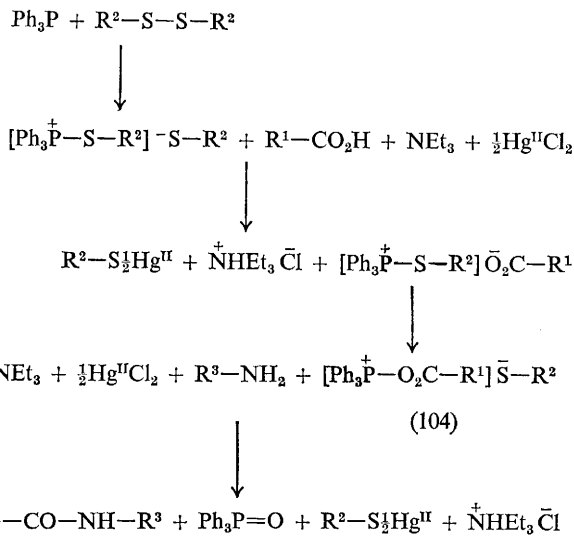
$$\downarrow$$

$$2R^1{-}CO{-}NH{-}R^3 + (R^2{-}S)_2Hg^{II} + 2MeO\cdot Ph + 2Ph_3P{=}O$$

Scheme 47

A more generally applicable variation of the method of 'oxidation–reduction condensation' involves the coupling of an *N*-protected amino-acid and an amino-acid ester mediated by a disulphide as oxidant and triphenylphosphine as reductant in the presence of mercury(II) chloride (Scheme 48).[97] The Young test revealed that this procedure leads to extensive racemization. In order to suppress this, the tertiary base was replaced by various organic mercury(II) compounds. Salts of acids with stable anions such as 2,4-dinitrophenol, which presumably react with the acyloxyphosphonium mercaptide (104) producing the metal mercaptide and another acyloxyphosphonium salt (105) as shown in Scheme 49, reduce racemization in proportion to the increasing stability of the anion. (Benzoyl-L-leucylglycine ethyl ester prepared by this variation had an optical purity of 46%.) If the additive is di-*p*-anisylmercury then the mercaptan produced after the coupling step is scavenged, obviating anion formation and thereby reducing the risk of racemization (Scheme 50). In the usual non-polar coupling

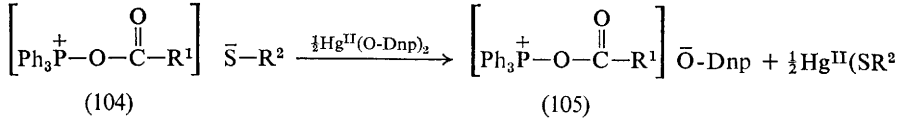
[97] R. Matsueda, H. Maruyama, M. Ueki, and T. Mukaiyama, *Bull. Chem. Soc. Japan*, 1971, **44**, 1373.



Scheme 48

solvents, the latter procedure gives benzoyl-L-leucylglycine ethyl ester with an optical purity of 90%.

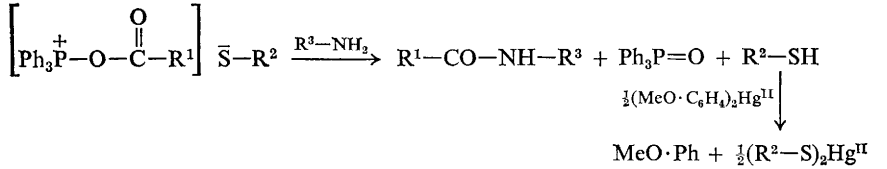
If this method of amide-bond formation were to be applied to the solid-phase method of synthesis, mercaptan scavengers would be required



(104) (105)

Scheme 49

which give rise to soluble mercaptides. Possible compounds are olefins such as 2,3-dihydropyran, sulphenate esters, and 2,2′-bipyridyl disulphide.[98] The optical purity of the Young peptide when prepared by the solution technique using the last compound was 92%.[97] This whole series of



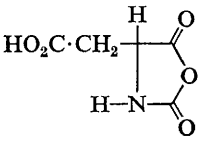
Scheme 50

[98] T. Mukaiyama, R. Matsueda, and H. Maruyama, *Bull. Chem. Soc. Japan*, 1970, **43**, 1271.

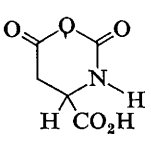
highly ingenious methods of 'oxidation–reduction' coupling clearly requires further study: it might well be that the use of a phosphorus-containing compound of higher basicity than triphenylphosphine might render racemization negligible.

A related method of peptide synthesis involves the interaction of an arylsulphenyl amino-acid, a carboxyl component, and triethyl phosphite in a non-chlorinated hydrocarbon solvent. The reaction proceeds under mild conditions in moderate yield but difficulties are experienced with asparagine and glutamine peptides. The method is claimed to be free of racemization by the Anderson and Young tests.[99]

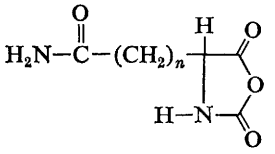
N-*Carboxy-anhydrides and* N-*Thiocarboxy-anhydrides.* The use of amino-acid *N*-carboxy-anhydrides and *N*-thiocarboxy-anhydrides in the controlled stepwise synthesis of peptides in aqueous media, without the isolation of each individual peptide, is outlined in Volume 1 of these Reports, p. 196. The syntheses of the novel *N*-carboxy-anhydrides[100] and *N*-thiocarboxy-anhydrides[101] used in this work have now been reported together with their evaluation in coupling reactions. The direct phosgenation of aspartic acid has been modified to give the corresponding Leuchs' anhydride (106) in the crystalline state, albeit in low yield, uncontaminated with compound (107). With ammonia (106) gives only isoasparagine, and in model coupling



(106)



(107)



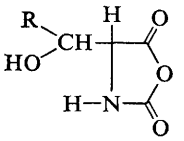
(108)

reactions it gives only α-linked peptides.[100] The *N*-carboxy-anhydrides of asparagine (108; $n = 1$) and glutamine (108; $n = 2$) are obtained in the crystalline state by the action of phosphorus tribromide on the corresponding benzyloxycarbonylamino-acids. Whereas the glutamine anhydride affords α-glutaminyl peptides without any evidence of rearrangement, the asparagine compound gives low yields of peptides, and the use of alternative reagents, such as t-butoxycarbonylasparagine succinimido-ester, is recommended.[100] An improved method is described for the preparation of the *N*-carboxy-anhydrides of serine (109; R = H) and threonine (109; R = Me) (the method has also been used for the Leuchs' anhydride of ε-t-butoxycarbonyl-L-lysine), which involves phosgenation of the silver salt of

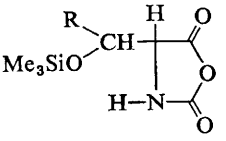
[99] Yu. V. Mitin and G. P. Vlasov, *J. Gen. Chem.* (*U.S.S.R.*), 1971, **41**, 419.

[100] R. Hirschmann, H. Schwam, R. G. Strachan, E. F. Schoenewaldt, H. Barkemeyer, S. M. Miller, J. B. Conn, V. Garsky, D. F. Veber, and R. G. Denkewalter, *J. Amer. Chem. Soc.*, 1971, **93**, 2746.

[101] R. S. Dewey, E. F. Schoenewaldt, H. Joshua, W. J. Paleveda, jun., H. Schwam, H. Barkemeyer, B. H. Arison, D. F. Veber, R. G. Strachan, J. Milkowski, R. G. Denkewalter, and R. Hirschmann, *J. Org. Chem.*, 1971, **36**, 49.

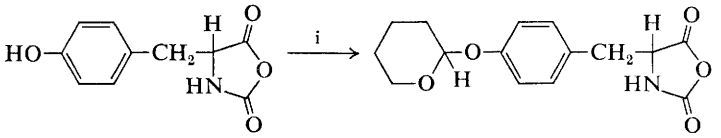


(109)



(110)

the corresponding amino-acid. Side-reactions can be avoided in the coupling of the serine and threonine derivatives if they are first converted into their *O*-trimethylsilyl derivatives (110; R = H or Me respectively), the *O*-protecting group being removed during the coupling reaction.[100] The direct preparation of (110; R = Me; DL) has been reported, using a new general method for forming *N*-carboxy-anhydrides, in which an *N*-silyloxycarbonylamino-acid trimethylsilyl ester is treated with phosgene.[102] In order to overcome side-reactions consequent upon the general insolubility of the Leuchs' anhydride of tyrosine (111) in aqueous coupling media, the phenolic group is protected as the tetrahydropyranyl ether (112) as shown in Scheme 51: the diastereoisomer with the lower negative optical rotation

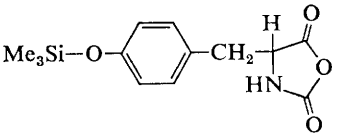


(111) (112)

Conditions: i, dihydropyran–Tos·Cl

Scheme 51

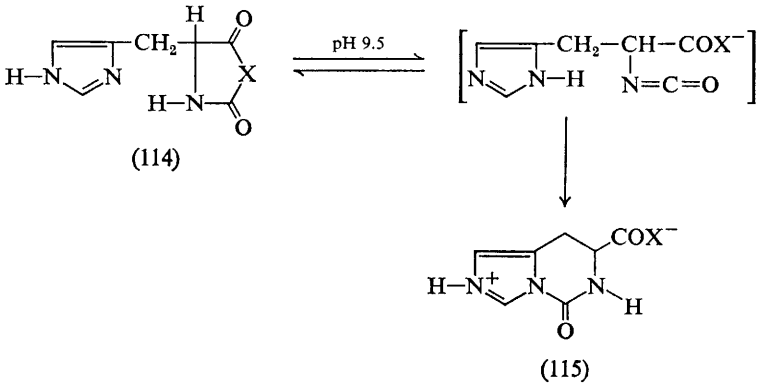
was used for peptide synthesis on account of its more favourable solubility. The protecting tetrahydropyranyl group is cleanly removed when the coupling solution is acidified to bring about decarboxylation of the carbamate.[100] The recently reported *O*-trimethylsilyl-L-tyrosine *N*-carboxy-anhydride (113) [102] might well become the reagent of choice for the introduction of tyrosyl residues by this method. The action of phosphorus tribromide on benzyloxycarbonylarginine yields a yellow tar containing *ca.* 50% of the desired *N*-carboxy-anhydride; despite this contamination the



(113)

[102] H. R. Kricheldorf, *Chem. Ber.*, 1971, **104**, 87.

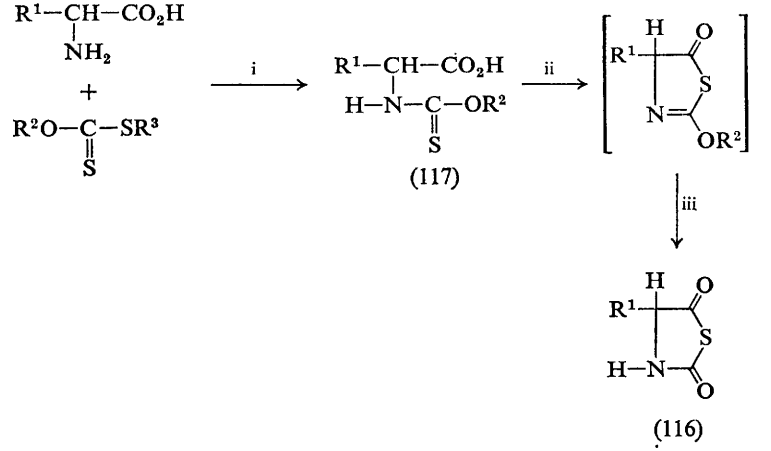
tar can be used successfully in peptide synthesis providing its anhydride content is known.[100] A similar method is used to prepare the crystalline hydrobromide salt of histidine *N*-carboxy-anhydride (114; X = O). This material could not be used successfully in repetitive peptide synthesis since the major product was an imidazo-tetrahydropyrimidinone (115; X = O), formed as shown in Scheme 52.[100]



(114) (115)

Scheme 52

The method of choice for the preparation of *N*-thiocarboxyamino-acid anhydrides (116) of high optical purity involves the cyclization of *N*-(alkoxythiocarbonyl) amino-acids (117) with phosphorus tribromide (Scheme 53).[101] Alternatively, the required anhydrides (116) can be

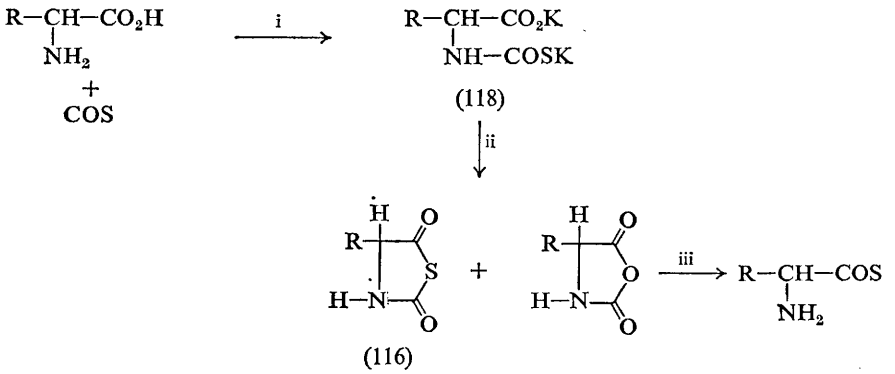


(117) (116)

Conditions: i, KOH; ii, PBr_3; iii, HBr

Scheme 53

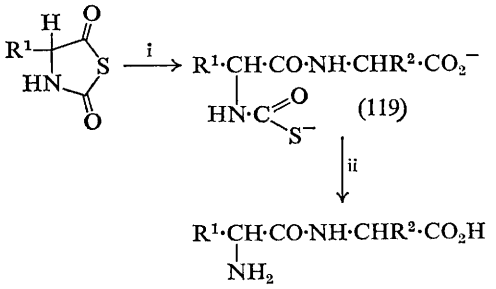
prepared by cyclization of amino-acid thiocarbamates (118) with phosphorus pentachloride (Scheme 54), or by the reaction of an amino-thio-acid with phosgene, but in general lower yields are obtained with these methods.[101] In general the *N*-thiocarboxy-anhydrides have an optical purity better than 98%, as revealed by hydrolysis to the corresponding amino-acid.



Conditions: i, KOH; ii, PCl_5; iii, H_2S

Scheme 54

In repetitive peptide synthesis the *N*-thiocarboxy-anhydrides in general give higher yields than do the corresponding *N*-carboxy-anhydrides, and on account of the greater stability of the thiocarbamate (119) produced, the coupling reaction is carried out at lower pH (Scheme 55) although side-



Conditions: i, $NH_2 \cdot CHR^2 \cdot CO_2^-$, pH 9, 0 °C; ii, H^+

Scheme 55

products are still observed. In contrast to the *N*-carboxy-anhydride of histidine, the *N*-thiocarboxy-anhydride can be used successfully for peptide synthesis provided the reagent is used in excess.[101] The *N*-thiocarboxy-anhydrides of glycine and alanine give significantly higher yields of product than do the corresponding *N*-carboxy-anhydrides.

Unfortunately, in contrast to the *N*-carboxy-anhydrides which give optically pure products, the *N*-thiocarboxy-anhydrides give products containing significant amounts of epimer. The optical purity of a series of products was measured by ^{1}H n.m.r. spectroscopy using the ^{13}C satellites as internal standards, or by hydrogen-isotope exchange, which showed the presence of 1—20% of the epimeric peptide.[101]

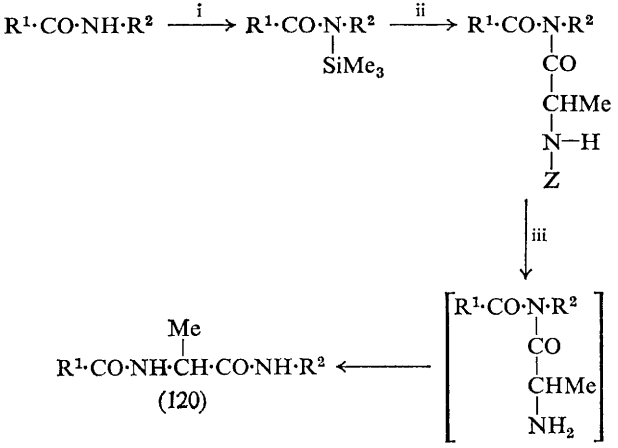
The controlled repetitive method of peptide synthesis using *N*-carboxy-anhydrides has been modified by carrying out the coupling reaction in a two-phase solvent system (acetonitrile–water, 60 : 50, v/v) containing sodium carbonate at an apparent pH of 11.6 (careful control of pH is unnecessary) at − 15 °C; raising the temperature to 40 °C enables the resulting carbamate to be decomposed, thus revealing a new amino-group for further chain elongation. The coupling reaction probably occurs at the solvent interface, the *N*-carboxy-anhydride being protected by the organic phase against side-reactions such as hydrolysis and polymerization. Furthermore, over-reaction is reduced since the carbamate produced appears to be stabilized in the mixed solvent system.[103]

A paper has appeared on the use of *N*-carboxy-anhydrides in the synthesis of oligopeptides containing ε-benzyloxycarbonyl-lysine,[104] and a study of the mechanism of Leuchs' anhydride formation from *N*-benzyloxycarbonyl-amino-acids has been reported.[105]

Other Methods. A symposium paper has appeared on the use of 1-hydroxy-benzotriazole as an additive in dicyclohexylcarbodi-imide-mediated couplings.[106] A report has now appeared on the evaluation of acyloxysilanes as acylating reagents (see vol. 2 of these Reports, p. 157) for peptide synthesis.[107] Carboxyl components are readily converted into the corresponding tetra-acyloxysilanes on treatment with silicon tetrachloride, and these react with amino-components to give moderate yields of simple dipeptides. These dipeptides were shown to be extensively racemized, and furthermore the benzyloxycarbonyl group is not compatible with this method of coupling since it undergoes cleavage.

A novel method has been developed for the insertion of amino-acid residues into peptides *via* the *N*-trimethylsilyl derivatives.[108] The reaction sequence is outlined in Scheme 56. Unambiguous synthesis of the product (120) showed that no significant degree of racemization had occurred in the course of the insertion process.

[103] Y. Iwakura, K. Uno, M. Oya, and R. Katakai, *Biopolymers*, 1971, **9**, 1419.
[104] T. D. Skalaban, I. M. Nazimov, S. S. Pankova, E. N. Zvonkova, R. P. Evstigneeva, and N. A. Preobrazhenskii, *Zhur. org. Khim.*, 1971, **7**, 47 (*Chem. Abs.*, 1971, **74**, 100 403e).
[105] I. Z. Siemion and D. Konopinska, *Roczniki Chem.*, 1970, **44**, 785 (*Chem. Abs.*, 1971, **74**, 23 116j).
[106] W. König and R. Geiger, ref. 7, p. 17.
[107] T. H. Chan and L. T. L. Wong, *J. Org. Chem.*, 1971, **36**, 850.
[108] J. S. Davies, C. H. Hassall, and K. H. Hopkins, *Chem. Comm.*, 1971, 1118.



Conditions: i, Me_3SiCl–NEt_3; ii, Z-Ala-Cl; iii, H_2–Pd/C

Scheme 56

A monograph[109] entitled 'Isonitrile Chemistry', which contains a chapter on the application of four-component condensations to peptide syntheses, has appeared. The formation of peptides from glycine in the presence of trimetaphosphate has been studied.[110] Calculations of the relative statistical distribution of peptides in stepwise synthesis (as a function of a constant coupling yield at each step) have been tabulated.[110a]

C. Racemization.—The results of racemization tests on various new coupling methods have been mentioned in the previous section.

The azide method of coupling, which has been shown previously to engender racemization when a vast excess of triethylamine is present, has now been shown to cause slight racemization even when the added base is not present in excess.[111] In the preparation of (121) by a 2 + 2 azide

Z-Val-His-Pro-Phe-OMe

(121)

coupling, triethylamine in amount equivalent to the hydrogen chloride present led to a peptide product of 90—95% optical purity; *N*-methylmorpholine led to 90—95% optical purity, and di-isopropylethylamine led to 97—98% optical purity. A number of examples were cited which

[109] G. Gokel, P. Hoffman, H. Kleimann, H. Klusacek, G. Luedke, D. Marquarding, and I. Ugi, in 'Isonitrile Chemistry', ed. I. Ugi, Academic Press, New York, 1971, p. 201.
[110] N. M. Chung, R. Lohrmann, L. E. Orgel, and J. Rabinowitz, *Tetrahedron*, 1971, **27**, 1205.
[110a] J. M. A. Baas, H. C. Beyerman, B. van de Graaf, and E. W. B. de Leer, ref. 7, p. 173.
[111] P. Sieber, M. Brugger, and W. Rittel, ref. 7, p. 60.

indicate the importance of carefully selecting the conditions to be used in a prospective azide synthesis.

A discussion has appeared on the base strengths of a series of tertiary amines in various organic solvents (measured spectrophotometrically using 2,4-dinitrophenol as indicator) and their relationships to the racemization problem.[112] A kinetic study of the rate of racemization and the rate of deuterium exchange of *N*-benzyloxycarbonyl-*S*-benzyl-L-cysteine pentachlorophenyl ester in non-polar solvent in the presence of triethylamine and monodeuteriomethanol indicates that base-catalysed α-hydrogen abstraction proceeds *via* isoracemization.[113] Preliminary kinetic data indicate that the base-catalysed racemization of *N*-benzyloxycarbonyl-L-phenylalanine pentachlorophenyl ester also proceeds *via* isoracemization.[113]

The Izumiya racemization test involves the determination of the amount of glycyl-D-alanyl-L-leucine in glycyl-L-alanyl-L-leucine by use of an amino-acid analyser. A simplified procedure has been described for the separation of these two peptides by ion-exchange chromatography on resin-coated chromatoplates.[114] A review has appeared on the use of gas–liquid chromatography in the determination of the optical purity of amino-acids and peptides.[115] A section of this review deals with the chromatographic resolution of diastereoisomeric peptides. The kinetics of the acid-catalysed reaction of DL-alanine ethyl ester with 2-phenyl-4,4-dimethyl-2-oxazolin-5-one have been studied.[116]

Treatment of α-melanotropin with hot alkali causes considerable racemization of some residues [82] (see p. 339).

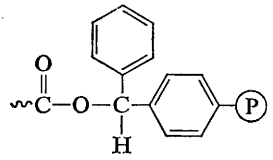
The effects of substituents in the ester aromatic nucleus of benzoyl-L-leucine phenyl ester on the rates of racemization by triethylamine and coupling with benzylamine have been investigated.[116a].

When determining the stereochemical purity of a peptide by acid hydrolysis and subsequent analysis of the hydrolysate, it is important to be able to correct for the racemization actually occurring during the hydrolysis. A new method for doing this is based on the use of tritiated hydrochloric acid for hydrolysis and measurement of the incorporation of radioactivity into each amino-acid by monitoring the effluent from an amino-acid analyser with a flow-cell scintillation counter.[117]

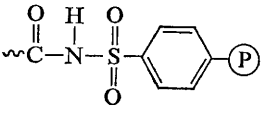
D. Repetitive Methods of Peptide Synthesis.—Brief reports of investigations of macroreticular resins as solid-phase supports have appeared.[118, 119]

[112] A. W. Williams and G. T. Young, ref. 7, p. 52.
[113] J. Kovacs, H. Cortegiano, R. E. Cover, and G. L. Mayers, *J. Amer. Chem. Soc.*, 1971, **93**, 1541.
[114] T. Devenyi, *Acta Biochim. Biophys.*, 1970, **5**, 441 (*Chem. Abs.*, 1971, **75**, 36 657g).
[115] J. W. Westley in 'Chemistry and Biochemistry of Amino-acids, Peptides and Proteins', ed. B. Weinstein, Marcel Dekker, New York, 1971, vol. 1, p. 1.
[116] H. Rodríguez, C. Chuaqui, S. Atala, and A. Márquez, *Tetrahedron*, 1971, **27**, 2425.
[116a] J. Morawiec, D. Konopinska, and I. Z. Siemion, *Roczniki Chem.*, 1971, **45**, 771 (*Chem. Abs.*, 1971, **75**, 98 786g).
[117] J. M. Manning, *J. Amer. Chem. Soc.*, 1970, **92**, 7449.
[118] M. A. Tilak and S. C. Hollinden, *Org. Prep. Proced.*, 1971, **3**, 183 (*Chem. Abs.*, 1971, **75**, 141 164s).
[119] S. Sano, R. Tokunaga, and K. A. Kun, *Biochim. Biophys. Acta*, 1971, **244**, 201.

Polymeric supports linked to the peptide through benzhydryl ester (122) [120, 121] offer the advantage that the final separation from the polymer requires milder acidolysis than does the Merrifield resin: very acid-labile α-amino-protecting groups are of course required for use in conjunction with (122), and the enamine type has been used.[121] The safety-catch principle has been incorporated into the design of the new polymeric support (123). The peptide–polymer link is completely acid stable, and nucleophilic

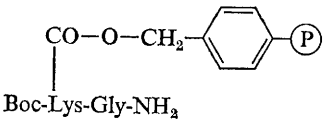


(122)



(123)

attack at the acylsulphonamide carbonyl group is prevented by the facile ionization of the N—H group. *N*-Methylation (with diazomethane), however, eliminates this inhibitory effect and the peptide–polymer link is then easily cleaved with alkali, ammonia, or hydrazine: the preliminary report describes the use of this support for synthesis of several oligopeptide derivatives which were suitable for further elaboration by classical means.[122] Some limitations were noted—including slight racemization of *C*-terminal residues—but with further development the method could become very useful. It has been suggested that the scope of solid phase peptide synthesis could be extended by using side-chain functionalities for attachment to the polymer, and the idea has been illustrated by a synthesis of lysine vasopressin by an essentially stepwise solid-phase procedure but starting from (124),



(124)

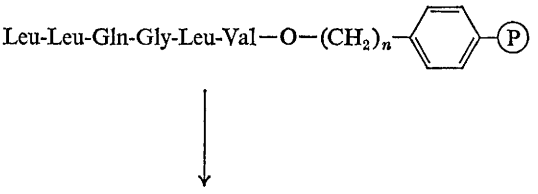
which was obtained *via* the corresponding polymeric chloroformate.[123] A symposium paper on the use of t-alkoxycarbonylhydrazide- (see vol. 2 of these Reports, p. 162) and t-alkyl alcohol-resins for the preparation of protected fragments has been published.[124]

The idea of using a soluble polymeric carrier so that coupling steps can be performed in homogeneous solution has been revived. In its first form

[120] G. L. Southard, G. S. Brooke, and J. M. Pettee, *Tetrahedron*, 1971, **27**, 2701.
[121] G. L. Southard, G. S. Brooke, and J. M. Pettee, ref. 7, p. 95.
[122] G. W. Kenner, J. R. McDermott, and R. C. Sheppard, *Chem. Comm.*, 1971, 636.
[123] J. Meienhofer and A. Trzeciak, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 1006.
[124] S. S. Wang and R. B. Merrifield, ref. 7, p. 74.

(see vol. 1 of these Reports, p. 164) this approach involved the separation of excess reagents *etc.* by washing the peptide-polymer conjugate with solvents which did not dissolve it: in the modified method which has been briefly described, polyethylene glycol is the polymeric carrier and ultrafiltration is used for this step.[125]

Many pharmacologically active polypeptides possess a *C*-terminal amide group. The obvious means of constructing this feature in a solid-phase synthesis involving a Merrifield resin is to use ammonolysis for scission of the peptide–resin ester link. This is fine if the *C*-terminal residue is glycine, as in oxytocin, where the method has been used with considerable success. Hindered *C*-terminal residues cause difficulty, although the *C*-terminal hexapeptide amide (125) of secretin, which has *C*-terminal valine, has been obtained using forcing ammonolysis conditions (Scheme 57; $n = 1$ or 6).[126, 127] In one of these examples (Scheme 57; $n = 6$)[127] a modified



(125)

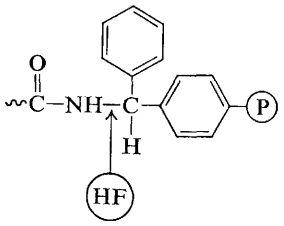
Conditions: NH_3–DMF (50 : 50 by vol.) at room temperature

Scheme 57

support was used: here the steric hindrance is presumably further reduced by greater distance between the site of ammonolysis and the polymeric web. An alternative approach to the solid-phase synthesis of *C*-terminal amides is to use a support which itself embodies an amide as the connecting functionality, removal from the carrier being accomplished by cleavage between the amide nitrogen and the polymer. Such is the principle of supports (126) derived from benzhydrylamine, which have recently proved useful in syntheses of Substance P and the luteinizing-hormone releasing factor (see Sections 3G and 3B, respectively).

Methods for the attachment of the first residue to Merrifield resins have received some attention. The most commonly used procedure has been refluxing the chloromethylated resin with the t-butoxycarbonyl derivative of the first residue in a solvent such as ethanol containing triethylamine. Some of the reactive groups on the resin become quaternized as a side-

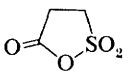
[125] M. Mutter, H. Hagenmaier, and E. Bayer, *Angew. Chem. Internat. Edn.*, 1971, **10**, 811.
[126] W. Parr, C. Yang, and G. Holzer, *Tetrahedron Letters*, 1972, 101.
[127] E. Bayer, E. Breitmaier, G. Jung, and W. Parr, *Z. physiol. Chem.*, 1971, **352**, 759.



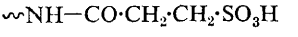
(126)

reaction—this can be avoided by using instead the tetramethylammonium salt of the t-butoxycarbonylamino-acid.[128] Treatment of the chloromethylated support with the t-butoxycarbonylamino-acid and methylamine in DMF at room temperature has been recommended as a mild procedure: this modification is suitable for use with t-butoxycarbonylasparagine, which is subject to side-reactions in the conventional technique.[129]

The problem of 'failure' and 'error' sequences in solid-phase products has been further considered by Bayer and his colleagues[130] who have developed[131] a sensitive method for detecting error sequences, based on ^{19}F n.m.r. spectroscopy after trifluoroacetylation of the solid-phase product: the ^{19}F shift in trifluoroacetylpeptides is very sensitive to the nature of the modified residue, so that contaminants with the wrong *N*-terminus are easily detected. Non-quantitative coupling has usually been blamed for the presence of failure and error sequences in solid-phase products, but this may not be the whole truth since incomplete deprotection has been identified as the difficulty in some cases,[132] and irreversible blocking of *N*-termini by impurities in methylene chloride has also been recognized.[133] Treatment with (127) after each coupling step has been recommended: this facilitates removal of the products of incomplete coupling at the end of the synthesis since these are converted by (127) into strongly acidic derivatives (128)—*cf.* the use of 3-nitrophthalic anhydride (see vol. 2 of these Reports, p. 163).



(127)



(128)

[128] A. Loffet, *Internat. J. Protein Res.*, 1971, **3**, 297.
[129] A. Marglin, *Tetrahedron Letters*, 1971, 3145.
[130] E. Bayer, H. Hagenmaier, G. Jung, W. Parr, H. Eckstein, P. Hunzicker, and R. E. Sievers, ref. 7, p. 65.
[131] E. Bayer, P. Hunziker, M. Mutter, R. E. Sievers, and R. Uhmann, *J. Amer. Chem. Soc.*, 1972, **94**, 265.
[132] F. C.-H. Chou, R. K. Chaivla, R. F. Kibler, and R. Shapira, *J. Amer. Chem. Soc.*, 1971, **93**, 267.
[133] K. Brunfeldt and T. Christensen, *F.E.B.S. Letters*, 1972, **19**, 345.

Other relevant papers include studies of the use of symmetrical anhydrides,[134] succinimido-esters,[135] triazole with active esters,[136] and 1-ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline ('EEDQ')[137] for coupling in solid-phase work, alcoholysis of the peptide resin ester link,[138] comparisons of the utility of different *N*-protecting groups,[139] of rates of coupling obtained using different resins,[140, 141] and of the reactivities of different t-butoxycarbonylamino-acids in solid-phase coupling,[142] process control,[143–145] a model system suitable for screening modified techniques,[146] and apparatus and instrumentation for automated solid-phase synthesis.[147–149]

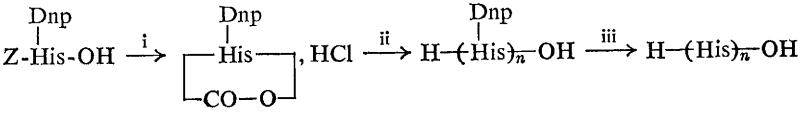
Other repetitive methods have been dealt with elsewhere—the use of *N*-carboxy-anhydrides in the section on formation of the peptide bond and the use of picolyl esters in the section on carboxy-group protection.

E. Synthesis of Polymeric Models for Studies in Protein Chemistry.—*Polyamino-acids.* By oversight, reference to an important book[150] which contains useful detailed chapters on the kinetics and mechanism of *N*-carboxy-anhydride polymerization was omitted from earlier volumes of these Reports.

The usefulness of the 2,4-dinitrophenyl protecting group for histidine side-chains has been further demonstrated by its employment in a synthesis of poly-L-histidine (Scheme 58):[70] a synthesis of poly-L-histidine which uses N^{im}-t-butoxycarbonyl protection has been discussed elsewhere (Scheme 33).

A synthesis of poly-(α-ethyl aspartate) has been described,[151] and poly-DL-serine has been obtained *via* serine azide hydrobromide.[152]

[134] T. Wieland, C. Birr, and F. Flor, *Angew. Chem. Internat. Edn.*, 1971, **10**, 336.
[135] A. Orlowska and S. Drabarek, *Roczniki Chem.*, 1971, **45**, 339 (*Chem. Abs.*, 1971, **75**, 64 246r).
[136] U. Ragnarsson, G. Lindeberg, and S. Karlsson, *Acta Chem. Scand.*, 1970, **24**, 3079.
[137] H. Yajima and H. Kawatani, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 1905.
[138] H. C. Beyerman, H. Hindricks, J. Hirt, and E. W. B. de Leer, ref. 7, p. 87.
[139] U. Ragnarsson, S. Karlsson, and G. Lindeberg, *Acta Chem. Scand.*, 1970, **24**, 2821.
[140] J. J. Maher, M. E. Furey, and L. J. Greenberg, *Tetrahedron Letters*, 1971, 27.
[141] A. Losse, *Tetrahedron Letters*, 1971, 4989.
[142] U. Ragnarsson, S. Karlsson, and B. Sandberg, *Acta Chem. Scand.*, 1971, **25**, 1487.
[143] H. C. Beyerman and H. Hindriks, ref. 7, p. 145.
[144] B. Mehlis, W. Fischer, and H. Niedrich, ref. 7, p. 146.
[145] K. Brunfeldt, P. Roepstorff, and J. Thomsen, ref. 7, p. 148.
[146] L. C. Dorman, L. D. Markley, and D. A. Mapes, *Analyt. Biochem.*, 1971, **39**, 492.
[147] C. Birr and W. Lochinger, *Synthesis*, 1971, 319.
[148] A. M. Tometsko, J. H. Garden, sec., and J. Tischio, *Rev. Sci. Instr.*, 1971, **42**, 331.
[149] R. Boni, G. M. Bonara, L. Ciceri, A. Gambini, A. Scatturin, and E. Scoffone, *Chimica e Industria*, 1971, **53**, 10 (*Chem. Abs.*, 1971, **75**, 62 842r).
[150] M. Szwarc, 'Carbanions, Living Polymers, and Electron Transfer Processes', Interscience, New York, 1968.
[151] S. E. Moschopedis and C. Mumford, *Canad. J. Chem.*, 1971, **49**, 2158.
[152] W. J. Bailey, N. Kawabata, and R. C. Capozza, 'Kinetics and Mechanisms of Polyreactions', International Symposium on Macromolecular Chemical Preparation, 1969, vol. 1, p. 195 (*Chem. Abs.*, 1971, **75**, 49 558a).

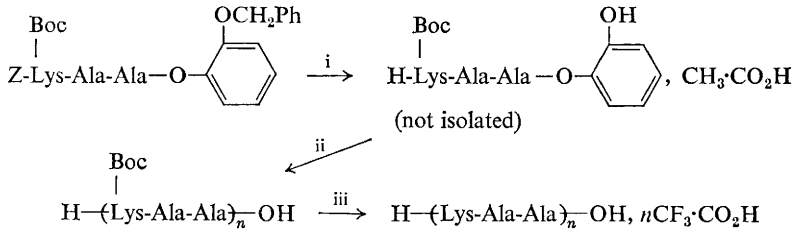


Conditions: i, $SOCl_2$; ii, Ag_2O followed by filtration then Et_3N–dioxan; iii, $HS{\cdot}CH_2{\cdot}CH_2{\cdot}OH$–DMF

Scheme 58

Sequential Polypeptides. The active-ester method of polymerization now enjoys exclusive application in sequential polypeptide synthesis: various active esters have been recommended for this purpose from time to time, but the succinimido-ester now seems the most favoured.[153–157]

A detailed paper describing the preliminary work on the catechol ester method of polymerization (see vol. 2 of these Reports, p. 71) has appeared.[158] Difficulties with the hydrogenolytic deprotection and activation of benzyloxycarbonylpeptide 2-benzyloxyphenyl esters have been traced to sulphur-containing catalyst poisons derived from reagents used in their preparation. Rigorous avoidance of such reagents has solved this problem and enabled the extension of the method to the synthesis of sequential polypeptides with side-chains protected by means of t-butyl alcohol-derived substituents[159] (*e.g.* Scheme 59). An alternative modification devised by Trudelle[160] has been illustrated by the homopolytripeptide synthesis shown in Scheme 60,

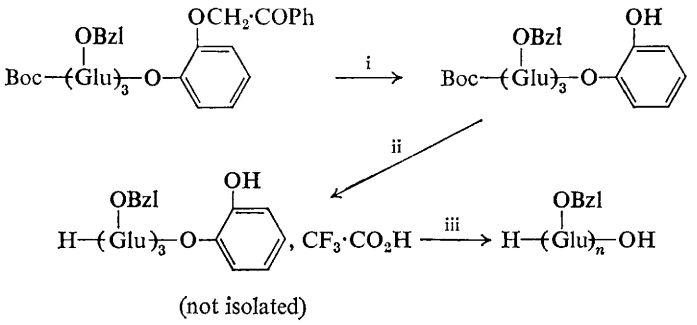


Conditions: i, H_2/Pd–AcOH; ii, Et_3N–Me_2SO; iii, 90% aq. $CF_3{\cdot}CO_2H$

Scheme 59

which incidentally provides some further evidence for the freedom of catechol ester polymerizations from racemization, since the specific rotation of the polymer obtained was the same as that yielded by the appropriate *N*-carboxy-anhydride.

[153] G. P. Lorenzi, B. B. Doyle, and E. R. Blout, *Biochemistry*, 1971, **10**, 3046.
[154] P. M. Hardy, J. C. Haylock, D. I. Marlborough, H. N. Rydon, H. T. Storey, and R. C. Thompson, *Macromolecules*, 1971, **4**, 435.
[155] J. Ramachandran, A. Berger, and E. Katchalski, *Biopolymers*, 1971, **10**, 1829.
[156] B. Schechter, I. Schechter, J. Ramachandran, A. Conway-Jacobs, and M. Sela, *European J. Biochem.*, 1971, **20**, 301.
[157] A. Yaron, N. Turkeltaub, and A. Berger, *Israel J. Chem.*, 1970, **8** (suppl.), 180p.
[158] R. D. Cowell and J. H. Jones, *J. Chem. Soc.* (*C*), 1971, 1082.
[159] R. D. Cowell and J. H. Jones, *Chem. Comm.*, 1971, 1009.
[160] Y. Trudelle, *Chem. Comm.*, 1971, 639.



Conditions: i, Zn–AcOH; ii, $CF_3 \cdot CO_2H$; iii, Et_3N–DMF

Scheme 60

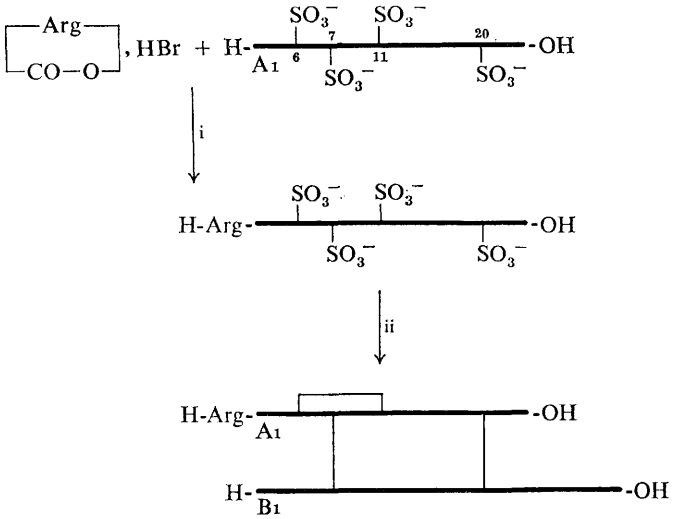
A complete list of the sequential polypeptides synthesized this year is to be found in Appendix A, part B.

F. Synthetic Operations with Peptides of Biological Origin.—As before, we confine ourselves here to work involving the controlled formation of new peptide bonds.

With Naked Natural Peptides. Further discussion of the extension of porcine β-MSH to [Lys[10]]-human β-MSH by reaction with a protected tetrapeptide azide (see vol. 3 of these Reports, p. 251) has been given at a symposium.[161] This particular case was complicated by the fact that porcine β-MSH contains two lysine residues and therefore three nucleophilic sites, so that a large number of coupling products were possible. The substantial pK_a difference between α- and ε-amino-groups encouraged the authors to attempt to direct coupling towards the α-amino-function by working at a pH such that most of the ε-amino-substituents were removed from the sphere of action by protonation. A complex mixture of mono- and poly-acyl derivatives was nevertheless obtained, but it fortunately proved possible to resolve this mixture on carboxymethylcellulose. The direct use of protein-degradation products in synthetic operations is much less complicated when there is no ambiguity about the point of attack. Such is the case when insulin A-chain tetra-*S*-sulphonate is involved as the nucleophilic partner in coupling, and a number of A-chain analogues extended at the *N*-terminus have been obtained in this way.[162–165] Acylamino-acid active esters or protected peptide azides were most commonly used as carboxyl components in solvents such as DMF–water. In one example,[163] however, an *N*-carboxy-anhydride was used, so that fully aqueous media were

[161] S. Lande and J. Burton, ref. 7, p. 109.
[162] M. Weinert, D. Brandenburg, and H. Zahn, *Z. physiol. Chem.*, 1969, **350**, 1556.
[163] M. Weinert, K. Kircher, D. Brandenburg, and H. Zahn, *Z. physiol. Chem.*, 1971, **352**, 719.
[164] V. K. Naithani and M. Weinert, unpublished work quoted in ref. 163.
[165] Y. Shimonishi, *Bull. Chem. Soc. Japan*, 1970, **43**, 3251.

appropriate (Scheme 61). Ion-exchange chromatography was of course necessary to separate the desired product from un- and over-reacted peptide, but this proved straightforward and the modified bovine A-chain obtained eventually yielded a crystalline bovine insulin analogue, after



Conditions: i, coupling at pH 10 followed by decarboxylation at pH 3, followed by ion-exchange chromatography; ii, reduction and oxidation in the presence of B-chain bis-*S*-sulphonate followed by gel chromatography and crystallization

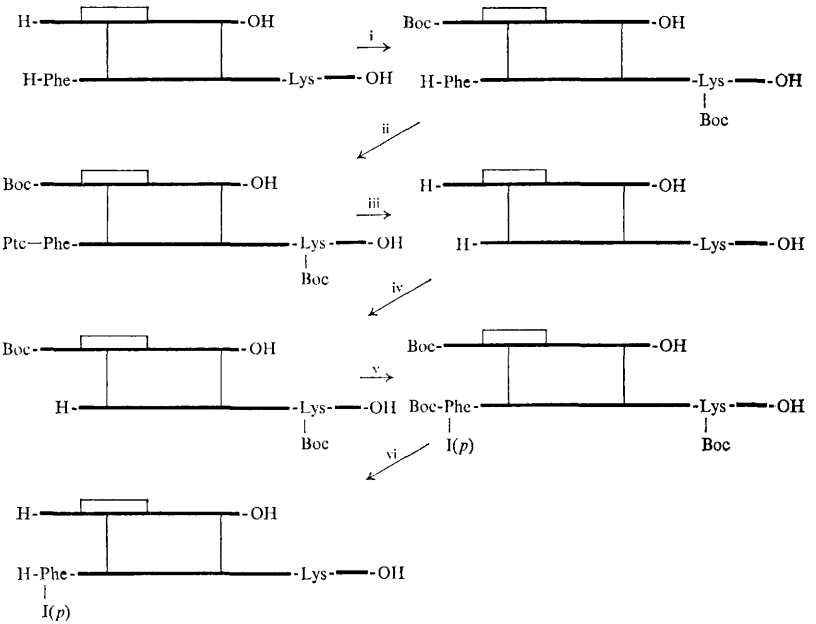
Scheme 61

reduction and oxidation together with natural bovine B-chain bis-*S*-sulphonate.

With Partially Blocked Peptides. It has been shown in the case of insulin [166] and of porcine β-MSH [167] that t-butoxycarbonylazide can be used for partial protection. Porcine β-MSH apparently reacts with this reagent at pH 10.5 or in water–pyridine–triethylamine (10 : 10 : 1) most rapidly at the ε-amino-groups, the differentiation presumably being simply a matter of relative nucleophilicity.[167] Insulin reacts with the same acylating agent in DMF–aqueous sodium bicarbonate to give the α(A1),ε(B29) bis-t-butoxycarbonyl derivative as the major product. Here the differentiation between the two α-amino-groups, which are presumably of comparable intrinsic nucleophilicity, is probably steric in origin, and this suggestion is borne out by the fact that the less hindered acylating agent methyl succinimidocarbonate gives a triacylinsulin (*cf.* recent studies [168] on the reaction of insulin with succinimido acetate). The bis-t-butoxycarbonylinsulin obtained

[166] R. Geiger, H.-H. Schöne, and W. Pfaff, *Z. physiol. Chem.*, 1971, **352**, 1487.
[167] S. Lande, *J. Org. Chem.*, 1971, **36**, 1267.
[168] D. G. Lindsay and S. Shall, *Biochem. J.*, 1971, **121**, 737.

has only one nucleophilic point, and the stage is therefore set for the preparation of semisynthetic insulins extended from the B-*N*-terminus. No realization of this possibility has been described yet, but the principle has been used in an ingenious sequence of reactions leading from insulin to a specifically iodinated derivative (Scheme 62).[169, 170]



Conditions: i, Boc-N_3–DMF–aq.$NaHCO_3$; ii, PhNCS; iii, $CF_3 \cdot CO_2H$; iv, Boc-N_3–DMF–aq.$NaHCO_3$; v, Boc-Phe(*p*-I)-OTcp–Me_2SO–*N*-methylmorpholine; vi, $CF_3 \cdot CO_2H$

Scheme 62

Anfinsen and his colleagues have outlined some of their plans for the resynthesis of fragment (6—48) of staphylococcal nuclease from its tryptic peptides (as a preliminary to a programme of study on semisynthetic analogues). They are considering the use of a water-soluble carbodi-imide for coupling, and have reported exploratory relevant experiments with a model system.[171]

3 Syntheses Achieved and Structure–Activity Correlations

As in previous volumes, we make apology for the fact that the discussion in this chapter is highly selective: the Reader is referred to Appendix A for a comprehensive list of syntheses described during the year.

[169] R. Geiger, *Z. physiol. Chem.*, 1971, **352**, 7.
[170] G. Krail, D. Brandenburg, H. Zahn, and R. Geiger, *Z. physiol. Chem.*, 1971, **352**, 1595.
[171] N. Izumiya, K. Noda, and C. B. Anfinsen, *Arch. Biochem. Biophys.*, 1971, **144**, 237.

A large number of books, reviews, and conference proceedings dealing with various aspects of polypeptide hormone chemistry and pharmacology have come to our notice, although our list is no doubt still incomplete, as it is difficult to keep systematic track of this kind of literature. Subjects covered include: methods of peptide hormone assay,[172] adrenocorticotropin,[173, 174] angiotensin,[175–177] bradykinin and other kinin hormones,[178, 179] calcitonin,[180] gastrointestinal hormones,[181] growth hormone,[174] hypothalamic releasing factors,[174, 182–185] insect venoms,[2] insulin,[174, 186] lipotropin,[174] neurohypophyseal hormones,[187] peptide hormones and the brain,[188] prolactin,[174] snake venoms,[189] Substance P,[190] and vasopressin.[191]

There has been a great deal of work recently on the involvement of cyclic adenosine monophosphate as a 'second messenger' in the mediation

[172] H. van Cauwenberge and P. Franchimont, 'Assay of Protein and Polypeptide Hormones', Pergamon Press, Oxford, 1970.

[173] L. D. Garren, G. N. Gill, H. Masui, and G. M. Walton, *Recent Progr. Hormone Res.*, 1971, **27**, 433.

[174] Proceedings of the Fourth Congress of the Hungarian Society of Endocrinology and Metabolism held in Budapest, 1969: 'Polypeptide Hormones', ed. E. Goth and J. Fövenyi, Akademiai Kiado, Budapest, 1971.

[175] G. W. Boyd and W. S. Peart, *Adv. Metabolic Disorders*, 1971, **5**, 77.

[176] J. W. Fisher, 'Kidney Hormones', Academic Press, 1971.

[177] F. Gross in 'Pharmacology of Naturally Occurring Polypeptides and Lipid-soluble Acids', ed. J. M. Walker, Pergamon Press, Oxford, 1971, vol. 1, p. 73.

[178] M. Rocha e Silva, 'Kinin Hormones', Charles C. Thomas, Springfield, Illinois, 1970.

[179] 'Bradykinin, Kallidin and Kallikrein', ed. E. G. Erdös, Springer Verlag, Berlin, 1970.

[180] Proceedings of the Second International Symposium on Calcitonin held in London, 1969: 'Calcitonin 1969', ed. S. Taylor and G. Foster, William Heinemann Medical Books, London, 1970.

[181] Proceedings of a Symposium held at Kingston, Ontario, 1969: 'The Exocrine Pancreas', ed. I. T. Beck and D. G. Sinclair, Churchill, London, 1971.

[182] R. Guillemin, *Adv. Metabolic Disorders*, 1971, **5**, 1; A. V. Schally and A. J. Kastin, *Adv. Steroid Biochem. and Pharmacol.*, 1970, **2**, 41.

[183] Proceedings of a conference on the Bioassay and Chemistry of the Hypophysiotropic Hormones of the Hypothalamus held at Tucson, Arizona 1969: 'Hypophysiotropic Hormones of the Hypothalamus: Assay and Chemistry', ed. J. Meites, Williams and Wilkins, Baltimore, 1970.

[184] 'Neurochemical Aspects of Hypothalamic Function', ed. L. Martini and J. Meites. Academic Press, London, 1971 (Proceedings of a Symposium held in Milan, Italy, 1969, during the 2nd Meeting of the International Society for Neurochemistry).

[185] 'The Hypothalamus', ed. L. Martini, M. Motta, and F. Fraschini, Academic Press, London, 1971 (Proceedings of the Workshop Conference on the Integration of Endocrine and Non-endocrine Mechanisms in the Hypothalamus held in Stresa, Italy, 1969).

[186] G. M. Grodsky, *Vitamins and Hormones*, 1970, **28**, 37.

[187] 'Pharmacology of the Endocrine System and Related Drugs: The Neurohypophysis', ed. H. Heller and B. T. Pickering, Pergamon Press, Oxford, 1970.

[188] Proceedings of a Conference on the Pituitary–Adrenal Axis and the Nervous System held at Vierhouten, The Netherlands, 1969: 'Pituitary Adrenal and the Brain', ed. D. de Weid and J. A. W. M. Weijnen, Elsevier, Amsterdam, 1970.

[189] S. B. Henriques and O. B. Henriques, in 'Pharmacology and Toxicology of Naturally Occurring Toxins', ed. H. Raskova, Pergamon Press, Oxford, 1971, vol. I, part II.

[190] F. Lembeck and G. Zetter, in 'Pharmacology of Naturally Occurring Polypeptides and Lipid Soluble Acids', ed. J. M. Walker, Pergamon Press, Oxford, 1971, vol. I, p. 29.

[191] N. A. Thorn, *Adv. Metabolic Disorders*, 1970, **4**, 40.

of the effects of peptide hormones, and a book on the subject is therefore very welcome.[192]

Biologists have been warned [193] that when synthetic peptides are used in quantitative work, 100% purity should not be assumed without supporting evidence: account must be taken of erroneous sequences if present, as well as of solvent and salt content.

A. Calcitonin.—A second synthesis of human calcitonin has been outlined at a symposium:[194] the synthesis of salmon calcitonin published [195] in 1969 was discussed [196] at the same symposium. The intriguing thing about the salmon hormone is that in the rat and in man it is very much more potent than calcitonins of mammalian origin.[197] Evidence has recently been presented [198] to support the view that much of this high potency is due to slow inactivation: infused salmon calcitonin disappears from the circulation of the anaesthetized dog more slowly than does porcine calcitonin. Such resistance to inactivation is, perhaps, not surprising—the salmon hormone is grossly different in amino-acid sequence from the mammalian calcitonins, although it has the same chain length and also incorporates a heptapeptide disulphide loop. An anonymous sage [199] has seen fit to enunciate a 'new concept' based on this finding. The 'new concept' is that 'the activity of a peptide hormone is dependent not only on those amino-acids which confer biological activity but also possibly on other amino-acids which affect its rate of destruction'. It has, of course, long been known that the deliberate introduction into a peptide hormone of structural features which can be expected to delay enzymatic inactivation often gives analogues with high apparent activity. This principle has been especially fruitful in the field of corticotropin analogues, a recent example [200] being provided by [1-β-alanine, 17-lysine]-β-corticotropin-(1—17)-heptadecapeptide-4-amino-n-butylamide which was designed to be resistant to amino- and carboxy-peptidases and was found to be more potent than the natural hormone despite its shorter chain length.

The results of a clinical trial of synthetic human calcitonin in five cases of Paget's disease have been published.[201] Paget's disease is a common

[192] G. A. Robison, R. W. Butcher, and E. W. Sutherland, 'Cyclic AMP', Academic Press, New York, 1971.
[193] J. R. Vane, *Nature*, 1971, **230**, 382; D. R. Bangham, D. H. Calam, J. A. Parsons, and C. J. Robinson, *Nature*, 1971, **232**, 631.
[194] H. M. Greven and L. J. W. M. Tax, ref. 7, p. 38.
[195] St. Guttmann, J. Pless, R. L. Huguenin, E. Sandrin, H. Bossert, and K. Zehnder, *Helv. Chim. Acta*, 1969, **52**, 1789.
[196] St. Guttmann, J. Pless, R. L. Hugenin, E. Sandrin, H. Bossert, and K. Zehnder, ref. 7, p. 54.
[197] H. T. Keutmann, J. A. Parsons, J. T. Potts, jun., and R. T. Schlueter, *J. Biol. Chem.*, 1970, **245**, 1491.
[198] J. F. Habener, F. R. Singer, L. J. Deftos, R. M. Neer, and J. T. Potts, jun., *Nature New Biol.*, 1971, **232**, 91.
[199] *Nature*, 1971, **232**, 156.
[200] R. Geiger, *Annalen*, 1971, **750**, 165.
[201] N. J. Y. Woodhouse, P. Bordier, M. Fisher, G. F. Joplin, M. Reiner, D. N. Kalu, G. V. Foster, and I. MacIntyre, *Lancet*, 1971 (I), 1139.

metabolic disorder characterized by very rapid bone formation and resorption in which the bones become weak and painful. Administration of synthetic calcitonin over a period of a year to the five patients restored biochemical normality and relieved the pain in all cases. In normal dogs, however, the administration of calcitonin is without long-term effect on strontium-85 (the deposition and release of which is very similar to that of calcium) whole-body retention [202] so that calcitonin may prove a disappointment as far as the treatment of bone-loss disease (osteoporosis) is concerned.

B. Hypothalamic Releasing Factors.—Activity continues to increase in this area following the outstanding recent success with thyroid-stimulating hormone releasing factor. Indeed the rate of publication is such that we recognize that much of the following section will be out of date (or be shown to be clearly wrong in some cases, perhaps) by the time it reaches the eyes of the Reader.

The prospects for diagnostic and therapeutic applications for synthetic hypothalamic releasing factors have been surveyed briefly.[203]

Growth-hormone Releasing Factor. Schally and his group have isolated a peptide from porcine hypothalami which has growth-hormone releasing factor (GH-RF) activity and they have assigned to it the structure (129).[204] This sequence has been synthesized by Denkewalter, Hirschmann, and their team,[205] together with the 9-glutamine analogue (130), since there was some

H-Val-His-Leu-Ser-Ala-Glu-Glu-Lys-Glu-Ala-OH

(129)

$$\begin{array}{l} \qquad\qquad\qquad\qquad\qquad\quad NH_2 \\ \qquad\qquad\qquad\qquad\qquad\quad | \\ \text{H-Val-His-Leu-Ser-Ala-Glu-Glu-Lys-Glu-Ala-OH} \end{array}$$

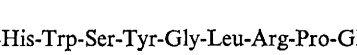
(130)

doubt about residue 9. The observation that (129) and (130) are very similar to the *N*-terminal sequence of porcine haemoglobin β-chain raises the possibility that haemoglobin may be the natural prohormone for growth-hormone releasing factor: the synthesis of the *N*-terminal decapeptide of human haemoglobin β-chain was therefore also performed. The syntheses were performed by azide condensations of fragments, a combination of *N*-carboxy-anhydride, succinimido-ester, and solid-phase

[202] J. J. B. Anderson. M. W. Balk, W. C. Crackel, M. K. Austin, E. C. Bollmeier, and R. Slusher, *Nature New Biol.*, 1971, **232**, 93.
[203] Editorial, *Brit. Med. J.*, 1972, **1**, 65.
[204] A. V. Schally, Y. Baba, R. M. G. Nair, and C. D. Bennett, *J. Biol. Chem.*, in the press (cited in ref. 205, but not available at the time of writing).
[205] D. F. Veber, C. D. Bennett, J. D. Milkowski, G. Gal, R. G. Denkewalter, and R. Hirschmann, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 235.

procedures being used for the preparation of the partial sequences. Unfortunately, details of the biological studies of these materials are not available at the time of writing, apart from the brief mention [205] that (129) is inactive, perhaps confirming suspicions that residue 9 is glutamine, and not glutamic acid.

Luteinizing-hormone Releasing Factor and Follicle-stimulating Hormone Releasing Factor. A peptide with both luteinizing-hormone releasing factor (LH-RF) and follicle-stimulating hormone releasing factor (FSH-RF) activities has been isolated by Schally and his colleagues from porcine hypothalamic extracts:[206–210] the two activities could not be separated, and so may reside in the same substance, but it should be noted that there may in fact be more than one peptide with LH-RF activity in porcine hypothalami since concurrent work by Geiger and his co-workers [211] has given two chromatographically well-separated preparations with this activity. The material isolated in Schally's laboratory was shown by degradation [212, 213]—using extremely small amounts of material at first but more substantial amounts after improvements [209] in the isolation procedures—to have the sequence (131). Since such small amounts of material were



(131)

initially available, the sequential analysis was not completely free of ambiguity and a solid-phase synthesis was therefore performed. A conventional strategy followed by ammonolysis from the resin, hydrogen fluoride deprotection, counter-current distribution, and finally carboxymethylcellulose chromatography gave a synthetic peptide with chromatographic and electrophoretic properties identical to the material of porcine origin.[214] The LH-RF activity of the synthetic preparation was somewhat greater than that of the natural substance, whereas the FSH-RF activity was about the same, but the natural material used for comparison was

[206] A. V. Schally, Y. Baba, A. Arimura, T. W. Redding, and W. F. White, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 50.
[207] A. V. Schally, A. Arimura, Y. Baba, R. M. G. Nair, H. Matsuo, T. W. Redding, and L. Debeljuk, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 393.
[208] A. V. Schally, A. Arimura, A. J. Kastin, H. Matsuo, Y. Baba, T. W. Redding, R. M. G. Nair, and L. Debeljuk, *Science*, 1971, **173**, 1036 (Summary paper).
[209] A. V. Schally, R. M. G. Nair, T. W. Redding, and A. Arimura, *J. Biol. Chem.*, 1971, **246**, 7230.
[210] Y. Baba, A. Arimura, and A. V. Schally, *J. Biol. Chem.*, 1971, **246**, 7581.
[211] R. Geiger, W. König, H. Wissmann, K. Geisen, and F. Enzmann, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 767.
[212] H. Matsuo, Y. Baba, R. M. G. Nair, A. Arimura, and A. V. Schally, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 1334.
[213] Y. Baba, H. Matsuo, and A. V. Schally, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 459.
[214] H. Matsuo, A. Arimura, R. M. G. Nair, and A. V. Schally, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 822.

impure, comprising only *ca.* 70% amino-acids by weight. Geiger and his colleagues have also synthesized [211] the sequence (131) using a strategy based on fragment condensation, mainly by the dicyclohexylcarbodi-imide–hydroxybenzotriazole technique developed in their laboratory: other points of interest are the use of 4,4′-dimethoxybenzhydryl and bis-benzyloxycarbonyl protection in the early stages for amide and guanidino-side-chains respectively, and the final purification by partition chromatography on Sephadex LH-20. Experienced workers will recognize the sequence (131) as a demanding exercise for unambiguous synthesis and will await full experimental detail with interest. The material synthesized by Geiger and his collaborators was shown by them to be chromatographically indistinguishable from one of the two different preparations with LH-RF activity obtained by them from porcine hypothalami: this synthetic material had LH-RF activity at nanogram levels, but no mention of examination for FSH-RF activity was made in the preliminary report.

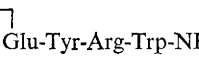
A second solid-phase synthesis of (131) has been described by Rivaille and his colleagues:[74] their synthesis is of special interest because of the use of a benzhydrylamine-type [215] resin support (126)* so that acidolysis after assembly of the peptide chain gave the required peptide amide directly. After scission from the resin, thiolysis was used to remove the 2,4-dinitrophenyl group which had been employed for protection of histidine, and, after purification by gel filtration, the product was homogeneous by electrophoretic and chromatographic criteria. The peptide thus obtained had greater LH-RF activity in normal man than was found [216] by Schally for the purified material of porcine origin: the FSH-RF activity of the synthetic peptide was on the other hand feeble.

Folkers and his colleagues were also engaged in the race for LH-RF but apparently in the first instance decided not to use the 'purify, isolate, analyse, and synthesize' tactics being simultaneously employed by Schally *et al.*, Geiger *et al.*, and Guillemin *et al.* Folkers' group based their approach on inactivation experiments using partially purified hypothalamic extracts with LH-RF activity. Investigations with specific enzymes and reagents [217, 218] suggested the following vital characteristics: a pyroglutamic acid *N*-terminal, an amide *C*-terminal, and the presence of tryptophan, tyrosine, and arginine residues. Since no other features could be shown to be essential by inactivation experiments it was reasoned that the releasing factor—or an active segment thereof—might comprise these amino-acids alone. All six possible tetrapeptide amides composed of pyroglutamic acid,

[215] P. G. Pietta and G. R. Marshall, *Chem. Comm.*, 1970, 650.
[216] A. J. Kastin, A. V. Schally, C. Gual, A. R. Midgley, jun., M. C. Miller, tert., and A. Cabeza, *J. Clin. Investigation*, 1971, **50**, 1551.
[217] Y. Baba, A. Arimura, and A. V. Schally, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 483.
[218] C. Bogentoft, B. L. Currie, H. Sievertsson, J.-K. Chang, K. Folkers, and C. Y. Bowers, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 403.

* Potential users of this type of resin will be interested to note that experimental detail of its preparation and use are given by Rivaille *et al.*[74]

arginine, tyrosine, and tryptophan were therefore synthesized.[219, 220] Only one (132) of the synthetic peptides had any activity but this singleton was relatively potent in assays for LH-RF, although no FSH-RF activity was detected, showing that the two actions are separable. Folkers and his co-workers recognized that (132) was not potent enough to be the natural



(132)

factor but suggested from their results that all or part of the sequence of (132) might appear in natural LH-RF. As far as the material isolated, analysed, and synthesized by Schally's group is concerned, Folkers' suggestion is clearly wrong since the relationship between (132) and (131) is highly tenuous, being confined to identical termini and limited similarity in functional side-chain composition. The possibility remains, however, that (132) may prove similar to the second substance in porcine hypothalami with LH-RF activity which was mentioned [211] in passing by Geiger and his colleagues. The finding that a relatively simple tetrapeptide derivative has LH-RF activity is of course of enormous potential clinical importance, and the demonstration that some gross structural changes are not necessarily inconsistent with the retention of LH-RF activity has probably already initiated the synthesis and screening of many analogues. The apparent insensitivity of LH-RF activity to structural change shown by this example also warns that caution must be exercised in 'proving' structures by synthesis in this area, especially if the solid-phase method is used for this purpose.

Ovine [221] and bovine [218] hypothalami yield LH-releasing factors which are very similar to and probably identical to that found in the pig.

The evidence [208] may seem persuasive that LH-RF and FSH-RF are identical, having the sequence (131). The security of this identification, however, remains to be fully tested. The notion that LH-RF and FSH-RF are identical goes against the simpler, and thus far tacitly accepted, hypothesis of one hypothalamic releasing factor for each pituitary hormone, although the generality of this hypothesis has very recently had doubt cast upon it from another direction by the demonstration [222] that TRF may also function as a releasing factor for prolactin. The report [211] that porcine hypothalami in fact contain at least two substances with LH-RF activity

[219] C. Y. Bowers, J.-K. Chang, H. Sievertsson, C. Bogentoft, B. L. Currie, and K. Folkers, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 414.
[220] J.-K. Chang, H. Sievertsson, C. Bogentoft, B. L. Currie, K. Folkers, and C. Y. Bowers, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 409.
[221] M. Amoss, R. Burgus, R. Blackwell, W. Vale, R. Fellows, and R. Guillemin, *Biochem. Biophys. Res. Comm.*, 1971, **44**, 205.
[222] C. Y. Bowers, H. C. Friesen, P. Hwang, H. J. Guyda, and K. Folkers, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1033.

is perhaps consistent with the control of LH and FSH release by a pair of peptides, one or both of which have some of the activity of the other.

Melanocyte-stimulating Hormone Release. It has been known for several years that the release of melanocyte-stimulating hormone (MSH) is under the predominant control of the hypothalamus, mediated by an MSH-release-inhibiting factor (MSH-RIF). It has recently been shown that this factor can be generated by the incubation of oxytocin with microsomal preparations from the stalk medium eminence of rats.[223] Further investigations demonstrated that prolyl-leucylglycinamide (133) is formed from

Pro-Leu-Gly-NH_2

(133)

oxytocin under these conditions, and Walter and his associates showed that a synthetic sample of this tripeptide did indeed possess MSH-RIF activity. The identification of the natural factor with (133) seems to be confirmed by concurrent work culminating in its isolation [224] from bovine hypothalami by Schally and his colleagues, but it should be noted that although bovine hypothalami do contain a potent MSH-RIF of structure (133) they also contain another potent material with this activity,[224] the structure of which has not yet been reported. In passing we note that also (133) has been found [225] to potentiate the behavioural effects of dopamine—an effect apparently not mediated by MSH because hypophysectomized mice reacted essentially as intact animals. Despite the apparently mutually corroborating reports from the groups of Walter and Schally that (133) has high MSH-RIF activity, Hruby *et al.*[226] report that in their hands pure synthetic (133) has no detectable MSH-RIF activity in rat or frog tissue. Hruby's group were however able to show that a synthetic peptide corresponding to the cyclic portion of oxytocin—'tocinoic acid' (134)—had

(134)

MSH-RIF activity in a rat pituitary assay at less than nanogram levels. The confusion may be in part due to the use of different assays and species in different laboratories: recent indications [226] are that there may be some species specificity for MSH-RF.

[223] M. E. Celis, S. Taleisnik, and R. Walter, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 1428.
[224] R. M. G. Nair, A. J. Kastin, and A. V. Schally, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 1376.
[225] N. P. Plotnikoff, A. J. Kastin, M. S. Anderson, and A. V. Schally, *Life Sci., Part I: Physiol. and Pharmacol.*, 1971, **10**, 1279.
[226] Sister A. Bower, MacE. Hadley, and V. J. Hruby, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 1185.

Although MSH release is under the predominant control of a release-inhibiting factor, a releasing factor (MSH-RF) has also been demonstrated: a metabolite of oxytocin was again suspected and the synthetic pentapeptide (135) was shown [227] to have MSH-RF activity, although identification of (135) with the natural factor is for the time being only compelling conjecture.

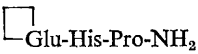
Assuming for the moment that the identification of MSH-RF as (135) is correct and making the further simplifying assumption that (133) or (134)

H-Cys-Tyr-Ile-Gln-Asn-OH

(135)

is the only MSH-RIF, the situation appears to be that the release of the peptide hormone MSH is controlled by two other peptides MSH-RF and MSH-RIF, each of which is formed by enzymic degradation of another peptide hormone. This is complicated enough, but there is yet more: pituitary MSH levels in the female rat are dependent on the oestrous cycle, possibly because the enzymic activity which generates MSH-RF fluctuates (that which generates MSH-RIF is apparently [223] constant)—presumably at the nod of yet another hormone.

Thyroid-stimulating Hormone Releasing Factor. The solution of the structure (136) of thyroid-stimulating hormone releasing factor (TSH-RF) and



(136)

preliminary related synthetic work were discussed in last year's Report. A considerable number of further syntheses of the hormone itself have been described in detail this year,[73, 228–232] and additional work on analogues has also appeared.[233, 234]

Endocrinologists have been quick to seize the opportunity offered by the availability of pure TSH-RF in quantity, and numerous biological studies using the synthetic material have already been published. Discussion of these is outside the scope of this Report but we note some of special interest, including the demonstration that normal women are more sensitive

[227] M. E. Celis, S. Taleisnik, and R. Walter, *Biochem. Biophys. Res. Comm.*, 1971, **45**, 564.
[228] A. C. Beyerman, P. Kranenburg, and J. L. M. Syrier, *Rec. Trav. chim.*, 1971, **90**, 791.
[229] J.-K. Chang, H. Sievertsson, C. Bogentoft, B. Currie, and K. Folkers, *J. Medicin. Chem.*, 1971, **14**, 481.
[230] J. Bøler, J.-K. Chang, F. Enzmann, and K. Folkers, *J. Medicin. Chem.*, 1971, **14**, 475.
[231] K. Inouye, N. Namba, and H. Otsuka, *Bull. Chem. Soc. Japan*, 1971, **44**, 1689.
[232] F. Enzmann, J. Bøler, K. Folkers, C. Y. Bowers, and A. V. Schally, *J. Medicin. Chem.*, 1971, **14**, 469.
[233] D. Gillessen, F. Piva, H. Steiner, and R. O. Studer, *Helv. Chim. Acta*, 1971, **54**, 1335.
[234] J.-K. Chang, H. Sievertsson, B. Currie, K. Folkers, and C. Y. Bowers, *J. Medicin. Chem.*, 1971, **14**, 484.

to TSH-RF than are normal men,[235] the finding [222] that TSH-RF also has prolactin-releasing activity, investigations of the metabolic fate of radioactive TSH-RF,[236] and experiments [237] on the infusion of TSH-RF into the anterior pituitary by way of the blood vessels which connect it to the hypothalamus: since this is a very efficient way of inducing TSH-release, direct support is provided for the idea that the releasing factor is formed in the hypothalamus and is carried to its site of action by the blood.

C. Oxytocin.—Oxytocin continues to be the subject of relentlessly patient synthetic alteration. Among the many new analogues reported this year (see Appendix A) are two which pose a conundrum: deamino-[4-threonine]-oxytocin and deamino-[4-threonine]-mesotocin.[238] [4-Threonine]-oxytocin has higher oxytocin-like activity than the parent hormone (see vol. 3 of these Reports, ch. 3, section 3E) and [4-threonine]-mesotocin is similar: since it is well known that removal of the amino-group from oxytocin enhances apparent biological activity, analogues in which both of these features were simultaneously present were investigated. The problem is that the deamino-[4-threonine] analogues were found to have low activity.

Elegant n.m.r. investigations [239–242] have led to a detailed proposal [241] for the conformation of oxytocin in solution. The essence of the proposal is that the cyclic part of the hormone contains a β-turn, and that the acyclic tail turns back to form a second β-turn. The compact folded molecule has one essentially hydrophobic side, lacking in notable chemical features, which is postulated to be involved in hormone-receptor binding: since most of the functionality present protrudes from the other side it is presumed that this is the side involved in the initiation of the biochemical events which eventually lead to the observed pharmacological effects. Many of the biological results obtained with synthetic modifications of oxytocin can be rationalized in terms of this conformation.[242, 243] For example, asparagine-5 occupies a central role in conformational stabilization since its N—H is involved in the formation of the ring β-turn, its C=O can hydrogen bond to the N—H of the tyrosine residue, and its side-chain stabilizes the folding back of the tail. The fact that 5-substituted analogues of oxytocin all have very low activities is therefore easily explained. The evidence for the hydrogen bond between the C=O of asparagine and the

[235] C. Y. Bowers, A. V. Schally, A. Kastin, A. Arimura, D. S. Schalch, G. Gual, E. Castineda, and K. Folkers, *J. Medicin. Chem.*, 1971, **14**, 477.
[236] T. W. Redding and A. V. Schally, *Endocrinology*, 1971, **89**, 1075.
[237] J. C. Porter, W. Vale, R. Burgus, R. S. Mical, and R. Guillemin, *Endocrinology*, 1971, **89**, 1054.
[238] M. Manning, E. J. Coy, and W. H. Sawyer, *Experientia*, 1971, **27**, 1372.
[239] L. F. Johnson, I. L. Schwartz, and R. Walter, *Proc. Nat. Acad. Sci., U.S.A.*, 1969, **64**, 1269.
[240] D. W. Urry, M. Ohnishi, and R. Walter, *Proc. Nat. Acad. Sci., U.S.A.*, 1970, **66**, 111.
[241] D. W. Urry and R. Walter, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 956.
[242] R. Walter in 'Structure–Activity Relationships of Protein and Polypeptide Hormones', ed. M. Margoulis and F. C. Greenwood, *Excerpta Medica*, 1971, p. 181.
[243] R. Walter, I. L. Schwartz, J. H. Darnell, and D. W. Urry, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 1355.

N—H of tyrosine is in fact a little uncertain in oxytocin itself but is unambiguous in the case of the deamino-analogue. It appears that this second transannular hydrogen bond is stronger in deamino-oxytocin than in the natural hormone: the higher potency of the former therefore finds a possible explanation in its greater conformational rigidity. Many other synthetic analogue activities can be similarly explained, and it is clear that this work is a fundamentally important step towards the understanding of peptide hormone action at the molecular level. Similar studies on lysine-vasopressin are also at a refined stage,[242, 244, 245] but whether conformational analysis will throw useful light on structure–action relations in other peptide hormones remains to be seen. The majority of the known peptide hormones are, unlike the neurohypophyseal group, acyclic and therefore presumably more conformationally mobile, possibly adopting their active conformations only on interaction with their receptors.

Synthetic analogues of oxytocin have been much used in studies of its *in vivo* inactivation: recent work [246–248] provides further support for the contention that the major initial step of the enzymic inactivation is proteolysis in the acyclic appendage. The recent demonstration that one definite and one possible oxytocin metabolite have potent pharmacological properties quite different from those of oxytocin itself (see p. 370) will no doubt stimulate systematic screening of peptide hormone partial sequences for various activities other than those of the parent.

D. Ribonuclease T_1.—Further discussion of the synthetic work in progress on ribonuclease T_1 has appeared,[249] and full details of the synthesis of a protected peptide (137) comprising residues (12—47) have been published.[250] The synthesis is outlined, in highly abbreviated form, in Scheme 63. Each

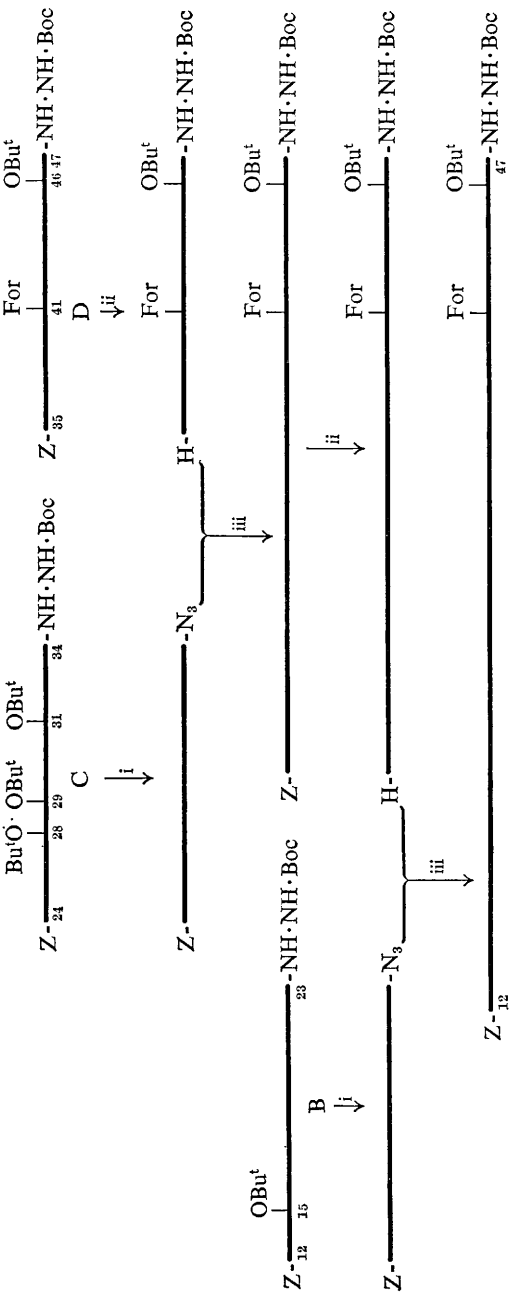
Z-Ser-Ser-Ser-Asp-Val-Ser-Thr-Ala-Gln-Ala-Ala-Gly-
12 **23**

-Tyr-Gln-Leu-His-Glu-Asp-Gly-Glu-Thr-Val-Gly-
24 **34**

For OBu^t
| |
-Ser-Asn-Ser-Tyr-Pro-His-Lys-Tyr-Asn-Asn-Tyr-Glu-Gly-NH·NH·Boc
35 **47**

(137)

[244] P. H. von Dreele, A. I. Brewster, H. A. Scheraga, M. F. Ferger, and V. du Vigneaud, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 1028.
[245] P. H. von Dreele, A. I. Brewster, F. A. Bovey, H. A. Scheraga, M. F. Ferger, and V. du Vigneaud, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 3088.
[246] R. Walter and H. Shlank, *Endocrinology*, 1971, **89**, 990.
[247] M. Loida, J. D. Glass, I. L. Schwartz, and R. Walter, *Endocrinology*, 1970, **88**, 633.
[248] R. Walter, H. Shlank, J. D. Glass, I. L. Schwartz, and T. D. Kerenyi, *Science*, 1971, **173**, 827.
[249] K. Hofmann, ref. 7, p. 130.
[250] J. Beacham, G. Dupuis, F. M. Finn, H. T. Storey, C. Yanaihara, N. Yanaihara, and K. Hofmann, *J. Amer. Chem. Soc.*, 1971, **93**, 5526.



(137)

Conditions: i, CF_3CO_2H, then HCl–DMF–Bu^tONO, − 10 °C; ii, Pd–H_2; iii, Et_3N–DMF, − 25 °C

Scheme 63

of the three protected peptide t-butoxycarbonylhydrazides B, C, and D were prepared by essentially stepwise procedures, mostly using trichlorophenyl or succinimido esters: acylamino-acid or acyldipeptide azides were used for the incorporation of (unprotected) hydroxyamino-acid residues. The synthesis of fragment B, which contains five unprotected hydroxy-side-chains and a glutamine residue, gave rise to solubility problems, and two of the hydrogenation steps were (successfully) performed in rather dilute solution in DMF at 70 °C. The protected hydrazides were purified as thoroughly as possible and were carefully examined for impurities: fragment C was obtained in high quality without special purification procedures, but the crude fragment D contained numerous impurities (some of which were possibly *O*-acyltyrosine derivatives) and partition chromatography was necessary for purification. It should be noted that the amino-acid analysis of the crude fragment D was 'good' and Hofmann and his colleagues stress that 'Amino-acid analysis in agreement with theory is a necessary, but not sufficient, criterion of purity'. Fragment B was too insoluble for sophisticated purification, but its state of purity was assessed by examination of deprotected derivatives.

The three fragments were condensed together in the order indicated by B + (C + D), using for each coupling Honzl–Rudinger azide reaction conditions, after exposure of the carboxyl component hydrazide by trifluoroacetic acid treatment and deprotection of the nucleophile by hydrogenolysis. The partially protected B–C–D product had numerous free side-chain carboxy-groups and could therefore be purified by ion-exchange chromatography. Hofmann *et al.* discuss suitable purity criteria for such fragment condensation products at some length, and point out that purity can be assessed reliably from the amino-acid analysis if attention is focused on 'diagnostic' residues which (*a*) are stable to acid hydrolysis and (*b*) are unique to one of the compounds involved in the fragment conjunction. If both partners in a fragment condensation contain a 'diagnostic' residue then the ratio between the diagnostic residues found on amino-acid analysis of the product provides a means of detecting contamination by either of the coupling components.* On this criterion and thin layer chromatography in two systems the partially protected B–C–D fragment appeared to be homogeneous.

E. Solid-phase Synthesis of High Molecular Weight Polypeptides.—The solid-phase approach is now firmly established as a satisfactory method for the rapid synthesis of oligopeptides of up to, say, 10—15 residues which can usually be purified with reasonable ease and be shown to be homogeneous by reliable criteria. This is far from true of the application of the

* Such an approach would obviously not detect coincidental contamination by equimolecular proportions of each component, and would give rise to misleading conclusions if both were present. In any case, contamination by peptides of very different size and composition is unlikely to present an intractable problem for detection or separation: much more serious is the problem of detecting contamination by slightly defective or damaged sequences of similar molecular weight and composition.

method to peptides and proteins of higher molecular weight. The principal limitations are, of course, (i) the presence in the synthetic product of 'error' and 'failure' sequences attributable to incomplete coupling, (ii) the presence of defects due to damage in the final stages by vigorous reagents, (iii) the inability of available methods to cope with the resolution of the complex mixtures produced, and (iv) the inadequacy of conventional purity criteria to differentiate reliably between very closely related macromolecules. However, some of these difficulties—especially (ii) and (iv)—also apply in varying measure to 'classical' syntheses of high molecular weight polypeptides and it is at this stage by no means clear that such approaches will be any more successful for protein synthesis than the solid-phase method.

Despite the difficulties, the year under review has seen several publications on applications of solid-phase synthesis to high molecular weight polypeptides: these are briefly discussed in alphabetical order below.

Acyl Carrier Protein. The acyl carrier protein (ACP) of *E. coli* is a 77-residue single-chain protein with a 4′-phosphopantetheine prosthetic group. No cysteine or tryptophan is present and the three residues at the *C*-terminus are not required for biological activity. Since the prosthetic group can be added to the apoprotein by means of an ACP synthetase which is specific for the purpose, residues 1—74 of ACP seemed a suitable simple objective for further evaluation of the use of the solid-phase method for protein synthesis. Preliminary reports of such a synthesis have appeared.[251] The procedure and tactics employed for assembling the peptide chain on the resin were essentially conventional. Removal of the peptide from the resin by means of hydrogen fluoride was not feasible as this reagent was found to cause complete inactivation of natural ACP apoprotein. Furthermore, 'it was observed that HF or HF-TFA mixtures gave incomplete removal of protecting groups', although the side-chain protection employed (benzyl esters and ethers; ε-benzyloxycarbonyl groups and ω-nitro-groups) was such that their removal by hydrogen fluoride would have been anticipated. Hydrogen bromide in trifluoroacetic acid (anisole and methionine as scavengers) was in fact used for scission of the peptide–resin link and the partially deprotected peptide was subjected to gel filtration, when some 5% of the applied material emerged in the expected position. Completion of the deprotection by hydrogenolysis, enzymatic introduction of the prosthetic group, and ion-exchange chromatography gave material which co-chromatographed with and was immunologically indistinguishable from natural ACP which had been deprived of its *C*-terminal tripeptide sequence. The synthetic material had some 30% biological activity when compared with the natural material. Since the activity of the natural material falls to 40% when it is similarly subjected to the conditions used for the deprotection and removal of the synthetic

[251] W. S. Hancock, D. J. Prescot, W. L. Nulty, J. Weintraub, P. R. Vagelos, and G. R. Marshall, *J. Amer. Chem. Soc.*, 1971, **93**, 1799; *Fed. Proc.*, 1971, **30**, 1273 Abs.

peptide from the resin, it would appear that the prospect for obtaining synthetic ACP of high activity by solid-phase synthesis is good if modifications involving less violent final steps prove applicable.

Basic Pancreatic Trypsin Inhibitor. Preliminary results of a solid-phase synthesis of bovine pancreatic trypsin inhibitor (58 residues, single chain, three disulphide bridges, nil tryptophan) have appeared.[252] Since it was found that the natural inhibitor could be recovered from solution in hydrogen fluoride without loss of activity,* this reagent (with anisole as scavenger) was used for deprotection and cleavage from the resin. Conversion to the *S*-sulphonate and gel filtration gave material which eluted at the same position as, and which was electrophoretically indistinguishable from, the hexa-*S*-sulphonate obtained from the natural peptide. Reduction of the synthetic *S*-sulphonate, further gel filtration and air oxidation gave a substance which had in two different assays *ca.* 30% of the trypsin-inhibitory activity of the natural inhibitor. Subjection of the natural inhibitor to the same reduction and reoxidation procedures regenerated only 40% of the activity. It might appear, therefore, that the solid-phase part of this experiment was reasonably satisfactory, and that the chance of ultimate success is fair if the purification can be improved (affinity chromatography is an obvious possibility). Unfortunately, trypsin-inhibitory activity is not an acceptable criterion of success, since the naturally occurring proteinase inhibitors are structurally rather diverse: it is possible, and some may think likely, that error and failure sequences may have some activity.

Human Growth Hormone. In 1969,[253] Li and his group reported a complete primary structure for human growth hormone comprising 188 residues in a single chain with two disulphide bridges and one tryptophan residue. A solid-phase synthesis of this sequence has now been outlined by Li and Yamashiro.[254] The assembly of the protected peptide on the resin followed the usual pattern, with some special features including the introduction of histidine residues using bis-t-butoxycarbonylhistidine and the protection of the tryptophan residue from oxidative degradation during acid-treatment steps by the addition of 2,5-dithiothreitol. Cleavage from the resin by means of hydrogen fluoride (anisole as scavenger), gel filtration, sodium–ammonia treatment, air oxidation, and further gel filtration gave material eluting 'as a single peak with a maximum close to the position of native human growth hormone'. The polymer thus prepared had *ca.* 10% growth-promoting activity, whereas the natural hormone which had been similarly abused by hydrogen fluoride and sodium in ammonia had *ca.* 35%

[252] K. Noda, S. Terada, N. Mitsuyasu, M. Waki, T. Kato, and N. Izumiya, *Naturwiss.*, 1971, **58**, 147.
[253] C. H. Li, J. S. Dixon, and W. K. Liu, *Arch. Biochem. Biophys.*, 1969, **133**, 70.
[254] C. H. Li and D. Yamashiro, *J. Amer. Chem. Soc.*, 1970, **92**, 7608.

* Quantitative recovery of activity was reported, which seems a little surprising: perhaps the relatively rigid structure which is presumably imposed by three disulphide bridges confers insensitivity to HF.

growth-promoting activity. Immunological testing was encouraging, since rabbit antiserum to human growth hormone did react with the synthetic preparation.

Shortly after the publication of this work, however, the sequence which had been the synthetic objective was shown to be grossly erroneous in several respects, the most serious being the mislocation of a 15-residue sequence incorporating the single tryptophan residue which in fact occupies position 85, not position 25.[255–257] The synthetic product obtained by Li and Yamashiro,[254] although possessing some biological activity, therefore could not have contained any material at all with the sequence of the natural hormone.

This abortive synthetic excursion is clearly a cautionary tale. In the first place the outlook for the eventual synthesis of growth hormone by this strategy is poor if experience with the erroneous sequence is any guide. More important, however, are the general warnings which arise. It is clear that greater caution than has hitherto been customary must be used in interpreting the results of biological activity determinations on solid-phase products: this example shows very clearly that the mere demonstration of some activity is only a necessary but not a sufficient criterion for synthetic success and does not necessarily provide good corroboration for a proposed structure. This point receives further emphasis from related work involving the solid-phase synthesis of some biologically active peptides obtained by limited proteolysis of growth hormone. The sections 1—21,[258] 81—121,[259] 122—153,[259] and 164—188[258] of Li's erroneous sequence were prepared and all were biologically active despite the fact that all four segments have subsequently been shown to contain sequence errors. It is also apparent from this work that immunological criteria are wholly unsatisfactory for establishing the success of a solid-phase synthesis. Antibodies to a native protein can be expected to react with another polypeptide if the antigenic determinant is present: an antigenic determinant can be quite a small part of a native macromolecule and it would be surprising indeed if an attempted solid-phase synthesis produced material containing none of the necessary determinants at all.

Lysozyme. Two brief abstracts[260] reporting a solid-phase attempt at egg-white lysozyme (129 residues, four disulphide bridges, and six tryptophan residues) are available. The relatively high tryptophan content and the known lability of lysozyme to hydrogen fluoride make this an especially

[255] H. D. Niall, *Nature New Biol.*, 1971, **230**, 90.
[256] H. D. Niall, M. L. Hogan, R. Sauer, I. Y. Rosenblum, and F. C. Greenwood, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 866.
[257] C. H. Li and J. S. Dixon, *Arch. Biochem. Biophys.*, 1971, **146**, 233.
[258] J. Bornstein, J. McD. Armstrong, F. Ng, B. M. Paddle, and L. Misconi, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 252.
[259] F. Chillemi and A. Pecile, *Experientia*, 1971, **27**, 385.
[260] J. J. Sharp, A. B. Robinson, and M. D. Kamen, *Fed. Proc.*, 1971, **30**, 1273 Abs.; L. E. Barstow, V. J. Hruby, A. B. Robinson, J. A. Rupley, J. J. Sharp, and T. Shimoda, *Fed. Proc.*, 1971, **30**, 1274 Abs.

difficult case for the present state of the art. Material with low lysozyme activity has been obtained, although the incorporation of tryptophan (several residues of which are involved in the catalytic process) was poor. A full evaluation must obviously await further work and the publication of experimental detail.

Parathyroid Hormone. Bovine parathyroid hormone is a single-chain polypeptide of 84 residues, but partial hydrolysis with dilute acid gives smaller peptides which retain high hormonal activity. A solid-phase synthesis of the biologically active *N*-terminal 1—34-tetratriacontapeptide (no disulphide bridges, one tryptophan residue) has recently been described.[261] The insoluble support used was a chloromethylated graft co-polymer of styrene and trifluorochloroethylene—no details are given of this new type of support for which 'unique advantages' are claimed, but publication of these elsewhere was promised. Otherwise the synthesis was mostly along well established lines: 2,4-dinitrophenyl protection was used for histidine side-chains and trifluoroacetylation for lysine ε-amino-functions. Mercaptoethanol was added at all acidic deprotection steps after addition of the tryptophan residue. The imidazole dinitrophenyl groups were removed by means of thiophenol before further partial deprotection and scission from the resin with hydrogen fluoride (anisole as scavenger): the trifluoroacetyl groups were removed with 1M piperidine–8M urea. Exposure of the native hormone to these deprotection conditions causes 'significant loss of biological activity'. Gel filtration followed by carboxymethylcellulose chromatography gave synthetic peptide which had essentially the expected amino-acid composition with the exception of histidine, the low content of which was attributed to incomplete deprotection. The biological activity of this material *in vitro* was somewhat greater than that of the fragments from partial hydrolysis of the native hormone. It seems that the *N*-terminal region of the hormone is the region responsible for biological activity—the synthetic work so far reported indicates that the *N*-terminal 34 residues are sufficient, but briefly mentioned unpublished work suggests that even shorter segments of this part of the molecule are enough. What then is the function of the remaining two-thirds or so of the natural hormone? A possible answer to this question is provided by the observation that the *in vitro* activity (relative to that of the whole natural hormone) of the synthetic 1—34 fragment was found to be considerably greater than the *in vivo* activity, suggesting that the function of the remainder of the native hormone is to retard metabolic inactivation in some fashion.

Ribonuclease A. Gutte and Merrifield have now described [262] in great detail the solid-phase synthesis of bovine pancreatic ribonuclease A which was briefly reported by them early in 1970 (see vol. 1 of these Reports).

[261] J. T. Potts, jun., G. W. Tregear, H. T. Keutmann, H. D. Niall, R. Sauer, L. J. Deftos, B. F. Dawson, M. L. Hogan, and G. D. Aurbach, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 63.

[262] B. Gutte and R. B. Merrifield, *J. Biol. Chem.*, 1971, **246**, 1922.

The purification has been much improved since the preliminary report, the crucial step being limited tryptic digestion of the synthetic protein. Natural ribonuclease is resistant to trypsin, but it was reasoned that defective sequences might have conformations which would not protect them from proteolysis. In the event this proved to be the case, since the specific activity of the synthetic enzyme increased 7.6-fold to *ca.* 60% on treatment with trypsin.* Fractional precipitation with ammonium sulphate after trypsin treatment gave material with a specific activity of 78% which was indistinguishable from the native enzyme by chromatographic and electrophoretic criteria. The amino-acid content, substrate specificity, and Michaelis constant of the natural and synthetic proteins were also in essential agreement, as were the peptide maps they gave after performic acid oxidation and tryptic digestion, and in addition to all this it was shown that rabbit bovine pancreatic ribonuclease antibodies neutralized the two enzymes with equal effectiveness.

Unfortunately, ribonuclease is obviously a special case in that it withstands chemical maltreatment well and regains its active structure after reduction and reoxidation. It follows that the kind of approach used in this first example will not prove so successful in less amenable cases. Even when the strategy seems feasible, the value of further excursions of this kind seems questionable, since it appears that currently used criteria of purity are quite inadequate for assessing the homogeneity of solid-phase products. This is well illustrated by the experience of Li's laboratory with growth hormone (see above), but perhaps even better by the very rigorous comparison made between their product and the native enzyme by Gutte and Merrifield. Their material is obviously impure, although it cannot be said whether the proportion of material with correct structure is greater or less than the apparent 78%—erroneous or damaged sequences might be inhibitory or active. Despite this evident contamination, a most careful comparison with the natural protein revealed a single spot on a peptide map as the sole significant difference. It seems then that further advances in practicable methods suitable for investigating the composition of solid-phase products are sorely needed. Until such time as these desiderata are attained it seems certain that the use of experiments on solid-phase products in protein chemistry will be unacceptable except in instances such as a demonstration that a solid-phase synthesis of a partial sequence of a protein gives active material, from which it can be unambiguously deduced that the omitted portion is inessential. Gutte and Merrifield have provided one such example. Samples were withdrawn from their ribonuclease synthesis after the addition of 99 and 104 residues, giving de-(21—25)-S-protein and S-protein respectively. Reduction and reoxidation of these

* This treatment also increased the total number of apparent units of enzymic activity, strongly suggesting the presence of erroneous sequences with inhibitory properties—yet another indication of the caution required in evaluating the biological activity of solid-phase products.

materials in the presence of natural or synthetic S-peptide gave ribonuclease S and de-(21—25)-ribonuclease S of approximately equal enzymatic activity. Even if the synthetic de-(21—25)-ribonuclease was grossly impure, it can still be confidently concluded that the sequence (21—25) is not required for biological activity since the entire sequence is an inconceivable contaminant.

Further discussion of the Merck group's semiclassical synthesis of ribonuclease S has appeared [263]—the highest specific activity mentioned in print to date is 40% [264] but no full experimental detail has yet been published.

Staphylococcal Nuclease. The work of Anfinsen's group on nuclease-T (see vol. 2, p. 177 and vol. 3, p. 257) and synthetic analogues of it continues.[265–267] The application of diverse purification techniques has culminated [265] in the isolation of semisynthetic nuclease-T with a specific activity of about 90% and other properties showing the 'essential identity' of the semisynthetic preparation with that obtained from the natural enzyme. A series of single-replacement analogues of P_2 with substitutions for aspartic acid-21, arginine-35, aspartic acid-40, and glutamic acid-43 have now been synthesized.[266] Results with these peptides confirm crystallographic evidence that the residues named are in the active site, since capacity to activate P_3 and produce nuclease-T activity was abolished in most of the analogues, although those which were conservative with respect to charge type were still able to bind to P_3 (see the Table).

Table[a] *Synthetic analogues of fragment P_2 of staphylococcal nuclease*

Peptide	*Does the peptide bind to P_3?*	*Does the peptide activate P_3?*
[Glu[21]]-(6—47)-P_2	Yes	No
[Asn[21]]-(6—47)-P_2	No	No
[Lys[35]]-(6—47)-P_2	Yes	No
[Cit[35]]-(6—47)-P_2	No	No
[Glu[40]]-(6—47)-P_2	Yes	No
[Asn[40]]-(6—47)-P_2	Yes	Slightly[b]
[Asp[43]]-(6—47)-P_2	Yes	No

[a] This Table is a continuation of Table 3, vol. 3, p. 259.
[b] As far as could be ascertained, the low activity of this analogue was intrinsic and not attributable to partial deamidation at residue 40.

An enzymically active complex is also formed on mixing the cyanogen bromide (99—149) and tryptic (1—126) fragments, which are separately inactive. This complex has a duplicated sequence between residues 99 and 126: the redundant portion appears to lie in the (99—149) component since on treatment of the active complex with trypsin in the presence of calcium

[263] R. Hirschmann, ref. 7, p. 138.
[264] R. Hirschmann and R. G. Denkewalter, *Naturwiss.*, 1970, **57**, 145.
[265] I. M. Chaiken, *J. Biol. Chem.*, 1971, **246**, 2948.
[266] I. M. Chaiken and C. B. Anfinsen, *J. Biol. Chem.*, 1971, **246**, 2285.
[267] C. B. Anfinsen, D. A. Ontjes, and J. M. Chaiken, ref. 7, p. 130.

ions and 3′,5′-thymidine phosphate (which protect native nuclease from tryptic degradation) only the (99—149) partner was attacked.[268] The first solid-phase synthetic explorations by means of modifications of the (99—149)-fragment of this system have been described.[269] Essentially the same procedures as used for P_2 and its analogues were employed, and as in their previous work, Anfinsen and his colleagues recognize the shortcomings of the solid-phase method and therefore confine themselves to seeking answers to questions which 'are essentially of the "yes-or-no" variety'. It appears that the (99—149)-fragment can be truncated by up to eighteen residues at its *N*-terminal end and by several residues at the other without destruction of the capacity to form an active complex on admixture with (1—126) from the natural protein. Also, the single tryptophan at position 140 can be replaced by phenylalanine without loss of enzymic activity.

F. Scotophobin.—It was reported in 1968 that extracts of the brains of rats which had been trained to avoid the dark could be used to induce the same behaviour in naive mice.[270] Evidence was obtained at that time indicating that the active principle was an oligopeptide, and this has now been confirmed [271] by its isolation ('scotophobin').* Brain material from 4000 trained rats yielded tiny amounts of the active peptide, which could not be detected in extracts obtained from untrained animals. Insufficient material for complete conventional wet sequence analysis was obtained, but fortunately amino-acid analysis, enzymic digestion, *N*-terminal analysis, and pyrolysis of a derivative on the mass-spectrometer probe gave sufficient information for its tentative formulation as a pentadecapeptide amide, but there was at that stage still ambiguity concerning the position and number of amide side-chains.[271, 272] Some of the possible sequences were therefore synthesized,[273, 274] and after elimination of two which were inactive, the sequence (138) was confirmed by a solid-phase preparation [273] of this structure, which gave material active in the behavioural bioassay.

Ser-Asp-Asn-Asn-Gln-Gln-Gly-Lys-Ser-Ala-Gln-Gln-Gly-Gly-Tyr-NH_2

(138)

[268] H. Taniuchi and C. B. Anfinsen, *J. Biol. Chem.*, 1971, **246**, 2291.
[269] I. Parikh, L. Corley, and C. B. Anfinsen, *J. Biol. Chem.*, 1971, **246**, 7392.
[270] G. Ungar, L. Galvan, and R. H. Clark, *Nature*, 1968, **217**, 1261.
[271] G. Ungar, D. M. Desiderio, and W. Parr, *Nature*, 1972, **238**, 198.
[271a] W. W. Stewart, *Nature*, 1972, **238**, 202.
[272] D. M. Desiderio, G. Ungar, and P. A. White, *Chem. Comm.*, 1971, 432.
[273] W. Parr and G. Holzer, *Z. physiol. Chem.*, 1971, **352**, 1043.
[274] A. Ali, J. H. R. Faesel, D. Sarantakis, D. Stevenson, and B. Weinstein, *Experientia*, 1971, **27**, 1138.

* It is hardly surprising that the suggestion that acquired information could be chemically transferred met with considerable scepticism. The much delayed long paper [271] on this work will not end the controversy, because it is accompanied [271a] by a detailed hypercritical dissection of practically every aspect.

G. Substance P.—Substance P, as its positively alchemical name suggests, has long been something of a mystery.[190] Forty years ago Gaddum found that a dried powder (P for Powder) obtained from, *inter alia*, brain extracts contained a substance with a wide range of pharmacological properties, including the ability to cause transient hypotension and stimulate salivary secretion on intravenous injection, and also to bring about contraction of smooth muscle preparations. Substance P is very widely distributed and has been the subject of much inconclusive research. The vague idea that it has some sort of generalized neural function seems to have established itself, but real progress in the area has been severely hampered by ignorance of the structure and the uncertainties which must necessarily accompany experiments using partially purified extracts.

The stage is now set for some of the long-standing questions about substance P to be answered since it has been isolated.[275] Its sequence (139)

Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met-NH_2

(139)

has been determined [276] and a solid-phase synthesis has been briefly outlined.[277] The synthesis was performed on a benzhydrylamine-type support (126) (*cf.* ref. 74). After purification the synthetic material was indistinguishable from natural substance P by chromatographic and electrophoretic criteria, and in four different bioassays the natural and synthetic substances P were equipotent.

4 Appendix A. A List of Syntheses Reported during 1971

A. Naturally Occurring Peptides, Proteins, Analogues, and Partial Sequences.—All syntheses are listed under the natural substance to which they are relevant.

Peptide	*Ref.*
Acyl carrier protein	
E. coli acyl carrier protein (see p. 376)	251
Adrenocorticotropins	
β-Corticotropin *N*-terminal decapeptide, with a ^{14}C label in the glycine residue	277*a*
[$Orn^{15,16,17,18}$]-β-corticotropin-(1—18)-octadecapeptide amide	277*b*
[D-Ser^{1},$Orn^{15,16,17,18}$]-β-corticotropin-(1—18)-octadecapeptide amide	277*b*
[$Orn^{11,15,16,17,18}$]-β-corticotropin-(1—18)-octadecapeptide amide	277*b*
[D-Ser^{1},$Orn^{11,15,16,17,18}$]-β-corticotropin-(1—18)-octadecapeptide amide	277*b*
[β-Ala^{1},Lys^{17}]-β-corticotropin-(1—17)-heptadecapeptide-4-amino-n-butylamide	200

[275] M. M. Chang and S. E. Leeman, *J. Biol. Chem.*, 1970, **245**, 4784.
[276] M. M. Chang, S. E. Leeman, and H. D. Niall, *Nature New Biol.*, 1971, **232**, 86.
[277] G. W. Tregear, H. D. Niall, J. T. Potts, jun., S. E. Leeman, and M. M. Chang, *Nature New Biol.*, 1971, **232**, 87.
[277a] K. Medzihradszky, H. S. Varga, S. Fittkau, and I. Marquardt, *Acta Chim. Acad. Sci. Hung.*, 1970, **65**, 449 (*Chem. Abs.*, 1971, **74**, 13 416t).
[277b] G. I. Tesser and J. T. Burs, *Rec. Trav. chim.*, 1971, **90**, 444.

Peptide	*Ref.*
[Lys(Dns)^{21}]-β-corticotropin-(1—24)-tetracosapeptide	277c
[D-Ser^{1},Lys^{17,18}]-β-corticotropin-(1—19)-nonadecapeptide and a series of *C*-terminal amide and ester derivatives thereof	278
[Leu^{7}]-β-corticotropin-(1—24)-tetracosapeptide	279
[MeTrp^{9}]-β-corticotropin-(1—24)-tetracosapeptide	279
[Aib^{1}]-β-corticotropin-(1—18)-octadecapeptide amide	280
[β-Ala^{1}]-β-corticotropin-(1—18)-octadecapeptide amide	280
β-Corticotropin-(6—24)-nonadecapeptide	281
β-Corticotropin-(4—23)-, -(5—23)-, and -(7—23)-peptide amides	281
Ovine β-corticotropin-(19—26)-heptapeptide	137
Angiotensin	
[Val^{5}]-angiotensin II	124; see also ref. 423
[Val^{5},MeAla^{7}]-angiotensin II	282
Derivatives of [Gly^{2},Val^{5}]-angiotensin II	283
Tetraglycyl-[Val^{5}]-angiotensin II-(5—8)-tetrapeptide	284
Diglycyl-[Ile^{5}]-angiotensin II-(3—8)-hexapeptide	285
[Ile^{5}]-angiotensin II	66
An extensive series of des-Asp^{1}-[Ile^{5}]-angiotensin II analogues; no synthetic details—biological activities only	286
A series of analogues of [Asn^{1},Ile^{5}]-angiotensin II with different alkyl side-chains on residue number 5 to investigate the importance of steric factors at this position	287
Other derivatives and analogues	288—291
Angiotensin-converting enzyme inhibitors	292
Angiotensinogen	
A series of analogues of the sequence around the renin-labile bond of angiotensinogen, some of which were inhibitory towards the angiotensin–renin system	293

[277c] R. Schwyzer and P. W. Schiller, *Helv. Chim. Acta.*, 1971, **54**, 897.
[278] M. Brugger, *Helv. Chim. Acta*, 1971, **54**, 1261.
[279] M. Fujino, C. Hatanaka, O. Nishimura, and S. Shinagawa, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 1075.
[280] K. Inouye, K. Watanabe, K. Namba, and H. Otsuka, *Bull. Chem. Soc. Japan*, 1970, **43**, 3873.
[281] M. Fujino, C. Hatanaka, and O. Nishimura, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 1066.
[282] R. H. Andreatta and H. A. Scheraga, *J. Medicin. Chem.*, 1971, **14**, 489.
[283] A. Pavars, G. Cipens, and Z. Auna, *Latv. P.S.R. Zinat. Akad. Vestis, Kim. Ser.*, 1970, 737 (*Chem. Abs.*, 1971, **74**, 142 332m).
[284] A. Pavars and G. Cipens, *Zhur. obshchei Khim.*, 1971, **41**, 467 (*Chem. Abs.*, 1971, **75**, 36 638b).
[285] S. A. Khan and K. M. Sivanandaiah, *Indian J. Chem.*, 1971, **9**, 184 (*Chem. Abs.*, 1971, **75**, 6314a).
[286] J. S. de Graaf, A. C. A. Jansen, K. E. T. Kerling, C. Schattenkerk, and E. Havinga, *Rec. Trav. chim.*, 1971, **90**, 301.
[287] E. C. Jorgensen, S. R. Rapaka, G. C. Windridge, and T. C. Lee, *J. Medicin. Chem.*, 1971, **14**, 899, 904.
[288] P. Ya. Romanovskii, G. Cipens, and Z. Auna, *Latv. P.S.R. Zinat. Akad. Vestis, Kim. Ser.*, 1971, 503 (*Chem. Abs.*, 1971, **75**, 118 597).
[289] J. A. Skorcz, *J. Medicin. Chem.*, 1971, **14**, 775.
[290] R. J. Weinkam and E. C. Jorgensen, *J. Amer. Chem. Soc.*, 1971, **93**, 7033.
[291] R. Paruszewski, *Roczniki Chem.*, 1971, **45**, 289 (*Chem. Abs.*, 1971, **75**, 64 244).
[292] M. A. Ondetti, N. J. Williams, E. F. Sabo, J. Pluscec, E. R. Weaver, and O. Kocy, *Biochemistry*, 1971, **10**, 4033.
[293] K. Shigezane, M. Muraki, T. Morikawa, and T. Mizoguchi, *Yakugaku Zasshi*, 1971, **91**, 987 (*Chem. Abs.*, 1971, **75**, 141 149r).

Peptide	*Ref.*
Bombesin	294
Bradykinin	295—297
[3 and 4-(*N*-2-aminoethylglycine)]-bradykinin; *i.e.* the derivative of [Gly^3]-bradykinin in which the carbonyl group of the peptide bond between Gly^3 and Gly^4 is reduced to a methylene group	298
1-Deamino-bradykinin	296
9-Decarboxy-bradykinin	296
1-Deamino, 9-decarboxy-bradykinin	296
[Har^1]-, [Har^9]-, and [Har^{1,9}]-bradykinins	295
[5-*threo*-β-phenylserine]-, [8-*threo*-β-phenylserine]-, and [5,8-*threo*-β-phenylserine]-bradykinins	299
[5-*erythro*-β-phenylserine]-, [8-*erythro*-β-phenylserine]-, and [5,8-*erythro*-β-phenylserine]-bradykinins	300
[5-β-(2-thienyl)alanine]-, [8-β-(2-thienyl)alanine]-, and [5,8-β-(2-thienyl)alanine]-bradykinins	301
[5,8-β-cyclohexylalanine]-bradykinin	297
Bradykinin-potentiating peptides	302
Caerulin	
Analogues	303, 303*a*
Calcitonin	
The human hormone	194
The salmon hormone	196
Cobrotoxin	
A partially protected hexapeptide fragment	304
Cytochromes	
Partially protected oligopeptides related to various cytochromes	305—309

[294] L. Bernardi, R. De Castiglione, O. Goffredo, and F. Angelucci, *Experientia*, 1971, **27**, 873.
[295] H. Arold and D. Stibenz, *J. prakt. Chem.*, 1970, **312**, 1161.
[296] W. H. Johnson, H. D. Law, and R. O. Studer, *J. Chem. Soc.* (*C*), 1971, 748.
[297] D. J. Schafer, G. T. Young, D. F. Elliot, and R. Wade, *J. Chem. Soc.* (*C*), 1971, 46.
[298] E. Atherton, H. D. Law, S. Moore, D. F. Elliot, and R. Wade, *J. Chem. Soc.* (*C*), 1971, 3393.
[299] H. Arold and S. Reissmann, *J. prakt. Chem.*, 1970, **312**, 1130.
[300] H. Arold and H. Feist, *J. prakt. Chem.*, 1970, **312**, 1145.
[301] F. W. Dunn and J. M. Stewart, *J. Medicin. Chem.*, 1971, **14**, 779.
[302] H. Kato and T. Suzuki, *Biochemistry*, 1971, **10**, 972.
[303] A. Anastasi, L. Bernardi, G. Bosisio, R. de Castiglione, O. Goffredo, G. Bertaccini, and V. Erspamer, ref. 7, p. 328.
[303a] A. Agosti, I. Orlandini, and R. de Castiglione, *Experientia*, 1971, **27**, 412.
[304] K. Noda, S. Terada, N. Mitsuyasu, M. Waki, T. Kato, and N. Izumiya, *Mem. Fac. Sci. Kyushi Univ. Ser. C.*, 1970, **7**, 189 (*Chem. Abs.*, 1971, **75**, 6297x).
[305] N. L. Alarkon, S. D. L'vova, and R. P. Evstigneeva, *Khim. prirod. Soedinenii*, 1971, **7**, 474 (*Chem. Abs.*, 1971, **75**, 141 153n).
[306] V. M. Kozhukhovskaya, S. D. L'vova, and R. P. Evstigneeva, *Khim. prirod. Soedinenii*, 1970, **6**, 599 (*Chem. Abs.*, 1971, **74**, 126 047s).
[307] S. D. L'vova, I. N. Sologubovskaya, and R. P. Evstigneeva, *Zhur. obshchei Khim.*, 1971, **41**, 216 (*Chem. Abs.*, 1971, **75**, 20 978b).
[308] S. D. L'vova, V. M. Kozhukhovskaya, T. A. Volkova, and R. P. Evstigneeva, *Zhur. obshchei Khim.*, 1971, **41**, 1393 (*Chem. Abs.*, 1971, **75**, 88 936s).
[309] S. D. L'vova, V. M. Kozhukhovskaya, M. N. Averin, and T. P. Evstigneeva, *Khim. prirod. Soedinenii*, 1971, **7**, 466 (*Chem. Abs.*, 1971, **75**, 152 065m).

Peptide	*Ref.*
Eledoisin	
Analogues	310—312
Follicle-stimulating hormone releasing factor: see luteinizing-hormone releasing factor (?)	
Ferredoxin	
Clostridal ferredoxin-(7—11)-pentapeptide and the analogue in which Cys^{8} and Cys^{11} are both replaced by Ser	313
Gastrin	
The *C*-terminal tetrapeptide and derivatives thereof	314—317
Glucagon	
Glucagon-(18—29)-dodecapeptide; passing mention	121
Glutathione	27, 52
Gramicidin S	
Linear decapeptide analogues	318, 319
Growth hormone	
Peptides and protected peptide fragments synthesized on the basis of the sequence since shown to be in error	254, 258, 259, 320, 321
Growth-hormone releasing factor (GH-RF: see p. 366)	
A decapeptide with the proposed sequence of porcine GH-RF	205
Haemoglobin	
Partial sequences of human haemoglobin β-chain	72, 205, 322—324
Insulin	
Ovine insulin A-chain, [Ala^{12}]-A-chain, [Glu^{5},$Ala^{12,18,21}$]-A-chain, and the formation of semisynthetic insulin analogues from these	57
A-arginyl bovine insulin (crystalline) from combination of modified natural A-chain with natural B-chain	163
An extensive series of analogues and partial sequences of bovine insulin designed to investigate the importance of arginine-B22	325, 326

[310] H. Niedrich, C. Berseck, and P. Oehme, ref. 7, p. 370.
[311] G. F. Zhukova, G. A. Ravdel, and L. A. Shchukina, *Zhur. obshchei Khim.*, 1970, **40**, 2573 (*Chem. Abs.*, 1971, **74**, 76 647q).
[312] R. Grupe and H. Niedrich, *J. prakt. Chem.*, 1970, **312**, 1087.
[313] H. Yajima, N. Shirai, and Y. Kiso, *Chem. and Pharm. Bull. (Japan)*, 1971, **19**, 1900.
[314] V. Vilka and G. Cipens, *Khim. prirod. Soedinenii*, 1971, **7**, 498 (*Chem. Abs.*, 1971, **75**, 141 161p).
[315] C. R. Sachatello, J. Sedwick, C. L. Moriarty, O. Grahl-Nielsen, and G. L. Tritsch, *Endocrinology*, 1971, **88**, 1300.
[316] M. Portelli, G. Renzi, and G. Borin, *Farmaco, Ed. Sci.*, 1971, **26**, 322 (*Chem. Abs.*, 1971, **75**, 209 832).
[317] O. V. Smirnov, L. K. Polevaya, and G. Cipens, *Khim. prirod. Soedinenii*, 1971, **7**, 94 (*Chem. Abs.*, 1971, **75**, 20 982y).
[318] S. Makisumi, S. Matsuura, M. Waki, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1971, **44**, 210.
[319] S. Makisumi, M. Waki, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1971, **44**, 143.
[320] C. H. Li and D. Chung, *Internat. J. Protein Res.*, 1971, **3**, 73; W. Datino and C. H. Li, *ibid.*, p. 81, 99; K. Kovacs, Y. Kovacs-Petres, and C. H. Li, *ibid.*, p. 93.
[321] K.-T. Wang and C. H. Li, *J. Org. Chem.*, 1971, **36**, 2419.
[322] V. P. Chernyshev, E. I. Filippovich, and R. P. Evstigneeva, *Zhur. obshchei Khim.*, 1971, **41**, 1620, 1623 (*Chem. Abs.*, 1971, **75**, 130 115a, 130 116b).
[323] P. G. Pietta and P. Cavallo, *Gazzetta*, 1970, **100**, 863.
[324] V. P. Chernyshev, E. I. Filippovich, R. P. Evstigneeva, and N. A. Preobrazhenskii, *Zhur. obshchei Khim.*, 1970, **40**, 2478 (*Chem. Abs.*, 1971, **75**, 20 985b).
[325] G. Weitzel, K. Eisele, H. Gugliemi, W. Stock, and R. Renner, *Z. physiol. Chem.*, 1971, **352**, 1735.
[326] G. Weitzel, U. Weber, J. Martin, and K. Eisele, *Z. physiol. Chem.*, 1971, **352**, 1005.

Peptide	*Ref.*
[B1-*p*-iodophenylalanine]-bovine insulin prepared semisynthetically from an appropriately protected des-Phe^{B1}-insulin	170
Insulin B-chain protected fragments, including symmetrical cystine derivatives	16, 32, 327—334, 334*a*
Des-(B28—B30)-bovine and porcine (human) insulins; resynthesis with materials from natural sources	78
Ovine (bovine, porcine) insulin B chain *S*-sulphonate	76
Human insulin B chain *S*-sulphonate	79
A series of fragments incorporating the A20—B19 disulphide bridge	331
Lysozyme (see p. 378)	260
[Ala^{76}]-Lysozyme-(64—82)-nonadecapeptide	335
Lysozyme: protected *N*-terminal sequences	88, 336, 337
Lipotropin	
Ovine β-lipotropin-(73—78)-hexapeptide	338
Luteinizing-hormone releasing factor (LH-RF: see p. 367)	
The decapeptide with the sequence proposed for LH-RF/FSH-RF	74, 211, 214
A series of pyroglutamyltri- and tetra-peptide amides, one of which (Glp-Tyr-Arg-Trp-NH_2) had LH-RF activity	220
Melanocyte stimulating hormones	
α-MSH	339
Semisynthetic [Lys^{10}]-human β-MSH	161
Melittin	
Melittin-(1—8)-, -(7—17)-, -(1—17)-, and -(14—17)-peptides	340
Melittins I and II (also some partial sequences)	341
Formyl melittins I and II	342
Comparison of natural and synthetic melittins	343

[327] P. G. Katsoyannis, J. Ginos, and M. Tilak, *J. Amer. Chem. Soc.*, 1971, **93**, 5866.
[328] P. G. Katsoyannis, M. Tilak, J. Ginos, and K. Suzuki, *J. Amer. Chem. Soc.*, 1971, **93**, 5862.
[329] P. G. Katsoyannis, M. Tilak, and K. Fukuda, *J. Amer. Chem. Soc.*, 1971, **93**, 5857.
[330] H. R. Bosshard, *Helv. Chim. Acta*, 1971, **54**, 951.
[331] B. Kamber, *Helv. Chim. Acta*, 1971, **54**, 398.
[332] R. G. Vdovina, A. K. Gracheva, M. G. Poznayak, and Yu. P. Shvachin, *Zhur. obshchei Khim.*, 1971, **41**, 239 (*Chem. Abs.*, 1971, **75**, 20 975y).
[333] M. I. Titov, L. I. Leont'eva, and Zh. D. Bespalova, *Zhur. obshchei Khim.*, 1971, **41**, 224 (*Chem. Abs.*, 1971, **75**, 20 989f).
[334] M. I. Titov, Zh.D. Bespalova, and L. I. Leont'eva, *Zhur. obshchei Khim.*, 1971, **41**, 928 (*Chem. Abs.*, 1971, **75**, 64 262t).
[334a] M. I. Titov and Z. A. Ardemasova, *Zhur. obshchei Khim.*, 1971, **41**, 1403 (*Chem. Abs.*, 1971, **75**, 98 811m).
[335] R. Arnon, E. Maron, M. Sela, and C. B. Anfinsen, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 1450.
[336] M. Guarneri, C. A. Benassi, R. Tomatis, A. Guggi, R. Ferroni, and R. Rocchi, *Gazzetta*, 1971, **101**, 363.
[337] M. Guaneri, C. A. Benassi, R. Ferroni, A. Guggi, R. Tomatis, and R. Rocchi, *Gazzetta*, 1971, **101**, 375.
[338] A. Loffett and J. Close, ref. 7, p. 92.
[339] J. Blake and C. H. Li, *Internat. J. Protein Res.*, 1971, **3**, 185.
[340] L. C. Dorman and L. D. Markley, *J. Medicin. Chem.*, 1971, **14**, 5.
[341] E. Schroder, K. Lübke, M. Lehmann, and I. Beetz, *Experientia*, 1971, **27**, 764.
[342] K. Lübke, S. Matthes, and G. Kloss, *Experientia*, 1971, **27**, 765.
[343] E. Habermann and G. Zeuner, *Naunyn-Schmiedebergs Arch. Pharmakol.*, 1971, **270**, 1 (*Chem. Abs.*, 1971, **75**, 49 574c).

Peptide	*Ref.*
Oxytocin	84, 343*a*
Oxytocin with a ^{14}C label in the glycine residue	344
[1-(L-2-hydroxy-3-mercaptopropionic acid)]-oxytocin	345
1-Deamino-[Thr^4]-oxytocin	238
1-Deamino-[Thr^4,Ile^8]-oxytocin; *i.e.* 1-Deamino-[Thr^4]-mesotocin	238
Oxytocin-(1—5)-pentapeptide	227
Analogues of 1-deamino-oxytocin in which the disulphide link is replaced by a single sulphur atom or by a methylene group	346
1-Deamino-carba^1-oxytocin	347
1-Deamino-carba^6-oxytocin	348
1-Deamino-dicarba-oxytocin	348
N-Acetyl-[Tyr(Me)^2]-oxytocin	349
[Phe(*p*-NO_2)^2]-oxytocin	350
[Tyr(3-NO_2)^2]-oxytocin	350
[D-Tyr(3-NO_2)^2]-oxytocin	350
[Arg^8]-oxytocin; *i.e.* [Arg^8]-vasotocin	351
[Arg^8]-oxytocinoic acid; *i.e.* [Arg^8]-vasotocinoic acid	351
1-Deamino-[Arg^8]-oxytocin; *i.e.* 1-deamino-[Arg^8]-vasotocin	351
1-Deamino-[Lys^8]-oxytocin; *i.e.* 1-deamino-[Lys^8]-vasotocin	352
[Leu^{2,4}]-oxytocin	353
[Pro^4]-oxytocin	354
[Pro^4,Ile^8]-oxytocin; *i.e.* [Pro^4]-mesotocin	354
[Pro^4,Gln^8]-oxytocin; *i.e.* [Pro^4]-glumitocin	354
[Lys^4,Ile^8]-oxytocin; *i.e.* [Lys^4]-mesotocin	354
S-Benzyl-oxytocin *C*-terminal tetrapeptide	84
Oxytocin-(1—6)-hexapeptide; *i.e.* tocinoic acid	226, 354*a*
Oxytocin-(1—6)-hexapeptide amide; *i.e.* tocinamide	355
1-Deamino-oxytocin-(1—6)-hexapeptide amide; *i.e.* 1-deamino-tocinamide	355
[1-(3-Methyl,3-mercaptobutyric acid)]-oxytocin; *i.e.* '[1-deamino-penicillamine]-oxytocin': deuteriated in position 1	356
Parathyroid hormone	
Bovine parathyroid hormone-(1—34)-tetratricontapeptide	261

[343a] O. S. Papsuevich and G. Cipens, *Zhur. obshchei Khim.*, 1970, **40**, 2768 (*Chem. Abs.*, 1971, **74**, 76 646p).
[344] R. Walter and R. T. Havran, *Experientia*, 1971, **27**, 645.
[345] D. B. Hope and M. Wälti, *Biochem. J.*, 1971, **125**, 909.
[346] K. Jošt and F. Šorm, *Coll. Czech. Chem. Comm.*, 1971, **36**, 2795.
[347] K. Jošt, *Coll. Czech. Chem. Comm.*, 1971, **36**, 218.
[348] K. Jošt and F. Šorm, *Coll. Czech. Chem. Comm.*, 1971, **36**, 234.
[349] K. Jošt and F. Šorm, *Coll. Czech. Chem. Comm.*, 1971, **36**, 297.
[350] Z. Auna, O. A. Kaurov, V. F. Martynov, and V. B. Morozov, *Zhur. obshchei Khim.*, 1971, **41**, 674 (*Chem. Abs.*, 1971, **75**, 49 560v).
[351] R. T. Havran, C. Meyer, I. L. Schwartz, and R. Walter, ref. 7, p. 161.
[352] M. Rimpler, *Annalen*, 1971, **745**, 8.
[353] V. J. Hruby and W. Y. Chan, *J. Medicin. Chem.*, 1971, **14**, 1050.
[354] M. Manning, J. W. M. Baxter, C. T. Wuu, V. Smart-Abbey, K. Morton, E. J. Coy, and W. H. Sawyer, *J. Medicin. Chem.*, 1971, **14**, 1143.
[354a] O. A. Kaurov, V. F. Martynov, Yu. V. Natochin, O. A. Popernatskii, and E. I. Shakhmatova, *Zhur. obshchei Khim.*, 1971, **41**, 213 (*Chem. Abs.*, 1971, **75**, 36 643z).
[355] V. J. Hruby, M. F. Ferger, and V. du Vigneaud, *J. Amer. Chem. Soc.*, 1971, **93**, 5539.
[356] V. du Vigneaud, J. D. Meador, M. F. Ferger, G. A. Allen, and A. T. Blomquist, *Bioorg. Chem.* 1971, **1**, 123.

Peptide	*Ref.*
Proinsulin	
Prolylprolyl-*N*-ε-benzyloxycarbonyl-lysyl-*N*-ω-nitroarginyl-bovine insulin A-chain; *i.e.* a partial proinsulin sequence obtained by semisynthetic means	165
Ribonuclease	
Ribonuclease T_1; further discussion of progress	249
Ribonuclease T_1; details for the 12—47 fragment	250
Ribonuclease S′; further discussion	263, 264
Ribonuclease A; full details	262
Ribonuclease *N*-terminal eicosapeptide (S peptide)	
[Ser^{12}]-S-peptide-(1—14)-tetradecapeptide	357
[12-β-(pyrazol-3-yl)-alanine]-S-peptide-(2—14)-and -(3—14)-peptides	357
[Nle^{13}]-S-peptide-(1—14)-, -(2—14)-, and -(3—14)-peptides	357
[Nle^{7}]-S-peptide-(1—14)-tetradecapeptide	357
[$Nle^{7,13}$]-S-peptide-(1—14)-tetradecapeptide	357
[Cpg^{8},Orn^{10}]-S-peptide*	358
Rubredoxin	
Rubredoxin: protected *N*-terminal decapeptide	359
Scotophobin (see p. 382)	126, 271, 273
Analogues	273, 274
Secretin	
Secretin *C*-terminal hexapeptide amide	126, 127
[$Orn^{12,14,18,21}$]-Secretin	360
Partially protected secretin-(18—27)-, -(12—27)- and -(7—11)-peptides	361—363
Staphylococcal nuclease	
Analogues of staphylococcal nuclease-(6—48)-peptide (P_2)	266, 267
Analogues and partial sequences of staphylococcal nuclease-(99—149)-peptide	269
Substance P	277
Thyroid-stimulating hormone releasing factor (TSH-RF)	73, 228—232
Tritiated TSH-RF	364
Numerous analogues of TSH-RF	233, 234
Tobacco mosaic virus (TMV) coat protein	
TMV coat protein: (98—102)-partially protected pentapeptide	365
Trypsin inhibitor	
Bovine basic pancreatic trypsin inhibitor (see p. 377)	252
Vasopressin	
Lysine-vasopressin	123
Lysine- and arginine-vasopressins with ^{14}C labels in their glycine residues	344

[357] K. Hofmann, R. Andreatta, F. M. Finn, J. Montibeller, G. Porcelli, and A. J. Quattrone, *Bioorg. Chem.*, 1971, **1**, 66.
[358] G. Borin, F. Marchiori, R. Rocchi, L. Moroder, and E. Scoffone, *Gazzetta*, 1971, **101**, 426.
[359] A. Ali and B. Weinstein, *J. Org. Chem.*, 1971, **36**, 3022.
[360] M. J. Smithers, ref. 7, p. 44.
[361] E. Wünsch and P. Thamm, *Chem. Ber.*, 1971, **104**, 2454.
[362] E. Wünsch, G. Wendlberger, and P. Thamm, *Chem. Ber.*, 1971, **104**, 2445.
[363] E. Wünsch, G. Wendlberger, and A. Högel, *Chem. Ber.*, 1971, **104**, 2430.
[364] T. W. Redding and A. V. Schally, *Internat. J. Appl. Radiation Isotopes*, 1970, **21**, 742.
[365] K. Brunfeldt and J. Halstrøm, *Acta Chem. Scand.*, 1970, **24**, 3013.

* Cpg = α-cyclopentylglycine.

Peptide	*Ref.*
1-Deamino-[Ile^3]-lysine-vasopressin; *i.e.* 1-deamino-[Lys^8]-vasotocin	352
[Hly^8]-vasopressin*	366
Arginine-vasopressin	367—369
Arginine-vasopressinoic acid	368
[Ile^3]-arginine-vasopressin; *i.e.* [Arg^8]-vasotocin	351
[Ile^3]-arginine-vasopressinoic acid; *i.e.* [Arg^8]-vasotocinoic acid	351
1-Deamino-[Ile^3]-arginine-vasopressin; *i.e.* 1-deamino-[Arg^8]-vasotocin	351
A partially protected nonapeptide amide with the sequence of arginine vasopressin	370
[Gln^8]-vasopressin	370*a*

B. Sequential Polypeptides

poly-[D-Glu(OBzl)-Glu(OBzl)]	371
poly-[Glu(OBu^t)-D-Glu(OBu^t)]	154
poly-[Glu(OBzl)-D-Glu(OBzl)]	154
poly-[D-Glu(OBzl)-Leu]	154
poly-(Ala-Gly)	372
poly-(Ala-Ala-Gly)	372
poly-(Ala-Gly-Gly)	372
poly-(Ala-Gly-Gly-Gly)	372
poly-(Ala-Gly-Lys)	373
poly-(Lys-Gly-Ala)	373
poly-(Lys-Ala-Ala)	157, 159
poly-(Gly-Lys-Gly)	374
poly-(Gly-Gly-Lys)	375
poly-(Lys-Lys-Gly)	374
poly-(Ala-Pro-Gly)	153
poly-(Gly-Pro-Ala)	158
poly-(Gly-Pro-Pro)	376
poly-(Pro-Ser-Gly)	377

[366] M. Bodanszky and G. Lindeberg, *J. Medicin. Chem.*, 1971, **14**, 1197.
[367] K. Lübke, ref. 7, p. 154.
[368] J. Meienhofer, A. Trzeciak, T. Dousa, O. Hechter, R. T. Havran, I. L. Schwartz, and R. Walter, ref. 7, p. 157.
[369] J. Meienhofer, A. Trzeciak, R. T. Havran, and R. Walter, *J. Amer. Chem. Soc.*, 1970, **92**, 7199.
[370] H. Wissmann and R. Geiger, *Angew. Chem. Internat. Edn.*, 1970, **9**, 908.
[370a] O. S. Papsuevich, G. Cipens, and A. Krikis, *Khim. prirod. Soedinenii*, 1971, **7**, 543 (*Chem. Abs.*, 1971, **75**, 141 152m).
[371] F. Heitz and G. Spach, *Macromolecules*, 1971, **4**, 429.
[372] A. Brack and G. Spach, *Compt. rend.*, 1970, **271**, *C*, 916.
[373] T. P. Chuvaeva, L. V. Morozova, V. A. Shibnev, and K. T. Poroshin, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1970, **13**(9), 28 (*Chem. Abs.*, 1971, **74**, 142 333n).
[374] A. V. Zakirov and V. G. Debabov, *Khim. prirod. Soedinenii*, 1970, **6**, 459 (*Chem. Abs*, 1971, **75**, 20 967x).
[375] V. K. Burichenko, V. G. Ovchinnikova, R. R. Dakhte, and K. T. Poroshin, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1971, **14**(3), 30 (*Chem. Abs.*, 1971, **75**, 88 941q).
[376] V. A. Shibnev, M. P. Finogenova, and K. T. Poroshin, *Doklady Akad. Nauk S.S.S.R.*, 1971, **198**, 862 (*Chem. Abs.*, 1971, **75**, 49 552u).
[377] V. A. Shibnev, Sh. Kh. Khahkov, M. P. Finogenova, and K. T. Poroshin, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1970, 2822 (*Chem. Abs.*, 1971, **75**, 6303w).

* Hly = homolysine.

Peptide	*Ref.*
poly-[Tyr(Me)-Glu-Ala-Gly]-[1-^{14}C]Gly-OEt	378
poly-(Glu-Tyr-Ala-Gly)-[1-^{14}C]Gly-OEt	379
poly-(Tyr-Glu-Gly-Gly)-Gly-OMe	380
poly-(Phe-Glu-Ala-Gly)-Gly-OMe	381
poly-(Phe-Glu-Ala-Gly)-[1-^{14}C]Gly-OEt	382
poly-(Tyr-Ala-Glu)	155, 156
Miscellaneous sequential polypeptides, intermediates in sequential polypeptide synthesis, *etc.*	158, 383—389
Repeating-sequence oligomers	
Derivatives of H-(Gly-Pro-Pro)$_n$-OH (n = 2—6)	390
Derivatives of H-(Ala-Pro-Gly)$_n$-OH (n = 2—6)	153

C. Miscellaneous Peptides

Leu-Ala-Gly-Val	36, 135, 146
Leu$_5$-Ala	391
Gly-Leu-Phe-Gly	119
Ala-Gly$_2$-Leu-Phe-Gly	119
Phe-Arg-Leu-Asp	171
Ile-Ala-Val-Gly	125
Chromophoric acyloligopeptide substrates for various proteolytic enzymes	392
A series of oligopeptide thrombin substrates	393, 394
Oligopeptide protocollagen hydroxylase substrates	395

[378] B. J. Johnson, *J. Medicin. Chem.*, 1971, **14**, 640.
[379] B. J. Johnson, *J. Medicin. Chem.*, 1971, **14**, 488.
[380] B. J. Johnson and E. G. Trask, *J. Medicin. Chem.*, 1971, **14**, 251.
[381] B. J. Johnson and F. Chen, *J. Pharm. Sci.*, 1971, **60**, 330.
[382] B. J. Johnson and C. Cheng, *J. Medicin. Chem.*, 1971, **14**, 1238.
[383] V. A. Shibnev, T. P. Chuvaeva, Sh. Kh. Khahkov, and K. T. Poroshin, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1970, 2566 (*Chem. Abs.*, 1971, **75**, 6310w).
[384] V. A. Shibnev, L. I. Mar'yash, V. S. Grechishko, and K. T. Poroshin, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1970, 2820 (*Chem. Abs.*, 1971, **75**, 6304x).
[385] J. Sperling and D. Elad, *J. Amer. Chem. Soc.*, 1971, **93**, 967.
[386] K. T. Poroshin, V. K. Buricenko, L. I. Mar'yash, and V. A. Shibnev, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1971, 1276 (*Chem. Abs.*, 1971, **75**, 77 253r).
[387] A. B. Zegel'man, I. M. Savranskaya, G. D. Bozhkova, and K. T. Poroshin, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1970, **13**(12), 28 (*Chem. Abs.*, 1971, **75**, 77 266x).
[388] A. B. Zegel'man, T. Yu. Yusupov, G. N. Demyanik, Z. Sharifova, and K. T. Poroshin, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1971, **14**(4), 31 (*Chem. Abs.*, 1971, **75**, 88 933p).
[389] V. K. Burichenko, N. I. Koryakina, and K. T. Poroshin, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1971, **14**(5), 16 (*Chem. Abs.*, 1971, **75**, 130 100s).
[390] V. A. Shibnev, K. T. Poroshin, and V. S. Grechishko, *Doklady Akad. Nauk S.S.S.R.*, 1971, **198**, 608 (*Chem. Abs.*, 1971, **75**, 64 247s).
[391] A. M. Tometsko, J. Tischio, and J. Garden, sec., *J. Pharm. Sci.*, 1970, **59**, 1655 (*Chem. Abs.*, 1971, **74**, 13 419w).
[392] E. Wünsch, G. Schönsteiner-Altmann, and E. Jäger, *Z. physiol. Chem.*, 1971, **352**, 1424; E. Wünsch, A. Högel, and E. Jäger, *ibid.*, p. 1553; E. Wünsch, E. Jäger and G. Schönsteiner-Altmann, *ibid.*, p. 1560, 1568.
[393] R. K. H. Liem, R. H. Andreatta, and H. A. Scheraga, *Arch. Biochem. Biophys.*, 1971, **147**, 201.
[394] R. H. Andreatta, R. K. H. Liem, and H. A. Scheraga, *Proc. Nat. Acad. Sci., U.S.A.*, 1971, **68**, 253.
[395] Y. Kikuchi, D. Fujimoto, and N. Tamiya, *Biochem. J.*, 1971, **124**, 695.

Peptide	*Ref.*
A series of oligopeptide pepsin substrates, each containing a residue of *p*-nitrophenylalanine	396
A decapeptide antamanide precursor containing a residue of *p*-azidophenylalanine	397
A series of oligolysines and their mono-ε-Dnp derivatives	398
Lysine oligomers	399
A series of t-butoxycarbonyl-oligo-isoleucine methyl esters	400, 401
Arginine oligomers	402
γ-Glutamyl peptides	403, 403*a*
Peptides containing arginine	404
Peptides containing β-cyanoalanine	405
Peptides containing ε-benzyloxycarbonyl-lysine	104
Peptides containing histidine	406
Peptides containing serine and glutamine	407
Peptides containing ornithine and proline	408
Peptides containing alanine and glutamic acid	409
Peptides containing β-lysine	410, 411
Peptides containing hydroxyproline	412, 413
Peptides containing 2-amino-3-sulphamoylpropionic acid	414

[396] K. Medzihradszky and J. S. Fruton, ref. 7, p. 285.
[397] T. Wieland, A. von Dungen, and C. Birr, *Annalen*, 1971, **752**, 109.
[398] Y. Stupp, W. E. Paul, and B. Benacerraf, *Immunology*, 1971, **21**, 575.
[399] G. P. Vlasov, V. R. Glushenkova, and T. N. Prokopyuk, *Zhur. obshchei Khim.*, 1971, **41**, 430 (*Chem. Abs.*, 1971, **75**, 20 969z).
[400] U. Widmer and G. P. Lorenzi, *Chimia* (*Switz.*), 1971, **25**, 236.
[401] C. Toniolo, *Biopolymers*, 1971, **10**, 1709.
[402] K. T. Poroshin, V. K. Burichenko, and G. F. Kasymova, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1970, **13**(8), 23 (*Chem. Abs.*, 1971, **75**, 20 965v).
[403] J. Meienhofer, P. M. Jacobs, H. A. Godwin, and I. H. Rosenberg, *J. Org. Chem.*, 1970, **35**, 4137.
[403a] D. E. Nitecki and J. W. Goodman, in 'Chemistry and Biochemistry of Amino-acids, Peptides, and Proteins', ed. B. Weinstein, vol. 1, p. 87.
[404] E. I. Vinogradova, Yu. B. Alakhov, V. M. Liplin, N. A. Aldanova, M. Yu. Feigina, Yu. B. Shvetsov, and L. A. Fonina, *Zhur. obshchei Khim.*, 1970, **40**, 1385 (*Chem. Abs.*, 1971, **74**, 54 165d).
[405] A. Michalik and B. Liberek, *Zeszyty Nauk., Mat., Fiz., Chem.*, 1970, **10**, 163 (*Chem. Abs.*, 1971, **74**, 13 418v).
[406] V. P. Chernyshev, E. I. Filippovich, R. P. Evstigneeva, and N. A. Preobiazhenskii, *Zhur. obshchei Khim.*, 1970, **40**, 1646 (*Chem. Abs.*, 1971, **75**, 6298y).
[407] N. A. Poddubnaya, G. N. Balandina, and O. A. Tambieva, *Zhur. obshchei Khim.*, 1971, **41**, 436 (*Chem. Abs.*, 1971, **75**, 20 967y).
[408] N. A. Poddubnaya and N. Ya. Krasnobrizhii, *Zhur. obshchei Khim.*, 1971, **41**, 441 (*Chem. Abs.*, 1971, **75**, 36 639c).
[409] C. Daicovicium, A. Balog, D. Breazu, L. Beu, and F. Gonczy, *Rev. Roumaine Chim.*, 1971, **16**, 751 (*Chem. Abs.*, 1971, **75**, 64 255t).
[410] L. I. Rostovtseva, E. A. Markelova, A. A. Kiryushkin, and A. S. Khokhlov, *Zhur. obshchei Khim.*, 1971, **41**, 1604; L. I. Rostovtseva, A. A. Kiryushkin, and A. S. Khokhlov, *Zhur. obshchei Khim.*, 1971, **41**, 1611 (*Chem. Abs.*, 1971, **75**, 130 099y, 130 117c).
[411] L. I. Rostovtseva, A. A. Kiryushkin, and A. S. Khokhlov, *Zhur. obshchei Khim.*, 1971, **41**, 1380 (*Chem. Abs.*, 1971, **75**, 88 923k).
[412] F. H. C. Stewart, *Austral. J. Chem.*, 1971, **24**, 1277.
[413] K. T. Poroshin, Sh. Kh. Khalikov, G. A. Pesina, and V. A. Shibnev, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1970, 1823 (*Chem. Abs.*, 1971, **75**, 6287u).
[414] B. Aleksiev, P. Nisanjan, S. Stoev, and V. Doseva, *Z. physiol. Chem.*, 1971, **352**, 1411.

Peptide	*Ref.*
Peptides containing a histidine analogue which incorporates a spin label	290
Peptides with tetrazole analogues at the *C*-termini	415, 416
Peptides containing *N*-(2-carboxamidoethyl)glycine	417, 418
Peptides containing *N*-(2-aminoethyl)glycine	298
Aza peptides	310, 312
Oxazoline peptides	415
Selenium-containing peptides	419
ε-(α-glutamyl)-, ε(γ-glutamyl)-, and ε(β-aspartyl)-lysine	420
Bis-cystine peptides	60
A series of cysteinyl-polyglycyl-cysteines	59
Antimitotic dipeptides	421
Three pentapeptides and a nonapeptide of unspecified relevance	422

5 Appendix B. A List of Some Useful New Synthetic Intermediates Described during 1971

As before, we warn that this list is selective and subjective. Compounds were taken to be new on the authority (whether explicit or implied) of the workers who described them: they are crystalline unless otherwise stated. One or two especially important intermediates which have been reported before have nevertheless been included in order to draw attention to some new point which we thought pertinent.

In this appendix, substitution into an aromatic protecting group is indicated by placing the substituent in brackets immediately after the symbol for the protecting group, *e.g.* Bzl(OMe) = methoxybenzyl.

Compound	*Ref.*
α-Aminobutyric acid (Abu)	
Abu-NH_2,HCl	234
α-Aminoisobutyric acid (Aib)	
Boc-Aib	280
α-Amino-β-ethylvaleric acid (Aev)	
Boc-Aev	287
α-Aminopimelic acid (Apm)	
Apm(OMe),HCl	346
Z-Apm(OMe),Dcha	346
Nps-Apm(OMe),Dcha	346
Alanine	
Ala-NH_2,HCl	234
Bpoc-Ala,Dcha	20
Z-Ala-SPy	83
Boc-Ala-SPy	83
Boc-Ala-OPy	84

[415] J. S. Morley, ref. 7, p. 361.
[416] Z. Grzonka, E. Rekowska, and B. Liberek, *Tetrahedron*, 1971, **27**, 2317.
[417] F. H. C. Stewart, *Austral. J. Chem.*, 1971, **24**, 1267.
[418] F. H. C. Stewart, *Austral. J. Chem.*, 1971, **24**, 1743.
[419] D. Theodoropoulis, ref. 7, p. 367.
[420] J. B. Caldwell, L. A. Holt, and B. Milligan, *Austral. J. Chem.*, 1971, **24**, 435.
[421] A. Jean and J. Anatol, *Bull. Soc. chim. France*, 1970, 3698.
[422] F. Sipos and D. W. Gaston, *Synthesis*, 1971, 321.

Compound	*Ref.*
β-Alanine	
Z-β-Ala-SPy	83
Arginine	
Arg *N*-carboxy-anhydride,HBr; impure but usable	100
Aoc-Arg(Tos)	66
Nps-Arg(NO_2)-ODnp	336
Nps-Arg(NO_2),Dcha	336
Nps-Arg(NO_2)-OPcp	337
Asparagine	
Asn *N*-carboxy-anhydride	100
Asn-OBzl(OMe),HCl	42
Asn-OBu^t	331
Bpoc-Asn: m.p. 40 °C higher than in a previous report	20
Boc-Asn-ONp: purification by counter-current distribution	50
Boc-Asn-OPy	84
Boc-Asn-SPy	83
Z-Asn-OTcp	250
Aspartic acid	
Asp *N*-carboxy-anhydride	100
Asp(OPic)-OPic,3HBr	423
Boc-Asp(OPic)	423
Z-Asp(OPic)	423
Z-Asp(OPic)-OTcp	423
Z-Asp(Bzl)-OPy	84
α-Cyclohexylglycine (Chg)	
Boc-Chg	287
α-Cyclopentylglycine (Cpg)	
Boc-Cpg	287
Cysteine	
Boc-Cys(Trt),Dcha	17
Boc-Cys(Trt)-ONp	16
Z-Cys(Trt); crystallization and physical constants of free acid	60
Cys(SBu^t)	54
Boc-Cys(SBu^t)	54
Nps-Cys(SBu^t),Dcha	54
Boc-Cys(Boc); useful detailed description	424
Boc-Cys(Boc)-ONSu	50
Cys(Boc),HCl	52
Cys(Acm)-OMe,HCl	53
Boc-Cys(Bzl),Dcha	17
Boc-Cys(Bzl)-ONSu: no characterization—used crude	17
Cys(Bzl)-OBzl(OMe),HCl	42
Z(OMe)-Cys(Bzl)-ONp	313
Cys(Bzl)-Bzl(NO_2),TosOH	355
Boc-Cys(Bzl)-OPy	84
Boc-Cys(Bzl)-SPy	83
Boc-Cys[Bzl(OMe)],Dcha	359
Nps-Cys[Bzl(OMe)],Dcha	336
Nps-Cys[Bzl(OMe)]-ONp	336

[423] R. Garner and G. T. Young, *J. Chem. Soc.* (*C*), 1971, 50.
[424] J. J. Ferraro, *Biochem. Prep.*, 1971, **13**, 39.

Compound	*Ref.*
Cystine	
Z-Cys,Dcha	
Z-Cys,Dcha	249
Boc-Cys	
Boc-Cys	52
Glutamic acid	
Boc-Glu(OPic)	423
Z-Glu(OPic)	423
Glu(OPic)-OPic,3HBr	423
Boc-Glu(OBu^t)-ONSu	357
Nps-Glu(OBzl)-ONp	336
Boc-Glu(OBzl)-OPy	84
Boc-Glu(OBzl)-SPy	83
Z-Glu(ONSu)-OBu^t	403
Glutamine	
Gln *N*-carboxy-anhydride	100
Gln-OBzl(OMe),HCl	42
Gln-NH·NH·Boc	250
Bpoc-Gln,Cha	20
Boc-Gln-SPy	83
Boc-Gln-OPy	84
Z-Gln-ONSu	250
Glycine	
Gly-OBzl(OMe),HCl	42
Gly-NH·NH·Z,HCl	294
Z-Gly-SPy	83
Boc-Gly-SPy	83
Boc-Gly-OPy	84
Histidine	
His *N*-carboxy-anhydride,HBr	100
His *N*-thiocarboxy-anhydride,HBr	101
His-OBzl,2HCl; improved procedure	44
Boc-His; alternative route	262
Boc-His(Z)	63
Z-His(Z); useful detailed description	425
Boc-His(Trt)	63, 67
His(Trt)-OMe,HCl	63
His(Trt),picrate	63
Boc-His(Bzh)	63, 67
His(Bzh)-OMe,HCl	63
His(Bzh),picrate	63
Z-His(Adoc)	65
Z-His(Boc); amorphous	62
His(Boc) *N*-carboxy-anhydride hydrochloride	62
His(Boc),2HCl	62
Boc-His(Boc),Dcha	254
Z-His(Dnp)	69
Z-His(Dnp)-ONSu	69
His(Dnp),2HBr	69

[425] A. A. Wieland, R. J. Albers, and D. F. DeTar, *Biochem. Prep.*, 1971, **13**, 28.

Compound	*Ref.*
His(Dnp) *N*-carboxy-anhydride hydrochloride	70
Boc-His(Piperidinocarbonyl)	67
Homoarginine	
Har(NO_2)-OMe,HCl	295
Har(NO_2)-OBzl(NO_2),HBr	295
Z-Har(NO_2)	295
Boc-Har(NO_2); crystalline	295
Nps-Har(NO_2),Dcha	295
Z-Har(NO_2)-ONp	295
Homolysine (Hly)	
Hly(Tos)	366
Z-Hly(Tos); oil	366
Z-Hly(Tos)-ONp	366
4-Hydroxyproline (4Hyp)	
A series of potentially useful *N*-nitroso-4Hyp derivatives (active esters, *etc.*)	412
4Hyp-OBzl(NO_2),HCl	412
4Hyp-OBzl(OMe),HCl	43
4Hyp-NH_2,HCl	412
Isoleucine	
Nps-Ile-ONSu	347
allo-Isoleucine	
Boc-*a*Ile,Dcha	287
Lysine	
Boc-Lys(Z)-OPcp	373
Lys(Z)-OPcp,HCl	373
Lys[Z(Cl)]	16, 32
Boc-Lys[Z(Cl)]	16
Boc-Lys[Z(Cl)]-ONp	16
Boc-Lys[Z(NO_2)]-ONp	16
Lys(Boc) *N*-carboxy-anhydride	100
D-Lys(Boc)	50
Boc-Lys(Boc)-ONSu	165
Boc-Lys(Tos)-ONSu	321
Z-Lys(Dns); amorphous	277*c*
Nps-Lys(Dnp),Dcha	398
ε-di-isopropylmethoxycarbonyl-lysine	32
Methionine	
Boc-Met,Dcha	280
Boc-Met-ONp	280
Boc-Met-OPcp	337
Boc-Met-SPy	83
Boc-Met-OPy	84
Z-Met-SPy	83
Norleucine	
Boc-Nle-ONSu	357
Z-Nle-ONSu	357
Ornithine	
Orn[Z(NO_2)]	426
Z(OMe)-Orn[Z(NO_2)],Dcha	426
Z(OMe)-Orn[Z(NO_2)]	426

[426] M. Ohno, K. Kuramizu, H. Ogawa, and N. Izumiya, *J. Amer. Chem. Soc.*, 1971, **93**, 5251.

Compound	*Ref.*
Orn(Z)-OMe,HCl	427
Z-Orn(Boc)-OMe	277*b*
Phenylalanine	
Bpoc-Phe,Dcha	20
Boc-Phe-SPy	83
Boc-Phe-OPy	84
Phenylalanine, substituted in the ring	
Boc-Phe(*p*-I)-OTcp	170
Boc-Phe(*p*-NO_2)	428
Boc-Phe(*p*-N_3)*	397, 428
Boc-Phe(*p*-N_3)-ONp	428
Boc-Phe(*p*-N_3)-ONSu	428
β-Phenylserine, *erythro*	
Erythro-β-phenylserine-OMe,HCl	300
Z-*erythro*-β-phenylserine	300
Boc-*erythro*-β-phenylserine	300
β-Phenylserine, *threo*	
Z-*threo*-β-phenylserine	299
Threo-β-phenylserine-NH·NH·Boc	299
Proline	
Pro-OBzl(OMe),HCl	43
Boc-Pro-SPy	84
Sarcosine	
Sar-OBzl(OMe),HCl	43
Serine	
Ser(Tms) *N*-carboxy-anhydride†	100
Bpoc-Ser	20
Bpoc-Ser(Bu^t),Dcha	20
Ser(Bzl): the resolution step improved in a large-scale preparation	51
Boc-Ser(Bzl),Cha	427
Boc-Ser(Bzl)-SPy	83
β-(2-Thienyl)alanine (Tha)	
Boc-Tha	301
Threonine	
Thr(Tms) *N*-carboxy-anhydride†	100
Thr-NH_2,HCl	234
Thr-OBzl,HCl	16
Boc-Thr-OBzl: oil	16
Bpoc-Thr,Cha	20
Tryptophan	
Boc-Trp-SPy	83
Boc-Trp-OPy	84
Tyrosine	
Tyr(Thp) *N*-carboxy-anhydride‡	100
Z-Tyr(Z)-ONSu	156

[427] M. Iwai, K. Nakajima, A. Uno, S. Hase, I. Takeuchi, and K. Okawa, *Bull. Chem. Soc. Japan*, 1970, **43**, 3246.
[428] R. Schwyzer and M. Caviezel, *Helv. Chim. Acta*, 1971, **54**, 1395.

* This derivative can be used in the synthesis of peptides which contain *p*-azidophenylalanine (a potential photoaffinity label) by the solid-phase method without any special precautions or procedures.
† Tms = Me_3Si.
‡ Thp = tetrahydropyranyl.

Compound	*Ref.*
Boc-Tyr(Bu^t)-OMe	347
Boc-Tyr(Bu^t),Dcha	347
Boc-Tyr(Bu^t)-ONSu	347
Boc-Tyr(Bzl)-ONp	280
Boc-Tyr(Bzl)-SPy	83
Boc-Tyr(Bzl)-ONSu	155
Valine	
Val-NH_2,HBr	363
Boc-Val-ONp; previously described as an oil	16
Bpoc-Val,Dcha	20
Z(OMe)-Val-ONp	313
Boc-Val-SPy	83
Miscellaneous derivatives	
Tetrazole analogues of some benzyloxycarbonylamino-acids	415, 429
A number of useful *N*-thiocarboxy-anhydrides	101
A series of *N*-(2-benzoyl-1-methylvinyl)amino-acid derivatives	121, 38

[429] Z. Grzonka and B. Liberek, *Tetrahedron*, 1971, **27**, 1783.

4

Peptides with Structural Features Not Typical of Proteins

BY J. S. DAVIES

1 Introduction

The aim this year has been to cover papers on this subject whose titles appeared in the 1971 volume of *Chemical Titles*. The coverage was augmented by scanning, over the same period, the sections on General Biochemistry, Microbial Biochemistry, and Synthesis of Amino-acids, Peptides, and Proteins in *Chemical Abstracts*. Although as many papers were retrieved as for Volume 3, fewer novel structures for peptide antibiotics were characterized sufficiently well for coverage in this compilation. Instead it has been a year dominated by the application of theoretical and spectroscopic techniques for the determination of the conformation of cyclic peptides and depsipeptides. Only a few years ago the dioxopiperazines presented almost the ultimate as models for such investigations of cyclic structures, but now cyclic nonapeptides and even larger structures such as alamethicin are undergoing detailed conformational analysis. However, the dioxopiperazines have not escaped detailed analysis and a number of interesting papers have appeared on their conformation. The occurrence of the epidithiadioxopiperazine ring system in a number of naturally occurring antibiotics has stimulated interest in the synthesis of these compounds, and although the methods reported for the synthesis of cyclic peptides in general have not produced novel approaches this year, it has been an active field with many important papers.

Of the new structures that have been reported a substantial number contain dehydroamino-acids, which could well be of biosynthetic significance, none less so than in the formation of the very interesting structure of nisin.

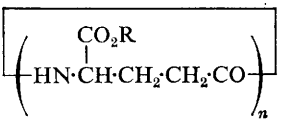
Less work than usual has appeared on the glycopeptide aspects of bacterial cell walls but this has been balanced by the increased number of glycosidic-asparaginyl linkages found in glycopeptides and glycoproteins from other sources. Epimerization, the conversion of penicillins into cephalosporins, and the formation of non-fused β-lactam derivatives dominate the interest in the penicillin–cephalosporin field. The synthesis of β-lactam derivatives without the fused thiazolidine ring could well be the beginning of interesting possibilities for the future. The first synthesis of a cyclol *via* an amide–amide interaction and the isolation of a naturally

occurring hydrazino-peptide for the first time were also achieved during the year.

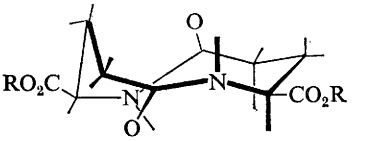
A comprehensive account of the earlier work on the isolation and chemistry of peptide antibiotics has appeared [1] but the review unfortunately contains very few papers published during the past five years. Most of the papers presented [2] at an antibiotics symposium in Quebec, Canada (May 1971), are very relevant to the subjects covered in this chapter. Recent developments in the biosynthesis [3] and synthesis [4] of peptide antibiotics have been reviewed and the significant contributions to cyclopeptide and cyclodepsipeptide chemistry made by the group led by the late Professor Shemyakin during the last decade have been recorded.[5] A comprehensive account of the chemistry and biochemistry of the mycobactins has also appeared.[6]

2 Cyclic Peptides

The application of various methods for determining conformation continues to dominate the interest in model cyclic peptides. Recent studies on the theoretical determination of the conformation of cyclic oligopeptides have been briefly discussed [7] as part of a more general review and the conformation of cyclic oligopeptides containing sarcosine has been discussed in a symposium lecture.[8] Circular dichroism (c.d.) studies [9] carried out on a series of cyclo-γ-oligoglutamic acids (1; n = 2, 3, or 4; R = H or Bu^t) show that the preferred conformation for the cyclodipeptides (n = 2) can be represented by structure (2) with both amide groups *trans*, and with



(1)



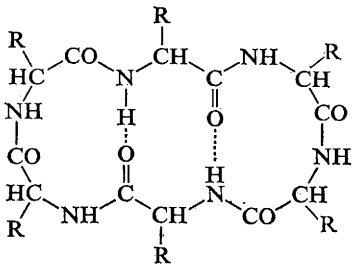
(2)

no indication of intramolecular hydrogen bonds. In the cyclo-tri- and -tetra-peptides there is more flexibility in the ring conformation but some cyclodipeptide character is still present. N.m.r. studies [10] at 220 MHz on *cyclo*-(Pro$_3$-) and derivatives have confirmed a previously calculated

[1] S. Sengupta, A. B. Banerjee, S. K. Majumder, and S. K. Bose, *J. Sci. Ind. Res., India*, 1970, **29**, 451.
[2] *Pure Appl. Chem.*, 1971, **28**, No. 4.
[3] D. Perlman and M. Bodanszky, *Ann. Rev. Biochem.*, 1971, **40**, 449; E. Katz, Ref. 2, p. 551.
[4] N. Izumiya, S. Matsuura, and K. Kuromizu, *J. Synthetic Org. Chem., Japan*, 1971, **29**, 1032.
[5] M. M. Shemyakin, *Pure Appl. Chem.*, 1971, **25**, 211.
[6] G. A. Snow, *Bacteriol. Rev.*, 1970, **34**, 99.
[7] H. A. Scheraga, *Chem. Rev.*, 1971, **72**, 195.
[8] J. Dale, *Pure Appl. Chem.*, 1971, **25**, 469.
[9] M. Kajtar, M. Hallosi, and G. Snatzke, *Tetrahedron*, 1971, **27**, 5659.
[10] C. M. Deber, D. A. Torchia, and E. R. Blout, *J. Amer. Chem. Soc.*, 1971, **93**, 4893.

Karplus-type analysis in that the conformation due to the planar *cis* peptide bonds requires the nitrogen atom to be out of the plane of the pyrrolidine carbon atoms in each proline residue.

Refinements to the conformations of a series of cyclohexapeptides previously investigated using n.m.r. spectroscopy have been deduced from c.d. studies.[11] It has been proposed that the conformation of *cyclo*-(Gly_5-Leu-) is an equilibrium between two forms, one of which is preferred in *cyclo*-(Gly_2-Leu-)$_2$. In *cyclo*-(Gly_5-Tyr-) and *cyclo*-(Gly_2-Tyr-Gly_2-His-) enhanced amide c.d. has been explained in terms of a coupling between both the tyrosyl and histidinyl chromophores and the amide bonds. N.m.r., i.r., and c.d. techniques have been applied in studies on cyclohexapeptides containing all possible combinations of L-alanyl and glycyl residues[12] and to all the diastereomeric cyclohexa-alanyls.[13] All the cyclohexapeptides revealed two groups of N—H signals: at 7.3—8.0 p.p.m. (two protons) and at 7.9—8.6 p.p.m. (four protons). The former has been assigned to protons involved in intramolecular transannular hydrogen-bonding as shown in (3)



(3)

(pleated sheet conformation) but complete randomness in the residues involved in hydrogen-bonding was revealed in the studies. Analysis of the coupling constants supports a structure for *cyclo*-(Gly-Ala_5-) as shown in Figure 1, with pseudo-axial methyl groups in positions 1, 3, 4, and 6. Changes in the polarity of the solvent[13] do not drastically affect the conformations of cyclohexa-alanyls. The solution conformations of *cyclo*-(Pro-Ser-Gly-Pro-Ser-Gly-) and its retro-isomers *cyclo*-(Ser-Pro-Gly-Ser-Pro-Gly-) have been investigated (220 MHz n.m.r.)[14, 15] and each

[11] S. M. Ziegler and C. A. Bush, *Biochemistry*, 1971, **10**, 1330.

[12] S. L. Portnova, V. V. Shilin, T. A. Balashova, J. Biernat, V. F. Bystrov, V. T. Ivanov, Yu. A. Ovchinnikov, *Tetrahedron Letters*, 1971, 3085; *Khim. prirod. Soedinenii*, 1971, **7**, 323, 339; V. T. Ivanov, G. A. Kogan, E. A. Meshcheryakova, V. V. Shilin, and Yu. A. Ovchinnikov, *ibid.*, p. 309.

[13] V. T. Ivanov, L. B. Senyavina, E. S. Efremov, V. V. Shilin, and Yu. A. Ovchinnikov, *Khim. prirod. Soedinenii*, 1971, **7**, 347; V. T. Ivanov, V. V. Shilin, G. A. Kogan, E. N. Meshcheryakova, L. B. Senyavina, E. S. Efremov, and Yu. A. Ovchinnikov, *Tetrahedron Letters*, 1971, 2841.

[14] D. A. Torchia, A. di Corato, S. C. K. Wong, C. M. Deber, and E. R. Blout, *J. Amer. Chem. Soc.*, 1972, **94**, 609.

[15] D. A. Torchia, S. C. K. Wong, C. M. Deber, and E. R. Blout, *J. Amer. Chem. Soc.*, 1972, **94**, 616.

shows two sets of resonances, indicating that the cyclic peptides rapidly interconvert between two conformations. Both conformations of the former cyclic peptide contain two glycyl–glycyl hydrogen bonds and two *trans* Gly-Pro peptide bonds, whereas in the retro-isomer the hydrogen bonds occur between the two seryl residues but with all peptide bonds *trans*. Changes in solvent affect the relative population of the conformations, *e.g.* in [$^2H_6$]DMSO the major conformation for the retro-isomer contains no intramolecular hydrogen bonds and has both Ser-Pro peptide bonds in the *cis* conformation.

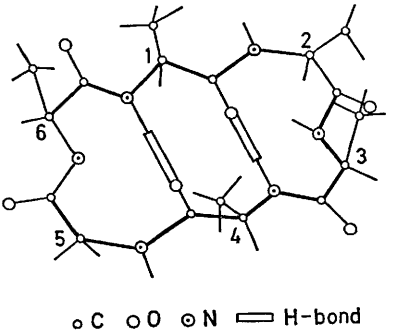
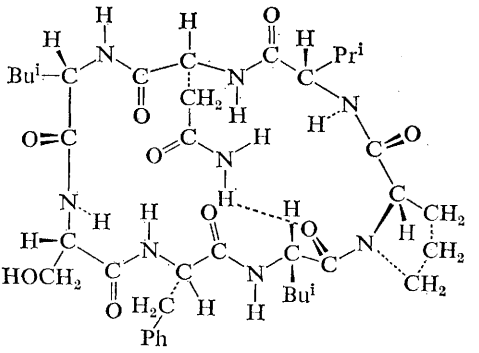


Figure 1

(Reproduced by permission from *Tetrahedron Letters*, 1971, 3085)

A detailed examination[16] of evolidine, *cyclo*-(Ser-Phe-Leu-Pro-Val-Asn-Leu-), again using 220 MHz n.m.r. spectroscopy, favours a conformation represented by (4) for the cycloheptapeptide. The temperature



(4)

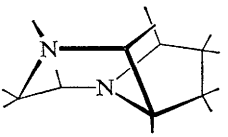
dependence of the chemical shifts of DMSO solutions shows that the peptide protons of the asparaginyl and phenylalanyl residues are shielded

[16] K. D. Kopple, *Biopolymers*, 1971, **10**, 1139.

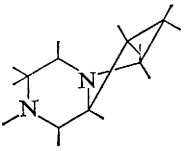
from the solvent. This case gives a timely note of warning against complete reliance on evidence from slow deuterium-exchange rates as a means of identifying internal hydrogen bonds, since only small changes in conformation would be required to make internal bonds exchange with solvent. In contrast to the cyclic peptides just described, a detailed study[17] of the all-L-cyclolinopeptide A, *cyclo*-(Leu-Ile-Ile-Leu-Val-Pro-Pro-Phe-Phe-), suggests that although four types of conformation are probable in solution (DMSO) none appears to possess intramolecular hydrogen-bonding, although two types of protons (five exposed and two internal) are implied from exchange reactions. The deductions for the cyclolinopeptide therefore appear to be in complete contrast to one of the proposals made for the conformation of antamanide (see later), a closely related case.

A pulse radiolysis study of cyclic peptides in solution has been reported.[18]

A. 2,5-Dioxopiperazines.—As in the case of the larger cyclic peptides, conformational aspects amount to an important part of the year's output of papers on dioxopiperazines. N.m.r. studies[19] have shown that the dioxopiperazine ring in *cyclo*-(Ala-Gly-) and *cyclo*-(Aib*-Gly-) is planar but that *cyclo*-(Pro-Gly-) exists as a boat conformation (5) which reverts to a planar ring system (6) in trifluoroacetic or deuteriotrifluoroacetic acids. In *cyclo*-(L-Pro-L-Pro-) the dioxopiperazine ring again reverts to the stable boat conformation as in (5) but in *cyclo*-(L-Pro-D-Pro-) the stable conformation is similar to (6). These deductions are also borne out in an *X*-ray



(5)



(6)

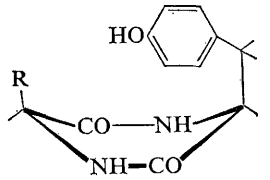
study[20] of *cyclo*-(L-Pro-L-Leu-) which shows that the molecule has a folded conformation with a dihedral angle of 143° between the two planar peptide units, with the leucyl side-chain fully extended. In this molecule the C^{β}-atom of the proline ring is out of the plane of the other four.

Calculation of optimum conformations together with ^1H n.m.r. and o.r.d. data favour[21] a non-planar conformation for *cyclo*-(L-Val-L-Val-) and a planar ring for the LD isomer.

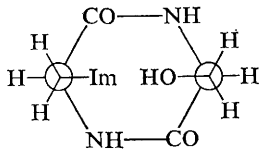
[17] F. Naider, E. Benedetti, and M. Goodman, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 1195; A. I. Brewster and F. A. Bovey, *ibid.*, p. 1199; A. E. Tonelli, *ibid.*, p. 1203.
[18] E. Hayon and M. Simic, *J. Amer. Chem. Soc.*, 1971, **93**, 6781.
[19] I. Z. Siemion, *Annalen*, 1971, **748**, 88.
[20] I. L. Karle, *J. Amer. Chem. Soc.*, 1972, **94**, 81.
[21] E. M. Popov, V. Z. Portnova, V. Z. Pletnev, V. T. Ivanov, P. V. Kostetskii, and Yu. A. Ovchinnikov, *Zhur. obshchei Khim.*, 1971, **41**, 420.

* Aib = α-aminoisobutyric acid.

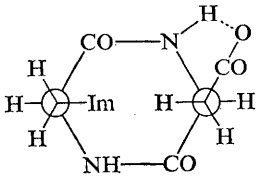
Previously reported n.m.r. studies on *cyclo*-(Gly-L-Tyr-) have been substantiated in a single-crystal *X*-ray study,[22] which shows that there is some buckling of the ring system towards a boat conformation with the aromatic ring above the dioxopiperazine ring as in (7). Quantum-mechanical calculations[23] support these results and also predict that the dioxopiperazine ring containing the side-chain methyl groups of valine would have the valine methyls stacked over the ring system. This stacking of side-chains above the ring is also favoured in the results obtained from a 220 MHz n.m.r. study[24] of *cyclo*-(L-His-L-Ser) (8) and *cyclo*-(L-Asp-L-His-) (9). *Cyclo*-(L-His-L-His) resembles[24] the cyclic anhydride of tyro-



(7)



(8)

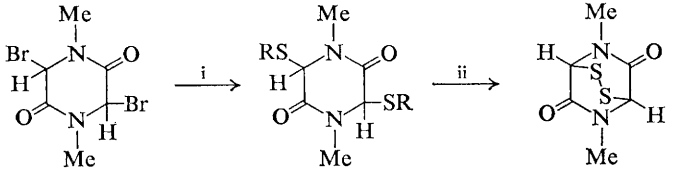


(9)

sine, suggesting that the imidazole rings are sharing the space over the dioxopiperazine ring. Minimum energy conformations derived from consistent force-field calculations[25] show that *cyclo*-(Gly-Gly-) and *cyclo*-(Ala-Ala-) have flexible structures with shallow non-planar conformations in the free state which average out to appear as planar except at low temperatures.

The occurrence of the epidithiadioxopiperazine ring system in the antibiotics gliotoxin, sporidesmin, aranotin, chaetocin, and, more recently, in verticillin A (see below) has initiated a number of approaches to the synthesis of this novel ring system. The direct introduction of the sulphur functions into the dioxopiperazine ring can be achieved[26] using a nucleophilic substitution reaction as outlined in Scheme 1. The reaction of 3,6-diethoxycarbonyl-2,5-dioxopiperazine with sulphur chloride S_2Cl_2 and

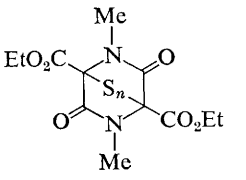
[22] L. E. Webb and C.-F. Lin, *J. Amer. Chem. Soc.*, 1971, **93**, 3818.
[23] J. Caillet, B. Pullman, and B. Maigret, *Biopolymers*, 1971, **10**, 221.
[24] Z. Kopple, K. D. Kopple, and C. A. Bush, *Tetrahedron Letters*, 1972, 483.
[25] S. Karplus and S. Lifson, *Biopolymers*, 1971, **10**, 1973.
[26] H. Poisel and U. Schmidt, *Angew. Chem. Internat. Edn.*, 1971, **10**, 130; *Chem. Ber.*, 1971, **104**, 1714.



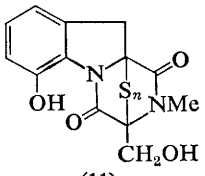
Conditions: i, MeOH–MeS$^-$ giving compound R = Me; ii; aqueous KI_3 starting from thiol R = H

Scheme 1

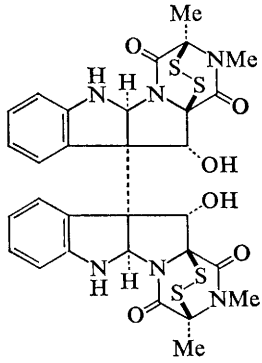
sodium hydride yields [27] a mixture of polysulphides (10; n = 1, 2, 3, or 4) but triphenylphosphine converts the tetrasulphide into the tri- and di-sulphides. This latter reagent has also been used [28] on compound (11; n = 2) and converts it into the monosulphide (11; n = 1). Degradative and spectroscopic evidence [29] on verticillin A is in agreement with structure (12) for this antibiotic, isolated from a *Verticillium* species. It is therefore similar to gliotoxin (13) and sporidesmin and is isomeric with chaetocin,



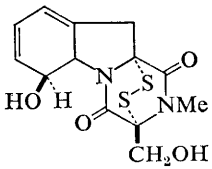
(10)



(11)



(12)

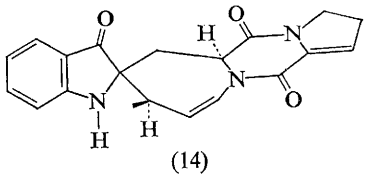


(13)

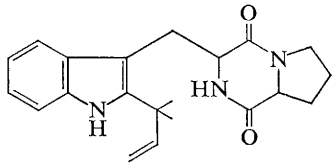
whose structure was recently determined by *X*-ray crystallography.[30] ^{14}C-Labelling studies [31] indicate that [3-^{14}C]phenylalanine is a more efficient precursor of gliotoxin (13) than is *m*-[2-^{14}C]tyrosine.

[27] T. Hino and T. Sato, *Tetrahedron Letters*, 1971, 3127.
[28] S. Safe and A. Taylor, *J. Chem. Soc.* (*C*), 1971, 1189.
[29] H. Minato, M. Matsumoto, and T. Katayama, *Chem. Comm.*, 1971, 44.
[30] D. Hauser, H. P. Weber, and H. D. Sigg, *Helv. Chim. Acta*, 1970, **53**, 1061.
[31] D. R. Brannon, J. A. Mabe, B. B. Molloy, and W. A. Day, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 588.

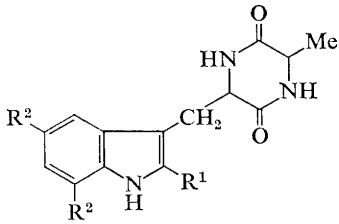
Austamide, a toxic metabolite from *Aspergillus ustus*, has been shown [32] to have the structure (14), and a minor component also extracted has been identified as (15) which is identical with the previously reported [33] deoxybrevianamide E. Ozonolysis and hydrolysis of echinulin (16) give L-aspartic acid, thus confirming that the echunin moiety in the molecule must have the L-configuration.[34] *cyclo*-L-Cystine (17) has been synthesized [35] for the first time by converting suitably protected *S*-acetamidomethylcysteine peptides into cystine by reaction with iodine.



(14)

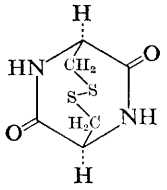


(15)



$R^1 = -CMe_2-CH=CH_2$
$R^2 = -CH_2-CH=CMe_2$

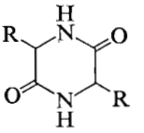
(16)



(17)

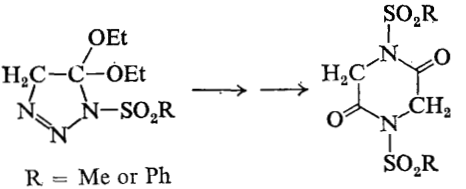
Cyclization of the appropriate cystine dipeptide ester followed by the reduction of the disulphide links with mercaptoethanol has yielded [36] *cyclo*-(Cys-Ser-), *cyclo*-(Cys-Asp-), and *cyclo*-(Cys-Trp-). Kinetic studies using *p*-nitrophenyl acetate indicate that the thiol groups of the cysteines in these systems have a greater nucleophilic reactivity than is expected on the basis of pK_{S-H} values. An investigation of the kinetics of hydrolysis of a series of dioxopiperazine derivatives, (18)—(20), has been reported.[37] In hydrochloric acid solution at pH 1—2 the cycloserine dimer (18) rapidly establishes an equilibrium with cycloserine monomer.

[32] P. S. Steyn, *Tetrahedron Letters*, 1971, 3331.
[33] A. J. Birch and J. J. Wright, *Tetrahedron*, 1970, **26**, 2329.
[34] R. Nakashima and G. P. Slater, *Tetrahedron Letters*, 1971, 2649.
[35] B. Kamber, *Helv. Chim. Acta*, 1971, **54**, 927.
[36] V. Rossbach, F. Schneider, and H. Wenck, *Z. Naturforsch.*, 1971, **26b**, 1144.
[37] C. H. Stammer and F. O. Lassen, *J. Org. Chem.*, 1971, **36**, 2631.



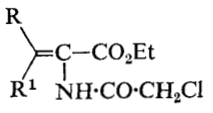
(18) R = CH_2OH
(19) R = $CH_2 \cdot ONH_2$
(20) R = Me

Two rather unconventional and specialized methods for preparing dioxopiperazines have been reported. The triazine intermediate (21) rearranges [38] at temperatures above 20 °C to give a dioxopiperazine acetal which can be hydrolysed to (22). An ethanolic solution of (23), on saturation with dry ammonia, gives [39] the dioxopiperazine (24). In a more conventional manner, albeit an unlikely pathway, the dioxopiperazine (25) has been formed [40] by treating the *p*-nitrophenyl ester of benzyloxycarbonylglycyl-proline with sodium bicarbonate or sodium carbonate solution.

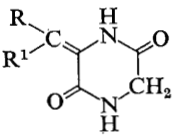


R = Me or Ph
(21)

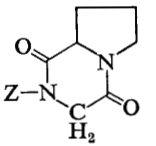
(22)



(23)



(24)



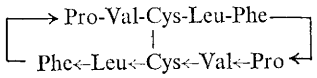
(25)

Various mechanisms have been proposed [41] for the fragmentation of 3,6-dibenzyl-2,5-dioxopiperazines in the mass spectrometer and new polymers have been reported [42] to be formed in the polymerization of 3,6-bifunctional-2,5-dioxopiperazines.

B. Gramicidins.—Interest continues in the conformation of gramicidin S with all the methods up to date confirming the well-established Hodgkin–Oughton–Schwyzer model (Vol. 3, p. 282). Thus calculations [43] based on

[38] G. J. Koves and I. G. Csizmadia, *Tetrahedron Letters*, 1971, 2599.
[39] C. Shin, M. Fujii, and J. Yoshimura, *Tetrahedron Letters*, 1971, 2499.
[40] G. Lucente, G. Fiorentini, and D. Rossi, *Gazzetta*, 1971, **101**, 109.
[41] K. Jankowski and L. Varfalvy, *Canad. J. Chem.*, 1971, **49**, 1583.
[42] U. Haberthuer and H. G. Elias, *Makromol. Chem.*, 1971, **144**, 183, 193, 213; V. Crescenzi, A. Ciana, V. Giancotti, E. Russo, L. Salvestrini, and L. Ciceri, *ibid.*, 1971, **141**, 199.
[43] P. De-Santis and A. M. Liquori, *Biopolymers*, 1971, **10**, 699.

available experimental evidence and computed c.d. properties [44] are again in agreement with a conformation containing four hydrogen bonds and a dyad axis relating the chemically equivalent halves of the molecule. Analogues of gramicidin S, namely [$Cys^{2,2'}$]gramicidin S (26) and [(*S*-acetamidomethyl-$Cys)^{2,2'}$] gramicidin S have been used to study [45] the chirality of the cystine disulphide group. The shielding of the valine peptide protons in (26) (from 7.20 to 7.00 p.p.m.) has been interpreted as



(26)

support for the *P*-helical chirality for the sulphur bridge. Attempts to correlate conformation with biological activity in gramicidin S analogues have also been reported.[46] A detailed o.r.d. study [47] of the copper complexes of gramicidin S establishes that the δ-amino-groups of the ornithine residues participate in complex formation.

Very impressive results with high yields (48—57% overall) of biologically active material have been reported [48] in the solid-phase synthesis of gramicidin S using the azide method (85% yield) for cyclization between leucyl and D-phenylalanyl residues. Quantitative detachment of the linear protected decapeptide from the polymer was achieved by hydrazinolysis. [D-$Ala^{4,4'}$]Gramicidin S and [D-Ala^{4}]cyclosemi-gramicidin S have been synthesized,[49] but only the former showed antibacterial activity, similar to that of gramicidin S. Thus the D-phenylalanine residues can be replaced without loss of activity. Linear decepeptide analogues with gramicidin S sequences and with a series of N- and C-terminal derivative groups have been synthesized.[50] Although o.r.d. measurements suggest that these linear peptides adopt certain conformational characteristics, their antibacterial activity does not match that of gramicidin S.

The elongation and cyclization of enzyme thioester-bound intermediates, recently proposed [51] as a mechanism for the biosynthesis of gramicidin S, included a hypothetical termination reaction involving an antiparallel doubling reaction between two carboxy-activated pentapeptide units. An

[44] P. M. Bayley, *Biochem. J.*, 1971, **125**, 90P.
[45] U. Ludescher and R. Schwyzer, *Helv. Chim. Acta*, 1971, **54**, 1637.
[46] T. Kato, M. Waki, S. Matsuura, and N. Izumiya, *J. Biochem.* (*Japan*), 1970, **68**, 751.
[47] N. A. Poddubnaya and N. Y. Krasnobrizhii, *Zhur. obshchei Khim.*, 1971, **41**, 46.
[48] M. Ohno, K. Kuromizu, H. Ogawa, and N. Izumiya, *J. Amer. Chem. Soc.*, 1971, **93**, 5251.
[49] S. Lee, R. Ohkawa, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1971, **44**, 158.
[50] S. Makisumi, M. Waki, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1971, **44**, 143; S. Makisumi, S. Matsuura, M. Waki, and N. Izumiya, *ibid.*, h. 210.
[51] W. Gevers, H. Kleinkauf, and F. Lipmann, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **63**, 1335; R. Roskoski, G. Ryan, H. Kleinkauf, W. Gevers, and F. Lipmann, *Arch. Biochem. Biophys.*, 1971, **143**, 485.

alternative mechanism now proposed [52] suggests that the two pentapeptide chains attached to the same enzyme cyclize *via* an intra- rather than an inter-molecular reaction. ^{14}C-Labelling studies [53] have confirmed the role of pantotheine as a transmitter of the growing peptide chain to thioester-linked amino-acids in both gramicidin S and tyrocidin biosynthesis, as summarized in Figure 2. A close link is therefore established between the biosynthesis of these compounds and Lynen's theory for fatty-acid biosynthesis.

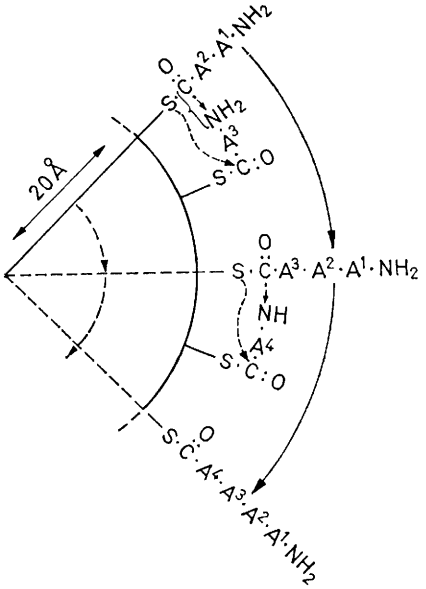


Figure 2 *A schematic representation of transpeptidation and transthiolation* (Reproduced by permission from *Science*, 1971, **173**, 875)

Studies on the interaction of gramicidin S (and polymyxin B) with membranes,[54] and on the production of gramicidin S in batch and continuous culture,[55] have been reported.

A series of helical structures have been proposed [56] for gramicidin A (27), which has alternating D and L residues. Because of the alternating directions of the carbonyl groups, head-to-head hydrogen-bonded dimers can be formed with the capacity to function as transmembrane channels.

[52] E. Stoll, O. Froyshov, H. Holm, T. L. Zimmer, and S. G. Laland, *F.E.B.S. Letters*, 1970, **11**, 348.
[53] H. Kleinkauf, R. Roskoski, and F. Lipmann, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2069; F. Lipmann, *Science*, 1971, **173**, 875.
[54] W. Pache, D. Chapman, and R. Hillaby, *Biochim. Biophys. Acta*, 1972, **255**, 358.
[55] H. W. Blanch and P. L. Rogers, *Biotechnol. and Bioeng.*, 1971, **13**, 843.
[56] D. W. Urry, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 672; D. W. Urry, M. C. Goodall, J. D. Glickson, and D. F. Mayers, *ibid.*, p. 1908.

HCO-L-Val-Gly-L-Ala-D-Leu-L-Ala-D-Val-L-Val-D-Val-L-Trp-D-Leu-L-Trp-D-Leu-L-Trp-D-Leu-L-Trp-NH·CH_2·CH_2·OH

(27)

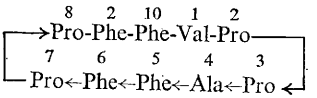
The chemistry and conformational integrity of the malonyl dimer, *NN'*-dideformyl-gramicidin A malonamide, as verified [56] by n.m.r. and lipid-layer studies, lend support to the head-to-head association.

C. Tyrocidins.—Further work [57] on the cell-free enzyme systems of *Bacillus brevis* ATCC 8185 has achieved the separation of the larger molecular weight component (*cf.* Vol. 3, p. 284) into two further sub-units with M.W. 335 000 and 237 000. These two sub-units are believed to be responsible for activating L-proline, L- and D-tryptophan, and L- and D-phenylalanine, and also for the formation of the sequence L-Pro → L-Trp (or L-Phe → D-Trp) in the biosynthesis of tyrocidins.

The mode of self-association of tyrocidin B has been studied [58] using a two-species plot technique in an ultracentrifuge; the results suggest that the process may be closely analogous to the formation of micelles by detergents.

D. Alamethicin.—From o.r.d. and c.d. measurements it has been deduced [59] that alamethicin exists as a monomer in 100% ethanol and that this monomer contains up to 40% of the peptide residues in a helical conformation. N.m.r. studies [60] of the interaction between alamethicin and phospholipids show that the interaction is largely hydrophobic involving the lipid chains. A mass spectrometric determination of the amino-acid sequence of alamethicin has been reported [61] but the details were not available in time for this compilation.

E. Peptides from *Amanita phalloides*.—More attention this year has been focused on the antitoxic cyclodecapeptide antamanide (28) rather than on



(28)

the other toxic principles extracted from this mushroom species. However, the conformation of antamanide has become a controversial subject in that two research groups have come to different conclusions about its structure

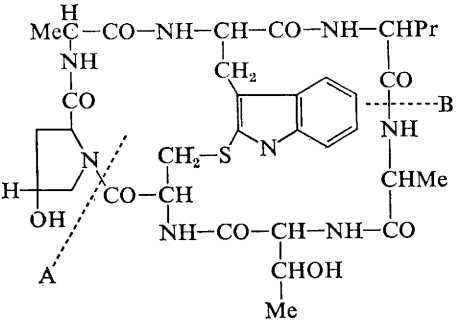
[57] K. Fujiwaka, Y. Sakamoto, and K. Kurahashi, *J. Biochem.* (*Japan*), 1971, **69**, 869; M. Kambe, Y. Sakamoto, and K. Kurahashi, *ibid.*, p. 1131.
[58] S. L. Laiken, M. P. Printz, and L. C. Craig, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 595; R. C. Williams, D. A. Yphantis, and L. C. Craig, *Biochemistry*, 1972, **11**, 70.
[59] A. I. McMullen and J. A. Stirrup, *Biochim. Biophys. Acta*, 1971, **241**, 807; A. I. McMullen, D. I. Marlborough, and P. M. Bayley, *F.E.B.S. Letters*, 1971, **16**, 278.
[60] H. Hauser, E. G. Finer, and D. Chapman, *J. Mol. Biol.*, 1970, **53**, 419.
[61] Yu. A. Ovchinnikov, A. A. Kiryushkin, and I. V. Kozhevnikova, *Zhur. obshchei Khim.*, 1971, **41**, 2085.

in solution. Tonelli *et al.*,[62] using 220 MHz n.m.r., c.d., and a theoretical treatment, claim that the conformation of antamanide is not stabilized by intramolecular hydrogen-bonding and that it possesses considerable symmetry, with a polar face (carbonyl groups capable of complexing with metal cations) on one side of the plane of the ring and a non-polar face on the other. Ivanov *et al.*[63] using 100 MHz n.m.r. with i.r. and o.r.d. data, have deduced that all N—H bonds participate in intramolecular hydrogen-bonding, and an n.m.r. examination of the Na^+ complex of antamanide [63] implies that the 'bracelet' form, similar to that of valinomycin complexes, is present in solution. Obviously more work is needed, and the answer may lie with *X*-ray crystallography; already a report on the crystallization and preliminary crystal data has been published.[64]

A number of antamanide analogues have been synthesized, in order to discover which residues are required for antibiotic activity against phalloidin (details of the cyclization step in each case are summarized in Table 1). Studies [65] on ten analogues possessing structural variants at positions 1 and 4 have shown that there is a requirement for amino-acids containing lipophilic side-chains in these positions, in order to retain antitoxic properties. Only [L-Abu[4]]antamanide in a series [66] of synthetic α-amino-butyric acid-containing analogues shows activity comparable to that of antamanide. *p*-Azidophenylalanine, tyrosine, *O*-methyltyrosine, and *O*-benzyltyrosine have in turn been incorporated [67, 68] into position 6 of antamanide. The first three analogues showed antitoxic activity but the *O*-benzyl analogue was ineffective. Antitoxic properties are also characteristic of [Phe[4], Val[6]]antamanide, which is an interesting analogue since it possesses *C*2 symmetry.[69] Most of the linear precursors of the cyclic decapeptides discussed above were synthesized using the automated solid-phase method. A new reactor for such a synthesis has also been tested in the synthesis of antamanide.[70] Other syntheses of antamanide, retro-antamanide, and perhydroantamanide have also been reported,[71] but details were not available in time for this review.

A synthesis of norphalloin (29) has been reported [72] and it involves two cyclization steps. The first cyclization at position A, using the anhydride

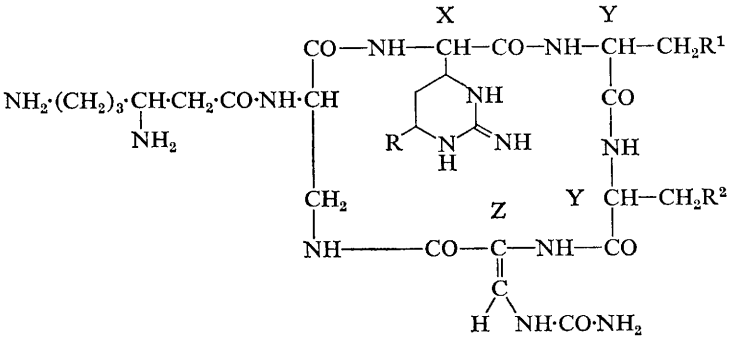
[62] A. E. Tonelli, D. J. Patel, M. Goodman, F. Naider, H. Faulstich, and Th. Wieland, *Biochemistry*, 1971, **10**, 3211.
[63] V. T. Ivanov, A. I. Miroshnikov, N. D. Abdullaev, L. B. Senyavina, S. E. Arkhipova, N. N. Uvarova, K. K. Khaliluna, V. F. Bystrov, and Yu. A. Ovchinnikov, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 654.
[64] W. Littke, *Tetrahedron Letters*, 1971, 4247.
[65] Th. Wieland, L. Lapatsanis, J. Faesel, and W. Konz, *Annalen*, 1971, **747**, 194.
[66] Th. Wieland, C. Birr, and A. von Dungen, *Annalen*, 1971, **747**, 207.
[67] Th. Wieland, A. von Dungen, and C. Birr, *Annalen*, 1971, **752**, 109.
[68] Th. Wieland and C. Rietzel, *Annalen*, 1971, **754**, 107.
[69] Th. Wieland, A. von Dungen, and C. Birr, *F.E.B.S. Letters*, 1971, **14**, 299.
[70] C. Birr and W. Lochinger, *Synthesis*, 1971, 319.
[71] Yu. A. Ovchinnikov, V. T. Ivanov, A. I. Miroshnikov, K. K. Khaliluna, and N. N. Uvarova, *Khim. prirod. Soedinenii*, 1971, **7**, 469.
[72] Th. Wieland, H. Faulstich, and F. Fahrenholz, *Annalen*, 1971, **743**, 77, 83.



(29)

method, yielded only 5% of the monocyclic compound, but using the same method for position B cyclization occurred in 25% yield.

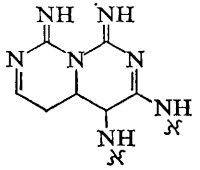
F. Viomycin, Capreomycin, and Tuberactinomycin.—Differences of opinion still exist concerning the structure of viomycin. In one study [73] the structure (30) has been put forward for viomycin and a related structure (31) [74] for



(30) R = R^1 = R^2 = OH
(31) R = R^1 = R^2 = NH_2
(33) R = H, R^1 = R^2 = OH, but with sequence of residues–YYZX instead of XYYZ shown in diagram

capreomycin IB. The structure of the chromophore in both cases has been proposed on the evidence that mild hydrolysis yields urea and a desurea derivative. Each desurea derivative re-formed the antibiotics in the presence of excess urea and dilute acid, thus giving strong support for the dehydroserine ureide unit as representing the chromophore. However, these deductions have been disputed [75] and another structure (32) for the

[73] B. W. Bycroft, D. Cameron, L. R. Croft, A. Hassanali-Walji, A. W. Johnson, and T. Webb, *Experientia*, 1971, **27**, 501.
[74] B. W. Bycroft, D. Cameron, L. R. Croft, A. Hassanali-Walji, A. W. Johnson, and T. Webb, *Nature*, 1971, **231**, 301.
[75] L. Lechowski, *Roczniki Chem.*, 1971, **45**, 581.

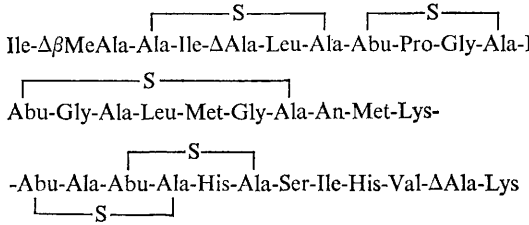


(32)

chromophore has been proposed. Nevertheless, *X*-ray data [76] on another related antibiotic, tuberactinomycin from *Streptomyces griseoverticillatus* var. *tuberacticus*, which also produces viomycin, confirm the dehydroserine ureide moiety for viomycin but lend support [76] to a totally different sequence for viomycin. The structure based on the *X*-ray results is represented by (33) for tuberactinomycin O, and from *X*-ray data viomycin appears to have a similar sequence.

Hydrogenated derivatives of viomycin and capreomycin have been reported [77] to have 30% and 43% respectively of the activity of the parent antibiotics. Detailed spectroscopic and degradative evidence [78] has been published for the structure of viomycidine, an artefact found in viomycin hydrolysates. Capreomycidine, a hydrogenation product of viomycidine, and an epimer have been synthesized,[79] and the configuration of *cis*-3-guanidinoproline, the other hydrogenation product of viomycidine, has been determined.[80]

G. Nisin.—Information obtained from cyanogen bromide cleavage, and from enzymic and Edman degradation, supports [81] the structure (34) for nisin. The structure, which contains a number of lanthionine residues, is

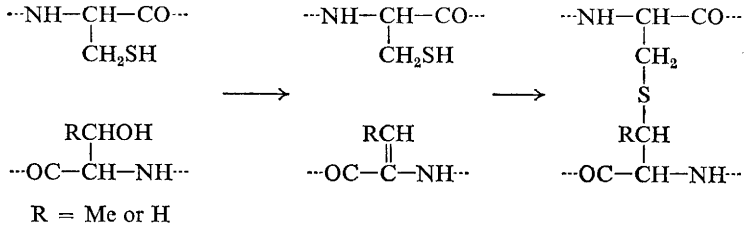


ΔβMeAla = β-methyl-dehydroalanine
ΔAla = dehydroalanine
Abu = α-aminobutyric acid

(34)

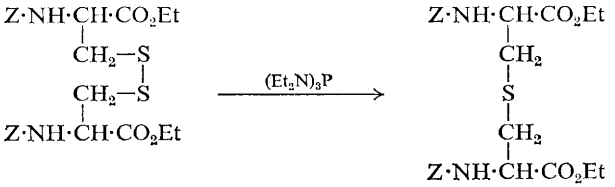
[76] H. Yoshioka, T. Aoki, H. Goko, K. Nakatsu, T. Noda, H. Sakakibara, T. Take, A. Nagata, J. Abe, T. Wakamiya, T. Shiba, and T. Kaneko, *Tetrahedron Letters*, 1971, 2043.
[77] J. R. Dyer, J. H. Carter, and P. J. van Wyk, *J. Medicin. Chem.*, 1971, **14**, 1120.
[78] G. Buchi and J. A. Raleigh, *J. Org. Chem.*, 1971, **36**, 873.
[79] B. W. Bycroft, D. Cameron, and A. W. Johnson, *J. Chem. Soc.* (*C*), 1971, 3040.
[80] C. Gallina, C. Marta, C. Colombo, and A. Romeo, *Tetrahedron*, 1971, **27**, 4681.
[81] E. Gross and J. L. Morell, *J. Amer. Chem. Soc.*, 1971, **93**, 4634.

the first discovered example of a pentacyclic heterodetic type peptide containing sulphur linkages. It is also interesting to note the presence of a number of $\alpha\beta$-unsaturated amino-acids which could well be biosynthetically important. In contrast to many microbial peptides, nisin is probably biosynthesized *via* a ribosomal mechanism, since lanthionine peptide synthesis can be inhibited [82] by chloroamphenicol, chlorotetracycline, and puromycin. Results using [^{14}C]cysteine show [82] that in addition to serine and threonine, cysteine also takes part in the biosynthesis of lanthionine as depicted in Scheme 2. Mass spectral properties of lanthionine



Scheme 2

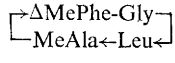
derivatives have been reported [83] together with a novel synthesis of L-lanthionine (36) by selective desulphurization of a cystine derivative (35).



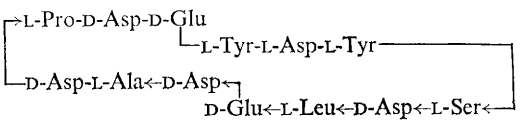
(35) (36)

H. Other Cyclic Peptides.—Tentoxin, a metabolite of *Alternaria tenius* Auct., has been tentatively shown [84] to have the structure (37) although the sequence and configuration is not completely defined. The results [85] from hypobromite oxidation, hydrazinolysis, and reduction support the structure (38) for mycobacillin, a peptide antibiotic isolated from cultures of *Bacillus subtilis* B_3. The solid-phase technique has been used extensively for the synthesis of linear protected lysine analogues of the cyclopeptide part of the polymyxins D,[86] M,[87] and E.[88]

[82] L. Ingram, *Biochim. Biophys. Acta*, 1970, **224**, 263.
[83] D. N. Harpp and J. G. Gleason, *J. Org. Chem.*, 1971, **36**, 73.
[84] W. L. Meyer, G. E. Templeton, C. I. Grable, C. W. Sigel, R. Jones, S. H. Woodhead, and C. Sauer, *Tetrahedron Letters*, 1971, 2357.
[85] S. Sengupta, A. B. Banerjee, and S. K. Bose, *Biochem. J.*, 1971, **121**, 839.
[86] M. A. Zevail and E. A. Morozova, *Vestnik. Moskov. Univ.*, 1971, **12**, 246.
[87] E. A. Morozova, M. A. Zevail, and G. F. Zhukova, *Zhur. obshchei Khim.*, 1970, **40**, 1376; E. A. Morozova and M. A. Zevail, *ibid.*, p. 1379.
[88] M. A. Zevail and E. A. Morozova, *Zhur. obshchei Khim.*, 1970, **40**, 2760.



(37)



(38)

I. Synthesis of Homodetic Cyclic Peptides.—Tables 1 and 2 summarize the details of the yields of cyclic peptides synthesized in 1971. It can be seen that the active ester and the dicyclohexylcarbodi-imide–hydroxysuccinimide methods for cyclization lead the list in popularity. However, the most impressive yield at the cyclization stage was achieved[48] by the azide

Table 1 *Syntheses of cyclic peptides achieved in* 1971
(Methods listed start from linear analogues unless otherwise stated)

Method for cyclization	*Peptide*	*Bond formed in cyclization step*	*Yield* (%)	*Ref.*
(*i*) *p-nitrophenyl ester method*				
	cyclo-(Pro$_3$-) and analogues	Pro-Pro	12	10
	cyclo-(Pro-Ser-Gly-)$_2$	Gly-Pro	16	14
	cyclo-(Ser-Pro-Gly-)$_2$	Gly-Ser[a]	4	15
	[Cys(Acm)^{2,7}]-gramicidin S[b]	Pro-Val	?	45
	[D-Ala$^{4,4'}$]-gramicidin S	Pro-Val	20	49
	cyclo-(Gly$_2$-L-Tyr-)$_2$	Gly-Tyr[a]	9	90
(*ii*) *DCCI–N-hydroxysuccinimide*				
	[Val^1, Ala^4]-antamanide	Phe-Phe	31	65
	[Ile^1, Val^4]-antamanide	Phe-Phe	48	65
	[Leu^1, Ala^4]-antamanide	Phe-Phe	46	65
	[Gly^1, Ala^4]-antamanide	Phe-Phe	49	65
	[Ala^1, Ala^4]-antamanide	Phe-Phe	45	65
	[Ala^1, Gly^4]-antamanide	Phe-Phe	49	65
	[Gly^1, Gly^4]-antamanide	Phe-Phe	47	65
	[Abu^1]-antamanide	Phe-Phe	26	66
	[Abu^4]-antamanide	Phe-Phe	16	66
	[Abu^{1,4}]-antamanide	Phe-Phe	8	66
	[Phe(pN$_3$)^6]-antamanide	Pro-Ala	14	67
	[Tyr^6]-antamanide	Phe-Tyr	33	68
(*iii*) *Azide method*				
	[Orn(Z(OMe))$^{2,2'}$]-gramicidin S	Leu-Phe	85	48
	cyclo-(Gly$_2$-L-Tyr)$_2$	Gly-Tyr[a]	6	90
	cyclo-(Gly$_2$-L-Tyr)$_2$	Gly-Gly[a]	17	90
	cyclo-(Gly$_2$-L-Tyr)$_2$	Tyr-Gly[a]	15	90
	cyclo-(Gly$_2$-L-Tyr)$_2$	Tyr-Gly	20	90
	cyclo-(Gly$_2$-Phe-)$_2$	Phe-Gly[a]	25	90
	cyclo-(Gly-Cys(Bzl)-L-Leu-)$_2$	Leu-Gly	26	92
	cyclo-(Gly-Cys(Bzl)-L-Leu-)$_2$	Leu-Gly[a]	20	92
	cyclo-(Gly-Cys(Bzl)-D-Leu-)$_2$	Leu-Gly	43	92
	cyclo-(Gly-Cys(Bzl)-D-Leu-)$_2$	Leu-Gly[a]	16	92

Table 1 (*cont.*)

Method for cyclization	*Peptide*	*Bond formed in cyclization step*	*Yield* (%)	*Ref.*
(*iv*) *Mixed anhydride method*				
	[Ala^1, Val^4]-antamanide	Phe-Phe	20	65
	[Val^1, Phe^4]-antamanide	Phe-Phe	21	65
	[Phe^1, Phe^4]-antamanide	Phe-Phe	30	65
	[Val^1, Gly^4]-antamanide	Gly-Phe	20	65
	norphalloin (29) 1st stage	*a*Hyp-Cys	5	72
	2nd stage	Nva-Ala	27	72
(*v*) *Tcp ester method*				
	(see Table 2)			89
(*vi*) *o-Phenylene chlorophosphite/imidazole*				
	cyclopeptide (40)	Gly-β-Ala	40	91
	cyclopeptide (41)	Gly-β-Ala	60	91

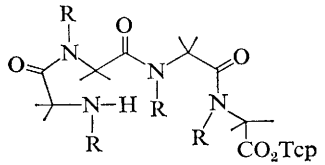
[a] By cyclodimerization of the corresponding monomeric sequences.
[b] Acm = acetamidomethyl.

Table 2 *Cyclization of linear sarcosine tetrapeptides (pyridine at* 115 °C)

1	2	3	4	% *Yield of cyclic tetrapeptide*	*Cyclic dipeptide* (% yield)
Sar	Sar	Sar	Sar-OTcp	43	—
Gly	Sar	Sar	Sar-OTcp	25	—
Sar	Gly	Sar	Gly-OTcp	40	—
Sar	Gly	Gly	Sar-OTcp	10	—
Sar	Sar	Gly	Gly-OTcp	—	*cyclo*-(Sar$_2$-) 25%; *cyclo*-(Gly$_2$-) 10%
Gly	Sar	Sar	Gly-OTcp	—	*cyclo*-(Gly-Sar-) 55%
L-Ala	Sar	Sar	Sar-OTcp	25	—
Sar-D-Ala		Sar-L-Ala	OTcp	30	—
Sar-L-Ala-L-Ala		Sar	OTcp	10	—
Sar	Sar-L-Ala	L-Ala	OTcp	—	*cyclo*-(Sar$_2$-) 30%; *cyclo*-(L-Ala$_2$-) 13%

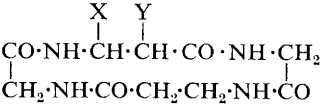
method in the synthesis of a derivative of gramicidin S. Table 2 shows that several closely related linear tetrapeptides containing sarcosine do not cyclize under the same conditions but split [89] into cyclic dipeptides. Cyclic dipeptides are formed more readily if the amide bond joining residues 1 and 2 is of the *cis*-allowed type (obtained readily in sarcosine) whereas the 3 → 4 amide group is of the *trans*-preferred type. Diagrammatically this is visualized as in structure (39) where the free amino end appears to be unable to reach the activated ester group. It has been suggested that a good yield of cyclic tetrapeptide is obtained whenever the linear peptide chain has a high probability of folding the same way as *cis,trans,cis,trans* sequence in the cyclic tetrapeptides.

[89] K. Titlestad, *Chem. Comm.*, 1971, 1527.



(39)

Studies [90] on the synthesis of *cyclo*-(Gly_2-L-Tyr-Gly_2-L-Tyr-) reveal that using the azide method the yields varied depending upon the position of the tyrosyl residue in the tripeptide precursor when a cyclodimerization method was used. Thus Tyr-Gly-Gly-N_3, Gly-Tyr-Gly-N_3, and Gly-Gly-Tyr-N_3 cyclodimerized in 6, 17, and 15% yield respectively. The best yield (20%) was obtained by the cyclization of the linear hexapeptide Gly_2-Tyr-Gly_2-Tyr-N_3 using the azide method. Treatment of the cyclic hexapeptide with chymotrypsin showed that the cyclic peptide was hydrolysed to Gly-Gly-Tyr after 6 h presumably *via* the linear hexapeptide analogue.



(40) X = H, Y = NH_2
(41) X = CO_2Me, Y = H

Very good yields of the model cyclic tetrapeptides (40) and (41) have been obtained [91] from the corresponding linear analogues using *o*-phenylene-chlorophosphite–imidazole. Two analogous cyclohexapeptides, *cyclo*-[Gly-Cys(Bzl)-L-Leu-Gly-Cys(Bzl)-L-Leu-] and *cyclo*-[Gly-Cys(Bzl)-D-Leu-Gly-Cys(Bzl)-D-Leu-] have been prepared [92] in 26% and 43% yield respectively from the t-butoxycarbonylhydrazide of the linear hexapeptide. The hydrazide without the t-butoxycarbonyl group attached gave a higher yield (55%) for the all-L compound, but lower yields were obtained when attempts were made to cyclodimerize from the corresponding tripeptides. Comparison of the c.d. spectrum of the all-L *S*-benzylcysteine peptide with that of an L-phenylalanine analogue showed that [92] the peptide backbones must have similar conformations. However, the comparison does not hold for the cyclo-[Gly-Cys(Bzl)-D-Leu-Gly-Cys(Bzl)-D-Leu-], which may be explained by the fact that the phenyl ring of the *S*-benzyl group is further away than the phenylalanine ring from the peptide backbone thus giving no phenyl–amide interaction.

[90] M. Konishi, N. Yoshida, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1971, **44**, 2801.
[91] C. H. Hassall, D. G. Sanger, and B. K. Handa, *J. Chem. Soc.* (*C*), 1971, 2814.
[92] K. Bláha, I. Frič, Z. Bezpalova, and O. Kaurov, *Coll. Czech. Chem. Comm.*, 1970, **35**, 3557.

Until this year, strict adherence to the title of this chapter has kept oxytocin analogues with their disulphide links outside the bounds of the chapter. However, synthetic studies [93] on analogues of deamino-oxytocin with ring systems not containing a disulphide bond have provided the necessary conditions for coverage in the present context. Thus *cyclo*-[Tyr-Ile-Gln-Asn-Cys($CH_2 \cdot CH_2 \cdot CO$)]-Pro-Leu-Gly-NH_2 with the cystine disulphide bridge replaced by a $-CH_2-S-$ bridge and the methylene analogue *cyclo*-[Tyr-Ile-Gln-Asn-Apim*]-Pro-Leu-Gly-NH_2 were obtained in up to 60% yield by cyclization at the tyrosyl–cysteinyl bond (or analogue) using the *p*-nitrophenyl ester method or pyridine in DMF.

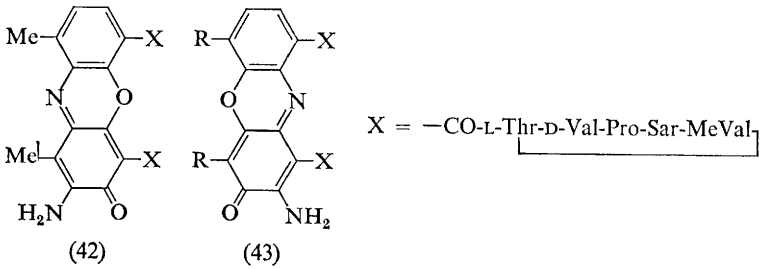
3 Depsipeptides (Heterodetic Peptides)

A. Actinomycins.—Actinomycins C_1(D) selectively deuteriated in both the α- or β-peptide lactone rings and in the chromophore have been synthesized [94] and used to good effect in the assignment [95] of chemical shifts in the n.m.r. spectra of actinomycin C_1(D) and closely related analogues. Details of the conformation of the single pentapeptide lactone unit derived from this study (Vol. 3, p. 294) have been questioned [96] mainly in connection with the fact that the *N*-methyl group of sarcosine appears to be too near the proline ring. Apart from this detail there is good agreement between this conformation and that obtained from theoretical calculations [96] based on experimental data. This model has also been supported by *X*-ray crystallographic data [97] on a complex formed between 7-bromoactinomycin C_1(D) and deoxyguanosine which confirms the two-fold symmetry of the actinomycin molecule as shown in Figure 3. A strong hydrogen bond exists between neighbouring cyclic pentapeptide chains connecting the N—H bond of one D-valine residue with the carbonyl groups of the other D-valine. It has also been shown that hydrogen bonds play an important role in the association between the actinomycin and the guanine residues and the 1 : 2 stoicheiometry of the complex is a direct consequence of the two-fold symmetry of the actinomycin. This appears to be a very interesting model for further work on actinomycin–DNA binding characteristics.

Pseudoactinomycin C_1 (42) [98] and a series of actinomycin derivatives [99] (43; R = H, OMe, Et, or Bu^t) have been synthesized. Yields of up to 30% in the cyclization step to form the lactone rings were achieved using *N*-acetylimidazole–acetyl chloride. In the derivatives (43) antibiotic activity was highest for R = H, but all substituents decreased the activity,

[93] K. Jost, *Coll. Czech. Chem. Comm.*, 1971, **36**, 218; K. Jost and F. Sorm, *ibid.*, p. 2795.
[94] H. Lackner, *Chem. Ber.*, 1971, **104**, 3653.
[95] H. Lackner, *Tetrahedron Letters*, 1971, 2221.
[96] P. De Santis, R. Rizzo, and G. Ughetto, *Tetrahedron Letters*, 1971, 4309.
[97] H. M. Sobell, S. C. Jain, T. D. Sakore, and C. E. Nordman, *Nature (New Biol.)*, 1971, **231**, 200.
[98] H. Brockmann and E. Schulze, *Tetrahedron Letters*, 1971, 1489.
[99] H. Brockmann and F. Seela, *Chem. Ber.*, 1971, **104**, 2751.

* Apim = α-aminopimelic acid.



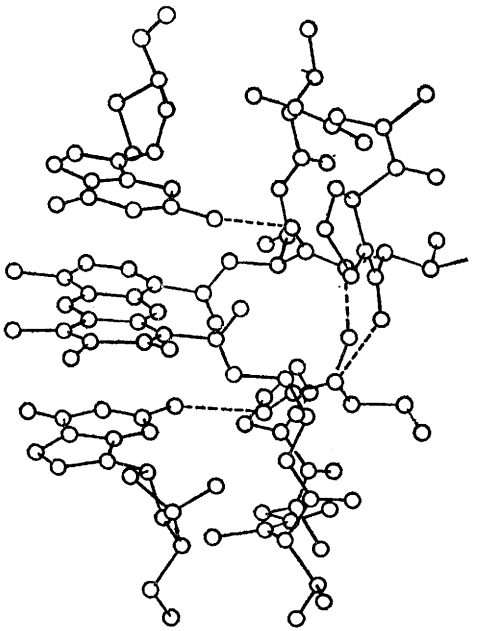
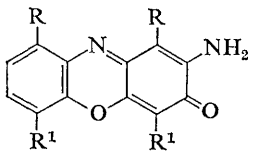


Figure 3 *Illustration of actinomycin–deoxyguanosine complex. An additional weak hydrogen bond (not shown) connects the guanine* N(3) *ring nitrogen with the* N—H *group in* L-*threonine residues*
[Reproduced by permission from *Nature* (*New Biol.*), 1971, **231**, 200]

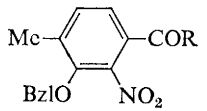
the extreme case being the 4,6-di-t-butyl compound with no activity. Model compounds (44),[100] (45), and (46),[101] based on the phenoxazone chromophore of actinomycin, have been synthesized in order to study actinomycin–DNA interactions. Attempts to link (47; R = Cl) to the protected penta-depsipeptide (48) have been complicated by the formation of oligomers,

[100] M. T. Wu and R. E. Lyle, *J. Heterocyclic Chem.*, 1971, **8**, 989.
[101] F. Seela, *Z. Naturforsch.*, 1971, **26b**, 875.



(44) R = H, R^1 = —CO·NH·CH_2·CO_2H or
R^1 = —CO·NH·CH(Me)·CO_2H

(45) R = CO·NH·$(CH_2)_2NEt_2$, R^1 = Me
(46) R = CO·NH·$(CH_2)_2NEt_2$, R^1 = Bu^t



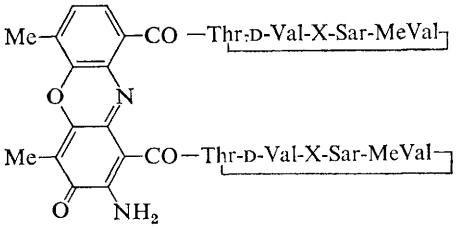
(47)

Z-Sar-MeVal
H-Thr-D-Val-Pro-OH

(48)

but an improved yield of the monomer can be obtained[102] using the symmetrical anhydride generated by reaction with dicyclohexylcarbodiimide.

A series of actinomycins has been used[103] to ascertain the usefulness of a pyrolysis-gas chromatography analysis of dioxopiperazines as a method for determining peptide sequences. All adjacent pairs of amino-acids in the actinomycins, except those containing threonine, formed dioxopiperazines and the sequences (49) and (50) have been confirmed for actinomycin II and actinomycin Pip 2 respectively.



(49) X = Sar
(50) X = Pipecolyl

Actinomycin mono-lactone isolated[104] from *Streptomyces antibioticus* 3720 shows only 1% of the biological activity of actinomycin C_1(D). Addition of D- or L-isoleucine or L-*allo*-isoleucine to the culture medium of *S. antibioticus* or *S. chrysomallus* has led[105] to the synthesis of new actinomycins. Hydrolysates of all these actinomycins contained *N*-methyl-L-*allo*-isoleucine, which is another interesting example of the isomerization of amino-acids by micro-organisms.

[102] J. Meienhofer, R. Cotton, and E. Atherton, *J. Org. Chem.*, 1971, **36**, 3746.
[103] A. B. Mauger, *Chem. Comm.*, 1971, 39.
[104] K. L. Perlman, J. Walker, and D. Perlman, *J. Antibiotics (Japan)*, 1971, **24**, 135.
[105] E. Katz, Y. Kawai, and J. Ohoji, *Biochem. Biophys. Res. Comm.*, 1971, **43**, 1035.

B. Valinomycin.—The ion-complexing properties and conformation of this cyclic depsipeptide continue to be fields of active interest. In a very comprehensive application of model calculations,[106] the complexation of alkali- and alkaline-earth-metal cations by valinomycin has been compared with the macrotetrolide antibiotics and other compounds. The fluorescent probes 1-anilino-8-naphthalene and 2-*p*-toluidino-6-naphthalenesulphonates have also been used [107] to determine relative cation affinities in valinomycin, the order of affinity being $Rb^+ > K^+ > Cs^+$. In a conductometric study [108] of complex formation with both valinomycin and its analogues, only analogues which formed potassium ion complexes showed antibiotic activity.

Details of the combined use of i.r., n.m.r., o.r.d., and c.d. techniques which support the 'bracelet' type of conformation for valinomycin have now appeared [109] (Vol. 2, p. 200). This 'bracelet' model has also been confirmed in conformational energy calculations [110] based on i.r. and n.m.r. studies. The most stable conformer exists with the methyl groups of the lactyl side-chain in the axial position with all the carbonyl groups pointing inwards. Valinomycin and its K^+-complex have also been studied [111] using Fourier-transform ^{13}C n.m.r. Resonances of the carbonyls directly co-ordinating the K^+-ion in valinomycin (and nonactin) shift downfield (4 p.p.m.) upon complex formation, but much smaller shifts (up- and down-field) appeared for carbons not involved in the binding site.

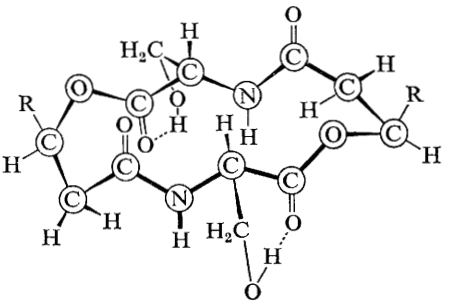
The linear sequence H-(D-Val-L-Lac-L-Val-D-Hyiv*)$_3$-OH of valinomycin has been synthesized [112] using three approaches: (*a*) in homogeneous phase, by stepwise lengthening of L-Val-D-Hyiv-OBzl with Boc-amino-acyl-hydroxy-acids (12% yield); (*b*) by solid-phase synthesis using the same unit (56% yield); (*c*) by a two-fold fragment condensation of D-Val-L-Lac-L-Val-D-Hyiv-O-Ⓟ with Boc-D-Val-L-Lac-L-Val-D-Hyiv-OH (66% yield). Cyclization to form valinomycin was carried out using thionyl chloride, giving yields of up to 24%. Topochemical analogues [113] of valinomycin and analogues differing in the number of amide and ester bonds have also been synthesized.[114] Ester bonds in the *meso* analogues, cyclo-(D-Val-L-Hyiv-L-

[106] W. E. Morf and W. Simon, *Helv. Chim. Acta*, 1971, **54**, 2683.
[107] M. B. Feinstein and H. Felsenfeld, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 2037.
[108] I. M. Andreev, G. G. Malenkov, A. A. Shkrob, and M. M. Shemyakin, *Mol. Biol.*, 1971, **5**, 614.
[109] V. T. Ivanov, I. A. Laine, N. D. Abdullaev, V. Z. Pletnev, G. M. Lipkind, S. F. Arkhipova, L. B. Senyavina, E. N. Meshcheryakova, E. M. Popov, V. F. Bystrov, and Yu. A. Ovchinnikov, *Khim. prirod. Soedinenii*, 1971, **7**, 221.
[110] D. F. Mayers and D. W. Urry, *J. Amer. Chem. Soc.*, 1972, **94**, 77.
[111] M. Ohnishi, M.-C. Fedarko, J. D. Baldeschwieler, and L. F. Johnson, *Biochem. Biophys. Res. Comm.*, 1972, **46**, 312.
[112] G. Losse and H. Klengel, *Tetrahedron*, 1971, **27**, 1423.
[113] V. T. Ivanov, I. A. Laine, I. D. Ryabova, and Yu. A. Ovchinnikov, *Khim. prirod. Soedinenii*, 1970, **6**, 744.
[114] L. A. Fonina, A. A. Sanasaryan, and E. I. Vinogradova, *Khim. prirod. Soedinenii*, 1971, **7**, 69.

* Hyiv = α-hydroxyisovaleric acid, Lac = lactic acid.

Ala-D-Hyiv-)$_3$, cyclo-(D-Ala-L-Hyiv-L-Ala-D-Hyiv-)$_3$, and cyclo-(D-Val-Glyc-L-Val-Glyc-*)$_3$ and eight diastereoisomers have been produced [115] by forming mixed anhydrides with benzenesulphonyl chloride, whereas the amide bonds were formed from the acid chloride or by means of dicyclohexylcarbodi-imide and *N*-hydroxysuccinimide.

C. Other Naturally Occurring Depsipeptides.—I.r. and n.m.r. studies [116] at 100 and 220 MHz on serratamolide derivatives suggest that the most likely conformation for the molecule is as shown in [51; R = $(CH_2)_6Me$] which contains 'free' N—H groups and intramolecularly bonded hydroxy-groups.



(51)

In an analogue of serratamolide in which two seryl groups have opposite configurations (*meso* form) and with the two hydroxy-groups protected as t-butyl ethers, a conformation similar to that proposed by Dale for cyclotetradecane has been proposed. Details of the elucidation of the structure of the macrocyclic lactone griselimycin (52) and a number of minor analogues (53), (54), and (55) have been reported.[117]

MeCO-MeVal-X-MeThr-Leu-Y-Leu-MeVal
Gly←D-MeLeu←Q

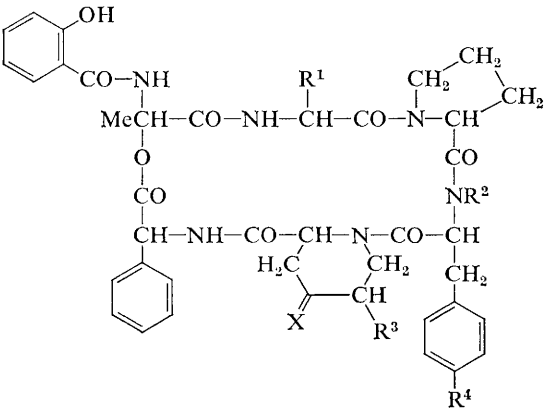
(52) X = Y = *trans*-4MePro, Q = Pro
(53) X = Y = Q = *trans*-4MePro
(54) X = Q = Pro, Y = *trans*-4MePro
(55) X = Y = Q = Pro

Purification of virginiamycin S has revealed [118] three other minor components, probably formed by use of a mutant strain (or due to the modification of the growth medium) of *Streptomyces virginiae*. Virginiamycin S was found to be identical with ostreogrycin B_1 (56) and virginiamycin S_3 was found to be identical with ostreogrycin B_2 (57). The third,

[115] L. A. Fonina, A. A. Sanasaryan, E. I. Vinogradova, and U. B. Shvetsov, *Khim. prirod. Soedinenii*, 1971, 7, 81.
[116] C. H. Hassall, M. C. Moschidis, and W. A. Thomas, *J. Chem. Soc.* (*B*), 1971, 1757.
[117] B. Terlain and J. P. Thomas, *Bull. Soc. chim. France*, 1971, 2349, 2357, 2363.
[118] H. Vanderhaeghe, G. Janssen, and F. Compernolle, *Tetrahedron Letters*, 1971, 2687.

* Glyc = glycollic acid.

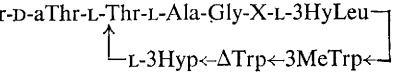
virginiamycin S_2, was found to have structure (58). Ostreogrycin B_3, a minor component from cultures of *Streptomyces ostreogrisens*, has been given [119] the structure (59).



(56) $R^1 = R^2 = Me$, $R^3 = R^4 = H$, $X = O$
(57) $R^1 = Et$, $R^2 = Me$, $R^3 = OH$, $R^4 = H$, $X = O$
(58) $R^1 = Et$, $R^2 = R^3 = R^4 = H$, $X = H,OH$
(59) $R^1 = Et$, $R^2 = Me$, $R^3 = OH$, $R^4 = NMe_2$, $X = O$

Full details [120] of the mass spectrometric breakdown of the destruxins A and B have become available (see Vol. 3, p. 299, for typical structures).

The results of the Edman sequencing method give support [121] to the structures (60) and (61) for the antibiotics A-128-OP and A-128-P respectively. The main structures of these compounds appear to differ from



(60) X = *trans*-3Hyp
(61) X = Pro

telomycin only in the choice of the threonine residue used to make up the depsipeptide ring. Mild alkaline hydrolysis of A-128-OP converted the *O*-substituted threonine into β-methyldehydroalanine.[122]

The origin of D-amino-acid residues in microbial peptides continues to be of interest but no satisfactory explanation has yet been obtained. Studies [123] on angolide cyclo-(L-Hyiv-L-Ile-L-Hyiv-D-*a*Ile-) from the fungus *Pithomyces sacchari* show that if the organism is grown on a medium rich in valine then a mixture of (L-Val^2, D-Val^4)-angolide together with

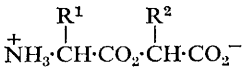
[119] B. R. Cox, F. W. Eastwood, B. K. Shell, and Lord Todd, *Chem. Comm.*, 1970, 1623.
[120] A. Suzuki, N. Takahashi, and S. Tamura, *Org. Mass Spectrometry*, 1970, **4**, 175.
[121] A. B. Silaev, G. S. Katrukha, Z. P. Trifonova, R. I. Li, and T. M. Melent'eva, *Khim. prirod. Soedinenii*, 1971, **7**, 130.
[122] I. G. Smirnova, A. B. Silaev, and G. S. Katrukha, *Khim. prirod. Soedinenii*, 1971, **7**, 544.
[123] R. O. Okotore and D. W. Russell, *Experientia*, 1971, **15**, 380.

homologues in which one isoleucine is replaced by valine are obtained. It has therefore been suggested, since exogeneous L-valine promotes the biosynthesis of L-Val[2]- and D-Val[4]-angolide in equal amounts, that the biosynthesis of angolide itself proceeds *via* an all-L cyclotetradepsipeptide with single D-residues being introduced by random inversion of one or other of two stereochemically equivalent L-amino-acid residues. ^{14}C-Labelling studies[124] on L-*N*,β-dimethyl-leucine in the antibiotic etamycin have shown that the amino-acid is derived from L-leucine and that the *N*-methyl and β-methyl groups are derived from L-methionine.

A theoretical conformational analysis of the enniatins has been reported.[125]

D. Miscellaneous.—The conformations of the substituted 2,5-dioxomorpholines cyclo-(L-Ala-L-Lac-), cyclo-(L-Ala-D-Lac-), cyclo-(L-Val-L-Hyiv-), and cyclo-(L-Val-D-Hyiv-) have been analysed[21,126] using ^{1}H n.m.r., o.r.d., and a theoretical calculation, and they all have non-planar conformations. Cyclo-(L-Val-L-Hyiv-), with an LL configuration and bulky side-chains, has a ring-folding angle > 180° with a pseudo-equatorial arrangement of substituents. Optimum conformations of the cyclic tetradepsipeptides cyclo-(Gly-Glyc-)$_2$ and cyclo-(MeAla-Lac-)$_2$ with all possible sets of C$^\alpha$-configurations have been calculated.[127] The compounds have *cis*-amide and *trans*-ester groups and the optimum form of cyclo-(L-MeAla-D-Lac)$_2$ corresponds to the conformation previously reported in an *X*-ray study on cyclo-(D-Hyiv-L-MeIle-)$_2$ (Vol. 2, p. 206). Conformational analyses have also been carried out[128] on various derivatives of *O*-acetyl-L-Lac-L-Ala and *N*-acetyl-L-Ala-L-Lac.

Synthesis of the LL or LD diastereoisomers of (62) has been achieved[129] using the triethylammonium salts of the benzyloxycarbonylamino-acid



(62) (R^1 = Pr^i or $CH_2 \cdot CHMe_2$; R^2 = Me or Et)

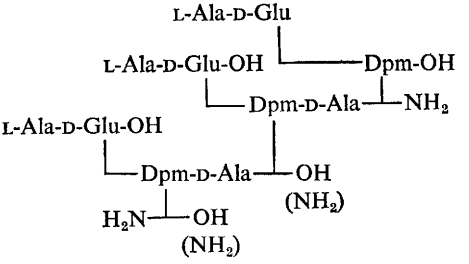
derivative and the t-butyl ester of racemic α-bromopropionic acid or α-bromobutyric acid. The free depsipeptides were obtained on catalytic hydrogenolysis and the diastereoisomers were separated[130] using paper chromatography.

[124] J. E. Walker and D. Perlman, *Biotechnol. and Bioeng.*, 1971, **13**, 371.
[125] E. M. Popov, V. Z. Pletnev, A. V. Evstratov, V. T. Ivanov, and Yu. A. Ovchinnikov, *Khim. prirod. Soedinenii*, 1970, **6**, 616.
[126] E. M. Popov, V. Z. Pletnev, G. M. Lipkind, and S. F. Arkhipova, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1971, 33.
[127] V. Z. Pletnev and E. M. Popov, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1970, 991; *Biofizika*, 1971, **16**, 407.
[128] E. M. Popov, V. Z. Pletnev, G. M. Lipkind, and S. F. Arkhipova, *Khim. prirod. Soedinenii*, 1971, **7**, 191, 184.
[129] C. Wasielewski and M. Hoffmann, *Roczniki Chem.*, 1971, **45**, 995.
[130] C. Wasielewski and M. Hoffmann, *Roczniki Chem.*, 1970, **44**, 465.

4 Peptide–Carbohydrate Linkages (See also the Specialist Periodical Reports on Carbohydrate Chemistry)

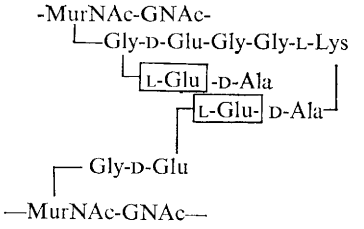
A. Glycopeptides from Bacterial Cell Walls.—Fewer papers than usual have appeared on this subject this year. Biosynthetic aspects have been the subject of a review [131] and a three-dimensional model of the bacterial cell wall has been proposed.[132]

A number of peptides making up the various partial structures of the peptide (63) have been isolated [133] from autolysates of the vegetative cell



(63)

walls of *Bacillus subtilis*. The main difference between this glycopeptide and the ones from *E. coli* and *B. megaterium* is the presence of the diaminopimelyl amide residues and the absence of C-terminal D-alanine. Partial hydrolysis [134] of purified UDP-activated precursors of the peptidoglycan of *Arthrobacter* strain J.39 has given peptides in agreement with the unit structure (64). This is the first example in which a dicarboxylic



(64)

amino-acid [*i.e.* L-Glu as indicated in (64)] occupies position 3: usually only diamino-acids are found there. Enzymic degradation [135] of cell walls of strains of *Aerococcus viridans* and *Gaffkya homari* reveal that the glycopeptides contain the peptide units L-Ala-γ-D-Glu-L-Lys-D-Ala and L-Ala-D-isoglutaminyl-L-Lys-D-Ala. Half of these peptide units occur as

[131] F. C. Neuhaus, *Accounts Chem. Res.*, 1971, **4**, 297.
[132] M. V. Kelemen and H. J. Rogers, *Proc. Nat. Acad. Sci. U.S.A.*, 1971, **68**, 992.
[133] A. D. Warth and J. L. Strominger, *Biochemistry*, 1971, **10**, 4349.
[134] B. Cziharz, K. H. Schleifer, and O. Kandler, *Biochemistry*, 1971, **10**, 3574.
[135] M. Nakel, J. M. Ghuysen, and O. Kandler, *Biochemistry*, 1971, **10**, 2170.

unlinked monomers whereas others form dimers *via* N^{ε}-(D-Ala)-L-lysyl linkages and constitute the first known examples of lysine-containing peptidoglycans of chemotype I (see Vol. 3, p. 302, for references defining chemotypes). L-Configurations have been assigned [136] to homoserine and diaminobutyric acid from mucopeptide precursor nucleotides and from the cell walls of some plant pathogenic corynebacteria. Chromatographically distinct fractions of the glycopeptide* GNAc-MurNAc-L-Ala-D-Glu-*meso*-Dpm-D-Ala have been isolated [137] from the murein of a venereal and intestinal strain of *Vibrio fetus* and were found to be identical. A simple elution system involving two (or three) buffer changes from pH 2.87 to 5.25 on a single-column amino-acid analyser has been used [138] in the analysis of all amino-acids and amino-sugars in some common cell-wall peptidoglycans.

B. Glycopeptides from Miscellaneous Sources.—The formation of trimethylsilyl derivatives of both amino-acids and carbohydrates from hydrolysates of known glycopeptides enable [139] g.l.c. analyses of both components to take place simultaneously. Preliminary reports [140] have appeared of a new type of carbohydrate–peptide linkage in the form of a link between the thiol group of cysteine and galactose as found in (galactosyl)-Cys-Glu-His-Ser-His-Asp-Gly-Ala isolated from human urine. The position of the carbohydrate–peptide link was inferred from the isolation of a galactosyl–cysteine from a leucine aminopeptidase digest. A similar type of glycopeptide has been found [141] in human erythrocyte membrane. The novel glucosyl–cysteine linkage was again proved using leucine aminopeptidase digestion and this, together with the use of Dansyl–Edman technique, confirmed the structure to be (65). Tryptic digests [142] of κ-type light chains

$$\begin{array}{c}(\text{Glucosyl})_3\\ |\\ \text{Cys-Glu-Gly-His-Ser-His-Asp-His-Gly-Ala}\end{array}$$

(65)

from human myeloma proteins have yielded the glycopeptide having the sequence Ala-Ser-Gln-Asn-Ile-Ser, corresponding to positions 25—31 in the protein chain, and another glycopeptide Phe-Ser-Gly-Ser-Gly-Ser-Gly-(Thr-Asp)-Phe-Thr-Leu-Asx-Ile-Ser-Arg, corresponding to positions 62—77. In both these peptides the carbohydrate link is probably attached to the asparagine residue.

Two large glycopeptides, one a unit containing 22 amino-acid residues with one carbohydrate link, and the other a unit with 65 amino-acid residues

[136] H. R. Perkins, *Biochem. J.*, 1971, **121**, 417.
[137] A. J. Winter, W. Katz, and H. H. Martin, *Biochim. Biophys. Acta*, 1971, **244**, 58.
[138] P. Guire, *Analyt. Biochem.*, 1971, **42**, 1.
[139] T. Bhatti and J. R. Clamp, *Biochim. Biophys. Acta*, 1971, **229**, 293.
[140] C. J. Lote and J. B. Weiss, *Biochem. J.*, 1971, **123**, 25P; *F.E.B.S. Letters*, 1971, **16**, 81.
[141] J. B. Weiss, C. J. Lote, and H. Bobinski, *Nature (New Biol.)*, 1971, **234**, 25.
[142] C. P. Milstein and C. Milstein, *Biochem. J.*, 1971, **121**, 211.

* GNAc = *N*-acetyl-D-glucosamine, MurNAc = *N*-acetylmuramic acid.

with four carbohydrate linkages, have been characterized[143] from the amino-terminus of α_1-acid glycoprotein (human plasma globulin). All the carbohydrate units were found to be attached to asparagine residues with the carboxy-terminal sequences; -Asn-Ala-Thr, -Asn-Lys-Ser, -Asn-Lys-Thr, -Asn-Thr-Thr, -Asn-Gly-Thr. This is in accord with a previous theory[144] that in globular glycoproteins the carbohydrate–peptide linkage is usually found in the environment -Asn-X-(Ser or Thr)-. On the basis of the structures deduced[145] for five glycopeptides [(66)—(70)] from the constant sequence region of μ heavy chain of human immunoglobulin M, it is quite probable that the same theory holds. The same situation arises in the structures of the glycopeptides (71)—(73) isolated[146] from tryptic

(66) Phe-Ser-Trp-Lys-Tyr[Ser, Lys, Asx, Asx, Asx(Glycan)]
(67) Gly-Leu-Thr-Phe-Gln-Glx-Asx(Glycan)-Ala-Ser-Ser-Met
(68) Val-Lys-Thr-His-Thr-Asx(Glycan)
(69) Ile-Ser-Glx[Ser, His, Pro, Asx(Glycan)]
(70) Leu-Tyr-Asx(Glycan)-Val-Ser-Leu-Val-Met
(71) Gly-Ser-Asn(Glycan)-Val-Thr-Asp-Cys-Ser-Gly-Asp-Phe
O_3H
(72) Asn(Glycan)-Lys-Ser-Asx_2(Thr,Asx,Pro,Cys(O_3H)Glx_2)-Ala-Gly-Tyr
(73) Gly-Ser-Asn(Glycan)-Val-Thr-Asp-Cys-Ser-Gly-Asp-Phe-Leu-Phe-Cys

hydrolysates of human transferrin. The carbohydrate linkage in a glycopeptide from taka-amylase A has been proved[147] to be GNAc-Asp-Ser with the sugar moiety β-anomeric. Two glycopeptides obtained from tryptic and pronase digests of fibroin of the silk worm *Bombyx mori* have both been found[148] to contain the Ser-Asn-(glucosaminyl)-Thr unit. Pronase digestion of phosvitin, the phosphoprotein from hens' egg yolk, has yielded[149] a glycopeptide having the sequence (74). Again all of the

Ser-Asn(Glycan)-Ser-Gly-Ser$(P)_8$-Arg-Ser-Val-Ser-His_2

(74)

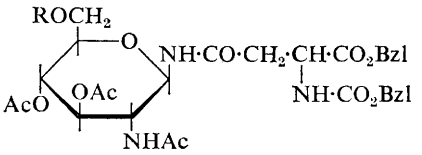
carbohydrate part of the molecule is attached to the aspartic acid residue and an interesting feature is the eight phosphoserine residues in sequence.

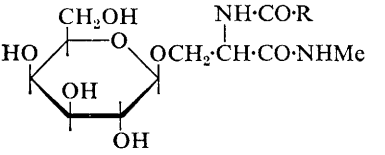
Alkaline degradation studies[150] have revealed linkages between 2-acetamido-2-deoxygalactose and serine and threonine in the antigenic

[143] K. Schmid, M. Ishiguro, J. Emura, S. Isemura, H. Kaufmann, and T. Motoyama, *Biochem. Biophys. Res. Comm.*, 1971, **42**, 280.
[144] A. Neuberger and R. D. Marshall, in 'Carbohydrates and their Roles', ed. H. W. Schultz, R. F. Cain, and R. W. Woolstad, Avi, Westport, Conn., 1969.
[145] A. Shimizu, F. W. Putnam, C. Paul, J. R. Clamp, and I. Johnson, *Nature* (*New Biol.*), 1971, **231**, 73.
[146] P. Charet and J. Montreuil, *Compt. rend.*, 1971, **273**, *D*, 533; P. Charet, G. Spil, and J. Montreuil, *ibid.*, p. 422.
[147] H. Yamaguchi, T. Ikenaka, and Y. Matsushima, *J. Biochem.* (*Japan*), 1971, **70**, 587.
[148] H. Sinohara, Y. Asano, and A. Fukui, *Biochim. Biophys. Acta*, 1971, **237**, 273.
[149] R. Shainkin and G. E. Perlmann, *J. Biol. Chem.*, 1971, **246**, 2278.
[150] J. M. How and J. D. Higginbotham, *Carbohydrate Res.*, 1970, **14**, 335.

sulphated glycopeptide from chick allantoic fluid. Details have also been reported of glycopeptides isolated from both low- and high-density platelet plasma membranes [151] and from enzymic digests [152] of normal and ulcerated gastric mucosae of pig.

C. Studies on Model Glycopeptide Linkages.—Syntheses have been reported [153] for *N*-glycyl, *N*-L-alanyl, *N*-L-valyl, *N*-L-glutamyl, and *N*-L-seryl derivatives of 2-acetamido-2-deoxy-β-D-glucopyranosyl starting from the amino-sugar. Only the aspartyl derivative was hydrolysed by an amido hydrolase. First-order rate constants for the alkaline reductive cleavage of 4-*N*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-L-asparagine and 4-*N*-(β-D-glucopyranosyl)-L-asparagine have been reported [154] to be 6.44×10^{-3} and 5.97×10^{-3} min^{-1} respectively in 0.2M-NaOH at 100 °C. Approximately 50% of each compound was degraded to aspartic acid and reduced sugars. The mono- and di-saccharide derivatives (75) and (76) have been synthesized [155] as intermediates for glycopeptide synthesis. A number of *O*-glycosides of dipeptides containing β-hydroxyamino-acids as given in (77) have been prepared [156] by the reaction of the appropriate glycosidyl

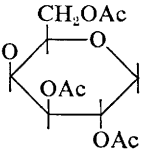




(77) R = CH·(CH$_2$)$_2$·CO$_2$Me, NH·CO$_2$·CMe$_3$; CH·CH$_2$OH, NHZ; CH·CHMe$_2$, NHZ

(75) R = H

(76) R = AcO

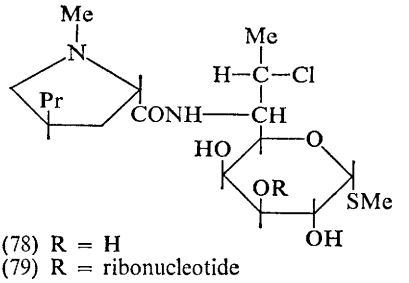


amino-acid derivative with an amino-acid derivative in aqueous pyridine–dicyclohexylcarbodi-imide at − 10 °C.

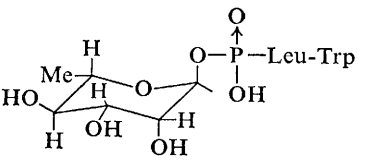
D. Other Carbohydrate-linked Compounds.—*Streptomyces coelicolor* transforms [157] clindamycin (78) into a ribonucleotide derivative (79). A new metabolite, phosphoramidon, isolated from a strain of *Streptomyces*

[151] A. J. Barber and G. A. Jamieson, *Biochemistry*, 1971, **10**, 4711.
[152] G. Pallavicini, G. Cetta, A. Quaroni, and A. Castellani, *Experientia*, 1971, **27**, 895.
[153] D. E. Cowley, L. Hough, and C. M. Peach, *Carbohydrate Res.*, 1971, **19**, 231.
[154] B. M. Austen and R. D. Marshall, *Biochem. J.*, 1971, **124**, 14P.
[155] M. Spinola and R. W. Jeanloz, *Carbohydrate Res.*, 1970, **15**, 361.
[156] V. A. Derevitskaya, I. M. Rotenberg, and N. K. Kochetkov, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1971, 2092.
[157] A. D. Argoudelis and J. H. Coats, *J. Amer. Chem. Soc.*, 1971, **93**, 534.

tanashiensis, has been identified[158] as (80). Mild acid hydrolysis of phosphoramidon yields L-leucyl-L-tryptophan.



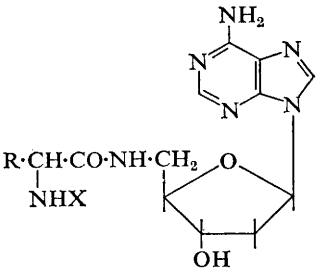
(78) R = H
(79) R = ribonucleotide



(80)

5 Peptides and Amino-acids Linked to Nucleosides and Nucleotides

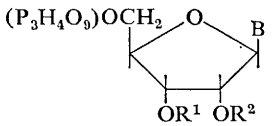
Several 5′-*N*-aminoacyl-5′-amino-5′-deoxy- and 5′-amino-2′,5′-dideoxy-9β-D-ribofuranosyl purine nucleoside peptides have been synthesized.[159] These represent a new class of peptide nucleosides of interest in studies of the nucleoside peptides of DNA and RNA. Peptide bond formation was achieved using the active ester and carbodi-imide methods, giving the compounds (81)—(84). ^{1}H N.m.r. data indicate that the phenyl and



(81) R = $PhCH_2$, X = H.
(82) R = $Me_2CH\cdot CH_2$, X = H
(83) R = H, X = H,
(84) R = H, X = $NH_2\cdot CH_2\cdot CO$

adenine rings are stacked over one another in (81). The imidazolide of *N*-benzyloxycarbonyl-L-Leu-Gly-Trp reacts readily[160] with the four main types of natural ribonucleoside-5′-triphosphates to give both the 3′- or 2′-*O*-peptidyl isomers (85) and (86). The *O*-peptidyl bond appears to be more stable than an *O*-amino-acyl bond when attached to the ribonucleosides.

[158] S. Umezawa, K. Tatsuta, O. Izawa, and T. Tsuchiya, *Tetrahedron Letters*, 1972, 97.
[159] M. J. Robins, L. N. Simon, M. G. Stout, G. A. Ivanovics, M. P. Schweizer, R. J. Rousseau, and R. K. Robins, *J. Amer. Chem. Soc.*, 1971, **93**, 1474.
[160] P. P. Purygin, A. A. Kraevskii, and B. P. Gottikh, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1970, 1369.

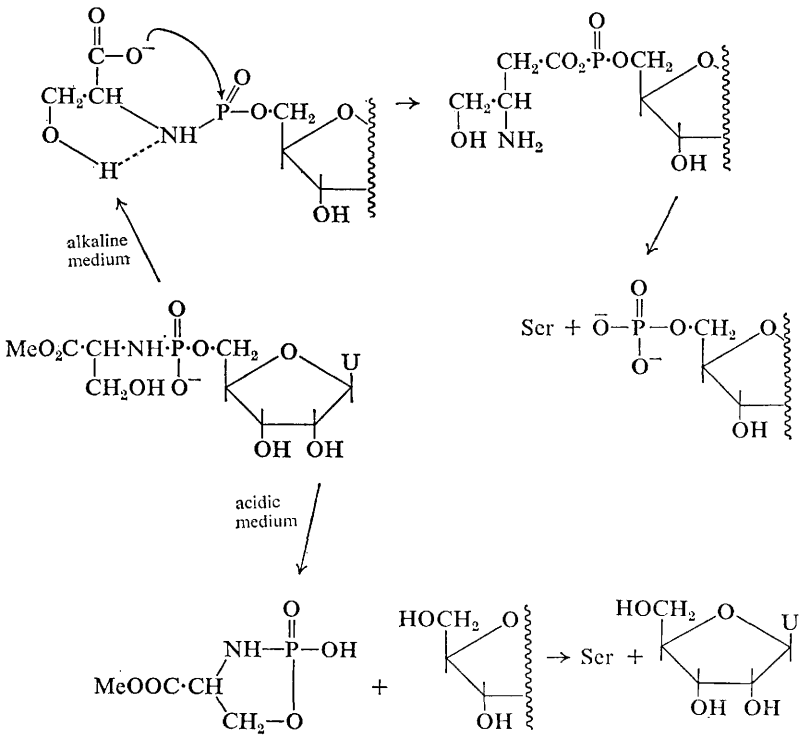


B = guanine or cytosine

(85) R^1 = Leu-Gly-Trp, R^2 = H
(86) R^1 = H, R^2 = Leu-Gly-Trp

The properties of phosphoramide nucleotide peptide bonds formed from serine are significantly different from the properties of this bond in derivatives of simple amino-acids. Hydrolysis studies[161] show that in acidic and alkaline media the reactions probably proceed according to the routes summarized in Scheme 3.

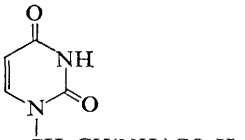
Several peptide derivatives based on the purine and pyrimidine systems have again been synthesized. Homo- and hetero-peptides based on



Scheme 3

[161] E. P. Savel'ev, E. S. Gromova, Z. A. Zhabarova, and M. A. Prokofiev, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1970, 1817.

willardiine (87) have been synthesized [162] using the mixed anhydride, active ester, and carbodi-imide methods for coupling. DL-Willardiyl-DL-willardiine was also prepared by these methods. Syntheses of β-pyridyl-α-alanine



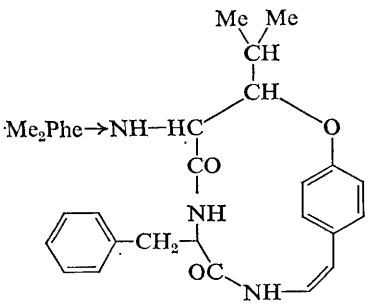
(87)

peptides,[163] α-amino-ε-(6-chloro-9-purinyl)caproic acid,[164] *N*-[(6-diphenylmethylthio)purin-9-yl]-, and *N*-[(6-thiopurin-9-yl)]-acetylamino-acids [165] have been reported.

6 Peptide Alkaloids

A comprehensive review dealing with this subject has recently been published.[166]

Another variation (88) on the *p*-alkoxystyrylamino-ring system has been identified,[167] using mass spectrometry, in the structure of scutianin B from *Scutia buxifolia* Reiss. A West Indian shrub, *Croton humilis* has yielded [168]



(88)

[162] M. Lidaks, R. Paegle, V. Straume, D. Snore, and Yu. P. Shvachkin, *Khim. geterotsikl. Soedinenii*, 1970, 1002; 1971, 404.

[163] B. L. Krainova, G. A. Agafonova, and E. S. Chapman, *Zhur. obshchei Khim.*, 1971, **41**, 1617; G. A. Agafonova, N. E. Gerasimova, M. V. Guseva, B. L. Krainova, T. V. Petrova, V. E. Pozdnev, and E. S. Chapman, *ibid.*, 1970, **40**, 2502.

[164] M. U. Lidak, Y. Y. Shluke, S. E. Poritere, and U. P. Shvachin, *Khim. geterotsikl Soedinenii*, 1971, 427.

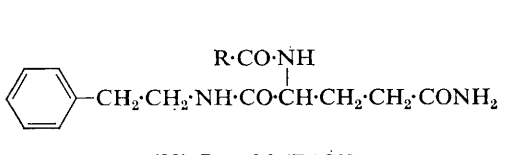
[165] E. D. Kaverzneva, V. V. Kiseleva, and L. I. Deeva, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1971, 2031.

[166] M. Pais and F.-X. Jarreau in 'Chemistry and Biochemistry of Amino-acids, Peptides, and Proteins', ed. B. Weinstein, Dekker, New York, 1971, vol. 1, p. 127.

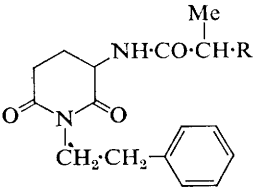
[167] R. Tschesche, E. Ammermann, and H.-W. Fehlaber, *Tetrahedron Letters*, 1971, 4405,

[168] J. P. Kutney, F. K. Klein, G. Knowles, and K. L. Stuart, *Tetrahedron Letters*, 1971, 3263; J. P. Kutney, F. K. Klein, G. Eigendorf, D. McNeill, and K. L. Stuart, *Tetrahedron Letters*, 1971, 4973.

a number of interesting peptidyl alkaloids (89)—(92). The α-glutamyl linkage in (89) and (90) was identified from the mass spectra and appears to be unique to this series. The glutarimide structures in (91) and (92) are also the first of this type to be isolated from higher plants.



(89) R = Me(Et)CH
(90) R = Me_2CH



(91) R = Me
(92) R = Et

7 Hydrazino Peptides

Negamycin, a new antibiotic isolated[169] from *Streptomyces purpeofuscus*, has been shown to have the structure (93). This novel structure has been confirmed by partial synthesis using the *N*-hydroxysuccinimide ester of di-*N*-Z-δ-*O*-tetrahydropyranyloxy-β-lysine with 1-methylhydrazine-acetic

(93)

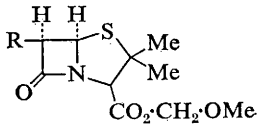
acid, $NH_2 \cdot NMe \cdot CH_2 \cdot CO_2H$. Selective acylation of the N^{α}-(with the acetyl group) and N^{β}-(with the benzyloxycarbonyl group) positions in α-hydrazino-β-phenylpropionic acid (NH_2-Phe) has provided[170] useful derivatives for the synthesis of hydrazine-containing peptides, as for example in the hetero-sequence Ala-L-NH-Phe-Ile-Gly-Leu-Met-NH_2, an analogue of the sequence found in eledoisin.

8 Penicillins and Cephalosporins

Once again the coverage of these compounds has been selective rather than comprehensive with the main emphasis on reactions involving the amide portion of the molecules rather than the heterocyclic aspects of their structures. A comprehensive review of the recent developments in the chemistry of the penicillins and cephalosporins (coverage up to 1970) has been published,[171] and the biochemistry and chemistry of the two series have been compared and contrasted in a symposium lecture.[172]

[169] S. Kondo, S. Shibahara, S. Takahashi, K. Maeda, H. Umezawa, and M. Ohno, *J. Amer. Chem. Soc.*, 1971, **93**, 6305.
[170] R. Grupe and H. Niedrich, *J. prakt. Chem.*, 1970, **312**, 1087.
[171] J. L. Lucke and G. Balavoine, *Bull. Soc. chim. France*, 1971, 2733.
[172] E. P. Abraham, Ref. 2, p. 399.

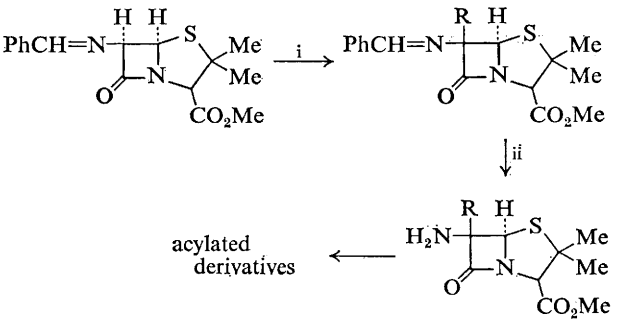
Epimerization at the C-6 position having a phthalimido side-chain proceeds *via* an enolate intermediate,[173] but in the penicillins, in which there is a secondary amide group present in that position, epimerization has proved difficult (Vol. 3, p. 315). Last year a breakthrough was made in the use of *NO*-bis(trimethylsilyl)acetamide (BSA) to induce epimerization, and further success using this reagent has now been achieved.[174] A mixture of BSA and 1,5-diazobicyclo[4,3,0]non-5-ene (DBN) gives 75% epimerization using either the esters or free acids of phenoxymethyl- and benzylpenicillin. Equilibration[175] of penicillanic acid derivatives (94) and (95)



(94) R = phthalimido
(95) R = 2-hydroxy-1-naphthylidenimino

using DBN in methylene chloride at room temperature gives complete conversion of the former into the 6α-isomer, but only 61% conversion in the latter case. It is therefore implied that the overwhelming thermodynamic preference for the α-isomer previously reported should not be assumed to apply to every penicillin derivative.

Attempts to substitute methyl groups in the 6-position of modified penicillins and the 7-position of cephalosporin have been accompanied by epimerization.[176] Alkylation of *N*-benzylidene-6-aminopenicillanic acid methyl ester with one equivalent of sodium hydride and excess of iodomethane gave a mixture of the 6-methyl epimers as shown in Scheme 4



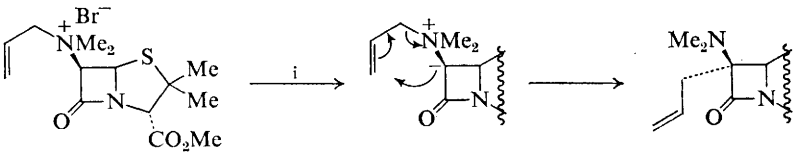
Conditions: i, NaH–MeI; ii, acid hydrolysis

Scheme 4

[173] B. G. Ramsey and R. J. Stoodley, *Chem. Comm.*, 1971, 450.
[174] A. Vlietinck, E. Roets, P. Claes, and H. Vanderhaeghe, *Tetrahedron Letters*, 1972, 285.
[175] J. R. Jackson and R. J. Stoodley, *Chem. Comm.*, 1971, 647.
[176] E. H. W. Bohme, H. E. Applegate, B. Toeplitz, J. E. Dolfini, and J. Z. Gougoutas, *J. Amer. Chem. Soc.*, 1971, **93**, 4324.

(R = Me). A similar reaction in the cephalosporin series gave a similar analogue. The methylated derivatives obtained were much less active than the unsubstituted analogues. In contrast, 6α-methyl (and -ethyl) penicillin G sodium salt and the cephalosporin analogues show appreciable activity.[177] It appears that the free carboxy-group is a requisite for activity, and the fact that these methyl-substituted analogues do show activity might be some support for the hypothesis that penicillins and cephalosporins react by mimicking D-Ala-D-Ala in enzyme transpeptidase reactions.

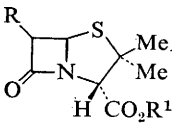
Epimerization at the penicillin C-6 position has been prevented in a rather specialized procedure[178] for alkylating this position as shown in Scheme 5. This is an interesting example of a carbanion centre retaining



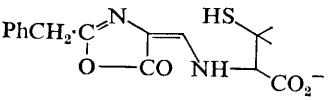
Conditions: i, NaH–Me_2N·CHO–benzene

Scheme 5

its configuration, which is probably the consequence of the adjacent geometry rather than being a general phenomenon. Benzyl 6-diazopenicillinate (96) has been converted[179] into 6-acetyl- and 6-phenylacetylhydrazonopenicillanic acids using triphenylphosphine in wet ether followed by acetylation. Stereoselective borohydride reduction gave the β-isomer (97) which is effective against a pencillin-resistant strain of *Staph. aureus*. Isomerization of benzylpenicillin into benzylpenicillenic acid (98) has been shown to be catalysed strongly by imidazole.[180]



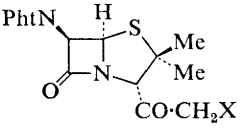
(96) R = N_2, R^1 = Bzl
(97) R = $R^2CO \cdot NH \cdot NH$, R^1 = Bzl



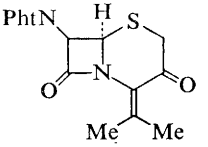
(98)

The economically important conversion of penicillins into cephalosporins continues to be a major field of research activity. Compounds such as (99) react[181] with the strong base DBN (1,5-diazabicyclo[4,3,0]-non-5-ene) to give (100) as a mixture of C-6 epimers. Replacement of the

[177] R. A. Firestone, N. Schelechow, D. B. R. Johnson, and B. G. Christensen, *Tetrahedron Letters*, 1972, 375.
[178] G. V. Kaiser, C. W. Ashbrook, and J. E. Baldwin, *J. Amer. Chem. Soc.*, 1971, **93**, 2342.
[179] D. M. Brunwin and G. Lowe, *Chem. Comm.*, 1972, 192.
[180] H. Bundgaard, *Tetrahedron Letters*, 1971, 4613.
[181] B. G. Ramsay and R. J. Stoodley, *J. Chem. Soc.* (*C*), 1971, 3859, 3864.

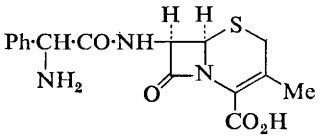


(99)

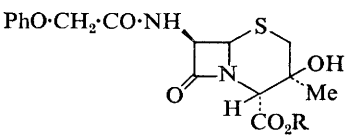


(100)

phthalimido-group by a phenoxyamido-group eliminates epimerization. More detailed analysis of the rearrangement of penicillin sulphoxide esters to desacetoxy-cephalosporins has been made [182] as a result of the discovery of the oral effectiveness of cephalexin (101). Acid-catalysed ring expansion of the penicillin sulphoxides yields (102), a novel class of 3-(*S*)-hydroxy-3-



(101)



(102)

methylcepham derivatives, as intermediates in these rearrangements. Complete details have now been published [183] of the preparation and rearrangements of 6β-phenylacetamido-penicillanic sulphoxides (103) into acetoxypenams and acetoxycephams. Thiols react [184] with the sulphoxide (103) to give (104; X = SSR) which, on treatment with trimethylphosphine, gives the thiol (104; X = SR). Similar results have been obtained [185] by treating an *N*-tritylpenicillin derivative with iodomethane in the presence of sodium hydride when a selective cleavage of the thiazolidine ring takes place giving (105). Successful attempts at removing the isopentenyl residue, as *e.g.* in (104), using diazomethane followed by t-butoxide,[186] and also for (106) [187] using permanganate or osmium tetroxide, have been reported. These methods for producing non-fused β-lactam compounds provide very interesting possibilities for future developments. A rather remarkable methylthio-group migration giving (107) occurs on treating (106) with lead tetra-acetate.[187]

[182] R. R. Chauvette, P. A. Pennington, C. W. Ryan, R. D. G. Cooper, F. L. Jose, I. G. Wright, E. M. Van Heyningen, and G. W. Huffman, *J. Org. Chem.*, 1971, **36**, 1259; G. E. Gutowski, C. M. Daniels, and R. D. G. Cooper, *Tetrahedron Letters*, 1971, 3429; G. E. Gutowski, B. J. Foster, C. J. Daniels, L. D. Hatfield, and J. W. Fisher, *ibid.*, 1971, 3433.

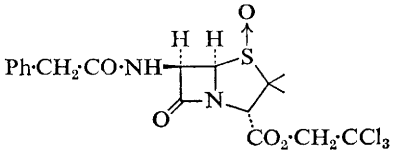
[183] D. H. R. Barton, F. Comer, D. G. T. Greig, P. G. Sammes, C. M. Cooper, G. Hewitt, and W. G. E. Underwood, *J. Chem. Soc.* (*C*), 1971, 3540.

[184] D. H. R. Barton, P. G. Sammes, M. V. Taylor, C. M. Cooper, G. Hewitt, B. E. Looker, and W. G. E. Underwood, *Chem. Comm.*, 1971, 1137.

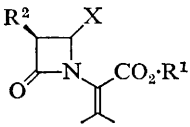
[185] J. P. Clayton, J. H. C. Nayler, R. Southgate, and P. Tolliday, *Chem. Comm.*, 1971, 590.

[186] D. H. R. Barton, D. G. T. Greig, P. G. Sammes, and M. V. Taylor, *Chem. Comm.*, 1971, 845.

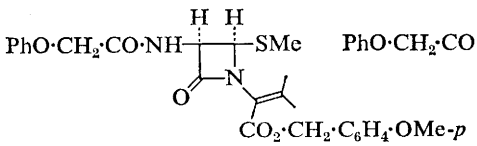
[187] E. G. Brain, A. J. Eglinton, J. H. C. Nayler, M. J. Pearson, and R. Southgate, *J.C.S. Chem. Comm.*, 1972, 229.



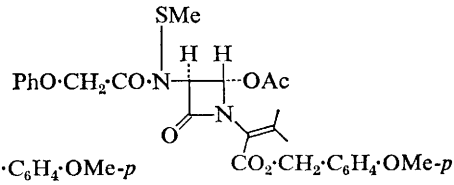
(103)



(104) $R^1 = CH_2 \cdot CCl_3$,
$R^2 = Ph \cdot CH_2 \cdot CO \cdot NH$
(105) $R^1 = p\text{-}MeO \cdot C_6H_4 \cdot CH_2$,
$R^2 = Trt\text{-}NH$, X = SMe

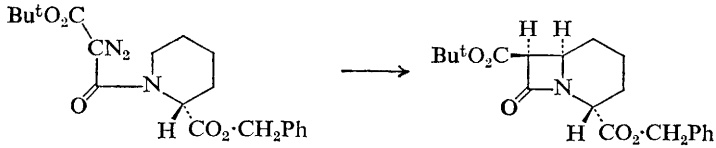


(106)



(107)

A preliminary report [188] has appeared on the synthesis of (+)-methyl-6-bromopenicillinate by a novel route using a mercury-induced carbene reaction. Analogues of the penicillins and cephalosporins have been synthesized [189] using the photolysis of the diazo-compound (108) as a key step in the reaction, but no antibacterial activity was found in the product.



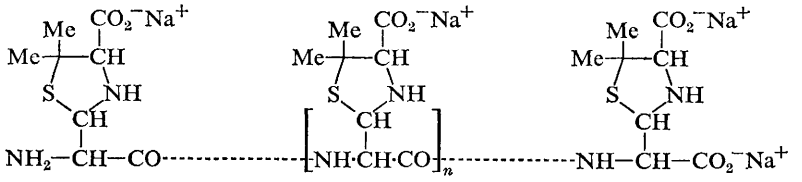
(108)

The key step in the formation of semisynthetic penicillins, namely the removal of the phenylacetyl group in penicillin G, has been accomplished [190] by means of phosphorus pentachloride and *N*-methylmorpholine followed by methanol. The phosphorus reagent with butanol at low temperatures is also capable of selective cleavage [191] of the amide bond in the silyl ester of benzylpenicillin. Semisynthetic penicillins and cephalosporins reported

[188] N. G. Johansson and B. Akermark, *Tetrahedron Letters*, 1971, 4785.
[189] D. M. Brunwin, G. Lowe, and J. Parker, *J. Chem. Soc.* (*C*), 1971, 3756; *Chem. Comm.*, 1971, 865.
[190] G. R. Fosker, K. D. Hardy, J. H. C. Nayler, P. Seggery, and E. R. Stove, *J. Chem. Soc.* (*C*), 1971, 1917, 1920.
[191] H. W. O. Weissenburger and M. G. Van der Hoeven, *Rec. Trav. chim.*, 1970, **89**, 1081.

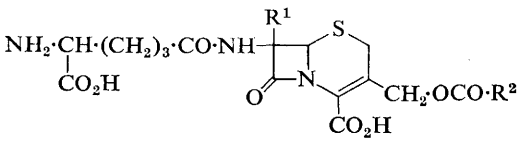
during the year include compounds derived from D-2-(1,4-cyclohexadienyl)-glycine,[192] 3-methoxymethyl-7-acylaminocephalosporin and derivatives,[193] and 2-thiomethyl- and 2-thiomethylene-cephalosporins.[194]

Successful attempts,[195] using Sephadex G-25, have been made to separate the polymers present in aqueous solutions of 6-aminopenicillanic acid, benzylpenicillin, ampicillin, and hetacillin. The evidence pointed to a general structure (109) for the polymers. Penicillin N and three new



(109)

cephalosporins (110)—(112) have been isolated[196] from *Streptomyces lipmanii* and *S. clavuligerus*. Compounds (110) and (112) exhibited greater

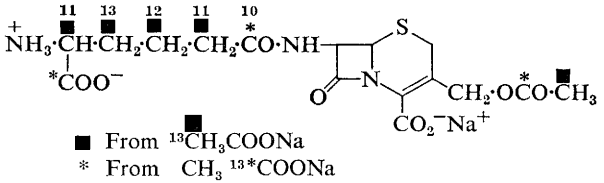


(110) R^1 = OMe, R^2 = Me
(111) R^1 = H, R^2 = NH_2
(112) R^1 = OMe, R^2 = NH_2

activity against Gram-negative organisms than does cephalosporin C. ^{13}C N.m.r. spectroscopic analysis[197] of the sodium salt of cephalosporin C obtained from cultures of *Cephalosporium acremonium* fed with sodium [1-^{13}C]- and [2-^{13}C]-acetate revealed enhanced ^{13}C-chemical shift intensities as shown in (113). Of particular importance is the distribution of labels in C-11, C-12, and C-13 in that it corresponds to the labelling expected for the formation of α-aminoadipic acid *via* the Krebs cycle.

[192] J. E. Dolfini, H. E. Applegate, G. Bach, H. Basch, J. Bernstein, J. Schwartz, and F. L. Weisenborn, *J. Medicin. Chem.*, 1971, **14**, 117.
[193] J. A. Webber, G. W. Huffman, R. E. Koehler, C. F. Murphy, C. W. Ryan, E. M. Van Heyningen, and R. T. Vasileff, *J. Medicin. Chem.*, 1971, **14**, 113.
[194] G. V. Kaiser, C. W. Ashbrook, T. Goodson, I. G. Wright, and E. M. Van Heyningen, *J. Medicin. Chem.*, 1971, **14**, 420, 426.
[195] H. Smith and A. C. Marshall, *Nature*, 1971, **232**, 45.
[196] R. Nagarajan, L. D. Boeck, M. Gorman, R. L. Hamill, C. E. Higgins, M. M. Hoehn, W. M. Stark, and J. G. Whitney, *J. Amer. Chem. Soc.*, 1971, **93**, 2308.
[197] N. Neuss, C. H. Nash, P. A. Lemke, and J. B. Grutzner, *J. Amer. Chem. Soc.*, 1971, **93**, 2337.

The effect of solvents on the ^1H n.m.r. chemical shifts of methyl esters of benzyl- and phenoxymethyl-penicillins has been reported.[198] A detailed



■ From $^{13}CH_3COONa$
* From $CH_3\ ^{13}*COONa$

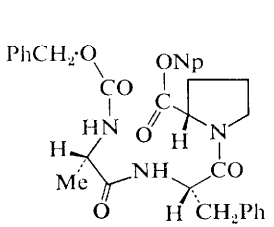
(113)

investigation[199] of the u.v. and c.d. spectra of a number of cephalosporins reveals an overlap between the π-system of the 3,4-double bond and the p- and π-orbitals of the β-lactam carbonyl group.

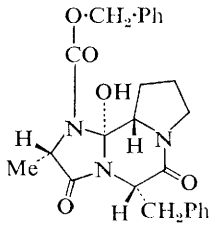
A new synthesis of the β-lactam unit has been achieved[200] by enlargement of the cyclopropanone ring.

9 Miscellaneous

The formation of cyclols due to amide–amide interaction, a subject of some discussion for a number of years, has been achieved for the first time[201] with the synthesis in 70% yield of (115) from (114) using a mildly alkaline



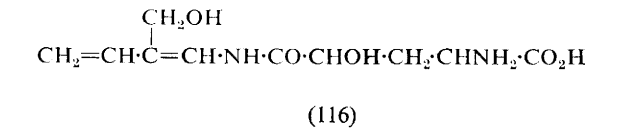
(114)



(115)

medium. The structure of cyclol (115) has been confirmed[202] by X-ray crystallography. The formation of the cyclol could be envisaged as occurring either *via* an acylalanyl dioxopiperazine or a cyclopeptide-type interaction. A novel amino-acid amide having the structure (116) has been found[203] in acid hydrolysates of seed meals from *Staphylea pinnata*.

[198] G. U. Pek, V. F. Bystrov, and I. N. Blinova, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1971, 1694.
[199] N. Nagarajan and D. O. Spry, *J. Amer. Chem. Soc.*, 1971, **93**, 2310.
[200] H. H. Wasserman, H. W. Adickes, and O. E. de-Ochoa, *J. Amer. Chem. Soc.*, 1971, **93**, 5586.
[201] G. Lucente and A. Romeo, *Chem. Comm.*, 1971, 1605.
[202] S. Cerrini, W. Fedeli, and F. Mazza, *Chem. Comm.*, 1971, 1607.
[203] M. D. Grove, M. E. Daxenbichler, D. Weisleder, and C. H. Van Etten, *Tetrahedron Letters*, 1971, 4477.



(116)

A preliminary degradative study [204] on the antibiotic complex S-520 from *Streptomyces diastaticus* shows that there is a mixture of peptides present which contains glycine, D-valine, ornithine and D-isoleucine, and L-threo-β-hydroxyglutamic acid, as well as previously unknown amino-acids. The

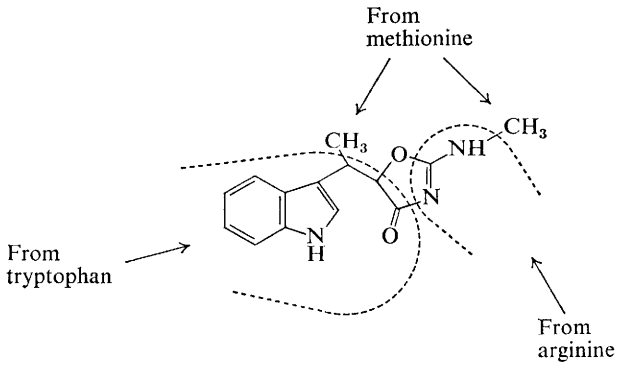


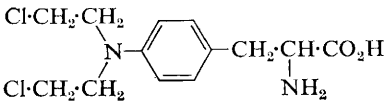
Figure 4

presence of D-isoleucine is interesting since it is only the second time that this isomer has been isolated from micro-organisms. Labelling studies [205] have shown that indolmycin is formed from the precursors shown in Figure 4. A biosynthetic pathway, involving an intermediate on the route to melanin, has at last been elucidated [206] for the bioconversion of tyrosine into the propylproline residue of lincomycin.

Homologues of D- and L-3-hydroxyacylhexapeptides have been prepared [207] by the azide coupling of Gly-Ser-Val-Thr-Leu with D- or L-3-hydroxyacyl-leucine hydrazides, but none of the compounds was identical with viscosin. An analogue of bradykinin, containing *N*-(2-aminoethyl) glycine instead of prolyl-glycyl residues, has been synthesized [208] but was

[204] J. Shoji, S. Kozuki, M. Mayama, and N. Shimaoka, *J. Antibiotics* (*Japan*), 1970, **23**, 429; J. Shoji and R. Sakazaki, *ibid.*, 1970, **23**, 432, 418.
[205] U. Hornemann, L. H. Hurley, M. K. Speedie, and H. G. Floss, *J. Amer. Chem. Soc.*, 1971, **93**, 3028.
[206] D. F. Witz, E. J. Hessler, and T. L. Miller, *Biochemistry*, 1971, **10**, 1128.
[207] M. Hiramoto, K. Okada, S. Nagai, and H. Kawamoto, *Chem. and Pharm. Bull.* (*Japan*), 1971, **19**, 1308.
[208] E. Atherton, H. D. Law, S. Moore, D. F. Elliott, and R. Wade, *J. Chem. Soc.* (*C*), 1971, 3393.

found to be inactive. Peptides of sarcolysin (117) with γ-aminobutyric acid and containing *C*- and *N*-terminal sarcolysin have been synthesized.[209]



(117)

The eleven peptide hydrogens in bacitracin A form[210] at least three distinct kinetic classes, including a single slow-exchanging proton—probably implying a single hydrogen bond in the macrocyclic ring.

[209] A. A. B. Paylyukonis, K. I. Karpavichus, O. V. Kilkisheva, and I. L. Knunyants, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1970, 161.
[210] R. E. Galardy, M. P. Printz, and L. C. Craig, *Biochemistry*, 1971, **10**, 2429.

5

Further Extracts from the Rules and Tentative Rules of the I.U.P.A.C.–I.U.B. Commission on Biochemical Nomenclature

In Volume 2, Chapter 5, we reprinted the following extracts from the I.U.P.A.C.–I.U.B. Recommendations relevant to amino-acids, peptides, and proteins:
Tentative Abbreviated Designation of Amino-acid Derivatives and Peptides; Tentative Rules for Naming Synthetic Modifications of Natural Peptides; Tentative Abbreviated Nomenclature of Synthetic Polypeptides (Polymerised Amino-acids). We reprint here the 1971 revision of the first-named recommendations, together with the new recommendations *Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains.*

I Symbols for Amino-Acid Derivatives and Peptides Recommendations (1971)

(*Reprinted by permission from* I.U.P.A.C. Information Bulletin, Appendices on Tentative Nomenclature, Symbols, Units, and Standards, No. 23, June 1972)

The revised Tentative Rules published by CBN in 1966 (1) were an attempt to achieve a broad systematization of various types of abbreviated notation already in use [e.g., Brand & Edsall (1947) *Annu. Rev. Biochem.; Report of the Committee on Abbreviations of the American Society of Biological Chemists*, 18 *December* 1959; *Report of the Committee on Nomenclature of the European Peptide Symposium* (1963) pp. 261–269, Pergamon Press, Oxford; 'Tentative Rules for Abbreviations and Symbols of Chemical Names of Special Interest in Biological Chemistry' (2)]. They sought to reconcile the needs of the protein chemist, i.e., indication of amino-acid sequences, with those of persons concerned more with the chemical reactions of proteins and the synthesis of polypeptides, i.e., the need for conveying more detailed chemical information in abbreviated form.

Recent progress in the field of peptide synthesis and in the chemical modification of proteins has made necessary a revision of these Tentative Rules. This revision has been aided by the work of an expert group consisting of J. S. Fruton, B. S. Hartley, R. R. Porter, J. Rudinger, R. Schwyzer and G. T. Young. They are greatly indebted to many colleagues, notably W. H. Stein, for helpful suggestions.

1. General Considerations

1.1

The symbols chosen are derived from the trivial names or chemical names of the amino acids and of chemicals reacting with amino acids and polypeptides. For the sake of clarity, brevity and listing in tables, the symbols for amino-acid residues have been, wherever possible, restricted to three letters, usually the first letters of the trivial names.

1.2

The symbols represent not only the names of the compounds but also their structural formulae.

1.3

The amino-acid symbols by themselves represent the amino acids. The use of the symbols to represent the free amino acids is *not* recommended in textual material, but such use may occasionally be desirable in tables, diagrams or figures. Residues of amino acids are represented by addition of hyphens in specific positions as indicated in Section 3.

1.4

Heteroatoms of amino-acid residues (e.g., O^3 and S^3 of serine and cysteine respectively, N^6 of lysine, N^2 of glycine etc.) do not explicitly appear in the symbol; such features are understood to be encompassed by the abbreviation.

1.5

Amino-acid symbols denote the L configuration unless otherwise indicated by D or DL appearing before the symbol and separated from it by a hyphen. When it is desired to make the number of amino-acid residues appear more clearly, the hyphen between the configurational prefix and the symbol may be omitted (see 6.3.1.1 *et seq.*). (Note: The designation of an amino-acid residue as DL is inappropriate for compounds having another amino-acid residue with an asymmetrical centre.)

1.6

Structural formulae of complicated features may be used along with the abbreviated notation wherever necessary for clarity.

1.7

All symbols listed below are to be printed or typed as one capital letter followed by two lower-case letters, e.g., Gln, not GLN or gln or GlN or glN, regardless of position in a sentence or structure. However, when used for purposes other than to represent an amino-acid residue (e.g., to designate a genetic factor), three lower-case italic letters (i.e., *gln*) should be used.

2. Symbols for Amino Acids

2.1 *Common amino acids*

Alanine	Ala	Leucine	Leu
Arginine	Arg	Lysine	Lys
Asparagine	Asn[2]	Methionine	Met
Aspartic acid	Asp	Phenylalanine	Phe
Cysteine	Cys	Proline	Pro
Glutamic acid	Glu	Serine	Ser
Glutamine	Gln[2]	Threonine	Thr
Glycine	Gly	Tryptophan	Trp
Histidine	His		(not Try)
Isoleucine	Ile	Tyrosine	Tyr
		Valine	Val

[2] Asparagine and glutamine may also be denoted as Asp(NH_2) or Asp and Glu(NH_2) or Glu, respectively, if necessary (as when the NH_2 is substituted, or its removal or modification is under discussion). See 4.2.

Glx may be used when the residue denoted could be 'glutamic acid or glutamine'; similarly, Asx for 'aspartic acid or asparagine'.

2.2 *Less-common amino acids*

Symbols for less-common amino acids should be defined in each publication in which they appear. The following principles and notations are recommended.

2.2.1 *Hydroxyamino acids.*

		Preferred alternatives	
5-Hydroxylysine	5Hyl	Lys(5OH) or	Lys[5]OH
3-Hydroxyproline	3Hyp	Pro(3OH) or	OH[3]Pro
4-Hydroxyproline	4Hyp	Pro(4OH) or	Pro[4]OH

2.2.2 *allo-Amino acids.*

allo-Isoleucine	*a*Ile	
allo-Hydroxylysine	*a*Hyl	*a*Lys(5OH) or OH[5]*a*Lys

2.2.3 *'Nor' and 'homo' amino acids.* 'Nor' (e.g., in norvaline) is not used in its accepted sense (denoting a lower homologue) but to change the trivial name of a branched-chain compound into that of a straight-chain compound (compare with 'iso', paragraph 2.1). 'Nor' should therefore be treated as part of the trivial name without special emphasis. 'Homo', used in the

sense of a higher homologue, may also be incorporated into the trivial name.

Norvaline	Nva	Homoserine	Hse
Norleucine	Nle	Homocysteine	Hcy

2.2.4 *Higher unbranched amino acids.* The functional prefix 'amino' is included in the symbol as the letter 'A', diamino as 'A_2'[3]. The trivial name of the parent acid is abbreviated to two letters. The word 'acid' ('-säure' etc.) is omitted from the symbol as carrying no significant information. Unless otherwise indicated, single groups are in the 2-position, two amino groups in the 2- and terminal positions (monocarboxylic acids) or 2- and 2′-positions (dicarboxylic acids). The location of amino groups in positions other than these is shown by appropriate prefixes.

Examples:

2-Aminobutyric acid	Abu
2-Aminoadipic acid	Aad
2-Aminopimelic acid	Apm
2,4-Diaminobutyric acid	A_2bu[3]
2,2′-Diaminopimelic acid	A_2pm[3]
2,3-Diaminopropionic acid	A_2pr[3] or Ala(3NH_2) or ^{3}Ala (see 4.3)
β-Alanine	βAla
Ornithine (2,4-diaminovaleric acid)	Orn
6-Aminohexanoic acid	εAhx[4]
3-Aminoadipic acid	βAad

[3] The symbols for diamino compounds previously (1) utilized the letter 'D' for 'diamino'. However, the overuse of 'D' as the initial letter for many compounds beginning with 'di' (and of 'T' for 'tri' and 'tetra'), in addition to the fact that standard chemical symbolism utilizes subscript numerals for multipliers, leads to the proposal that diamino should be represented by A_2. This eliminates the ambiguity attached to 'D' and makes more clear the chemical relationship between the diamino and monoamino derivatives. It is in keeping with the increasing use of Me_2SO instead of DMSO and of Me_3Si- in place of TMS-, and with the earlier proposal of H_4 for tetrahydro (4).

[4] Recommended in place of the previous (1) εAcp, in which 'cp' for caproic may be confused with capric and caprylic.

2.2.5 *N^3-Alkylated amino acids.* N^3-Alkylamino acids are becoming more and more common (e.g., in the large group of depsipeptides). This justifies special symbols.

Examples:

N-Methylglycine (sarcosine) (see 6.2)	MeGly or Sar
N-Methylisoleucine	MeIle
N-Methylvaline etc.	MeVal etc.
N-Ethylglycine etc.	EtGly etc.

2.3 *Non-amino-acid residues linked to peptides*

For residues of muramic acid, sialic acid, neuraminic acid etc. linked to amino-acid residues, as in bacterial-cell-wall components, the symbols Mur, Sia, Neu etc. (preceded by Ac if *N*-acetylated) are recommended. The symbols for sugar residues (Glc, Gal etc.) (2) and nucleotides (Ado, Cyd etc.) (3) may also be used.

3. Amino-Acid Residues

The links between residues have frequently been shown by peptide chemists as full points (periods, dots: ·) and by carbohydrate chemists (generally) as short strokes (dashes, hyphens: -). At times, special symbols have been used (> or →) to show the direction of what is in all cases an unsymmetrical link (peptide or glycoside).

For consistency and ease of typing as well as economy in printing, the hyphen, representing the peptide bond, should be the standard connecting symbol (2).

The simple usage by which Gly-Gly-Gly stands for glycylglycylglycine appears to involve the employment of the same three letters (Gly) for three different residues or radicals (*b*), (*c*), (*d*) below. However, if the dashes or hyphens are considered as part of each symbol, we have four distinct forms, for the free amino acid and the three residues, viz.:

(*a*) Gly = $NH_2—CH_2—CO_2H$ the free amino acid
(*b*) Gly- = $NH_2—CH_2—CO—$ the left-hand unit
(*c*) -Gly- = $—NH—CH_2—CO—$ the middle unit
(*d*) -Gly = $—NH—CH_2—CO_2H$ the right-hand unit

For peptides, a distinction may be made between the *peptide*, e.g., Gly-Glu (shown *without* dashes at the ends of the symbols), and the *sequence*, e.g., -Gly-Glu- (shown *with* dashes at the ends of the symbols).

3.1 *Lack of hydrogen on the 2-amino group*

The 2-amino group is understood to be at the left-hand side of the symbol when hyphens are used, and — in special cases — at the point of the arrow when arrows are used to indicate the direction of the peptide bond (-CO → NH-, -NH ← CO-). (For substitution for 2-amino hydrogen, see 4.1.)

Examples:

-Gly $—HN—CH_2—CO_2H$

> Gly or —Gly: $\rangle N—CH_2—CO_2H$

-Ala: $—HN—\underset{\displaystyle CH_3}{\underset{|}{CH}}—CO_2H$

>Ala or ⊥Ala: $\rangle N—\underset{\displaystyle CH_3}{\underset{|}{CH}}—CO_2H$

3.2 *Lack of hydroxyl on the 1-carboxyl group*

The 1-carboxyl group is understood to be on the right-hand side of the symbol when hyphens are employed and — in such special cases as 6.3.1.3 — at the tail of the arrow when arrows are used to indicate the direction of the peptide bond (-CO → NH-, -NH ← CO-).

Example:

Gly-: $H_2N—CH_2—CO—$.

It is generally convenient to use the same abbreviated formula for a polypeptide no matter what its state of ionization. To show that a peptide is acting as a cation or anion, the amino-terminal and carboxyl-terminal ends of the peptide are amplified with H and OH respectively (I); these may be modified to show the appropriate state of ionization (II or III).

H-Gly-Val-Thr-OH		or Gly-Val-Thr	(I)
$^+H_2$-Gly-Val-Thr-OH	or	$^+$HGly-Val-Thr	(II)
H-Gly-Val-Thr-O$^-$		or Gly-Val-ThrO$^-$	(III)

3.3 *Lack of hydrogen on amino, imino, guanidino, hydroxyl and thiol functions in the side chain*

(For substitution in such positions, see 4.2.)

Lys or Lys: $H_2N—CH(—(CH_2)_4—NH—)—CO_2H$

His or His:[5] $H_2N—CH(—CH_2—\pi N—CH=N—)—CO_2H$ or $H_2N—CH(—CH_2—N=CH—\tau N—)—CO_2H$

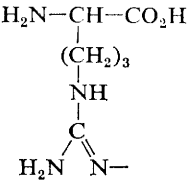
Trp or Trp: $H_2N—CH(—CH_2—\text{indole}—)—CO_2H$

[5] The prolonged and well-entrenched ambiguity in the nomenclature of the *N*-methylhistidines (the chemist's N-1 being the biochemist's N-3 and vice versa) leads to the proposal that a new trivial system for designating these substances is necessary. It is therefore proposed that the imidazole N *near*er the alanine residue be designated *pros* (symbol π) and the one *far*ther *tele* (symbol τ), to give the following names and symbols:

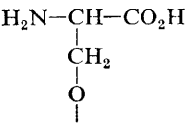
*pros*methylhistidine or N^{π}-methylhistidine, His(πMe);
*tele*methylhistidine or N^{τ}-methylhistidine, His(τMe).

See also 4.2, final example.

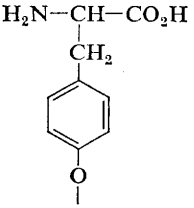
Arg or Arg:



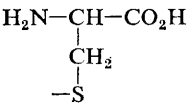
Ser or Ser:

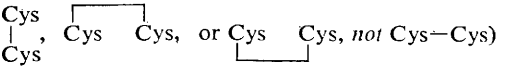


Tyr or Tyr:



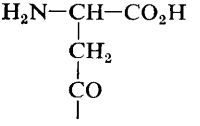
Cys or Cys:


('half-cystine')

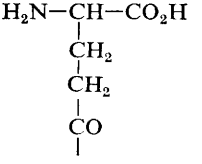
(Cystine would be:  , *not* Cys—Cys)

3.4 *Lack of hydroxyl on carboxyl groups in the side chain*

Asp or Asp:



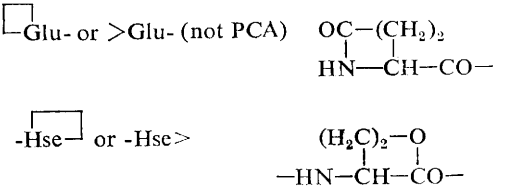
Glu or Glu:



3.5 *Cyclic derivatives of amino acid residues*

For the special cases of the residues derived from pyrrolid-2-one-5-carboxylic acid (also known as pyroglutamic acid) and from homoserine

lactone, the following are recommended:

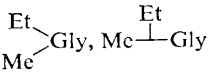


4. Substituted Amino Acids

4.1 *Substitution in the* 2*-amino and* 1*-carboxyl groups*

This follows logically from 3.1 and 3.2. The following examples will make the usage clear. (See also 6.2.)

N-Acetylglycine	Ac-Gly
Glycine ethyl ester	Gly-OEt
N^2-Acetyllysine	Ac-Lys
Serine methyl ester	Ser-OMe
O^1-Ethyl *N*-acetylglutamate	Ac-Glu-OEt
Isoglutamine	Glu-NH_2
O^1-Methyl hydrogen aspartate	Asp-OMe
N-Ethyl-*N*-methylglycine	Et-MeGly,



4.2 *Substitution in the side chain*

Side-chain substituents may be portrayed above or below the amino-acid symbol (see 3.3 and 3.4), or by placing the symbol for the substituent in parentheses immediately after the amino-acid symbol.

The use of parentheses should be reserved for a *single* symbol denoting a side-chain substituent. When a more complex substituent is involved, it is recommended that the vertical stroke and the two-line abbreviation be used (5). In general, the one-line abbreviation should be used only when the structure of a substituted peptide is given in textual material.

O^4-Methyl hydrogen aspartate	OMe Asp or Asp or Asp(OMe) OMe
O^5-Ethyl hydrogen *N*-acetylglutamate	Ac-Glu(OEt)
N^6-Acetyllysine	Ac Lys or Lys or Lys(Ac) Ac

O^3-Acetylserine	Ac | Ser or Ser or Ser(Ac) | Ac
O^4-Methyltyrosine	Me | Tyr or Tyr or Tyr(Me) | Me
S-Ethylcysteine	Et | Cys or Cys or Cys(Et) | Et
S-Sulfocysteine (*S*-cysteinesulfonic acid)	SO_3H Cys | or | Cys SO_3H or $Cys(SO_3H)$
Cysteinesulfinic acid	OH Cys | or | Cys OH or Cys(OH)
Cysteinesulfenic acid	O_2H Cys | or | Cys OH_2 or $Cys(O_2H)$
Cysteic acid (3-sulfoalanine)	O_3H Cys | or | Cys O_3H or $Cys(O_3H)$
S-Cyanocysteine	CN Cys | or | Cys CN or Cys(CN)
Methionine sulfoxide	O | Met or Met or Met(O) | O
Methionine sulfone	O_2 | Met or Met or $Met(O_2)$ | O_2
O^3-Phosphonoserine (phosphoserine)	P | Ser or Ser or Ser(P) | P
N^{τ}-Methylhistidine[5] (see 3.3) (*tele*methylhistidine)	Me τ| His or His or His(τMe) τ| Me

similarly for N^{π}-substitution (*pros*methylhistidine).

4.3 *Substitution on carbon side chain*

This may use the same convention as in 4.2, with the addition of locant numerals where necessary, e.g.:

3-Nitrotyrosine	NO_2 (3) Tyr or Tyr or Tyr(3NO_2) (3) NO_2
2,3-Diaminopropionic acid (see 2.2.4) (3-aminoalanine)	NH_2 (3) Ala or Ala (3) NH_2 or Ala(3NH_2)
Di-iodotyrosine	Tyr(I_2)

5. Symbols for Substituents

Groups substituted for hydrogen or for hydroxyl may be indicated either by their structural formulae or by symbols or by combinations of both, e.g.:

Benzoylglycine (hippuric acid)	-CO-Gly* or C_6H_5CO-Gly or [6]Bz-Gly or PhCO-Gly
Glycine methyl ester	Gly-OCH_3 or Gly-OMe
Trifluoroacetylglycine	CF_3CO-Gly

Suggestions for symbols designating substituent (or protecting) groups common in polypeptide and protein chemistry follow.

5.1 *N-Substituents (protecting groups) of the urethane type*

Benzyloxycarbonyl-	Z- or Cbz-
p-Nitrobenzyloxycarbonyl-	Z(NO_2)-
p-Bromobenzyloxycarbonyl-	Z(Br)-
p-Methoxybenzyloxycarbonyl-	Z(OMe)-
p-Methoxyphenylazobenzyloxycarbonyl-	Mz-
p-Phenylazobenzyloxycarbonyl-	Pz-
t-Butoxycarbonyl-	Boc- or Bu^tOCO-
Cyclopentyloxycarbonyl-	Poc- or *c*PeOCO-

5.2 *Other N-substituents*

Acetyl-	Ac-
Benzoyl- (C_6H_5—CO—)	PhCO- or Bz-
Benzyl- (C_6H_5—CH_2—)	$PhCH_2$- or [6]Bzl
Benzylthiomethyl-	$PhSCH_2$- or Btm-†
Carbamoyl-	NH_2CO- (preferred to Cbm)
1-Carboxy-2-nitrophenyl-5-thio-	[7]Nbs-

[6] Bz- is the symbol generally used for *benzoyl* in organic chemistry. It should not be used for *benzyl* ($C_6H_5CH_2$- or $PhCH_2$-), for which the symbol is Bzl-. However, $PhCH_2$- is unambiguous.

[7] See Comment following 5.3.

* *Senior Reporter's note:* In this formulation the phenyl group was to be repeated by its strnctural formulae.

† *Senior Reporter's nnte:* Benzylthiomethyl is $PhCH_2 \cdot SCH_2^-$. It is used as an S-, not N-substituent. Section 5.2 includes groups used for O- and S-protein.

3-Carboxypropionyl- ($HO_2C—CH_2—CH_2—CO—$)[8]	Suc-
Dansyl- (5-dimethylaminonaphthalene-1-sulfonyl)	Dns-
Dinitrophenyl-	$^9N_2$ph- or Dnp
Formyl-	HCO- or CHO-
p-Iodophenylsulfonyl- (pipsyl)	Ips
Maleoyl- ($—OC—CH=CH—CO—$)	-Mal- or Mal <
Maleyl- ($HO_2C—CH=CH—CO—$)	Mal-
Methylthiocarbamoyl-[10]	MeNHCS- or ^{10}Mtc-
o-Nitrophenylthio-	Nps-
Phenylthiocarbamoyl-[10]	PhNHCS- or ^{10}Ptc
Phthaloyl-	-Pht- or Pht <
Phthalyl-	Pht-
Succinyl-[11] ($—OC—CH_2—CH_2—CO—$)	-Suc- or Suc <
Tetrahydropyranyl-	H_4pyran- (preferred to Thp[9])
Tosyl- (*p*-tolylsulfonyl)	Tos-
Trifluoroacetyl-	$^9CF_3CO$-
Trityl- (triphenylmethyl)	Ph_3C- or Trt-

5.3 *Substituents at carboxyl group*

Benzyloxy- (benzyl ester)	-OCH_2Ph or -OBzl
Cyanomethoxy-	-OCH_2CN
Diphenylmethoxy- (benzhydryl ester)	-$OCHPh_2$ or -OBzh
Ethoxy- (ethyl ester)	-OEt
Methoxy- (methyl ester)	-OMe
p-Nitrophenoxy- (*p*-nitrophenyl ester)	-ONp
p-Nitrophenylthio-	-SNp
Phenylthio- (phenylthiol ester)	-SPh
1-Piperidino-oxy-	-OPip
8-Quinolyloxy-	-OQu
Succinimido-oxy-	-ONSu
t-Butoxy- (*t*-butyl ester)	-OBu^t

[8] Not succinyl, although it is the univalent radical of succinic acid. See succinyl and footnote 11.

[9] The use of 'D' for 'di' and 'T' for 'tri' or 'tetra' (and 'DH' and 'TH' for 'dihydro' and 'tetrahydro' respectively) is discouraged. Recognized symbols and subscripts are recommended. See also footnote 3.

[10] The symbol Pth has been used to denote a phenylthiohydantoin (e.g., Pth-Leu). Since this incorrectly implies the substitution of an amino acid by a 'phenylthiohydantoyl' group, it is suggested that the abbreviated symbol for such compounds be of the type CS-Leu-NPh or PhNCS-Leu (or Leu > PhNCS in textual meterial).

[11] Not succinoyl (6).

Comment

Many reagents used in peptide and protein chemistry for the modification (protection) of amino, carboxyl and side-chain groups in amino-acid residues have been designated by a variety of acronymic abbreviations, too numerous to be listed here. Extensive and indiscriminate use of such abbreviations is discouraged, especially where the accepted trivial name of

a reagent is short enough, e.g., tosyl chloride, bromosuccinimide, trityl chloride, dansyl chloride etc., or may be formulated in terms of the group transferred, e.g., $^{9}N_2$ph-F instead of FDNB for 1-fluoro-2,4-dinitrobenzene, Dns-Cl or dansyl-Cl in place of DNS, $^{9}Nbs_2$ in place of DTNB for 5,5′-dithiobis-(2-nitrobenzoic acid) (Ellman's reagent), $(Pr^iO)_2PO$-F, $Pr^i{}_2P$-F, iPr_2P-F, or Dip-F[3, 9] instead of DFP for di-isopropylfluorophosphate. Other commonly used substances that may be expressed more clearly in terms of symbols are MalNEt (instead of NEM) for *N*-ethylmaleimide, Tos-PheCH_2Cl (instead of TPCK) for L-1-tosyl-amido-2-phenylethyl chloromethyl ketone, Tos-Arg-OMe (instead of TAME) for tosyl-L-arginine methyl ester, Me_3Si- (instead of TMS-) for trimethylsilyl, CF_3CO- (instead of TFA) for trifluoroacetyl (see 5.2), H_4furan (instead of THF) etc. (See also footnotes 3 and 9.)

Some additional symbolic terms for substituents (and reagents), as examples, are:

2-Aminomethyl-	$-(CH_2)_2NH_2$ (preferred to Aet)
Carbamoylmethyl-	$-CH_2CONH_2$ (preferred to Cam)
Carboxymethyl-	$-CH_2CO_2H$ (preferred to Cm)
Chloroethylamine	$Cl(CH_2)_2NH_2$
Ethyleneimine	$(CH_2)_2NH$
Chloroacetamide	$ClCH_2CONH_2$
Chloroacetic acid	$ClCH_2CO_2H$
p-Carboxyphenylmercuri-	-HgBzOH
p-Chloromercuribenzoate	$pCl\text{-}HgBzO^-$
Diazoacetyl-	N_2CHCO-
Hydroxyethyl-	$-(CH_2)_2OH$
Ethylene oxide	$(CH_2)_2O$

6. Polypeptides

6.1 *Polypeptide chains* (5)

Polypeptides may be dealt with in the same manner as substituted amino acids, e.g.:

Glycylglycine — Gly-Gly

N-α-Glutamylglycine — Glu-Gly

N-γ-Glutamylglycine — Glu(—Gly) or Glu(—Gly)

or Glu(—Gly) or Glu(Gly)

Glutathione — Glu(—Cys-Gly) or Glu(γ—Cys-Gly)

or Glu(—Cys-Gly or Glu(Cys-Gly)

(Note that Glu would represent the corresponding thiol ester with a
|
Cys-Gly
bond between the γ-carboxyl of glutamic acid and the thiol group of cysteine.)

N^2-α-Glutamyllysine	Glu-Lys
N^6-α-Glutamyllysine	Glu(Lys) or Glu(Lys)
N^2-γ-Glutamyllysine	Glu(Lys) or Glu(Lys)
N^6-γ-Glutamyllysine	Glu(Lys) or Glu(Lys) or Glu(Lys)

The presence of free, substituted, or ionized functional groups can be represented (or stressed) as follows:

Glycyllysylglycine	H-Gly-Lys(H)-Gly-OH
Its dihydrochloride	$^{+}H_2$-Gly-Lys($^{+}H_2$)-Gly-OH·2Cl$^-$
Its sodium salt	Gly-Lys-Gly-O$^-$Na$^+$
Its N^6-formyl derivative	Gly-Lys(CHO)-Gly
	Gly-Lys(CHO)-Gly
etc.	

6.2 *Peptides substituted at N^2 (see* 4.1)

Glycylnitrosoglycine	Gly(NO)Gly or Gly-(NO)Gly
Glycylsarcosine (see 2.2.5)	Gly(Me)Gly or Gly-MeGly
	or Gly-Sar
N-Glycyl-*N*-acetylglycine	Gly(Ac)Gly or Gly-(Ac)Gly
N,*N*-Diglycylglycine	Gly(Gly)Gly
	or $(Gly)_2$-Gly etc.

6.3 *Cyclic polypeptides*

6.3.1 *Homodetic cyclic polypeptides* (*the ring consists of amino-acid residues in peptide linkage only*). Three representations are possible:

6.3.1.1 The sequence is formulated in the usual manner but placed in parentheses and preceded by (an italic) *cyclo*.

Example:

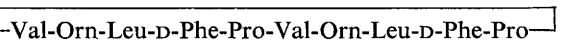
Gramicidin S

cyclo(-Val-Orn-Leu-D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro-)

or (see 1.5, sentence 2)

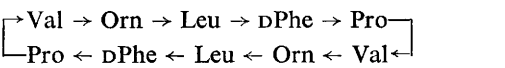
cyclo(-Val-Orn-Leu-DPhe-Pro-Val-Orn-Leu-DPhe-Pro-)

6.3.1.2 The terminal residues may be written on one line, as in 6.3.1.1, but joined by a lengthened bond. Using the same example in the two forms (see 1.5):



or

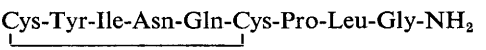
6.3.1.3 The residues are written on more than one line, in which case the CO → NH direction must be indicated by arrows, thus (in the optional manner of 1.5):



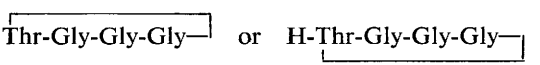
6.3.2 *Heterodetic cyclic polypeptides* (*the ring consists of other residues in addition to amino-acid residues in peptide linkage*). These follow logically from the formulation of substituted amino acids.

Examples:

Oxytocin



Cyclic ester of threonylglycylglycylglycine



References

(1) Abbreviated Designation of Amino Acid Derivatives and Peptides: *J. Biol. Chem.* (1966) **241**, 2491; *Biochemistry* (1966) **5**, 2485; *Biochim. Biophys. Acta* (1966) **121**, 1; *Biochem. J.* (1967) **102**, 23; *Arch. Biochem. Biophys.* (1967) **121**, 1; *Eur. J. Biochem*, (1967) **1**, 375; *Hoppe-Seyler's Z. Physiol. Chem.* (1967) **348**, 256; *Bull. Soc. Chim. Biol.* (1967) **49**, 121; *Molek. Biol.* (1968) **2**, 282

(2) Abbreviations and Symbols for Chemical Names of Special Interest in Biological Chemistry (Section 5 revised by 3 below): *J. Biol. Chem.* (1966) **241**, 527; *Biochemistry* (1966) **5**, 1445; *Biochem. J.* (1966) **101**, 1; *Virology* (1966) **29**, 480; *Arch. Biochem. Biophys.* (1966) **115**, 1; *Eur. J. Biochem.* (1967) **1**, 259; *Hoppe-Seyler's Z. Physiol. Chem.* (1967) **348**, 245; *Bull. Soc. Chim. Biol.* (1968) **50**, 3; *Molek. Biol.* (1967) **1**, 872

(3) Abbreviations and Symbols for Nucleic acid, Polynucleotides and the Constituents: *Biochem. J.* (1970) **120**, 449; *Biochemistry* (1970) **9**, 4022; *Eur. J. Biochem.* (1970) **15**, 203; *J. Biol. Chem.* (1970) **245**, 5171; *Hoppe-Seyler's Z. Physiol. Chem.* (1970) **351**, 1055; *J. Mol. Biol.* (1971) **55**, 299; in the press elsewhere

(4) Nomenclature of Vitamins, Coenzymes and Related Compounds: Trivial Names of Miscellaneous Compounds of Importance in Biochemistry, Nomenclature of Quinones with Isoprenoid Side Chains, Nomenclature and Symbols for Folic Acid and Related

Compounds, Nomenclature of Corrinoids: *Arch. Biochem. Biophys.* (1967) **118**, 505; *Biochem. J.* (1967) **102**, 15; *Biochim. Biophys. Acta* (1965) **107**, 1; (1966) **117**, 285; *Eur. J. Biochem.* (1967) **2**, 1; *J. Biol. Chem.* (1966) **241**, 2987; *Bull. Soc. Chim. Biol.* (1967) **49**, 331; *Hoppe-Seyler's Z. Physiol. Chem.* (1967) **348**, 266

(5) Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino Acids): *Biopolymers* (1969) **8**, 161; *Arch. Biochem. Biophys.* (1968) **123**, 633; *Biochem. J.* (1968) **106**, 577; *Biochemistry* (1968) **7**, 483; *Biochim. Biophys. Acta* (1968) **168**, 1; *Eur. J. Biochem.* (1968) **3**, 129; *J. Biol. Chem.* (1968) **243**, 2451; *Bull. Soc. Chim. Biol.* (1969) **51**, 205; *Hoppe-Seyler's Z. Physiol. Chem.* (1969) **349**, 1013 (Revision in the press)

(6) IUPAC: Nomenclature of Organic Chemistry (Definitive Rules for Section C), Rule C-404 and Table VI: *Pure Appl. Chem.* (1965) **11**, nos. 1–2

II Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains

(*Reprinted by permission from* I.U.P.A.C. Information Bulletin, Appendices on Tentative Nomenclature, Symbols, Units, and Standards, No. 10, February 1971)

Preamble

These Rules are based on *A proposal of standard conventions and nomenclature for the description of polypeptide conformation* (Edsall *et al.*, 1966*a,b,c*), and have been prepared by a Subcommission set up by the I.U.P.A.C.–I.U.B. Commission on Biochemical Nomenclature in 1966. The original proposals have been modified so as to bring them as far as possible into line with the system of nomenclature current in the fields of organic and polymer chemistry.

Two Recommendations are appended to the Rules, the first dealing with the terms configuration and conformation, and the second with primary, secondary and tertiary structure. These are formulated as recommendations rather than rules, because there is at present no general agreement about their definition.

Note: Two alternative notations are recommended throughout. That with superscripts and subscripts may be used when it is unlikely to cause confusion, e.g. in printed or manuscript material; that without is to be used where superscripts or subscripts may cause confusion, or are technically difficult or impossible, e.g. in computer outputs. In the latter connexion the following Roman equivalents of Greek letters are recommended:

α	A	τ	T
β	B	υ	U
γ	G	ϕ	F
δ	D	χ	X
ε	E	ψ	Q
ζ	Z	ω	W
η	H		

Rule 1. General Principles of Notation

1.1. *Designation of atoms.* The atoms of the main chain are denoted thus:

$$-\mathrm{NH}-\mathrm{C}^{\alpha}\mathrm{H}^{\alpha}-\mathrm{CO}-$$

Where confusion might arise, the following additional symbolism may be used:

$$-\mathrm{N'H'-C^{\alpha}H^{\alpha}-C'O'-}$$
$$\qquad\qquad\quad |$$

1.2. *Amino-acid residues*, —NH–CHR–CO—, are numbered sequentially from the amino-terminus to the carboxyl-terminus of the chain, the residue number being denoted *i*.

Example: C^{α} of the *i*th residue is written C_i^{α} or Cα(*i*).

1.3. *Peptide units.* For some purposes it is more convenient to group together the atoms —CHR–CO–NH—. These groups are described as "peptide units", and the peptide unit number, like the residue number, is denoted *i*. It will be noted that the two numbers are identical for all atoms except NH; generally there will be no confusion, because a single document will use either "residues" alone, or "peptide units" alone, but in the latter case explicit reference must be made to this usage at the beginning. If confusion might arise, the symbols N_i^* and H_i^* are to be used for these atoms in the *i*th peptide unit, which are N_i and H_i in the *i*th residue (so that $N_i^* = N_{i+1}$).

Example: $NH_2-CHR_1-CO+NH+CHR_2-CO+NH+CHR_3-$

Peptide unit no. 2

Residue no. 2

Residue notation	N_2	C_2^{α}	C_2	N_3
Peptide unit notation	N_1^*	C_2^{α}	C_2	N_2^*

Notes: (i) Residue notation is used throughout these Rules.
(ii) Whether "residues" or "peptide units" are being used, ϕ_i and ψ_i always refer to torsion angles about bonds of the same C_i^{α}.

1.4. *Bond lengths.* If a bond A—B be denoted A_i—B_j or A_i (see Rules 3.1, 4.5), the bond length is written $b(A_i, B_j)$ (or $b(Ai, Bj)$), or b_i^A (or $bA(i)$). An abbreviated notation for use in side chains is indicated in Rule 4.5.

Note: The symbol previously recommended for bond length was *l*. This symbol is no longer recommended, partly because it is easily confused with 1 in many type fonts, and and partly because it is also used for vibration amplitude in electron diffraction and spectroscopy.

1.5. *Bond angles.* The bond angle included between three atoms $A_i \overset{B_j}{\diagdown\diagup} C_k$ is written $\tau(A_i, B_j, C_k)$, which may be abbreviated, if there is no ambiguity, to $\tau(B_j)$ or τ_j^B or $\tau B(j)$.

1.6. *Torsion angles.* If a system of four atoms $A \diagdown B-C \diagup D$ is projected on to a plane normal to bond B—C, the angle between the projection of

A—B and the projection of C—D is described as the *torsion angle*† of A and D about bond B—C; this angle may also be described as the angle between the plane containing A, B and C, and the plane containing B, C and D. The torsion angle is written in full as $\theta(A_i, B_j, C_k, D_l)$, which may be abbreviated, if there is no ambiguity, to $\theta(B_j, C_k)$, $\theta(B_j)$ or θ_j^B, etc. In

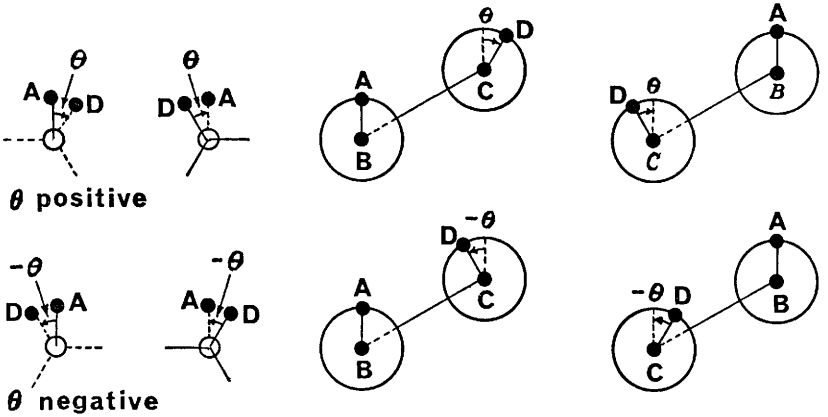


FIG. 1. Newman and perspective projections illustrating positive and negative torsion angles. Note that a right-handed turn of the bond to the front atom about the central bond gives a positive value of θ from whichever end the system is viewed.

Notes: (i) Angles are measured in the range $-180° < \theta \leqslant +180°$, rather than from 0° to 360°, so that the relationship between enantiomeric configurations or conformations can be readily appreciated.

(ii) The symbols actually used to describe the various torsion angles important in polypeptides are ϕ, ψ, ω, υ and χ (see Rules 3.2, 4.5.2). In the above, θ is used simply as an illustrative generic symbol covering all these.

the eclipsed conformation in which the projections of A—B and C—D coincide, θ is given the value 0° (synplanar conformation). A torsion angle is considered positive (+ θ) or negative (− θ) according as, when the system is viewed along the central bond in the direction B → C (or C → B), the bond to the front atom A (or D) requires to be rotated to the right or to the left, respectively, in order that it may eclipse the bond to the rear atom D (or A); note that it is immaterial whether the system be viewed from one end or the other. These relationships are illustrated in Figure 1.

† The terms *dihedral angle* and *internal rotation angle* are also used to describe this angle, and may be regarded as alternatives to *torsion angle*, though the latter has been used throughout these Rules.

Rule 2. The Sequence Rule, and Choice of Torsion Angle

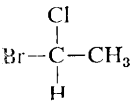
2.1. The Rules here enunciated for use in the field of synthetic polypeptides and proteins are in general harmony with the Sequence Rule of Cahn, Ingold & Prelog†, with the exceptions of Rules 2.1.1 and 2.2.2 (Cases II and III), and later Rules dependent upon these. The Sequence Rule was formulated as a universal and unambiguous means of designating the "handedness" or chirality of an element of asymmetry. It includes Subrules for the purpose of arranging atoms or groups in an order of precedence or preference, and this system may conveniently be used in the description of steric relationships across single bonds (see Klyne & Prelog, 1960). Here its function is to determine the priority or precedence of different atoms or groups attached to the same atom. However, Rule 2.1.1 below overrides the precedences of the Sequence Subrules, providing a new 'local' (specialist) system for use with the general Sequence Rule‡. After application of Rule 2.1.1, the normal procedure of the Sequence Rule is applied, but modified by Rule 2.2.2; in this connexion the only parts of the Sequence Rule required are given in Rules 2.1.2 to 2.1.5.

2.1.1. The main chain is given formal priority over branches, notwithstanding any conflict with the following rules. Thus the main chain has precedence at C^{α} over the side chain, and at C′ over O′.

Note: This rule has not yet been formally accepted except in the present context.

2.1.2. The order of (decreasing) priority is the order of (decreasing) atomic number.

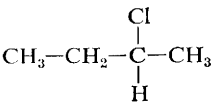
Example: in



the order of priority is Br, Cl, CH_3, H.

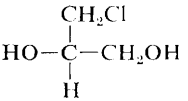
2.1.3. If two atoms attached to the central atom are the same, the ligands attached to these two atoms are used to determine the priority.

Examples: (i) in



the order is Cl, (CH_2–CH_3), CH_3, H. (C^xH_2–CH_3 takes precedence over C^yH_3 because C^x is bonded to C,H,H and C^y to H,H,H.)

(ii) in

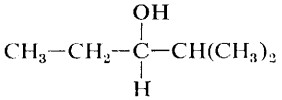


† See Cahn, Ingold & Prelog (1966*a*,*b*,*c*), IUPAC Tentative Rules for the Nomenclature of Organic Chemistry, Section E, IUPAC Information Bulletin no. 35, pp. 36–80 (1969). Earlier papers: Cahn & Ingold (1951), Cahn, Ingold & Prelog (1956). For a partial, simplified account see Cahn (1964) and Eliel (1962).

‡ Other local systems are available analogously for steroids, carbohydrates and cyclitols, where the Sequence Rule is applied when the local system does not suffice.

the order is OH, CH_2Cl, CH_2OH, H.

(iii) in



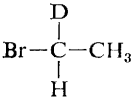
the order is OH, $CH(CH_3)_2$, CH_2–CH_3, H.

2.1.4. A double bond is formally treated as though it were split. Thus > C=O is treated as > C(O)–C.

Example: in CH_3–CO–OH the order is =O, −OH, CH_3.

2.1.5. If two ligands are distinguished only by having different masses (e.g. deuterium and hydrogen), the heavier takes precedence.

Example: in

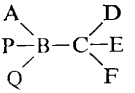


the order is Br, CH_3, D, H.

Note: This rule is to be used only if the two previous rules do not give a decision.

2.2. *Choice of torsion angle and numbering of branches (tetrahedral configurations)*

2.2.1. If, in a compound



the Sequence Rule gives the priorities A > P, Q and D > E > F, then the *Principal Torsion Angle* θ is that measured by reference to the atoms A—B—C—D as in Rule 1.6 above.

The branches beginning at C are numbered $C\underset{1}{-}D$, $C\underset{2}{-}E$ and $C\underset{3}{-}F$.

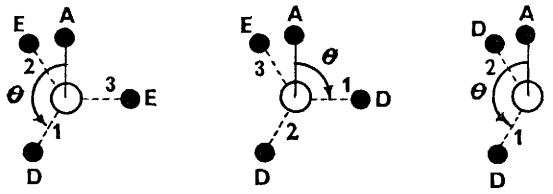
2.2.2. If two branches are identical, and the third is different (or non-existent), they are numbered in a clockwise sense when viewed in the direction B → C, as follows (see Fig. 2).

Case I. D > E = E. D has the highest priority and is given the smallest number (1).

Case II. D = D > E. E has the lowest priority and is given the largest number (3).

Case III. D = D, numbered 1 and 2 (E non-existent).

In each case the Principal Torsion Angle is measured between A—B and branch 1.



I II III

FIG. 2. Tetrahedral configurations. Case I: D > E = E. Case II: D = D > E. Case III: D = D (E non-existent).

Notes: (i) The rule given in Case II differs from Conformational Selection Rule (b) of the Sequence Rule (see Cahn *et al.*, 1966*a*, p. 406), according to which if an identity among the groups of a set leaves one group unique, the unique group is fiducial. The reason for the difference is that the Sequence Rule would define Principal Torsion Angle in terms of a hydrogen atom whenever a single such atom formed part of the set; in the X-ray technique, nearly always used to establish structures of the type under discussion, hydrogen atoms are usually unobservable, and even at the best cannot be located accurately, so that the position of one used to define a Principal Torsion Angle could only be established by calculation based on (perhaps unjustified) assumptions about the bond angles concerned. These considerations apply with even more force to Case III, where one branch is non-existent: the "phantom atom" of zero atomic number would be given highest priority because it is unique.

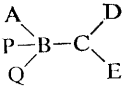
(ii) in case III the clockwise passage from CD^1 to CD^2 shall be by the shorter of the two possible routes.

2.2.3. If all three branches are identical, that giving the smallest positive or negative value of the Principal Torsion Angle is normally† assigned the highest priority and the lowest number (1) (see Fig. 3, IV, V); if two branches have torsion angles respectively + 60° and − 60°, the former is chosen (see Fig. 3, VI). The others are numbered in a clockwise sense when viewed in the direction B → C.

Note: Rule 2.2.3 introduces a new principle, not invoked in 2.2.1 or 2.2.2, that the precedence depends on the conformation. This must necessarily be done since in this case the branches are distinguishable only in this respect. (The same applies to Rule 2.3.2 below.)

2.3. *Choice of torsion angle and numbering of branches (planar trigonal configurations)*

2.3.1. If, in a compound

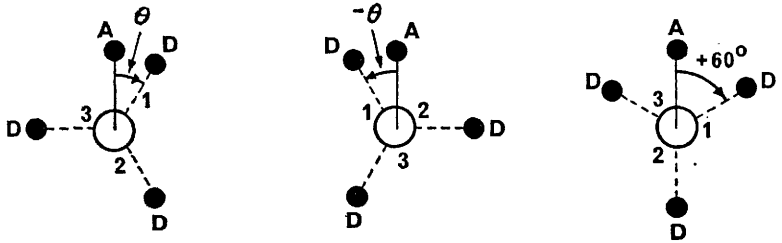


such that B, C, D and E are coplanar, or nearly so, the Sequence Rule gives the priorities A > P, Q and D > E, then the Principal Torsion

† The qualification "normally" is added to avoid the need to renumber the branches if by chance the rule would demand this in consequence of a movement during refinement of a structure. In this or similar cases the symbolism should remain unchanged.

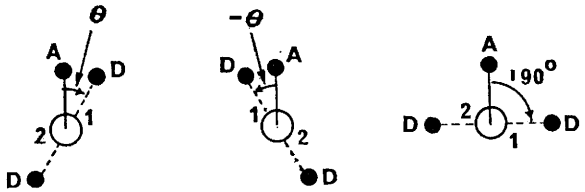
Angle is that measured by reference to atoms A—B—C—D as in Rule 1.6 above.

The branches beginning at C are numbered C_1—D, C_2—E.



IV V VI

FIG. 3. Tetrahedral configurations. Three identical branches; IV, general case, θ positive; V, general case, θ negative; VI, $\theta = +60°$.



VII VIII IX

FIG. 4. Planar trigonal configurations. Identical branches; VII, θ positive, VIII, θ negative, IX, $\theta = +90°$.

2.3.2. If the branches are identical, that giving the smaller positive or negative value of the Principal Torsion Angle is normally assigned the highest priority and the lowest number (1); if the two branches have torsion angles respectively + 90° and − 90°, the former is chosen (see Fig. 4).

Rule 3. The Main Chain (or Polypeptide Backbone)

3.1. *Designation of bonds.* Bonds between main-chain atoms are denoted by the symbols of the two atoms terminating them, e.g. N_i—C_i^α, C_i^α—C_i, C_i—N_{i+1}, C_i—O_i, N_i—H_i. Abbreviated symbols should not be used. Bond lengths are written $b(C_i, N_{i+1})$, etc.

3.2. *Torsion angles*

3.2.1. The Principal Torsion Angle describing rotation about N—C^α is denoted by ϕ, that describing rotation about C^α—C is denoted by ψ, and that describing rotation about C—N is denoted by ω. The symbols ϕ_i, ψ_i and ω_i are used to denote torsion angle of bonds within the ith residue in the case of ϕ and ψ, and between the ith and $(i + 1)$th

residues in the case of ω; specifically, ϕ_i refers to the torsion angle of the sequence of atoms C_{i-1}, N_i, C_i^α, C_i; ψ_i to the sequence N_i, C_i^α, C_i, N_{i+1}; and ω_i to the sequence C_i^α, C_i, N_{i+1}, C_{+i1}^α (see Fig. 5). In accordance with Rules 1.6 and 2.1.1, these torsion angles are ascribed zero values for eclipsed conformations of the main-chain atoms N, C^α and C, that is, for the so-called *cis*-conformations (see Table 1).

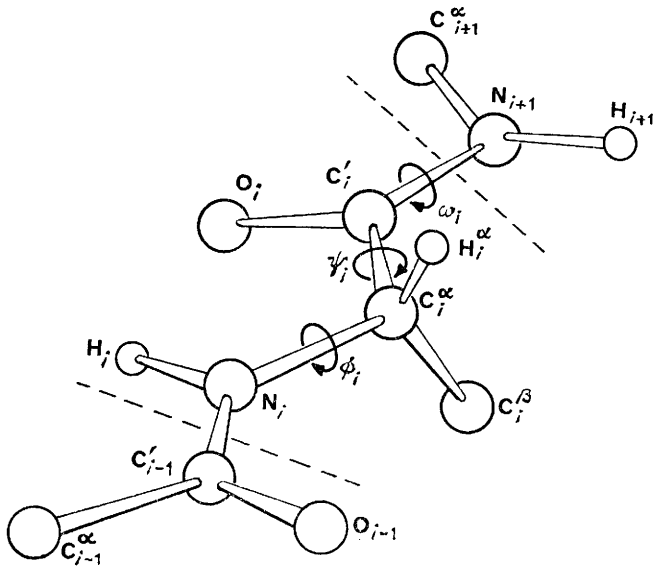


FIG. 5. Perspective drawing of a section of polypeptide chain representing two peptide units. The limits of a *residue* are indicated by dashed lines, and recommended notations for atoms and torsion angles are indicated. The chain is shown in a fully extended conformation ($\phi_i = \psi_i = \omega_i = +180°$), and the residue illustrated is L-.

Notes: (i) This convention differs from that proposed by Edsall *et al.* (1966*a*,*b*,*c*). The new designation of angles may be derived from the old by adding 180° to, or subtracting 180° from, the latter. (This statement is precisely correct only if the peptide bond is exactly planar, which is not generally the case in experimentally determined structures.)

(ii) Owing to the partial double-bond character of CO···NH, it is normally possible for ω to assume values only in the neighbourhood of 0° or 180°. $\omega \sim 180°$ is the value which is generally found (i.e. the *trans*-conformation).

(iii) A "fully-extended" polypeptide chain is characterized by $\phi = \psi = \omega = +180°$. The case of $\phi = \psi = 0°$ would involve the relations indicated in Table 1.

(iv) Table 2 gives values of ϕ and ψ for various well-known regular structures. It is noteworthy that a right-handed α-helix has negative torsion angles.

(v) Figure 6 is a typical conformational map ($\phi - \psi$ plot) using the Rules enunciated above.

3.2.2. There may occasionally be a need to consider torsion angles differing from zero for the sequences of atoms O=C—N—C^{α} and C^{α}—C—N—H, in cases where C=O or N—H lies out of the peptide plane. These angles may be represented υ^{O} and υ^{H} (Greek *upsilon*).

TABLE 1

Main-chain torsion angles for various conformations in peptides of L-*amino acids*

ϕ	Rotation about N—C^{α}		ψ	Rotation about C^{α}—C	
0°	C^{α}—C *trans*	to N—H	0°	C^{α}—N *trans*	to C—O
+ 60°	C^{α}—H *cis*		+ 60°	C^{α}—R *cis*	
+ 120°	C^{α}—R *trans*		+ 120°	C^{α}—H *trans*	
+ 180°	C^{α}—C *cis*		+ 180°	C^{α}—N *cis*	
− 120°	C^{α}—H *trans*		− 120°	C^{α}—R *trans*	
− 60°	C^{α}—R *cis*		− 60°	C^{α}—H *cis*	

Notes: (i) *Trans* to N_i—H_i is the same as *cis* to N_i—C_{i-1}; *trans* to C_i—O_i is the same as *cis* to C_i—N_{i+1} (see Fig. 5).

(ii) For the description of D-amino acids, interchange C^{α}—H and C^{α}—R in the Table.

TABLE 2

Approximate torsion angles for some regular structures

	ϕ (deg.)	ψ (deg.)	ω (deg.)	Reference
Right-handed α-helix (α-poly(L-alanine))	− 57	− 47	+ 180	Arnott & Dover (1967)
Left-handed α-helix	+ 57	+ 47	+ 180	Arnott & Dover (1967)
Parallel-chain pleated sheet	− 119	+ 113	+ 180	Schellman & Schellman (1964)
Antiparallel-chain pleated sheet (β-poly(L-alanine))	− 139	+ 135	− 178	Arnott, Dover & Elliott (1967)
Polyglycine II	− 80	+ 150	+ 180	Ramachandran, Sasisekharan & Ramakrishnan (1966)
Collagen	− 51, − 76, − 45	+ 153, + 127, + 148	+ 180	Yonath & Traub (1969)
Poly(L-proline) I	− 83	+ 158	0	Ramachandran & Sasisekharan (1968), calculated from Traub & Schmucli (1963)
Poly(L-proline) II	− 78	+ 149	+ 180	Arnott & Dover (1968)

(*Note*: for a fully extended chain $\phi = \psi = \omega = + 180°$.)

3.3 *Chain terminations*

3.3.1. If the terminal amino group of the chain is protonated, the three hydrogen atoms are denoted H_1^1, H_1^2 and H_1^3; the hydrogen atom giving the smallest (positive or negative) value of the Principal Torsion Angle H—N—C^α—C is denoted H_1^1 and the others are numbered in

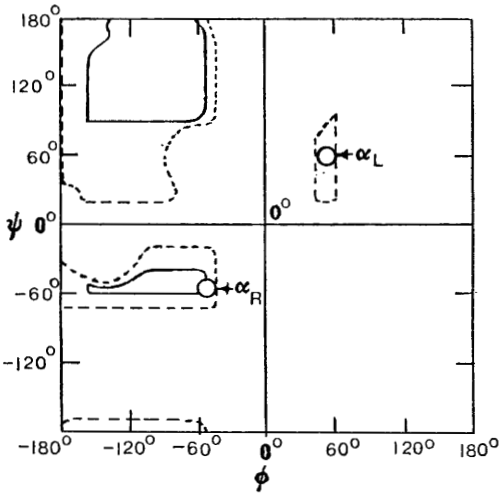
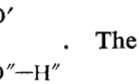


FIG. 6. Typical conformational map (Ramachandran, Ramakrishnan & Sasisekharan, 1963) transposed into the standard conventions.

Note: This diagram is identical with that of Edsall *et al.* (1966*a*,*b*,*c*), except that the origin is now at the centre, instead of at the lower left-hand corner. The solid lines enclose the freely allowed values of ϕ and ψ for an alanyl residue in a polypeptide; the dotted lines enclose "outer limit" values based on the shortest known van der Waals radii in related structures. Analogous diagrams for other residues, and for slightly different assumptions, are given by Ramachandran & Sasisekharan (1968; note that these authors used the earlier convention with the origin at the corner).

a clockwise sense when viewed in the direction $C^\alpha \rightarrow N$. The corresponding torsion angles are denoted ϕ_1^1, ϕ_1^2 and ϕ_1^3. If the terminal amino group is not protonated, the hydrogen atoms are denoted H_1^1 and H_1^2 in accordance with Rule 2.2.2, and the corresponding torsion angles ϕ_1^1 and ϕ_1^2.

3.3.2. At the carboxyl-terminus of the chain (i = T) the double-bonded oxygen is written as O′ and the other oxygen as O″, thus



. The torsion angles about the C^α—C bond are written ψ_T^1 and ψ_T^2 (or ψ1(T), ψ2(T)); the torsion angle about the C—O″ bonds, defining the orientation of the hydrogen atom of the

hydroxyl group relative to C^{α}, is written θ^{C}_{T} (or $\theta C(T)$). If the terminal carboxyl group is ionized, the oxygen atoms are denoted O′ and O″, the precedence being determined by Rule 2.3.2, and the torsion angles are written as before.

Note: Instead of O′ and O″ the alternative notations O^1 and O^2 may be used. ψ_T may be used instead of ψ^{1}_{T}, in conformity with the convention for the middle of the chain, so long as confusion does not arise.

3.3.3. *Substituted terminal groups.* Natural extensions of the above rules may be devised, e.g.:

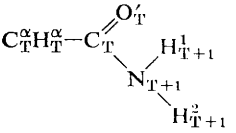
(i) *N-formyl group*

$$H_0—C_0O_0—N_1H_1—C^{\alpha}_1H^{\alpha}_1—.....$$

(ii) *N-acetyl group*

$$C_0(H^1_0,H^2_0,H^3_0)—C_0O_0—N_1H_1—C^{\alpha}_1H^{\alpha}_1—.....$$

(iii) *C-amido group*



Rule 4. Side Chains

4.1. Atoms are lettered, or lettered and numbered, from C^{α}, and bonds are numbered from C^{α}, working outwards away from the main chain.

4.2. *Designation of atoms other than hydrogen.* Atoms other than hydrogen are designated in the usual way by Greek letters, β, γ, δ, etc. C^{β}_{i} (or $C\beta(i)$), N^{ζ}_{il} (or $N\zeta(i)$).

Note: The notations for the amino acids normally occurring in proteins are given in Table 3.

TABLE 3

Symbols for atoms and bonds in the side chains of the commonly occurring L-*amino acids*

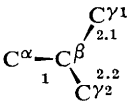
(a) Unbranched side chains

Alanine	$C^{\alpha}_{1}—C^{\beta}$
Serine	$C^{\alpha}_{1}—C^{\beta}_{2}—O^{\gamma}$
Cysteine	$C^{\alpha}_{1}—C^{\beta}_{2}—S^{\gamma}$
Cystine	$C^{\alpha}_{1_i}—C^{\beta}_{2_i}—S^{\gamma}_{3_i}—S^{\gamma}_{2_k}—C^{\beta}_{1_k}—C^{\alpha}_{k}$
Methionine	$C^{\alpha}_{1}—C^{\beta}_{2}—C^{\gamma}_{3}—S^{\delta}_{4}—C^{\epsilon}$
Lysine	$C^{\alpha}_{1}—C^{\beta}_{2}—C^{\gamma}_{3}—C^{\delta}_{4}—C^{\epsilon}_{5}—N^{\zeta}$

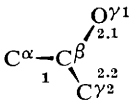
TABLE 3—*continued*

(b) Branched side chains

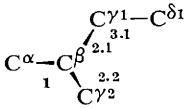
Valine



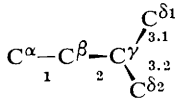
Threonine



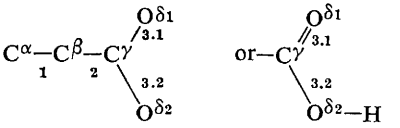
Isoleucine



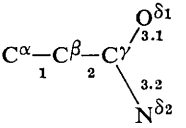
Leucine



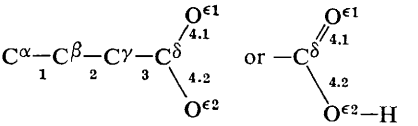
Aspartic acid



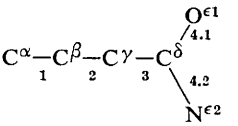
Asparagine



Glutamic acid



Glutamine



Arginine

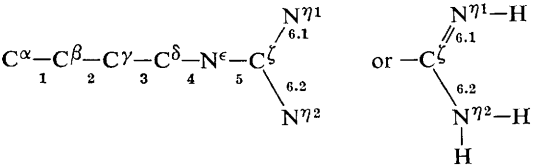
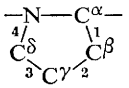


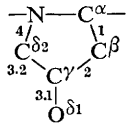
TABLE 3—*continued*

(c) Cyclic side chains

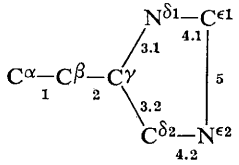
Proline



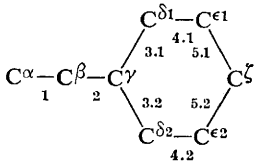
Hydroxyproline



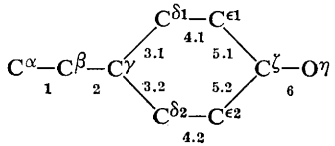
Histidine



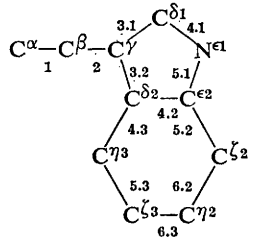
Phenylalanine



Tyrosine



Tryptophan



4.3. *Designation of branches.* If a side chain is branched, the branches are numbered 1 and 2, the order being determined:

(i) in cases where the branches are different, by application of Rule 2.2.1 or 2.3.1,

(ii) in cases where two branches are identical (e.g. in valine, phenylalanine), by the application of Rule 2.2.2 (valine) or 2.3.2 (phenylalanine).

Non-hydrogen atoms in different branches are designated by the Greek letter indicating their degree of remoteness from C^{α} and by the number of their branch (see Rules 2.2 and 2.3); e.g. in valine $C_i^{\gamma 1}$ and $C_i^{\gamma 2}$ (or $C\gamma 1(i)$, $C\gamma 2(i)$). The branch number need not be indicated where no ambiguity results: e.g. in threonine O^{γ} and C^{γ} instead of $O^{\gamma 1}$ and $C^{\gamma 2}$, hydroxyproline O^{δ}, C^{δ} instead of $O^{\delta 1}$, $C^{\delta 2}$, and in histidine C^{δ}, N^{ϵ}, etc. instead of $C^{\delta 2}$, $N^{\epsilon 2}$. For asparagine or glutamine, in cases where nitrogen and oxygen in the amide group have not yet been distinguished, these atoms may be written $(NO)^{\delta 1}$, $(NO)^{\delta 2}$, or $(NO)^{\epsilon 1}$, $(NO)^{\epsilon 2}$, the indices 1 and 2 being determined by Rule 2.3.2.

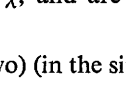
4.4. *Designation of hydrogen atoms.* Hydrogen atoms are designated by the Greek letter and/or number of the atom to which they are attached; e.g. in valine H_i^{β} (or $H\beta(i)$). Where three hydrogen atoms are attached to a single non-hydrogen atom, they are designated 1, 2 and 3; in the situation

the hydrogen atom giving the smallest (positive or negative) value of the Principal Torsion Angle is designated 1, and the others are numbered in a clockwise sense when viewed in the direction B → C (see Rule 2.2.3, which also covers the case where $\theta = \pm 60°$). E.g. in valine $H_i^{\gamma 11}$, $H_i^{\gamma 12}$, $H_i^{\gamma 13}$ and $H_i^{\gamma 21}$, $H_i^{\gamma 22}$, $H_i^{\gamma 23}$ (or $H\gamma 11(i)$, etc.). Where only two hydrogen atoms are present, they are designated in accordance with Rule 2.2.2, case I for $-CH_2-R$ and case III for $-NH_2$.

4.5. *Designation of bonds and torsion angles* (see Table 3)

4.5.1. Bonds are designated by means of the two atoms terminating them, e.g. $C_i^{\alpha}-C_i^{\beta}$, $N_i^{\zeta}-H_i^{\zeta 2}$, or, if no ambiguity results, by the symbol of the first atom of the bond, e.g. C_i^{α}, $C_i^{\gamma 1}$. In superscripts, the bond may be denoted either by α; β; $\gamma 1$; $\gamma 2$ etc. or by, 1; 2; 3,1; 3,2; etc. Bond lengths are denoted $b(C_i^{\alpha}, C_i^{\beta})$, bC_i^{α}, b_i^{1}, $b_i^{3,1}$ etc.

4.5.2. Torsion angles are denoted by χ, and are specified by two (or three) superscripts, the first one (or two) (in the situation



indicating the bond B—C about which the angle is measured, and the last indicating whether the angle is measured relative to D, E or F. The Principal Torsion Angle is defined by Rule 2.2.1, and if there is no ambiguity the last superscript may be omitted in referring to it.

Thus, in valine, $\chi_i^{2,1}$ and $\chi_i^{2,2}$ refer to the torsion angles specifying atoms $C_i^{\gamma 1}$ and $C_i^{\gamma 2}$; in leucine $\chi_i^{3,1,1}$, $\chi_i^{3,1,2}$ and $\chi_i^{3,1,3}$ refer to the torsion

angles specifying the three hydrogen atoms attached to $C^{\delta 1}$. If there is no ambiguity the Principal Torsion Angles may be referred to, in valine and leucine, as χ_i^1 and $\chi_i^{3,1}$, respectively. Corresponding notations without subscripts are $\chi 2,1(i)$, $\chi 2(i)$; $\chi 3,1,1(i)$, $\chi 3,1(i)$.

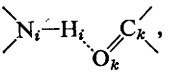
Note: By the Sequence Rule, when $\chi_1 = 0$, C^{γ} (or $C^{\gamma 1}$) is in the eclipsed position relative to N.

Rule 5. Hydrogen Bonds

5.1. *Polarity of hydrogen bonds.* In specifying a hydrogen bond as directed from residue i to residue k (or from atom X_i to atom Y_k), the direction X—H to :Y is implied; i.e. the atom covalently linked to the hydrogen atom is mentioned first.

Example: in the α-helix, the N—H of residue i is hydrogen-bonded to the O=C of residue $(i - 4)$. Therefore, the α-helix is described as having i to $(i - 4)$, or $(5 - 1)$, hydrogen-bonding.

5.2. *Dimensions of hydrogen bonds.* Dimensions may be denoted by natural extensions of the nomenclature given above. For example, in



the following symbols might be used:
$b(H_i, O_k)$, $\tau(N_i, H_i, O_k)$, $\tau(H_i, O_k, C_k)$, $\theta(H_i, O_k)$, $\theta_i(N, H)$, $\theta_k(C, O)$.

Rule 6. Helical Segments

A regular helix is strictly of infinite length, with all ϕ's identical and all ψ's identical. A helical *segment* of polypeptide chain may be defined *either* in terms of ϕ and ψ, *or* in terms of symmetry and hydrogen-bond arrangement.

6.1. In the description of helices or helical segments, the following symbols should be used:

n = number of residues per turn;
h = unit height (translation per residue along the helix axis);
$t = 360°/n$ = unit twist (angle of rotation per residue about the helix axis).

6.2. *Definition in terms of ϕ and ψ.* Under this definition a helical segment is referred to as a (ϕ, ψ) helix; thus a right-handed α-helix would be a $(-57°, -47°)$ helix. The *first* and *last* residues of the helical segment are taken to be the first and last residues which have ϕ and ψ values equal to those defining the helix, within limits which should be defined in the context. No account is taken of hydrogen-bonding arrangements.

6.3. *Definition in terms of symmetry and hydrogen-bond arrangement.* A helix is referred to as an n_r helix,

where n = number of residues per turn;

r = number of atoms in ring formed by a hydrogen bond and the segment of main chain connecting its extremities.

Thus an α-helix would be 3.6_{13}. The *first* helical residue is taken as the first whose CO group is *regularly* bonded to NH along the helix (in the case of an α-helix, to the NH of the fifth residue); the *last* helical residue is the last whose NH is *regularly* hydrogen-bonded to CO along the helix (in the case of an α-helix, to the CO of the residue last but four). Irregular hydrogen-bonding arrangements are not considered to form part of the helix.

Notes: (i) A helical segment defined by Rule 6.2 may, but need not necessarily, be two residues shorter than the same segment defined by Rule 6.3.

(ii) These rules prescribe no definitions for irregular helical segments.

APPENDIX

Recommendation A. Conformation and Configuration

There is at present no agreed definition of these two terms for general stereochemical usage.

In polypeptide chemistry, the term "conformation" should be used, in conformity with current usage, to describe different spatial arrangements of atoms produced by rotation about covalent bonds; a change in conformation does not involve the breaking of chemical bonds (except hydrogen bonds) or changes in *chirality* (see Cahn, Ingold & Prelog, 1966*a,b,c*).

On the other hand, in polypeptide chemistry the term "configuration" is currently used to describe spatial arrangements of atoms whose interconversion requires the formal breaking and making of covalent bonds (*Note*: this usage takes no account of the breaking or making of hydrogen bonds). For a more extensive discussion see IUPAC Tentative Rules for the Nomenclature of Organic Chemistry, Section E, Fundamental Stereochemistry, *IUPAC Information Bull. no.* 35 (1969), pp. 71–80.

Recommendation B. Definitions of Primary, Secondary, Tertiary and Quaternary Structure

These concepts, originally introduced by Linderstrøm-Lang (1952)†, cannot be defined with precision, but the definitions given below may be helpful.

B.1. The *primary structure* of a segment of polypeptide chain or of a protein is the amino-acid sequence of the polypeptide chain(s), without regard to spatial arrangement (apart from configuration at the α-carbon atom).

Note: This definition does not include the positions of disulphide bonds, and is therefore not identical with "covalent structure".

† The use of the terms "primary, secondary, tertiary and quaternary structure" has been criticized as being imprecise by Wetlaufer (1961). He has proposed an alternative terminology.

B.2. The *secondary structure* of a segment of polypeptide chain is the local spatial arrangement of its main-chain atoms without regard to the conformation of its side chains or to its relationship with other segments.

B.3. The *tertiary structure* of a protein molecule, or of a subunit of a protein molecule, is the arrangement of all its atoms in space, without regard to its relationship with neighbouring molecules or subunits.

B.4. The *quaternary structure* of a protein molecule is the arrangement of its subunits in space and the ensemble of its intersubunit contacts and interactions, without regard to the internal geometry of the subunits.

Note: A protein molecule not made up of at least potentially separable subunits (not connected by covalent bonds) possesses no quaternary structure. Examples of proteins without quaternary structure are ribonuclease (1 chain) and chymotrypsin (3 chains).

REFERENCES

Arnott, S. & Dover, S. D. (1967). *J. Mol. Biol.* **30**, 209.
Arnott, S. & Dover, S. D. (1968). *Acta Cryst.* B, **24**, 599.
Arnott, S., Dover, S. D. & Elliott, A. (1967). *J. Mol. Biol.* **30**, 201.
Cahn, R. S. (1964). *J. Chem. Educ.* **41**, 116.
Cahn, R. S. & Ingold, C. K. (1951). *J. Chem. Soc.* p. 612.
Cahn, R. S., Ingold, C. K. & Prelog, V. (1956). *Experientia*, **12**, 81.
Cahn, R. S., Ingold, C. K. & Prelog, V. (1966*a*). *Angew. Chem. intern. Edit.* **5**, 385.
Cahn, R. S., Ingold, C. K. & Prelog, V. (1966*b*). *Angew. Chem. intern. Edit.* **5**, 511.
Cahn, R. S., Ingold, C. K. & Prelog, V. (1966*c*). *Angew. Chem.* **78**, 413.
Edsall, J. T., Flory, P. J., Kendrew, J. C., Liquori, A. M., Némethy, G., Ramachandran, G. N. & Scheraga, H. A. (1966*a*). *J. Biol. Chem.* **241**, 1004.
Edsall, J. T., Flory, P. J., Kendrew, J. C., Liquori, A. M., Némethy, G., Ramachandran, G. N. & Scheraga, H. A. (1966*b*). *Biopolymers*, **4**, 130.
Edsall, J. T., Flory, P. J., Kendrew, J. C., Liquori, A. M., Némethy, G., Ramachandran, G. N. & Scheraga, H. A. (1966*c*). *J. Mol. Biol.* **15**, 339.
Eliel, E. L. (1962). *Stereochemistry of Carbon Compounds*, pp. 92–4. New York: McGraw Hill.
IUPAC Information Bulletin. (1969). No. 35, pp. 36–80.
Klyne, W. & Prelog, V. (1960). *Experientia*, **16**, 521.
Linderstrøm-Lang, K. (1952). *Proteins and Enzymes, Lane Memorial Lectures*, p. 54. Stanford, Calif.: Stanford University Press.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Advanc. Protein Chem.* **23**, 283.
Ramachandran, G. N., Sasisekharan, V. & Ramakrishnan, C. (1966). *Biochim. biophys. Acta*, **112**, 168.
Schellman, J. A. & Schellman, C. (1964). In *The Proteins*, ed. by H. Neurath, vol. 2, p. 1. New York: Academic Press.
Traub, W. & Schmueli, U. (1963). In *Aspects of Protein Structure*, ed. by G. N. Ramachandran, p. 81. London: Academic Press.
Wetlaufer, D. B. (1961). *Nature*, **190**, 1113.
Yonath, A. & Traub, W. (1969). *J. Mol. Biol.* **43**, 461.

Author Index