Amino-acids, Peptides, and Proteins—Volume 2

Specialist Periodical Reports

The Chemical Society

SBN: 85186 014 1

A Specialist Periodical Report

Amino-acids, Peptides, and Proteins

Volume 2

A Review of the Literature Published during 1969

Senior Reporter

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

SBN: 85186 014 1

The Chemical Society
Burlington House, London, W1V 0BN

Reporters

R. Blagrove, *University of Oxford*
C. C. F. Blake, *University of Oxford*
J. R. Brocklehurst, *University of Oxford*
B. W. Bycroft, *University of Nottingham*
D. G. Dalgleish, *University of Oxford*
J. S. Davies, *University College of Swansea, University of Wales*
R. Henson, *University of Oxford*
J. H. Jones, *University of Oxford*
P. H. Lloyd, *University of Oxford*
A. R. Peacocke, *University of Oxford*
R. N. Perham, *Cambridge University*
N. C. Price, *University of Oxford*
R. M. Stephens, *Portsmouth Polytechnic*
I. O. Walker, *University of Oxford*

Organic formulae composed by Wright's Symbolset Method

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

Preface

The scope of this Report, as for Volume 1, is essentially the chemistry of amino-acids, peptides, and proteins, and the intention has been to review all the relevant papers listed in *Current Chemical Papers* during 1969, together with any appearing in the main journals that year even if not so listed. It was, however, considered that the extent of the new work on metal derivatives did not justify a separate chapter this year, and the papers appearing in 1969 and 1970 will be reviewed together in the next Report. Volume 1 contained a discussion of selected aspects of the relationship between the structure and biological activity of peptides and proteins. Systematic complete coverage of the literature in this field would constitute a major extension of the scope of this Report; this has not been attempted and no separate section has been included this year. On the other hand, biological aspects have naturally been discussed in connection with the chemistry in many parts of this volume. Comment should also be made on the change in the title of Chapter 4. The term 'peptides of abnormal structure' has been used in the past for peptides having structural features (*e.g.* ester linkages) other than those typical of proteins, but of course such peptides are by no means uncommon in Nature. In order to avoid possible misinterpretation we have changed this year to the more cumbersome but more accurate title 'Peptides with Structural Features Not Typical of Proteins'. This chapter now includes work on the penicillins and cephalosporins. We shall be grateful if our attention is drawn to any important omissions in the bibliography, in order that the references may be included in the next volume.

As in Volume 1, there is an author index but not a subject index, and the extended list of contents will, we hope, assist the reader in finding the section he requires. Reference numbers apply to the chapter in which they occur or, if the chapter is divided into parts, to the part concerned. For the convenience of readers, we have reprinted as Chapter 5 the main recommendations of the I.U.P.A.C.–I.U.B. Commission on Biochemical Nomenclature relevant to this field.

I would again thank warmly the contributors who have made my editorial task so pleasant.

G. T. YOUNG

Contents

Chapter 2 Structural Investigation of Peptides and Proteins
I Primary Structures and Chemical Modification
By R. N. Perham

Abbreviations

Abbreviations for amino-acids and their use in the formulation of derivatives follow the relevant Tentative Rules of the I.U.P.A.C.-I.U.B. Commission on Biochemical Nomenclature, relevant extracts of which are reprinted in Chapter 5 of this Report.

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-biphenyl)-isopropoxycarbonyl
BSA	bovine serum albumin
Cha	cyclohexylamine
c.d.	circular dichroism
Cm	carboxymethyl
Dce	2,2-diethoxycarbonyl
Dcha	dicyclohexylamine
DMF	*NN*-dimethylformamide
DNA	deoxyribonucleic acid
Dnp	2,4-dinitrophenyl
Dns	1-dimethylaminonaphthalene-5-sulphonyl (dansyl)
Ec	ethylcarbamoyl
EDTA	ethylenediamine tetra-acetate
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^+, oxidised, 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
Pipoc	piperidino-oxycarbonyl
Pth-Gly	the phenylthiohydantion derived from glycine, *etc.*
RNA	ribonucleic acid
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

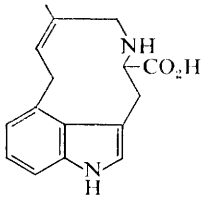
The presentation of this chapter conforms, in general, to the pattern established in last year's Report. New amino-acids with novel structures continue to be isolated from natural sources and the chemistry of amino-acids still attracts considerable attention from all branches of the subject. The emphasis remains on α-amino-acids and it is only for these that a comprehensive coverage has been attempted. Biochemical aspects have been covered only where they relate directly to the chemistry.

1 Naturally Occurring Amino-acids

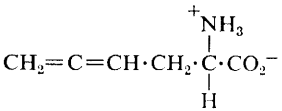
A. Introduction.—The literature on non-protein naturally occurring amino-acids has been reviewed [1] up to the end of 1968 in a valuable article which also discusses their probable origin and function. Speculation concerning the origin of D-amino-acids continues: the observation that for amino-acids containing more than one centre of chirality inversion of configuration occurs only at the α-centre has led to the rule of α-epimerisation.[2, 3] Amino-acids have been isolated during the year which conform to the rule, together with one apparent exception (see below). Two possible explanations to account for the origin of D-amino-acids, both of which accommodate most of the known information relating to D-amino-acids and are in accord with the above described rule, have been proposed.[4, 5]

B. New Free Occurring Amino-acids.—The continuing usefulness of physical methods in structural elucidation is apparent, and several new amino-acids have been characterised during the year. Those whose structure has been confirmed by synthesis are included in the list of newly synthesised amino-acids in section 2. Especially noteworthy is the characterisation by spectroscopic methods of the novel amino-acid (1), which is produced by *Claviceps* species inhibited with ethionine and is probably an intermediate in the biosynthesis of lysergic acid.[6] Several other amino-acids have also been isolated from fungi or micro-organisms. The

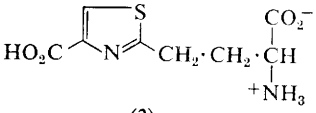
[1] J. F. Thompson, C. J. Morris, and L. K. Smith, *Ann. Rev. Biochem.*, 1969, **38**, 137.
[2] M. Bodanszky and D. Perlman, *Nature*, 1968, **218**, 291.
[3] M. Bodanszky and D. Perlman, *Science*, 1969, **163**, 352.
[4] A. B. Mauger, *Experientia*, 1968, **24**, 1068.
[5] B. W. Bycroft, *Nature*, 1969, **224**, 595.
[6] J. E. Robbers and H. G. Floss, *Tetrahedron Letters*, 1969, 1857.



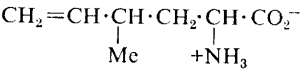
(1)



(2)



(3)



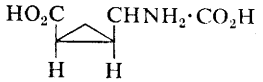
(4)

mushrooms *Morchella esculenta* and related species produce *cis*-3-amino-L-proline.[7] β-Methylene-L-norvaline has been obtained from *Lactarius helvus*,[8] and the closely related allenic amino-acid (2) occurs in *Amanita solitaria*.[9] A new sulphur-containing amino-acid (3) from *Xeromus subtomentosus* has been reported,[10] but this is probably a modified dipeptide derived from glutamic acid and cysteine. The antimetabolic antibiotic (4), produced by an unidentified *Streptomyces* species, has been shown to be active against Gram-positive organisms.[11]

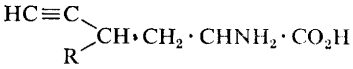
A number of new plant amino-acids have been described. The phytochemical similarities between *Aesculus* and *Blighia* species have been extended: *exo*-3,4-methanoproline and *cis*-α-(carboxycyclopropyl)glycine (5) occur in *Aesculus parviflora*, while the *trans*-isomer of (5) is found in *Blighia sapida*.[12] The structures of three new acetylenic amino-acids (6), (7), and (8) from the seeds of *Euphoria longan* have been determined [13] and the structure of the pyrimidine amino-acid isolated from pea seedlings has been revised to (9).[14] Selenomethylselenocysteine has been shown to be the principal selenium-containing amino-acid in *Astragalus bisulcatus*.[15]

C. New Amino-acids from Peptide Hydrolysates.—Hydrolysis of the antibiotic monamycin affords the novel piperidazine amino-acid (10), the structure and absolute configuration of which have been confirmed by a

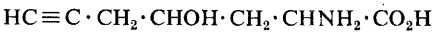
[7] S. Hatanaka, *Phytochemistry*, 1969, **8**, 1305.
[8] B. Levenberg, *J. Biol. Chem.*, 1968, **243**, 6009.
[9] W. S. Chilton, G. Tsou, L. Kirk, and R. G. Benedict, *Tetrahedron Letters*, 1968, 6283.
[10] J. Jadot, J. Casimir, and R. Warin, *Bull. Soc. chim. belges*, 1969, **78**, 299.
[11] R. B. Kelly, D. G. Martin, and L. J. Hanka, *Canad. J. Chem.*, 1969, **47**, 2504.
[12] L. Fowden, A. Smith, D. S. Millington, and R. C. Sheppard, *Phytochemistry*, 1969, **8**, 437.
[13] M. L. Sung, L. Fowden, D. S. Millington, and R. C. Sheppard, *Phytochemistry*, 1969, **8**, 1227.
[14] E. G. Brown and B. S. Mangat, *Biochim. Biophys. Acta*, 1969, **177**, 427.
[15] S. N. Nigam, J. I. Tu, and N. B. McConnell, *Phytochemistry*, 1969, **8**, 1161.



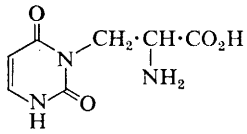
(5)



(6) R = Me
(7) R = CH_2OH

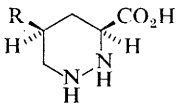


(8)

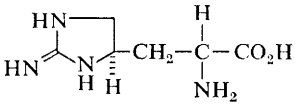


(9)

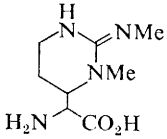
direct correlation with L-ornithine, as well as the structurally related D-amino-acids (11) and (12) (opposite absolute configuration).[16] In addition, D-isoleucine, an exception to the rule of α-epimerisation, has also been isolated from the acid hydrolysate.[17] In contrast, the acid hydrolysates of the antibiotics aspartocin and enduracidin yield L-*threo*- and D-*erythro*-αβ-diaminobutyric acid [18] and enduracididine (13) and *allo*-enduracididine



(10) R = OH
(11) R = H
(12) R = Cl



(13) L-α
(14) D-α



(15)

(14) [19] respectively; both the D-amino-acids are in accord with the rule. The structure, but not the stereochemistry, of the basic amino-acid stendomycidine (15) has been determined,[20] and the close similarity between it and the previously isolated guanidine amino-acids capreomycidine and viomycidine is noteworthy. Further details on the chemistry of viomycidine have been published.[21] Hydrolysis of diatom cell walls affords *cis*-3,4,*trans*-3,4-dihydroxy-L-proline,[22] the structure of which has been confirmed by an *X*-ray crystallographic analysis (see section 3). *N*-(2-Hydroxyethyl)alanine was isolated from the alkali hydrolysate of a phospholipid obtained from rumen protozoa.[23]

[16] C. H. Hassall, R. B. Morton, Y. Ogihara, and W. A. Thomas, *Chem. Comm.*, 1969, 1079.
[17] K. Bevan, J. S. Davies, C. H. Hassall, and D. A. S. Phillips, *Chem. Comm.*, 1969, 1246.
[18] W. K. Hausmann, D. B. Borders, and J. E. Lancaster, *J. Antibiotics*, 1969, **22**, 207.
[19] S. Horii and Y. Kameda, *J. Antibiotics*, 1968, **21**, 665.
[20] M. Bodanszky, G. Marconi, and A. Bodanszky, *J. Antibiotics*, 1969, **22**, 40.
[21] B. W. Bycroft, L. R. Croft, A. W. Johnson, and T. Webb, *J. Antibiotics*, 1969, **22**, 133.
[22] T. Nakajima and B. E. Volcani, *Science*, 1969, **164**, 1400.
[23] P. Kemp and R. M. C. Dawson, *Biochim. Biophys. Acta*, 1969, **176**, 678.

D. Occurrence of Known Amino-acids.—β-Cyanoalanine, the neurotoxin, is the product of cyanide fixation and an intermediate in asparagine biosynthesis in a number of *Lathyrus* and *Vicia* species.[24] Canaline (L-γ-aminoxy-α-aminobutyric acid) has been obtained from the unripe seeds of *Astragalus suricus*[25] and *trans*-3-hydroxy-L-proline from the seeds of *Delonix regia*.[26] N^{α}-Acetyl-L-arginine has been reported to be present in cattle brain.[27]

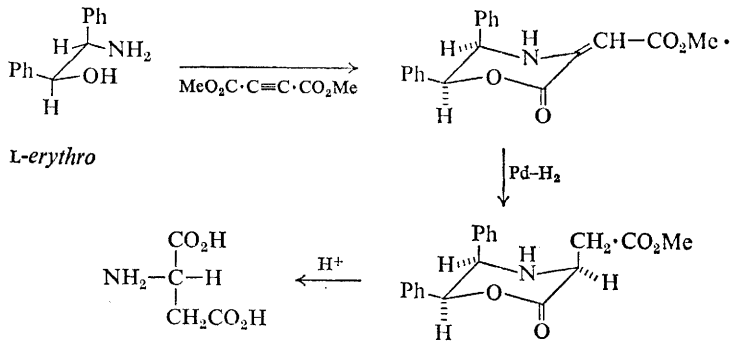
2 Chemical Synthesis and Resolution of Amino-acids

A. Introduction.—The majority of new syntheses, with notable exceptions, were performed by variations of well established routes or involved elaboration of available amino-acids. Therefore, only those syntheses which have points of particular interest will be discussed and the remainder merely mentioned or incorporated into the list of newly synthesised amino-acids.

B. Protein Amino-acids.—New syntheses have been reported for DL-lysine,[28] DL-tryptophan,[29] DL-cystine,[30] D- and L-cysteine,[31] and D-phenylalanine,[32] and the advances in asymmetric synthesis of α-amino-acids reviewed.[33] Both forms of alanine have been obtained from oxaloacetic acid by an asymmetric transamination,[34, 35] and an elegant asymmetric synthesis (Scheme 1) of optically pure L-aspartic acid has been reported.[36] The detailed studies, first reported last year, on the resolution of glutamic acid and alanine by fractional crystallisation have been extended.[37, 38]

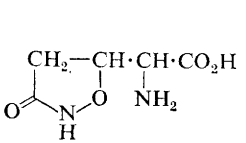
C. Other Naturally Occurring Amino-acids.—Interest in L-dopa (3,4-dihydroxyphenylalanine) continues: syntheses from L-tyrosine using crystalline β-tyrosinase[39] and from *N*-blocked L-tyrosine derivatives employing micro-organisms[40] have been published. DL-Indospicine,[41] the heptatotoxic amino-acid, first reported last year, and all the four isomers of

[24] C. Ressler, Y. H. Giza, and S. N. Nigam, *J. Amer. Chem. Soc.*, 1969, **91**, 2766.
[25] H. Inatomi, F. Inugai, and T. Murakami, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 2521.
[26] M. L. Sung and L. Fowden, *Phytochemistry*, 1968, **7**, 2061.
[27] H. Ohkusu and A. Mori, *J. Neurochem.*, 1969, **16**, 1485.
[28] T. I. Samoilovich, A. S. Polyanskaya, and V. V. Perekalin, *Zhur. org. Khim.*, 1969, **5**, 579.
[29] I. Maeda and R. Yoshida, *Bull. Chem. Soc. Japan*, 1968, **41**, 2975.
[30] T. Furuta and T. Ishimaru, *J. Chem. Soc. Japan*, 1968, **89**, 716.
[31] A. Schöberl, M. Rimpler, and K. H. Magosch, *Chem. Ber.*, 1969, **102**, 1767.
[32] J. Schawartz and G. Eibel, *Chem. and Ind.*, 1968, 1698.
[33] K. K. Babievskii and V. K. Latov, *Uspekhi Khim.*, 1969, **39**, 1009.
[34] K. Matsumoto and K. Harada, *J. Org. Chem.*, 1968, **33**, 4526.
[35] K. Harada and K. Matsumoto, *J. Org. Chem.*, 1968, **33**, 4467.
[36] J. P. Vigneron, H. Kagan, and A. Horeau, *Tetrahedron Letters*, 1968, 5681.
[37] T. Watanabe and G. Noyori, *J. Chem. Soc. Japan, Ind. Chem. Sect.*, 1969, **72**, 1080.
[38] T. Watanabe and G. Noyori, *J. Chem. Soc. Japan, Ind. Chem. Sect.*, 1969, **72**, 1083.
[39] H. Kumagai, H. Matsui, H. Ohgishi, K. Ogata, H. Yamada, T. Ueno, and H. Fukami, *Biochem. Biophys. Res. Comm.*, 1969, **34**, 266.
[40] C. J. Sih, P. Foss, J. Rosazza, and M. Lemberger, *J. Amer. Chem. Soc.*, 1969, **91**, 6204.
[41] C. C. J. Culvenor, M. C. Foster, and M. P. Hegarty, *Chem. Comm.*, 1969, 1091.

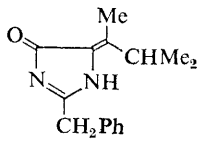


Scheme 1

tricholomic acid [42–46] (16) have been synthesised. Synthetic DL-α-amino-δ-(guanylureido)-*n*-valeric acid is identical with gigartinine in all respects except optical activity,[47] and the structures of β-methylene-norvaline,[8] the



(16)

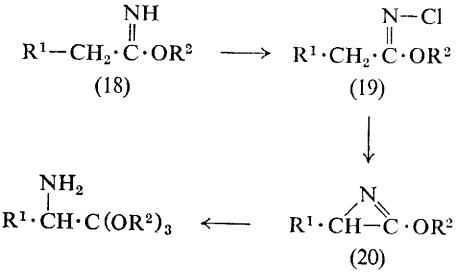


(17)

sulphur-containing amino-acid [10] (3), and *N*-2,3-dihydroxybenzoyl-serine [48] have all been confirmed by synthesis. Reduction of (17) followed by hydrolysis and methylation affords mainly DL-*threo*-*N*,β-dimethyl-leucine, indicating that the predominant geometrical isomer of (17) is that shown.[49] A new synthesis of L-αβ-diaminopropionic acid has been described.[50]

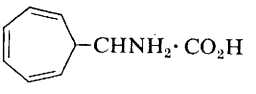
D. α-Alkyl-α-amino-acids.—A brief preliminary description of a new route [51] to α-amino-acid orthoesters and hence α-amino-acids has appeared: the three-step synthesis (Scheme 2) involves conversion of a nitrile to an imino-ester (18), chlorination of the latter to an *N*-chloroimidate (19), and alkoxide treatment of (19) to form the α-amino orthoester. The synthesis

[42] H. Iwasaki, T. Kamiya, O. Oka, and J. Ueyanagi, *Chem. and Pharm. Bull.* (*Japan*), 1969, **17**, 866.
[43] H. Iwasaki, T. Kamiya, C. Hatanaka, Y. Sunada, and J. Ueyanagi, *Chem. and Pharm. Bull.* (*Japan*), 1969, **17**, 873.
[44] T. Kamiya, *Chem. and Pharm. Bull.* (*Japan*), 1969, **17**, 879.
[45] T. Kamiya, *Chem. and Pharm. Bull.* (*Japan*), 1969, **17**, 886.
[46] T. Kamiya, *Chem. and Pharm. Bull.* (*Japan*), 1969, **17**, 890.
[47] K. Ito and Y. Hashimoto, *Agric. and Biol. Chem.* (*Japan*), 1969, **33**, 237.
[48] I. G. O'Brien, G. B. Cox, and F. Gibson, *Biochim. Biophys. Acta*, 1969, **177**, 321.
[49] H. Kotake, T. Saito, and K. Okubo, *Bull. Chem. Soc. Japan*, 1969, **42**, 1367.
[50] T. Kitagawa, T. Ozasa, and H. Taniyama, *J. Pharm. Soc. Japan*, 1969, **89**, 285.
[51] W. H. Graham, *Tetrahedron Letters*, 1969, 2223.

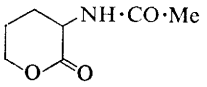


Scheme 2

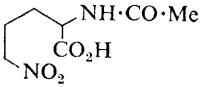
was unsuccessful when α-disubstituted acetonitriles were employed, and it is suggested that the reaction proceeds through the intermediacy of (20). A convenient synthesis of the novel racemic amino-acid (21) from cycloheptatrienylium tetrafluoroborate and dimethyl formamidomalonate has been reported.[52] Further syntheses of 3- and 4-substituted α-glutamic and 4-substituted-α-aminoadipic acid derivatives have also been described.[53–55]



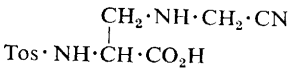
(21)



(22)



(23)

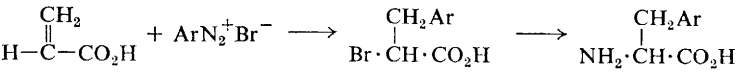


(24)

The reaction of (22) with hydrobromic acid followed by silver nitrate affords (23), which can be resolved by stereospecific enzymic deacylation to give (*S*)-α-amino-δ-nitrovaleric acid,[56] an important intermediate in the synthesis of ferrichrome. Reduction of (24), prepared by cyanomethylation of either the D- or L- form of αβ-diaminopropionic acid, yields the D- and L- forms of 4-azalysine.[57] This method has also been used to make [6-^{14}C]-L-4-azalysine.

[52] S. Hanessian and G. Schütze, *J. Medicin. Chem.*, 1969, **12**, 347.
[53] G. M. Shakhnazaryan, L. A. Saakyan, and M. T. Dangyan, *Zhur. org. Khim.*, 1968, **4**, 1914.
[54] L. A. Saakyan, M. T. Dangyan, and G. M. Shakhnazaryan, *Armyan. khim. Zhur.*, 1968, **21**, 971.
[55] L. V. Alekseeva, N. L. Burde, and B. N. Lundin, *Zhur. obshchei Khim.*, 1968, **38**, 1687.
[56] B. Maurer and N. Keller-Schierlein, *Helv. Chim. Acta*, 1969, **52**, 388.
[57] J. Kolc, *Coll. Czech. Chem. Comm.*, 1969, **34**, 630.

Interest continues in substituted phenylalanines because of their potential biological activity, and a number of new derivatives have been prepared by standard methods.[58–62] An extensive study of the application of the Meerwein reaction for a general synthesis (Scheme 3) of substituted phenylalanines has been reported.[63] Twenty-one substituted aniline derivatives were investigated and it was observed that for those with electron-withdrawing groups in *para*, and, particularly, *ortho* positions, yields in the ammonolysis step were low. A similar observation was made in an attempted synthesis of pentafluorophenylalanine *via* the Meerwein arylation route: the intermediate α-bromo-acid afforded *trans*-4-amino-2,3,5,6-tetrafluorocinnamic acid rather than the phenylalanine on treatment with ammonia. Pentafluorophenylalanine was ultimately synthesised by the azlactone method.[64]

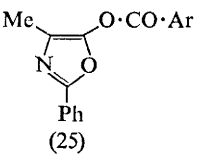


Scheme 3

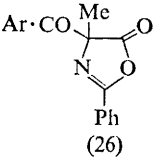
E. α-Dialkyl-α-amino-acids.—The oxazole (25), readily prepared from the corresponding azlactone and benzoyl chloride, rearranges on warming in pyridine to the azlactone (26), which on reduction gives a mixture of *erythro*- and *threo*-*N*-benzoyl-2-methyl-3-phenylserine esters.[65] Separation of the isomers was effected by chromatography. The advantage of this method over a similar synthesis,[66] reported last year, is that it provides access to those amino-acids whose aromatic moiety is more readily available as the carboxylic acid chloride than as a Grignard reagent.

An interesting conversion of (*S*)-2-methyl-3-phenylpropionic acid to (*R*)-α-methylphenylalanine has been described.[67] The key step in this sequence was the photolysis of the azidoformate (27). This generates a nitrene which then undergoes a stereospecific intramolecular insertion to give (28). A number of bicyclic α-amino-acids of the type (29) have been synthesised using a Diels–Alder addition of cyclopentadiene to substituted α-nitrocinnamates.[68]

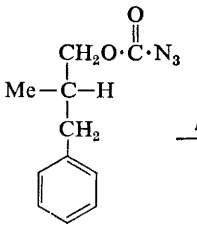
[58] H. H. Ong, C. R. Creveling, and J. W. Daly, *J. Medicin. Chem.*, 1969, **12**, 458.
[59] A. Langemann and M. Scheer, *Helv. Chim. Acta*, 1969, **52**, 1095.
[60] P. Crooij and J. Eliaers, *J. Chem. Soc.* (*C*), 1969, 559.
[61] A. Jean and J. Anatol, *Compt. rend.*, 1969, **268**, *C*, 1307.
[62] M. Frankel, A. Shenhar, D. Gertner, and A. Zilkha, *Israel J. Chem.*, 1968, **6**, 921.
[63] G. H. Cleland, *J. Org. Chem.*, 1969, **34**, 744.
[64] R. Filler, N. R. Ayyangar, W. Gustowski, and H. H. Kang, *J. Org. Chem.*, 1969, **34**, 534.
[65] S. H. Pines and M. Sletzinger, *Tetrahedron Letters*, 1969, 727.
[66] S. H. Pines, S. Karady, and M. Sletzinger, *J. Org. Chem.*, 1968, **33**, 1758.
[67] S. Terashima and S. Yamada, *Chem. and Pharm. Bull.* (*Japan*), 1968, **16**, 2064.
[68] M. Kinoshita, H. Yanagisawa, S. Doi, E. Kaji, and S. Umezawa, *Bull. Chem. Soc. Japan*, 1969, **42**, 194.



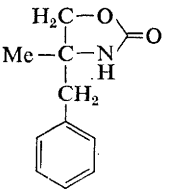
(25)



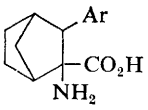
(26)



(27)



(28)



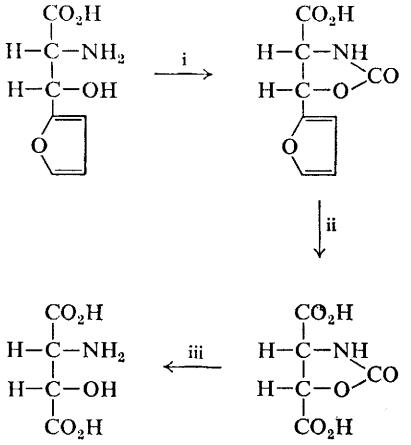
(29)

F. α-Amino-acids with Aliphatic Hydroxy-groups in the Side-chain.—A new synthesis [69] of either the *threo-* or the *erythro-* form of β-hydroxy-DL-aspartic acid is shown in Scheme 4. The starting materials, *threo-* and *erythro-*β-furylserine, are readily prepared from furfural and glycine and the oxidation step proceeds in tolerable yield. Resolution of *erythro-*β-hydroxyaspartic acid has been achieved *via* the *N*-benzyl derivative.[70] Amination of *erythro-*β-methoxy-α-bromohexanoic acid proceeds with retention of configuration when ammonia is used, but treatment with sodium azide followed by reduction results in inversion. This reaction was employed for the synthesis of *threo-* and *erythro-*β-hydroxylysines.[71] The *erythro-* but not the *threo-* form of α-amino-β-hydroxy-γ-benzyloxybutyric acid can be resolved enzymatically.[72] Resolution of the latter can be accomplished using the method of Vogler.[73]

The preparation of α-methylserines was discussed in the previous section.

G. *N*-Substituted-α-amino-acids.—Interest continues on *N*-hydroxy-α-amino-acids, and further work [74] describes the synthesis of a number of optically active forms, not previously reported, by the action of hydroxylamine on α-bromo-esters. A number of *N*-β-hydroxyethyl-α-amino-acids have been synthesised by a modified Strecker procedure.[75] This synthesis is of interest since the first naturally occurring derivative of this type

[69] T. Inui, Y. Ohta, T. Ujike, H. Katsura, and T. Kaneko, *Bull. Chem. Soc. Japan*, 1968, **41**, 2148.
[70] Y. Liwschitz, A. Singerman, and Y. Wiesel, *Israel J. Chem.*, 1968, **6**, 647.
[71] C. H. Stammer and R. G. Webb, *J. Org. Chem.*, 1969, **34**, 2306.
[72] K. Okawa, K. Hori, K. Hirose, and Y. Nakagawa, *Bull. Chem. Soc. Japan*, 1969, **42**, 2720.
[73] K. Vogler and P. Lanz, *Helv. Chim. Acta*, 1966, **49**, 1348.
[74] T. LaNoce, E. Bellasio, and E. Testa, *Ann. Chim.* (*Italy*), 1968, **58**, 393.
[75] D. Giraud-Clénet and J. Anatol, *Compt. rend.*, 1969, **268**, *C*, 117.

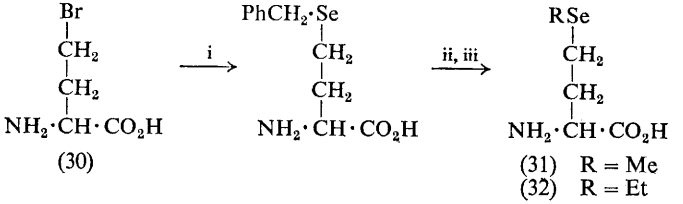


Reagents: i, $COCl_2$; ii, $KMnO_4$; iii, hydrolysis

Scheme 4

has been reported.[23] A more convenient reduction of 2-pyrrolecarboxylic acid to 3,4-dehydroproline is possible using a mixture of hypophosphorous and hydriodic acids as the reducing agent.[76]

H. α-Amino-acids Containing Sulphur or Selenium.—Alkylation of homocysteine thiolactone with a primary alkyl halide in the presence of sodium methoxide provides a more facile route to *S*-alkyl-DL-homocysteines, since it avoids the use of sodium in liquid ammonia necessary in the usual route.[77] L-Selenomethionine (31) and L-selenoethionine (32) have been



Reagents: i, $PhCH_2SeNa$; ii, $Na–NH_3$; iii, RI

Scheme 5

prepared from L-α-amino-γ-bromobutyric acid (30) using the method outlined in Scheme 5.[78]

[76] A. Corbella, P. Gariboldi, G. Jommi, and F. Mauri, *Chem. and Ind.*, 1969, 583.
[77] H. M. Kolenbrander, *Canad. J. Chem.*, 1969, **47**, 3271.
[78] H. D. Jakubke, J. Fischer, K. Jošt, and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1968, **33**, 3910.

I. A List of α-Amino-acids which have been Synthesised for the First Time.—

Compound	*Ref.*
DL-β-methylene-norvaline	8
L-α-amino-γ-(4-carboxythiazol-2-yl)butyric acid	10
D- and L-*erythro*-α-amino-3-oxo-5-isoxazolidineacetic acid	42
D- and L-*threo*-α-amino-3-oxo-5-isoxazolidineacetic acid	43
DL-α-amino-5-(guanylureido)-n-valeric acid (gigartinine)	47
DL-*N*-2,3-dihydroxybenzoylserine	48
DL-2-(cyclohepta-2,4,6-trien-1-yl)glycine	52
DL-γ-ethyl-α-glutamic acid	53
DL-γ-propyl-α-glutamic acid	53
DL-γ-isoamyl-α-glutamic acid	53
DL-γ-phenyl-α-glutamic acid	53
DL-3-chloroglutamic acid	55
(*S*)-α-amino-δ-nitrovaleric acid	56
D- and L-4-azalysine	57
DL-β-(2,4,5-trihydroxyphenyl)alanine (6-hydroxy-dopa)	58, 59
DL-β-(3-bromo-5-methoxyphenyl)alanine	60
DL-β-(3,5-dimethyl-4-methoxyphenyl)alanine	61
DL-β-(4-trimethylsilylphenyl)alanine	62
DL-β-(4-triethylsilylphenyl)alanine	62
DL-β-[2,4-di(trimethylsilyl)phenyl]alanine	62
DL-β-(4-cyanophenyl)alanine	63
DL-β-(4-acetylphenyl)alanine	63
DL-β-(4-carboxamidophenyl)alanine	63
DL-2-amino-3-(*p*-aminophenyl)bicyclo[2,2,1]heptane-2-carboxylic acid (diastereoisomers separated and identified)	68
DL-2-amino-3-(*m*-aminophenyl)bicyclo[2,2,1]heptane-2-carboxylic acid (diastereoisomers not separated)	68
DL-2-amino-3-(*p*-chlorophenyl)bicyclo[2,2,1]heptane-2-carboxylic acid (diastereoisomers not separated)	68
DL-β-hydroxylysine(*threo* and *erythro*)	71
DL-*N*-β-hydroxyethylvaline	75
DL-*N*-β-hydroxyethylphenylalanine	75
DL-*N*-β-hydroxyethyl-2-aminovaleric acid	75
L-selenoethionine	78
DL-α-(1,3-dithiolan-2-yl)glycine	79
DL-β-(1,3-dithiolan-2-yl)alanine	79
DL-α-amino-δ-(1,3-dithiolan-2-yl)butyric acid	79
L-3,5-di(phenylthio)-4-(4′-hydroxyphenoxy)phenylalanine	80
L-3,5-di(phenylthio)-4-(3′-isopropyl-4′-hydroxyphenoxy)phenylalanine	80
L-3,5-di(ethylthio)-4-(4′-hydroxyphenoxy)phenylalanine	80
L-3,5-di(ethylthio)-4-(3′-isopropyl-4′-methoxyphenoxy)phenylalanine	80
β-pyrrolyl-L-alanine	81
L-2-amino-4-pyrrolylbutyric acid	81
L-2-amino-5-pyrrolylvaleric acid	81
DL-β-(2-pyridyl)alanine	82, 83

[79] M. P. Mertes and A. A. Ramsey, *J. Medicin. Chem.*, 1969, **12**, 342.
[80] E. C. Jorgensen, R. O. Muhlhauser, and R. A. Wiley, *J. Medicin. Chem.*, 1969, **12**, 689.
[81] K. Poduška, J. Rudinger, J. Gloede, and H. Gross, *Coll. Czech. Chem. Comm.*, 1969, **34**, 1002.
[82] P. T. Sullivan, M. Kester, and S. J. Norton, *J. Medicin. Chem.*, 1968, **11**, 1172.
[83] H. Watanabe, S. Kuwata, K. Naoe, and Y. Nishida, *Bull. Chem. Soc. Japan*, 1968, **41**, 1634.

Compound	*Ref.*
DL-β-(3-pyridyl)alanine	82
DL-β-(8-quinolinol-5-yl)alanine	84
DL-β-(6-aminopurin-9-yl)alanine	85, 86
DL-β-(4-amino-2-hydroxypyrimidin-1-yl)alanine	85
DL-β-(2,4-dihydroxy-5-methylpyrimidin-1-yl)alanine	85
DL-β-(6-hydroxypurin-9-yl)alanine	86
DL-β-(6-dimethylaminopurin-9-yl)alanine	86
DL-β-(6-mercaptopurin-9-yl)alanine	86
DL-β-(6-methylthiopurin-9-yl)alanine	86
DL-β-(*o*-carboranyl)alanine	87
DL-4-azatryptophan	88
DL-4-methyl-6-chloro-7-azatryptophan	88
L-N^{ε}-(1,4-dihydro-6-methyl-3-hydroxy-4-oxo-1-pyridyl)lysine	89
DL-*N*-(1,2-dihydro-2-oxo-4-pyrimidyl)serine	90
DL-*N*-(1,2-dihydro-2-oxo-4-pyrimidyl)valine	91
DL-*N*-(1,2-dihydro-2-oxo-4-pyrimidyl)leucine	91
DL-*N*-(1,2-dihydro-2-oxo-4-pyrimidyl)methionine	91
N-(1,2-dihydro-5-methyl-2-oxo-4-pyrimidyl)glycine	91
DL-*N*-(1,2-dihydro-5-methyl-2-oxo-4-pyrimidyl)valine	91
DL-*N*-(1,2-dihydro-5-methyl-2-oxo-4-pyrimidyl)leucine	91
DL-*N*-(1,2-dihydro-5-bromo-2-oxo-4-pyrimidyl)valine	91
DL-*N*-(1,2-dihydro-5-bromo-2-oxo-4-pyrimidyl)leucine	91
DL-*N*-(1,2-dihydro-5-iodo-2-oxo-4-pyrimidyl)leucine	91
N-(9-methylpurin-6-yl)glycine	92

J. Labelled Amino-acids.—The Strecker synthesis still continues to be an important route to labelled amino-acids: synthesis of [1-^{14}C]-DL-glutamic acid, [1-^{14}C]-DL-ornithine, and [1-^{14}C]-DL-arginine using this method and starting with 3-cyanopropionaldehyde has been reported.[93] A study of the reaction of α-ketocarboxylic acids with radioactive cyanide has shown that decarboxylation of the intermediate α-cyano-α-hydroxyacetic acid is much more rapid than hydrolysis. The reaction can therefore be employed for labelling, and [1-^{14}C]-DL-alanine and [1-^{14}C]-DL-glutamic acid have been synthesised from pyruvic and α-oxoglutaric acids respectively.[94] An alternative synthesis[95] of [1-^{14}C]-L-asparagine and [4-^{14}C]-L-aspartic acid from [4-^{14}C]-β-L-cyanoalanine has been described, and the preparation and enzymic resolution of *o*-hydroxy-[2-^{14}C]-DL-phenylalanine, starting from

[84] K. Matsumura, T. Kasai, and H. Tashiro, *Bull. Chem. Soc. Japan*, 1969, **42**, 1741.
[85] M. T. Doel, A. S. Jones, and N. Taylor, *Tetrahedron Letters*, 1969, 2285.
[86] M. Yu. Lidak, J. Sluke, and P. Yu. Shvachkin, *Khim. geterotsikl. Soedinenii*, 1968, 955.
[87] V. A. Brattsev and V. I. Stanko, *Zhur. obshchei Khim.*, 1969, **39**, 1175.
[88] V. A. Azimov, M. Ya. Uritskaya, and L. N. Yakhontov, *Zhur. Khim.-Farm.*, 1968, **2**, 16.
[89] P. A. Finot, R. Viani, J. Bricout, and J. Mauron, *Experientia*, 1969, **25**, 134.
[90] S. Hoffmann and E. Mühle, *Z. Chem.*, 1969, **9**, 112.
[91] R. A. Paegle, M. G. Plata, and M. Yu. Lidak, *Khim. geterotsikl. Soedinenii*, 1969, 558.
[92] G. B. Chheda, R. H. Hall, and H. Ross, *J. Org. Chem.*, 1969, **34**, 3492.
[93] I. Mező, I. Teplán, and J. Márton, *Acta Chim. Acad. Sci. Hung.*, 1969, **60**, 399.
[94] I. Teplán, I. Mező, L. Bursics, and J. Márton, *Acta Chim. Acad. Sci. Hung.*, 1969, **60**, 301.
[95] Y. H. Giza and C. Ressler, *J. Labelled Compounds*, 1969, **5**, 142.

ethyl-[2-^{14}C]-acetamidocyanoacetate published.[96] Syntheses of [4-^{14}C]-DL-tyrosine,[97] [β-^{14}C, ^{15}N, $^{14}CH_3$]-DL-*N*-α-methyltryptophan,[98] and [6-^{14}C]-L-4-azalysine [57] have also been reported.

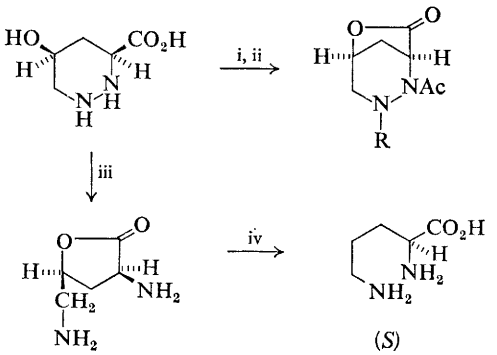
Further studies on enzymic transamination have resulted in an improved synthesis of [^{15}N]-L-aspartic acid and [^{15}N]-L-glutamic acid.[99] Efficient utilisation of [^{15}N]ammonium chloride is claimed, allowing preparation of any desired $^{15}N^{14}N$ ratio. [$\alpha\beta\beta$-$^2H_3$]-L-Glutamic acid has been obtained by incubating L-glutamic acid with deuterium oxide in the presence of pig-heart glutamate–oxaloacetate aminotransferase and catalytic amounts of pyridoxal-5′-phosphate and oxaloacetic acid.[100] However, under similar conditions with glutamate–pyruvate aminotransferase in the presence of pyruvic acid, [α-^{2}H]-L-glutamic acid is produced. Chemical synthesis of glutamic acid specifically labelled in the α and β positions with either deuterium or tritium has also been reported.[101] The absolute configuration of enzymatically tritiated glycine has been determined using D-amino-acid oxidase. The enantiomer prepared by incubating [α-$^3H_2$]glycine with serine hydroxymethylase in the absence of formaldehyde is (*R*)[2-^{3}H]glycine.[102] A synthesis of [2,3,3,4,4,5,5-$^2H_7$]-DL-lysine,[103] but without experimental details, has been reported, and the preparations of tritium-labelled *N*-methylated lysine derivatives [104] and tritiated phenylalanine [105] have been published.

3 Physical and Stereochemical Studies of Amino-acids

A. Determination of Absolute Configuration.—Perhaps the most interesting development in this field has been the reported application of n.m.r. spectroscopy for the determination of absolute configuration.[106] The spectra of the enantiomers of a given α-amino-acid methyl ester have been shown to differ appreciably in optically active 2,2,2-trifluorophenylethanol. This spectral non-equivalence is ascribed to strong solvent–solute interactions resulting in the formation of transitory diastereomeric solvates. It has been demonstrated that the method is widely applicable, not only to the determination of absolute configuration but also to the determination of optical purity. It is claimed that the method is more versatile than o.r.d. and c.d. in correlating absolute configuration. The decision as to whether or not this claim is justified must await wider application of the technique.

[96] C. Petitclerc, A. D'iorio, and N. L. Benoiton, *J. Labelled Compounds*, 1969, **5**, 265.
[97] J. H. Kim, C. R. Creger, and J. R. Couch, *J. Labelled Compounds*, 1969, **5**, 35.
[98] D. Gross, A. Unverricht, and H. R. Schütte, *Z. Chem.*, 1969, **9**, 64.
[99] J. A. Zintel, A. J. Williams, and R. S. Stuart, *Canad. J. Chem.*, 1969, **47**, 411.
[100] D. J. Whelan and G. J. Long, *Austral. J. Chem.*, 1969, **22**, 1779.
[101] M. C. Hochreiter and K. A. Schellenberg, *J. Labelled Compounds*, 1969, **5**, 270.
[102] M. Akhtar and P. M. Jordan, *Tetrahedron Letters*, 1969, 875.
[103] D. H. Rich, *Diss. Abs.*, 1968, **29**, *B*, 553.
[104] K. Hempel, H. W. Lange, and L. Birkhofer, *Z. physiol. Chem.*, 1969, **350**, 867.
[105] C. Gerday and W. G. Verly, *J. Labelled Compounds*, 1968, **4**, 334.
[106] W. H. Pirkle and S. D. Beare, *J. Amer. Chem. Soc.*, 1969, **91**, 5150.

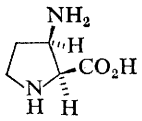
Chemical correlation with L-ornithine has established that the novel amino-acid (10) possesses the configuration (*S*) at C-3, and chemical considerations, such as facile lactone formation (Scheme 6), and n.m.r.



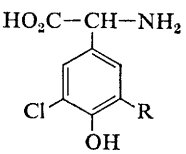
Reagents: i, 2,4-Dinitrofluorobenzene; ii, Ac_2O; iii, H_2–Pt; iv, P–HI. R = 2,4-dinitrophenyl

Scheme 6

data have been used to assign the (*S*) configuration at C-5.[16] The identification and configurational assignment of D-isoleucine, isolated from monamycin, was achieved by oxidation to the corresponding enantiomer of α-methylbutyraldehyde.[17] N.m.r. studies have enabled an assignment of the relative and hence the absolute configuration of L-*cis*-3-aminoproline (33) by a comparison of the coupling constant ($J_{2,3}$) with those of *cis*- and *trans*-3-hydroxyproline.[7] The β-hydroxyasparagine isolated from human urine has been identified as *erythro*-β-hydroxy-L-asparagine.[107] Hydrolysis of the antibiotic actinoidine yields *p*-hydroxyphenylglycine and 3-chloro-4-hydroxyphenylglycine (34), both of which have been shown to have the configuration (*R*).[108]



(33)



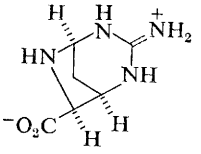
(34) R = H
(35) R = Cl

The determination of the absolute configuration of stereospecifically tritiated glycine was reported during the discussion on labelled amino-acids in section 2.

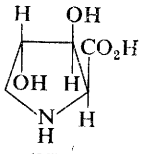
[107] H. Okai and N. Izumiya, *Bull. Chem. Soc. Japan*, 1969, **42**, 3550.
[108] N. N. Lomakina, V. A. Zenkova, and M. S. Yurina, *Khim. prirod. Soedinenii*, 1969, 43.

B. Crystal Structures of Amino-acids.—(See also Chapter 2, part II, section 1.) The crystal structures of L-cysteic acid,[109] L-valine hydrochloride,[110] L-ornithine hydrochloride,[111] DL-acetyl-leucine-*N*-methylamide,[112] and L-azetidine-2-carboxylic acid [113] have been described and the cell dimensions and space groups of L-tyrosine and L-tryptophan have been published.[114] A comparison of the β-synthesis and heavy-atom synthesis in the structure determination of L-arginine mono-hydrobromide monohydrate has also been reported.[115]

The structure (35), but not the absolute configuration, of an unusual amino-acid isolated from the hydrolysate of a new antibiotic was reported in a preliminary publication,[116] and it is of interest to note the similarity to the amino-acid (34) from the antibiotic actinoidine.[108] The proposed structures and configurations of viomycidine (36) [117, 118] and the proline



(36)



(37)

derivative (37) from diatom cell walls [119] have been confirmed by *X*-ray crystallographic analyses.

C. Optical Rotatory Dispersion (O.r.d.) and Circular Dichroism (C.d.).—(See also Chapter 2, part III, section 2B.) Extensive work in this area continues since a detailed knowledge of the optical rotatory properties of amino-acids is important for an understanding of similar properties of peptides and proteins. The o.r.d. has been measured for a series of derivatives of L-phenylalanine substituted in the aromatic ring. All the amino-acids exhibited positive Cotton effects corresponding to one of the aromatic electronic transitions and the n–π^* transition of the carboxy-group.[120] On the basis of these results it was concluded that the rotation of these compounds is hardly affected by the absorption bands of the aromatic

[109] Z. M. El Saffar, W. A. Hendrickson, and W. S. Koski, *Acta Cryst.*, 1969, **25**, *B*, 160.
[110] S. Thyagaraja Rao, *Z. Krist.*, 1969, **128**, 339.
[111] S. Guha, S. K. Mazumdar, and N. N. Saha, *Z. Krist.*, 1969, **129**, 84.
[112] T. Ichikawa and Y. Iitaka, *Acta Cryst.*, 1969, **B25**, 1824.
[113] H. M. Berman, *J. Amer. Chem. Soc.*, 1969, **91**, 6177.
[114] B. Khawas and G. S. R. K. Murti, *Acta Cryst.*, 1969, **B25**, 1006.
[115] K. K. Chacko and S. K. Mazumdar, *Z. Krist.*, 1969, **128**, 315.
[116] K. Kamiya, M. Nishikawa, H. Matsumaru, M. Asai, and K. Mizuno, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 2303.
[117] J. C. Floyd, J. A. Bertrand, and J. R. Dyer, *Chem. Comm.*, 1969, 998.
[118] G. Koyama, H. Nakamura, S. Omoto, T. Takita, K. Maeda, and Y. Iitaka, *J. Antibiotics*, 1969, **22**, 35.
[119] I. L. Karle, J. W. Daly, and B. Witkop, *Science*, 1969, **164**, 1401.
[120] I. Frič, V. Špirko, and K. Bláhá, *Coll. Czech. Chem. Comm.*, 1968, **33**, 4008.

chromophore but is dominated by optically active absorption bands below 230 nm. An observed linear relationship between the change in the molecular rotation differences and the Hammett σ values for a series of *para*-substituted phenylalanines tends to support this proposal.[121] A detailed analysis of the vibrational fine structure in the c.d. of L-phenylalanine and its derivatives has led to the suggestion that a conformational equilibrium exists at room temperature.[122]

A study of the solvent dependence of the o.r.d. of *N*-benzoyl- and *N*-benzyloxycarbonyl-L-α-amino-acids has been described, and it is claimed that it reflects the influence of the solvent polarity on the amide–iminol equilibrium.[123, 124] Similar studies on *N*-acetyl-L-phenylalanine-amide[125] and *N*-thiobenzoyl-L-α-amino-acids[126] have been reported. Further work on the c.d. of L-cystine and its derivatives has been published[127] in relation to investigations concerning the contribution of disulphide linkages to the c.d. of proteins.

D. Nuclear Magnetic Resonance (N.m.r.) Spectra.—The use of an optically active solvent in n.m.r. spectroscopy as a means of direct determination of absolute configuration of α-amino-acids has been discussed in the section on absolute configuration. Further details of a n.m.r. method of distinguishing α-amino-acids from isomers having amino-groups β or γ to the carboxyl function have been reported.[128] The methyl ester signal of the *N*-trityl α-amino-ester appears significantly upfield of the corresponding peak in the untritylated amino-ester. No significant shift is observed in the corresponding β- and γ-isomers. It is claimed that this difference can be used diagnostically for structure assignment.

A preliminary account of the ^{19}F n.m.r. of *N*-trifluoroacetyl derivatives of α-amino-acids has appeared.[129] The difference in the ^{19}F chemical shifts can be used to identify various amino-acids. These studies are primarily concerned with possible application to the peptide field and they will not be discussed in detail here. ^{23}Na n.m.r. studies have revealed, in contrast to a previous report,[130] that histidine is ineffective in complexing sodium.[131]

Inconsistencies between the chemical and physical properties of the anhydrides of argininosuccinic acid have been clarified by a detailed ^1H n.m.r. investigation.[132] N.m.r. techniques have been frequently employed

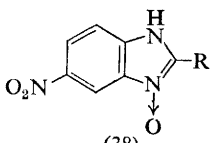
[121] K. Bláhá and I. Frič, *Coll. Czech. Chem. Comm.*, 1969, **34**, 2852.
[122] J. Horwitz, E. H. Strickland, and C. Billups, *J. Amer. Chem. Soc.*, 1969, **91**, 184.
[123] I. Z. Siemion and D. Konopinska, *Bull. Acad. Polon. Sci.*, 1968, **16**, 401.
[124] I. Z. Siemion, *Roczniki Chem.*, 1968, **42**, 741; 1969, **43**, 513.
[125] N. S. Simmons, A. O. Barbel, and A. N. Glazer, *Biopolymers*, 1969, **7**, 275.
[126] G. C. Barrett and A. R. Khokhar, *J. Chem. Soc.* (*C*), 1969, 1120.
[127] A. Imanishi and T. Isemura, *J. Biochem.* (*Japan*), 1969, **65**, 309.
[128] R. G. Webb, M. W. Haskell, and C. H. Stammer, *J. Org. Chem.*, 1969, **34**, 576.
[129] R. E. Sievers, E. Bayer, and P. Hunziker, *Nature*, 1969, **223**, 179.
[130] O. Jardetzky and J. E. Wertz, *J. Amer. Chem. Soc.*, 1969, **82**, 318.
[131] T. L. James and J. H. Noggle, *J. Amer. Chem. Soc.*, 1969, **91**, 3424.
[132] A. Kowalsky and S. Ratner, *Biochemistry*, 1969, **8**, 899.

for the characterisation of new amino-acids, particularly in differentiating between diastereoisomers (see, *e.g.*, references 7, 16, 19, and 119).

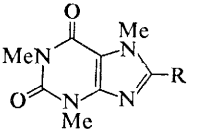
E. Mass Spectrometry.—The considerable current interest in the application of mass spectrometry to peptide sequence determination (see Chapter 2, part I, section 2C) prompted an investigation on the mass spectra of the volatile trimethylsilyl derivatives of a number of amino-acids and di- and tri-peptides. Trimethylsilylation appears to be unsuitable for sequence work, but is possibly useful in determining the molecular weights of unusual amino-acids.[133] The fragmentation processes observed were not distinctive, but all the compounds investigated gave fairly pronounced parent ions. The mass spectra of methyl- and phenyl-hydantoin amino-acids have been reported in relation to sequence analysis in peptides.[134, 135]

4 Chemical Studies of Amino-acids

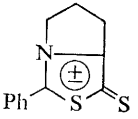
A. General Reactions.—Details of the photolysis of *N*-2,4-dinitrophenyl-amino-acids (Dnp-amino-acids) have been reported:[136] the structural requirements for the formation of 6-nitrobenzimidazole-1-oxide (38) and the precise reaction conditions are recorded. Photolysis of caffeine in the presence of aliphatic α-amino-acids leads to 8-alkylcaffeine derivatives (39).[137] The reaction of amino-acids and their *N*-acyl and *N*-thioacyl



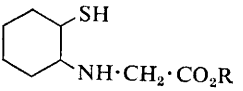
(38)



(39)



(40)

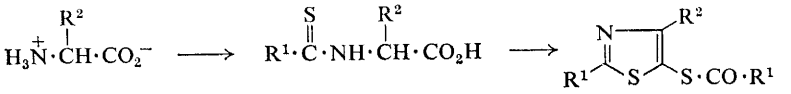


(41)

derivatives with thioacetic acid offers a simple new route to nitrogen- and sulphur-containing heterocyclic systems (Scheme 7).[138] For example, *N*-thiobenzoylproline gave the meso-ionic 2-phenylthiazole-5-thione (40).

[133] K. M. Baker, M. A. Shaw, and D. H. Williams, *Chem. Comm.*, 1969, 1108.
[134] F. F. Richards, W. T. Barnes, R. E. Lovins, R. Salomone, and M. D. Waterfield, *Nature*, 1969, **221**, 1241.
[135] B. W. Melvås, *Acta Chem. Scand.*, 1969, **23**, 1679.
[136] D. J. Neadle and R. J. Pollitt, *J. Chem. Soc.* (*C*), 1969, 2127.
[137] D. Elad and I. Rosenthal, *Chem. Comm.*, 1969, 905.
[138] G. C. Barrett, A. R. Khokhar, and J. R. Chapman, *Chem. Comm.*, 1969, 818.

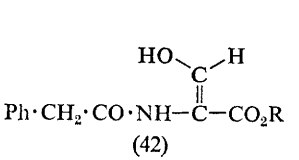
Cyclohexene sulphide and glycine ester react to give (41) and the products of its further reaction with cyclohexene sulphide.[139]



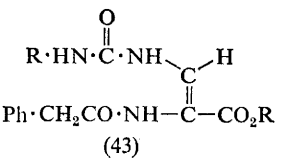
Scheme 7

NN-Dialkylamino-acids can be prepared in good yield from amino-acids by condensation with formaldehyde and catalytic reduction.[140] Oxidation of the products with aqueous hydrogen peroxide and acetic acid affords the corresponding *N*-oxides. Further work on the guanylation of amino-acids and derivatives using *N*-acyl-*S*-methylisothiourea has been published,[141] and a wide variety of amino-alcohols have been synthesised from α- and β-amino-acids and their esters.[142–144]

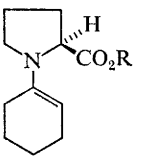
Interest continues in dehydro-amino-acid chemistry: condensation of urea and monosubstituted ureas with *N*-phenylacetyldehydroserine ethyl ester (ethyl penaldate) (42) affords the corresponding ureides (43), whose spectral properties are very similar to those of the chromophore of the antibiotic viomycin.[145] Dehydro-amino-acid derivatives have also been



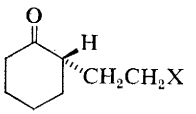
(42)



(43)



(44)



(45)

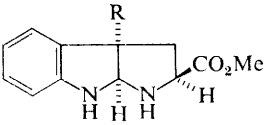
[139] K. Jankowski, R. Gauthier, and C. Berse, *Canad. J. Chem.*, 1969, **47**, 3179.
[140] Y. Ikutani, *Bull. Chem. Soc. Japan*, 1968, **41**, 1679.
[141] K. Nowak, *Roczniki Chem.*, 1969, **43**, 231.
[142] S. Hayashi, M. Furukawa, Y. Fujino, and T. Ohkawara, *Chem. and Pharm. Bull. (Japan)*, 1969, **17**, 1054.
[143] S. Hayashi, M. Furukawa, Y. Fujino, N. Matsuishi, and T. Ohkawara, *Chem. and Pharm. Bull. (Japan)*, 1969, **17**, 145.
[144] J. Szammer, *Acta Chim. Acad. Sci. Hung.*, 1969, **61**, 417.
[145] B. W. Bycroft, D. Cameron, A. Hassanali-Walji, and A. W. Johnson, *Tetrahedron Letters*, 1969, 2539.

prepared *via* 4-alkylidene-2-chloromethyloxazol-5-ones [146] and by β-elimination of the thiol group from cysteine derivatives.[147]

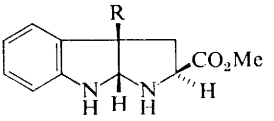
Amino-acids are being increasingly used in asymmetric syntheses: alkylation of the enamine (44) has been shown to occur from the least hindered side giving rise, on hydrolysis, to optically active cyclohexane derivatives (45) possessing the absolute configuration shown.[148, 149]

Acetylation of αω-diamino-acids can be achieved using 2-*N*-diacetylamino-cyclohexen-2-one,[150] and improved methods of esterification of *N*-acylamino-acids have been reported.[151, 152]

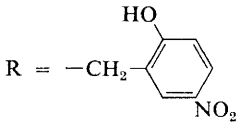
B. Other Reactions.—The increasing use of Koshland's reagent (2-hydroxy-5-nitrobenzyl bromide) for the modification and quantitative estimation of tryptophan in proteins has prompted two groups to investigate the reaction with L-tryptophan ester itself. [153, 154] The main products formed have been identified as (46) and (47), and acid treatment of both isomers results in

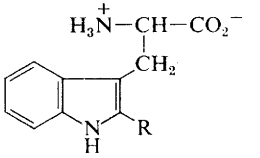


(46)



(47)





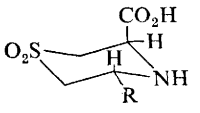
(48)

the formation of (48). The significance of these results, in relation to protein chemistry, was discussed but will not be elaborated here.

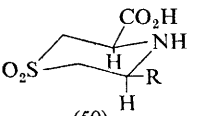
A detailed study of the reaction of *erythro*- and *threo*- forms of β-arylserine derivatives with thionyl chloride has revealed that both forms undergo

[146] H. Kurita, Y. Chigira, M. Masaki, and M. Ohta, *Bull. Chem. Soc. Japan*, 1968, **41**, 2758.
[147] L. Kisfaludy, A. Patthy, and M. Löw, *Acta Chim. Acad. Sci. Hung.*, 1969, **59**, 159.
[148] S. Yamada, K. Hiroi, and K. Achiwa, *Tetrahedron Letters*, 1969, 4233.
[149] S. Yamada and G. Otani, *Tetrahedron Letters*, 1969, 4237.
[150] K. M. Ermolaev, N. A. Popova, and A. I. Tochilkin, *Zhur. Vsesoyuz. Khim. obshch. im. D. I. Mendeleeva*, 1969, **14**, 357.
[151] V. V. Korshak, S. V. Rogozhin, and Yu. A. Davidovich, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1969, 977.
[152] S. V. Rogozhin, V. V. Korshak, and Yu. A. Davidovich, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1969, 2086.
[153] G. M. Loudon, D. Portsmouth, A. Lukton, and D. E. Koshland jun., *J. Amer. Chem. Soc.*, 1969, **91**, 2792.
[154] B. G. McFarland, Y. Inque, and K. Nakanishi, *Tetrahedron Letters*, 1969, 857.

substitution with retention of configuration. Chlorination of the *threo*-form proceeds directly, but the evidence suggests that the *erythro*- form goes through an intermediate *trans*-oxazoline.[155] The conversion of methionine sulphoxide to methionine sulphimine has been shown to take place predominantly with retention of configuration.[156] Further studies on the ease of formation of oxazolones from *N*-benzoyl-L-phenylalanine have been reported[157] and a number of ethyleneimides of *N*-acylated amino-acids prepared.[158] Reduction of selenocysteine can be achieved with either cysteine or glutathione,[159] and further work on *S*-sulpho-derivatives of amino-acids has shown that they can be cleanly desulphurised with Raney nickel.[160] The action of cyanide on the mixed disulphide of cysteine and penicillamine affords 2-amino-2-thiazoline-4-carboxylic acid, the reaction being considerably slower than with cystine.[161] When vinylic L-cysteine sulphones are treated with base, cyclisation occurs and two epimeric cyclic sulphones are formed.[162] The structures (49) and (50) have been proposed largely on the basis of n.m.r. evidence.



(49)

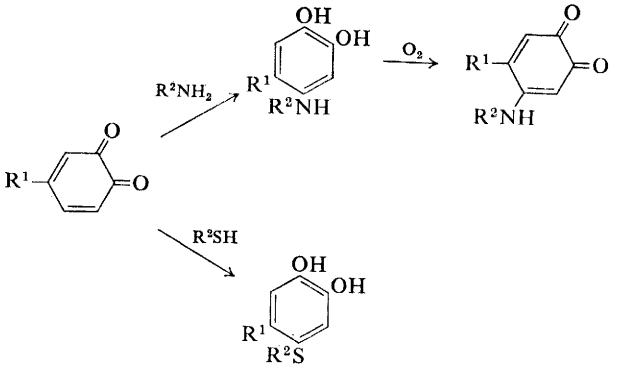


(50)

C. Non-enzymic Models of Biochemical Processes Involving Amino-acids.—The reactions of *o*-quinones with a number of amino-acids have been studied in relation to the biosynthesis of products resulting from quinone–amino-acid and quinone–protein interactions. The *o*-quinones produced by enzymic oxidation of chlorogenic and caffeic acids react with the α-amino-group of amino-acids, with the exception of cysteine and lysine (Scheme 8).[163] The ε-amino-group of lysine undergoes addition and oxidation in a similar manner, but cysteine reacts through the thiol group and the product is not oxidised further. Similar observations have been reported for the reaction of cysteine and *N*-acetyl-3,4-dihydroxyphenylalanine quinone, and the product may be implicated in the biogenesis of phaeomelanins.[164]

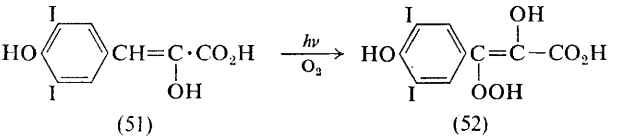
Further work on non-enzymic transamination reactions has shown that histidine is the most effective amino-acid to achieve this with α-oxoglutaric

[155] S. H. Pines, M. A. Kozlowski, and S. Karady, *J. Org. Chem.*, 1969, **34**, 1621.
[156] B. W. Christensen and A. Kjaer, *Chem. Comm.*, 1969, 934.
[157] M. Goodman and C. B. Glaser, *Tetrahedron Letters*, 1969, 3473.
[158] A. A. Yamontaite, G. K. Krasil'nikova, K. I. Karpavichus, and O. V. Kil'disheva, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1969, 1856.
[159] R. C. Dickson and A. L. Tappel, *Arch. Biochem. Biophys.*, 1969, **130**, 547.
[160] B. V. Aleksiev, Ch. P. Ivanov, P. G. Nišanjan, and S. Minčev, *Compt. rend. Acad. bulg. Sci.*, 1969, **22**, 281.
[161] T. R. C. Boyde, *J. Chem. Soc.* (*C*), 1968, 2751.
[162] J. F. Carson, L. E. Boggs, and R. Lundin, *J. Org. Chem.*, 1968, **33**, 3739.
[163] W. S. Pierpoint, *Biochem. J.*, 1969, **112**, 609.
[164] G. Prota, G. Scherillo, F. Napolano, and R. A. Nicolaus, *Gazzetta*, 1968, **98**, 495.

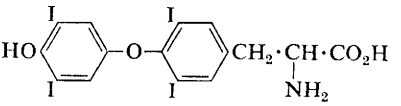


Scheme 8

acid.[165] The effectiveness of histidine in this reaction is attributed to the formation of the sparingly soluble metal ion–imidazolylpyruvic acid complexes which are end-products. The reaction of pyridoxal with amino-acids has been studied in methanol, and the position of equilibrium of the keto-enamine and enol-imine forms of the Schiff's bases followed spectroscopically.[166, 167] Amino-acids possessing a β-hydroxy-group form cyclic carbinolamines whereas cysteine afforded the product containing a thiazolidine ring. The effects of divalent metals were investigated and the change of the equilibrium was again followed spectroscopically. The nonenzymic decarboxylation employing pyridoxal-5-phosphoric acid has been reported for a number of amino-acids.[168] L-Serine-*O*-sulphate and L-threonine-*O*-sulphate are both deaminated in the presence of pyridoxal-5-phosphate to equimolar amounts of the corresponding keto-acid, ammonia, and sulphate.[169] The hydroperoxide (52) has been obtained in reasonable yield



(51) (52)



(53)

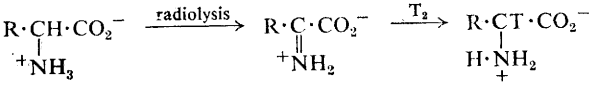
[165] V. M. Doctor and J. Oró, *Biochem. J.*, 1969, **112**, 691.
[166] Y. Matsushima, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 2046.
[167] Y. Matsushima, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 2143.
[168] H. A. Fischer, N. Seiler, J. Thobe, and G. Werner, *Naturwiss.*, 1968, **55**, 445.
[169] J. H. Thomas, K. S. Dodgson, and N. Tudrall, *Biochem. J.*, 1968, **110**, 687.

by the photo-oxidation of (51) and reacts with 4-hydroxy-3,5-di-iodophenylalanine to give thyroxine (53).[170] This represents a new synthetic route to thyroxine and a further example of the resemblance between biological and photochemical oxidations.

D. Effects of Electromagnetic Radiation on Amino-acids.—The effect of ionising radiation on amino-acids is still an area of active research. Although a number of studies in the solid state have been reported [171, 172] the emphasis appears to have shifted to reactions in aqueous solution. Particular interest has centred on sulphur-containing [173–175] and aromatic amino-acids.[176–178] The damage is predominantly caused by oxidising species (·OH), but tryptophan appears also to react rapidly with hydrated electrons and hydrogen atoms mainly by addition to the indole nucleus.[176] The effect of concentration on the products of radiolysis of phenylalanine [178] and the effect of the presence of metal ions on the irradiation of glycine have been published.[179] The comparison of chemically produced hydroxyl radicals with those resulting from radiolysis continues: in many cases the e.s.r. studies indicate that the same radicals are formed in both cases.[180–182]

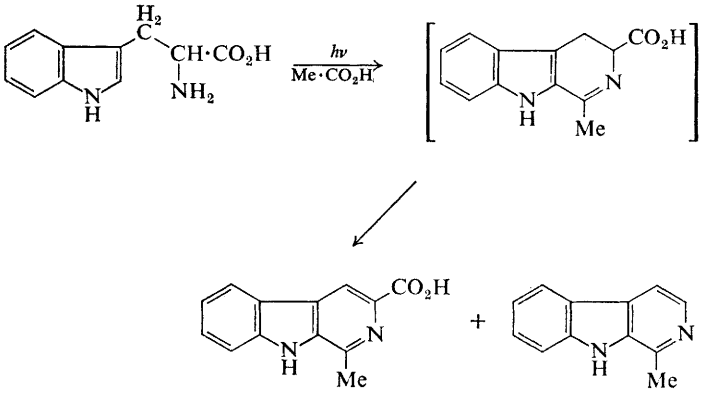
The labelling of optically active amino-acids by the Wilzbach technique [183] by exposure of the compound to tritium gas is of interest both to biological and fundamental radiation studies. Initial observation on a number of optically active amino-acids in the solid phase demonstrated that tritium is incorporated with appreciable racemisation.[184, 185] The mechanism outlined in Scheme 9 has been suggested to account for the racemisation, and the interesting observation that L- or D-phthaloylglutamic acid is tritiated in the solid phase with retention of configuration is cited as evidence supporting the proposed mechanism.[186] It is argued that in the case of the phthaloyl derivatives the formation of the imine intermediate is not possible, but whether or not the two reactions proceed by the same mechanism or not must await further study.

[170] T. Matsuura, K. Omura, and A. Nishinaga, *Chem. Comm.*, 1969, 366.
[171] D. B. Peterson, J. Holian, and W. M. Garrison, *J. Phys. Chem.*, 1969, **73**, 1568.
[172] M. Fujimoto, W. A. Seddon, and D. R. Smith, *J. Phys. Chem.*, 1968, **48**, 3345.
[173] J. Hähn, A. Peil, M. Pohl, S. Richter, G. Siegel, and E. Tapp, *Z. Naturforsch.*, 1968, **23b**, 1550.
[174] V. A. Sharpatyi, N. V. Zakatova, and G. A. Brodskaya, *Khim. vysok. Energii*, 1968, **2**, 473.
[175] W. Freyland and A. Müller, *Internat. J. Radiation Biol.*, 1968, **14**, 483.
[176] R. C. Armstrong, and A. J. Swallow, *Radiation Res.*, 1969, **40**, 563.
[177] O. H. Wheeler and R. Montalvo, *Radiation Res.*, 1969, **40**, 1.
[178] G. A. Brodskaya and V. A. Sharpatyi, *Zhur. fiz. Khim.*, 1969, **43**, 2390.
[179] G. A. Shagisultanova, L. A. Il'yukevich, and I. G. Sidorova, *Khim. vysok. Energii*, 1969, **3**, 364.
[180] G. K. Burk and G. Schoffa, *Internat. J. Protein Res.*, 1969, **1**, 113.
[181] D. J. Whelan, *J. Chem. Phys.*, 1968, **49**, 4734.
[182] R. Poupko, B. L. Silver, and A. Lowenstein, *Chem. Comm.*, 1968, 453.
[183] K. E. Wilzbach, *J. Amer. Chem. Soc.*, 1957, **79**, 1013.
[184] E. A. Evans, 'Tritium and its Compounds', Butterworths, London, 1966, p. 121.
[185] J. H. Parmentier, *J. Labelled Compounds*, 1966, **2**, 367.
[186] J. L. Garnett, S. W. Law, J. O'Keefe, B. Halpern, and K. Turnbull, *Chem. Comm.*, 1969, 323.



Scheme 9

The importance of the photochemistry of the aromatic amino-acids in relation to protein chemistry was pointed out in last year's Report and the interest in this field continues: in particular, the photochemical behaviours of tyrosine and tryptophan have come under detailed scrutiny.[187–190] It has recently been shown that when L-tryptophan is irradiated with visible light, in the presence of a sensitizer in acetic acid solution, it undergoes a photocyclodehydrogenation (Scheme 10), and it is claimed that the reaction may be of synthetic utility.[191] U.v. irradiation of glycine adsorbed on silica gel results in the formation of glycylglycine, an interesting example of photochemical peptide bond formation, although the possibility that the reaction might be thermally induced on the silica gel surface is not completely discounted.[192]



Scheme 10

5 Analytical Methods (See also Chapter 2, part I, section 2A)

The annual output of papers on amino-acids continues to contain a high percentage (approx. 20—25%) which are devoted to analytical methods. The majority of these describe modifications of established techniques or the determination of a specific amino-acid under certain conditions. In this section it is intended to do no more than to cite the majority of publications and outline any advances of general interest.

[187] M. E. McCarville and S. P. McGlynn, *Photochem. and Photobiol.*, 1969, **10**, 171.
[188] K. S. Asquith and D. E. Rivett, *Textile Res. J.*, 1969, **39**, 633.
[189] J. Feitelson, *Photochem. and Photobiol.*, 1969, **9**, 401.
[190] H. B. Steen, *Photochem. and Photobiol.*, 1969, **9**, 479.
[191] G. Jori, G. Galiazzo, and G. Gennari, *Photochem. and Photobiol.*, 1969, **9**, 179.
[192] M. M. Cosgrove, M. A. Collins, and R. A. Grant, *Photochem. and Photobiol.*, 1969, **10**, 141.

A. Ion-exchange Chromatography.—Interest in the use of trinitrobenzenesulphonic acid with automatic analysers for quantitative determination continues,[193, 194] and a new sampling device, which is claimed to give considerable reduction in analysis time, has been reported.[195] Several applications of computer analysis for data handling have been reported.[196–198] A simplified buffer system for gradient elution is claimed to reduce analysis by a third without sacrificing resolution or precision.[199] The determination of D- and L-amino-acids on ion-exchange resins by first converting to L,D- and L,L-dipeptides has been described.[200] A new method of separation of hydantoins on an anion-exchange resin appears effective,[201] and separations of water-soluble dinitrophenylamino-acids [202] and selenium-containing amino-acids [203] have also been reported.

B. Thin-layer Chromatography.—A further valuable general review on paper and thin-layer chromatography has appeared,[204] supplementing a more specific review of amino-acid t.l.c. reported last year.[205] Several papers have appeared on various aspects of t.l.c. on different types of cellulose powder [206–208] and an interesting application of silicic acid–glass fibre sheets for bio-autography reported.[209] Other papers dealing with t.l.c. of amino-acids have been published, *viz.*, a new solvent system for the separation of the 2,4-dinitrophenyl derivatives of leucine and isoleucine,[210] separation of dansyl derivatives of a number of amino-acids,[211] separation of sulphur-containing amino-acids,[212] the characterisation of iodo-amino-acids and their derivatives,[213] and finally the separation of the aromatic amino-acids on poly-(*N*-vinylpyrrolidone).[214]

C. Other Methods.—The preparation of suitable derivatives of amino-acids and their analysis by g.l.c. has been reviewed recently.[215] It has been

[193] J. Harmeyer, H. P. Sallmann, and L. Ayoub, *J. Chromatog.*, 1968, **32**, 258.
[194] H. H. Brown, *Clinical Chem.*, 1968, **14**, 967.
[195] J. J. T. Gerding and K. A. Peters, *J. Chromatog.*, 1969, **43**, 256.
[196] A. Mondino, *J. Chromatog.*, 1969, **41**, 156.
[197] G. Ertingshausen, H. J. Adler, and A. S. Reichler, *J. Chromatog.*, 1969, **42**, 355.
[198] R. E. Exss, H. D. Hill, and G. K. Summer, *J. Chromatog.*, 1969, **42**, 442.
[199] J. P. Ellis jun. and J. M. Prescott, *J. Chromatog.*, 1969, **43**, 260.
[200] J. M. Manning and S. Moore, *J. Biol. Chem.*, 1968, **243**, 5591.
[201] P. Hagel and J. J. T. Gerding, *Analyt. Biochem.*, 1969, **28**, 47.
[202] M. R. Heinrich and E. Bugna, *Analyt. Biochem.*, 1969, **28**, 1.
[203] J. L. Martin and M. L. Gerlach, *Analyt. Biochem.*, 1969, **29**, 257.
[204] K. Macek, I. M. Hais, J. Kopecký, and J. Gasparič, *J. Chromatog.*, 1968, Supplementary volume, 1.
[205] G. Patakai, *Chromatog. Rev.*, 1967, **9**, 23.
[206] C. L. De Ligny and E. C. M. Kok, *J. Chromatog.*, 1968, **38**, 224.
[207] C. Haworth and J. G. Heathcote, *J. Chromatog.*, 1969, **41**, 380.
[208] J. G. Heathcote and C. Haworth, *J. Chromatog.*, 1969, **43**, 84.
[209] G. H. Wagman and J. V. Bailey, *J. Chromatog.*, 1969, **41**, 263.
[210] J. Blass, *Chromatographia*, 1969, 178.
[211] D. Stehelin and H. Duranton, *J. Chromatog.*, 1969, **43**, 93.
[212] B. States and S. Segal, *Analyt. Biochem.*, 1969, **27**, 323.
[213] R. H. Osborn and T. H. Simpson, *J. Chromatog.*, 1969, **40**, 219.
[214] T. M. Dougherty and A. I. Schepartz, *J. Chromatog.*, 1969, **42**, 415.
[215] C. W. Gehrke and D. L. Stalling, *Separation Sci.*, 1968, **2**, 101.

claimed that an initial purification of *N*-trifluoroacetyl-amino-acid esters on silica removes any traces of unreacted amino-acid, which frequently reduce the resolution and occasionally produce multiple peaks in g.l.c. analysis.[216] Further studies on pertrimethylsilyl derivatives for the g.l.c. of iodine- and sulphur-containing amino-acids have been reported.[217] Other topics in analytical chemistry discussed include: separation of acidic amino-acids by high-voltage electrophoresis,[218] negative chromatogram photometry,[219] determination and identification of amino-acids by thermometric titration and n.m.r. spectroscopy,[220] direct fluorometric scanning,[221] determination of cysteine and cystine using *N*-bromosuccinimide,[222] and improvements in analysis in the nanomole range.[223]

D. Determination of Specific Amino-acids.—Papers on the determination of the following amino-acids have appeared: lysine,[224–226] asparagine,[227–229] arginine,[229] tryptophan,[230, 231] $\alpha\varepsilon$-diaminopimelic acid,[232, 233] histidine,[234] hydroxyproline,[235] phenylalanine,[236] cystine,[237] selenocysteine and selenomethionine,[238] 3-hydroxyglutamic acid,[239] and thyroxine.[240] The free amino-acid contents of the following have been examined: the peel and flesh of *Solanum tuberosum*,[241] grass pollen,[242] seeds of *Abrus precatorius*,[243] and humic acid.[244]

[216] M. D. Waterfield and A. Del Favero, *J. Chromatog.*, 1969, **40**, 294.
[217] F. Shahrokhi, *Diss. Abs.*, 1969, **29**, *B*, 2793.
[218] P. J. Peterson, *J. Chromatog.*, 1968, **38**, 301.
[219] S. Burzyński and Z. Czerniak, *Chem. analit.*, 1969, **14**, 667.
[220] B. Sen and W. C. Wu, *Analyt. Chim. Acta*, 1969, **46**, 37.
[221] G. Pataki and K. T. Wang, *J. Chromatog.*, 1968, **37**, 499.
[222] R. J. Thibert, M. Sarwar, and J. E. Carroll, *Mikrochim. Acta*, 1969, 615.
[223] K. Krejci and W. Machleidt, *Z. physiol. Chem.*, 1969, **350**, 981.
[224] F. Quevedo and C. Vega, *Bol. Soc. quím. Peru*, 1968, **34**, 42.
[225] E. Villegas, C. E. McDonald, and K. A. Gilles, *Cereal Chem.*, 1968, **45**, 432.
[226] M. L. Kakade and I. E. Liener, *Analyt. Biochem.*, 1969, **27**, 273.
[227] A. K. Saxena and O. C. Saxena, *Microchem. J.*, 1969, **14**, 224.
[228] M. L. Verma and O. C. Saxena, *Chim. analyt.*, 1969, **51**, 217.
[229] O. C. Saxena, *Microchem. J.*, 1968, **13**, 571.
[230] R. D. Lolashvili, *Priklad. Biokhim. i. Microbiol.*, 1969, **5**, 335.
[231] P. Slump and H. A. W. Schreuder, *Analyt. Biochem.*, 1969, **27**, 182.
[232] N. P. Sen, E. Somers, and R. C. O'Brien, *Analyt. Biochem.*, 1969, **28**, 345.
[233] N. P. Sen, E. Somers, and R. C. O'Brien, *Analyt. Biochem.*, 1968, **26**, 457.
[234] J. A. Ambrose, A. Crimm, J. Burton, K. Paullin, and C. Ross, *Clinical Chem.*, 1969, **15**, 361.
[235] I. Bergman and R. Loxley, *Analyst*, 1969, **94**, 575.
[236] J. A. Ambrose, *Clinical Chem.*, 1969, **15**, 15.
[237] G. R. Limb and A. M. Dollar, *Analyt. Biochem.*, 1969, **29**, 100.
[238] J. V. Benson jun. and J. A. Patterson, *Analyt. Biochem.*, 1969, **29**, 130.
[239] G. Lindstedt, *J. Chromatog.*, 1969, **40**, 316.
[240] O. V. Tarasonva, Z. A. Shevchenko, and I. A. Favorskaya, *Vestnik Leningrad. Univ.* (*Fiz. Khim.*), 1969, 154.
[241] E. Bancher, J. Washüttl, and G. Widtmann, *Z. Lebensm.-Untersuch.*, 1968, **138**, 201.
[242] E. J. Shellard and G. H. Jolliffe, *J. Chromatog.*, 1968, **38**, 257.
[243] L. D. Sharma, D. S. Misra, and B. K. Soni, *J. Chinese Chem. Soc.* (*Formosa*), 1968, **15**, 67.
[244] T. J. Piper and A. M. Posner, *Soil Sci.*, 1968, **106**, 188.

2
Structural Investigation of Peptides and Proteins

By R. N. PERHAM, C. C. F. BLAKE, A. R. PEACOCKE

PART I: Primary Structures and Chemical Modification *by R. N. Perham*

1 Introduction

The past twelve months, in the nature of things, have witnessed significant additions to our knowledge of the primary structures of peptides and proteins, as a glance at the ever more bulky yet ever more admirable 'Atlas of Protein Sequence and Structure'[1] will confirm. This has been achieved by the application and refinement, often elegant, of the principles and methods reviewed last year,[2] rather than by the introduction of some startling new technique. That revolution at least, welcome though it might be to many, has not yet reached the campus!

Last year, too, your Reporter sought forgiveness for his inability to cover in depth, within the confines of a single article, every development in what continues to be a rapidly expanding field. If such a plea was necessary then, the more so is it now, and attention will be focussed again on a broad selection of topics. The general distinction between methods and results, though they often will properly overlap, is retained.

2 Methods

The twin aims of amino-acid sequence determination, to multiply the sensitivity and the automation of the methods and, thereby, unlike the nation, to increase the joy,[3] have been well served. Reports of such improvements constitute a large part of the relevant literature.

A. Amino-acid Analysis.—(See also chapter 1, section 5.) The inclusion of 2-mercaptoethanol during acid hydrolysis of proteins has been reported[4] to give improved recovery of methionine without effect on other amino-acids. Similarly, the presence of thioglycollic acid (2%) during acid hydrolysis under high vacuum has been shown[5] to allow high recoveries

[1] 'Atlas of Protein Sequence and Structure,' ed. M. O. Dayhoff, National Biomedical Research Foundation, Maryland, vol. 4, 1969.
[2] 'Amino-acids, Peptides and Proteins', ed. G. T. Young, vol. 1, The Chemical Society, London, 1969.
[3] Isaiah, Ch. 9, v. 3.
[4] H. T. Keutmann and J. T. Potts jun., *Analyt. Biochem.*, 1969, **29**, 175.
[5] H. Matsubura and R. M. Sasaki, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 175.

(up to 85%) of tryptophan from proteins: if generally applicable, this would simplify tryptophan analysis considerably.

A micro method has been described [6] for the determination of the total sulphur content of proteins by flame spectrophotometry which has the advantage that it is suitable for measuring the 'labile' sulphur such as one finds in the ferredoxins additional to the sulphur-containing amino-acids.[7] The quantitative micro-determination of thiol groups in proteins with the mercurial Mercury Orange, used previously in histochemical work, has also been reported,[8] although it appears to offer little advantage over the classical well-documented mercurials.[9] Another approach to thiol group analysis in proteins which allows spectral impurities in the original reaction mixture is the use of ^{14}C-labelled chloromercuribenzoate.[10] After reaction with the protein, the excess reagent is separated from the labelled protein by gel filtration and the bound reagent is then estimated from its specific radioactivity. A very rapid fluorimetric assay of amino-acids, capable of analysis of 0·01—50 μg of amino-acid, has also been described.[11] It uses D- or L-amino-acid oxidase to generate hydrogen peroxide from the amino-acid, which in turn oxidises homovanillic acid to a fluorescent derivative.

Ion-exchange Chromatography. Automatic amino-acid analysis on ion-exchange columns continues to hold the affection of a majority of protein chemists, understandably in view of the continual reports of efforts to improve it.[12] Using standard stepwise elution conditions but operating at higher temperatures than normal, full separations have been achieved in as little as 1·5 hr.[13] A simplified buffer for gradient elution of amino-acids has been described [14] and it has been suggested [15] that making up the buffers from citric acid rather than sodium citrate reduces the rise in the base-line caused by dissolved ammonia. Simple improvements to increase the sensitivity and rapidity of analysis on single-column machines have been reported.[16] Gradient elution of amino-acids in just over 1 hour, with up to 40 samples run sequentially unattended,[17] illustrates the trend to automation and plans have been given of an automatic sample loader for amino-acid analysers.[18] To increase the sensitivity, chromatography on columns

[6] K. Gersonde, *Analyt. Biochem.*, 1968, **25**, 459.
[7] R. Malkin and J. C. Rabinowitz, *Ann. Rev. Biochem.*, 1967, **36**, 113.
[8] H. Sakai, *Analyt. Biochem.*, 1968, **26**, 269.
[9] J. F. Riordan and B. L. Vallee, 'Methods in Enzymology', vol. 11, ed. C. H. W. Hirs, Academic Press, New York, 1967, p. 541.
[10] V. G. Erwin and P. L. Pedersen, *Analyt. Biochem.*, 1968, **25**, 477.
[11] G. G. Guilbault and J. E. Hieserman, *Analyt. Biochem.*, 1968, **26**, 1.
[12] D. L. Iur and I. E. Suggs, *Chromatographia*, 1968, **1**, 488.
[13] V. A. Bairos and M. A. Bairos, *Biochim. Biophys. Acta*, 1968, **168**, 567.
[14] J. P. Ellis jun. and J. M. Prescott, *J. Chromatog.*, 1969, **43**, 260.
[15] A. Wainer and J. J. Richter, *J. Chromatog.*, 1968, **39**, 79.
[16] H. Tschesche and C. Frank, *J. Chromatog.*, 1969, **40**, 296.
[17] G. Ertingshausen, H. J. Adler, and A. S. Reichler, *J. Chromatog.*, 1969, **42**, 355.
[18] A. R. Thomson and J. W. Eveleigh, *Analyt. Chem.*, 1969, **41**, 1073.

0·5 mm internal diameter has been attempted:[19] 25 picomoles may yet be sufficient for quantitative analysis.

A new method of automatic analysis in which quantification of the amino-acids is achieved simply by measuring peak heights has been described [20] and the value of lithium citrate buffers in resolving asparagine and glutamine emphasized.[21] This latter observation was first made by Moore and Stein [22] some years ago and is confirmed in a number of recent papers.[23–25] Lithium citrate buffers may have other advantages in the resolution of the rarer basic amino-acids *e.g.* 1- and 3-methylhistidine.[21] Glutamine may also be determined with reasonable accuracy by the action of the enzyme transglutaminase from guinea-pig liver on suitable digests of the protein.[26] The use of trinitrobenzenesulphonic acid as an alternative to ninhydrin in automatic analysers continues to receive attention.[27]

'Making dictionaries', said the good Doctor, 'is dull work.'[28] The same might well be said of calculating amino-acid analyses. To counter this, much effort still goes into devising suitable computer procedures.[20, 29] Some methods allow calculation of the minimum molecular weight consistent with the recorded analysis.[30, 31]

Suitable chromatographic conditions for resolving many unusual amino-acids have been reported during the year. Conditions for the fractionation [32] of the *N*-methylated lysines and their preparation [33] in a tritiated form have been described: so have the resolution of ornithine and lysine [34] and the separation of the isologous sulphur- and selenium-containing amino-acids.[35] A system has been developed for use in the amino-acid analysis of collagen,[36] which is ordinarily complicated by unusually large amounts of glycine and the presence of rare amino-acids such as 3- and 4-hydroxyproline and hydroxylysine. The simultaneous determination of amino-acids, glucosamine, and galactosamine in hydrolysates of glycoproteins has been described [37] and the detection of other less

[19] K. Krejci and W. Machleidt, *Z. physiol. Chem.*, 1969, **350**, 981.
[20] A. Mondino, *J. Chromatog.*, 1969, **41**, 156.
[21] A. Mondino, *J. Chromatog.*, 1969, **39**, 262.
[22] S. Moore and W. H. Stein, *J. Biol. Chem.*, 1951, **192**, 663.
[23] J. V. Benson, M. J. Gordon, and J. A. Patterson, *Analyt. Biochem.*, 1967, **18**, 228.
[24] T. L. Perry, D. Stedman, and S. Hansen, *J. Chromatog.*, 1968, **38**, 460.
[25] J. H. Peters, B. J. Berridge jun., J. G. Cummings, and S. C. Lin, *Analyt. Biochem.*, 1968, **23**, 459.
[26] H. Toda and J. E. Folk, *Biochim. Biophys. Acta*, 1969, **175**, 427.
[27] H. H. Brown, *Clinical Chem.*, 1968, **14**, 967.
[28] S. Johnson, 'Dictionary of the English Language': Dull; 8.
[29a] J. J. T. Gerding, *Internat. J. Protein Res.*, 1969, **1**, 169; [b] R. E. Exss, H. D. Hill, and G. K. Summer, *J. Chromatog.*, 1969, **42**, 442.
[30] E. P. Katz, *Analyt. Biochem.*, 1968, **25**, 417.
[31] M. DeLaage, *Biochim. Biophys. Acta*, 1968, **168**, 573.
[32] K. Hempel, H.-W. Lange, and L. Birkhofer, *Z. physiol. Chem.*, 1969, **350**, 966.
[33] K. Hempel, H.-W. Lange, and L. Birkhofer, *Z. physiol. Chem.*, 1969, **350**, 867.
[34] K. Pfordte, *J. Chromatog.*, 1969, **39**, 506.
[35] R. Walter, D. H. Schlesinger, and I. L. Schwartz, *Analyt. Biochem.*, 1969, **27**, 231.
[36] W. Bandlow and A. Nordwig, *J. Chromatog.*, 1969, **39**, 328.
[37] M. Monsigny, *Bull. Soc. Chim. biol.*, 1968, **50**, 2188.

common amino-acids continues. For example, L-5-hydroxylysine is found in shell membranes of the hen's egg,[38] 2,3-dihydroxyproline in the diatom cell wall,[39, 40] and hydroxyproline in glycopeptides from the primary cell walls of plants.[41]

The versatility of ion-exchange chromatography is further revealed in special cases. It has been pointed out [42] that if weak (carboxylic) cation exchangers are used in the analyser instead of strong (sulphonic) exchangers, the same apparatus may be used for rapid analysis of amines. Further, the precise determination of D- and L-isomers in a mixture is possible following conversion to the diastereoisomeric mixture of dipeptides with an L-amino-acid *N*-carboxyanhydride. The conversion is easy and quantitative and elution conditions for all the expected dipeptides have been found.[43]

High-voltage Electrophoresis and Thin-layer Chromatography. In view of their speed, simplicity, and cheapness, these methods continue to be developed. Clearly, to achieve maximal separation, electrophoresis should be conducted in a buffer of pH near a pK_a common to the components of a mixture. This principle has been recognised in the separation of acidic amino-acids found in plant material by paper electrophoresis at pH 3·4 or 5·3 followed by chromatography in butan-1-ol–acetic acid–water.[44] A bibliography of paper and thin-layer separations in the years 1961—1965 has been published [45] and new methods of quantitative analysis of amino-acids by t.l.c. have been described.[46, 47] The particular separation of aromatic amino-acid derivatives on polyamide layers [48] and the characterisation of iodo-amino-acids by t.l.c.[49] are nice examples of the high resolution and sensitivity attainable. Like ion-exchange chromatography, t.l.c. is also readily developed into a micro method for determining amino-acid configuration: the diastereoisomeric dipeptides are first generated and then resolved.[50]

Gas–Liquid Chromatography. The potential advantages of amino-acid analysis by g.l.c. are well known and recent advances have been discussed.[51, 52] A single-column analysis of amino-acids as the *N*-trifluoroacetyl

[38] J. K. Candlish and K. K. Scougall, *Internat. J. Protein Res.*, 1969, **1**, 299.
[39] T. Nakajima and B. E. Volcani, *Science*, 1969, **164**, 1400.
[40] F. Irreverre and B. Witkop, *J. Chromatog.*, 1969, **43**, 127.
[41] D. T. A. Lamport, *Biochemistry*, 1969, **8**, 1155.
[42] R. A. Wall, *J. Chromatog.*, 1968, **37**, 549.
[43] J. M. Manning and S. Moore, *J. Biol. Chem.*, 1968, **243**, 5591.
[44] P. J. Peterson, *J. Chromatog.*, 1968, **38**, 301.
[45] K. Macek, I. M. Hais, J. Kopecky, and J. Gasparic, *J. Chromatog.*, 1968, Supplementary Volume, 1.
[46a] C. Haworth and J. G. Heathcote, *J. Chromatog.*, 1969, **41**, 380; [b] J. G. Heathcote and C. Haworth, *ibid.*, 1969, **43**, 84.
[47] M. E. Clark, *Analyst*, 1968, **93**, 810.
[48] J. D. Sapira, *J. Chromatog.*, 1969, **42**, 134.
[49] R. H. Osborn and T. H. Simpson, *J. Chromatog.*, 1969, **40**, 219.
[50] T. Wieland and A. Buku, *Analyt. Biochem.*, 1969, **27**, 378.
[51] S. F. Herb, *J. Amer. Oil Chemists' Soc.*, 1968, **45**, 784.
[52] D. Roach and C. W. Gehrke, *J. Chromatog.*, 1969, **43**, 303.

derivatives of their n-butyl esters in nanomolar quantities has been described [53] and the chromatography of the same derivatives of radioactive amino-acids reported.[54] To facilitate analysis, the *N*-trifluoroacetyl derivatives of the amino-acid n-butyl esters can be separated from unreacted material by chromatography on columns of silica gel before submission to g.l.c.[55] The methyl esters are, in fact, potentially better, since their greater volatility means chromatography can be effected at lower temperatures. On the other hand, the greater volatility makes quantitative preparation of the derivatives more difficult. Successful chromatography of the trifluoroacetyl derivatives of the methyl esters has, however, been accomplished recently.[56] Moreover, some of the difficulties involved in the chromatography of histidine have been resolved.[57] Another analysis of iodine-containing amino-acids by g.l.c. separation of the trimethylsilyl derivatives has also been reported.[58]

However, the fact remains that another year has gone by without a substantial number of workers shifting to g.l.c. for routine amino-acid analysis. Doubtless the reasons adduced for this last year [2] still obtain: it would still be rash to suppose that they will always do so, although there is no immediate prospect of change.

B. End-group Analysis.—In view of the continual discovery of blocked *N*-terminal residues in proteins, it is worth noting that a highly sensitive technique for the analysis of acetyl groups in biological material by g.l.c. has been published.[59] A micro method for the detection of formyl and acetyl groups has also been described.[60] The hydrazide resulting from hydrazinolysis of the acylated material is allowed to react with dansyl chloride and the fluorescent derivative produced may then be detected and identified by t.l.c. Another common blocked *N*-terminal residue is that of pyrrolidonecarboxylic acid. While it may most conveniently be removed specifically by an enzyme, pyrrolidonylpeptidase,[61] a chemical method for dealing with it has been suggested. The pyrrolidonecarboxylic acid may be converted to *N*-terminal proline by reduction with diborane in tetrahydrofuran or tetramethylurea in yields of up to 46%. Unfortunately, the reaction is not entirely specific and some carboxy-groups and peptide bonds may be affected.[62]

There have been promising descriptions of the separation of Pth-amino-acids by g.l.c. and subsequent identification by mass spectrometry [63] and

[53] W. J. McBride jun. and J. D. Klingman, *Analyt. Biochem.*, 1968, **25**, 109.
[54] A. Del Favero, A. Darbre, and M. Waterfield, *J. Chromatog.*, 1969, **40**, 213.
[55] M. D. Waterfield and A. Del Favero, *J. Chromatog.*, 1969, **40**, 294.
[56] A. Islam and A. Darbre, *J. Chromatog.*, 1969, **43**, 11.
[57] D. Roach, C. W. Gehrke, and R. W. Zumwalt, *J. Chromatog.*, 1969, **43**, 311.
[58] K. Funakoshi and H. J. Cahnmann, *Analyt. Biochem.*, 1969, **27**, 150.
[59] G. R. Shepherd and B. J. Noland, *Analyt. Biochem.*, 1968, **26**, 325.
[60] G. Schmer and G. Kreil, *Analyt. Biochem.*, 1969, **29**, 186.
[61] R. W. Armentrout and R. F. Doolittle, *Arch. Biochem. Biophys.*, 1969, **132**, 80.
[62] S. Takahashi and L. A. Cohen, *Biochemistry*, 1969, **8**, 864.
[63] B. W. Melvas, *Acta Chem. Scand.*, 1969, **23**, 1679.

of the separation by g.l.c. of all Pth-amino-acids, with the usual exception of arginine, as their trimethylsilyl derivatives.[64] Such methods, though employing what might be considered perhaps rather expensive equipment for the task, do offer rapid, unequivocal, and quantitative analysis. In this connexion, it is interesting that the virtues of methyl isothiocyanate, rather than the more usual phenyl compound, for the Edman degradation have been stressed.[65, 66] In its favour, it is noted that it is quite volatile (and therefore easier than phenyl isothiocyanate to remove when necessary), that the methyl thiohydantoins of all amino-acids except arginine can be readily separated by g.l.c., and that very sensitive analyses are possible by mass spectrometry.

In studies of collagen and various structural proteins of hair, techniques have been developed for the separation of the water-soluble [67] and ether-soluble [68] Dnp-amino-acids on nylon powder columns, and the ion-exchange chromatography on IRC-50 of some Dnp-amino-acids has been reported.[69] More solvent systems for the separation of Dnp-amino-acids by t.l.c. have been described [70] and some attention paid to the direct fluorimetric scanning of Dns-, Dnp-, and Pth-amino-acid derivatives.[71] Both fluorescence and fluorescence-quenching techniques were employed and chromatography on silica gel and polyamide thin layers compared. The latter were found preferable since they gave more constant base-lines in quenching experiments. Such direct analysis could be very valuable when perfected.

Separations of Dns-amino-acids continue to abound: by chromatography on close-textured paper,[72] by t.l.c.,[73] and by thin-layer electrophoresis on cellulose.[74] However, little improvement seems to have been made on the original highly sensitive two-dimensional chromatography on polyamide layer sheets [75] which is now being extended, quite remarkably, to the identification of Dns-amino-acids in the picomole range.[76] A warning has also been sounded that calculation of the number of dansyl groups bound to proteins from some optical measurements may well be incorrect and certainly at variance with the number estimated from the specific radioactivity of ^{14}C-labelled dansyl chloride.[77]

[64] J. J. Pisano and T. J. Bronzert, *J. Biol. Chem.*, 1969, **244**, 5597.
[65] F. F. Richards, W. T. Barnes, R. E. Lovins, R. Salomone, and M. D. Waterfield, *Nature*, 1969, **221**, 1241.
[66] M. Waterfield and E. Haber, *Biochemistry*, 1970, **9**, 832.
[67] H. Beyer and U. Schenk, *J. Chromatog.*, 1969, **39**, 482.
[68] H. Beyer and U. Schenk, *J. Chromatog.*, 1969, **39**, 491.
[69] M. R. Heinrich and E. Bugna, *Analyt. Biochem.*, 1969, **28**, 1.
[70] H. Tanaka, K. Yoshikawa, and H. Shinomura, *J. Agric. Chem. Soc. Japan*, 1968, **42**, 735.
[71] G. Pataki and K.-T. Wang, *J. Chromatog.*, 1968, **37**, 499.
[72] L. Munier, C. Thommegay, and A. Drapier, *Compt. rend.*, 1969, **268**, *D*, 1882.
[73] D. Stehelin and H. Duranton, *J. Chromatog.*, 1969, **43**, 93.
[74] B. P. Sloan, *J. Chromatog.*, 1969, **42**, 426.
[75] K. R. Woods and K.-T. Wang, *Biochim. Biophys. Acta*, 1967, **133**, 369.
[76] V. Neuhoff, F. von der Haar, E. Schlimme, and M. Weise, *Z. physiol. Chem.*, 1969, **350**, 121.
[77] R. F. Chen, *Analyt. Biochem.*, 1968, **25**, 412.

For use with other techniques of *N*-terminal analysis, the ion-exchange chromatography of amino-acid hydantoins produced in the cyanate method has been described [78] as have the c.d. properties of *N*-thiobenzoyl-L-α-amino-acids [79] and the separation by t.l.c. of the *N*-thiobenzoylamino-acid anilides.[80]

A further paper has appeared [81] on *C*-terminal analysis by means of conversion of the protein to its peptidylthiohydantoin at the *C*-terminus with ammonium thiocyanate and acetic anhydride and subsequent cleavage of the thiohydantoin. The use of 12M-HCl as the cleavage reagent was investigated and, not surprisingly, a maximum of two or three steps of unequivocal sequential degradation could be achieved. As a routine method it seems unlikely to attain wide currency.

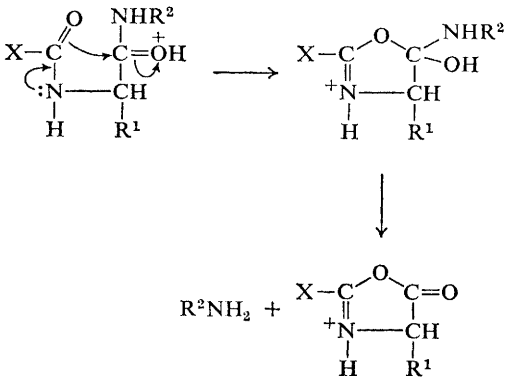
C. Sequential Degradation and Mass Spectrometry.—(See also chapter 1, section 3E.) The Edman degradation and its high-sensitivity version, the dansyl-Edman procedure, account for a large proportion of the amino-acid sequences established last year. Commercial versions of the automatic 'sequenator' are now on the market—a doubtful mark of respectability, perhaps—and encouraging reports of sequences established with the original machine continue to come from Edman's laboratory. Some improvements to the machine in the way of reagent systems have been briefly reported [82] and it is worth recalling that the more volatile methyl isothiocyanate could perhaps be substituted to advantage.[65, 66]

A drawback to the Edman degradation is the gradual random cleavage of the polypeptide chain brought about during the exposure to anhydrous acid required to secure release of the *N*-terminal residue as the 2-anilinothiazolin-5-one. A possible mechanism for this comes from an examination of the effect of anhydrous trifluoroacetic acid on *N*-acylated peptides.[83] The *N*-terminal residue is cleaved off to yield an oxazolone and a new *N*-terminal group (Scheme 1). The rate of cleavage varies with the acylating group and it might be that some interior peptide bonds could be labilised by a similar mechanism. It is sometimes, though fortunately rarely, found that an Edman degradation refuses to go through an Asp-Gly or Asn-Gly sequence and it has been suggested [84] that the lack of a side-chain in glycine might be a contributing factor in allowing the Asx-Gly sequence to form the cyclic imide and thence the more stable β-aspartyl linkage, thereby inhibiting the degradation (although it must be confessed that this does not explain why the rearrangement happens only in isolated cases). It is

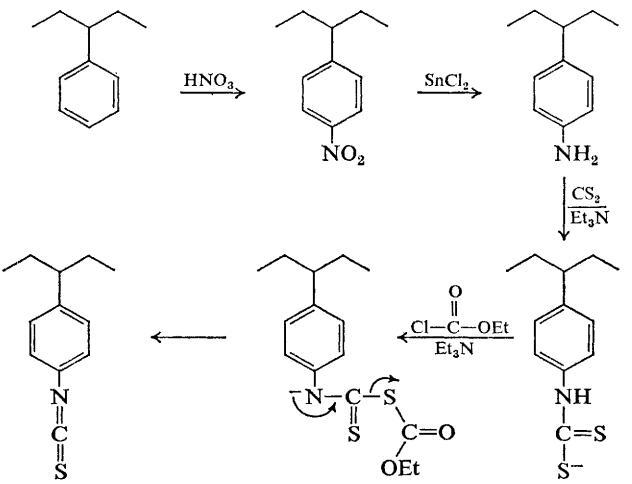
[78] P. Hagel and J. J. T. Gerding, *Analyt. Biochem.*, 1969, **28**, 47.
[79] G. C. Barrett and A. R. Khokhar, *J. Chem. Soc.* (*C*), 1969, 1120.
[80] G. C. Barrett and A. R. Khokhar, *J. Chromatog.*, 1969, **39**, 47.
[81] L. D. Cromwell and G. R. Stark, *Biochemistry*, 1969, **8**, 4735.
[82] H. D. Niall, H. Penhasi, P. Gilbert, R. C. Myers, F. G. Williams and J. T. Potts jun., *Fed. Proc.*, 1969, **28**, 661.
[83] A. Previero, M.-A. Coletti-Previero, and L.-G. Barry, *Biochim. Biophys. Acta*, 1969, **181**, 361.
[84] R. N. Perham, Ph.D. Thesis, University of Cambridge, 1964.

interesting that the same reason has been adduced for problems with the Asp-Gly sequence during solid-phase peptide synthesis.[85]

The use of an insoluble Edman reagent, based on polystyrene and prepared according to Scheme 2, has been investigated.[86] Reaction of some



Scheme 1



Scheme 2

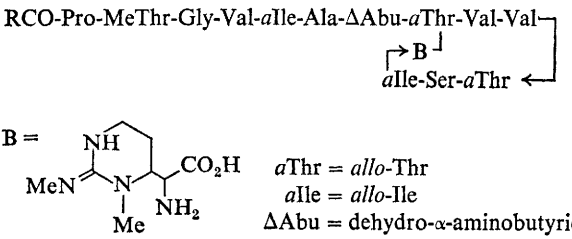
60% of the functional groups introduced with glucosaminol was required to increase the hydrophilic properties of the material sufficiently, after which several steps of peptide degradation were achieved. However, the yields were somewhat low and large amounts of peptide were required by

[85] M. A. Ondetti, A. Deer, J. T. Sheehan, J. Pluščec, and O. Kocy, *Biochemistry*, 1968, **7**, 4069.

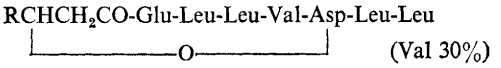
[86] L. M. Dowling and G. R. Stark, *Biochemistry*, 1969, **8**, 4728.

comparison with the dansyl-Edman procedure. The method, though interesting, would appear unlikely to be widely used.

Amino-acid sequence determination by high-resolution mass spectrometry, on the other hand, is developing well and has been the subject of numerous reviews.[87–90] Attempts are even being made with nucleotide sequences.[91] Typical structures determined are those of the peptide antibiotics stendomycin (1)[92] and esperin (2).[93] The fatty acid component



(1)



$R = C_{12}H_{25}(45\%),\ C_{10}H_{23}(35\%),\ C_{10}H_{21}(20\%)$

(2)

in stendomycin is isomyristic acid and there is a 30% replacement of *allo*-isoleucine at position 13 with valine. In such peptides, sequence-determining ions up to 10 residues or so can be clearly distinguished. The traditional difficulties with arginine-containing peptides have previously been circumvented by converting the arginine to ornithine or by blocking the guanidino-group with 1,3-diketones. The successful use of 1,2-cyclohexanedione for the same purpose has now been described[94] and the desulphurisation of sulphur-containing peptides before mass spectrometry has been documented.[95]

[87] M. M. Shemyakin, *Pure Appl. Chem.*, 1968, **17**, 313.
[88] E. Lederer, *Pure Appl. Chem.*, 1968, **17**, 489.
[89] G. VanLear and F. W. McLafferty, *Ann. Rev. Biochem.*, 1969, **38**, 289.
[90] M. M. Shemyakin, Yu. A. Ovchinnikov, E. I. Vinogradova, A. A. Kiryushkin, M. Yu. Feigina, N. A. Aldonova, Yu. B. Alakhov, V. M. Lipkin, A. I. Miroshnikov, B. V. Rosinov, and S. A. Kazaryan, *F.E.B.S. Letters*, 1970, **7**, 8.
[91] D. F. Hunt, C. E. Hignite, and K. Biemann, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 378.
[92] D. W. Thomas, E. Lederer, M. Bodanszky, J. Izdebski, and I. Muramatsu, *Nature*, 1968, **220**, 580.
[93] D. W. Thomas and T. Ito, *Tetrahedron*, 1969, **25**, 1985.
[94] J. Lenard and P. M. Gallop, *Analyt. Biochem.*, 1969, **29**, 203.
[95] A. A. Kiryushkin, V. A. Gorlenko, B. V. Rosinov, Yu. A. Ovchinnikov and M. M. Shemyakin, *Experientia*, 1969, **25**, 913.

It is encouraging to see that more conventional peptides are now having their sequences successfully analysed by mass spectrometry. Among those recently described are peptides from silk fibroin,[96] feline gastrin,[97] and a peptide from the λ-chain of pig immunoglobulin.[98] In the latter case, 10 residues of an 18-residue peptide were placed in sequence, which exemplifies how rapidly techniques for rendering large peptides sufficiently volatile for analysis are advancing. This is due in large part to permethylation of the peptide, and a new technique for permethylation with methyl iodide and sodium hydride in dimethylformamide[99] would seem most valuable. Certainly it has been used to good effect in sequence analysis of another immunoglobulin peptide in which the *N*-terminal pyrrolidonecarboxylic acid residue presented no problems and was readily identified.[100]

High-resolution mass spectrometry is clearly turning into a viable method for sequence determination in oligopeptides. Coupled with computer analysis of the spectra, it could obviously provide a powerful automatic sequence method. However, it must be borne in mind that the apparatus is not simple and is expensive and that not every laboratory, therefore, will be able to afford to set up the technique. Some thought might be given, perhaps, to several laboratories sharing an instrument and its facilities. Parenthetically, the same is true, though the financial problem less acute, with the sequenator. As a spur to further research, it would be nice to see many more attempts being made to couple the separation of suitably modified peptides on g.l.c. and their subsequent analysis in the mass spectrometer. In view of the high resolution available in modern mass spectrometry and the possibility of computer processing of the spectra, it may well be that the individual components in simple peptide mixtures could still be recognised and analysed. Since peptide fractionation is likely to remain the rate-limiting step in protein sequence determination, the resultant time-saving could be of immense value.

D. Peptide Chain Cleavage.—*Proteolytic Enzymes.* Numerous proteinases have been purified and examined during the past year *e.g.* those from *Serratia*,[101] from *Aspergillus*,[102] a carboxypeptidase from citrus fruit,[103] and aminopeptidases from mammalian connective tissue[104a] and bovine lens.[104b] Among the more interesting enzymes is a thermostable amino-peptidase from *B. stearothermophilus* which is stable even at 80 °C for

[96] H. R. Morris, A. J. Geddes, and G. N. Graham, *Biochem. J.*, 1969, **111**, 38P.
[97] K. L. Agarwal, G. W. Kenner, and R. C. Sheppard, *J. Amer. Chem. Soc.*, 1969, **91**, 3096.
[98] F. Franěk, B. Keil, D. W. Thomas, and E. Lederer, *F.E.B.S. Letters*, 1969, **2**, 309.
[99] J. Coggins and L. Benoiton, Abstracts of papers, BIOL. 18, 156th A.C.S. National Meeting, Atlantic City, 1968.
[100] G. H. de Haas, F. F. Franěk, B. Keil, D. W. Thomas, and E. Lederer, *F.E.B.S. Letters*, 1969, **4**, 25.
[101] A.-C. Rydén and B. v. Hofsten, *Acta Chem. Scand.*, 1968, **22**, 2803.
[102] M. Ito and M. Sugiura, *J. Pharm. Soc. Japan*, 1968, **88**, 1591.
[103] H. Zuber, *Z. physiol. Chem.*, 1968, **349**, 1337.
[104a] C. Schwabe, *Biochemistry*, 1969, **8**, 783; [b] U. Kettman, K. Kretschmer, and H. Hanson, *Z. physiol. Chem.*, 1968, **349**, 1537.

several hours.[105] The molecular basis of this remarkable stability remains obscure.

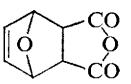
Reports continue to be made of the value of another enzyme from thermophilic bacteria, thermolysin, in structural studies. For example, much use has been made of it in the derivation of the amino-acid sequence of carboxypeptidase,[106] where quite large peptides were specifically obtained. The specificity is similar to that of the α-protease of *Crotalus atrox* snake venom, which cleaved preferentially linkages involving the amino-groups of valine, isoleucine, and leucine in melittin.[107] The proteolytic enzymes of *Streptomyces griseus* K-1 that together comprise the so-called 'Pronase' have also been characterised:[108] 11 enzymes were found in the mixture, including endo- and exo-peptidases, which underlines the average protein chemist's concern not to get any on his fingers! Improvements have been made to the purification of another remarkable enzyme, dipeptidyl amino-peptidase I, from rat liver or ox spleen.[109] The enzyme removes amino-acids two at a time from the *N*-terminus of proteins as dipeptides, exemplified by the reaction with the hormone glucagon.[109]

Limiting tryptic cleavage by modification of lysine or arginine residues in the intact polypeptide is a valuable method of securing large 'overlaps'. Thus, guanidination of the lysine residues has been used in continuing studies of the amino-acid sequence of horse heart myoglobin.[110] A full description of the reversible reaction of maleic anhydride with protein amino-groups has also been given.[111] However, a word of caution is necessary here. Though the reaction of maleic anhydride with a number of proteins appears to be specific for amino-groups,[111] in the case of tobacco mosaic virus protein at least, substantial ester formation with the hydroxy-groups of serine and threonine residues occurs under the conditions necessary to achieve full reaction of the protein amino-groups.[112] These esters, contrary to previous speculation, do not break down rapidly at alkaline pH, presumably because the unprotonated hydroxy-groups of serine and threonine, unlike that of tyrosine, are not good leaving groups. They may be cleaved specifically, however, by treatment of the reacted protein with 1M-hydroxylamine at pH 8·5, leaving the maleylamino-groups intact. Moreover, they are also cleaved at acid pH under the conditions used to regenerate the protein amino-groups. Somewhat similar results have been noted with the exhaustive maleylation of antibodies,[113] and the same considerations doubtless apply to other reversible anhydride reactions

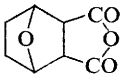
[105] G. Roncari and H. Zuber, *Internat. J. Protein Res.*, 1969, **1**, 45.
[106] R. A. Bradshaw, *Biochemistry*, 1969, **8**, 3871.
[107] J. Jentsch, *Z. Naturforsch.*, 1969, **24b**, 415.
[108] Y. Narahashi, K. Shibuya, and M. Yanagita, *J. Biochem.* (*Japan*), 1968, **64**, 427.
[109] J. K. McDonald, P. X. Callahan, B. B. Zeitman, and S. Ellis, *J. Biol. Chem.*, 1969, **244**, 6199.
[110] K. Han, M. Dautrevaux, Y. Boulanger, and G. Biserte, *F.E.B.S. Letters*, 1969, **3**, 141.
[111] P. J. G. Butler, J. I. Harris, B. S. Hartley, and R. Leberman, *Biochem. J.*, 1969, **112**, 679.
[112] L. King and R. N. Perham, unpublished work.
[113] M. H. Freedman, A. L. Grossberg, and D. Pressman, *Biochemistry*, 1968, **7**, 1941.

with proteins.[2] It has also been pointed out [2] that unblocked sulphydryl groups might add irreversibly across the double bond of both maleic and 2-methylmaleic (citraconic) anhydride. A study of the reaction of aldolase with citraconic anhydride [114] has revealed that although the rate of alkylation of thiol groups is slow under the conditions used for acylation of protein amino-groups (possibly because of steric hindrance by the substituent methyl group at the double bond compared with maleic anhydride), irreversible loss of protein thiol groups does occur and is almost certainly the reason for the subsequent difficulties in regenerating enzyme activity.

In an effort to circumvent the difficulties caused by the presence of an activated olefinic bond in maleic anhydride and its derivatives, two other suitable anhydrides, *exo-cis*-3,6-endoxo-Δ^4-tetrahydrophthalic anhydride (3) and its fully reduced form (4), have been synthesised and their reaction



(3)



(4)

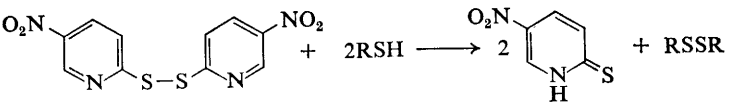
with model compounds and with lysozyme studied.[115] The half-life of hydrolysis of the amino-group derivative for each anhydride falls conveniently between that of the maleylamino- and citraconylamino-groups and, though the possibility of diastereoisomerism occurs in reaction with optically active compounds, with ε-amino-groups the asymmetric induction is negligible and the two products are kinetically identical in hydrolysis.

Methods of chemical cleavage of peptide chains have been the subject of an interesting review by Witkop.[116] The cyanogen bromide cleavage method continues to remain supreme, however, finding very wide use. A typical example is the cleavage of bacterial flagellin.[117] It has been reported that in studies of the enzyme catalase, difficulty was observed in obtaining good cleavage of a Met—Thr bond when the reaction was carried out in the conventional 70% formic acid.[118] For some reason, the reaction went very much better in 70% trifluoroacetic acid.

E. Fractionation Methods.—Trends in paper chromatography have been the subject of a recent review:[119] so have recent advances in gradient t.l.c.[120] The possibilities of direct spectrophotometry and fluorimetry of thin-layer chromatograms in amino-acid and peptide chemistry have been explored [121] and an apparatus has been devised for electrophoresis on

[114] I. Gibbons and R. N. Perham, *Biochem. J.*, 1970, **116**, 843.
[115] M. Riley and R. N. Perham, *Biochem. J.*, in the press.
[116] B. Witkop, *Science*, 1968, **162**, 318.
[117] C. R. Parish and G. L. Ada, *Biochem. J.*, 1969, **113**, 489.
[118] W. A. Schroeder, J. B. Shelton, and J. R. Shelton, *Arch. Biochem. Biophys.*, 1969, **130**, 551.
[119] V. C. Weaver, *Adv. Chromatog.*, 1968, **7**, 87.
[120] A. Niederwieser, *Chromatographia*, 1969, **2**, 23.
[121] G. Pataki, *Chromatographia*, 1968, **1**, 492.

polyamide layers.[122] In two-dimensional peptide mapping by electrophoresis and chromatography, electrophoresis on sheets of paper up to 150 cm long has been described [123] to increase separation. It may well be that such mapping would be better carried out on cellulose thin layers: it has been reported that higher resolution is obtainable.[124] Moreover, high resolution and very highly sensitive separations of Dns-peptides can be achieved on polyamide layer sheets.[124] It is also perhaps worth mentioning that a new selective reagent for thiols has been described, 2,2′-dithiobis-(5-nitropyridine), which could form the basis of a good spray test for thiol compounds, producing a yellow colour on a white background (Scheme 3).[125]



Scheme 3

Methods for the thin-layer gel filtration of proteins continue to be improved. Thus, standard densitometric equipment has been used to give quantitative evaluation of the filtration pattern,[126] in which the usual linear relationship was observed to apply between mobility and log (molecular weight), as for the standard column method.[127] An unusual 'tandem' column procedure for the purification of beef heart cytochrome *c* has also been described in which gel filtration succeeds ion-exchange chromatography in the same column.[128]

The advent of standardised loosely cross-linked gel filtration media has meant that many larger proteins are now amenable to purification by gel filtration. As a corollary, the molecular weights of larger proteins than hitherto can now be estimated by the method. Several recent papers have documented this.[129, 130] Further, since multimeric proteins are ordinarily broken down to their constituent subunits in the presence of strong solutions of guanidine hydrochloride, the inclusion of this salt in the gel filtration media should enable the molecular weights of subunits to be ascertained.[131] Several reports of such experiments have been published.[132a, b] It is possible that detergent could be substituted for the guanidine or urea.[132c]

[122] K. T. Wang and P. H. Wu, *Chemistry (Quart. Chinese Chem. Soc., Formosa)*, 1968, 85.
[123] G. Custer, *J. Chromatog.*, 1969, **42**, 429.
[124] C. Gerday, E. Robyns, and C. Gosselin-Rey, *J. Chromatog.*, 1968, **38**, 408.
[125] D. R. Grassetti and J. F. Murray, jun., *J. Chromatog.*, 1969, **41**, 121.
[126] B. J. Radola, *J. Chromatog.*, 1968, **38**, 61.
[127] P. Andrews, 'Protides of the Biological Fluids', vol. 14, Proceedings of the 14th Colloquium, Bruges, 1966, ed. H. Peeters, Elsevier, Amsterdam, 1967, p. 573.
[128] E. Soru and K. Rudescu, *J. Chromatog.*, 1969, **41**, 236.
[129] G. A. Locascio, H. A. Tigier, and A. M. del C. Battle, *J. Chromatog.*, 1969, **40**, 453.
[130] J. Marrink and M. Gruber, *F.E.B.S. Letters*, 1969, **2**, 242.
[131] P. F. Davison, *Science*, 1968, **161**, 906.
[132a] W. W. Fish, K. G. Mann, and C. Tanford, *J. Biol. Chem.*, 1969, **244**, 4989; [b] L.-P. Chao and E. R. Einstein, *J. Chromatog.*, 1969, **42**, 485; [c] M. Pagé and C. Godin, *Canad. J. Biochem.*, 1969, **47**, 401.

A neat method of selective purification of thiol peptides on columns has been described.[133] A water-insoluble organomercurial copolymer of ethylene and maleic acid is prepared and only peptides and proteins containing free thiol groups can be bound. These may subsequently be recovered by elution with mercaptoethanol. Along the same lines, but embodying the principle of reversible binding to a specific substrate or competitive inhibitor, is the method of enzyme purification described by Anfinsen and his colleagues.[134] The inhibitor or substrate is bound covalently to a cross-linked polymer or gel (the cross-linked dextran agarose is found to be convenient since attachment of the appropriate compound containing unprotonated amino-groups is readily achieved after 'activation' of the gel with cyanogen bromide) and after elution of the contaminating proteins, the desired enzyme is eluted by changes in ionic strength or pH or by adding substrate or competitive inhibitor to the eluting buffer. Applications of the technique to the purification of the avidin from egg-white (the substance attached to the polymer is ε-*N*-biotinyl-L-lysine)[135] and the purification of synthetic S-peptide derivatives of ribonuclease (the substance attached is S-protein)[136] have been described, in addition to the initial experiments with staphylococcal nuclease, chymotrypsin, and carboxypeptidase A.[134] Another approach to the same principle has been chemically to modify preformed polyacrylamide beads.[137] A wide variety of functional groups can be introduced, *e.g.* conversion of the stable amide groups to aminoethyl derivatives or hydrazides. Such experiments may well find additional application in the insolubilisation of enzymes.

More pieces of apparatus in which to carry out gel electrophoresis have been described[138] and a simple technique for applying two samples to the same polyacrylamide gel has been reported.[139] The samples are applied side by side on a paper disc placed on top of the column of gel and are prevented from diffusing across by a silicone water-repellent barrier on the disc. Several useful methods are now available for performing the isoelectric focussing separation of proteins usually carried out in columns[140] in supporting polyacrylamide gels,[141] which is useful for analytic purposes. Preparative electrophoresis of proteins in polyacrylamide gels, suitable for up to perhaps 1 g of protein, has been described in detail[142] for those

[133] I. E. Liener and L.-P. Chao, *Analyt. Biochem.*, 1968, **25**, 317.
[134] P. Cuatrecasas, M. Wilchek, and C. B. Anfinsen, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 636.
[135] P. Cuatrecasas and M. Wilchek, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 235.
[136] I. Kato and C. B. Anfinsen, *J. Biol. Chem.*, 1969, **244**, 5849.
[137] J. K. Inman and H. M. Dintzis, *Biochemistry*, 1969, **8**, 4074.
[138] M. De Mets, A. Lagasse, and M. Rabaey, *J. Chromatog.*, 1969, **43**, 145; D. P. Blattler, *Analyt. Biochem.*, 1969, **27**, 73.
[139] E. W. Johns, *J. Chromatog.*, 1969, **42**, 152.
[140] D. G. Brown, C. Baron, and A. Abrams, *Biochim. Biophys. Acta*, 1968, **168**, 386.
[141] D. H. Leaback and A. C. Rutter, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 447; Z. L. Awdeh, A. R. Williamson, and B. A. Askonas, *Nature*, 1968, **219**, 66.
[142a] S. Hjertén, S. Jerstedt, and A. Tiselius, *Analyt. Biochem.*, 1969, **27**, 108; [b] A. D. Brownstone, *ibid.*, 25; [c] W. S. Bont, J. Geels, and G. Rezelman, *ibid.*, 99.

who do not wish to buy the commercial apparatus. At the other end of the scale, experiments have been described on the DNA-dependent RNA polymerase from *E. coli* in which polyacrylamide gel electrophoresis has been carried out in 5 μl capillaries with as little as 0·4—0·5 μg of protein.[143] Such efforts merit close attention.

The most useful single advance recently has probably been the introduction of the sodium dodecyl sulphate (SDS)–gel electrophoresis system.[144] The presence of large amounts of SDS in the gels ensures that all polypeptide chains are strongly denatured and, presumably because the peptide chains are saturated with bound SDS, a linear relationship is found to hold between the electrophoretic mobility and the log (molecular weight) of the peptide chain, analogous to the gel filtration system referred to above. Thus, the subunit molecular weight of a given protein may be readily determined by calibration of the gels with proteins of known molecular weight.[144] The reliability and reproducibility of the results seem beyond question.[145, 146] Many protein chemists should now be able to bid the ultracentrifuge thankfully goodbye for this purpose.

3 Structural Proteins

The term 'structural proteins' is taken here to include the fibrous proteins, myosin, and certain globular proteins to which no substantial enzymic activity has yet been accorded. Immunoglobulins are dealt with separately in section 6.

A. The Crystallins.—Calf α-crystallin, a water-soluble lens protein, has been separated by ion-exchange chromatography into two acidic and two basic chains.[147] The *S*-carboxymethylated acidic chains are identical with respect to molecular weight and amino-acid composition but are distinguishable in peptide maps.[147, 148] The *N*-terminal sequence of both acidic chains is *N*-acetyl-Met-Asp- and the *C*-terminal residue is serine.[147] In independent experiments, it has been shown that both main kinds of chain in ox α-crystallin (A and B) have acetylated *N*-terminal residues in the following sequences:[149]

A. *N*-acetyl-Met-Asp-Ile-Ala-Ile-Gln-His-Pro-Trp-Phe-Lys-
B. *N*-acetyl-Met-Asp-Ile-Ala-Ile——His(Pro,Trp)Ile -Arg-

suggesting a common evolutionary precursor.

[143a] V. Neuhoff, W.-B. Schill, and H. Sternbach, *Z. physiol. Chem.*, 1968, **349**, 1126; [b] V. Neuhoff, W.-B. Schill, and H. Sternbach, *Z. physiol. Chem.*, 1969, **350**, 335.
[144] A. L. Shapiro and J. V. Maizel jun., *Analyt. Biochem.*, 1969, **29**, 505.
[145] K. Weber and M. Osborn, *J. Biol. Chem.*, 1969, **244**, 4406.
[146] A. K. Dunker and R. R. Rueckert, *J. Biol. Chem.*, 1969, **244**, 5074.
[147] J. G. G. Schoenmakers, J. J. T. Gerding, and H. Bloemendal, *European J. Biochem.*, 1969, **11**, 472.
[148] J. G. G. Schoenmakers, R. Matze, M. van Poppel, and H. Bloemendal, *Internat. J. Protein Res.*, 1969, **1**, 19.
[149] P. H. Corran and S. G. Waley, *Biochem. J.*, 1969, **115**, 789.

Rabbit lens γ-crystallin, on the other hand, has been suggested to comprise two different chains, each of molecular weight 20,000, one having *C*-terminal tyrosine and the other *C*-terminal valine.[150]

B. Muscle Proteins.—The role of the muscle proteins in contraction has been excellently reviewed by Huxley.[151] Physical studies on the molecular weight of myosin [152] show that it comprises two heavy chains each of molecular weight approximately 212,000 and two, or perhaps three, light chains of molecular weight approximately 20,000. Of these light chains, two appear to be bound specifically and a third perhaps non-specifically. These results are in good agreement with the analysis of the fragments of myosin produced by controlled papain digestion [153] in which the major portion of the myosin molecule is shown to consist of two subunits.

Proteins associated with motility in more commonplace systems often turn up somewhat unexpectedly in other situations.[154] Thus, slime mould actin, myosin, and actomyosin have been extracted from the plasmodium of *Physarum polycephalum* and it has been shown that slime mould myosin will form actomyosin with rabbit actin,[155] suggesting some structural similarity. The plasmodium exhibits very large protoplasmic streaming but a direct connexion with the actomyosin remains to be demonstrated. Actin has also been isolated from amoeba and shown to be very like rabbit skeletal muscle actin, even to the extent of containing one mole of 3-methylhistidine per mole of protein.[156] This sort of study is very interesting and is revealing a remarkable degree of what might be termed protein parsimony in Nature.

C. Collagen.—*Primary Structure.* Cyanogen bromide cleavage continues to provide the major method of attack on the α-chains of collagen and much of the work has recently been reviewed.[157] Thus, ten fragments accounting for the whole of the α1-chain [158] and six fragments accounting for the whole of the α2-chain [159] have been isolated from cyanogen bromide digests of chick skin collagen α-chains. The peptides from the α1-chain of chick skin collagen are very similar to peptides from the α1-chain of rat skin collagen and identical in all respects examined to the peptides from the α1-chain of chick bone collagen, except for the degree of hydroxylation of the lysine residues.[158, 160] Presumably, therefore, the α1-chains of

[150] A. S. Manalaysay and M. C. Hines, *Biochim. Biophys. Acta*, 1968, **168**, 383.
[151] H. E. Huxley, *Science*, 1969, **164**, 1356.
[152] L. C. Gershman, A. Stracher, and P. Dreizen, *J. Biol. Chem.*, 1969, **244**, 2726.
[153] S. Lowey, H. S. Slayter, A. G. Weeds, and H. Baker, *J. Mol. Biol.*, 1969, **42**, 1.
[154] I. Gibbons, *Ann. Rev. Biochem.*, 1968, **37**, 521.
[155] M. R. Adelman and E. W. Taylor, *Biochemistry*, 1969, **8**, 4964, 4976.
[156] R. R. Weihung and E. D. Korn, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 906.
[157] K. A. Piez, H. A. Bladen, J. M. Lane, G. J. Miller, P. Bornstein, W. T. Butler, and A. H. Kang, *Brookhaven Symp. Biol.*, 1968, **21**, 345.
[158] A. H. Kang, K. A. Piez, and J. Gross, *Biochemistry*, 1969, **8**, 1506.
[159] A. H. Kang, S. Igarishi, and J. Gross, *Biochemistry*, 1969, **8**, 3200.
[160] E. J. Miller, J. M. Lane, and K. A. Piez, *Biochemistry*, 1969, **8**, 30.

chick skin and bone collagens are derived from identical structural genes. Similarly, cyanogen bromide fragments accounting for the whole of the α2-chain of rat skin collagen [161] and chick bone collagen [162] have been isolated and characterised and shown to be homologous. It is interesting that the aldehyde-precursor lysine residue of peptide α2-CB1 from chick bone collagen is partially hydroxylated, as was noted for the corresponding residue in the α1-chain of chick bone collagen.[163] On the other hand, the same residue in the α2-chain of chick skin collagen is not hydroxylated.[159] Similar investigations into the cyanogen bromide cleavage of the α1-chain of cod skin collagen are also in progress [164] and homologies with selected peptides of chick bone collagen demonstrated.

The amino-acid sequences of some of these cyanogen bromide peptides have been reported. The sequence of a 37-residue peptide, α1-CB5, which is the major hexose-containing peptide of the α1-chain of rat skin collagen, has been given as follows:[165]

Glc-Gal- Hyl-Gly-His-Arg-Gly-Phe-Ser-Gly-Leu-Asp-Gly-Ala-Lys-Gly-Asn-
1 10
Thr-Gly-Pro-Ala-Gly-Pro-Lys-Gly-Glu-Hyp-Gly-Ser-Hyp-Gly-Glx-Asx-
20 30
(Gly,Ala,Hyp)-Gly-Gln-Met-
35

The disaccharide of glucose and galactose is attached to the hydroxy-group of the *N*-terminal hydroxylysine residue in glycosidic linkage. The peptide represents residues 99—135 of the α1-chain. The *N*-terminal 55 residues of the α1-chain of chick skin collagen have now been established [166] and are listed below with the *N*-terminal sequence of the α2-chain [166] for comparison:

α1 Glp-Met-Ser-Tyr-Gly-Tyr-Asp-Glu-Lys-Ser-Ala-Gly-Val-Ala-Val-Pro-
1 10
Gly-Pro-Met-Gly-Pro-Ala-Gly-Pro-Arg-Gly-Leu-Hyp-Gly-Pro-Hyp-
20 30
Gly-Ala-Hyp-Gly-Pro-Gln-Gly-Phe-Gln-Gly-Pro-Hyp-Gly-Glu-Hyp-
40
Gly-Glu-Hyp-Gly-Ala-Ser-Gly-Pro-Met-
50

α2 Glp-Tyr-Asp-Pro-Ser-Lys-Ala-Ala-Asp-Phe-Gly-Pro-Gly-Pro-Met-
1 10

These sequences are closely homologous with those of the α1- and α2-chains of rat skin collagen previously published. Few differences are apparent in

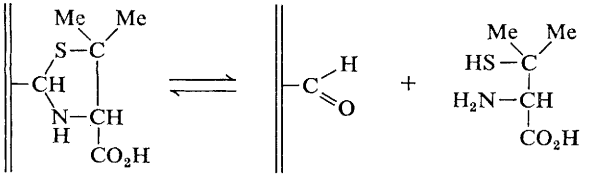
[161] P. P. Fietzek and K. A. Piez, *Biochemistry*, 1969, **8**, 2129.
[162] J. M. Lane and E. J. Miller, *Biochemistry*, 1969, **8**, 2134.
[163] E. J. Miller, J. M. Lane, and K. A. Piez, *Biochemistry*, 1969, **8**, 30.
[164] F. Laszlo and B. R. Olsen, *European J. Biochem.*, 1969, **11**, 140.
[165] W. T. Butler, *Biochemistry*, 1970, **9**, 44.
[166] A. H. Kang and J. Gross, *Biochemistry*, 1970, **9**, 796.

the α1-chains. Alanine (chick) replaces serine (rat) at positions 14 and 22 and the only other change is that the sequence Glp-Met-Ser-Tyr- is absent at the *N*-terminus of the rat skin α1-chain (although, strangely enough, it is present in rat tail tendon α1-chain).[167] It is also apparent that the α2-chains differ more in this region than do the α1-chains, which may be of some evolutionary significance. Finally, it should be mentioned that if acid extraction rather than salt extraction is used to isolate chick skin collagen, an α1′-chain type is found that lacks some 19 residues at the *N*-terminus of the normal α1-chain.[168] Either it is an artefact of the isolation procedure or the residues could be removed enzymically. If the latter is the case, since acid-extracted collagen is older than the more newly synthesized material extracted by salt, this may represent some stage in the maturation of the collagen.

Cross-links. Intramolecular cross-linking by means of an aldol condensation and subsequent dehydration between two lysine-derived δ-semialdehydes near the *N*-termini of adjacent chains[157] appears to have won wide acceptance. Some progress has also been made with the problem of the intermolecular bonds. There is evidence that intermediate Schiff base formation might occur. For example, the cross-links can be stabilised by reduction with sodium borohydride, giving a rise in stability of the collagen similar to that which occurs on ageing.[169] Schiff base formation by lysine and hydroxylysine side-chains of collagen with various aldehydes has been studied by a number of workers[170–172] and it would appear that there is a discrete family of reactive residues. A defect in inter- and intra-chain cross-links in collagen is caused by penicillamine.[172] Unlike lathyritic collagen, penicillamine collagen has an aldehyde content greater than normal and rapidly forms stable cross-links *in vitro*. A possible mechanism for reversible formation of the penicillamine complex is suggested (Scheme 4).[173] An effect of copper deficiency in diminishing cross-linking has also been reported and attributed to the catalytic role normally played by copper in the formation of an intermediate aldehyde compound.[174]

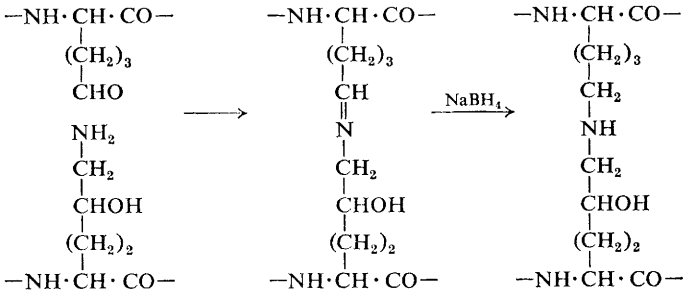
To establish the chemical nature of the intermolecular cross-link in detail is clearly more difficult. However, if young, native collagen is reduced with tritium-labelled sodium borohydride and the hydrolysates compared with those of collagen reduced after cleavage of the intermolecular bonds,[175] the label is found partly in hydroxylysino-norleucine.[176] A

[167] P. Bornstein, *Biochemistry*, 1969, **8**, 63.
[168] A. H. Kang, K. A. Piez, and J. Gross, *Biochemistry*, 1969, **8**, 3648.
[169] G. A. Balian, J. H. Bowes, and C. W. Cater, *Biochim. Biophys. Acta*, 1969, **181**, 331.
[170] R. C. Page and E. P. Benditt, *Science*, 1969, **163**, 578.
[171] J. H. Bowes and C. W. Cater, *Biochim. Biophys. Acta*, 1968, **168**, 341.
[172] R. C. Page, E. P. Benditt, and C. R. Kirkwood, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 752.
[173] K. Deshmukh and M. E. Nimni, *J. Biol. Chem.*, 1969, **244**, 1787.
[174] W. S. Chou, J. E. Savage, and B. J. O'Dell, *J. Biol. Chem.*, 1969, **244**, 5785.
[175] A. J. Bailey and D. Lister, *Nature*, 1968, **220**, 280.
[176] A. J. Bailey and C. M. Peach, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 812.



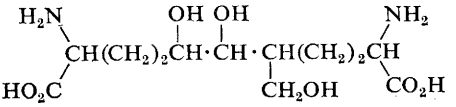
Complex Normal Collagen

Scheme 4

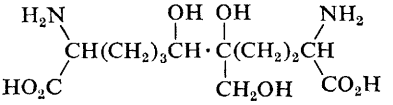


Scheme 5

possible mechanism is shown in Scheme 5. Since hydroxylysino-norleucine is only found after reduction, it probably exists as the labile Schiff base in native collagen. Moreover, since the lysine-derived aldehyde appears to reside only in the *N*-terminal region of the α-chains,[157] it is likely that Schiff base formation occurs between an aldehyde near the *N*-terminus of one molecule and some ε-amino-group of another molecule alongside in the collagen fibril, in accord with the staggered alignment of molecules in the native fibril. Another, more complicated, cross-link in bone and dentine collagen has also been established[177a] and two of the possible isomers are shown, (5) and (6). They are also observed in hydrolysates of



(5)



(6)

[177a] A. J. Bailey, L. J. Fowler, and C. M. Peach, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 663; [b] M. A. Paz, R. W. Lent, B. Faris, C. Franzblau, O. O. Blumenfeld, and P. M. Gallop, *ibid.*, 1969, **34**, 221.

the collagen after reduction with sodium borohydride and would derive from reduction of aldol condensation products of the δ-semialdehydes of lysine and hydroxylysine. The location of this cross-link is unspecified [177a] but could presumably occur as an intramolecular link between two adjacent chains. Similar experiments have been reported from another laboratory.[177b]

The presence of any cross-links involving peptide bond formation through the γ-carboxy-groups of glutamic acid residues has been ruled out by analyses of enzymic digests of collagen.[178] Treatment of denatured collagen with hydroxylamine leads to its degradation [179] and it has been suggested that the collagen chains consist of subunits linked together by (ester-like) bonds sensitive to hydroxylamine. However, this would seem to be at variance with all the work on collagen primary structure and it has now been established that one of the hydroxylamine-sensitive linkages in the α1-chain of rat skin collagen is a specific Asn—Gly bond.[180] The reasons for this peculiar susceptibility remain obscure. Serine phosphate peptides have been isolated from digests of bovine enamel, with the sequences Glu-Ser(P)-Leu and Glu-Ser(P)-Tyr.[181] Phosphorylation plays some role in initiating calcification.

D. Elastin.—Work on the primary structure of elastin has been bedevilled by the inability to recognise and isolate the soluble precursor subunits from which it is assumed to derive. However, it has recently been reported that in animals raised on a copper-free diet, the soluble protein fraction of elastic tissues may contain substantial amounts of elastin in a non-cross-linked form.[182] The amino-acid compositions of the soluble and insoluble elastin from pigs have now been shown to be essentially identical, with the exception of a lower lysine content in the cross-linked protein.[183] This difference has been ascribed to the formation of cross-links derived from the δ-semialdehydes of lysine residues, *e.g.* desmosine, isodesmosine, lysino-norleucine, and others. One of these could be an aldol condensation product comparable to that of collagen.[184a] Conclusive structural evidence for the presence of lysino-norleucine in hydrolysates of elastin has been provided.[184b] It is presumably biosynthesised *via* the Schiff base intermediate formed between α-amino-adipic acid δ-semialdehyde and a lysine residue (*cf.* Scheme 5 for collagen). Partial characterisation of the soluble protein has been achieved, for which the name tropoelastin is suggested.[185] The molecule, of molecular weight 67,000, contains no

[178] H. B. Bensusan, *Biochemistry*, 1969, **8**, 4716.
[179] D. Volpin, H. Hörmann, and K. Kühn, *Biochim. Biophys. Acta*, 1968, **168**, 389.
[180] W. T. Butler, *J. Biol. Chem.*, 1969, **244**, 3415.
[181] J. Seyer and M. J. Glimcher, *Biochim. Biophys. Acta*, 1969, **181**, 410.
[182] D. W. Smith, N. Weissman, and W. H. Carnes, *Biochem. Biophys. Res. Comm.*, 1968, **31**, 309.
[183] J. A. Petruska and L. B. Sandberg, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 222.
[184a] R. W. Lent, B. Smith, L. L. Salcedo, B. Faris, and C. Franzblau, *Biochemistry*, 1969, **8**, 2837; [b] C. Franzblau, B. Faris, and R. Papaioannou, *ibid.*, 2833.
[185] L. B. Sandberg, N. Weissman, and D. W. Smith, *Biochemistry*, 1969, **8**, 2940.

histidine, no methionine, and no isodesmosine or desmosine, and has the *N*-terminal sequence Gly-Gly-Val-Ile-Gly-. Now that such a soluble protein is available, we may look forward to rapid strides being made in its chemistry. The role of copper and lysine δ-semialdehydes in the cross-linking of elastin is closely comparable to that of collagen and strengthens the idea that these structural proteins might have had a common evolutionary precursor.

E. Chromosomal Proteins.—These have been the subject of a recent review,[186] and the stepwise removal of histones from chromosomes by acid treatment has been investigated.[187] The results of the latter analysis are intermediate between the two extreme models; that all histones are alike in a given region of the chromosome or, alternatively, that all histones are different in a given region. Further developments in the purification of histones from calf thymus and posterior silk glands by countercurrent distribution have been reported,[188] and the fractionation of chicken erythrocyte histones on carboxymethylcellulose has been described.[189] Depending on the method of extraction, a pure histone peculiar to the erythrocyte can be isolated in a pure form. The *N*-terminal residue is threonine and the *C*-terminal residue is lysine.[189] The large-scale purification of the avian-specific histone from erythrocytes has also been described[190] and the presence of 3-methylhistidine in avian erythrocyte histones reported.[191] The presence of various *N*-methylated and *N*-acetylated lysine residues in histones is, of course, well known. Other experiments have been concerned with the phosphorylation of serine residues in histones. Thus, specific phosphokinases are known to phosphorylate serine residues and a specific phosphatase for phosphorylated histones and protamine has been described.[192] These enzymes can give control of the degree of phosphorylation of the histones, which may play some role in altering the functional state of the molecules. A serine phosphate residue has been shown to exist at the *N*-terminus of trout testis histone IV in the sequence, *N*-acetyl-Ser(P)-Gly-Arg-.[193]

A curious fact about the histones, that must be borne in mind when considering their role in chromatin structure and gene action, is their limited heterogeneity. Another report on this for calf histones has been published.[194] The five main fractions of calf thymus histones have been

[186] R. H. Stellwagen and R. D. Cole, *Ann. Rev. Biochem.*, 1969, **38**, 951.
[187] K. Murray, *J. Mol. Biol.*, 1969, **39**, 125.
[188] K. Yokotsuka, M. Yoshida, and K. Shimura, *J. Biochem. (Japan)*, 1968, **64**, 129.
[189] M. Champagne, A. Mazen, and X. Wilhelm, *Bull. Soc. Chim. biol.*, 1968, **50**, 1261.
[190] E. W. Johns and J. H. Diggle, *European J. Biochem.*, 1969, **11**, 495.
[191] E. L. Gershey, G. W. Haslett, G. Vidali, and V. G. Allfrey, *J. Biol. Chem.*, 1969, **244**, 4871.
[192] M. H. Meisler and T. A. Langan, *J. Biol. Chem.*, 1969, **244**, 4961.
[193] G. W. Dixon, K. Marushige, V. Ling, M. Sung, and D. T. Wigle, *Fed. Proc.*, 1969, **28**, 599.
[194] S. Panyim and R. Chalkley, *Biochemistry*, 1969, **8**, 3972.

shown to have the following *C*-terminal sequences:[195]

F1	-Lys-Ala-Lys
F2a2	-Val-Gly-Ala-Arg-His-Lys
F3	-Arg-Ala
F2b	-Lys
F2a1	-Gly

The evolutionary conservation of histone sequence is indicated by the fact that the rat thymus histones have the same *C*-terminals.[195] The remarkable conservation of sequence between calf thymus histone IV and pea seedling histone IV points this up. The sequence of this glycine-rich arginine-rich histone from calf thymus has been determined in two independent studies.[196, 197] These agree exactly except that the arginine residue previously placed at position 44 [196] is now put at position 40.[197b, 198] The sequence of the corresponding histone from pea seedlings is identical for the whole 102 residues with three exceptions:[198]

Position	20	60	77
Calf thymus	MeLys	Val	Lys
Pea seedling	Lys	Ile	Arg

In addition, in the calf thymus histone, Lys-16 is approximately 50% acetylated and is the only residue so modified, whereas in the pea histone residues 5, 8, 12, and 16 can be acetylated, with a total acetylation of only 6%. None the less, these changes between two such phylogenetically distinct sources are very few indeed and such strict conservation of sequence must have some biological meaning in terms of histone function. Further experiments have also indicated that the corresponding glycine-rich, arginine-rich histone from bovine lymphosarcoma, Novikoff hepatoma, and foetal calf thymus share the same *C*-terminal amino-acid sequence as the pea and calf thymus histone IV,[199] to emphasise the remarkable homology.

A glance at the sequence of histone IV [1, 2] indicates that there are considerably more basic amino-acids in the *N*-terminal half than the *C*-terminal half of the molecule, which may be significant in the interaction with nucleic acid. The reverse asymmetry has been detected in the lysine-rich histone of rabbit thymus.[200] Cleavage at the single tyrosine residue with *N*-bromosuccinimide reveals that the *C*-terminal fragment contains much more lysine than the *N*-terminal fragment.[201] The same is probably true of calf thymus lysine-rich histones.

[195] D. M. P. Phillips and P. Simson, *Biochim. Biophys. Acta*, 1969, **181**, 154.
[196] R. J. DeLange, D. M. Fambrough, E. L. Smith, and J. Bonner, *J. Biol. Chem.*, 1969, **244**, 319.
[197a] G. Quagliarotti, Y. Ogawa, C. W. Taylor, P. Sautiere, J. Jordan, W. C. Starbuck, and H. Busch, *J. Biol. Chem.*, 1969, **244**, 1796; [b] Y. Ogawa, G. Quagliarotti, J. Jordan, C. W. Taylor, W. C. Starbuck, and H. Busch, *ibid.*, 4387.
[198] R. J. DeLange, D. M. Fambrough, E. L. Smith, and J. Bonner, *J. Biol. Chem.*, 1969, **244**, 5669.
[199] L. Desai, Y. Ogawa, C. M. Mauritzen, C. W. Taylor, and W. C. Starbuck, *Biochim. Biophys. Acta*, 1969, **181**, 146.
[200] M. Bustin, S. C. Rall, R. H. Stellwagen, and R. D. Cole, *Science*, 1969, **163**, 391.
[201] M. Bustin and R. D. Cole, *J. Biol. Chem.*, 1969, **244**, 5291.

The amino-acid sequences of the protamines salmine, from Hokkaido salmon, and iridine, from rainbow trout, have also been reported.[202]

F. Miscellaneous.—*Serum and Egg Proteins.* Two polypeptide components of the high density lipoproteins of human sera have been isolated and partially characterised.[203] One has a molecular weight of 14,900 and the *C*-terminal residue is glutamine. The other, with *C*-terminal threonine, is reported to exhibit the astonishing property of dimerisation in solutions of guanidine hydrochloride.

Ovalbumin is converted to plakalbumin by limited proteolysis with subtilisin. The peptide generated, which remains attached to the remainder of the protein by non-covalent interactions, has now been examined in more detail.[204] It has a molecular weight of 3800 and the *N*-terminal sequence is Ser-Val-Ser-Glu-Glu-Phe-Arg-Ala-Asp-. The *C*-terminal residue is proline, the same as intact ovalbumin, and the peptide would appear to derive from the *C*-terminal end of the albumin molecule. More information has become available about the ovotransferrin molecule. Transferrin binds two iron atoms per molecule of protein (molecular weight 80,000) with equal binding constants. Amino-acid analyses indicate a total of 31 half-cystine residues per mole of protein and 34 unique sequences around cysteic acid residues have now been established for oxidised transferrin, showing that the molecule is not a dimer of identical halves.[205] Some ambiguity of sequence was found in some peptides, for which several explanations are possible. The hens used might not be genetically homogeneous or they might be heterozygous for the transferrin gene. A third explanation, that of ambiguity during protein biosynthesis, seems unlikely (see section 5C).

The amino-acid sequences of the *N*-terminal 24 residues of the serum albumin from rat, human, and ox have been compared [206] and are listed in Figure 1. The homologies are considerable and the replacements generally conservative.

Ox	Asp-Thr-**His-Lys-Ser-Glu**-Ile	**-Ala-His-Arg-Phe-Lys-Asp-Leu-Gly-**
Rat	Glu-Ala-**His-Lys-Ser-Glu**-Ile	**-Ala-His-Arg-Phe-Lys-Asp-Leu-Gly-**
Human	Asp-Ala-**His-Lys-Ser-Glu**-Val	**-Ala-His-Arg-Phe-Lys-Asp-Leu-Gly-**
	1	**10**
Ox	**Glu**-Glu-His	**-Phe-Lys-Gly-Leu-Val-Leu-**
Rat	**Glu**-Gln-His	**-Phe-Lys-Gly-Leu-Val-Leu-**
Human	**Glu**-Glu-Asn	**-Phe-Lys-Ala-Leu-Val-Leu-**
	20	

Figure 1 *Amino-acid sequences of the* N*-terminal* 24 *residues of serum albumin from ox, rat, and human.*[206]

[202] T. Ando and S. Watanabe, *Internat. J. Protein Res.*, 1969, **1**, 221.
[203] V. Shore and B. Shore, *Biochemistry*, 1968, **7**, 3396.
[204] R. Sleigh, R. Hosken, M. B. Smith, and E. O. P. Thompson, *Austral. J. Biol. Sci.*, 1969, **22**, 239.
[205] T. C. Elleman and J. Williams, *Biochem. J.*, 1970, **116**, 515.
[206] R. A. Bradshaw and T. Peters jun., *J. Biol. Chem.*, 1969, **244**, 5582.

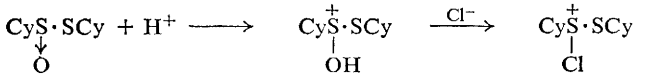
A partial characterisation of the γ-chain of human fibrinogen has been carried out on the *S*-sulpho-derivative.[207] The *N*-terminal residue has been shown to be tyrosine and the *C*-terminal sequence to be (Ser, Ala, Gly)-Asp-Val. Other experiments have been directed at proving the existence of ε-(γ-glutamyl)lysine in fibrin polymerized by Factor XIII. Cross-linked fibrin is reacted with acrylonitrile and then hydrolysed. With all free ε-amino-groups cyanoethylation occurs and carboxyethyl-lysine is produced on acid hydrolysis. If the ε-amino-group is blocked by peptide bond formation in the cross-linked fibrin, free lysine results on hydrolysis. The supposed cross-link has been demonstrated in this fashion and also by examining enzymic digests.[208] Glutamate appears to be the sole acceptor.

Wool and Others. The amino-acid sequences of 34 peptides from the chymotryptic digest of urea-soluble fraction (U.S.3) of oxidised wool have been reported.[209] The *N*-terminal sequence of a fragment of wool component 8 obtained by cyanogen bromide cleavage has been shown to be:[210]

$$\begin{array}{c} \text{Tyr} \\ \text{N-acetyl-Ser-Phe-Asp-Phe-Cys(Cm)-Leu-Pro-Asp-Leu-Ser-Phe-Arg-} \\ \text{Pro} \qquad \text{Leu} \end{array}$$

that is, a basic sequence with variations. It has also been suggested by other workers[211] that the high-sulphur proteins of wool are families of proteins with common structural features. In addition to the acetylated *N*-terminus found in component 8 of the low-sulphur proteins of wool,[210] pyrrolidonecarboxylic acid has also been detected as an *N*-terminal residue.[212]

The unusual sequence -Cys-Cys- found in keratin has led to a search for the presence of cyclocystine -Cys-Cys- but it has not yet been found.[213] These experiments have, however, suggested an interesting explanation for the peculiar disulphide exchange reaction that takes place in cystine peptides in strong acid. This exchange, curiously, is inhibited by thiol compounds such as thioglycollic acid. It is suggested[213] that the mechanism might involve the formation of the *S*-monoxide and sulphonium ions (Scheme 6).



Scheme 6

[207] D. A. Mills and I. E. Liener, *Arch. Biochem. Biophys.*, 1969, **130**, 629.
[208] J. J. Pisano, J. S. Finlayson, and M. P. Peyton, *Biochemistry*, 1969, **8**, 871.
[209] M. C. Corfield and J. C. Fletcher, *Biochem. J.*, 1969, **115**, 323.
[210] I. J. O'Donnell, *Austral. J. Biol. Sci.*, 1969, **22**, 471.
[211] R. L. Darskus, J. M. Gillespie, and H. Lindley, *Austral. J. Biol. Sci.*, 1969, **22**, 1197.
[212] I. J. O'Donnell, *Austral. J. Biol. Sci.*, 1968, **21**, 1327.
[213] H. Lindley and T. Haylett, *Biochem. J.*, 1968, **108**, 701.

S-Monoxides enter very readily into exchange reactions and these are catalysed by halide ions. The inhibitory role of a thiol compound would then be as an antioxidant to prevent the initial formation of the *S*-monoxide or to break it down if already formed (Scheme 7). Thus, the thiol com-

$$\begin{matrix} R^1S\cdot SR^2 \\ \downarrow \\ O \end{matrix} \quad + \; 2R^3SH \longrightarrow \; R^1S\cdot SR^3 \; + \; R^2S\cdot SR^3 \; + \; H_2O$$

Scheme 7

pound would only delay the interchange rather than inhibit it entirely, as is found to be the case, and the mechanism would also explain why the exchange is favoured in HCl rather than H_2SO_4.

The amino-acid sequence of the protein from Holmes rib grass strain of tobacco mosaic virus has been determined and compared with other natural strains.[214] It is interesting in that it contains 156 rather than the usual 158 amino-acid residues and the sequence shows it to be the most distantly related of the strains *vulgare*, U2, and *dahlemense*. Tryptic peptides that account for the whole sequence of 162 residues in α-lactoglobulin A and B have also been isolated.[215]

4 Peptides and Hormones

A. Growth Hormone.—Ox growth hormone has two *N*-terminal sequences in approximately equal amounts. These have now been characterised and compared with the *N*-terminal sequence of the human hormone:[216]

Ox	Ala-**Phe-Pro**-Ala-Met-Ser-**Leu-**
	Phe-Pro-Ala-Met-Ser-**Leu-**
Human	**Phe-Pro**-Thr-Ile -Pro-**Leu-**

Why the ox hormone has two *N*-terminal sequences is not clear. The additional *N*-terminal residue could possibly be removed enzymically or, alternatively, the reason might be genetic (gene doubling or allelic forms).

The results of a study of the cyanogen bromide cleavage of ox growth hormone have also been published.[217]

B. Insulin.—(See also this chapter, part II, section 3B.) The isolation and properties, including tryptic activation, of bovine proinsulin have been reported.[218] Two forms were isolated. One consisted of the B-chain separated by 29 residues from the *C*-terminal A-chain: the other comprised two chains linked by disulphide bridges, containing the normal A- and B-chains together with 27 additional residues. A single fish insulin has been

[214] H. G. Wittmann, I. Hindennach, and B. Wittmann-Liebold, *Z. Naturforsch.*, 1969, **24b**, 877.
[215] G. Frank and G. Braunitzer, *Z. physiol. Chem.*, 1968, **349**, 1456.
[216] M. Wallis, *F.E.B.S. Letters*, 1969, **3**, 118.
[217] R. E. Fellows and A. D. Rogol, *J. Biol. Chem.*, 1969, **244**, 1567.
[218] D. D. Schmidt and A. Arens, *Z. physiol. Chem.*, 1968, **349**, 1157.

isolated from a bonito–tuna–swordfish insulin mixture and its complete sequence derived.[219] It is interesting in that the B-chain maintains its conventional length of 30 residues by deleting the usual *C*-terminal alanine residue and adding an additional residue, valine, at the *N*-terminus. This seems to be generally true of fish insulins, exemplified by that of cod,[220] where the additional *N*-terminal residue is methionine.

The isolation of insulin from the turkey (*Meleagris gallopavo*) has been described and the amino-acid composition shown to be identical to that of chicken. A sequence is awaited.[221]

C. Interstitial-cell Stimulating Hormone (ICSH).—Sheep ICSH is a glycoprotein of molecular weight 30,000 containing two subunits, C1 and C2. Both have proline at the *N*-terminus, whereas the *C*-terminal sequence of C1 is -Tyr-Tyr-Ser and that of C2 is -Cys-Ile-Leu.[222] More detailed study of the cyanogen bromide digest of subunit C2 shows that its *N*-terminal sequence is:[223]

Pro-Gln-Arg-Val-Cys-Thr-Tyr-His-Gln-Leu-Arg-Phe-Ala-Ser-Val-Arg-Leu-
1 **10**
Pro-Gly-Pro-Cys-Pro-Val-Asp-Pro-Gly-Met-Lys-Arg-Val-Leu-Pro-Val-Pro-
20 **30**
Pro-Leu-Ile-Pro-Met-
39

D. Other Hormones.—A simple procedure for isolating calcitonin from the pig thyroid gland in good yield has been described.[224] A new preparation has also been given for human somatotropin and the molecular weight estimated to be 22,400. The *N*-terminal sequence is Phe-Pro-Thr-Ile-Pro-Leu-Ser- and the *C*-terminal residue is phenylalanine.[225]

The *N*- and *C*-terminal sequences of sheep lactogenic hormone have been shown to be:[226]

C-terminal: -Leu-Asn-Cys-Arg-Ile-Tyr-Asn-Asn-Asn-Cys
N-terminal: Thr-Pro-Val-Cys-Pro-Asx-Gly-Pro-Gly-Asx-Cys-Glx-

The presence of small disulphide loops close to the *N*- and *C*-terminus of the molecule is an unusual and interesting feature.

[219] P. Neumann and R. E. Humbel, *Internat. J. Protein Res.*, 1969, **1**, 125.
[220] K. B. M. Reid, P. T. Grant, and A. Youngson, *Biochem. J.*, 1968, **110**, 289.
[221] G. Weitzel, W. Oertei, K. Rager, and W. Kemmler, *Z. physiol. Chem.*, 1969, **350**, 57.
[222] T. S. Anantha Samy, H. Papkoff, and C. H. Li, *Arch. Biochem. Biophys.*, 1969, **130**, 674.
[223] T. S. Anantha Samy, H. Papkoff, and C. H. Li, *Arch. Biochem. Biophys.*, 1969, **132**, 315.
[224] C. D. Arnand and H. S. Tsao, *Biochemistry*, 1969, **8**, 449.
[225] P. Fønss-Bech and K. D. Schmidt, *Internat. J. Protein Res.*, 1969, **1**, 85.
[226] C. H. Li, J. S. Dixon, K. D. Schmidt, Yu. A. Pankov, and T. Lo, *Nature*, 1969, **222**, 1268.

A third form of neurophysin from bovine pituitary posterior lobes has been reported.[227] The molecular weight is approximately 10,000 and the amino-acid composition is very similar to that of the other two bovine neurophysins previously described. All three have alanine as *N*-terminal residue and form complexes with the hormones oxytocin and vasopressin.

E. Peptides.—A good deal of work has been reported on the primary structures of toxins and other pharmacologically active peptides. Viscotoxin A3 from the European mistletoe *Viscum album*, L. has the sequence:[228]

Lys-Ser-Cys-Cys-Pro-Asn-Thr-Thr-Gly-Arg-Asn-Ile-Tyr-Asn-Ala-Cys-Arg-
1 **10**
Leu-Thr-Gly-Ala-Pro-Arg-Pro-Thr-Cys-Ala-Lys-Leu-Ser-Gly-Cys-Lys-Ile-
20 **30**
Ile-Ser-Gly-Ser-Thr-Cys-Pro-Ser-Tyr-Pro-Asp-Lys
40 **47**

The neurotoxin, Toxin α, from Egyptian cobra (*Naja haje haje*) venom has been shown to have the sequence:[229]

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

In common with many of the extracellular enzymes, *e.g.* lysozyme, ribonuclease, and the pancreatic serine proteinases, such toxins seem to have an abundance of disulphide bridges, presumably to increase their stability in unfavourable environments.

The encephalitogenic A1 protein from myelin capable of inducing allergic encephalomyelitis has been reported to have a molecular weight of 16,400 and to have the *N*-terminal sequence *N*-acetyl-Ala-Ser-Ala-Gln-Lys-. The *C*-terminal sequence is -(Lys, Ala, Ile)-Leu-Val-His-Phe-Met-Ala-Arg-Arg.[230]

Various peptides of bee venom have been studied. One of them has the sequence [231]

Ala-Gly-Pro-Ala-Gln-Histamine

which is of interest in that it contains the physiologically interesting histamine in peptide linkage. Other venom peptides that have been

[227] R. Rauch, M. D. Hollenberg, and D. B. Hope, *Biochem. J.*, 1969, **115**, 473.
[228] G. Samuelsson, L. Seger, and T. Olson, *Acta Chem. Scand.*, 1968, **22**, 2624.
[229] D. P. Botes and D. J. Strydom, *J. Biol. Chem.*, 1969, **244**, 4147.
[230] G. A. Hashim and E. H. Eylar, *Biochem. Biophys. Res. Comm.*, 1969, **34**, 770.
[231] D. A. Nelson and R. O'Connor, *Canad. J. Biochem.*, 1968, **46**, 1221.

investigated include a mast-cell degranulating peptide [232] and various bradykinin-potentiating peptides.[233]

5 Enzymes

(See also this chapter, part II, section 3B.) The amino-acid sequences, partial and complete, of many enzymes continue to be collected together most valuably [1] and a review of advances in the chemistry of enzyme active sites has been published.[234] Specific reviews of ribonucleases [235] and the mechanism of action of proteolytic enzymes [236] have appeared.

A. Proteolytic Enzymes.—The veritable torrent of information about proteolytic enzymes acquired by numerous techniques has prompted two timely conferences. The first, organised by the British Biophysical Society, has been succinctly reviewed.[237] The proceedings of the second, held by the Royal Society, constitute a seminal volume that will remain a source-book for some years to come.[238] One fact, among others, that is emerging very clearly is that the proteolytic enzymes are falling coherently into groups, or families, the members of which are distinct and yet have much in common in terms of structure and mechanism of action.

For example, a comparison of the pancreatic serine proteinases and thrombin [239] reveals that, although the 49-residue A-chain of thrombin shows no detectable homology with the other serine proteinases, the B-chain of about 250 residues shows distinct homology with the pancreatic enzymes, in particular with respect to some of the disulphide bridges. In fact, it appears to resemble trypsin more closely than the others (its specificity in fibrinogen is for an Arg—Gly bond) and shares the ion-pair charge-relay mechanism of activation reviewed in these pages last year.[2] There can be little doubt that its tertiary structure must be very close to that of the pancreatic proteinases. Since liver is the source of prothrombin, it is also of interest that liver and pancreas have a common embryological origin. In line with these observations is the fact that thrombin is inhibited by the classic trypsin inhibitor, α-*N*-tosyl-L-lysyl chloromethyl ketone, and that the inhibition is accompanied by a loss in histidine content.[240] Similarly, since the amino-acid sequence of pig elastase is closely similar to that of ox trypsin and chymotrypsin,[241] it comes as no surprise that the three-dimensional structures are also closely alike.[242]

[232] P. Haux, *Z. physiol. Chem.*, 1969, **350**, 536.
[233] H. Kato and T. Suzuki, *Experientia*, 1969, **25**, 694.
[234] B. L. Vallee and J. F. Riordan, *Ann. Rev. Biochem.*, 1969, **38**, 733.
[235] E. A. Barnard, *Ann. Rev. Biochem.*, 1969, **38**, 677.
[236] A. Williams, *Quart. Rev.*, 1969, **23**, 1.
[237] L. N. Johnson, *F.E.B.S. Letters*, 1969, **2**, 201.
[238] *Phil. Trans.*, 1970, **B257**, 63—266 (No. 813).
[239] B. S. Hartley, *Phil. Trans.*, 1970, **B257**, 77.
[240] T. M. Chulkova and V. N. Orekhovich, *Biokhimiya*, 1968, **33**, 1222.
[241] D. M. Shotton and B. S. Hartley, *Nature*, 1970, **225**, 802.
[242] D. M. Shotton and H. C. Watson, *Nature*, 1970, **225**, 811.

The alliance of protein chemistry and *X*-ray crystallography exemplified in these studies sometimes brings unexpected results. Thus, what almost constitutes a game of musical chairs has been played with the amide groups assigned to certain aspartic acid residues, with striking consequences for theories of the mechanism of enzyme action. Reinvestigation of the amino-acid sequence around residue 102 of chymotrypsin revealed that this residue is aspartic acid, not asparagine, which led in turn to the charge-relay theory of the activity of the enzyme [243] (and indeed of the analogous trypsin, elastase, and thrombin). Residue 189 in trypsin, formerly thought to be asparagine, is now known to be aspartic acid, nicely accounting for the specificity of trypsin by the change in charge in this region of the protein accommodating the formation of a salt bridge with the positively charged side-chain of trypsin substrates.[239] (The corresponding residue in chymotrypsin and elastase is serine, in the 'tosyl' hole.) Similarly, residues 64 and 175 in papain have been shown definitely to be asparagine,[244] which makes redundant previous speculation about simple electrostatic interactions involving these residues. All of which goes to show that considerable care should be exercised in assigning amide groups even to seemingly innocuous aspartic and glutamic acid residues.

By modification with acetic anhydride, the α-amino-group of the *N*-terminal isoleucine residue of trypsin has been identified as a group controlling the enzymic activity at alkaline pH, analogous to chymotrypsin.[245] The activation of ox trypsinogen has also been more closely investigated.[246] When trypsinogen reacts with glycinamide under mild conditions in the presence of a water-soluble carbodi-imide, approximately 2·5 carboxy-groups are modified. The modified zymogen may now be activated in the absence of Ca^{2+} ions and the activation peptide, Val-Asp-Asp-Asp-Asp-Lys is found to be a mixture of the mono- and di-glycinamide derivatives. Each of the four side-chain carboxy-groups appears to be randomly modified and it is suggested that the role of Ca^{2+} ions in activation may be to bind these carboxyls, enhancing the susceptibility of the Lys(6)—Ile(7) bond to cleavage by trypsin in the activation process. Trypsin has been isolated from the shrimp,[247] and a trypsin-like protease from the crayfish, *Astacus leptodactylus*, has been studied and shown to have the trypsin sequence -Asp-Ser-Gly- around the functional serine residue.[248] Similarly, a chymotrypsin-like protease from the sea anemone *Metridium senile* has been found to contain amino-acid sequences around the critical serine and histidine residues identical to those of mammalian chymotrypsin.[249]

[243] D. M. Blow, J. J. Birktoft, and B. S. Hartley, *Nature*, 1969, **221**, 337.
[244] S. S. Husain and G. Lowe, *Biochem. J.*, 1970, **116**, 689.
[245] J. Chevallier, J. Yon, and J. Labouesse, *Biochim. Biophys. Acta*, 1969, **181**, 73.
[246] T. M. Radhakrishnan, K. A. Walsh, and H. Neurath, *Biochemistry*, 1969, **8**, 4020.
[247] B. J. Gates and J. Travis, *Biochemistry*, 1969, **8**, 4483.
[248] V. Tomasek, F. Šorm, R. Zwilling, and G. Pfleiderer, *F.E.B.S. Letters*, 1970, in the press.
[249] D. Gibson and G. H. Dixon, *Nature*, 1969, **222**, 753.

On the other hand, the subtilisins (proteinases from *Bacillus* that share the catalytic mechanism of the mammalian serine proteinases) show no homology in primary or tertiary structure with the mammalian enzymes,[238] which has been taken to mean that these two groups of enzymes have distinct genetic origins despite their functional similarity. None the less, there does exist in *B. sorangium* a serine proteinase that has the mammalian type of structure,[250] an interesting observation which serves to strengthen the idea perhaps that these two classes of serine proteinase have led separate evolutionary lives.

It is known that subunit II of ox procarboxypeptidase A resembles in many respects the chymotrypsinogens A and B from ox and pig pancreas. It also seems likely now, on the basis of amino-acid sequence analysis of the short A-chain produced in the activation step, that subunit II of procarboxypeptidase A and chymotrypsinogen C are one and the same.[251] The sequence evidence for this assumption is given below, where it is clear that ox subunit II appears to be closer to pig chymotrypsinogen C than to ox or pig chymotrypsinogens A or B:

Ox A	**Cys-Gly-Val-Pro-Ala-Ile -Gln-Pro-Val -Leu-Ser-Gly-Leu-Ser -Arg-**
Ox B	**Cys-Gly-Val-Pro-Ala-Ile -Gln-Pro-Val -Leu-Ser-Gly-Leu-Ala-Arg-**
Pig A	**Cys-Gly-Val-Pro-Ala-Ile -Pro-Pro-Val -Leu-Ser-Gly-Leu-Ser -Arg-**
Pig B	**Cys-Gly-Val-Pro-Ala-Ile -Pro-Pro-Val -Leu-Ser-Gly-Leu-Ser -Arg-**
Pig C	**Cys-Gly-Val-Pro-Ser -Phe-Pro-Pro-Asn-Leu-Ser————Ala-Arg-**
Ox Subunit II	**Cys-Gly-Ala-Pro-Ile -Phe-Gln-Pro-Asn-Leu-Ser————Ala-Arg-**

Other types of proteinase have continued to receive attention. A 105-residue fragment derived from the *N*-terminus and a 158-residue peptide containing the single histidine residue have been isolated from pig pepsin [252] and the structural homologies between pepsin and rennin have been confirmed and extended.[253] The structure of the carbohydrate moiety of stem bromelain has been reported [254] and the uncomfortable differences between the primary structure of papain interpreted from the *X*-ray-crystallographic analysis [255] and earlier chemical analyses have been satisfactorily resolved.[256]

The chromatographic purification of carboxypeptidase A(Cox) has been described [257] and allotypic forms of the enzyme have been characterised involving interchanges of Ile/Val at position 179, Ala/Glu at position 228

[250] L. B. Smillie and D. R. Whitaker, *J. Amer. Chem. Soc.*, 1967, **89**, 3350.
[251] R. J. Peanasky, D. Gratecos, J. Baratti, and M. Rovery, *Biochim. Biophys. Acta*, 1969, **181**, 82.
[252] V. A. Trufanov, V. Kostka, B. Keil, and F. Šorm, *European J. Biochem.*, 1969, **7**, 544.
[253] B. Foltmann, *Phil. Trans.*, 1970, **B257**, 147.
[254a] Y. Yasuda, N. Takahashi, and T. Murachi, *Biochemistry*, 1970, **9**, 25; [b] J. Scocca and Y. C. Lee, *J. Biol. Chem.*, 1969, **244**, 4852.
[255] J. Drenth, J. N. Jansonius, R. Koekoek, L. A. A. Sluyterman, and B. G. Wolthers, *Phil. Trans.*, 1970, **B257**, 231.
[256] S. S. Husain and G. Lowe, *Biochem. J.*, 1969, **114**, 279.
[257] P. H. Pétra and H. Neurath, *Biochemistry*, 1969, **8**, 5029.

and Val/Leu at position 305.[258] As usual, the replacements in 'internal' residues (179 and 305) are conservative, hydrophobic ones, whereas the only 'surface' change (228) can be more radical. The isolation of the cyanogen bromide cleavage fragments [259a] and the amino-acid sequence of some of them [259b] have been reported, leading to the derivation of the complete amino-acid sequence.[260] Considerable knowledge of the three-dimensional structure of the enzyme is, of course, already available,[261] and more detailed analysis of the correlations between structure and activity will now be possible.

Carboxypeptidase B may be inactivated by treatment with the substrate analogue, 4-bromo-acetamidobutylguanidine, with the alkylation of a specific tyrosine residue. The peptide containing the tyrosine residue affected has been isolated and compared with the corresponding peptide from carboxypeptidase A:[262]

A **Thr-Ile-Tyr-**Gln-**Ala-Ser-Gly-Gly-Ser-**Ile **-Asp-Trp**
B **Thr-Ile-Tyr-**Pro-**Ala-Ser-Gly-Gly-Ser-**Asp-**Asp-Trp**

The similarity of these sequences is good evidence that the two carboxypeptidases, not surprisingly in the light of our knowledge of other proteolytic enzymes, have derived from a common ancestral gene. A similar conclusion has been reached elsewhere.[263]

B. Other Enzymes.—The isolation and characterisation of human milk lysozyme has been reported [264] and goose egg-white lysozyme has been shown to contain only three tryptophan residues compared with the six found in the enzyme from chicken egg-white.[265] The goose enzyme is rather less stable than that from chicken, perhaps because of the lower cystine content. The fitting of the amino-acid sequence of the milk protein, α-lactalbumin, to the three-dimensional model of egg-white lysozyme has also been described.[266] Work such as this has far-ranging significance in that the sequences of these apparently unrelated proteins are homologous

[258] P. H. Pétra, R. A. Bradshaw, K. A. Walsh, and H. Neurath, *Biochemistry*, 1969, **8**, 2762.
[259a] M. Nomoto, N. G. Srinivasan, R. A. Bradshaw, R. D. Wade, and H. Neurath, *Biochemistry*, 1969, **8**, 2755; [b] R. A. Bradshaw, D. R. Babin, M. Nomoto, N. G. Srinivasan, L. H. Ericsson, K. A. Walsh, and H. Neurath, *ibid.*, 3859; R. A. Bradshaw, *ibid.*, 3871.
[260] R. A. Bradshaw, L. H. Ericsson, K. A. Walsh, and H. Neurath, *Proc. Nat. Acad. Sci., U.S.A.*, 1969, **63**, 1389.
[261] G. N. Reeke, J. A. Hartsuck, M. L. Ludwig, F. A. Quiocho, T. A. Steitz, and W. N. Lipscomb, *Proc. Nat. Acad. Sci., U.S.A.*, 1967, **58**, 2220.
[262] T. H. Plummer, *J. Biol. Chem.*, 1969, **244**, 5246.
[263] R. A. Bradshaw, H. Neurath, and K. A. Walsh, *Proc. Nat. Acad. Sci., U.S.A.*, 1969, **63**, 406.
[264] R. M. Parry, R. C. Chandan, and K. M. Shahani, *Arch. Biochem. Biophys.*, 1969, **130**, 59.
[265] A.-C. Dianoux and P. Jollès, *Helv. Chim. Acta*, 1969, **52**, 611.
[266] W. J. Browne, A. C. T. North, D. C. Phillips, K. Brew, T. C. Vanaman, and R. L. Hill, *J. Mol. Biol.*, 1969, **42**, 65.

and the lactalbumin sequence fits the lysozyme model with the need of no more than informed guesswork in a few places. If, as surely must be, detailed crystallographic analysis of α-lactalbumin confirms the proposed structure, lysozyme and α-lactalbumin will join the proteolytic enzymes and haemoglobin and myoglobin as examples of how similar amino-acid sequences define similar three-dimensional structures. Further weight will thereby be lent to the gathering opinion that the *X*-ray crystallographic analysis of one member of a family of proteins of related amino-acid sequence will suffice to build plausible models of them all.

On the other hand, fig and papaya lysozymes are single polypeptide chains with a molecular weight of approximately 25,000 which must differ substantially from the egg-white lysozymes.[267] The papaya enzyme has the *N*-terminal sequence Gly-Ile-Ser-Lys-Ile- and the *C*-terminal sequence -Ser-Phe-Gly.[268] It would be most interesting if it turned out that these enzymes are unrelated structurally to the egg-white enzymes and yet share a common bond-breaking mechanism, a situation found so far only in the pancreatic and microbial serine proteinases and the result, presumably, of convergent evolution. The complete amino-acid sequences of T2 and T4 bacteriophage lysozymes have also been described.[269] Both contain 164 residues, including two thiol groups and no disulphide bridges, and show no obvious sequence homology with the enzyme from egg-white. On the other hand, they are almost identical one with the other, only three residues differing:

T2 = Ser -40, Val-41, Ala-151
T4 = Asn-40, Ala-41, Thr-151

all of which are single-base changes in the *E. coli* codon dictionary. The lysozyme of bacteriophage λ has also been studied. It contains 159 residues and the cyanogen bromide peptides have been characterised.[270]

The *C*-terminal sequences of human carbonic anhydrases B and C have been compared and found to be closely homologous (Figure 2).[271, 272] They are curious in that the Arg—Pro bond (position 16—15) in the human C and bovine B enzymes is cleaved, albeit slowly, by trypsin whereas the corresponding bond in the human B enzyme is not split. The reason for this is obscure. It will be seen that the human C enzyme more closely resembles the bovine B enzyme than the human B enzyme in this region, a fact which is in line with the kinetic properties of the various enzymes. In

[267] A. N. Glazer, A. O. Barel, J. B. Howard, and D. M. Brown, *J. Biol. Chem.*, 1969, **244**, 3583.
[268] J. B. Howard and A. N. Glazer, *J. Biol. Chem.*, 1969, **244**, 1399.
[269] M. Inouye and A. Tsugita, *J. Mol. Biol.*, 1968, **37**, 213; A. Tsugita and M. Inouye, *ibid.*, 1968, **37**, 201.
[270] W. L. Black and D. S. Hogness, *J. Biol. Chem.*, 1969, **244**, 1982.
[271] P. O. Nyman, L. Strid, and G. Westermark, *European J. Biochem.*, 1968, **6**, 172.
[272] B. Andersson, P. O. Göthe, T. Nilsson, P. O. Nyman, and L. Strid, *European J. Biochem.*, 1968, **6**, 190.

another study,[273] the amino-acid sequence around a reactive histidine residue in human carbonic anhydrase B has been shown following carboxymethylation to be -Thr-His(Cm)-Pro-Pro-Leu-: the carboxymethylation is accompanied by substantial loss of enzymic activity.

Human B	-Ser-Leu-Leu-Ser-Asn-Val-Glu-Asp-Asn-Gly-Ala-Val-Pro-Met-Glx-	
Human C		-Val-
Bovine B		-Leu-
	30	**20**
Human B	His-**Asn**-Asn-**Arg**-**Pro**-Thr-**Gln**-**Pro**-**Leu**-**Lys**-Gly-**Arg**-Thr-Val-Arg-	
Human C	Asp-**Asn**-Trp-**Arg**-**Pro**-Ala-**Gln**-**Pro**-**Leu**-**Lys**-Asn-**Arg**-Gln-Ile -Lys-	
Bovine B	Ala-**Asn**-Trp-**Arg**-**Pro**-Ala-**Gln**-**Pro**-**Leu**-**Lys**-Asn-**Arg**-Gln-Val-Arg-	
	10	
Human B	Ala-Ser-Phe	
Human C	Ala-Ser-Phe-Lys	
Bovine B	Gly-Phe-Pro-Lys	
	1	

Figure 2 *C-Terminal amino-acid sequences of human and bovine carbonic anhydrases* [271, 272]

The amino-acid sequence of 36 residues at the *C*-terminus of pig heart lactate dehydrogenase has been shown to be:[274]

-Arg-Leu-Lys-Asp-Asp-Glu-Val-Ala-Gln-Leu-Lys-Gly-Leu-Thr-Ser-Asn-Val-Ile-Gln-Lys-Asn-Ser-Ala-Asp-Thr-Leu-Trp-Gly-Ile-Gln-Lys-Asp-Leu-Lys-Asp-Leu

This sequence is different from that of the muscle enzyme and shows that the enzymes are the products of different structural genes. Moreover, the four chains of the pig heart enzyme are identical, at least in the *C*-terminal region. In view of the progress with the *X*-ray crystallographic analysis of lactate dehydrogenase,[275] the complete amino-acid sequence of the protein is awaited with interest. Comparable work on the dimeric horse-liver alcohol dehydrogenase continues. Three main forms of the enzyme are known, resulting from the association of two possible types (E and S) of protein chain in the dimer. These chains, though very similar, differ as follows:[276]

E = Glu-17, Thr-130(?), Arg-130(?)+7
S = Gln-17, Ile -130(?), Ser -130(?)+7

Other differences, as yet undetected, may also exist. Studies on the amino-acid sequence of streptokinase (M 48,000) have revealed that the *N*-terminal

[273] S. L. Bradbury, *J. Biol. Chem.*, 1969, **244**, 2002.
[274] K. Mella, H.-J. Torff, E. T. J. Fölsche, and G. Pfleiderer, *Z. physiol. Chem.*, 1969, **350**, 28.
[275] M. J. Adams, D. J. Haas, B. A. Jeffery, A. McPherson jun., H. L. Mermall, M. G. Rossman, R. W. Schevitz, and A. J. Wonacott, *J. Mol. Biol.*, 1969, **41**, 159.
[276] H. Jörnvall, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 542.

sequence is Ile-Ala-Gly- and that the *C*-terminal residue is lysine:[277] investigation of the cyanogen bromide cleavage fragments is in hand. The tetrameric glycogen phosphorylases from rat and rabbit muscle have been shown to be highly homologous by examination of the amino-acid sequence around the serine phosphate residue in the active site.[278]

After some controversy, the amino-acid sequence around the serine phosphate residue in the active site of rabbit phosphoglucomutase has now been shown to be:[279]

Thr-Ala-Ser(P)-His-Asp-Pro-Gly-Gly-Pro-Asn-Gly-Asn-Phe-Gly-Ile-Lys-

Similarly, the amino-acid sequence around the catalytically essential cysteine residue of arginine kinase from lobster (*Homarus vulgaris*) muscle has been determined after alkylation with *N*-ethylmaleimide and compared with that found for the creatine kinase from rabbit muscle:[280]

Lobster arginine kinase -Gln-**Thr-Cys-Pro**-Thr-**Ser-Asn-Leu-Gly-Thr**——Val-**Arg**-
Rabbit creatine kinase -Leu-**Thr-Cys-Pro**——**Ser-Asn-Leu-Gly-Thr**-Gly-Leu-**Arg**-

Despite the postulated deletion and insertion, the sequence homology is marked, evidence perhaps of a common genetic origin once more. Work on a number of bacterial enzymes has been described. The *C*-terminal amino-acid of an amylase from *B. subtilis* has been reported to be lysine [281] and extensive sequence studies have been carried out on bacterial penicillinases.[282] For sheer length of sequence, however, pride of place must go this year to the investigation of the tetrameric bovine liver catalase.[283] The almost complete sequence of 505 residues in the subunit of molecular weight 57,500 has been established although the disulphide bridges remain to be determined. The haem prosthetic group is linked non-covalently to the protein.

There have been several reports of further structural studies on non-haem iron-containing proteins. The complete amino-acid sequences of ferredoxin from *Micrococcus aerogenes* (54 residues) [284] and from *Clostridium acidi-urici* (55 residues) [285] have been established and shown to resemble closely the ferredoxins previously isolated from *Clostridia*. The bacterial ferredoxins are of interest also in that the primary structure indicates evolution, by gene doubling, from a more primitive molecule of approximately half the length. Another group of ferredoxins with somewhat longer sequences is found in algae and higher plants: that from

[277] F. J. Morgan and A. Henschen, *Biochim. Biophys. Acta*, 1969, **181**, 93.
[278] C. L. Sevilla and E. H. Fischer, *Biochemistry*, 1969, **8**, 2161.
[279] S. Harshman, H. R. Six, and V. A. Najjar, *Biochemistry*, 1969, **8**, 3417.
[280] E. der Terrossian, L. A. Pradel, R. Kassab, and N. V. Thoai, *European J. Biochem.*, 1969, **11**, 482.
[281] H. Kojima and K.-I. Sugae, *J. Biochem.* (*Japan*), 1968, **64**, 713.
[282] R. P. Ambler and R. J. Meadway, *Nature*, 1969, **222**, 24.
[283] W. A. Schroeder, J. R. Shelton, J. B. Shelton, B. Robberson, and G. Apell, *Arch. Biochem. Biophys.*, 1969, **131**, 653.
[284] J. N. Tsunoda, K. T. Yasunobu, and H. R. Whiteley, *J. Biol. Chem.*, 1968, **243**, 6262.
[285] S. C. Rall, R. E. Bolinger, and R. D. Cole, *Biochemistry*, 1969, **8**, 2486; K. Sugeno and H. Matsubura, *J. Biol. Chem.*, 1969, **244**, 2979.

Scenedesmus (96 residues),[285] *Leucaena glauca* (96 residues),[286] and alfalfa (97 residues)[287] have recently been described. There is some evidence that the degree of amino-acid sequence homology between these two classes of ferredoxins is greater than chance would permit. Details of the derivation of the amino-acid sequence of the 113 residues of the subunit of hemerythrin from the sipunculid worm *Golfingia gouldii* have now been given.[288] The molecule is octameric and each subunit binds two atoms of iron. The steroid hydroxylase from the adrenal cortex has also been isolated (M 12,000) and shown to be very similar to ferredoxin in the nature of the linkage of the iron in the protein.[289] These non-haem iron proteins form an unusual class of metalloenzymes, the mechanism of which is proving of much interest.

C. Haemoglobin, Myoglobin, and Cytochromes.—Much of the recent work on the structure of haemoglobin and its bearing on the mechanism of action has been covered in a didactic review.[290] The particular role of the imidazole groups of the *C*-terminal histidine residues of the β-chains, together with the α-amino-groups of the α-chains, in the Bohr effect has also been described.[291, 292] It is greatly encouraging to see how answers to the problems of haemoglobin function are now becoming available in real molecular terms, since haemoglobin has served for many years as a paradigm of such investigation.

What appear to be the final nails in the coffin of ambiguous translation of the genetic code [2] have also been produced. Further evidence in favour of the alternative explanation, that of gene duplication, has been found in the α-chain of goat haemoglobin.[293] The fact that the α-chain variants of rabbit haemoglobin act as alleles[294] also points to these chains being the products of different structural genes rather than the result of ambiguous translation of a single gene.

Improved methods for the chromatography of human normal and abnormal haemoglobins on CM-Sephadex have been published,[295] as have improved techniques for the isolation of the γ-chain of human foetal haemoglobin.[296] New abnormal human haemoglobins continue to be

[286] A. M. Benson and K. T. Yasunobu, *J. Biol. Chem.*, 1969, **244**, 955.
[287] S. Keresztes-Nagy, F. Perini, and E. Margoliash, *J. Biol. Chem.*, 1969, **244**, 981.
[288] G. L. Klippenstein, J. W. Holleman, and I. M. Klotz, *Biochemistry*, 1968, **7**, 3868.
[289] T. Kimura, K. Suzuki, R. Padmanabhan, T. Samejima, O. Taruntani, and N. Ui, *Biochemistry*, 1969, **8**, 4027.
[290] M. F. Perutz, *Proc. Roy. Soc.*, 1969, **B173**, 113.
[291] M. F. Perutz, H. Muirhead, L. Mazzarella, R. A. Crowther, J. Greer, and J. V. Kilmartin, *Nature*, 1969, **222**, 1240.
[292] J. V. Kilmartin and L. Rossi-Bernardi, *Nature*, 1969, **222**, 1243.
[293] M. D. Garrick and T. H. J. Huisman, *Biochim. Biophys. Acta*, 1968, **168**, 585.
[294] T. Hunter and A. Munro, *Nature*, 1969, **223**, 1271.
[295] A. M. Dozy and T. H. J. Huisman, *J. Chromatog.*, 1969, **40**, 62.
[296] A. Kajita, K. Taniguchi, and R. Shukuya, *Biochim. Biophys. Acta*, 1969, **175**, 41; C. Ioppolo, E. Chiacone, E. Antonini, and J. Wyman, *Arch. Biochem. Biophys.*, 1969, **132**, 249.

described: among them are $I^{Toulose}$ (β66 Lys → Ser),[297] *Richmond* (β102 Asn → Lys),[298] *Tacoma* (β30 Arg → Ser),[299] and a new haemoglobin *Lepore*.[300] Details of the amino-acid sequence of the β-chain of rabbit haemoglobin have now been given[301] and two types of chain have been separated by counter-current distribution from frog (*Rana esculenta*) haemoglobin.[302] The one, of the α-chain series, has the *N*-terminal sequence *N*-acetyl-Ala-Leu- and the *C*-terminal sequence -Lys-Tyr; the other, of the β-chain series, has the *N*-terminal residue Gly- and the *C*-terminal sequence -Lys-Ala-Tyr-His. The kangaroo haemoglobin has also been investigated.[303] Two haemoglobins were found, I and II, differing only in the interchange of His(II) and Gln(I) in the β-chains at position 2. Subsequent work has established the complete amino-acid sequence of the β-chains.[304] The kangaroo β-chain appears to be very different from that of other mammals, suggesting early divergence of the marsupials in evolution. None the less, the residues thought to be important in function, in the haem-pockets and the contacts between subunits, are highly conserved.[304] Preliminary characterisation of the haemoglobin of the domestic fowl has also been carried out:[305] no acetyl groups were found, in conflict with earlier literature on the subject.

The preparation of pure myoglobin by gel chromatography has been described[306] and detailed studies of the amino-acid sequence of horse myoglobin continue.[307] The complete amino-acid sequence of the protein is now available.[1] So indeed are the complete sequences of myoglobin from the porpoise and the harbour seal.[308] All are very similar. In another study[309] the amino-acid sequence of human myoglobin has been examined. Out of 152 residues, 124 have been placed in unique sequence and the remainder have been placed by analogy with sperm whale myoglobin (153 residues). To maximise the homology, one deletion has to be postulated in the human protein at position 9, corresponding to a leucine residue in the sperm whale protein. The high degree of sequence homology between these proteins indicates very similar tertiary structures. This is confirmed for the human and sperm whale myoglobins by the similarity of

[297] R. Cabanes, D. Labie, J. Rosa, J. Ruffie, and R. Bierme, *Compt. rend.*, 1969, **269**, *D*, 111.
[298] G. D. Efremov, T. H. J. Huisman, L. L. Smith, J. B. Wilson, J. L. Kitchens, R. N. Wrightstone, and H. R. Adams, *J. Biol. Chem.*, 1969, **244**, 6105.
[299] B. Brimhall, R. T. Jones, E. W. Baur, and A. G. Motulsky, *Biochemistry*, 1969, **8**, 2125.
[300] C. Baglioni and V. Ventruto, *European J. Biochem.*, 1968, **5**, 29.
[301] J. S. Best, U. Flamm, and G. Braunitzer, *Z. physiol. Chem.*, 1969, **350**, 563.
[302] J. P. Chauvet and R. Acher, *F.E.B.S. Letters*, 1968, **1**, 305.
[303] E. O. P. Thompson, R. Hosken, and G. M. Air, *Austral. J. Biol. Sci.*, 1969, **22**, 449.
[304] G. M. Air and E. O. P. Thompson, *Austral. J. Biol. Sci.*, 1969, **22**, 1437.
[305] B. A. Moss and E. O. P. Thompson, *Austral. J. Biol. Sci.*, 1969, **22**, 1455.
[306] K. Bünnig and R. Hamm, *J. Chromatog.*, 1969, **43**, 450.
[307] Y. Boulanger, M. Dautrevaux, K. Han, and G. Biserte, *Bull. Soc. Chim. biol.*, 1968, **50**, 1651; K. Han, Y. Boulanger, M. Dautrevaux, and G. Biserte, *ibid.*, 1968, **51**, 439,
[308] R. A. Bradshaw and F. R. N. Gurd, *J. Biol. Chem.*, 1969, **244**, 2167.
[309] R. L. Hill, C. M. Harris, J. F. Naylor, and W. M. Sams, *J. Biol. Chem.*, 1969, **244**, 2182.

the alkylation of the methionine, lysine, and histidine side-chains in the two proteins.[310]

Details of the derivation of the complete amino-acid sequence of baker's yeast cytochrome *c*, first published some years ago, have now been given [311] and the sequence of the 110 residues in cytochrome b_{562} has been established.[312] Certain similarities with haemoglobin and myoglobin are apparent (Figure 3) but the significance of this is somewhat unclear as yet.

Cytochrome b_{562}	*Arg*-**His**-**Gly**-*Phe*-Asp-*Ile* -*Leu*-Val -*Gly*-Glu-
	65 **70**
Myoglobin	*Lys*-**His**-**Gly**-*Val*-Thr -*Val*-*Leu*-Thr-*Ala*-Leu-
Haemoglobin α-	Ala-**His**-**Gly**-Lys-Lys -Val-Ala-Asp-Gly-Leu-
Haemoglobin β-	Ala-**His**-**Gly**-Lys-Lys -Val-Leu-His -Ser -Phe-
Cytochrome b_{562}	Ile -Asp-Asp-Ala-Leu-*Lys*-
	75 **80**
Myoglobin	Gly-Ala -Ile -Leu-Lys-*Lys*-
Haemoglobin α-	Thr-Leu-Ala -Val -Gly-His-
Haemoglobin β-	Gly-Glu-Gly -Val -His -His-

Figure 3 *Amino-acid sequences of corresponding regions of cytochrome b_{562} from* E. coli, *sperm whale myoglobin, and the α- and β-chains of horse haemoglobin* [312]

Table *Quaternary structure of some individual enzymes*

Enzyme	*Source*	*Molecular weight*	*No. of subunits*
Lactic dehydrogenase[a]	Dogfish	144,000	4
Glyceraldehyde 3-phosphate dehydrogenase[b]	*B. stearothermophilus*	160,000	4
Glutamic dehydrogenase[c]	Pig liver, ox liver	310,000	6
Glucose 6-phosphate dehydrogenase[d]	Yeast	101,000	2
Malic dehydrogenase[e]	Ox heart	72,000	2
	Rat liver mitochondria	66,000	2
L-Glycerol 3-phosphate dehydrogenase[f, g]	Chicken breast muscle and chicken liver, rabbit muscle	60,000 or 77,000	2
	Honeybee	65,000	2
Aldolase[h]	Rabbit liver	158,000	4
	Spinach	120,000	4
	Mammalian muscle	160,000	4
	Yeast	80,000	2
Crotonase[i]	Ox	164,000	6
RNA polymerase[j]	*E. coli*	495,000 ($\alpha = 39{,}000$) ($\beta = 155{,}000$) ($\beta^1 = 165{,}000$) ($\sigma = 95{,}000$)	$\alpha_2\beta\beta^1\sigma$
Flagellin[k]	*Prot. vulgaris*	40,000	1
Arginine decarboxylase[l]	*E. coli*	820,000	10

[310] C. M. Harris and R. L. Hill, *J. Biol. Chem.*, 1969, **244**, 2195.
[311] K. Narita and K. Titani, *J. Biochem.* (*Japan*), 1969, **65**, 259.
[312] E. Itagaki and L. P. Hager, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 1013.

Table (*cont.*)

Enzyme	*Source*	*Molecular weight*	*No. of subunits*
UDP-glucose pyrophosphorylase[m]	Calf liver	480,000	8
Tryptophan oxygenase[n]	*Pseud. acidovorans*	123,000	4
L-Arabinose isomerase[o]	*E. coli*	360,000	6
Pyruvate kinase[p]	Rabbit muscle	230,000	4
UDP-galactose 4-epimerase[q]	*E. coli*	79,000	2
Caeruloplasmin[r]	Human	160,000	$8(\alpha_4\beta_4)$
Acetylcholinesterase[s]	*Electrophorus electricus*	260,000	$4(\alpha_2\beta_2)$
Enolase[t]	Yeast	67,000	2
Fructose 1,6-diphosphatase[u]	Rabbit liver	131,000 ($\alpha = 30{,}000$) ($\beta = 37{,}000$)	$4(\alpha_2\beta_2)$
Hexokinase[v]	Yeast	111,000	4
Ornithine aminotransferase[w]	Rat liver	132,000	4
Glutamine synthetase[x]	Sheep brain	525,000	8
Carboxylesterase[y]	Pig liver	160,000	2
Tryptophan synthetase B protein[z]	*E. coli*	90,000	2
Erythrogenic toxin[aa]	*Streptococcus pyogenes*	30,000	1
Lysyl-*t*RNA synthetase[bb]	*E. coli*	100,000	2
Pyruvate decarboxylase[cc]	*E. coli* K12	180,000	2
Luciferases[dd]	Bacteria	80,000	$2(\alpha, \beta)$
α-Amylase[ee]	Rabbit	54,000	1

[a] W. S. Allison, J. Admiraal, and N. O. Kaplan, *J. Biol. Chem.*, 1969, **244**, 4743. [b] R. Singleton, J. R. Kimmel, and R. E. Amelunxen, *J. Biol. Chem.*, 1969, **244**, 1623. [c] P. Dessen and D. Pantaloni, *European J. Biochem.*, 1969, **8**, 292. [d] R. H. Yue, E. A. Noltmann, and S. A. Kuby, *J. Biol. Chem.*, 1969, **244**, 1353. [e] K. G. Mann and C. S. Vestling, *Biochemistry*, 1969, **8**, 1105; C. Wolfenstein, S. Englard, and I. Listowsky, *J. Biol. Chem.*, 1969, **244**, 6415. [f] H. B. White tert., and N. O. Kaplan, *J. Biol. Chem.*, 1969, **244**, 6031; T. P. Fondy, C. R. Ross, and S. J. Sollohub, *J. Biol. Chem.*, 1969, **244**, 1631. [g] R. W. Brosemer and R. W. Kuhn, *Biochemistry*, 1969, **8**, 2095. [h] R. W. Gracy, A. G. Lacko, and B. L. Horecker, *J. Biol. Chem.*, 1969, **244**, 3913; G. Rapoport, L. Davis, and B. L. Horecker, *Arch. Biochem. Biophys.*, 1969, **132**, 286; C. E. Harris, R. D. Kobes, D. C. Teller, and W. J. Rutter, *Biochemistry*, 1969, **8**, 2442. [i] G. M. Hass and R. L. Hill, *J. Biol. Chem.*, 1969, **244**, 6080. [j] R. R. Burgess, *J. Biol. Chem.*, 1969, **244**, 6158; A. A. Travers and R. R. Burgess, *Nature*, 1969, **222**, 537. [k] J. Y. Chang, D. M. Brown, and A. N. Glazer, *J. Biol. Chem.*, 1969, **244**, 5196. [l] E. A. Bocker, E. H. Fischer, and E. E. Snell, *J. Biol. Chem.*, 1969, **244**, 5239. [m] S. Levine, T. A. Gillett, E. Hageman, and R. G. Hansen, *J. Biol. Chem.*, 1969, **244**, 5729. [n] W. N. Poillon, H. Macno, K. Koike, and P. Feigelson, *J. Biol. Chem.*, 1969, **244**, 3447. [o] J. W. Patrick and N. Lee, *J. Biol. Chem.*, 1969, **244**, 4277. [p] G. L. Cottam, P. F. Hollenberg, and M. J. Coon, *J. Biol. Chem.*, 1969, **244**, 1481. [q] D. B. Wilson and D. S. Hogness, *J. Biol. Chem.*, 1969, **244**, 2132. [r] I. M. Vasilets, L. A. Konnova, V. P. Kushner, V. M. Bozhkov, E. P. Zdrodovskaya, and G. V. Mukha, *Biokhimiya*, 1968, **33**, 1285. [s] W. Lenzinger, M. Goldberg, and E. Cauvin, *J. Mol. Biol.*, 1969, **40**, 217. [t] J. M. Brewer, T. Fairwell, J. Travis, and R. E. Lovins, *Biochemistry*, 1969, **9**, 1011. [u] C. L. Sia, S. Traniello, S. Pontremoli, and B. L. Horecker, *Arch. Biochem. Biophys.*, 1969, **132**, 325. [v] J. S. Easterby and M. A. Rosemeyer, *F.E.B.S. Letters*, 1969, **4**, 84. [w] C. Peraino, L. G. Bunville, and T. N. Tahmisian, *J. Biol. Chem.*, 1969, **244**, 2241. [x] S. Wilk, A. Meister, and R. H. Haschemeyer, *Biochemistry*, 1969, **8**, 3168. [y] D. J. Horgan, J. R. Dunstone, J. K. Stoops, E. C. Webb, and B. Zerner, *Biochemistry*, 1969, **8**, 2006. [z] G. M. Hathaway, S. Kida, and I. P. Crawford, *Biochemistry*, 1969, **8**, 989. [aa] C. Nauciel, J. Blass, R. Mangalo, and M. Raynaud, *European J. Biochem.*, 1969, **11**, 160. [bb] R. Stern and A. Peterkofsky, *Biochemistry*, 1969, **8**, 4346. [cc] G. Dennert and D. Eaker, *F.E.B.S. Letters*, 1970, **6**, 257. [dd] J. W. Hastings, K. Weber, J. Freidland, A. Eberhard, G. W. Mitchell, and A. Gunsalus, *Biochemistry*, 1969, **8**, 4681. [ee] G. M. Malacinski and W. J. Rutter, *Biochemistry*, 1969, **8**, 4382.

The cytochrome *c* from *Neurospora crassa* is interesting in that it contains two residues of ε-*N*-trimethyl-lysine.[313] No mono- or di-methyl-lysine is observed and, like comparable modifications in other proteins, the methylation takes place enzymically after the biosynthesis of the peptide chain.[314] It is possible that conversion to the methylated form may reflect binding of the cytochrome *c* to the mitochondrion.[314]

D. Studies on Quaternary Structure.—The quaternary structures of many proteins have recently been collated.[315] Some newer ones have been mentioned elsewhere in this Report; others are listed in the Table.

6 Immunoglobulins

The structure of immunoglobulins and the biological implications continue to be well served by recent reviews.[316] The structural feat of the year has doubtless been the elucidation of the entire covalent structure of a myeloma IgG molecule, the heavy chain of which contains 446 residues and the light chain 214 residues.[317] Comparison with another similar molecule suggests that the variable region of the heavy chain is comparable in length with that of the light chain, *i.e.* approximately 110 residues from the *N*-terminus. To reassure protein chemists that expenditure of energy on this scale serves more than academic interest, it has been shown, more convincingly than previously, that isolated heavy chains will recover specific antibody activity on reoxidation after complete reduction and unfolding in solutions of guanidine hydrochloride.[318] Variations in amino-acid sequence must therefore form the basis of antibody specificity, in accordance with current expectation.

A. Light Chains.—The complete amino-acid sequence of a κ-type human Bence-Jones protein has been reported.[319] The idea, previously suggested for κ-chains, that the sequence changes in light chains can be ascribed to variations of basic sequences, or sub-groups, has now been extended to the λ-chains,[320, 321] and the complete amino-acid sequence of a human λ-chain allotted to one of these sub-groups has been established.[320]

Bence-Jones proteins may be cleaved *in vitro* by a proteolytic factor in urine as well as by several endopeptidases, suggesting a particular susceptibility to proteolysis at the switch region between the variable and invariant

[313] R. J. DeLange, A. N. Glazer, and E. L. Smith, *J. Biol. Chem.*, 1969, **244**, 1385.
[314] W. A. Scott and H. K. Mitchell, *Biochemistry*, 1969, **8**, 4282.
[315] I. M. Klotz and D. W. Darnall, *Science*, 1969, **166**, 126.
[316a] G. M. Edelman and W. E. Gall, *Ann. Rev. Biochem.*, 1969, **38**, 415; [b] F. W. Putnam, *Science*, 1969, **163**, 633.
[317] G. M. Edelman, B. A. Cunningham, W. E. Gall, P. D. Gottlieb, U. Rutishauser, and M. J. Waxdal, *Proc. Nat. Acad. Sci. U.S.A.* 1969, **63**, 78.
[318] J.-C. Jaton, N. R. Klinman, D. Givol, and M. Sela, *Biochemistry*, 1968, **7**, 4185.
[319] K. Titani, T. Shinoda, and F. W. Putnam, *J. Biol. Chem.*, 1969, **244**, 3550.
[320] H. Ponstingl and N. Hilschmann, *Z. physiol. Chem.*, 1969, **350**, 1148.
[321] L. Hood and D. Ein, *Nature*, 1968, **220**, 764.

regions.[322] This presumably explains the occasional presence of half-molecules of Bence-Jones proteins in some urine. Since κ- and λ-chains differ in their *C*-terminal sequences, digestion with carboxypeptidase may be used to assign human Bence-Jones proteins to their appropriate type.[323] The light chains of six Waldenström macroglobulins, four of which show some antibody activity, appear to be entirely comparable with the light chains found in other classes of immunoglobulins (IgG and IgA), in accord with former ideas.[324] The variable regions of pig λ-chains have been examined and shown to be very similar to those of human λ-chains[325] and the light chains of goat IgG have also been found to correspond to the human λ-chain since they contain *C*-terminal serine and a blocked *N*-terminal residue.[326]

The *C*-terminal sequence of the guinea pig κ-chain has been shown to be -Thr-Ile-Asn-Arg-Ser-Glu-Cys,[327] closely similar to that of other animals. Variations in the *C*-terminal sequence of rabbit light chains have been correlated with the light chain allotypes:[328]

*b*4	-Ser-Phe-Asn-**Arg**-Gly-Asn-**Cys**
*b*5	-Phe-Ser -**Arg**-Lys-Asn-**Cys**
*b*6	-Ser -**Arg**-Lys-Ser -**Cys**

Other changes associated with the allotype cannot be excluded: the presence of multiple variations in the allotypes is, however, clearly established by this work. Multiple variation has been found previously with other allotypes too. As with the subgroups referred to above in human light chains, there is now evidence that subtypes of rabbit κ-chains, differing among other things in the number of disulphide bridges, may be found associated with the *b* locus.[329]

B. Heavy Chains.—Further progress has been reported on the examination of the primary structure of the *N*-terminal half (F_d region) of pooled rabbit IgG. It has now been found[330] that an underlying sequence for almost all the F_d region can be established with the exception of a short region of some 10—15 residues at about position 100 from the *N*-terminus, suggesting that amino-acid sequence variation in this region may be closely associated with antibody specificity. The major *C*-terminal residue of the heavy chain of goat IgG is glycine: the major *N*-terminal sequence is

[322] A. Solomon and C. I. McLaughlin, *J. Biol. Chem.*, 1969, **244**, 3393.
[323] A. B. Edmundson, N. B. Simonds, F. A. Sheber, T. Johnson, and B. Bangs, *Arch, Biochem. Biophys.*, 1969, **132**, 502.
[324] A. P. Kaplan and H. Metzger, *Biochemistry*, 1969, **8**, 3944.
[325] F. Franěk and J. Novotný, *European J. Biochem.*, 1969, **11**, 165; F. Franěk, B. Keil, and F. Šorm, *ibid.*, 1969, **11**, 170.
[326] D. Givol and E. Hurwitz, *Biochem. J.*, 1969, **115**, 371.
[327] M. E. Lamm and B. Lisowska-Bernstein, *Nature*, 1968, **220**, 712.
[328] E. Appella, J. Rejnek, and R. A. Reisfeld, *J. Mol. Biol.*, 1969, **41**, 473.
[329] J. Rejnek, E. Appella, R. E. Mage, and R. A. Reisfeld, *Biochemistry*, 1969, **8**, 2712.
[330] R. G. Fruchter, S. A. Jackson, L. E. Mole, and R. R. Porter, *Biochem. J.*, 1970, **116**, 249.

Glp-Val-Gln-.[326] The A11 and A12 allotypes, detected by haemagglutination techniques, in the heavy chains of rabbit IgG have also been investigated.[331] In the A11 allotype, the residue on the *N*-terminal side of the inter-heavy-chain disulphide bridge is methionine: in the A12 allotype, the corresponding residue is threonine. These allotypes should be distinguished from the group *a* allotypes correlated with sequence changes elsewhere in the polypeptide chain.[332]

Other work has centred on the heavy chain from IgM. Four heavy chains from human IgM have been examined by peptide mapping techniques with the conclusion that, like the heavy and light chains of other classes of immunoglobulins, these chains contain a region of common amino-acid sequence and a region of variation.[333] There appear to be three inter-heavy-chain disulphide bridges in human IgM together with a single disulphide bridge linking the heavy and light chains.[334] The variation in the amino-acid sequence at the *N*-terminal end of heavy chains from IgG and IgM has been reviewed.[316b] Examination of the *N*-terminal sequences of four different μ-chains from human IgM shows that variation at the *N*-terminal end occurs and that considerable homology is shown with the γ-chains (and λ-chains) of human IgG, in accord with the suspected common genetic origin:[335]

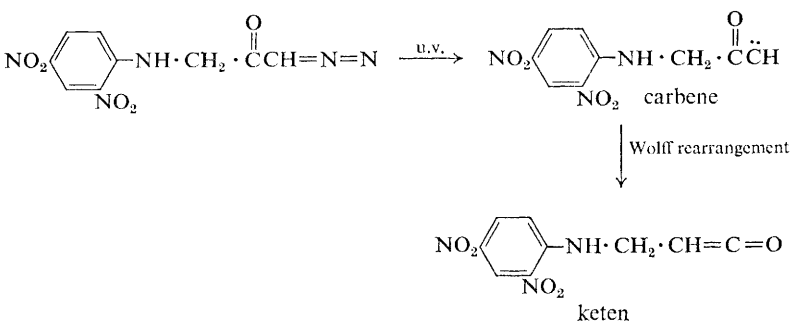
μ-{ **Glp-Ser-Val**-Ala-Asx-
Glp-Ser-Val-Leu-Asx-
Glp-Ser-Val-Ala-Glx-
Glp-Ser-Val-Leu-Asx-

Typical γ-{ **Glp——Val**-Thr-Leu-Arg-
Glp——Val-Thr-
Glp-Ser-Val-Glu-Glu-

C. Antibody Binding Sites.—A major problem in immunochemistry awaiting solution is that of the nature of the interaction of antigen and antibody. The problem is compounded, of course, by the inability as yet to prepare a chemically homogeneous antibody against a chemically defined antigen. Two possible approaches present themselves. In the first, one seeks to discover antibody activity for any of the homogeneous myeloma proteins isolated in large numbers from man and mouse. Further hopeful steps in this direction are being made.[324] It scarcely needs pointing out that assigning an antibody activity to a myeloma protein can be a very tedious process, even with a reliable and simple screening procedure. However, once an activity is discovered, there is the great advantage of dealing with a chemically unique protein.

[331] J. W. Prahl, W. J. Mandy, and C. W. Todd, *Biochemistry*, 1969, **8**, 4935.
[332] J. M. Wilkinson, *Biochem. J.*, 1969, **112**, 173; *Nature*, 1969, **223**, 616.
[333] J. C. Bennett, *Arch. Biochem. Biophys.*, 1969, **131**, 551.
[334] D. Beale and N. Buttress, *Biochim. Biophys. Acta*, 1969, **181**, 250.
[335] J. C. Bennett, *Biochemistry*, 1968, **7**, 3340.

An alternative technique for defining the active site of an antibody is that of affinity labelling. For example, if anti-Dnp antibody is raised in rabbits or mice and then treated with the tritiated diazonium reagent, [^{3}H]*m*-nitrobenzenediazonium fluoroborate, it may be hoped that susceptible amino-acid side-chains in the antibody active site will become labelled.[336] In this particular case, it is found that tyrosine residues in both heavy and light chains are modified, suggesting that both chains participate in the active site. After proteolytic digestion with the enzyme Nagarse, the tritium is found in dipeptides of the form, X—label, where the labelled residue is *m*-nitrobenzene azotyrosine. In the heavy chain of rabbit IgG antibody, X is mostly threonine and in the light chain X is mostly valine. These findings are in accord with the postulated common genetic origin of heavy and light chains and, as expected, comparison with the published amino-acid sequences of human Bence-Jones proteins reveals that the labelled amino-acid must be in the variable, *N*-terminal, half of the antibody light chain. An interesting new approach to the problem of devising affinity labels is that of using photosensitive reagents.[337, 338] Thus, when photolysed, diazoketones yield carbenes which potentially may react even with hydrocarbon side-chains of the antibody or may rearrange to give ketens with more limited reactivity. A typical reaction using dinitrophenylglycine diazoketone, an affinity label for anti-Dnp antibodies, is illustrated in Scheme 8.[337]



Scheme 8

The use of 4-azido-2-nitrophenyl antigenic groups[338] is particularly attractive, because the specific antibody is produced against the precursor (azide) of the labelling reagent itself (the nitrene derived by irradiation of the azide hapten), and the nitrene is capable of reacting with virtually any group within its reach. Further developments with this type of reagent are awaited with much interest. Of course, the use of myeloma

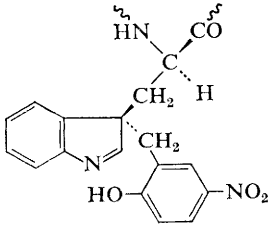
[336] N. O. Thorpe and S. J. Singer, *Biochemistry*, 1969, **8**, 4523.
[337] C. A. Converse and F. F. Richards, *Biochemistry*, 1969, **8**, 4431.
[338] G. Fleet, J. R. Knowles, and R. R. Porter, *Nature*, 1969, **224**, 511.

proteins with antibody activity and the techniques of affinity labelling are not mutually exclusive and one can look forward to a happy combination of the methods in the future.

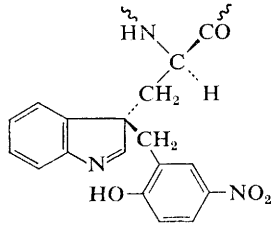
7 Chemical Modification of Proteins

Many aspects of the chemical modification of proteins have already been alluded to in earlier sections of this Report. However, it might be useful to collect together some further recent information on the subject. Although it cannot properly be called a chemical modification, it is worth noting that the methionine residues of staphylococcal nuclease can be replaced by norleucine in a most elegant way by growing a methionine auxotroph of *S. aureus* in the presence of limiting methionine and a large excess of norleucine.[339] An active variant of the nuclease can then be isolated in which norleucine replaces methionine.

A. Tryptophan, Tyrosine, and Histidine Residues.—A single tryptophan residue at position 59 in ribonuclease T_1 has been specifically oxidised with *N*-bromosuccinimide with associated loss of enzymic activity.[340] The same residue will also react with 2-hydroxy-5-nitrobenzyl bromide.[341] Until recently, the nature of the reaction of this reagent with tryptophan residues has been obscure. However, it now appears [342, 343] that the simple alkylation products (7) and (8) are formed first, and may then undergo ring



(7)



(8)

closure with any of a number of suitable nucleophiles, *e.g.* free amino-groups, depending on the environment of the tryptophan residue undergoing modification.

Further experiments on the reaction of tryptophan residues with sulphenyl halides have been described. For example, the 2-thioaryltryptophan

[339] C. B. Anfinsen and L. G. Corley, *J. Biol. Chem.*, 1969, **244**, 5149.
[340] S. Kawashima and T. Ando, *Internat. J. Protein Res.*, 1969, **1**, 185.
[341] T. Terao and T. Ukita, *Biochim. Biophys. Acta*, 1969, **181**, 347.
[342] G. M. Loudon, D. Portsmouth, A. Lukton and D. E. Koshland jun., *J. Amer. Chem. Soc.*, 1969, **91**, 2792.
[343] B. G. McFarland, Y. Inoue, and K. Nakanishi, *Tetrahedron Letters*, 1969, 857.

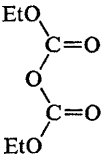
residue resulting from the modification is converted into 2-hydroxytryptophan under the conditions of peptide bond hydrolysis [344] and *o*-nitrophenylsulphonyl chloride has been used to modify the tryptophan residues of dolphin myoglobin.[345]

Tetranitromethane continues to be widely used for the modification of tyrosine residues. Thus, the specific reaction of certain tyrosine residues in trypsin,[346] ribonuclease,[347] and dolphin myoglobin [345] have been reported. Nitration or iodination of tyrosine residues in subtilisin Carlsberg leads to a six- or seven-fold increase in the rate of hydrolysis of the substrate clupeine, although hydrolysis of low-molecular-weight ester substrates remains unchanged.[348] Subsequent reduction of the nitrotyrosyl residues decreased the rate of clupeine hydrolysis to that found for the unmodified enzyme. Unfortunately, the nitration of tyrosine residues is not the sole reaction of tetranitromethane with proteins. The concomitant oxidation of cysteine residues was recognised in the earliest use of the reagent and this reaction has now been more fully documented.[349] The primary reaction is to form disulphide bridges, although the oxidation can proceed further to sulphinic acids. At pH 5·5, moreover, the reaction with thiol groups occurs without the simultaneous nitration of tyrosine residues. These authors also warn that other side-reactions with histidine, tryptophan, and methionine residues can occur under certain conditions. The reaction with tryptophan residues has, in fact, been noted before [350] and, indeed, the inhibition of papain by tetranitromethane has been attributed to the formation of the 7-nitro derivative of tryptophan residues rather than modification of the catalytic thiol group.[351] The iodination of tyrosine residues in lysozyme has been described, two residues (positions 20 and 23) both becoming di-iodinated and the other (position 53) proving unreactive.[352]

The use of *N*-acetylimidazole and 1*H*-diazonium-tetrazole to determine the 'free' and 'buried' tyrosine and histidine residues in bovine fibrinogen has been reported [353] and the reaction of trypsin with *N*-acetylimidazole has also been described.[354] The photo-oxidation of histidine to aspartic acid and urea *via* several intermediates gives the same products whether riboflavin, Rose Bengal, or Methylene Blue is used as the sensitiser,

[344] F. M. Veronese, A. Fontana, E. Boccu, and C. A. Benassi, *Z. Naturforsch.*, 1968, **23b**, 1319.
[345] P. Nedkov and B. Meloun, *Coll. Czech. Chem. Comm.*, 1969, **34**, 2021.
[346] V. Holeyšovský, B. Keil, and F. Šorm, *F.E.B.S. Letters*, 1969, **3**, 107.
[347] G. H. Beaven and W. B. Gratzer, *Biochim. Biophys. Acta*, 1968, **168**, 456.
[348] I. Svendsen, *Compt. rend. Trav. Lab. Carlsberg*, 1968, **36**, 347.
[349] M. Sokolovsky, D. Harell, and J. F. Riordan, *Biochemistry*, 1969, **8**, 4740.
[350] P. Cuatrecasas, S. Fuchs, and C. B. Anfinsen, *J. Biol. Chem.*, 1968, **243**, 4787.
[351] K. Morihara and K. Nagami, *J. Biochem.* (*Japan*), 1969, **65**, 321.
[352] K. Hayashi, T. Shimoda, T. Imoto, and M. Funatsu, *J. Biochem.* (*Japan*), 1968, **64**, 365.
[353] R. M. Huseby and M. M. Murray, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 169.
[354] L. L. Houston and K. A. Walsh, *Biochemistry*, 1970, **9**, 156.

although the relative yields may vary.[355] Histidine may also be treated with diethyl pyrocarbonate (9) to form the carbethoxylated derivative, and



(9)

some results with the modification of actin have been given.[356] The same reagent may also be used for the spectroscopic determination of histidine in proteins.[357]

B. Cysteine Residues and Disulphide Bridges.—The properties of subtilisin in which the catalytically active serine residue has been turned into a thiol group have been documented:[358] although the modified enzyme is inactive towards normal substrates, substrate binding remains essentially unchanged. This has been ascribed to the lack of formation of the normal acyl enzyme, contrary to what one would expect from the behaviour of model thiols and alcohols in nucleophilic reactions. A warning has been given that 'thiol' enzymes may be inactivated by phenyl thiocyanate, perhaps according to Scheme 9.[359] The selective photo-oxidation of cysteine residues can be

$$R^1S^- + R^2 \cdot SCN \rightleftharpoons R^1S \cdot SR^2 + CN^-$$

Scheme 9

achieved using Cresol Red and Crystal Violet as sensitisers,[360] In acid solution, cysteine is converted into cysteic acid: in neutral or alkaline solution one gets cystine as the product of a competing dark reaction. The oxidation of the thiol groups of creatine kinase by treatment with iodine to produce, among other things, sulphenyl oxides has also been described.[361]

Treatment of glyceraldehyde 3-phosphate dehydrogenase with fluorodinitrobenzene under mild conditions in the absence of NAD^+ leads to loss of enzymic activity.[362a] All the label is found as *S*-Dnp-cysteine and the label may be completely removed, with restoration of enzyme activity, by

[355] M. Tomita, M. Irie, and T. Ukita *Biochemistry*, 1969, **8**, 5149.
[356] A. Mühlrad, G. Hegyi, and M. Horanyi, *Biochim. Biophys. Acta*, 1969, **181**, 184.
[357] J. Ovadi, S. Libor, and P. Elodi, *Acta Biochim. Biophys. Acad. Sci. Hung.*, 1967, **2**, 455.
[358] K. E. Neet, A. Nanci, and D. E. Koshland jun., *J. Biol. Chem.*, 1968, **243**, 6392.
[359] S. Mahadevan, P. S. Shukla, V. S. Kalyanaraman, and S. A. Kumar, *F.E.B.S. Letters*, 1969, **2**, 149.
[360] G. Jori, G. Galiazzo, and E. Scoffone, *Internat. J. Protein Res.*, 1969, **1**, 289.
[361] D. Trundle and L. W. Cunningham, *Biochemistry*, 1969, **8**, 1919.
[362a] S. Shaltiel and M. Soria, *Biochemistry*, 1969, **8**, 4411; [b] K. Wallenfels and B. Eisele, *European J. Biochem.*, 1968, **3**, 267.

treatment with 2-mercaptoethanol. The inactivation may be prevented by the presence of NAD^+ and provides a neat reversible modification of thiol groups at the active site of the enzyme. Stereospecific alkylation with asymmetric reagents can be used to probe the environment of a cysteine residue. Thus, the D(+) and L(−) antipodes of α-iodopropionic acid and its amide react at different rates with different enzymes, presumably because the orientation of the reagent is affected by neighbouring amino-acid side-chains.[362b]

In addition to their reaction with tryptophan residues, the sulphenyl halides will also react with cysteine residues and, under certain conditions, the reaction is exclusively with sulphydryl groups. Thus, azobenzene-2-sulphenyl bromide, a water-soluble sulphenyl halide, reacts selectively with protein sulphydryl groups at pH 5.[363] It is worth noting, too, that the number of so-called 'accessible' functional groups and their individual rates of reaction may well vary with the reagent used to probe their accessibility. For example, in rabbit muscle aldolase the reactivity of given sulphydryl groups towards different reagents can vary widely.[364]

The chemical modification of protein disulphide bridges has a major advantage in that it can be made entirely specific. The reduction of protein disulphide bridges with dithiothreitol in the absence of denaturing agents has been the subject of a recent study.[365] It was observed that lysozyme, prolactin, and insulin can be totally reduced, bovine serum albumin can only be partly reduced, and ribonuclease is not affected at all. One disulphide bridge, linking positions 179 and 203, out of a total of six in trypsinogen can be specifically reduced with sodium borohydride.[366] After carboxymethylation of the two cysteine residues thereby generated, the zymogen can still be activated despite the fact that the loop reduced contains the serine residue of the active site. The slight changes in enzyme activity suggest, however, that there are subtle changes in the conformation of the protein. A single disulphide bridge linking positions 43 and 152 can also be opened with 2-mercaptoethanol in the enzyme papain [367] and then closed again with the formation of a —S—Hg—S— link by treatment with mercury chloride.[368] Similar experiments have also been reported to be possible with ribonuclease.[369] It may well be that crystallographers will be able to take advantage of reactions such as these for the preparation of specific isomorphous heavy-atom derivatives.

In other experiments [369] it has been shown that the derivatives of human pituitary growth hormone formed by total reduction and alkylation with

[363] A. Fontana, F. M. Veronese, and E. Scoffone, *Biochemistry*, 1968, **7**, 3901.
[364] P. J. Anderson and R. N. Perham, *Biochem. J.*, 1970, **117**, 291.
[365] T. A. Bewley and C. H. Li, *Internat. J. Protein Res.*, 1969, **1**, 117.
[366] A. Light, B. C. Hardwick, L. M. Hatfield, and D. L. Sondack, *J. Biol. Chem.*, 1969, **244**, 6289.
[367] E. Shapira and R. Arnon, *J. Biol. Chem.*, 1969, **244**, 1026.
[368] R. Arnon and E. Shapira, *J. Biol. Chem.*, 1969, **244**, 1033.
[369] T. A. Bewley, J. Brovetto-Cruz, and C. H. Li, *Biochemistry*, 1969, **8**, 4701.

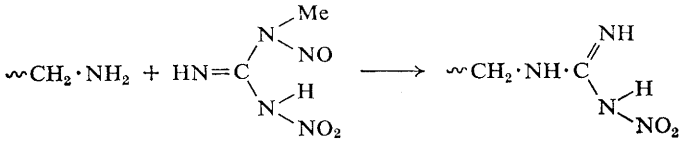
either iodoacetic acid or iodoacetamide retain lactogenic activity, although only the latter derivative has growth-promoting potency. In all other respects, the derivatives and the native hormone appear to be very similar in structure, in line with the view that, in general, disulphide bridges are not required to form the three-dimensional structure of a protein but serve to increase stability.

C. Amino-groups.—A method for analysing kinetically the reactivity of protein amino-groups towards trinitrobenzenesulphonic acid has been given in detail [370] and a refined technique for measuring the number of amino-groups in a protein using the same reagent has been described.[371] Following modification with such reagents, some proteins may become poorly soluble and it has been suggested that a better reagent might be β-naphthoquinone-4-sulphonic acid with additional sulphonic acid groups introduced at the 6- or 7-position of the napthoquinone to promote solubility.[372]

Under carefully controlled conditions, acetic anhydride can be used to obtain cytochrome *c* in which certain ε-amino-groups are modified.[373] A development of this type of approach allows the p*K* of an amino-group to be measured.[374] The protein is allowed to react with a limiting amount of radioactive acetic anhydride and then reacted to completion with a huge excess of unlabelled acetic anhydride. The protein is digested enzymically and the specific activity of each lysine-containing peptides is then measured, giving the relative rates of reaction of individual ε-amino-groups. By determining the rates of reaction as a function of pH, one can calculate the p*K* of the reacting group.

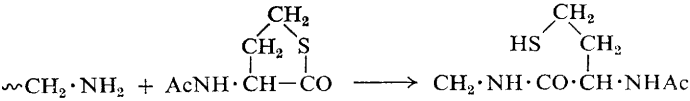
The reaction of the three amino-groups of insulin with diketen has been examined.[375] The ε-amino-group appears not to react at low concentrations of reagent but difficulty was noted in securing the reversal of the aceto-acetylation by treatment with hydrazine. It has been recorded that during the cyanoethylation of insulin, acrylonitrile reacts not only with the α- and ε-amino-groups but also with the histidine residues.[376] Reaction of histidine residues can also occur during coupling of chymotrypsin and chymotrypsinogen with *p*-nitrobenzenediazonium fluoroborate to form the diazo-derivatives of the ε-amino-groups.[377] The use of nitroguanidino derivatives for reaction with the amino-groups of proteins has been outlined by two groups of workers (Scheme 10).[378] The nitrohomoarginine

[370] R. B. Freedman and G. K. Radda, *Biochem. J.*, 1969, **114**, 611.
[371] R. E. Fields, Ph.D. Thesis, University of Cambridge, 1970.
[372] A. Matsushima, K. Sakurai, M. Nomoto, Y. Inada, and K. Shibata, *J. Biochem. (Japan)*, 1968, **64**, 507.
[373] K. Wada and K. Okunuki, *J. Biochem. (Japan)*, 1968, **64**, 667.
[374] H. Kaplan, B. S. Hartley, and K. J. Stevenson, *Fed. Proc.*, 1969, **28**, 533.
[375] D. G. Lindsay and S. Shall, *Biochem. J.*, 1969, **115**, 587.
[376] H. R. Bosshard, K. H. Jørgensen, and R. E. Humbel, *European J. Biochem.*, 1969, **9**, 353.
[377] V. T. Maddaiah, *Canad. J. Biochem.*, 1969, **47**, 423.
[378] D. R. McCalla and A. Reuvers, *Canad. J. Biochem.*, 1968, **46**, 1411.



Scheme 10

formed from lysine residues carries no charge on the side-chain and is also unstable to normal acid hydrolysis.[378a] The reaction of *N*-acetyl-DL-homocysteine thiolactone with the amino-groups of proteins (Scheme 11)



Scheme 11

has been explored with a view to the formation of isomorphous heavy-atom derivatives for crystallographic analysis.[379] The reaction is catalysed by Ag^+, which may then be removed and the newly exposed sulphydryl group allowed to react with a suitable mercurial.

D. Carboxy-groups.—The use of water-soluble carbodi-imides for coupling protein carboxy-groups and suitable amines, *e.g.* glycine methyl ester or glycinamide, has received further attention. By this method the functional significance of the carboxyl side-chains of Asp-52 and Glu-35 in lysozyme has been demonstrated[380] and it has been shown that almost all the carboxy-groups of chymotrypsin can be modified without loss of enzyme activity. The carboxy-group of Asp-194 is found to react only after denaturation of the protein in urea,[381] in accord with its function in the enzyme active site,[243] and another inaccessible carboxy-group (Asp-102, perhaps) appears to exist. In separate experiments, lysozyme has been treated under mild conditions with triethyloxonium fluoroborate with the production of a single ester derivative of the enzyme that is inactive and yet binds the competitive inhibitor chitotriose.[382] Chemical analysis of the modified enzyme reveals that it is the carboxy-group of Asp-52 that is esterified, in accord with the suspected enzyme mechanism.

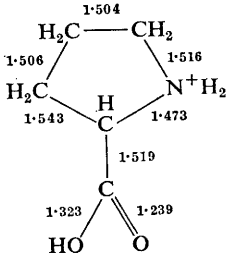
A similar approach may also be possible by using the reaction of *N*-alkyl-5-phenylisoxazolium salts with the carboxy-groups of proteins to form enol esters that can subsequently be caused to react with a suitable nucleophile, *e.g.* glycine methyl ester or hydroxylamine.[383]

[378a] T. Sugimura, S. Fujimura, M. Nagao, T. Yokoshima, and M. Hasegawa, *Biochim. Biophys. Acta*, 1968, **170**, 427.
[379] S. Shall and E. A. Barnard, *J. Mol. Biol.*, 1969, **41**, 237.
[380] T. Y. Lin and D. E. Koshland jun., *J. Biol. Chem.*, 1969, **244**, 505.
[381] K. L. Carraway, P. Spoerl, and D. E. Koshland jun., *J. Mol. Biol.*, 1969, **42**, 133.
[382] S. M. Parsons and M. A. Raftery, *Biochemistry*, 1969, **8**, 4199.
[383] P. Bodlaender, G. Feinstein, and E. Shaw, *Biochemistry*, 1969, **8**, 4941.

PART II: *X*-Ray Studies *by C. C. F. Blake*

1 Amino-acids and Oligopeptides (See also Chapter 1, section 3B.)

A. DL-Proline.—The crystal structure of DL-proline hydrochloride[1] has now been determined. The bond lengths of the molecule are shown on structure (1). The pyrrolidine ring is unusual in that while C_β, C_γ, C_δ and



(1)

the N are coplanar to within 0·02 Å, the α-carbon is 0·5 Å away from the ring plane. In comparison L-proline, hydroxyproline, tosylprolylhydroxyproline, and leucylprolylglycine have the γ-carbon out of the plane of the ring. It is suggested that the puckering of the pyrrolidine ring occurs because of the repulsion of non-ring atoms and that in this molecule the out-of-plane movement of the α-carbon relieves the overcrowding caused by the bulky carboxy-group. It appears that, in general, the pyrrolidine ring is rather flexible and its conformation is affected by various factors, including intermolecular interactions.

B. DL-Valine.—In this determination[2a] the molecule is in the normal zwitterion form, with a $C—N^+H_3$ bond length of 1·483 Å in good agreement with the average value listed by Marsh and Donohue,[2b] and with almost exactly equal C—O bond lengths of 1·248 and 1·249 Å in the carboxylate group. The two C_γ atoms have χ values of 82 and 209° corresponding to the staggered positions I and II. These same positions are occupied in L-valine hydrochloride, while I and III are occupied in L-valine hydrochloride monohydrate. It is interesting to note that positions II and III, which are favourable for the conformation of poly-L-valine in α-helical form, have not yet been observed in the crystal structures of the amino-acid.

C. Dipeptides.—The crystal structures of a number of dipeptides have been reported this year, glycylglycine hydrochloride,[3] alanylalanine hydro-

[1] Y. Mitsui, M. Tsuboi, and Y. Iitaka, *Acta Cryst.*, 1969, **B25**, 2182.
[2a] M. Mallikarjunan and S. Thyagaraja Rao, *Acta Cryst.*, 1969, **B25**, 296; [b] R. E. Marsh and J. Donahue, *Advances in Protein Chemistry*, 1967, **22**, 235.
[3] R. Parthasarathy, *Acta Cryst.*, 1969, **B25**, 509.

chloride,[4] *N*-methyl-leucylglycine hydrobromide,[5] threonylphenylalanine *p*-nitrobenzyl ester hydrobromide,[6] and one pseudo-dipeptide, *N*-acetyl-leucine *N*-methylamide.[7] As might have been expected the peptide groups in all these compounds are planar, with bond lengths and angles very close to the average values listed by Marsh and Donohue.[2b]

The arrangement of hydrogen bonds in the structure of the glycylglycine molecule includes two short contacts for each of the three hydrogens of the terminal $-NH_3^+$ ion. In a brief survey of structures containing bifurcated hydrogen bonds the authors show that while the relative strength of the two interactions varies widely, the hydrogen atom always lies close to the plane of the donor atom and the two acceptors.

The conformation of the alanylalanine dipeptide is very close to that of the antiparallel pleated sheet. The conformational parameters are $\psi_1 = 334{\cdot}2°$, $\phi = 26{\cdot}5°$, $\psi_2 = 341{\cdot}3°$, $\omega = 0{\cdot}8°$ close to the expected values for the pleated sheet of $\phi = 38°$, $\psi = 325°$, $\omega = 0°$, and those found in β-poly-L-alanine of $\phi = 41{\cdot}4°$, $\psi = 314{\cdot}7°$, $\omega = 1{\cdot}5°$.

The leucine side-chain of *N*-methyl-DL-leucylglycine hydrobromide has very similar dihedral angles ($\chi^{11} = 181°$, $\chi^{21} = 63°$, $\chi^{22} = 185°$) to those found in leucine hydrobromide whose corresponding angles are 188°, 58°, and 182°. The angles in two other structures containing the leucine side-chain, leucylglycine hydrobromide and leucylglycylproline, are close to 300, 180, and 300°.

The ϕ and ψ values for the threonyl residue of threonylphenylalanine *p*-nitrobenzyl ester hydrobromide are 70 and 318° respectively, while the ψ_1 and ψ_2 for the *C*-terminal phenylalanine are 119 and 301°. These latter values are almost invariably *ca.* 180 and 360° for peptides. However, for glycylasparagine they are 60—70° larger while in this study they are smaller by about the same amount. It is interesting to observe that in both these molecules the *C*-terminal group is not a carboxy but an amide or an ester. The threonyl side-chain has its C_γ in staggered position II and the hydroxy in position III. The C_γ of the phenylalanine goes into position II.

In DL-acetyl-leucine *N*-methylamide the ϕ and ψ angles are 86·2 and 318·6° respectively, lying between the values for the parallel and anti-parallel pleated sheet structures. This is interesting because in the crystal structure of this molecule the hydrogen-bonding scheme of both types of pleated sheet occurs. D- and L-molecules are arranged alternately head to tail to form chains which are hydrogen-bonded together into sheets. The leucine side-chains are nearly perpendicular to the sheets, forming hydrophobic regions between them. It has been suggested that in β-structures composed only of L-residues, the characteristic puckering of the pleated

[4] Y. Tokuma, T. Ashida, and M. Kakuda, *Acta Cryst.*, 1969, **B25**, 1367.
[5] R. Chandrasekaran and E. Subramanian, *Acta Cryst.*, 1969, **B25**, 2599.
[6] M. Mallikarjunan, S. Thyagaraja Rao, K. Venkatesan, and V. R. Sarma, *Acta Cryst.*, 1969, **B25**, 220.
[7] T. Ichikawa and Y. Iitaka, *Acta Cryst.*, 1969, **B25**, 1824.

sheet is caused by overcrowding of the side-chains. Overcrowding is not observed in this structure, but nevertheless the peptide chains have the puckered conformation which, it is suggested, results from interactions of main-chain atoms with the β-carbons of the side-chains.

D. Glycylprolyl-leucylglycine.—An interesting structural feature of the *p*-bromobenzyloxycarbonyl derivative of this tetrapeptide [8] is the fact that the chain folds back on itself to form an intramolecular hydrogen bond between the two terminal peptide groups. Such a folded structure has not been previously reported for linear peptides. It causes the peptide to have many conformational features of cyclic peptides such as cyclohexaglycine and ferrichrome A. In particular the dihedral angles ϕ and ψ of the proline and leucine residues are quite close to those for two of the ornithine residues of ferrichrome A. The ϕ and ψ values of the *N*-terminal glycine are not far from those of polyglycine II.

The pyrrolidine ring of the proline residue has C_α, C_β, C_δ, and N in a plane while the C_γ is 0·26 Å out of the plane. The ϕ and ψ values of this residue are not far from those of the right-handed α-helix and the $3{\cdot}0_{10}$ helix but quite different from those of collagen and polyproline II. This results from an internal rotation around the C_α—C′ bond which causes the C_α—H and C′—O bonds to be *cis*, as in myoglobin and lysozyme, whereas they are *trans* in collagen. It is intriguing that this may account for the fact that the tetrapeptide is not a substrate of collagenase.

2 Polypeptides

In an attempt to decide between the three proposed models of collagen, Traub and his colleagues have examined the structures of three collagen-like poly-tripeptides and four poly-hexapeptides.[9]

Poly-(Ala-Pro-Gly) [10] in the solid state can exist in several forms characterised by different *X*-ray patterns. The structure of one form, stabilised by water, is in the polyproline II conformation, corresponding to an axial translation of 9·4 Å and a rotation of 360°, approximately equally divided between the three amino-acid residues. Poly-(Pro-Gly-Gly) [11] has a similar structure, the chains of which are held together by two NH—O bonds per tripeptide to form double-layered sheets, with the two layers antiparallel. All the pyrrolidine rings of the proline residues are on the outside and the associated carbonyls are those not involved in hydrogen bonding.

Poly-(Pro-Gly-Pro) [12] closely resembles collagen in its *X*-ray pattern. A detailed conformational analysis has been made of various sterically

[8] T. Ueki, T. Ashida, M. Kakudo, Y. Sasada, and Y. Katsube, *Acta Cryst.*, 1969, **B25**, 1840.
[9] W. Traub, A. Yonath, and D. M. Segal, *Nature*, 1969, **221**, 914.
[10] D. M. Segal and W. Traub, *J. Mol. Biol.*, 1969, **43**, 487.
[11] W. Traub, *J. Mol. Biol.*, 1969, **43**, 479.
[12] A. Yonath and W. Traub, *J. Mol. Biol.*, 1969, **43**, 461.

reasonable models. These computations indicate that there is a unique conformation, shown in Figure 1, which resembles the collagen II model in its mode of interchain hydrogen bonding. Conformations of the collagen I type were found to make short intermolecular contacts and those in which the glycyl imido-group is hydrogen-bonded in the manner of the two-bonded model proved incompatible with the observed intensities. This suggests that the conformation of poly-(Pro-Gly-Pro) may be representative of the structure of collagen as a whole.

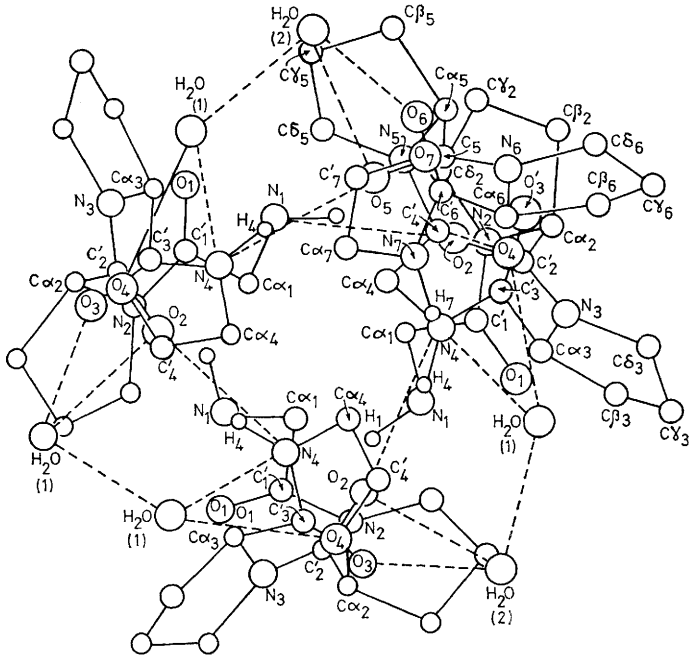


Figure 1 *Projection down the* c*-axis of the structure of* $(Pro\text{-}Gly\text{-}Pro)_n$ *including two water molecules per tripeptide. Dashed lines indicate hydrogen bonds*
(Reproduced with permission from *J. Mol. Biol.*, 1969, **43**, 461)

All four of the polyhexapeptides [13] poly-(Gly-Pro-Ala-Gly-Pro-Pro), poly-(Gly-Ala-Pro-Gly-Pro-Pro), poly-(Gly-Ala-Pro-Gly-Pro-Ala), and poly-(Gly-Ala-Ala-Gly-Pro-Pro) show collagen-like diffraction patterns. A systematic investigation of the sterically allowed collagen-like structures showed that none of the polyhexapeptides adopts the conformation of the two-bonded model of collagen. On the other hand, all four can form sterically acceptable one-bonded structures very similar to poly-(Pro-Gly-Pro). This again suggests that collagen itself has the same one-bonded structure as these synthetic model compounds.

[13] D. M. Segal, W. Traub, and A. Yonath, *J. Mol. Biol.*, 1969, **43**, 519.

Keith, Padden, and Giannoni[14] have described structure analyses obtained from *X*-ray analysis of powder patterns and selected area electron diffraction patterns, of small chain-folded single crystals of polyglutamic acid and its alkaline-earth salts. All the structures were found to be based on the antiparallel pleated sheet.

The calcium salt of polyglutamic acid crystallises in the monoclinic system with $a = 9{\cdot}40$ Å, $b = 6{\cdot}83$ Å, $c = 12{\cdot}82$ Å, and $\beta = 100° 20'$. The planes of the carboxylate ions are parallel to the chain axis; one half of these ions are in contact with one cation and the other half in contact with three cations. The cations can be removed from the crystals by hydrogen exchange. The resulting crystals of the free acid are still monoclinic but the unit cell dimensions have changed to $a = 9{\cdot}66$ Å, $b = 6{\cdot}98$ Å, $c = 9{\cdot}10$ Å, and $\beta = 105°$. The carboxy-groups maintain their original orientation but become intermeshed so that the intersheet spacing is reduced from 12·5—13 Å to 8·8 Å. However, when crystals of the free acid are obtained directly from aqueous solution of the sodium salt by a gradual lowering of the pH the structure is different. The carboxy-groups lie in planes normal to the chain direction giving rise to an even smaller intersheet spacing of 7·8 Å.

3 Globular Proteins

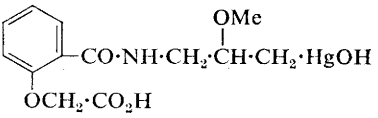
A. Methods.—*Isomorphous Derivatives.* Trial-and-error continues to be the most important means of preparing the isomorphous derivatives needed to solve the phase problem. The new structures reported this year have mostly supplemented one or more specifically prepared derivatives with several obtained by trial and error. A very valuable rationalisation of the chemistry of interaction of $PtCl_4^{2-}$, one of the most useful derivatives, has been made by Dickerson and his colleagues.[15] They surveyed all six protein structures solved at high resolution which had used $PtCl_4^{2-}$ as one of the isomorphous derivatives. In each case there is a strong binding site close to the sulphur atom of an exposed methionine residue. In addition there were two cases of binding to an exposed sulphur of a disulphide bridge and three cases of weak binding to a histidine residue. From these data they have been able to draw the important conclusion that all exposed methionines in the six proteins form powerful binding sites for $PtCl_4^{2-}$.

An interesting proposal on the chemistry of the interaction of $PtCl_4^{2-}$ with methionine residues has been deduced from the nature of its complex formed with Met-65 of horse heart cytochrome *c*. The platinum atom occupies two sites corresponding to the positions of the two lone pairs of electrons on the sulphur atom of the methionine, and located *ca.* 2·3 Å from the sulphur. It is suggested that the platinum is oxidised from Pt^{II} to

[14] H. D. Keith, F. J. Padden, and G. Giannoni, *J. Mol. Biol.*, 1969, **43**, 423.
[15] R. E. Dickerson, D. Eisenberg, J. Varnum, and M. L. Kopka, *J. Mol. Biol.*, 1969, **45**, 77.

Pt^{IV}, accepting the sulphur lone pair into one of the new octahedral co-ordination sites, and picking up the sixth ligand from the solution.

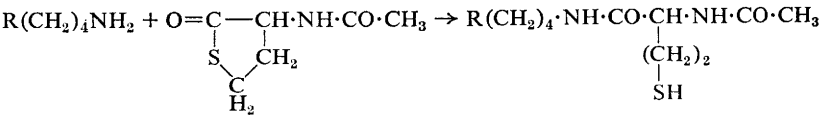
On the basis of these observations Dickerson and his colleagues suggest that it would be useful to examine three particular heavy-atom compounds at the beginning of a crystallographic structure determination of a protein; *p*-chloromercuribenzoate to search for —SH groups, $PtCl_4^{2-}$ to investigate exposed methionines, and mersalyl (2) as a histidine reagent. However, the



(2)

latter compound is large and may well cause non-isomorphism, and in addition its reputation as a histidine reagent rests solely on its behaviour in subtilisin. The list of derivatives could with advantage be lengthened to include uranyl salts such as $UO_2(NO_3)_2$ and $K_3UO_2F_5$ as possible carboxylic acid reagents and K_2HgI_4 to search for accessible hydrophobic areas.

A useful modification of the specific reaction of protein amino-groups with *N*-acetylhomocysteine thiolactone to form new —SH groups, first suggested by Benesch and Benesch, has been examined by Shall and Barnard [16] (Scheme). Ribonuclease was treated with the thiolactone in the



Scheme

presence and absence of silver ions at neutral pH. In the presence of silver ions the reaction is faster and more specific for particular lysine residues. Shall and Barnard suggest that the reaction may involve a nearby histidine in a catalytic role, and thus lysines with histidine neighbours tend to be favoured in the reaction. The reaction of ribonuclease with the thiolactone in the presence of silver ions leads to two products each containing a free thiol group. The two products could be separated chromatographically; they were crystallised and caused to react with *p*-chloromercuribenzoate. Avey and Shall [17] have examined the two derivatives crystallographically. They have found the mercury atoms in each derivative and have shown that they are located at different sites. However, in each derivative the *b*-axis of the crystals is 3·2 Å longer than in the native enzyme. It may be

[16] S. Shall and E. A. Barnard, *J. Mol. Biol.*, 1969, **41**, 237.
[17] H. P. Avey and S. Shall, *J. Mol. Biol.*, 1969, **43**, 341.

possible to overcome the non-isomorphism by using one of the lactone-reacted ribonuclease derivatives as the parent protein.

Direct Methods. Two recent investigations have been made into the possibility of replacing or combining multiple isomorphous replacement with statistical phasing. These methods have been used for the solution of increasingly complex structures first in centrosymmetric crystals and more recently for non-centrosymmetric crystals as well. Little, however, has been published on their possible uses in the problem of protein structures.

Weinzierl, Eisenberg, and Dickerson [18] have examined the usefulness of Kerle and Hauptman's tangent formula:

$$\tan \alpha_H \sim \frac{\sum_K |E_K E_{H-K}| \sin(\alpha_K + \alpha_{H-K})}{\sum_K |E_K E_{H-K}| \cos(\alpha_K + \alpha_{H-K})},$$

where α_K = phase of reflexion K,

and E_K = normalised structure factor of reflexion K,

and $\sim$ = probably equal to,

in a model protein structure and on horse heart cytochrome *c*, both at medium resolution. The model structure was made up of the polypeptide backbone, α-carbons, and haem group of sperm whale myoglobin. Calculated structure amplitudes and phases to 4 Å resolution for this structure were used to test the accuracy of the tangent formula, at low and medium resolutions. Phases calculated from the tangent formula were found to be very precise for the largest E values but the error increased with decreasing E values. The formula also begins to fail at higher E values when the resolution of the data is reduced from 4 to 5 Å. This seems to suggest that the formula may be of greater value at resolutions of *ca.* 2 Å. Refinement of the phases led to a rapid reduction of errors followed by a drift into an incorrect but self-consistent set of phases.

Tests carried out on horse heart cytochrome *c* used the observed data phased by two heavy-atom derivatives to 3·7 Å resolution. The appearance of the haem group was used as the criterion of improvement. Use of the tangent formula to refine phases obtained by the two derivatives led to inconclusive results. However, much more encouraging results were obtained by combining isomorphous replacement with the tangent formula. They used only one of the derivatives for phase determination and brought in the tangent formula to choose between the two ambiguous phase values predicted by the heavy atom. The results from both cytochrome *c* and from the model myoglobin, in which a single hypothetical heavy atom was used to give the phase ambiguity, were sufficiently good for the authors to suggest that this method could be used in practice when only one very good isomorphous derivative exists.

[18] J. E. Weinzierl, D. Eisenberg, and R. E. Dickerson, *Acta Cryst.*, 1969, **B25**, 380.

In the second study of the usefulness of the tangent formula in protein crystallography, Reeke and Lipscomb[19] come to rather less optimistic conclusions. They applied the formula to the carboxypeptidase data to determine its usefulness in refining multiple isomorphous phases at various resolutions and for the extension of these phases to higher resolution. The results were judged by comparing electron density maps from the various phase sets with the known 2 Å structure of carboxypeptidase, and by examining the average differences between isomorphous phases and tangent formula phases. Contrary to theoretical predictions and the expectations of Weinzierl, Eisenberg, and Dickerson, the tangent formula did not work better as the resolution was increased from 6 to 2 Å. This may be related to the poor initial phasing inevitably given by isomorphous replacement at very high resolutions. It may also correlate with the observation that the usefulness of the tangent formula in extending the resolution of data is limited to *ca.* 3·5 Å. However, Reeke and Lipscomb have found that the formula is of definite value in improving low resolution phases by refinement. Like Weinzierl, Eisenberg, and Dickerson, they found that the best phases were obtained in the first three or four cycles of refinement at low resolution and after only one cycle at high resolution. Further cycles of refinement improved the index of self-consistency but the phases themselves deteriorated.

Diffraction Techniques. An investigation of the advantages of using Cu K_α *X*-rays monochromatised by a plane graphite crystal as compared with the normal Ni-filtered radiation has been carried out on cytochrome *c* peroxidase.[20] The experiments were performed on a Buerger–Supper precession camera, at a constant temperature of 4° C, and the intensity data measured with a computer-controlled film scanner. It was found that a 25% longer exposure was needed when using the monochromator to obtain the same intensities as with the Ni-filter. This value depends very critically on the thickness of the filter and the longer exposure is in any case completely offset by the advantages of using the monochromator. These include a marked decrease in background scattering on the film which resulted in an increase of 10% in the number of observed reflexions and an increase in the precision of measuring reflexions. Even more important was the reduction in radiation damage; a factor of 2 is quoted as the increase in the stability of the crystal to the monochromatised as opposed to the Ni-filtered radiation. It is to be hoped that these considerable benefits are not offset by any systematic error caused by the monochromator. It will be interesting to see how monochromators of this kind behave on diffractometers, where the homogeneity of the *X*-ray beam, often a particular problem with monochromators, is of possibly greater importance than with film methods.

[19] G. N. Reeke and W. N. Lipscomb, *Acta Cryst.*, 1969, **B25**, 2614.
[20] L.-O. Hagman, L. O. Larsson and P. Kierkegaard, *Internat. J. Protein Res.*, 1969, **1**, 283.

From the protein structure point of view, one of the chief disadvantages of *X*-ray diffraction is its inability to locate hydrogen atoms. This can be particularly important for some enzymes in which the actual state of protonation of a particular side-chain may decide between one catalytic mechanism and another. The use of neutron diffraction in these circumstances would, in principle, allow the positions of hydrogen atoms to be determined as precisely as for any other atom in a protein. This arises because the neutron scattering amplitudes of the atoms normally present in proteins vary only slightly with atomic species, the lowest being sulphur with a value of 3·2 and the highest nitrogen with a value of 9·4. Hydrogen is distinctive since it is the only atom normally present in proteins which has a negative scattering amplitude, its value being −3·8.

A study by Schoenborn [21] on sperm whale myoglobin has shown that neutron analysis of proteins is feasible. He used very large crystals of myoglobin, whose volumes ranged between 5 and 25 mm^3, grown from aqueous ammonium sulphate. The crystals were repeatedly soaked in ammonium sulphate–D_2O solutions to exchange the water of crystallisation and the exchangeable protein hydrogens for deuterium. This was necessary to reduce the background caused by the incoherent scattering from the hydrogen atoms. Monochromatic neutrons of 1·527 Å wavelength were obtained from a germanium monochromator, which gave a flux of *ca.* 10^6 neutrons cm^{-2} s^{-1}. Data for the centric *h*0*l* zone, collected on a four-circle diffractometer, showed that the background was *ca.* 100 c/min with peaks up to 15,000 c/min. A three-dimensional set of data to 2·8 Å was collected and used to calculate a three-dimensional map of the protein, using the phases obtained from the *X*-ray study. Although these phases are only approximately correct for the neutron study, the molecule was nevertheless well defined. It remains to be seen if neutron diffraction can play a part in protein structure investigations; its role (if any) would seem to lie not in the determination of protein structure as such, where *X*-rays have many advantages, but rather in the precise definition of certain features of proteins, particularly the positions of hydrogen atoms.

B. Results.—*Insulin.* The structure of pig insulin has been determined at 2·8 Å by isomorphous replacement using five derivatives.[22] Each asymmetric unit of the rhombohedral unit cell contains two insulin molecules nearly but not exactly related by a non-crystallographic two-fold axis. The whole cell contains three such dimers to make up the insulin hexamer which is associated with two atoms of zinc.

The electron density map clearly shows the two insulin molecules in the asymmetric unit to be very similar but not quite identical, as shown in Figure 2. The 21-residue A chain is compact with the 30-residue B chain wrapped around it. The A chain contains two near-helical

[21] B. P. Shoenborn, *Nature*, 1969, **224**, 143.
[22] M. J. Adams, T. L. Blundell, E. J. Dodson, G. G. Dodson, M. Vijayan, E. N. Baker, M. M. Harding, D. C. Hodgkin, B. Rimmer, and S. Sheat, *Nature*, 1969, **224**, 491.

conformations involving residues A2—6 and A13—19. [The structure around residues A12—18 has in fact now been revised (personal communication from Professor Dorothy Hodgkin), but the conformation of the backbone as shown in Figures 2 and 3 is not greatly altered.] The loop

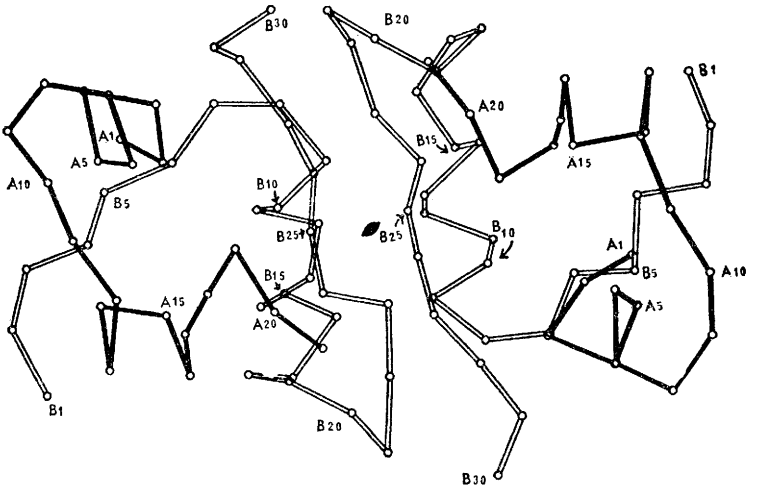


Figure 2 *View of the α-carbon positions in the peptide backbone of the insulin dimer, projected along the non-crystallographic two-fold axis. A chain black, B chain open* (Reproduced with permission from *Nature*, 1969, **224**, 491)

composed of residues A6—11 that are between the intrachain disulphide bridge lies with residues A8—10 (the three residues that vary most in mammalian species) on the surface of the molecule, while the bridge itself is internal. The A chain is stabilised by hydrogen bonds which link the two tyrosines A14 and A19 in the *C*-terminal half of the chain to residues A1 and A15 respectively in the *N*-terminal section. The central part of the B chain forms three turns of helix. The two interchain disulphide bridges B7—A7 and B19—A20 are located one at each end of this helix. The remaining residues of the B chain, B1—6 and B21—30, are largely extended and pack loosely around the A chain.

The contacts between insulin molecules to form the dimer involve only B chain residues. The closest contacts are between the extended sections B23—28 which are arranged antiparallel and form a hydrogen-bonded pleated sheet. These residues are arranged about the non-crystallographic two-fold axis which has been taken as the dimer two-fold axis. There is also an aromatic cage on the axis formed by Tyr-B26 and Phe-B24 and their partners related by the axis. The residue Phe-B25 in each of the two molecules also forms part of this system but can only do so by violating the two-fold symmetry. The arrangement causes neighbouring atoms to move in such a way that the exact two-fold symmetry is destroyed in their

vicinity. The dimer as a whole is an elongated cylinder 20 Å across and 40 Å long.

The insulin hexamer shown in Figure 3 (facing p. 84) is formed by the co-ordination of three insulin dimers around the two zinc atoms which are located 8·9 Å above and below the two-fold axis. Each zinc atom is linked to the three B10 histidine residues which, together with one water molecule, give the metal a very distorted octahedral co-ordination. Tyr-B16 and His-B5 may be linked to the zinc co-ordination sphere by a system of hydrogen-bonded water molecules. The three dimers interact in neighbouring pairs about a second set of two-fold axes. Important interactions around these axes appear to be formed by the phenylalanines B1. There is hydrogen-bond contact between the six glutamic acid residues, B13, which enclose a space around the crystallographic three-fold axis.

So far the structure sheds little light on the biological function of insulin, apart from the fact that certain residues that seem important for activity, for example Asn-A21, are on the surface.

Carboxypeptidase. Following the preliminary report of the structure of carboxypeptidase at 2 Å, Lipscomb and his colleagues [23] have now published a more detailed description of the molecule and have extended the analysis of some enzyme–substrate complexes to 2 Å. Bovine carboxypeptidase A is a zinc-containing enzyme of 34,600 molecular weight, which catalyses the hydrolysis of the *C*-terminal peptide bond of peptides and proteins, and does so most efficiently when the *C*-terminal residue is aromatic. The molecule is a single chain of 307 or 308 residues. Lipscomb had had only partial sequences of the molecule for the comparison with his Fourier maps which he reports in this paper, but now the complete sequence of the molecule has been reported.[24]

Perhaps the most interesting feature of the molecule (Figure 4) (after p. 84) is an extensive pleated sheet which, twisting through about 120°, runs through the centre of the molecule from top to bottom. The sheet contains four pairs of parallel and three pairs of antiparallel chains. The residues involved in this structure and its hydrogen-bond arrangement are shown in Figure 5 (after p. 84).

About 30% of the residues are contained in helical segments, nearly all of which lie on the surface of the molecule on one side of the pleated sheet and lined up approximately parallel to the chains of the sheet. There are four major helical residues, 14—29, 72—88, 215—233, and 288—305. Four other regions contain a few turns of helix of various degrees of perfection, including residues 94—103, 115—123, 174—184, and 254—262. Also α_{II} helices are observed at the *C*-terminal ends of three helices, residues 26—29, 100—103, and 260—262. On the other side of the pleated sheet is a

[23] W. N. Lipscomb, J. A. Hartsuck, G. N. Reeke, F. A. Quiocho, P. H. Bethge, M. L. Lubwig, T. A. Steitz, H. Muirhead, and J. C. Coppola, *Brookhaven Symp. Biol.*, 1968, **21**, 24; *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **64**, 28.

[24] R. A. Bradshaw, L. H. Ericsson, K. A. Walsh, and H. Neurath, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **63**, 1389.

tortuous chain possessing little secondary structure, which involves about 20% of the molecule (122—174 and 247—254). It is this chain that is involved in conformational changes on binding substrate.

Lying between the pleated sheet and the irregular coil is the deep cavity with a groove leading to it, which marks the active site. The essential zinc atom is located in the surface of this cavity linked to the protein by three side-chains, His-69, Glu-72, and Lys-196. At least one further ligand is present but it is not part of the enzyme molecule. In spite of chemical evidence which has indicated two —SH groups in the molecule, one forming a zinc ligand and the other a buried unreactive group, the crystallography shows clearly that both cysteines are involved in a disulphide bridge, residues 138—161, located in the irregular coil region of the molecule. The carboxylate of the *C*-terminal Asn-307 is linked to the guanidinium group of Arg-265 by an ion-pair bond.

The sequence of all 307 residues of carboxypeptidase has been tentatively assigned from the electron density after a preliminary exercise in fitting the known sequence of residues 1—22. Later, a further part of the sequence, residues 23—103, was obtained chemically. When they were compared with the *X*-ray sequence, 64% were found to be correct, 9% were the first choice among alternatives, in a further 17% the correct assignment occurred among several equally probable identifications, and only 10% appeared to be incorrect.

The main investigation of the active site of carboxypeptidase has been made on the unusually stable glycyltyrosine–enzyme complex. The dipeptide probably forms a non-productive competitive inhibitor complex. The extension of the resolution from 2·8 to 2·0 Å resolution, together with the more extensive analysis of the enzyme, has allowed a more detailed interpretation of the complex to be made. The *C*-terminal tyrosine residue is located in the large pocket which does not appear to contain any specific binding groups. This is in agreement with the enzyme's moderate but not high specificity for the *C*-terminal residue. The terminal carboxy-group interacts with the guanidinium group of Arg-145, which moves *ca.* 2 Å towards the inhibitor to make the contact. Although the peptide group of the inhibitor cannot be precisely located, presumably because it displaces water molecules that occupy the active site of the native enzyme, its most probable position is with the carbonyl oxygen acting as the fourth ligand of the integral zinc atom. The *N*-terminus of the inhibitor is linked to Glu-270 through a water molecule. This interaction may account for the 1000-fold decrease in the rate of hydrolysis of dipeptides having a free amino-group compared with the rate for other peptides, since the interaction with Glu-270 may prevent the acid from fulfilling its true catalytic function.

A number of conformational changes occur on inhibitor binding. The phenolic hydroxy-group of Tyr-248 moves *ca.* 12 Å, by means of a rotation of 120° about its C_α—C_β bond and a motion of the polypeptide backbone. This movement brings the hydroxy-group into the vicinity of the peptide

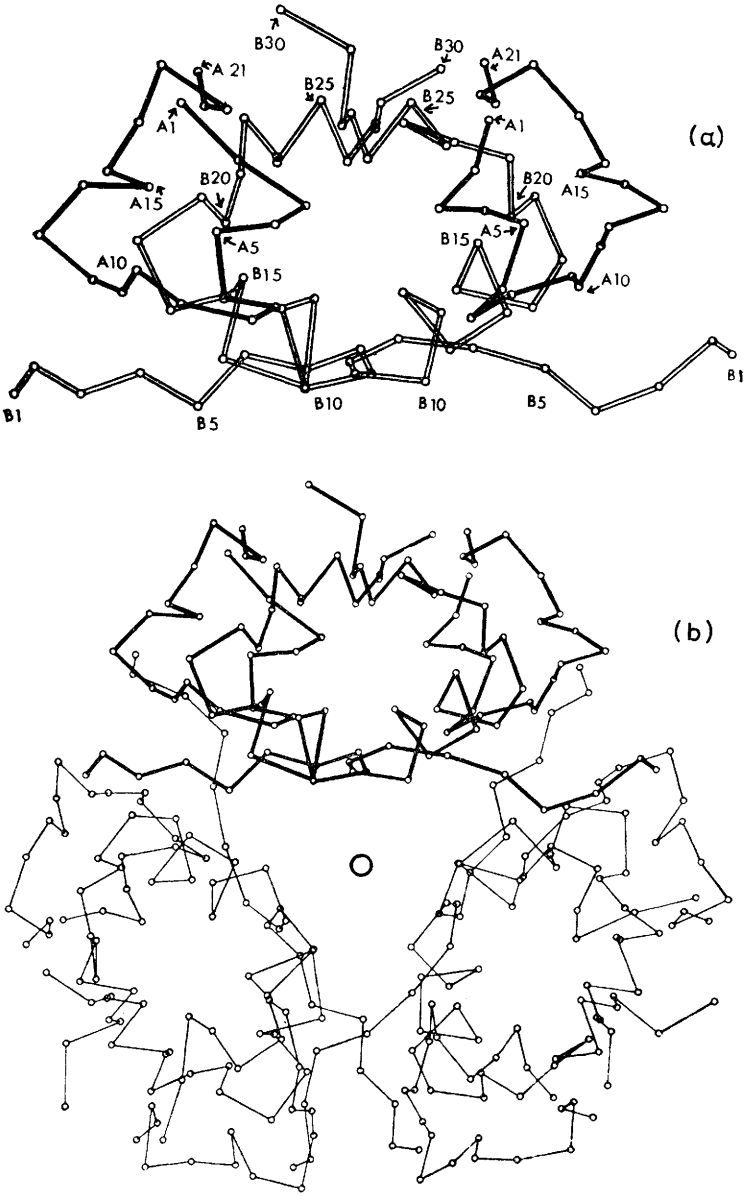


Figure 3 (*a*) *View of the insulin dimer projected along the crystallographic three-fold axis, showing* α*-carbons and backbone only.* (*b*) *View of the insulin hexamer projected down the three-fold axis. One dimer is in bold lines*
(Reproduced with permission from *Nature*, 1969, **224**, 491)

[*Facing p. 84*

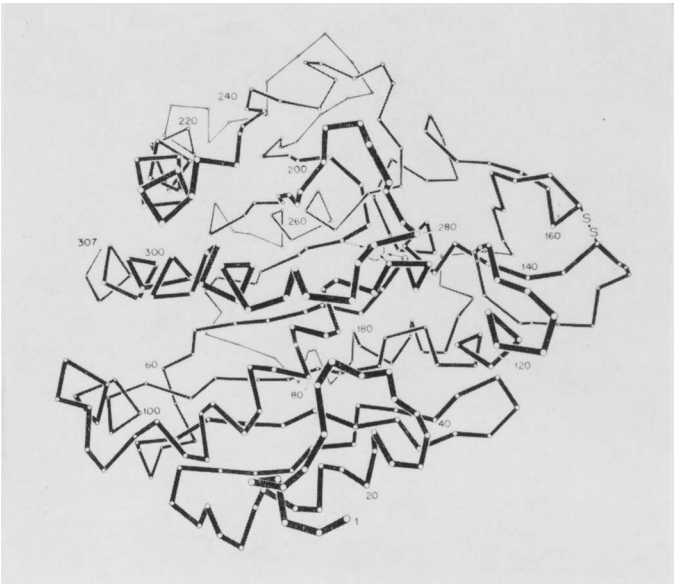


Figure 4 *The polypeptide chain of carboxypeptidase*
(Reproduced with permission from *Brookhaven Symp. Biol.*, 1968, **21**, 24)

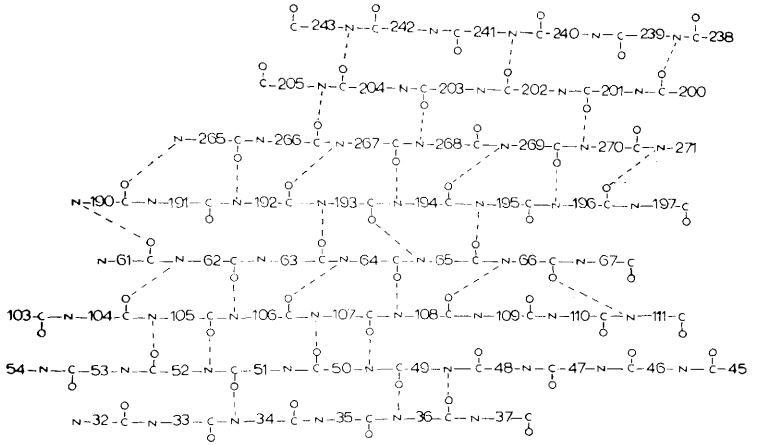


Figure 5 *The parallel and antiparallel pleated sheet structure of carboxypeptidase. Only the most certain hydrogen bonds are shown by broken lines*
(Reproduced with permission from *Brookhaven Symp. Biol.*, 1968, **21**, 24)

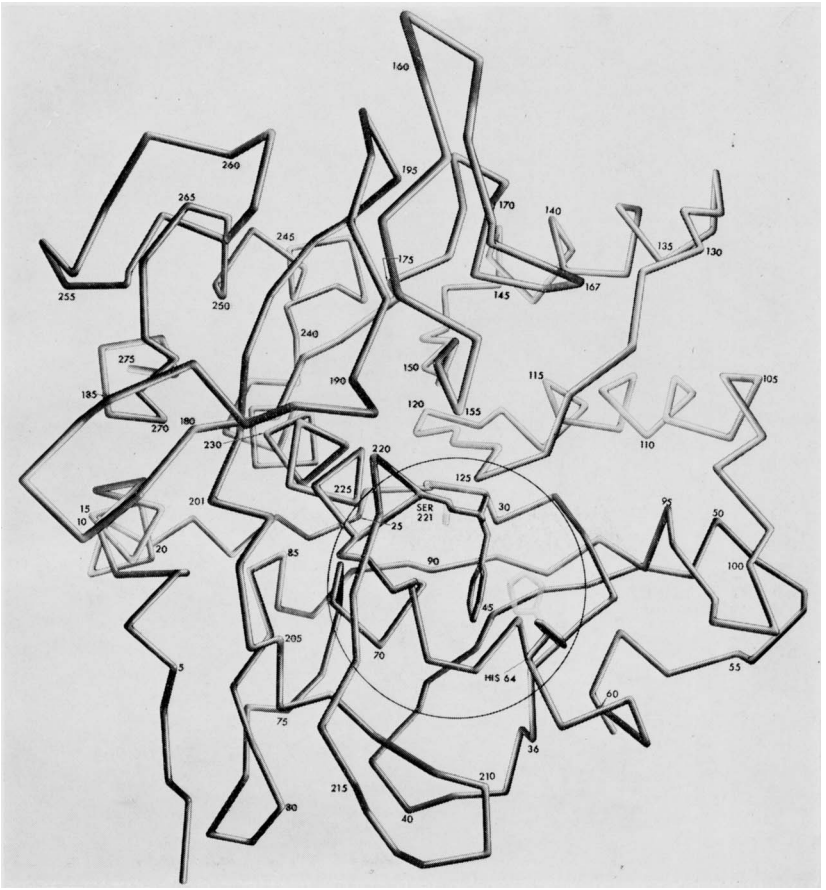


Figure 7 *The backbone of PMS-subtilisin BPN. The vertices represent the α-carbon positions. The active site lies within the circled area. The side-chain orientations of His-64 and Ser-221 in the native enzyme are shown in pale outline*
(Reproduced with permission from *Nature*, 1969, **221**, 235)

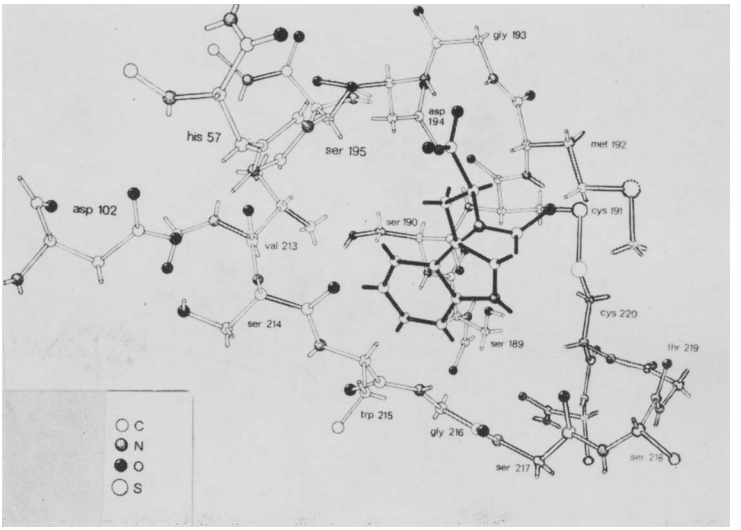


Figure 8 *The active site region of* α*-chymotrypsin showing the position occupied by formyl-*L*-tryptophan*
(Reproduced with permission from *J. Mol. Biol.*, 1969, **46**, 337)

bond of the inhibitor. In the native enzyme Arg-145 and Tyr-248 are linked by a system of hydrogen bonds. When glycyltyrosine is bound, this system is broken up and the two amino-acid residues move in a co-ordinated manner to bring a group (Tyr-248) required for catalysis into contact with the substrate. This provides a clear example of 'induced fit'.

As is usual, no direct evidence of the catalytic event can be obtained from *X*-ray analysis, and Lipscomb and his colleagues have been forced to extrapolate the results obtained from the inhibitor binding, by model building, and to take account of the known chemical properties in an attempt to deduce the mechanism of catalysis of carboxypeptidase. All the groups involved in binding glycyltyrosine have been assumed to occupy the same sites when true substrates are bound, with the exception of Glu-270 which no longer forms the link to the amino-terminus through a water molecule when the substrate is larger than a dipeptide. Nevertheless, Glu-270 will still be very close to the carbonyl carbon of the hydrolysable bond.

It has been assumed that longer chain substrates will be located in the groove leading to the enzyme pocket as well as in the pocket itself. This is shown schematically in Figure 6. Interactions in the groove can occur between the carbonyl groups of residues S_3, and perhaps S_4, and the guanidinium group of Arg-71 and the side-chains of S_3, and perhaps S_2 and S_4 associate with Tyr-198 and His (Phe)-279.

A variety of mechanisms has been examined and found to be possible; indeed it is not impossible that different mechanisms are used under different conditions. However, one mechanism seems at the present state of knowledge to be slightly favoured; the details of which are as follows (see Figure 6). A polypeptide substrate is bound to the enzyme with its *C*-terminal residue in the pocket and with its terminal carboxylate bound to Arg-145. The adjacent residues are bound in the groove as described above. The oxygen of the carbonyl of the susceptible peptide bond is co-ordinated to the zinc, while the phenolic hydroxy-group of Tyr-248, after its conformational shift, is able to hydrogen bond to the —NH of the susceptible peptide. At the same time it is able to hydrogen bond to the penultimate peptide bond. Glu-270 also takes part in the conformational changes on substrate binding, because to preserve normal contacts it has to move 2 Å away from the susceptible peptide bond. It is assumed that, in binding to the enzyme, the hydrolysable peptide bond is strained in making the interactions detailed above. Cleavage of the peptide bond is brought about by Tyr-248 donating a proton to the NH of the peptide bond (the resulting phenoxide is stabilised by the hydrogen it accepts). Glu-270 then forms an anhydride with the carbon of the polarised carbonyl group; a real alternative here is that Glu-270 promotes the attack of a water molecule on the carbonyl carbon. These concerted reactions cause the peptide bond to split, the anhydride is cleaved by a water molecule facilitated by the neighbouring phenoxide and the polypeptide chain, and the newly formed *C*-terminal amino-acid moves away. At the very least this proposed mechanism should

suggest suitable chemical tests which should lead either to the establishment of this mechanism or to a rapid elucidation of the correct mechanism.

Subtilisin. Subtilisin BPN′ is an extracellular protease produced by the soil organism *Bacillus amyloliquefaciens* (formerly thought to be a strain of

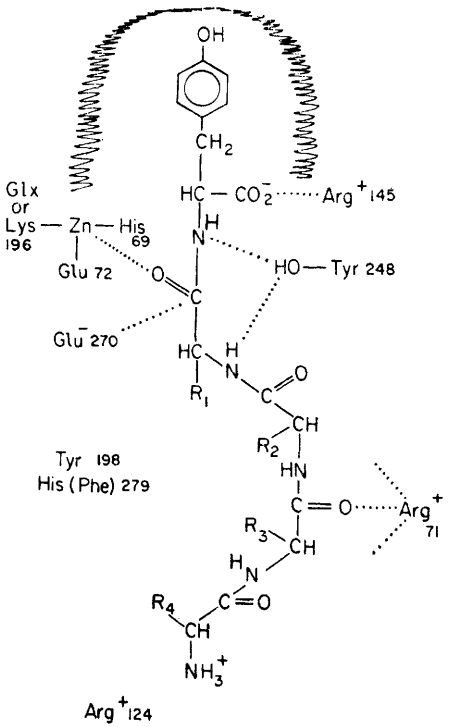


Figure 6 *The carboxypeptidase–substrate complex in the region of the active site. The plane of the susceptible peptide bond is perpendicular to that of the drawing* (Reproduced with permission from *Brookhaven Symp. Biol.*, 1968, **21**, 24)

B. subtilis). The enzyme has two notable features. The first is that the sequence of the single chain of 275 amino-acid residues contains a number of repeating homologous segments. The second is its relation to chymotrypsin. Both enzymes contain a highly reactive serine that can be phosphorylated, sulphonated, and acylated by such specific reagents as phenylmethanesulphonyl fluoride (PMSF), di-isopropyl phosphofluoridate and *trans*-cinnamoylimidazole. The active sites of both enzymes also contain a histidine residue. Further, both are reactive against synthetic small molecule substrates such as *N*-acetyl-L-tyrosine ethyl ester and *N*-benzoyl-L-tyrosine ethyl ester. This suggests quite strongly that the active sites of the two enzymes are essentially similar, as are their mechanisms of catalysis.

However, in spite of these similarities, the amino-acid sequences, including the sequences around the reactive serines of the two enzymes, are entirely unrelated. The *X*-ray analysis of subtilisin [25] has revealed the enzyme's unusual relationship to chymotrypsin and has shed some light on its own evolutionary history.

Subtilisin crystallises from 1M-ammonium sulphate at pH 5·8—6·0 in the monoclinic space group $C2$ with $a = 66{\cdot}7$ Å, $b = 54{\cdot}4$ Å, $c = 62{\cdot}9$ Å, and $\beta = 91{\cdot}9°$. There is one molecule in the asymmetric unit. The parent protein chosen was PMS-subtilisin since it was more stable than the native enzyme, and more isomorphous with the heavy-atom derivative in which the reactive serine was labelled with *p*-(carboxymethylmercaptomercuri)-benzenesulphonyl fluoride. Other derivatives were prepared by soaking the crystals in $6{\cdot}5 \times 10^{-4}$M K_2PtCl_4, $2{\cdot}5 \times 10^{-2}$M mersalyl (2), and iodinating some of the tyrosines with $9{\cdot}2 \times 10^{-4}$M KI_3.

The molecule, shown in Figure 7, (after p. 84) is approximately spherical with an overall diameter of 42 Å. There is no especially pronounced cleft or depression at the active site, which was labelled with the PMS molecule. The polypeptide chain contains eight helical residues defined as follows:

Helix *A* Residues 5—10	Helix *E* Residues 132—145
Helix *B* Residues 14—20	Helix *F* Residues 223—238
Helix *C* Residues 64—73	Helix *G* Residues 242—252
Helix *D* Residues 103—117	Helix *H* Residues 269—275

With the exception of helix *G*, all the helices lie within 15° of a common direction. The longest helix *F* runs approximately through the centre of the molecule. In addition to these helical sections there is an extensive twisted pleated sheet structure. Residues 148—152, 120—124, 28—32, 89—94, and 45—50 form a parallel hydrogen-bonded sheet, the bottom chain of which runs at approximately 45° to the top chain. The side-chains of these segments form a large part of the hydrophobic core of the molecule. It is very interesting to observe that this apparently important structural element is also present in carboxypeptidase and appears to fulfil the same function. It may be significant that these are the two largest single-chain protein molecules so far analysed and it is possible that other molecules that contain polypeptide chains of molecular weight greater than 30,000 will contain a large pleated sheet as an important (or indeed the most important) structural element. The overall folding of the backbone chain of subtilisin appears to have occurred in three distinct pieces, residues 1—100, 100—176, and 176—275.

In order to determine the structure of the active site of the native enzyme a difference Fourier map was calculated between the parent PMS-subtilisin and the native enzyme. This showed that the sulphonation of the

[25] C. S. Wright, R. A. Alden, and J. Kraut, *Nature*, 1969, **221**, 235.

active serine, residue 221, by PMSF causes the neighbouring histidine, residue 64, to rotate some 80°, without any appreciable movement of the backbone, from the position it occupies in the native enzyme. It also suggested that in its native conformation the side-chain of Ser-221 is hydrogen-bonded to the $N_{\epsilon 2}$ atom of His-64. It is extremely interesting in view of the structure of the active site of α-chymotrypsin that the carboxy-group of Asp-32 is within 3 Å of the $N_{\delta 1}$ atom of His-64 and could be hydrogen-bonded to it. This structural information on the active site of subtilisin taken together with the chemical evidence of the enzyme's homology with α-chymotrypsin provides the first documented example of the convergent evolution of two distinct enzyme molecules towards a common active site and catalytic mechanism.

Two of the segments of repeating sequences in the molecule, residues 67—75, His-Val-Ala-Gly-Thr-Val-Ala-Ala-Leu, and residues 226—233, His-Val-Ala-Gly-Ala-Ala-Ala-Leu, are found in the *C*-terminal portions of helices *C* and *F* respectively. Both Ser-221 and His-64 are close to where the three parts of the chain come close together, and each is in one of the *N*-terminal sections of the two helices containing the repeating sequences. It seems that these observations are related in some way to the evolutionary history of the subtilisins.

Chymotrypsin. The success of the structure determination of α-chymotrypsin has enabled Blow and his colleagues [26] to investigate the active site of the enzyme. In this case most of the data have been obtained not from enzyme–competitive inhibitor complexes but from a complex of the enzyme with *N*-formyl-L-tryptophan, a virtual substrate of the enzyme since the substrate is turned over very slowly indeed. In the usual way 2·5 Å resolution data were collected from the enzyme–substrate complex and a three-dimensional difference map was calculated.

The location of the substrate on the α-chymotrypsin molecule is shown in Figure 8. (facing p. 85) No alteration in the position of backbone atoms takes place on binding, apart from a shift in the position of Tyr-146 in the molecule related by the non-crystallographic dyad, and this is undoubtedly an artefact of the particular crystal form. The indole ring of the substrate fits into an irregular pocket near Ser-195. The dimensions of this pocket are *ca.* 10—12 Å by 5·5—6·5 Å by 3·5—4·0 Å. It is long enough to accommodate tryptophan and deep enough for tyrosine. The narrowness of the pocket allows only one orientation for the plane of the bound aromatic residue. The most prominent interactions between the indole and the protein atoms in the pocket involve peptide bonds whose planes are parallel to the plane of the indole ring and *ca.* 3·5—4·0 Å from it. They include peptide bonds 190—192 and 215—216. The side-chain of Met-192 appears to function as a flexible hydrophobic lid to the pocket. The amide hydrogen of the substrate points towards the carbonyl oxygen of Ser-214 but is at greater than hydrogen-bonding distance. The carbon of the carboxy-group of the

[26] T. A. Steitz, R. Henderson, and D. M. Blow, *J. Mol. Biol.*, 1969, **46**, 337.

substrate is close to the hydroxy-group of Ser-195 and one of the carboxy-oxygens appears to be hydrogen-bonded to Ser-195 and the peptides of Gly-193 and Ser-195, while the other oxygen is hydrogen-bonded to His-57. *X*-Ray analysis of the complex with *N*-formyl-L-phenylalanine indicates that the binding of this molecule is very similar to that of its tryptophan analogue.

The extension of the binding of the virtual substrate to that of the real polypeptide substrates has been complicated by two factors. The first is the interference of the hydroxy-group of Tyr-146 from a neighbouring molecule which is close enough to interact with the amide group of the substrate. The second is that the pH at which the virtual substrate is bound is nearly two units below the optimum pH of the enzyme, 7·5. At pH 5·7 the active site is protonated while the carboxy-group of the substrate is unprotonated.

It is believed that in the productive complex the substrate's aromatic side-chain is located in the pocket, as shown in Figure 8, but that the NH of the aromatic side-chain is moved slightly so that it can form a hydrogen bond with the carbonyl oxygen of Ser-214. On the basis of model building, the NH of the susceptible peptide bond occupies the position of the carboxy-oxygen of the virtual substrate that is near His-57. This places the carbonyl carbon of the susceptible bond in a position for nucleophilic attack by the oxygen of Ser-195, and places the NH of the susceptible bond favourably for protonation by the proton shared by His-57 and Ser-195. This is exactly the structural arrangement required by the 'charge relay' mechanism proposed earlier by Blow and his colleagues.

Information on the structure of the acyl-chymotrypsin intermediate has been provided by the high resolution *X*-ray structure of tosyl-chymotrypsin. This showed that some repositioning of the side-chains of His-57 and Ser-195 takes place on tosylation of the enzyme. The oxygen of Ser-195 rotates about its C_α—C_β bond from the 'up' position shown in Figure 8 to the 'down' position. His-57 moves out slightly to allow a water molecule to bind between its $N_{\varepsilon 2}$ and one of the sulphonyl oxygens of the tosyl group. It is possible for the same conformational changes and water binding to take place in the acylated enzyme, while the aromatic side-chain remains bound in the pocket and maintains its hydrogen bond to Ser-214.

The deacylation step can proceed by a reversal of the acylation mechanism. Model-building shows that the water molecule can be placed between His-57 and the acyl group so that when it is hydrogen-bonded to the $N_{\varepsilon 2}$ of the histidine its lone pair is pointing towards the carbonyl carbon of the acyl group. The water molecule replaces the serine hydroxy-group in the charge relay system and can attack the carbonyl carbon of the acyl group. This model suggests that the stability of sulphonyl- and phosphoryl-chymotrypsin results from the steric exclusion of this activated water molecule from the vicinity of the atom to be attacked.

The definition of the substrate binding site of α-chymotrypsin allows the specificity of the homologous enzymes trypsin and elastase to be explained. The amino-acid sequence of trypsin shows two differences with respect to chymotrypsin. One is the replacement of Met-192 by Gln, and the second is the exchange of Ser-189, located deep in the pocket, to Asp (177 in the trypsin sequence). This latter change is, in itself, enough to explain the specificity of trypsin towards basic side-chains. Asp-177 is in such a position that, assuming the substrate binding to be analogous to that in chymotrypsin, it can form an ion pair with the guanidinium of arginine or with the ε-amino of lysine residues bound in the pocket.

In elastase the pocket is blocked by two amino-acid replacements as compared with chymotrypsin. Gly-216 is replaced by valine, which blocks the entrance to the pocket, and Gly-226 is threonine in elastase and also fills the pocket. This explains why elastase does not hydrolyse the same substrates as chymotrypsin and trypsin.

The α-chymotrypsin molecule can undergo a slow reversible pH-dependent transformation into γ-chymotrypsin. Crystals of this modification can be obtained at pH 5·6 from ammonium sulphate solutions. The structure of γ-chymotrypsin in these crystals has now been obtained at 5·5 Å resolution.[27] Three of the four derivatives used were obtained by chemically modifying the enzyme; *p*-iodobenzenesulphonyl fluoride and *p*-methoxy-*m*-chloromercuri-benzylsulphonyl fluoride were used to label the reactive serine, and two tyrosine residues were iodinated with $2{\cdot}5 \times 10^{-3}$M I_2 in 10^{-2}M KI. The resulting Fourier map of the enzyme could not be directly interpreted but it was found that the electron density could be transformed into the α-chymotrypsin cell to give a very satisfactory fit. This suggests that the α- and γ-chymotrypsin molecules are essentially the same. However, some differences of behaviour are found on tosylation of the active serine. A difference Fourier at 5·5 Å resolution between the native and tosylated γ-form of the enzyme shows the expected peak at the sulphonyl position of the tosyl group, while nearby there is a hole and a smaller peak. These changes are not the same as those observed under the same conditions in α-chymotrypsin and unfortunately cannot be interpreted at this resolution.

Staphylococcal Nuclease. The structure of the calcium-activated extracellular nuclease of *Staphylococcus aureus* has been obtained at 4 Å resolution.[28] The enzyme attacks the phosphodiester bonds of both RNA and DNA. The molecule is composed of a single chain of 149 residues, corresponding to a molecular weight of 16,800, without either disulphide bridges or sulphydryl groups. Crystals were obtained from 29—32% 2-methyl-2,4-pentandiol at pH 8·6—8·7 in phosphate buffer. The crystals

[27] G. H. Cohen, E. W. Silverton, B. W. Matthews, H. Braxton, and D. R. Davies, *J. Mol. Biol.*, 1969, **44**, 129.

[28] A. Arnone, C. J. Bier, F. A. Cotton, E. E. Hazen, D. C. Richardson, and J. S. Richardson, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **64**, 420.

are tetragonal, space group $P4_1$ with $a = b = 47{\cdot}7$ Å and $c = 63{\cdot}5$ Å. Derivatives were prepared by soaking the crystals in (i) 1×10^{-4}M K_2PtCl_4, (ii) 1×10^{-4}M *p*-acetoxymercuri-aniline, and (iii) 3×10^{-4}M *p*-chloromercuribenzenesulphonic acid. In the resulting Fourier map calculated at 4 Å resolution it could be seen that the molecule contained about 15% α-helix in one run of $3\frac{1}{2}$ turns and a second less clear section of 2 or 3 turns. There is also a pleated sheet with three antiparallel sections. The path of the backbone can be traced for about two thirds of its length but in some places its direction is unclear.

Crystals of the nuclease grown in the presence of equimolar concentrations of a potent inhibitor, thymidine-3,5-diphosphate, and two moles per mole of calcium ion, were found to be not strictly isomorphous with the native enzyme. Therefore, isomorphous crystals were obtained from solutions of the heavy-atom-labelled inhibitor 5-iododeoxyuridine-3,5-diphosphate and used to calculate an independent map of the inhibited nuclease by the combined isomorphous and anomalous probability distribution from the single iodine atom. This map showed that the molecule was generally very similar in its native and inhibited forms. The inhibitor molecule was clearly seen located in a well-defined pocket in the enzyme's surface, with its 5′-phosphate near the calcium ion. The pyrimidine ring is close to and parallel with a tyrosine residue, which undergoes a rotation of *ca.* 90° from the position it occupies in the native enzyme. Several other small, less well-defined, conformational changes occur near the active site.

Lactate Dehydrogenase. The structure of lactate dehydrogenase is the first report,[29] strictly speaking, of the structure of an intracellular enzyme and is by a considerable factor the largest molecule to be analysed by *X*-ray methods. However, because of the very advantageous symmetry of the crystals the crystallographic unit represents only a single subunit with a fairly modest molecular weight of 35,000.

The M_4 isozyme of dogfish lactate dehydrogenase crystallises in the tetragonal system. The space group has been termed for convenience *F*422 with $a = 146{\cdot}9$ and $c = 155{\cdot}1$, but is more conventionally *I*422. The asymmetric unit contains one quarter of the molecule, that is a single subunit, which immediately indicates that the four subunits of lactate dehydrogenase are identical. The molecule contains four thiols which can react with *p*-chloromercuribenzoate without loss of activity. Four further thiols can be caused to react but this results in rapid inactivation. The non-essential thiol provides the locus of three derivatives; *p*-chloromercuribenzoate itself and the two dimercurials, dimercuriacetic acid and 'Baker's mercurial'. This thiol also forms sites of low occupancy with $AuCl_4^-$ and $Pt(NH_2 \cdot CH_2 \cdot CH_2 \cdot NH_2)_2Cl_2$. The gold compound has a second site also

[29] M. J. Adams, D. J. Haas, B. A. Jefferey, A. McPherson, H. L. Mermall, M. G. Rossmann, R. W. Schevitz, and A. J. Wonacott, *J. Mol. Biol.*, 1969, **41**, 159.

of low occupancy at the essential thiol, while the platinum compound has a highly occupied site elsewhere. These five heavy-atom derivatives have been used to calculate an electron density map of the enzyme at 5 Å resolution.

The map indicated that the molecule is centred on a 222 symmetry point in the space group and thus exhibits tetrahedral symmetry. There are liquid regions 10 Å in diameter along the two-fold axes which result in a large cavity in the centre of the molecule. The individual subunits show one remarkable feature in a large groove which divides the subunit into two sections joined by a narrow 'neck'. The essential thiol is found in the vicinity of the 'neck' while the non-essential thiol is situated very close to the surface between subunits. Several regions of continuous polypeptide chain can be seen but with evidence for only a few turns of helix, and in general the course of the chain cannot be followed for any distance.

At this resolution considerably more interesting results have been obtained from a study of the binding of the co-factor nicotinamide-adenine dinucleotide (NAD). When the co-factor itself (or its reduced form) is added to the crystals a change in space group occurs with only a very small change in one of the cell dimensions. The space group changes from $I422$ to the very closely related $P4_22_12$. The essential point is that the molecular symmetry expressed by the space group is reduced from 222 to 2. The change is probably caused by either a small shift in the subunits so that the molecular symmetry is reduced or a rotation of the molecules as a whole about the four-fold axis so that they maintain their tetrahedral symmetry, but the space group ceases to express it. Examination of the changes on co-factor binding by Patterson and Fourier methods is consistent with a translation of the subunits by about 5 Å.

The effect of a large number of the constituent parts of the co-factor on the enzyme crystals has also been examined. Of the purine end of NAD, adenosine binds but the symmetry is not altered, while both adenosine monophosphate and adenosine diphosphate alter the symmetry but are less effective than NAD. As examples of the pyridine end of the co-factor, nicotinamide binds without changing the space group while no binding at all is observed for nicotinamide mononucleotide. No change in symmetry is observed for the analogues of the central part of the co-factor, ribose, ribose phosphate, and pyrophosphate. In general there appears to be a correlation between those compounds that bind or change the symmetry and those that cause competitive inhibition.

A Fourier map of the adenosine–enzyme complex indicated only partial binding of the adenosine. The purine ring was found to be located near the non-essential thiol. Since adenosine is a competitive inhibitor this suggests that the active site is near the non-essential thiol. On the other hand, in the presence of NAD, mercuric chloride is unable to react with either thiol and conversely NAD is unable to bring about the conformational change if the essential thiol has been mercuriated. Clearly this system needs further investigation.

An abortive ternary complex of the enzyme can be formed by the addition of the oxidized co-factor and oxidized product, pyruvate. Crystals of this complex were found [30] to belong to yet another tetragonal space group, $P422$ with $a = 95{\cdot}4$ Å and $c = 86{\cdot}1$ Å. The subunits of the enzyme are probably centred in this space group on 222 symmetry points, and thus the subunit arrangement has the same symmetry as the apo-enzyme. The proposed obligatory order of binding appears to be associated with changes in the quaternary structure as follows:

$$(E)_{222} + C + S \longrightarrow (EC)_2 \longrightarrow (ECS)_{222},$$

where E = enzyme,

C = co-factor,

S = substrate,

and the subscripts represent the point symmetry of the molecule.

The diffraction pattern of the ternary complex shows only some general features in common with the apo-enzyme and is otherwise rather different.

Haem Proteins. The study of the erythrocruorin (insect haemoglobin) of *Chironomus thummii* has now resulted in a Fourier map at 2·8 Å.[31] The molecule is essentially similar in tertiary structure to myoglobin and the individual subunits of mammalian haemoglobins. Some differences do exist; in particular the helix equivalent to the *H* helix of myoglobin and haemoglobin has a pronounced kink which causes it to deviate by *ca.* 35° from the linear structure found in the other two proteins. Only the environment of the haem group has so far been reported from the high resolution map. In general its situation in the erythrocruorin molecule is equivalent to that in myoglobin and haemoglobin. The haem is located in a hydrophobic pocket between helices *E* and *F*, with its iron atom linked to the protein through a histidine, and with a second histidine hydrogen-bonded to the haem-linked water molecule on its distal side. There are differences however, most notable amongst which is the much higher proportion of aromatic residues, mostly phenylalanines, in contact with the haem group. Some of these additional contacts are made by the side-chains of the *H* helix, which because of the kink gets closer to the haem than it does in myoglobin and haemoglobin.

A report of a 5 Å resolution Fourier map of the myoglobin of the common seal [32] shows that, as expected, its tertiary structure is very similar to that of sperm whale myoglobin. Seal myoglobin has 25 amino-acid substitutions in the 153-residue polypeptide chain as compared with the sperm whale molecule. The substitutions are found to be fairly evenly distributed along the chain, apart from a concentration in the short *D* helix.

[30] R. Leberman, I. E. Smiley, D. J. Haas, and M. G. Rossmann, *J. Mol. Biol.*, 1969, **46**, 217.
[31] R. Huber, O. Epp, and H. Formanek, *J. Mol. Biol.*, 1969, **42**, 591.
[32] H. Scouloudi, *J. Mol. Biol.*, 1969, **40**, 353.

At this resolution the effect of the amino-acid substitutions cannot be very fully interpreted. The two heavy-atom derivatives used for phasing the seal myoglobin map are, apart from one minor site, found to occupy the same sites as they do in the sperm whale molecule. The $AuCl_4^-$ ion is bound in the vicinity of histidines B5 and GH1 in sperm whale myoglobin, and is found at the equivalent site in the seal molecule, in which neither of these histidines is substituted. The main binding site of the HgI_3^- ion is close to the haem group in both myoglobins.

Preliminary Data for Other Proteins. Preliminary crystallographic data have been reported for a number of proteins. In favourable cases useful structural and functional information has been deduced from such data.

(*a*) *Aldolase.*[33] Rabbit muscle aldolase has been crystallised in two distinct forms. Crystals grown from about 50% saturated ammonium sulphate at pH 6·0 were hexagonal, space group $P6_222$ with $a = 120{\cdot}6$ Å and $c = 169{\cdot}0$ Å. Estimates of the protein content of these crystals were consistent with half a molecule in the asymmetric unit with the molecule situated on a crystallographic two-fold axis. The hexagonal crystals are very susceptible to radiation damage. Aldolase crystals grown under similar conditions become monoclinic when the pH is increased to 7·3. The space group is $P2_1$ and the cell dimensions of $a = 164{\cdot}5$ Å, $b = 57{\cdot}3$ Å, $c = 85{\cdot}0$ Å, and $\beta = 102° 40'$ are consistent with one molecule per asymmetric unit. The diffraction pattern gives some indication of molecular symmetry in the form of a mirror plane of symmetry in the *h0l* zone normal to the c^* axis. Further indications of molecular symmetry were obtained from a three-dimensional Patterson synthesis calculated at 8·5 Å resolution. The Patterson contained only one significant peak corresponding to an accumulation of inter-subunit vectors consistent with a non-crystallographic two-fold axis parallel to the b^* axis. These data combined with results obtained from the use of the rotation function indicate that the aldolase molecule is a tetramer possessing a fair degree of 222 symmetry. Similar monoclinic crystals have been obtained after reacting four thiols per molecule with *p*-chloromercuribenzoate, but these crystals have cell dimensions which differ from the native values by as much as 4 Å. This seems to suggest that the molecule undergoes some kind of conformational change when the thiols are reacted.

(*b*) *Asparaginase.*[34] The L-asparaginase of *Erwinia cartovora* has been crystallised from ethanol–water mixtures at pH 9. The crystals are orthorhombic, space group $P2_12_12_1$ with $a = 91{\cdot}8$ Å, $b = 177{\cdot}1$ Å, and $c = 78{\cdot}4$ Å. The molecular weight obtained by ultracentrifuge studies is 128,000—142,000 while the molecular weight in the asymmetric unit of the crystals appears to be in the range 116,000—140,000. No molecular symmetry is expressed in the space group.

[33] P. A. M. Eagles, L. N. Johnson, M. A. Joynson, C. H. McMurray, and H. Gutfreund, *J. Mol. Biol.*, 1969, **45**, 533.

[34] A. C. T. North, H. E. Wade, and K. A. Cammack, *Nature*, 1969, **224**, 594.

(*c*) *Cytochrome* b_5.[35] Calf liver cytochrome b_5 has a molecular weight of 10,280 and contains a non-covalently bound haem group. Crystals grown from phosphate at pH 7·5 are orthorhombic, space group $P2_12_12_1$ with $a = 64{\cdot}55$ Å, $b = 46{\cdot}01$ Å, $c = 29{\cdot}89$ Å, and contain one molecule per asymmetric unit.

(*d*) *Cytochrome* c *peroxidase*.[20] Crystals of the yeast enzyme grown from 2-methyl-2,4-pentanediol buffered with 0·05M phosphate are orthorhombic, space group $P2_12_12_1$ with cell dimensions $a = 108{\cdot}0$ Å, $b = 77{\cdot}8$ Å, and $c = 51{\cdot}4$ Å. There appears to be one molecule of *ca.* 40,000 molecular weight in the asymmetric unit.

(*e*) *Flavodoxin*.[36] Flavodoxin from *Clostridium pasteurianum* is an electron-transfer flavoprotein of molecular weight *ca.* 15,000 containing a single flavin mononucleotide. Crystals of the protein were grown from 65% ammonium sulphate at pH 6·4—6·8. Under anaerobic conditions in the presence of ethylenediaminetetra-acetic acid, light reduces the flavin to a neutral semiquinone. Deep-red crystals of the reduced derivative were grown as before but under a nitrogen atmosphere. Both kinds of crystal were trigonal, of space group either $P3_121$ or $P3_221$ with $a = 61{\cdot}7$ Å and $\gamma = 120°$; the length of the *c*-axis, however, depends on the state of the flavin, the oxidised form being 70·4 Å while the reduced form is 70·9 Å. There is one molecule in the asymmetric unit. The intensity patterns of the oxidised and reduced forms differ to some extent, which suggests that the conformation of the protein is dependent on the oxidation state of the flavin.

(*f*) *Human lysozyme*.[37] Large quantities of lysozyme can be obtained from the urine of patients suffering from some types of leukemia. Lysozyme obtained in this way appears to be identical with normal human lysozyme. There is some chemical evidence that human lysozyme may not have the expected homology with hen egg-white lysozyme. Crystals of the urinary enzyme grown from 5% protein solutions in 3M sodium chloride at pH 4·5 are orthorhombic, space group $P2_12_12_1$ with cell dimensions $a = 65{\cdot}3$ Å, $b = 110{\cdot}5$ Å, and $c = 43{\cdot}7$ Å. The asymmetric unit comprises two molecules. Crystals can be obtained in the same space group but with only one molecule in the asymmetric unit by lowering the protein concentration to 2%. These crystals have cell dimensions $a = 57{\cdot}1$ Å, $b = 61{\cdot}0$ Å, and $c = 33{\cdot}0$ Å when grown from either 3M sodium chloride or 7M ammonium nitrate. Neither of these crystal forms is similar to any of the crystal forms of hen egg-white lysozyme.

(*g*) *Myeloma protein Fab fragments*.[38] Poljak and his colleagues have crystallised two more Fab fragments of human myeloma proteins. Fab

[35] F. S. Matthews and P. Strittmatter, *J. Mol. Biol.*, 1969, **41**, 295.
[36] M. L. Ludwig, R. D. Andersen, S. G. Mayhew, and V. Massey, *J. Biol. Chem.*, 1969, **244**, 6047.
[37] E. F. Osserman, S. J. Cole, I. D. A. Swan, and C. C. F. Blake, *J. Mol. Biol.*, 1969, **46**, 211.
[38] R. L. Humphrey, H. P. Avey, L. N. Becka, R. J. Poljak, G. Rossi, T. K. Choi, and A. Nisonoff, *J. Mol. Biol.*, 1969, **43**, 223.

'Hil' crystallises in the orthorhombic space group $P2_12_12_1$ with $a = 111{\cdot}1$ Å, $b = 127{\cdot}8$ Å, and $c = 66{\cdot}1$ Å, while Fab 'Smo' crystallises in the trigonal space group $P3_121$ with $a = 120{\cdot}7$ Å and $c = 110{\cdot}6$ Å. Both crystals have two Fab molecules in the asymmetric unit. It is interesting to note that while the Fc fragments of the two major Gm allotypes crystallise isomorphously, the three Fab fragments so far examined crystallise in different space groups. This is probably a reflexion of the fact that the Fab fragments examined include the two major types of light chain, κ and λ, but is probably also related to the differences in sequence in the *N*-terminal half of the light chain and in the heavy chain (Fd) fragment. It is suggested that the molecular packing of the Fab fragments is different because of these differences in sequence, but there are probably extensive structural homologies in the fragment molecules as a whole.

(*h*) *Peptidase A.*[39] The protease, peptidase A, from *Penicillium janthinellum* resembles porcine pepsin in amino-acid composition, molecular weight, pH optimum, and specificity. The enzyme crystallises from 0·9% ammonium sulphate at pH 4·4 in the monoclinic space group *C*2 with $a = 97{\cdot}8$ Å, $b = 46{\cdot}7$ Å, $c = 65{\cdot}7$ Å, and $\beta = 115° 30'$. There is one molecule in the asymmetric unit.

(*i*) *α-Lytic protease.*[40] The soil bacillus *Sorangium* produces a protease that shows remarkable similarities with bovine α-chymotrypsin. The amino-acid sequence around the reactive serine is identical with the sequence around Ser-195 in chymotrypsin, while the sequence around the sole histidine in the protease molecule is similar to that around His-57 of chymotrypsin. However, the molecular weight of 20,000 appears to be significantly lower than chymotrypsin. The bacterial enzyme crystallises from 1·7M ammonium sulphate at pH 7·3 in the trigonal space group $P3_121$ (or $P3_112$) with $a = b = 66{\cdot}5$ Å, $c = 80{\cdot}2$ Å, and $\gamma = 120°$. There appears to be one molecule in the asymmetric unit.

PART III: Spectroscopic and Solution Studies on the Conformation and Interactions of Polypeptides and Proteins

by A. R. Peacocke, with contributions by R. Blagrove, J. R. Brocklehurst, D. G. Dalgleish, R. Henson, P. H. Lloyd, R. M. Stephens, and I. O. Walker

1 Introduction

by A. R. Peacocke

The scope of the report on 'secondary structures' in the first volume (pp. 112 following) has in this second volume been widened to include other aspects of the structure of polypeptides and proteins amenable to study by spectroscopy and by methods applicable to these macromolecules

[39] M. Camerman, T. Hoffman, S. Jones, and S. C. Nyburg, *J. Mol. Biol.*, 1969, **44**, 569.
[40] M. N. G. James and L. B. Smillie, *Nature*, 1969, **224**, 694.

in solution. These methods, applications of which are reported on in the succeeding sections of this chapter, have been selected principally on the basis of their prominence in the literature from the end of 1968 to the end of 1969, or a little further in a few instances. In particular, a very large number of circular dichroism investigations have been stimulated by the availability of accurate commercial measuring instruments and these studies have been surveyed in section 3. This report and the others in sections 1—5 cover papers characterised by the application of a particular spectroscopic or other technique.

What used to be called the 'tertiary' and 'quaternary' structure of proteins, the assembly of internally covalently linked units into structures of higher order, has been investigated in solution principally by employing agents which cause dissociation, and less frequently association, and by observing concomitant changes in macromolecular parameters, such as molecular weight. Studies of this kind which appeared in the period covered by the Report are surveyed in section 6. In section 7 attention is drawn to published developments in one particular dynamic technique, that of diffusion measurements, with respect to which there are already signs of interesting new developments possibly applicable to the study of association equilibria in proteins. Two very useful multi-author volumes have appeared which describe the application of physico-chemical techniques to the study of biological macromolecules, in general, and proteins in particular.[1, 2] A very valuable volume, whose fascination arises not least because of its reflection of the many facets of him whom it honours, is the set of short surveying essays dedicated to Professor L. Pauling: these touch on many of the problems which underlie the contents of this chapter and indeed of this whole Report.[3]

2 Infrared

contributed by R. M. Stephens

It has for several years been possible to characterise the conformation of polypeptides from their infrared spectra. This has involved the observation of the frequencies and dichroism of the fundamental vibrations of the peptide group, which occur in the range 3400—700 cm^{-1}, and comparison of the results either with the theoretical calculations for the vibrational frequencies of the various conformations or with the conformations obtained from other techniques. Advances in instrument design have allowed these conformational studies to be extended to the far-i.r. region, 700—50 cm^{-1}, where the absorption bands due to backbone skeletal vibrations appear.

[1] 'Physical Techniques in Biological Research,' vol. 2, Parts A and B, 2nd edn., ed. D. H. Moore on 'Physical Chemical Techniques,' Academic Press, New York and London, 1968 and 1969.

[2] 'Physical Principles and Techniques of Protein Chemistry,' Part A, ed. S. J. Leach, Academic Press, New York and London, 1969.

[3] 'Structural Chemistry and Molecular Biology', ed. A. Rich and N. Davidson, W. H. Freeman & Co., San Francisco, 1968.

An understanding of the vibrations which occur in this low frequency region could be useful in conformational studies on proteins whose amide vibrations in the i.r. region are generally masked by water absorption bands. The frequency of the amide V band, which results from bending of the N—H out of plane, and of its deuteriated analogue, have been analysed for various conformations of poly-γ-methyl-L-glutamate, poly-γ-ethyl-L-glutamate, poly-L-alanine, sodium poly-α-L-glutamate, and poly-γ-isoamyl-L-glutamate.[4] It was found that this absorption band occurs at about 610—620, 650, and 700—705 cm^{-1} for the α-helical, random chain, and antiparallel pleated sheet conformations, respectively. These bands occurred in the deuteriated analogues at 445—465, 510, and 515—530 cm^{-1} respectively.

The far-i.r. spectra of α-helical poly-L-α-amino-n-butyric acid, poly-L-norvaline, poly-L-norleucine, poly-L-leucine, and the pleated sheet forms of poly-L-α-amino-n-butyric acid, poly-L-valine, poly-DL-α-amino-n-butyric acid, poly-DL-norvaline, and poly-DL-norleucine have shown[5] that the α-helix gives rise to characteristic absorption bands at 690, 650, 610, 380, 150, and 100 cm^{-1} whereas the pleated sheet has characteristic bands at 700, 240, and 120 cm^{-1}. The interpretation of these absorption bands is difficult because the main-chain vibrations are strongly coupled to the side-chain deformation vibrations. The characteristic absorption bands at about 1600 and 600 cm^{-1} are especially useful for conformational diagnosis of DL polypeptides which do not exhibit optical dispersion characteristics. Samples of poly-γ-methyl-DL-glutamate have been synthesised by Leuch's method with varying anhydride to initiator (A/I) ratios.[6] For high A/I ratios the polymers were mainly α-helical whereas at low A/I ratios the conformations were a mixture of the random and β forms.

Only a little is known about the absolute intensities of the characteristic vibrations of the peptide group. From a study of the i.r. spectra of α-helical poly-γ-benzyl-L-glutamate, the absolute intensities and molecular absorption coefficients for the amide I, II, and V bands were determined and the electro-optical parameters of the peptide group vibrations calculated.[7]

The amide A region at about 3300 cm^{-1} is used to study the vibrations of the peptide N—H stretching band, particularly in solutions where the other amide bands may be obscured. Three model compounds have been examined, $MeCO \cdot NH \cdot CHR^1 \cdot CO \cdot NHR^2$, $MeCO \cdot N(Me) \cdot CH_2R^1$, and $MeCO \cdot NH \cdot CHR^1 \cdot CONHR^2$, in dilute carbon tetrachloride, and the N—H stretching bands have been characterised for two different chelated compounds.[8] The frequency of the unbonded N—H absorption band was dependent upon the substituent on the N atom.

[4] Y. Masuda, K. Fukushima, T. Fujii, and T. Miyazawa, *Biopolymers*, 1969, **8**, 91.
[5] K. Itoh, T. Shimanouchi, and M. Oya, *Biopolymers*, 1969, **7**, 649.
[6] Y. Masuda and T. Miyazawa, *Bull. Chem. Soc. Japan*, 1969, **42**, 570.
[7] Yu. N. Chirgadze and E. P. Rashevskaya, *Biofizika*, 1969, **14**, 608.
[8] M. Avignon, P. V. Huong, J. Lascombe, M. Marraud, and J. Neel, *Biopolymers*, 1969, **8**, 69.

The configuration of the peptide group can be either *cis* or *trans* and it is possible to distinguish them by means of their i.r. spectra. Lactams are useful models for studying these *cis* and *trans* forms as they are capable of forming both configurations in solution.[9] The results showed that it was difficult to distinguish whether the *cis* or *trans* configuration was present by observing only the amide A band and that the amide I and II bands should also be examined. The effect of temperature, solvent, and concentration must be taken into account as well. Other studies on *cis–trans* configurations have been carried out using the amide A band only.[10]

N-Deuteriation of the amide group is useful for locating vibrations involving the N—H bond, and partial deuteriation often allows one to obtain the unperturbed vibrational frequencies as regular interactions are broken down. The i.r. spectra of orthorhombic crystals of $MeCONH_2$, CD_3CONH_2, and their deuteriated and partially deuteriated derivatives have been recorded so that the assignment of the *trans* and *cis* NHD vibrational bands could be made.[11] For the 50% *N*-deuteriated samples, the absorption bands near 1530 and 1360 cm^{-1} were assigned to the amide II and III respectively of the *trans* NHD species, and those near 1480 and 1400 cm^{-1} to the NH in-phase deformation and the CN-stretching mode of the *cis* NHD species. Confirmation of these assignments came from a normal co-ordinate analysis of eight isotropic species of acetamide based on the Urey–Bradley force field.

Earlier i.r. studies of polyamides have helped in the understanding of the intermolecular and intramolecular attractional forces between amide groups. The effect on the i.r. spectra of a unique folding that can occur in polyhexamethyleneadipamide, Nylon 66, in which there is a tightly folded chain with adjacent re-entry, has recently been investigated.[12] Two new absorption bands appear at 1224 and 1323 cm^{-1} and have been assigned to a $N—CH_2$ bond vibration. These bands are also present in the cyclic and dimer form of Nylon 66 indicating that this bond is also present. The bands, however, were not present in the spectra of the linear oligomers. Observation of the amide A region of the spectra of Nylon 6, 610, 1010, and 11 over a temperature range of 25 to 250 °C has shown that the amount of non-hydrogen-bonded NH groups is less than 1% well below their melting point, and that this proportion rises to between 10 and 20% at 230 °C.[13]

An analysis of the i.r. spectra of a series of glycine dipeptides, tryptophan dipeptides, and L-alanine-L-leucine dipeptides allowed an assignment of the absorption bands from the end and amide groups, and it was suggested

[9] C. Y. Chen and C. A. Swenson, *J. Phys. Chem.*, 1969, **73**, 2999.
[10] L. A. Dement'era, A. V. Logansen, and G. A. Kurkchi, *Zhur. priklad. Spektroskopii*, 1968, **10**, 625.
[11] T. Uno, K. Machida, and Y. Saito, *Bull. Chem. Soc. Japan*, 1969, **42**, 897.
[12] P. Frayer, J. L. Koenig, and J. B. Lando, *J. Macromol. Sci.*, 1969, **B3**, 329.
[13] E. Bessler and G. Bier, *Makromol. Chem.*, 1969, **122**, 30.

that information about the sequence, especially those containing tryptophan, could be obtained from the 1500 to 1700 cm^{-1} region.[14] The spectra of oligo-glycines, $H_3N^+(CH_2 \cdot CONH)_nCH_2CO_2^-$, and oligo-alanines, $H_3N^+[CH(Me)CONH]_nCH(Me)CO_2^-$, and their N-1 deuteriated derivatives have suggested that these oligopeptides exist in an extended conformation in the solid state.[15] The effect of the isotope ^{15}N on the i.r. spectra of $RCONH_2$, where R = CF_3, CCl_3, Ph, *p*-ClC_6H_4, and *p*-$O_2N \cdot C_6H_4$, indicated that all the amide absorption bands which occurred in the region 1435—1328 cm^{-1} shifted to lower frequencies when ^{15}N was present, but was not sensitive to ^{18}O tagging.[16] The existence of a hydrogen-bond between the carboxy-group of poly-L-proline I and II and the hydroxy-group of alcohols was verified by observing the shifts in the amide band vibrations and the OH-stretching bands of the alcohols.[17]

Internal reflectance spectroscopy is useful when samples are not readily adaptable to normal i.r. techniques. The intensity changes that occurred at the amide I and II bands of insoluble fibrous collagen when bathed in D_2O have been related to hydrothermal shrinkage. These i.r. changes together with a differential thermal analysis gave a detailed view of the fibre-melting process.[18]

A detailed analysis of the secondary structure of poly-L-arginine and its derivatives using i.r., *X*-ray, o.r.d., and c.d. showed the presence of the α-helical conformation in poly-$N^{\omega,\omega'}$-dibenzyloxycarbonyl-L-arginine and poly-N^{ω}-benzyloxycarbonyl-L-arginine.[19]

3 Optical Rotatory Dispersion and Circular Dichroism

contributed by D. G. Dalgleish

A. General.—Three reviews are of note: the first two are general surveys of the present state of o.r.d. and c.d. practice, containing sections on polypeptides and proteins,[20] and the third is more specialised, dealing with model compounds, such as peptides, dioxopiperazines, and homopolypeptides.[21] Several other papers, of more general interest, have also appeared in the last year. The problem, as yet unresolved, of determining the helix, pleated sheet, and random coil contents of proteins has been discussed in the light of new and more accurate parameters for the c.d. of the regular structures of poly-L-lysine.[22] Use of these parameters is,

[14] H. S. Kimmel and A. Saifer, *Analyt. Biochem.*, 1969, **32**, 1.
[15] A. Theoret, Y. Grenie, and C. Garrigou-Lagrange, *J. Chim. phys.*, 1969, **66**, 1196.
[16] V. A. Shokol, A. A. Kisilenko, and G. I. Derkach, *Zhur. obshchei Khim.*, 1969, **39**, 1492.
[17] H. Stassmair, J. Engel, and G. Zundel, *Biopolymers*, 1969, **8**, 237.
[18] R. J. Warren, W. E. Smith, W. J. Tillman, and A. Veis, *J. Amer. Leather Chemists' Assoc.*, 1969, **64**, 4.
[19] T. Hayakawa, Y. Kondo, and H. Yamatota, *Bull. Chem. Soc. Japan*, 1969, **42**, 1937.
[20] W. Gratzer and D. A. Cowburn, *Nature*, 1969, **222**, 426; C. W. Deutsche, D. A. Lightner, R. W. Woody, and A. Moscowitz, *Ann. Rev. Phys. Chem.*, 1969, **20**, 407.
[21] D. W. Urry, *Ann. Rev. Phys. Chem.*, 1969, **19**, 477.
[22] N. Greenfield and G. D. Fasman, *Biochemistry*, 1969, **8**, 4108.

however, not completely satisfactory, especially for proteins with low contents of ordered structure. The c.d. spectra of unordered chains from homopolypeptides and fibrous and globular proteins have been determined,[23] and it has been shown that the unordered states all give qualitatively the same shape of spectrum, with a broad minimum at about 200 nm. However, the magnitude of this trough is variable, illustrating that it is difficult, if not impossible, to derive a suitable general parameter for the c.d. spectrum of random coil structures in proteins. An investigation on the Moffitt parameters of 107 proteins has shown that there is a degree of correlation between b_0 and the sum of the residue percentages of those amino-acids whose homopolymers can form helices in solution.[24] A nomogram is presented in this work which allows semiquantitative determination of the α- and β-structure from a plot of b_0 against $(a_0-a_0{}^{\mathrm{R}})$, where $a_0{}^{\mathrm{R}}$ refers to the fully denatured protein.

Calculations on the rotational strength of the peptide n–π^* transitions in α-helical polypeptides show that the residue rotational strength is influenced by the side-chains, since they in turn affect the backbone configuration.[25, 26] The geometry of the hydrogen bonds in the helix is important: this depends upon the torsion angles ϕ and ψ, which may enforce non-linear hydrogen bonds. The effect of chain length is also of importance,[25, 27] and it is therefore difficult to adopt a value of $\mathrm{R}_n{}^{\pi^*}$ which will apply to all peptide α-helices. Further estimates show that the mixing of the n–π^* transition with charge-transfer states is unlikely to affect the rotational strength, but a good agreement with observed and calculated c.d. spectra can be obtained by using a secular determinant to describe the mixing of n–π^* and π–π^* states, instead of the usual method of first-order perturbations.[28] The positions of the bands considered are critical for the success of this method, and small alterations in either position or band width alter the fit. Calculations on parallel and antiparallel β-structure [29] show that poly-L-lysine and poly-L-serine are in the antiparallel configuration. The exciton component of the absorption shifts to longer wavelength as the width of the antiparallel structure increases, and the position of the $\pi\pi^*$ band may be used as a criterion of sheet width. One interesting paper has calculated the i.r. c.d. which should be obtained from a helical polymer,[30] and it is shown that such determinations would be useful: it is to be hoped that production of an instrument capable of measuring c.d. spectra in the vibrational region of the spectrum will be stimulated by this publication.

[23] M. L. Tiffany and S. Krimm, *Biopolymers*, 1969, **8**, 347.
[24] D. E. Goldsack, *Biopolymers*, 1969, **7**, 299.
[25] J. N. Vournakis, J. F. Yan, and H. A. Scheraga, *Biopolymers*, 1968, **6**, 1531.
[26] J. F. Yan, G. Vanderkooi, and H. A. Scheraga, *J. Chem. Phys.*, 1968, **49**, 2713.
[27] J. A. Schellman and M. J. Lowe, *J. Amer. Chem. Soc.*, 1968, **90**, 1070.
[28] R. W. Woody, *J. Chem. Phys.*, 1968, **49**, 4797.
[29] R. W. Woody, *Biopolymers*, 1969, **8**, 669.
[30] C. W. Deutsche and A. Moscowitz, *J. Chem. Phys.*, 1968, **49**, 3257.

B. Amino-acids and Small Model Compounds.—The amino-acids of particular interest are those which possess absorption bands in the near u.v., namely tyrosine, tryptophan, phenylalanine, and cystine. The last is perhaps of most interest since it is itself a dissymmetric chromophore, and it is important to ascertain the sign of the c.d. which should be associated with the chirality of the disulphide group. Three studies have recently appeared, all of which reach the same general conclusion. From studies on cystine and its dihydrochloride in KBr discs,[31] and on substituted 1,2-dithiane derivatives,[32, 33] it is generally concluded that a left-handed chirality of the disulphide group gives rise to a negative peak in the near-u.v. c.d.

Phenylalanine has also been studied in the free form and as an amide.[34, 35] *N*-Acetyl-L-phenylalanineamide possesses seven bands in the c.d. between 195 and 268 nm, some of which correspond to peaks or shoulders in the absorption spectrum. A trough in the c.d. spectrum at 240 nm increased markedly in size when solvent polarity was altered, an effect due to the presence of the primary amide. Studies on L-phenylalanine, and *N*-acetyl-L-phenylalanine and its amide, methyl, and ethyl esters showed variations in the signs and intensities of the various c.d. bands, but all are basically similar. Only non-totally-symmetrical-vibrations affect the c.d., and the total rotational strength increased three- to eight-fold when, the temperature was reduced to 77 K, suggesting that a conformational equilibrium occurs at 298 K. A similar study using *N*-acetyl-L-tryptophanamide showed an eighteen-fold increase in the c.d. when the temperature was lowered over the same range, again suggesting conformational mobility,[36] but the tryptophan peak in the c.d. of chymotrypsinogen varies only by about 25% over this temperature range, showing that the side-chain is not conformationally mobile in the protein.

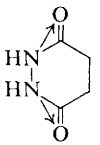
N-Thiobenzoylamino-acids exhibit an optically active band in the region of 270 nm,[37] and this is also true for terminal *N*-thiobenzoyl oligopeptides.[38] The terminal amino-acid operates upon the rotational strength of the aromatic transition, and L-amino-acids give positive and L-imino-acids (proline) give negative rotational strength. The penultimate amino-acids in the oligo-peptides also affect the observed rotational strength, by influencing both the transition and conformational equilibrium.

Dipeptides and cyclic dipeptides have been studied as models of amide chromophores in proteins. A theoretical treatment of molecules containing two peptide groups,[39] considering the n–π^* and π–π^* transitions, shows

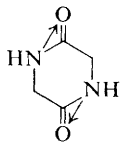
[31] A. Imanishi and T. Isemura, *J. Biochem.* (*Japan*), 1969, **65**, 309.
[32] G. Claeson, *Acta. Chem. Scand.*, 1968, **22**, 2429.
[33] R. M. Dodson and V. C. Nelson, *J. Org. Chem.*, 1968, **33**, 3966.
[34] J. Horowitz, E. H. Strickland, and C. Billups, *J. Amer. Chem. Soc.*, 1969, **91**, 184.
[35] N. S. Simmons, A. O. Barel, and A. N. Glazer, *Biopolymers*, 1969, **7**, 275.
[36] E. H. Strickland, J. Horowitz, and C. Billups, *Biochemistry*, 1969, **8**, 3205.
[37] G. C. Barrett and A. R. Khokar, *J. Chem. Soc.* (*C*), 1969, 1120.
[38] G. C. Barrett, *J. Chem. Soc.* (*C*), 1969, 1123.
[39] P. M. Bayley, E. B. Nielsen, and J. A. Schellman, *J. Phys. Chem.*, 1969, **73**, 228.

that the observed o.r.d. and c.d. are dependent upon the angles ϕ and ψ, and that a map of the sign of the rotational strength against these angles can be drawn. Other simple compounds studied are those in which the amide group is held in a rigid conformation, such as cyclic amides and dioxopiperazines. Thus 3-aminopyrrolidin-2-one has a c.d. spectrum which can be attributed to an n–π^* transition at 220 nm, and a π–π^* transition at 190 nm, the latter of which is split in hydroxylic solvents.[40] In acetonitrile, a third band is found at about 200 nm, which may be due to aggregation, as was also found in the c.d. spectrum of L-pyrrolidin-2-one-5-carboxylic acid.[41]

Substituted 2,5-dioxopiperazines have been shown to be non-planar,[42] and this has an effect upon the o.r.d.[43] Urea treatment also affects the o.r.d. and c.d. and reduces the rotations at any wavelength.[44] This effect is thought to arise from the distortions produced by hydrophobic interactions of the 2- and 5-substituents, since L-serine dioxopiperazine is largely unaffected, as is L-valine-glycine dioxopiperazine, whereas the dioxopiperazines of L-alanine, L-lysine, and L-valine are increasingly sensitive, in that order. Solvent effects upon the transitions of rigid molecules, such as L-pyrrolidin-2-one-5-carboxylic acid, can be separated from those due to conformational changes, since the molecule is rigid.[40] The c.d. spectra of dioxopiperidazines (I) differ from those of dioxopiperidines (II) in that the π–π^* transitions are split since the transition moments (indicated by arrows) are not parallel,[45]



(I)



(II)

The cyclic oligopeptides tyrocidin and gramicidin S have been studied by c.d.[46] The latter molecule shows minima at 207 and 215 nm, whose shapes and intensities are altered by hydrogenation, showing that the aromatic side-chains (from D-phenylalanine) contribute in this region either by interaction with the backbone or by spectral overlap. This result is of importance, since it illustrates the production of a c.d. spectrum not unlike that of an α-helix in a molecule which cannot be helical; it appears

[40] N. J. Greenfield and G. D. Fasman, *Biopolymers*, 1969, **7**, 595.
[41] M. Goodman, C. Toniolo, and J. Falcetta, *J. Amer. Chem., Soc.* 1969, **91**, 1816.
[42] E. Benedetti, P. Corradini, M. Goodman, and C. Pedone, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 650.
[43] E. B. Nielsen and J. A. Schellman, *J. Phys. Chem.*, 1967, **71**, 2297.
[44] M. D'Alagni, B. Pispisa, and F. Quadrifoglio, *Ricerca sci.*, 1968, **38**, 910; M.D' Alagni, and B. Pispisa, *J. Biol. Chem.*, 1969, **244**, 5843.
[45] C. G. Overburger, G. Montaudo, J. Šebenda, and R. A. Veneski, *J. Amer. Chem. Soc.*, 1969, **91**, 1256.
[46] S. Laiken and L. C. Craig, *J. Biol. Chem.*, 1969, **244**, 4454.

that such loops may interfere with helix estimations in proteins. A similar result can be obtained from a study of bacitracin A and the peptide hormone ACTH, where the α-helix type of c.d. may arise from a local rigidity of the molecule.[47] The tyrocidins A, B, and C also show the similar type of far-u.v. c.d., but in addition the B and C molecules have a negative peak and shoulder at 230 nm, which disappear on hydrogenation and are attributed to tryptophan.[46]

C. Polypeptides.—Poly-L-lysine of chain length greater than 20—35 residues is known to be in the form of a β-pleated sheet at 50 °C and high pH. This poly-amino-acid in these conditions forms large aggregates of diameter greater than 5 μ, which are very large pleated sheet structures,[48] 85% of the polymer exists in this form, the remainder being α-helical. Addition of dodecyl sulphate produces smaller aggregates, with the consequence that the rotational strength of the β-sheet decreases, since this is known to be dependent upon the size of the regular structure.[49] The β-form in dodecyl sulphate is thought to be a better model for β-sheet in proteins than the form at 50 °C. Formation of the α-helix in poly-L-lysine is also dependent upon the length of the polymer, and it is found that about 75 residues are necessary before the typical c.d. spectrum is found.[48] Poly-L-arginine has a helix–coil transition at about 0·5 pH units higher than that of poly-L-lysine in water and water–dioxan mixtures, because of the increased pK of the guanidinium group, and it is thought that this may help to explain the difference between lysine-rich and arginine-rich histone fractions and their interactions with nucleic acids.[50]

Poly-L-glutamic acid (PGA) and its derivatives are polymers commonly used for studying conformational changes. A temperature-dependent change is found when PGA is cooled to −35 °C, which shows a hysteresis effect on heating, and is consistent with a helix–coil transition with aggregation.[51] The effect of ethanol on the helix–coil transition is to increase the stability of the helix for both charged and uncharged PGA, the stabilisation of the uncharged form being essentially an entropic effect.[52] A comparison of the unordered forms of PGA and polyhydroxymethyl-L-glutamine shows that the two differ markedly in their c.d. spectra, probably because of the greater extension of the PGA coil which results from its charge. The c.d. spectra of the helical forms are similar, although there is some evidence for aggregation in PGA.[53] A calculation of the counterion effect on the extended-helix structure of unordered PGA shows that the added salt should alter

[47] L. C. Craig, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 152.
[48] L.-K. Li and A. Spector, *J. Amer. Chem. Soc.*, 1969, **91**, 220.
[49] D. W. Urry, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 394.
[50] J. M. Rifkind, *Biopolymers*, 1969, **8**, 685.
[51] F. Travers, M. Debey, P. Douzon, and A. M. Michelson, *Biochim. Biophys. Acta*, 1969, **194**, 265.
[52] G. Conio and E. Patrone, *Biopolymers*, 1969, **8**, 57,
[53] A. J. Adler, R. Hoving, J. Potter, M. Wells, and G. D. Fasman, *J. Amer. Chem. Soc.*, 1968, **90**, 4736.

the conformation of the polymer by decreasing repulsive interactions, and it is indeed found that there are changes in the c.d. of the charged form when 1·8M LiCl is added.[54] Left- as well as right-handed helix is formed in copoly-γ-methyl-DL-glutamate, even when the content of D-amino acid is small:[55] this should be contrasted with the behaviour of copoly-γ-benzyl-DL-glutamate, which possesses appreciable left-handed helix only when the content of D-amino-acid is high.

A calculation of the relative stability of right- and left-handed helices in polypeptides shows that there is only a slight energy difference between the two forms of poly-γ-*p*-chlorobenzyl-L-glutamate,[26] but o.r.d. measurements in various solvents show that the molecule exists in a right-handed helix.[56] The helix–coil transition of the related molecule poly-γ-benzyl-L-glutamate (PBLG) provides a value of ΔH for the transition which is in good agreement with calorimetric determinations,[57] and in contrast to PBLG, which forms helices on heating, poly-γ-*N*-benzyloxycarbonyl-L-α,γ-diaminobutyric acid behaves more 'normally', in that the helix is the favoured form at lower temperatures, the transition being at about 10 °C.[58] The protonation behaviour of poly-γ-ethyl-L-glutamate in sulphuric acid–water mixtures shows a transition from helix to coil at 57—62% sulphuric acid.[59] It is thought that only about one-half of the amide groups are protonated early in the transition, and the co-operativity of the transition is demonstrated by the narrow range of acid concentrations over which it occurs.

Studies on a block copolymer of γ-ethyl-L-glutamate and L-tryptophan, with the tryptophan being replaced with varying amounts of 2-(*o*-nitrophenylsulphenyl)-L-tryptophan showed a c.d. spectrum typical of an α-helix. As the degree of modification was decreased, a gradual change was observed in the o.r.d. from which it was thought that no conformational change was occurring.[60] This is to be compared with a similar situation in unmodified copolymers of γ-ethyl-L-glutamate and L-tryptophan.[61]

Another such polypeptide is of interest as being a model for collagen. Poly(glycyl-L-prolyl-L-alanine) shows an increased far-u.v. c.d. as the temperature is decreased from 24 to −112 °C, and at the lower temperature gives a c.d. spectrum similar in shape to, but of smaller magnitude than, that of collagen.[62] Comparison of this polymer with both collagen and poly-L-proline II shows that the increased c.d. is not the result of

[54] S. Krimm, J. E. Mark and M. L. Tiffany, *Biopolymers*, 1969, **8**, 695.
[55] Y. Masuda, T. Miyazawa, and M. Goodman, *Biopolymers*, 1969, **8**, 515.
[56] E. H. Erenrich, R. H. Andrietta, and H. A. Scheraga, *Biopolymers*, 1969, **7**, 805.
[57] M. Cortijo, A. Roig, and F. G. Blanco, *Biopolymers*, 1969, **7**, 315.
[58] F. Gaskin, S. Kubota, and J. T. Yang, *J. Amer. Chem. Soc.*, 1969, **91**, 6526.
[59] J. Steigman, E. Peggion, and A. Cosani, *J. Amer. Chem. Soc.*, 1969, **91**, 1822.
[60] E. Peggion, A. Fontana, and A. Cosani, *Biopolymers*, 1969, **7**, 517.
[61] E. Peggion, A. Cosani, A. S. Verdini, A. del Pra, and M. Mammi, *Biopolymers*, 1968, **6**, 1477.
[62] F. R. Brown tert., J. P. Carver, and E. R. Blout, *J. Mol. Biol.*, 1969, **39**, 307.

temperature-dependent solvent effects, but must be due to a conformational transition of the molecule, which gives an increase in periodic structure.

D. Metal Ion Complexes.—Complexes of cupric ions with amino-acids show optical activity and c.d. bands in the visible,[63] and complexes of the type CuA_2 and CuA^+ have been described.* Similarly, cobalt(II) and nickel(II) give visible c.d. spectra from such complexes, with, in the case of cobalt, an additional complex[63] of the type CoA_3^-. Cupric and nickel ions bound to amino-acids show a negative c.d. band at 210 nm, in agreement with the octant rule for n–π^* transitions,[64] although this is obscured by a charge-transfer band in cupric ion complexes. Complexes of cupric ions with di- and tri-peptides exhibit c.d. peaks in the region of 300—315 nm, where there are no peaks in the absorption spectrum, and the c.d. of an aromatic amino-acid, which may be enhanced by incorporation into a dipeptide, may also be enhanced by the chelation of cupric ion.[64] Cupric ions complexed with pentapeptides of glycine and alanine gave a visible c.d. spectrum similar to that obtained with the tripeptide $(Ala)_3$, and contributions at about 300 nm have been found.[65] In addition, the optical rotatory properties of the complex of cupric ion with the *N*-terminal pentapeptide from sperm whale myoglobin can be distinguished qualitatively from those of histidine-containing peptides.

Complexes of copper(II) and nickel(II) with proteins show that the first equivalent of either ion binds to the *N*-terminal end of lysine-vasopressin, conalbumin, α-chymotrypsin, and bovine albumin, while cupric ions bind to the *N*-terminal of lysozyme and nickel(II) to the *N*-terminal of ribonuclease.[66] A second or third equivalent of ion binds to bovine albumin sulphydryl groups.

Iron also binds to bovine serum albumin to give artificial non-haem iron proteins.[67] Three types of such proteins are found, depending on the method of incubation of the complex mixture, and whether β-mercaptoethanol or sulphide is present in the reaction mixture. The complex prepared in the presence of sulphide has spectroscopic similarities to ferredoxin, but can be distinguished from the latter molecule by its different c.d. spectrum. Cotton effects appearing at 300—350 nm can be assigned to an iron–protein interaction, and those between 350 and 400 nm to an iron–sulphur interaction.

Pyrocatechase, a natural non-haem iron protein, contains trivalent iron, as estimated from changes occurring in the c.d. spectrum.[68] Certain of the

[63] A. Bonniol. *J. Chim. phys.*, 1969, **66**, 336, 340, 344.
[64] J. M. Tsangaris, J. W. Chang, and R. B. Martin, *J. Amer. Chem. Soc.*, 1969, **91**, 726.
[65] C. R. Hartzell and F. R. N. Gurd, *J. Biol. Chem.*, 1969, **244**, 147.
[66] J. M. Tsangaris, J. W. Chang, and R. B. Martin, *Arch. Biochem. Biophys.*, 1969, **130**, 53.
[67] K. McCarthy and W. Lovenberg, *J. Biol. Chem.*, 1968, **243**, 6436.
[68] A. Nakazawa, T. Nakazawa, S. Kotani, M. Nozaki, and O. Hayaishi, *J. Biol. Chem.*, 1969, **244**, 1527.

* A = an amino-acid residue.

c.d. bands in this protein were found to be associated with the enzyme activity, since they changed in the presence of the substrate, catechol, or were destroyed by removal of the iron atom. The c.d. of xanthine oxidase shows similarities to non-haem iron proteins, such as ferredoxin, in both the native and deflavo state,[69, 70] and since changes on oxidation or reduction occur in two stages, it has been inferred that there are two types of iron atoms in the molecule.

E. Haem Proteins.—Calculations show that the origin of the rotational strength of the haem transitions in myoglobin is probably the interaction with the π–π^* transitions of aromatic side-chains, rather than with the peptide n–π^* or π–π^* or side-chain σ–σ^* transitions.[71] The side-chains with most influence appear to be the histidines (64) and (97), phenylalanine (33), and the tyrosines (103) and (146). In agreement with this, the c.d. spectrum of the side-chain region changes on the liganding of the haem, showing the likelihood of some interaction.[72] Haematin-*a*, the prosthetic group of cytochromes, exists in an aggregated state in solution,[73] but by addition of pyridine can be brought to a state approaching monomeric. This latter state is still optically active, and it may be that the observed optical activity is a reflection of the dissymmetry of the conformation of the molecule around the haem-iron.

Haemoglobin recombination has been studied in several cases, and the c.d. spectra of abnormal haemoglobins have been reported. As in myoglobin,[72] the helical content of haemoglobin is decreased by the removal of the haem, but the helix content is fully restored by the binding of one mole of haem per globin.[74] The α and β chains do not give the same c.d. spectra in the Soret region under any conditions,[75–77] and recombination leads to a c.d. spectrum indistinguishable from native haemoglobin, but differing from the sum of the subunits.[75–78] Formation of the β_4 complex also alters the haem environment, and the abnormal β chain of haemoglobin-Yakima (His for Asp at β99) causes the c.d. spectrum of haemoglobin to differ from that of haemoglobin-A.[78] This must arise because of altered interactions between the subunits, since isolated chains from the abnormal haemoglobin have identical c.d. spectra with those from normal haemoglobin.

The c.d. spectra of haemoglobins show side-chain effects between 260 and 300 nm. Foetal haemoglobin and haemoglobin-M_{Boston}, which have tryptophan replacing the tyrosine in position (130) and tyrosine replacing

[69] H. Komal, V. Massey, and G. Palmer, *J. Biol. Chem.*, 1969, **244**, 1692.
[70] G. Palmer and V. Massey, *J. Biol. Chem.*, 1969, **244**, 2614.
[71] M. C. Hsu and R. W. Woody, *J. Amer. Chem. Soc.*, 1969, **91**, 3679.
[72] T. Samejima and M. Kita, *J. Biochem. (Japan)*, 1969, **65**, 759.
[73] F. C. Young and T. E. King, *J. Biol. Chem.*, 1969, **244**, 509, 515.
[74] K. Javaherian and S. Beychok, *J. Mol. Biol.*, 1968, **37**, 1.
[75] G. Geraci and T.-K. Li, *Biochemistry*, 1969, **8**, 1848.
[76] M. Nagai, Y. Sugita, and Y. Yoneyama, *J. Biol. Chem.*, 1969, **244**, 1651.
[77] Y. Ueda, T. Shiga, and I. Tyuma, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 1.
[78] P. T. Goodall and E. M. Shooter, *J. Mol. Biol.*, 1969, **39**, 675.

the distal histidine respectively, show marked changes from haemoglobin-A in this region.[79] A transition at 260 nm has been found to be sensitive to the state of the haem, and a further transition at 280 nm has been associated with configurational changes linked to co-operative interactions,[80] both of these being linear with oxygen saturation. The ellipticity of haemoglobin-A is the same as the ellipticities of the subunits so that no perturbation of these transitions occurs during recombination. The band at 260 nm is sensitive to the spin state of the iron and increases from high-spin to low-spin; this is the opposite of the effect of spin on the Soret bands, showing that the two must have different electronic origins.[81]

Deoxygenation or liganding of haemoglobin leads to small changes, of the order of 8—10%, in the magnitude of the 233 nm trough in the o.r.d. This change is independent of the nature of the ligand, and the change is not found for isolated α or β chains.[82] It occurs, however, for $\alpha\beta$ dimers and for haemoglobins modified to prevent co-operative interactions,[82] as well as for the variant haemoglobins S and F.[79]

Catalase possesses a small negative c.d. contribution in the Soret region, which is destroyed by denaturation by acid, alkali, or urea, although the molecule still contains 20—25% helix in the acid-denatured form, and 15—20% in the alkali-denatured state.[83] Addition of azide or cyanide to the molecule reduces the helical content from about 50% to 30% and gives a completely inactive enzyme.

The helical contents of cytochrome *c'* and *cc'* were estimated to be about 63%, and showed no variation with the state of the haem.[84] This high helical content may in part result from the large amount of alanine in the molecules. Exposure of the proteins to 6M urea for 20 hr reduces the helix content, but has no effect upon the c.d. spectrum above 240 nm, suggesting that the haem environment is protected by sections of helix.

Peroxidase-*a* from Japanese radish contains about 35% helix, which does not alter with the state of the haem, nor is it affected by combining with fluoride or cyanide. On acidification or removal of the haem, there is evidence only for pleated sheet as the regular structure. This may arise either from an helix–pleated sheet (α–β) conformational change, or because the conditions used destroy the α-helix, but leave intact a portion of β-structure whose c.d. contribution is masked by that of helix in the intact molecule.[85] The c.d. of the protein shows some fine structure in the aromatic region, underlying which there appears to be a broad band due to a haem transition. Peroxidase-*a* from horseradish shows haem-associated

[79] T.-K. Li and B. P. Johnson, *Biochemistry*, 1969, **8**, 2083.
[80] S. R. Simon and C. Cantor, *Proc. Nat. Acad. Sci., U.S.A.*, 1969, **63**, 205.
[81] T.-K. Li and B. P. Johnson, *Biochemistry*, 1969, **8**, 3638.
[82] G. Hänisch, J. Engel, M. Brunori, and H. Fasold, *European J. Biochem.*, 1969, **9**, 335.
[83] T. Samejima and M. Kita, *Biochim. Biophys. Acta*, 1969, **175**, 24.
[84] Y. Imai, K. Imai, K. Ikeda, K. Hamaguchi, and T. Horio, *J. Biochem.* (*Japan*), 1969, **65**, 629.
[85] K. Hamaguchi, K. Ikeda, C. Yoshida, and Y. Morita, *J. Biochem.* (*Japan*), 1969, **66**, 191.

optical activity which is greatly diminished by reduction of the haem,[86] and it is thought that there is a local conformational change of the protein around the haem on reduction.

Lactate dehydrogenase contains both haem and flavin as prosthetic groups. The o.r.d. and c.d. of the enzyme below 250 nm are unaffected by the removal of flavin mononucleotide (FMN), but the c.d. shows a negative band at 270 nm which is missing when FMN is removed,[87] this latter behaviour not being observed by o.r.d.[88] From the far-u.v. c.d. spectrum, a helix content of 17—25% is calculated, which is not affected by the oxidation state of the haem,[87, 89, 90] nor by the removal of the haem from a low-molecular-weight core obtained from proteolytic degradation.[90] The Soret-band Cotton effect of the haem is sensitive to the presence of FMN, and the optical and catalytic properties of the molecule may be due to haem–flavin interaction, although whether this is direct,[89] or *via* protein side-chains.[90] is uncertain. By calculating the o.r.d. of the enzyme by a Kronig–Kramers transform from the c.d. spectrum and comparing it with the observed o.r.d., a further band may be found which contributes to the o.r.d. This is centred at 150 nm, and is not sensitive to the state of the haem, and may result from a helix contribution.[91] The above studies used lactate dehydrogenase extracted from the yeast *Saccharomyces cerevisiae*: studies on the enzyme from *Hanseluma anomala* show a similar c.d. spectrum below 300 nm, but altered haem transitions.[87, 90] A further study, on the enzyme isolated from pig heart, gave indications of a higher helix content (about 30%, obtained from the Moffitt b_0), which was unaltered by cofactor binding.[92]

F. Binding of Other Molecules to Proteins.—Other common prosthetic groups are flavin and pyridoxal. Riboflavin, when in a free state, shows bands in the c.d. at 480, 450, 370, and 330 nm, which are inverted and enhanced by binding to egg-white flavoprotein.[93] In contrast to free riboflavin, the bound molecule is not sensitive to pH-induced conformational changes. Studies on reduced and oxidised diphospho-pyridine nucleotide dehydrogenase show that the interaction of the flavin groups and the protein is dependent on the oxidation state of the enzyme, through changes in the protein conformation which depend upon the oxidation state of the enzyme.[94] Pig heart lipoamide dehydrogenase changes conformation markedly when flavin adenine dinucleotide (FAD) is removed, but in the presence of equimolar FMN no change is found in the far-u.v. c.d. spectrum,

[86] G. E. Willick, G. R. Schonbaum, and C. M. Kay, *Biochemistry*, 1969, **8**, 3729.
[87] J. M. Sturtevant and T. Y. Tsong, *J. Biol. Chem.*, 1969, **244**, 4942.
[88] J. M. Sturtevant and T. Y. Tsong, *J. Biol. Chem.*, 1968, **243**, 2359.
[89] T. Y. Tsong and J. M. Sturtevant, *J. Biol. Chem.*, 1969, **244**, 2397.
[90] M. Iwatsubo and J. L. Risler, *European J. Biochem.*, 1969, **9**, 280.
[91] T. Y. Tsong and J. M. Sturtevant, *J. Amer. Chem. Soc.*, 1969, **91**, 2382.
[92] H. d'A. Heck, *J. Biol. Chem.*, 1969, **244**, 4375.
[93] M. Nishikimi and K. Yagi, *J. Biochem.* (*Japan*), 1969, **66**, 427.
[94] T. E. King, P. M. Bayley, and B. Mackler, *J. Biol. Chem.*, 1969, **244**, 1890.

although it can be shown that no FMN is bound.[95] The longer-wavelength bands of the enzyme show inversion and enhancement when FAD binds. α-Glucan phosphorylase [96] and amine oxidase [97] contain bound pyridoxal, which gives a band in the c.d. spectrum at 333 nm, and which interacts by a Schiff-base mechanism.

Binding of NAD^+ to glyceraldehyde phosphate dehydrogenase gives a c.d.-active band at 350 nm, which increases linearly with the amount of NAD^+ bound up to a value of 2 moles per mole of protein, but does not level off until there are 3 moles of NAD^+ per mole of protein.[98] This binding is also accompanied by a change in b_0. The apo- and holo-enzyme do not differ in conformation, as shown by far-u.v. c.d., and the band at 300—450 nm is due to a charge-transfer between NAD^+ and a sulphydryl group, and to interaction between adenine and the enzyme.

1-Anilino-8-naphthalenesulphonic acid binds to bovine serum albumin, becoming optically active in the process. A study of the resulting c.d. spectra shows that there are five primary binding sites of two types: such heterogeneity might not have been detectable by other methods.[99] Poly-L-proline I and II bind alcohols, and the binding of benzyl alcohol is revealed by the production of extrinsic Cotton effects in the near-u.v. Although this indicates a homogeneity of binding sites on the regular structures, it is not certain whether the observed optical activity results from the asymmetry of the binding site or from interaction between the alcohol molecules themselves.[100]

A complex between two different proteins is illustrated by the complex of haemoglobin and haptoglobin.[101] It is probable that the observed lack of Bohr effect is due to the alteration of the haem environment, since the complex possesses an altered Soret-band Cotton effect and also shows changed c.d. spectra in the region of 240—310 nm, showing altered side-chain environments.

Binding of protein to polynucleotide is illustrated by poly-L-lysine binding to polyadenylic acid. The o.r.d. of the polynucleotide is red-shifted and decreased when the polypeptide binds, and complex formation proceeds in two steps, namely association of a single chain of each followed by aggregation, the latter step being maximal when the polylysine is 50% helical:[102] no complex formation is observed when the polypeptide is fully helical. The binding of histone fractions to nucleic acids has also been studied;[103] the conformation of individual histones is ionic-strength

[95] A. H. Brady and S. Beychok, *J. Biol. Chem.*, 1969, **244**, 4634.
[96] A. Kamogawa, T. Fukui, and Z. Nikuni, *J. Biochem.* (*Japan*), 1968, **64**, 709.
[97] O. Adachi and H. Yamada, *J. Biochem.* (*Japan*), 1969, **65**, 639.
[98] J. J. M. de Vijlder and B. J. M. Harmsen, *Biochim. Biophys. Acta*, 1969, **178**, 434.
[99] S. R. Anderson, *Biochemistry*, 1969, **8**, 4838.
[100] H. Strassmair, J. Engel, and G. Zundel, *Biopolymers*, 1969, **8**, 237.
[101] H. Hamaguchi, A. Isomoto, and H. Nakajima, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 6.
[102] B. Davidson and G. D. Fasman, *Biochemistry*, 1969, **8**, 4116.
[103] D. Y. H. Tuan and J. Bonner, *J. Mol. Biol.*, 1969, **45**, 59.

dependent, and it appears that the conformation of the histones in association with DNA approximates to that at high ionic strength. The conformation of the DNA is altered by the binding of whole histone, but the lysine-rich histone is not solely responsible for this effect. A conformational change also occurs in the protein of tobacco mosaic virus when the RNA is removed,[104] and it is thought that the aromatic Cotton effects observed show that the tryptophan-52 of the coat protein is in the vicinity of the RNA in the intact virus, by being near to the surface of a protein subunit.

G. Aggregation, Disaggregation, and Denaturation.—O.r.d. changes do not occur on dissociation of chicken egg-white macroglobulin into subunits in acid solution, although changes are observed in the absorption spectrum.[105] This behaviour is similar to that found for α-gliadin, which aggregates with no change in the far-u.v. c.d. spectrum.[106]

On the other hand, changes in the far-u.v. c.d. spectra of virus coat proteins are observed on aggregation.[107–110] Depolymerisation of the protein coat may be effected by 11M acetic acid, and conformational changes may be observed as the acid is dialysed away. In the case of bacteriophage *fr* protein, the molecules appear to become more helical on disaggregation, and to form an intermediate random form as the acid is removed, finally refolding to the native conformation as the acid concentration falls below 0·3M. Tobacco mosaic virus protein, on the other hand, does not go through the intermediate state, and, although the helical content is increased in 11M acetic acid, the native structure is attained at concentrations of less than 5M acid.

In general, it appears that the majority of proteins which have been recently studied show little or no detectable change of conformation when aggregated or dissociated into monomeric units. Alkaline phosphatase[109] shows only minor changes when dissociated by acid, nor is a change found when the zinc atom is removed. Similarly, caeruloplasmin when dissociated in 5M urea shows no change in the far-u.v. c.d. spectrum,[110] and haemocyanin dissociated by oxygen shows only minor changes in the near-u.v.[111] Glutamate dehydrogenase[112] and haemerythrin[113] likewise

[104] P.-Y. Cheng, *Biochemistry*, 1968, **7**, 3367.
[105] J. W. Donovan, C. J. Mapes, J. G. Davis, and R. D. Hamburg, *Biochemistry*, 1969, **8**, 4190.
[106] D. D. Kasarda, J. E. Bernardin, and W. Gaffield, *Biochemistry*, 1968, **7**, 3950.
[107] D. Schubert, *Biochim. Biophys. Acta*, 1969, **188**, 147.
[108] D. Schubert and B. Krafczyk, *Biochim. Biophys. Acta*, 1969, **188**, 155.
[109] M. L. Applebury and J. E. Coleman, *J. Biol. Chem.*, 1969, **244**, 308.
[110] Y. Hibino, T. Samejima, S. Kajiyama, and Y. Nosoh, *Biochim. Biophys. Acta*, 1968, **168**, 411.
[111] H. A. de Phillips, K. W. Nickerson, M. Johnson, and K. E. van Holde, *Biochemistry*, 1969, **8**, 3665.
[112] P. Dessen and D. Pantaloni, *European J. Biochem.*, 1969, **8**, 292.
[113] D. W. Darnall, K. Garbett, I. M. Klotz, S. Aktipis, and S. Keresztes-Nagy, *Arch. Biochem. Biophys.*, 1969, **133**, 103.

show no evidence of conformational changes on aggregation of their subunits.

Glucagon, however, shows a concentration-dependent increase of helix content, with the appearance of a new Cotton effect at 290 nm, due to tryptophan,[114] and the concentration-dependence of the helix persists even in a helix-inducing solvent (chloroethanol).[115] The aggregation appears to occur *via* two linked equilibria, the associated state acting as a thermodynamic trap for structural changes. Phycoerythrin and phycocyanin show changes in the visible c.d. spectrum when dissociated by mercurials: this may arise from interaction of the chromophoric groups in the intact molecules.[116]

Most proteins show some unfolding and loss of structure when exposed to high concentrations of urea or guanidinium chloride. An exception to this is phospholipase A_2 and its zymogen, which are unaffected by 8M urea, probably because the large number of disulphide bonds stabilise the conformation, which remains at about 50% helical.[117] If urea-denaturation is accompanied with reduction of disulphide bonds of the protein, then correct reoxidation of these is necessary to reverse the denaturation. Thus, ribonuclease can be refolded to its native conformation from the reduced denatured state in the presence of mercaptoethanol,[118, 119] but in the absence of mercaptoethanol, recovers only 83% of its activity and refolds to a mixture of different conformations.[118]

Caeruloplasmin is not unfolded in 5M urea, but is dissociated; as the urea concentration is raised, the β-structure of the subunits is destroyed, leaving only a random structure in 8M urea.[110] Guanidinium chloride causes complete unfolding of *E. coli* acyl carrier protein,[120] and 8M urea produces a complete loss of the ordered structure of catalase,[83] but there is a residual c.d. contribution from α-gliadin in 8M urea at 220 nm, showing that ordered structure still exists.[106]

The denaturation caused by urea is not always reversible, as shown by carbonic anhydrases B and C, which irreversibly lose side-chain optical activity.[121] Also, α-amylase has about 20% helix, which is lost on treatment with urea and edta, the denaturation only being reversible in the presence of cupric ions.[122] A further form of α-amylase is unstable in acid solution and is denatured with a shift of the o.r.d. minimum from 235 to 228 nm.[123]

[114] W. B. Gratzer and G. H. Beaven, *J. Biol. Chem.*, 1969, **244**, 6675.
[115] P. A. Srere and G. C. Brooks, *Arch. Biochem. Biophys.*, 1969, **129**, 708.
[116] J. Pecci and E. Fujimori, *Biochim. Biophys. Acta*, 1969, **188**, 230.
[117] A. M. Scanu, L. L. M. van Deenen, and G. H. DeHaas, *Biochim. Biophys. Acta*, 1969, **181**, 471.
[118] M. N. Pflumm and S. Beychok, *J. Biol. Chem.*, 1969, **244**, 3982.
[119] T. Isemura, K. Yutani, A. Yutani, and A. Imanishi, *J. Biochem. (Japan)*, 1968, **64**, 411.
[120] T. Tagaki and C. Tanford, *J. Biol. Chem.*, 1968, **243**, 6432.
[121] J. A. Verpoorte and C. Lindblow, *J. Biol. Chem.*, 1968, **243**, 5993.
[122] A. Yutani, K. Yutani, and T. Isemura, *J. Biochem. (Japan)*, 1969, **65**, 201.
[123] T. Takagi, M. Arai, Y. Minoda, T. Isemura, and K. Yamada, *Biochim. Biophys. Acta*, 1969, **175**, 438.

Studies of myoglobin in urea and ethanol solutions show that the random coil is only formed at high urea molarity, and that myoglobin is still as helical in 70% ethanol as it is in the native form, although the sedimentation coefficient is that of a swollen molecule.[124] The induction of helix in denatured lysozyme by addition of ethanol occurs at higher ethanol concentrations than for myoglobin and it is suggested that the secondary structure of denatured proteins is similar to the native secondary structure, thus providing a nucleus for the folding process.[124]

H. Lipoproteins and Membranes.—Following on observations that human serum high-density lipoprotein appeared to be destabilised to solvent effects by removal of lipid,[125] the dependence on temperature of the conformation of the protein was studied by c.d.[126] It was found that the changes dependent on temperature did occur, and that the effect of temperature was increased by removal of lipid, which suggested that the lipid moieties have a role in maintaining the protein conformation. Similar results have been found for low-density lipoprotein,[127] and the appearance of the c.d. spectra suggests that the proteins contain both α- and β-structure, whose amounts are dependent on the temperature and the amount of lipid bound to the molecule.[128] High concentrations of dodecyl sulphate have the same effect as lipid in stabilising the proteins.[126] C.d. and i.r. studies on human β-lipoprotein suggest that β-structure is present as well as α-helix, but that the β-structure is destroyed by dodecyl sulphate.[129] The β-structure is partly retained on removal of lipid, but some transforms to random coil.

Plasma membranes show different far-u.v. c.d. from normal proteins.[130] The difference consists in a red-shift of the π–π^* amide transition and a broadening of the n–π^* transition. This deviation may result from the presence of distorted helices, caused by lipid–protein interactions. On the other hand, the observed changes may be due to light-scattering and absorption flattening due to the large size of the membrane particles.[131] It has been shown that the far-u.v. c.d. of membranes is blue-shifted and increased by dispersion into smaller particles by some irradiation.[131]

Human erythrocyte membranes have been found in one case not to show this anomalous c.d. spectrum, but show a typical c.d. with about 30% helix content.[132] Binding of human growth hormone to these

[124] J. Hermans jun., D. Puett, and G. Acampora, *Biochemistry*, 1969, **8**, 22.
[125] A. Scanu and R. Hirz, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 890.
[126] A. M. Scanu, *Biochim. Biophys. Acta*, 1969, **181**, 268.
[127] A. M. Scanu, H. Pollard, R. Hirz, and K. Kothary, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 171.
[128] D. G. Dearborn and D. B. Wetlaufer, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 179.
[129] A. M. Gotto, R. I. Levy, and D. S. Frederickson, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 1436.
[130] A. S. Gordon, D. F. H. Wallach, and J. H. Strauss, *Biochim. Biophys. Acta*, 1969, **183**, 405.
[131] T. H. Ji and D. W. Urry, *Biochem. Biophys. Res. Comm.*, 1969, **34**, 404.
[132] M. Sonnenberg, *Biochem. Biophys. Res. Comm.*, 1969, **36**, 450.

membranes gives a conformational change amounting to a loss of about 8% of the helical structure, which does not occur on the addition of bovine growth hormone or bovine serum albumin to the membrane.[132]

I. Native and Modified Proteins.—Helical contents of about 100% are reported for light meromyosin, tropomyosin, paramyosin,[133] and for an organ-specific liver membrane protein.[134] Heavy meromyosin appears to have a helix content of 45—48%,[133] which is not altered by the addition of ATP,[135] although sedimentation shows that small structural changes have occurred. Rhodopsin contains about 60% of helix, one-fifth of which is irreversibly lost on bleaching,[136] and silk sericine appears to contain some β-structure.[137] An α–β conformational change is shown when uridine 5′-phosphate or D-fucose is added to epimerase,[138] and a random coil to β change occurs in phosvitin when the *O*-phosphoserine groups are neutralised at low pH.[139] No conformational change is found when papain changes from an inactive to active state,[140] nor is there any change in the conformation of pituitary growth hormone when the disulphides are reduced and alkylated.[141]

Removal of the zinc from alkaline phosphatase produces no conformational changes in the enzyme subunits,[109] but replacement of the zinc by cobalt produces an active enzyme with optically active visible absorptions.[142] Binding of phosphate affects the c.d. spectrum of these bands, but does not produce changes in the overall enzyme conformation: by gaussian fitting it was found that phosphate does not introduce new bands into the cobalt transitions, but changes the oscillator strengths and rotational strengths of already existing bands. Removal of copper from caeruloplasmin similarly shows little effect on the backbone structure of the enzyme subunits.[143]

Modification of fetuin by digestion of the non-protein portion with neuraminidase does not greatly alter the conformation of the molecule, although there are only small amounts of helix, which can be estimated at 11%[144] or 16%[145] Dodecyl sulphate appears to increase the helix content,[145] and over the pH range 8·0—4·5, the reduced and carboxymethylated derivatives show a random coil to β transition.

Ribonucleases A and S have identical c.d. spectra over the range 198—300 nm, but ribonuclease S-protein has a different c.d. spectrum, with

[133] K. Oikawa, C. M. Kay, and W. D. McCubbin, *Biochim. Biophys. Acta*, 1968, **168**, 164.
[134] D. M. Neville, jun., *Biochem. Biophys. Res. Comm.*, 1969, **34**, 60.
[135] W. B. Gratzer and S. Lowey, *J. Biol. Chem.*, 1969, **244**, 22.
[136] H. Schichi, M. S. Lewis, F. Irreverre, and A. L. Stone, *J. Biol. Chem.*, 1969, **244**, 529.
[137] E. Iizuka, *Biochim. Biophys. Acta*, 1969, **181**, 477.
[138] A. W. Bertland and H. M. Kalckar, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 629.
[139] G. Taborsky, *J. Biol. Chem.*, 1968, **243**, 6014.
[140] A. O. Barel and A. N. Glazer, *J. Biol. Chem.*, 1969, **244**, 268.
[141] T. A. Bewley, J. Brovetto-Cruz, and C. H. Li, *Biochemistry*, 1969, **8**, 4701.
[142] M. L. Applebury and J. E. Coleman, *J. Biol. Chem.*, 1969, **244**, 709.
[143] Y. Hibino, T. Samejima, S. Kajiyama, and Y. Nosoh, *Arch. Biochem. Biophys.*, 1969, **130**, 617.
[144] Y. Oshiro and E. H. Eylar, *Arch. Biochem. Biophys.*, 1969, **130**, 227.
[145] A. C. Murray, K. Oikawa, and C. M. Kay, *Biochim. Biophys. Acta*, 1969, **175**, 331.

diminished intensity in the far-u.v.,[146] and a shift of the 240 nm extremum to the blue by 3 nm.[146, 147] The far-u.v. bands of ribonuclease A can be fitted to give 11·5% helix and 33% β-structure, if it is assumed that there is a positive band at 226 nm. Peptides 1—8, 1—13, 1—15, and 1—20, when isolated, show evidence for helical structure, and the c.d. spectra of the three longer peptides are essentially similar at low temperatures, suggesting that the residues 2—12 are as helical in these fragments as in native ribonuclease.[148] The positive extremum at 240 nm in ribonuclease may be due to titratable tyrosines, but this could not be conclusively established.[146] This band changes linearly in magnitude between 15 and 50 °C, and then decreases sharply[147] above 50 °C; on cooling from 65 °C, the 240 nm extremum is not fully recovered, although the enzyme is fully active. A similar situation exists for ribonuclease which has been reduced in 8M urea and reoxidised in the absence of mercaptoethanol.[118]

The negative c.d. extremum at 275 nm of ribonuclease A is temperature-sensitive above 50 °C, but is obtained reversibly on cooling.[147] Ribonuclease S does not show this reversible behaviour. It is assumed that the bands at 275 and 240 nm do not arise from the same residues, since their temperature dependences are different. It is not conclusively established whether the 275 nm band is derived from titratable tyrosines,[146] but it has been suggested, from studies on nitrated ribonuclease, that the effect arises from buried tyrosines at neutral pH,[149] and that the red-shift obtained on raising the pH may be due to titratable residues which acquire optical activity when ionised.

A comparison between *N*-acetyl-L-tryptophanamide and chymotrypsinogen shows that the tryptophan residues responsible for the optical activity of the protein in the near-u.v. are rigidly positioned, since their optical activity varies by only a small amount between 238 and 77 K, compared with the large variation of the c.d. of the model compound over this temperature range.[36] Modification of the two tyrosine residues in chymotrypsinogen accessible to cyanuric fluoride has no effect on the near-u.v. c.d. spectrum, and a similar result is obtained when seven tyrosine residues in trypsinogen are modified by the same reagent.[150] Ovalbumin, however, when it has reacted with cyanuric fluoride, shows marked changes in the near-u.v. c.d. spectrum, depending on how many of the five available tyrosine residues are attacked. Alkylation of chymotrypsinogen at methionine-192 with phenacyl bromide causes the formation of a new optically active absorption band in the region of 290—365 nm, which is present in neither the enzyme nor the reagent.[151] By studies of model compounds this band is found to be due to the formation of an ylide.

[146] M. N. Pflumm and S. Beychok, *J. Biol. Chem.*, 1969, **244**, 1692.
[147] E. R. Simons, E. G. Schneider, and E. R. Blout, *J. Biol. Chem.*, 1969, **244**, 4023.
[148] J. E. Brown and W. A. Klee, *Biochemistry*, 1969, **8**, 2876.
[149] G. H. Beaven and W. B. Gratzer, *Biochim. Biophys. Acta*, 1968, **168**, 456.
[150] M. J. Gorbunoff, *Biochemistry*, 1969, **8**, 2591.
[151] D. S. Sigman, D. A. Torchia, and E. R. Blout, *Biochemistry*, 1969, **8**, 4560.

Deacylation of trimethylacetyl-chymotrypsin produces a conformational change, or possibly a change in the perturbation of aromatic or cystinyl residues in the active site of the enzyme,[152] and it has been shown from a study of the c.d. spectrum that at high pH the carboxy-group of aspartic acid 194 (which forms an ion-pair with isoleucine-16 at neutral pH) may protrude into the active site and interfere with substrate binding.[153]

Carbamylation and succinylation of pepsinogen alter the conformation of the molecule by destroying interactions between lysine residues (the site of the modification) and dicarboxylic acids which stabilise the native molecule.[154] Changing the pH also causes conformational changes, shifting the far-u.v. c.d. spectrum to shorter wavelengths as the pH is increased.

Nitration of carbonic anhydrases B and C yields products with the same o.r.d. and c.d. as the native enzymes, apart from the introduction of new bands at 420—430 nm: no change is produced in the enzyme activity.[121] However, iodination of anhydrase B decreases esterase activity and causes a change in the active structure of the molecule.

p-Azobenzenearsonate couples with tyrosine and lysine, principally with the former, and this property has been used to modify carboxypeptidase-A. The modified enzyme has similar physico-chemical properties to the native enzyme, but new optically active bands are found in the absorption spectrum.[155] Denaturation of the enzyme with guanidine produces a c.d. spectrum similar to that of the model compound arsanilazotyrosine, and the c.d. spectrum of the modified protein is affected by substrate, inhibitors, and the removal of zinc. The modification thus produces a probe which is sensitive to the conformation of the active site.

Tyrosyl residues in insulin and β-lactoglobulin are susceptible to modification by *N*-acetylimidazole, and the effects of the reagent in altering the near-u.v. c.d. spectra have been followed. The change in the insulin c.d. spectrum when the octapeptide B-23 to B-30 is removed by trypsin amounts to about 20% of the original at 275 nm. This effect must result from the removal of the (previously buried) tyrosine B-26.[156] Although acetylation of two accessible tyrosines in β-lactoglobulin results in only small changes in the c.d. spectrum, acetylation of a third, buried, residue almost destroys the optical activity above 260 nm, and causes changes in the backbone conformation in addition.[157] Studies on the backbone conformation of a genetic variant of lactoglobulin have shown that there is no difference from native lactoglobulin in conformation.[158]

[152] M. Volini and P. Tobias, *J. Biol. Chem.*, 1969, **244**, 5105.
[153] J. McConn, G. D. Fasman, and G. P. Hess, *J. Mol. Biol.*, 1969, **39**, 551.
[154] K. Grizzuti and G. E. Perlmann, *J. Biol. Chem.*, 1969, **244**, 1764.
[155] H. M. Kagan and B. L. Vallee, *Biochemistry*, 1969, **8**, 4223.
[156] C. J. Menendez and T. T. Herskovits, *Biochemistry*, 1969, **8**, 5052.
[157] R. Townend, T. T. Herskovits, S. N. Timasheff, and M. J. Gorbunoff, *Arch. Biochem. Biophys.*, 1969, **129**, 567.
[158] G. Brignon, R. Ribadeau-Dumas, J. Garnier, D. Pantaloni, S. Guinand, J. G. Basch, and S. N. Timasheff, *Arch. Biochem. Biophys.*, 1969, **129**, 720.

Neurotoxins from two snake species, cobra and *Bungarus*, show distinct similarities in their c.d. spectra.[159, 160] They both show evidence of β-structure, and show a positive extremum at 228—232 nm, which may be due to tyrosine (*cf.* ref. 126). This positive extremum is destroyed in *S*-carboxymethylated cobrotoxin,[159] and the c.d. spectrum is consistent with that of a protein which is mainly unordered in conformation, but possesses small amounts of α- and β-structure.

Kunitz trypsin inhibitor shows bands in the near- and far-u.v. c.d. spectrum, the latter giving little evidence for ordered backbone structure. This is confirmed by the small changes induced by succinylation, acetylation and denaturation with 6M guanidinium hydrochloride.[161] A positive band at 226 nm is shown in the c.d. spectrum, which is destroyed by sulphitolysis,[161] or by acetylation of all four tyrosines in the molecule in 8M urea, a procedure which prevents refolding of the protein.[162] The bands in the near-u.v. c.d. are reduced by the acetylation of the one surface tyrosine or by sulphitolysis, and the reaction of the accessible tryptophan residue with *N*-bromosuccinimide destroys the c.d. about 294 nm, showing that the tryptophan residue is the source of this contribution.

4 Nuclear Magnetic Resonance

contributed by R. Henson

Although high resolution proton magnetic resonance (p.m.r.) has again made the greatest number of n.m.r. contributions to interaction studies, halide, alkali-metal, and ^{13}C probes have also been used effectively. The smaller the molecular weight of the molecule, generally the easier is the assignment of the resonances to particular nuclei, so it is convenient to discuss the applications of n.m.r. in order of increasing molecular weight.

There has been a little work on amino-acid systems[163—165] and the structures of manganese(II) and copper(II) histidine complexes have been proposed.[166] With a view to complementing the deuterium substitution technique[167, 168] for simplifying proton spectra, ^{13}C spectra of enriched amino-acids have been obtained and compared with the corresponding proton spectra.[169]

[159] C. C. Yang, C. C. Chang, K. Hayaishi, T. Suzuki, K. Ikeda, and K. Hamaguchi, *Biochim. Biophys. Acta*, 1968, **168**, 373.
[160] K. Hamaguchi, K. Ikeda, and C.-Y. Lee, *J. Biochem. (Japan)*, 1968, **64**, 503.
[161] B. Jirgensons, M. Kawabata, and S. Capetillo, *Makromol. Chem.*, 1969, **125**, 126.
[162] M. Baba, K. Hamaguchi, and T. Ikenaka, *J. Biochem. (Japan)*, 1969, **65**, 113.
[163] T. L. James and J. H. Noggle, *J. Amer. Chem. Soc.*, 1969, **91**, 3424.
[164] W. H. Pirkle and S. D. Beare, *J. Amer. Chem. Soc.*, 1969, **91**, 5150.
[165] R. Garner and W. B. Watkins, *Chem. Comm.*, 1969, **8**, 386.
[166] H. Sigel, R. Griesser, and D. B. McCormick, *Arch. Biochem. Biophys.*, 1969, **134**, 217.
[167] H. L. Crespi, H. F. Daboll, and J. J. Katz, *Biochim. Biophys. Acta*, 1970, **200**, 266.
[168] J. L. Markley, I. Putter, and O. Jardetzky, *Z. analyt. Chem.*, 1968, **243**, 367.
[169] W. Horsley, H. Sternlicht, and J. S. Cohen, *Biochem. Biophys. Res. Comm.*, 1969, **37**, 47.

I.r. and p.m.r. spectra of alanine dipeptides and their *N*-methyl derivatives indicate that the dipeptides are about 70% in the intramolecularly hydrogen-bonded form, and analysis of the spin–spin coupling constants suggests a *cis* arrangement of the CH and NH hydrogens [170] in the CONH–CH moiety. The variation of J_{NH-CH} with dihedral angle has served as a basis for discussing the conformation of the extended form of dipeptides in polar solvents.[170, 171] In small glycine-containing peptides conformational preferences cause the α-protons to be non-equivalent,[172] while aryl methyl side-chains of histidine and tryptophan in cyclic dipeptides stack on to the face of the dioxopiperazine ring.[173] Of the six amide protons in certain cyclic hexapeptides, two are shielded from the solvent and four are exposed to it.[174–176] Similar conformational studies on the dependence on temperature of the chemical shifts of the amide proton have been made on gramicidin S, valinomycin,[177] and actinomycin D,[178] and Conti [179] has revised his previous analysis of gramicidin S after measuring spectra at 220 MHz and effecting double- and triple-resonance experiments. Broadening of the proton resonance of the lactate methyl group in valinomycin has yielded preliminary rates for the complex formation with potassium ions [180] and well-integrated n.m.r., o.r.d., and i.r. experiments [181] have provided the conformations in the pure and complexed forms, of which the latter resembles the potassium complex of nonactin.[182] The variation of tyrocidin B spectra with concentration [183] is explained in terms of association: the intact ring is necessary for this association but not the aromatic residues, and protection of polar groups, for example succinylation of the ornithine basic group, failed to eliminate association properties; this indicated predominantly hydrophobic bonding in the agglomerate.

The *gauche* conformer obtained by rotation around the CH_2—CH_2 bond has a lower enthalpy than the *trans* conformer in poly-β-alanine, which suggests that the CO and NH groups undergo some kind of positive interaction in trichloroacetic acid solution. Measurements [184] were made

[170] V. F. Bystrov, S. L. Portnova, V. I. Tsetlin, V. T. Ivanov, and Yu. A. Ovchinnikov, *Tetrahedron*, 1969, **25**, 493.
[171] C. M. Thong, D. Canet, P. Granger, M. Marraud, and J. Néel, *Compt. rend.*, 1969, **269**, *C*, 580.
[172] K. D. Kopple, T. Saito, and M. Ohnishi, *J. Org. Chem.*, 1969, **34**, 1631.
[173] K. D. Kopple and M. Ohnishi, *J. Amer. Chem. Soc.*, 1969, **91**, 962.
[174] K. D. Kopple, M. Ohnishi, and A. Go, *J. Amer. Chem. Soc.*, 1969, **91**, 4264.
[175] K. D. Kopple, M. Ohnishi, and A. Go, *Biochemistry*, 1969, **8**, 4087.
[176] R. Schwyzer and U. Ludescher, *Helv. Chim. Acta*, 1969, **52**, 2033.
[177] M. Ohnishi and D. W. Urry, *Biochem. Biophys. Res. Comm.*, 1969, **36**, 194.
[178] T. A. Victor, F. E. Hruska, K. Hikichi, S. S. Danyluk, and C. L. Bell, *Nature*, 1969, **223**, 302.
[179] F. Conti, *Nature*, 1969, **221**, 777.
[180] D. H. Haynes, A. Kowalsky, and B. C. Pressman, *J. Biol. Chem.*, 1969, **244**, 502.
[181] V. T. Ivanov, I. A. Laine, N. D. Abdlaev, L. B. Senyavina, E. M. Popov, Yu. A. Ovchinnikov, and M. M. Shemyakin, *Biochem. Biophys. Res. Comm.*, 1969, **34**, 803.
[182] J. H. Prestegard and S. I. Chan, *Biochemistry*, 1969, **8**, 3921.
[183] A. Stern, W. A. Gibbons, and L. C. Craig, *J. Amer. Chem. Soc.*, 1969, **91**, 2794.
[184] M. Guaita and L. F. Thomas, *Makromol. Chem.*, 1968, **119**, 113.

at 116 °C, because at lower temperatures the signals were too broad owing to slow segmental motions, and this made it impossible to measure the corresponding ΔH value of the thermally more labile poly-α-methyl-β-alanine. Spectra of poly-L-alanine in chloroform–trifluoroacetic acid mixtures measured at 220 MHz under various conditions of temperature and concentration contained separate lines for helix and random coil α-CH and peptide NH protons, thereby enabling the helix content to be determined directly.[185] The two observed peaks for α-CH protons are assigned to two different structures in poly-L-proline [186] on the basis of their relative stability in different solvents. Only one broadened peak at 4·6 p.p.m. was observed on dissolution in acetic acid, but after 43 min two signals are apparent at 4·6 and 4·8 p.p.m., and after 10 days the first peak had decreased in intensity to insignificance. Thus structure II is the only one stable in formic acid as only one peak is present at 4·8 p.p.m. Bradbury [187] suggests that the down-field shifts accompanying the helix-to-coil transition of poly-γ-benzyl-L-glutamate in strong acid is due, first, to the collapse of the helix to uncharged random coil and, secondly, to the charging of the amide groups of the random coil. The linewidths of the methyl resonances in denatured proteins are usually constant for a particular solvent and independent of molecular weight. However, for catalase and fibrinogen in dichloroacetic acid, for example,[188] broadening indicates incomplete unfolding and/or association. Poly-L-alanine and poly-DL-alanine show broadened lines at low temperatures; it is denied that this indicates a conformational change, but it is thought to be caused by aggregation.[189]

Studies made at 60, 100, and 220 MHz of α_{S1}-, β-, and κ-caseins have been reported.[190–193] These are highly aggregated systems but the observation of n.m.r. signals implies some internal flexibility. Spectra of β- and κ-casein differ markedly since signals from the aromatic and terminal methyl protons are fully developed only after heating or treatment with urea. These groups are restricted in motion in contrast to the hydrophilic groups.

Resonances in lysozyme and ribonuclease have been assigned by studying their pH dependence,[168, 194, 196–198] by selectively deuteriating the enzyme,[167, 168] by simulating spectra comprising overlapping lines [168, 195] by

[185] J. A. Ferretti and L. Paolillo, *Biopolymers*, 1969, **7**, 155.
[186] F. Conti, M. Piatelli, and P. Viglino, *Biopolymers*, 1969, **7**, 411.
[187] J. H. Bradbury and M. D. Penn, *Austral. J. Chem.*, 1969, **22**, 357.
[188] J. H. Bradbury and N. L. R. King, *Austral. J. Chem.*, 1969, **22**, 1083.
[189] A. Takahashi, L. Mandelkern, and R. E. Glick, *Biochemistry*, 1969, **8**, 1673.
[190] R. B. Leslie, L. Irons, and D. Chapman, *Biochem. Biophys. Acta*, 1969, **188**, 237.
[191] S. H. Koenig and W. E. Schiller, *J. Biol. Chem.*, 1969, **244**, 3283.
[192] W. J. O'Sullivan, R. Virden, and S. Blethan, *European J. Biochem.*, 1969, **8**, 562.
[193] P. G. Schmidt, G. R. Stark, and J. D. Baldeschwieler, *J. Biol. Chem.*, 1969, **244**, 1860.
[194] H. Rüterjans, H. Witzel and O. Pongs, *Biochem. Biophys. Res. Comm.*, 1969, **37**, 247.
[195] C. C. McDonald and W. D. Phillips, *J. Amer. Chem. Soc.*, 1969, **91**, 1513.
[196] G. C. K. Roberts, D. H. Meadows, and O. Jardetzky, *Biochemistry*, 1969, **8**, 2053
[197] H. Rüterjans and H. Witzel, *European J. Biochem.*, 1969, **9**, 118.
[198] J. S. Cohen, *Nature*, 1969, **223**, 43.

means of a digital computer, and by studying the stepwise denaturation of these proteins.[168] Selective deuteriation is potentially a very important method for assigning particular protons to peaks under the envelope of the peaks due to aromatic groups. Jardetzky *et al.*[168] have shown that the up-field shifts of the tryptophan multiplet of staphylococcal nuclease are consistent with fluorescence results in indicating that the tryptophan is buried in a stack of aromatic rings. The variations with pH of the C-2 H shifts of histidine in ribonuclease[194] and lysozyme[198] are caused by the proximity of charged groups or hydrogen-bonding. With lysozyme, differences in the histidine titration curves for human and hen egg-white lysozyme and similar activities towards the same substrate indicate that the histidine residue has no functional significance.

The binding of *N*-acetyl-β-D-glucosamine to lysozyme is appreciably different from that of other monosaccharide inhibitors and the rate constant for binding is different from that of di-*N*-acetyl-D-glucosamine.[199] These results confirm conclusions from temperature-jump experiments. Sykes[200] has also determined the rates of formation of complexes of lysozyme with *N*-acetyl α- and β-D-glucosamine as well as the corresponding activation parameters: the results indicate that this is not a simple bimolecular process and that the anomers bind at the same site and in a similar configuration. Similar studies[201] have been made of the binding of trifluoroacetyl-D-phenylalanine to α-chymotrypsin, DFP-treated chymotrypsin and chymotrypsinogen.*

Raftery *et al.*[202, 203] have continued a study of binding of inhibitors and related substrates to lysozyme and the results are in good agreement with *X*-ray studies. Three contiguous sub-sites to which acetamidopyranoside residues bind have been detected in the enzyme, and the relative modes of binding of the inhibitors have been described on the basis of the chemical shifts observed.

McDonald and Phillips[204] report a technique for displacing some of the peaks from the main mixed band. Initially they observed the 220 MHz spectrum of lysozyme of hen egg-white, and the peaks at the extremity of the spectrum were assigned to CH protons of those amino-acid side-chains that are close to the faces of the conjugated rings of aromatic side-chains, since they move to high field positions. Relatively few protons contribute to this region which is therefore amenable to detailed analysis, although some resonances still overlap. Since addition of cobalt(II) ions to the protein solution perturbs resonances to high and low field, and the shifts

[199] B. D. Sykes and C. Parravano, *J. Biol. Chem.*, 1969, **244**, 3900.
[200] B. D. Sykes, *Biochemistry*, 1969, **8**, 1110.
[201] B. D. Sykes, *J. Amer. Chem. Soc.*, 1969, **91**, 949.
[202] M. A. Raftery, F. W. Dahlquist, S. M. Parsons, and R. G. Wolcott, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 44.
[203] F. W. Dahlquist and M. A. Raftery, *Biochemistry*, 1969, **8**, 713.
[204] C. C. McDonald and W. D. Phillips, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 43.

* DFP = di-isopropyl phosphofluoridate.

increase regularly with cobalt concentration, such addition aids assignment of previously unresolved peaks, especially as circular dichroism studies indicate no conformational changes on addition of Co^{2+}. From a plot of the resonance position against concentration of Co^{2+}, an association constant can be determined and it was concluded that the binding of cobalt occurred mainly at a single site. These contact shifts are strongly dependent on pH and the site probably involves aspartic or glutamic acids. Peaks due to Leu-17, the methyl group of Ile-98, Leu-8, and Met-105 have been assigned.

Effects of paramagnetic ions are also observed in haemoglobins and myoglobin.[205, 206] Shifts due to unpaired electrons and ring currents can be distinguished by means of temperature studies since the former varies as the reciprocal of temperature. The contact coupling constants of the aromatic protons and methyl groups attached to the aromatic ring are related to the unpaired electron density on the neighbouring carbon atom so that spin densities at various points in the haem group can be derived from the hyperfine interaction. Other studies showed that no change occurred in the haem groups themselves when neighbouring haem groups bound ions and consequent conformational changes had occurred. It is such changes in tertiary structure which are related to the co-operativity of oxygen binding in Koshland's model; the ligand-induced changes are even propagated into neighbouring subunits but not as far as the next haem.

Diamagnetic effects of the haem ring in ferrocytochrome *c* cause shifts in resonances that have been assigned to Met-80 in the sixth co-ordination position, and the proximity of Ile-81 to the ring is indicated.[207]

The dependence on magnetic field of the spin–lattice relaxation rate of solvent protons in solutions of diamagnetic apotransferrin has been measured as a function of concentration, temperature, and pH, and the results have been compared with the predictions of several models for the interaction of solvent protons with protein molecules. Proton exchange between bulk water and protein surface occurs through the exchange of water molecules at a pH between 5 and 8, and, at higher pH, proton exchange with basic amino-acid residues may be important. The theory shows that water need not be irrotationally bound to the protein and also that the variation of the spin–lattice relaxation rate is related to the rotational diffusion relaxation time.[191]

[205] D. G. Davis, N. L. Mock, and V. R. Laman, C.Ho, *J. Mol. Biol.*, 1969, **40**, 311.

[206] K. Wüthrich, R. G. Shulman, and T. Yamane, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 1199; R. G. Shulman, K. Wüthrich, T. Yamane, E. Antonini, and M. Brunori, *ibid.*, 1969, **62**, 623; K. Wüthrich, R. G. Shulman, B. J. Wyluda, and W. S. Caughey, *ibid.*, p. 636; R. G. Shulman, S. Ogawa, K. Wüthrich, T. Yamane, J. Peisach, and W. E. Blumberg, *Science*, 1969, **165**, 251.

[207] C. C. McDonald and W. D. Phillips, *Biochem. Biophys. Res. Comm.*, 1969, **36**, 442.

Proton relaxation enhancement studies[192] show that $MnATP^-$ and $MnATP^{2-}$, rather than Mn^{2+}, bind to arginine kinase, and an observed correlation between the enhancement and values of the maximum velocities of binding for various nucleotides could be indicative of a gradation in conformational changes induced by substrate. Linewidths for protons linked to carbon in analogues of carbamyl phosphate in the presence of the catalytic subunit of aspartate transcarbamylase measure the rotational freedom of the bound analogue. The linewidths of methylene protons of succinate, a competitive inhibitor, increased a little when catalytic subunit was added, but much more upon further addition of carbamyl phosphate, which also caused an increase in the lifetime of the succinate–enzyme complex.[193]

The halide ion probe is a sensitive method for detecting binding to some proteins.[208, 209] A comparison of linewidth of ^{35}Cl resonance for carboxypeptidase A and bovine serum mercaptalbumin in the presence of mercuric ions before and after dialysis showed that the 'cystine' residues in the former must be linked by a disulphide bond; the possibility of a sulphydryl group at the active site, masked by the zinc atom, was eliminated by repetition of the experiment with apo-enzyme.[210] This result was confirmed by *X*-ray studies. Zinc and cadmium ions have also been suggested for this type of assay.[211] Cottam and Ward[212] found that four zinc atoms bound per molecule of pyruvate kinase, that all four sites had the same association constant of $0{\cdot}81 \pm 0{\cdot}07 M^{-1}$. Addition of bicarbonate ion, ATP, or ADP to the zinc–pyruvate kinase binary complex decreased the ^{35}Cl linewidth and this could be explained in terms of the formation of a ternary complex or of conformational changes. Mildvan and Cohn have already shown that Mn^{2+} binds to two sites in pyruvate kinase when K^+ is the activating cation, but when a non-activating cation is present, *e.g.* NMe_4^+, then Mn^{2+} occupies four sites.

A comparison of metallo-enzymes with metal-free proteins[213] gives evidence that $^{81}Br^-$ binds independently of metallic sites. The broadening was caused by electric field gradients and not by viscosity or exchange effects; these possibilities were eliminated by repeating the experiments with $^{85}Rb^+$ and $^{79}Br^-$ present, respectively. Blocking of zinc atoms in horse liver alcohol dehydrogenase by NAD had no detectable influence upon the line widths. The binding of anionic surfactants to human serum albumin has also been studied by the halide probe method.[214]

[208] R. L. Ward, *Biochemistry*, 1969, **8**, 1879.
[209] W. D. Ellis, H. B. Dunford, and J. S. Martin, *Canad. J. Biochem.*, 1969, **47**, 157.
[210] R. G. Bryant, H. J. C. Yeh, and T. R. Stengle, *Biochem. Biophys. Res. Comm.*, 1969, **37**, 603.
[211] R. G. Bryant, *J. Amer. Chem. Soc.*, 1969, **91**, 976.
[212] G. L. Cottam and R. L. Ward, *Arch. Biochem. Biophys.*, 1969, **132**, 308.
[213] M. Zeppezauer, B. Lindman, S. Forsen, and J. Lindquist, *Biochem. Biophys. Res. Comm.*, 1969, **37**, 137.
[214] G. Gillberg-La force and S. Forsen, *Biochem. Biophys. Res. Comm.*, 1970, **38**, 137.

5 Fluorescence

contributed by J. R. Brocklehurst

The many fluorescence techniques used for studying biomolecular structure make use both of the intrinsic fluorescence of the system and of the extrinsic fluorescence of added probes. (See also last year's Report.[215])

Non-covalent probes, *e.g.* 2-*p*-toluidino-naphthalene-6-sulphonate (TNS) and 1-anilino-naphthalene-8-sulphonate (ANS) are very widely used now, largely because they are not fluorescent in water, but are highly fluorescent in non-aqueous environments (*e.g.* when bound to a protein). By studying the properties of anhydrous and hydrated crystals of TNS, it has been possible to understand these properties.[216] *X*-Ray diffraction studies show that in the anhydrous crystal TNS is planar, allowing electronic delocalisation over the whole molecule, while the hydrated molecule is bent. The lack of fluorescence of aqueous TNS could be the result of hydrogen-bonding as well as of a solvent polarity effect.

Changes in the fluorescence of ANS indicate conformational changes in glutamate dehydrogenase brought about by NADH and GTP, and by diethylstilboestrol in the absence of NADH[217] (in contrast to earlier work[218]). There are two stages: a rapid change in subunit structure followed by a slow rearrangement of subunits. Use of ANS also detects the desensitisation of glutamate dehydrogenase by chemical modification of the tyrosine residues.[219] The binding of ANS to horse serum butylcholinesterase has been examined and ANS may prove to be an uncompetitive inhibitor.[220] The use of ANS for detecting structural changes has now been extended to membrane systems.[221, 222] One difficulty in using ANS is that under certain conditions it can form dimers. In the presence of nitrite (*e.g.* from nitric acid cell cleaning baths) a new fluorescent species is formed which has a larger fluorescence than ANS, and a higher affinity for binding to BSA. BSA catalyses the formation of this species (thought to be an ANS dimer).[223] The conformation of protein binding sites has been investigated by binding 6-aminonaphthalene-2-sulphonate to BSA and an antibody, and quenching the fluorescence with bromate ion. Both proteins shield the probe from the bromate, and BSA provides a much less polar binding site.[224]

[215] P. M. Hardy in 'Amino-acids, Peptides and Proteins', The Chemical Society, London, 1969, vol. I, p. 112.
[216] A. Camerman and L. H. Jensen, *Science*, 1969, **165**, 493.
[217] G. H. Dodd and G. K. Radda, *Biochem. J.*, 1969, **114**, 407.
[218] W. Thompson and K. L. Yielding, *Arch. Biochem. Biophys.*, 1968, **126**, 399.
[219] N. C. Price and G. K. Radda, *Biochem. J.*, 1969, **114**, 419.
[220] S. T. Christian and R. Janetzko, *Biochem. Biophys. Res. Comm.*, 1969, **37**, 1022.
[221] J. Vanderkooi and A. Martonosi, *Arch. Biochem. Biophys.*, 1969, **133**, 153.
[222] J. R. Brocklehurst, R. B. Freedman, D. J. Hancock, and G. K. Radda, *Biochem. J.*, 1970, **116**, 721.
[223] C.-G. Rosen and G. Weber, *Biochemistry*, 1969, **8**, 3915.
[224] M. H. Winkler, *Biochemistry*, 1969, **8**, 2586.

The covalent attachment of a fluorescent probe to a protein or polypeptide has many uses, *e.g*, tracing and identification. Such probes can also yield structural information. By studying the fluorescence depolarisation of the attached probe, the rotational relaxation time of the whole conjugate may be obtained. This gives useful information on the overall size and shape of the conjugate.[225] 5-Dimethylamino-naphthalene-1-sulphonyl (dansyl) conjugates of various proteins have been studied. A slightly modified labelling procedure has yielded beef heart lactate dehydrogenase molecules labelled with 1—4 dansyl groups. The rotational relaxation time is independent of the extent of labelling and of the protein concentration (3 μg/ml—1 mg/ml). This indicates that the protein retains its shape and size throughout the concentration range and does not dissociate as had been thought previously.[226] The state of aggregation of β-lactoglobulin A has been studied by measurement of the depolarisation of dansyl fluorescence.[227] Flexibility has been detected in immunoglobulins by this method.[228] However, the rotational relaxation times of thyroglobulin conjugates suggest that they have rigid structures (but these deviate from an anhydrous sphere model).[229]

The use of the dansyl group is restricted to proteins with molecular weights less than 150,000 because of its short fluorescence lifetime. A new long-lifetime label, pyrene-1-butyryl, has been used successfully on systems with known rotational relaxation times and has given results in agreement with previous measurements.[229, 230] Rotational relaxation times of human immunoglobulin M conjugates were consistent with molecular weight determinations before and after alkylation, and with sedimentation coefficients.[230] The rate of decay of fluorescence polarisation of dansyl-γ-globulin is biphasic. The slow phase is due to rotation of the whole molecule, while the fast phase is due to some flexibility within the molecule.[231] Similar studies on fluorescein isothiocyanate γ-globulin show that polarisation varies with the extent of labelling, due to energy transfer between the fluorophores. The rotational time of the molecule, and the presence of flexibility both agree with the dansyl results.[232]

The binding of non-covalent probes has also been studied by this technique. The orientational anisotropy (ρ) for ANS bound to BSA has been investigated: it increases with the number of bound ANS molecules for excitation at 366 nm, but not for excitation at 436 nm. This is explained in terms of depolarisation due to energy transfer below 436 nm and leads to the conclusion that there is a preferred orientation of the ANS molecules.[233]

[225] G. Weber and S. R. Anderson, *Biochemistry*, 1969, **8**, 361.
[226] S. R. Anderson, *Biochemistry*, 1969, **8**, 1394.
[227] P. Wahl and S. N. Timasheff, *Biochemistry*, 1969, **8**, 2945.
[228] Y. A. Zagyansky, R. S. Nezlin, and L. A. Tumerman, *Immunochem.*, 1969, **6**, 787.
[229] A. B. Rawitch, E. Hudson, and G. Weber, *J. Biol. Chem.*, 1969, **244**, 6543.
[230] J. A. Knopp and G. Weber, *J. Biol. Chem.*, 1969, **244**, 6309.
[231] P. Wahl, *Biochim. Biophys. Acta*, 1969, **175**, 55.
[232] M. Fayet and P. Wahl, *Biochim. Biophys. Acta*, 1969, **181**, 373.
[233] S. R. Anderson and G. Weber, *Biochemistry*, 1969, **8**, 371.

As BSA becomes saturated with ANS, the randomness of orientation of the ANS molecules decreases.[234] Small differences in binding constant for the different ANS binding sites on BSA have been detected by c.d. measurements.[235] Fresh analysis of the fluorescence measurements (which were previously thought to indicate homogeneous binding) gives a similar result.[236]

The various techniques discussed above are not without their drawbacks. For various γ-globulin conjugates, the rotational relaxation time varies with the particular fluorophore used. This is partly due to the contribution of the rotation of the probe to the observed rotation time.[237] Many workers have used average values for the fluorescence lifetime in their calculations, not realising the importance of changes in the lifetime brought about by the particular system under examination. If lifetime variation is not excluded, a linear Perrin plot does not necessarily mean the absence of intramolecular rotation.[238] One of the major difficulties of the use of extrinsic probes is the possibility of disturbing the biological system by the introduction of the probe. Thus dansylation of BSA changes the BSA–anti-BSA binding constant by a factor of 2.[239]

By using the intrinsic fluorescence of the biological system this last difficulty is overcome. The fluorescence may be that of a coenzyme, or that of the protein, polypeptide, membrane, *etc.*, itself. The binding of NADH to liver alcohol dehydrogenase has been studied by following the enhancement of NADH fluorescence by D_2O. This enhancement is reduced by liver alcohol dehydrogenase and is further reduced when inhibitor is added.[240] Flavinyl peptides containing aromatic amino-acids have been made as model compounds for flavoproteins, and the relationship between fluorescence quenching (both of the flavin and of the aromatic amino-acid) and interfluorophore distance has been studied.[241]

With the development of more sophisticated spectrofluorimeters it has become possible to distinguish readily between tyrosine and tryptophan fluorescence in proteins and polypeptides. When G-actin is polymerised, the tyrosine fluorescence is unchanged while that of tryptophan decreases by 10—13%. This change is not due to changes in ionic strength.[242] Two groups at the active site of papain have been identified by studying the dependence on pH of the tyrosine and tryptophan fluorescence of active and unactivated papain.[243]

[234] B. Witholt and L. Brand, *Fed. Proc.*, 1969, **28**, 470.
[235] S. R. Anderson, *Biochemistry*, 1969, **8**, 4841.
[236] T. L. Pasby and G. Weber, *Biochemistry*, 1970, in the press.
[237] R. F. Chen, *Arch. Biochem. Biophys.*, 1969, **133**, 263.
[238] P. Johnson and A. C. R. Thornton, *Biochem. J.*, 1969, **115**, 725.
[239] F. Kierszenbaum, J. Dandliker, and W. B. Dandliker, *Immunochem.*, 1969, **6**, 125.
[240] C. H. Blomquist, *J. Biol. Chem.*, 1969, **244**, 1605.
[241] R. E. MacKenzie, W. Föry, and D. B. McCormick, *Biochemistry*, 1969, **8**, 1839.
[242] N. S. Vedenkina, M. N. Ivkova, V. N. Leonova, and E. A. Burshtein, *Biofizika*, 1968, **13**, 847.
[243] A. O. Barel and A. N. Glazer, *J. Biol. Chem.*, 1969, **244**, 268.

Two series of model peptides, one with both tryptophan and tyrosine, and the other with just tyrosine, have been studied to find out how pH, salt concentration, temperature, *etc.*, affect the fluorescence of the two amino-acids separately and together.[244] Similar studies on polypeptide hormones have revealed that they have no precise conformation in solution.[245] The same conclusion was reached by studying the fluorescence polarisation of tryptophan in these hormones and in synthetic polypeptides.[246]

When glutamate and lactate dehydrogenases are stoicheiometrically inactivated by silver nitrate the protein fluorescence decreases simultaneously. It is suggested that this is due to small structural changes at the active sites.[247, 248] Similarly, titration of the —SH groups of uridine-diphosphate-galactose-4-epimerase with *p*-chloromercuribenzoate abolishes the activity and diminishes the protein fluorescence by 40%. Complete titration of the available —SH groups abolishes the fluorescence too.[249] The binding of pyridoxamine-5-phosphate to the apo-enzyme aspartate aminotransferase is accompanied by a biphasic quenching of the protein fluorescence. The slow phase parallels the recovery of activity and may be associated with a conformational change.[250] When Cu^{2+} and Fe^{2+} bind to transferrin, the protein fluorescence is quenched by long-range transfer of energy to the absorption bands of the metal–protein interaction.[251]

Large-scale structural changes, such as denaturation, produce very marked effects on protein fluorescence. For an extensive series of proteins the quantum yields of fluorescence vary from 0·05 to 0·275 for native proteins and 0·07 to 0·166 for denatured proteins.[252] Increasing concentrations of dioxan bring about considerable structural changes as detected by the nitrate quenching of tryptophan fluorescence in chymotrypsinogen, chymotrypsin, and trypsin. Above 70% dioxan the protein appears to invert to a hydrophilic interior and a hydrophobic exterior.[253]

The polarisation of fluorescence of tryptophan in proteins is a good measure of its mobility. In solution the dependence of polarisation on temperature (2—60 °C) of a large number of proteins shows discontinuities which may have biological significance.[254] In crystalline proteins

[244a] H. Edelhoch, R. L. Perlman, and M. Wilchek, *Ann. New York Acad. Sci.*, 1969, **158**, 391; [b] R. W. Cowgill, *Biochim. Biophys. Acta*, 1968, **168**, 417–439.
[245] H. Edelhoch and R. E. Lippoldt, *J. Biol. Chem.*, 1969, **244**, 3876.
[246] R. S. Bernstein, M. Wilchek, and N. Edelhoch, *J. Biol. Chem.*, 1969, **244**, 4398.
[247] K. S. Rogers, *Enzymologia*, 1969, **36**, 153.
[248] K. S. Rogers, *Enzymologia*, 1969, **37**, 174.
[249] A. N. Bhaduri, *Science and Culture*, 1968, **34**, 49.
[250] J. E. Churchich and J. G. Farrelly, *J. Biol. Chem.*, 1969, **244**, 3685.
[251] S. S. Lehrer, *J. Biol. Chem.*, 1969, **244**, 3613.
[252] M. J. Kronman and L. G. Holmes, *Fed. Proc.*, 1969, **28**, 470.
[253] M. N. Ivkova, V. V. Mosolov, and E. A. Burshtein, *Molec. Biol.* (*U.S.S.R.*, *Eng. Trans.*), 1968, **2**, 661.
[254] S. V. Konev, V. M. Mazhul, and E. A. Chernitskii, *Doklady Acad. Nauk. Beloruss. S.S.R.*, 1968, **12**, 1122.

similar measurements reveal different mobilities of tryptophan in different proteins.[255]

Studies of protein fluorescence *in situ* may in the future provide a great deal of information about protein interactions. The changes in tryptophan fluorescence of myosin ATPase in frog tissue exactly parallel changes in activity induced by urea.[256] The polarisation of tryptophan fluorescence in muscle fibres depends on the orientation of the fibre in the plane of polarisation of the exciting light: $p_{\perp}$ varies with the state of tension of the muscle while $p_{\parallel}$ is constant. This indicates changes in the average orientation of the tryptophan molecules in muscle protein during contraction and relaxation.[257] The development of a spectrofluorimetric microscope able to measure fluorescence (and phosphorescence) spectra of specimens as small as 1 μg will prove useful in investigations of this type.[258]

Measurement of energy transfer between chromophores can yield information about the inter-chromophore distances. This is a potentially powerful tool in studying protein and peptide structure. Energy transfer in poly-L-tyrosine has been studied as a function of phenolic ionisation.[259] Proflavin bound to α chymotrypsin acts as a singlet acceptor for triplet tryptophan. Two types of tryptophan can be detected, and this technique will prove useful in the 15—60 Å range of distances.[260] Fluorescent substrate analogues covalently attached to pepsin accept energy from the tryptophan groups with different efficiencies possibly indicating a non-uniform distribution of tryptophan about the active site.[261] Myosin A fluorescence is quenched by bound ANS (by energy transfer). The transfer efficiency is larger in the native protein than in a modified version, indicating a small conformational change (which is not detectable by o.r.d. in such a large molecule).[262] Energy transfer from BSA to bound chlorotriazine dyes gives inter-chromophore distances which change when the conformation is modified (*e.g.* by heat).[263]

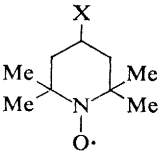
6 Spin Labels

contributed by N. C. Price

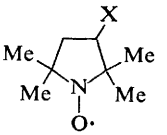
The term 'spin label' was first used by McConnell and his co-workers [264] to describe the very stable nitroxide radicals prepared by Rozantzev.[265] Spin

[255] S. V. Konev, V. M. Mazhul, and E. A. Chernitskii, *Doklady Acad. Nauk. S.S.S.R.*, 1968, **183**, 1201.
[256] S. V. Konev, E. A. Chernitskii, and E. I. Lin, *Biofizika*, 1968, **13**, 1040.
[257] J. F. Aronson and M. F. Morales, *Biochemistry*, 1969, **8**, 4517.
[258] C. A. Parker, *Analyst*. 1969, **94**, 161.
[259] J. J. ten Bosch, J. W. Longworth, and R. O. Rahn, *Biochim. Biophys. Acta*, 1969, **175**, 10.
[260] W. C. Galley and L. Stryer, *Biochemistry*, 1969, **8**, 1831.
[261] R. A. Badley and F. W. J. Teale, *J. Mol. Biol.*, 1969, **44**, 71.
[262] H. C. Chueng and M. F. Morales, *Biochemistry*, 1969, **8**, 2177.
[263] G. I. Likhtenshtein, A. P. Pivovarov, and N. B. Smolina, *Molec. Biol.* (*U.S.S.R., Eng. Trans.*), 1968, **2**, 235.
[264] T. J. Stone, T. Buckman, P. L. Nordio, and H. M. McConnell, *Proc. Nat. Acad. Sci. U.S.A.*, 1965, **54**, 1010.
[265] M. B. Neiman, E. G. Rozantzev, and Yu. G. Mamedova, *Nature*, 1962, **196**, 472.

labels are normally compounds of one of two types:



Type I label
N-(1-oxyl-2,2,6,6-tetramethyl-4-piperidinyl) derivative



Type II label
N-(1-oxyl-2,2,5,5-tetramethyl-3-pyrrolidinyl) derivative

The importance of these compounds lies in the possibility of modifying the group X to meet the requirements of a particular biochemical problem, without affecting the free radical properties of the nitroxide moiety. Three particular properties of the unpaired electron in such nitroxides have been used to gain information on macromolecular structure. These can be dealt with as (i) the sensitivity of the electron spin resonance (e.s.r.) spectrum to correlation time or rotational mobility, (ii) interaction of the unpaired electron with neighbouring paramagnetic ions or nuclei, and (iii) interaction of the unpaired electrons in a biradical. Two reviews of the uses of spin labels have appeared.[266, 267]

A. Measurements of Correlation Time.—When the correlation time (τ_c) of the spin label is 10^{-10}—10^{-11} sec (*e.g.* the free label in aqueous solution),[266] a very sharp three-line e.s.r. spectrum is observed, reflecting the hyperfine interaction of the unpaired electron with the ^{14}N nucleus. In media of high viscosity (*e.g.* glycerol) or when the spin label is rigidly attached to a macromolecule, either covalently or non-covalently, the correlation time of the label is lengthened ($\tau_c \sim 10^{-8}$ sec) resulting in considerable broadening of the e.s.r. spectrum. Changes in the secondary (or higher) structure of the macromolecule may well, therefore, be detected from changes in the e.s.r. spectrum of a bound label. From the exact shape of the spectrum an estimate of τ_c can be obtained.[266]

Model studies on spin-labelled polypeptides and small proteins have been described. The known helix–coil transitions in poly-L-lysine [268] and poly-L-tyrosine [269] brought about by a change of pH have been detected by changes in the e.s.r. spectra of the spin-labelled polypeptides.[264, 270] Bovine serum albumin (BSA) has been shown [271] from studies of viscosity and optical rotation to undergo a conformational change between pH 4 and 2.

[266] C. L. Hamilton and H. M. McConnell, in ref. 3, p. 115.
[267] O. H. Griffith and A. S. Waggoner, *Accounts Chem. Res.*, 1969, **2**, 17.
[268] J. Applequist and P. Doty, in 'International Symposium on Polyamino Acids, Polypeptides and Proteins, Madison, Wisconsin', ed. M. A. Stahmann, University of Wisconsin Press, 1962.
[269] G. D. Fasman, E. Bodenheimer, and C. Lindblow, *Biochemistry*, 1964, **3**, 1665.
[270] M. D. Barratt, G. H. Dodd, and D. Chapman, *Biochim. Biophys. Acta*, 1969, **194**, 600.
[271] J. T. Yang and J. F. Foster, *J. Amer. Chem. Soc.*, 1954, **76**, 1588.

Reaction of BSA with a spin label derivative of *N*-ethylmaleimide gave a labelled protein whose e.s.r. spectrum above pH 4 had substantial contributions from both 'weakly' and 'strongly' immobilised label sites.[264, 272] In the pH range mentioned above, there was a striking diminution of the 'strongly' immobilised portion of the spectrum, showing that the label can report the conformational change.[272]

The sensitivity of the shape of the e.s.r. spectrum to τ_c was used in an experiment to define the depth of the binding site of an antibody protein.[273] The antigen can be replaced, in so far as binding to the correct antibody is concerned, by a small group (hapten) such as 2,4-dinitrophenyl (Dnp). A number of such haptens with a nitroxide spin label at various, calculable distances from the Dnp group were synthesised. When bound to the corresponding antibody, the e.s.r. spectrum of the labelled hapten showed a dramatic change-over to a less strongly immobilised type in those molecules where the distance between the Dnp group and the spin label was greater than 12 Å. In these cases the spin label could rotate more freely outside the cleft in the antibody where the binding site is located. This value for the depth of the binding site is in very good agreement with that derived from electron microscopy measurements of fixed antibodies in the solid state.[274]

Spin label analogues of di-isopropyl phosphofluoridate (DIFP), which is a powerful irreversible inhibitor of many enzymes, have been used to show significant differences between the active sites of the enzymes, α-chymotrypsin and acetylcholinesterase. The e.s.r. spectrum shows that the spin label attached to chymotrypsin is very strongly immobilised.[275] Spin-labelled acetylcholinesterase, by contrast, shows a much less immobilised type of spectrum, suggesting that the active site is in a less confining region of the enzyme.[275] This conclusion was also reached in spin label studies on acetylcholinesterase bound to membranes and is consistent with independent evidence.[276] It is interesting to note that those enzymes, which are not bound to membranes *in vivo*, contain the active site in a deep cleft of the protein.[277] This may highlight an important difference between the two classes of enzymes.

Spin label derivatives of *p*-nitrophenyl esters and amides act as substrates for chymotrypsin. An intermediate in which the spin label was covalently linked to the enzyme (presumably at the 'essential' serine 195) could be isolated by rapidly lowering the pH after mixing enzyme and substrate. The effects of oxidation of the neighbouring methionine residues

[272] O. H. Griffith and H. M. McConnell, *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **55**, 8.
[273] J. C. Hsia and L. H. Piette, *Arch. Biochem. Biophys.*, 1969, **129**, 296.
[274] R. C. Valentine and N. M. Green, *J. Mol. Biol.*, 1967, **27**, 615.
[275] J. C. Hsia, D. J. Kosman, and L. H. Piette, *Biochem. Biophys. Res. Comm.*, 1969, **36**, 75.
[276] J. D. Morrisett, C. A. Broomfield, and B. E. Hackley jun., *J. Biol. Chem.*, 1969, **244**, 5758.
[277] M. F. Perutz, *European J. Biochem.*, 1969, **8**, 455.

180 and 192 on the e.s.r. spectrum of the labelled enzyme, together with the effects of indole on the rate of breakdown of the intermediate, were combined to suggest a hypothesis of the mode of action of chymotrypsin and clarify the interrelationship between the 'aryl' and 'amide' regions of the enzyme active site.[278]

The nature of the interaction of the subunits in haemoglobin (Hb) has been investigated by the spin labelling technique. As *X*-ray studies show, the cysteine residues at the 93 position of the β chains are very close to the $\alpha_1\beta_2$, $\alpha_2\beta_1$ contact regions between subunits.[279, 280] Previous work had shown that a spin label of Type II attached to the β 93 cysteines was sensitive to a conformational change when oxygenation of the haem systems of the β chain occurred.[281] When a label of Type I was linked to the β 93 cysteines two types of conformational change could be detected on oxygenation; one due to oxygenation of the haems of the β chain haem, the other due to oxygenation of the haems of the α chain.[282] [This latter change was detected after 'fixation' of the separated labelled β chains in the cyanomet form, reconstitution of the $(\alpha\beta)_2$ molecule, and alteration of the state of oxygenation of the α chain haems.] This type of effect is important in theories of 'co-operative' or 'allosteric' phenomena,[266, 283] and provides good evidence for the importance of the $\alpha_1\beta_2$, $\alpha_2\beta_1$ contacts between subunits in the haem–haem interaction.

The e.s.r. spectra of spin-labelled derivatives of oxy-, carbonmonoxy-, met-, met-fluoride-, met-cyano-, and met-azide-haemoglobin (Type I or Type II labels used) show evidence for an equilibrium between two isomeric states of the label relative to the protein.[282, 284] Spin-labelled deoxy-Hb, however, shows no evidence for such an equilibrium. The results of these studies of Hb in solution support the conclusion from *X*-ray studies of crystals that the structure of deoxy-Hb is distinctly different from that of the other derivatives listed.[279, 285, 286] Further work has characterised these two isomeric states in more detail, and also some small differences between the e.s.r. spectra of spin-labelled carbonmonoxy- and met-Hb.[287] These latter differences are present in the polycrystalline and single crystal forms as well as in solution. The principal axes of orientation (in the single crystals) were found to be very nearly identical for Type I or Type II

[278] D. J. Kosman, J. C. Hsia, and L. H. Piette, *Arch. Biochem. Biophys.*, 1969, **133**, 29.
[279] H. Muirhead, J. M. Cox, L. Mazzarella, and M. F. Perutz, *J. Mol. Biol.*, 1967, **28**, 117.
[280] M. F. Perutz, H. Muirhead, J. M. Cox, and L. C. G. Goaman, *Nature*, 1968, **219**, 131.
[281] S. Ogawa and H. M. McConnell, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 19.
[282] S. Ogawa, H. M. McConnell, and A. Horwitz, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 401.
[283] D. E. Koshland jun., G. Nemethy, and D. Filmer, *Biochemistry*, 1966, **5**, 365.
[284] H. M. McConnell, S. Ogawa, and A. Horwitz, *Nature*, 1968, **220**, 787.
[285] W. Bolton, J. M. Cox, and M. F. Perutz, *J. Mol. Biol.*, 1968, **33**, 283.
[286] M. F. Perutz and F. S. Mathews, *J. Mol. Biol.*, 1966, **21**, 199.

labelled met- and carbonmonoxy-Hb. Hence the differences in protein structure between these two Hb derivatives must be very small. This is then not inconsistent with the *X*-ray studies which showed that these two Hb derivatives must have very similar secondary and tertiary structures and identical quaternary structures.[286]

B. Interaction of the Unpaired Electron with Neighbouring Paramagnetic Ions or Nuclei.—In an elegant series of experiments, Mildvan and Weiner used the ability of the unpaired electron of the spin label to enhance the relaxation rate of neighbouring protons, to obtain considerable information about the active site of liver alcohol dehydrogenase (ADH).[288–290] They described the synthesis of a spin label analogue of β-nicotinamide adenine dinucleotide (NAD^+)—a substrate for ADH. In the (enzymatically inactive) analogue, the unpaired electron is in a position sterically equivalent to the bond between the ring nitrogen of the pyridine and the C-1 atom of the ribose of NAD^+. The spin label analogue (ADP-R$^\bullet$) bound non-covalently to ADH at two strong binding sites and at a number (5—6) of weaker binding sites. A study of the kinetics of the enzymatic reaction showed that ADP-R$^\bullet$ was a competitive inhibitor with respect to NAD^+. However, the broadening of the e.s.r. spectrum of ADP-R$^\bullet$ on binding to ADH was too great to allow an estimate of the correlation time to be derived.

The unpaired electron of ADP-R$^\bullet$ bound to ADH enhanced the relaxation rate ($1/T_1$) of water protons 81 times as efficiently as did free ADP-R$^\bullet$. This enhancement factor was found to be sensitive to the degree of occupancy of the tight binding sites by the label, suggesting a site–site interaction which leads to an 'opening up' of the active site (this may be necessary to permit substrate entry). Formation of the ternary complex ADH/ADP-R$^\bullet$/ethanol on addition of ethanol to the binary complex was detected (i) by a reduction in the enhancement factor for relaxation of water protons, and (ii) by the relaxation of the protons of ethanol caused by bound ADP-R$^\bullet$. The relaxation effect is proportional to r^{-6}, where r is the distance between the unpaired electron and the particular nucleus.[291, 292] Measurement of the relaxation times in various ternary complexes led to the assignment of the distances from the unpaired electron (and hence from the pyridine–ribose bond of NAD^+) to (i) the methylene and methyl sets of protons in ethanol, (ii) the unique proton of acetaldehyde, and (iii) the methyl protons of isobutyramide (a competitive inhibitor of the enzyme) when these compounds were separately bound to the ADH/ADP-R$^\bullet$ complex.[290]

[287] H. M. McConnell, W. Deal, and R. T. Ogata, *Biochemistry*, 1969, **8**, 2580.
[288] H. Weiner, *Biochemistry*, 1969, **8**, 526.
[289] A. S. Mildvan and H. Weiner, *Biochemistry*, 1969, **8**, 552.
[290] A. S. Mildvan and H. Weiner, *J. Biol. Chem.*, 1969, **244**, 2465.
[291] I. Solomon, *Phys. Rev.*, 1955, **99**, 559.
[292] I. Solomon and N. Bloembergen, *J. Chem. Phys.*, 1956, **25**, 261.

A spin–spin interaction between the Mn^{II} ion and the free electron of a spin label was observed in labelled creatine kinase. The two sulphydryl groups of the enzyme molecule essential for catalytic activity were combined with a Type II spin label derivative.[293] From the e.s.r. spectrum of the labelled enzyme an estimate of τ_c of 5×10^{-8} sec was derived, compared with $\tau_c \sim 10^{-11}$ sec for the free spin label in aqueous solution.[266] Alkaline-earth ions, which activate the enzyme, diminish the amplitude of the e.s.r. spectrum of the bound label. In the presence of adenosine-5′-diphosphate, spin–spin interactions between the bound label and the paramagnetic ions Mn^{II}, Ni^{II}, and Co^{II} were observed. From the magnitude of the interaction, the distance between the Mn^{II} ion and the unpaired electron of the spin label was estimated as 7—10 Å.

C. Interaction of the Unpaired Electrons in a Biradical.—The exchange of electrons between the two sites of a biradical results in changes in the e.s.r. spectrum. Various types of biradicals with rigid or flexible backbones have been synthesised and their e.s.r. spectrum recorded.[294] Two types of electron exchange are possible (i) direct, *via* the medium, and (ii) indirect, *via* the biradical backbone; (ii) is unimportant in biradicals of suitably chosen length and geometry. The electron exchange may give further information on the environment of the biradical. Calvin and his co-workers diffused a biradical, of the flexible backbone type, into a resting nerve (from the walking leg of the lobster).[294] The e.s.r. spectrum indicated that the biradical within the nerve had a high tumbling rate (*i.e.* low τ_c) and that its backbone had become more rigid; possibly as a result of being situated in a lipid portion of the nerve membrane. On excitation of the nerve, the e.s.r. spectrum remained unchanged, consistent with the suggested location of the biradical. A biradical analogue of di-isopropyl phosphofluoridate has been caused to react with chymotrypsin, but no detailed study of the electron exchange was made.[275]

Among the many other papers on spin labels which appeared in 1969 are those in the following fields: polypeptides,[295] local conformational transitions in proteins,[296, 297] lipid–protein complexes,[298] muscle components,[299–301]

[293] J. S. Taylor, J. S. Leigh jun., and M. Cohn, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **64**, 219.
[294] M. Calvin, H. H. Wang, G. Entine, D. Gill, P. Ferruti, M. A. Harpold, and M. P. Klein, *Proc. Nat. Acad. Sci., U.S.A.* 1969, **68**, 1.
[295] P. L. Nordio, A. Scatturin, and A. Tamburro, *Ricerca sci.*, 1968, **38**, 832.
[296] G. I. Likhtenshtein and P. Kh. Bobodzhanov, *Biofizika*, 1968, **13**, 757 (*Chem. Abs.* 1969, **70**, 34,905).
[297] G. I. Likhtenshtein, T. V. Troshkina, Yu. D. Akhmedov, and V. F. Shuvalov, *Mol. Biol.*, 1969, **3**, 413. (*Chem. Abs.* 1969, **71**, 35,954.)
[298] M. D. Barratt, D. K. Green, and D. Chapman, *Biochim. Biophys. Acta*, 1968, **152**, 20.
[299] Y. Tonomura, S. Watanabe, and M. F. Morales, *Biochemistry*, 1969, **8**, 2171.
[300] R. Cooke and M. F. Morales, *Biochemistry*, 1969, **8**, 3188.
[301] J. Quinlivan, H. M. McConnell, L. Stowring, R. Cooke, and M. F. Morales, *Biochemistry*, 1969, **8**, 3644.

membranes and model membranes,[302–307] lipoproteins,[308] steroids,[309] valyl-transfer ribonucleic acid,[310] and vitamin B_{12}.[311]

Spin labels have thus been used to investigate many problems with different types of proteins. It is clearly essential to know more quantitatively what the terms 'strongly' and 'weakly' immobilised labels mean in any particular case, and to be able to characterise the conformational changes of macromolecules (often very localised) which the e.s.r. spectrum of the bound label reflects. The experiments of Mildvan and Weiner [288–290] show that the use of spin label bio-analogues can give detailed information on the geometry of active sites, and suggest that this technique can be widely used in investigations of the structure and function of biological molecules.

7 Dissociation and Association of Proteins

contributed by R. Blagrove, P. H. Lloyd, and I. O. Walker

A. Dissociating Agents.—It is now almost a routine procedure for anyone purifying a new protein to measure its molecular weight in both dilute buffer and in a denaturing solvent and to deduce the number of subunits in the protein from the ratio of the two. The most popular denaturing solvent is 6M guanidine hydrochloride.[312–317] In none of these studies have the authors checked to see if this concentration of guanidine hydrochloride was sufficient to dissociate the protein completely, and Woods [318] has shown that for paramyosin a concentration of guanidine hydrochloride of at least 7M was needed. Careful control of pH may also be necessary.[319] On the other hand, several authors have found that lower concentrations

[302] W. L. Hubbell, and H. M. McConnell, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 12.
[303] D. Chapman, M. D. Barratt, and V. B. Kamat, *Biochim. Biophys. Acta*, 1969, **173**, 154.
[304] W. L. Hubbell and H. M. McConnell, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **63**, 16.
[305] L. J. Libertini, A. S. Waggoner, P. C. Jost, and O. H. Griffith, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **64**, 13.
[306] W. L. Hubbell and H. M. McConnell, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **64**, 20.
[307] H. E. Sandberg, R. G. Bryant, and L. H. Piette, *Arch. Biochem. Biophys.*, 1969, **133**, 144.
[308] A. M. Gotto and H. Kon, *Biochem. Biophys. Res. Comm.*, 1969, **37**, 444.
[309] J. C. Espie, H. Lemaire, and A. Rassat, *Bull. Soc. chim. France*, 1969, 399.
[310] B. M. Hoffman, P. Schofield, and A. Rich, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 1195.
[311] T. Buckman, F. S. Kennedy, and J. M. Wood, *Biochemistry*, 1969, **8**, 4437.
[312] D. L. Barker and W. P. Jencks, *Biochemistry*, 1969, **8**, 3879.
[313] G. M. Malacinski and W. J. Rutter, *Biochemistry*, 1969, **8**, 4382.
[314] F. J. Castellino and R. Barker, *Biochemistry*, 1968, **7**, 2207.
[315] W. N. Poillon, H. Maeno, K. Koike, and P. Feigelson, *J. Biol. Chem.*, 1969, **244**, 3447.
[316] R. W. Gracy, A. G. Lacko, and B. L. Horecker, *J. Biol. Chem.*, 1969, **244**, 3913.
[317] W. Leuzinger, M. Goldberg, and E. Gauvin, *J. Mol. Biol.*, 1969, **40**, 217.
[318] E. F. Woods, *Biochem. J.*, 1969, **113**, 39.
[319] D. B. Millar, V. Fratalli, and G. E. Willick, *Biochemistry*, 1969, **8**, 2416.

of guanidine hydrochloride were sufficient.[320–324] Such are the difficulties, both theoretical and practical, of working with concentrated salt solutions that the use of the minimum in any particular case is desirable.

The effect of Cl^-, Br^-, I^-, and SCN^- salts of guanidinium, carbamoylguanidinium, and guanylguanidinium cations on the water solubility of benzoyl-L-tyrosine ethyl ester, a model hydrophobic compound, and acetyltetraglycine ethyl ester, a model peptide and amide compound, have been examined.[325] Regardless of the cation used, the solubility of both model compounds increases progressively through the series $Cl^- < Br^- < I^- < SCN^-$, which parallels the effectiveness of these anions as denaturants of many proteins. The solubility of both compounds has been compared at constant concentration of the same anions. It is concluded that the best model for denaturation by salts of the guanidinium type involves the disruption of the water structure, the loosening of hydrophobic interactions, and the solubilisation both of the interior of the protein due to specific interactions with the peptide bonds and of the hydrophobic regions.

Another widely used dissociating and denaturing solvent is 8M urea,[326–330] but again it has been shown that for some proteins this is either excessive or insufficient.[320, 331–333] For example, histidine decarboxylase is not dissociated by 8M urea but is broken into ten subunits by 5M guanidine hydrochloride with 1% sodium dodecyl sulphate,[320] and ceruloplasmin is incompletely dissociated by 8M urea.[331] This latter protein could be reassembled after dissociation by 4·5M urea with *p*-chloromercuribenzoate, but not after treatment with 8M urea and *p*-chloromercuribenzoate.[332] The effect of urea on a series of model compounds (dioxopiperazines) has been studied to see whether destruction of the characteristic structure of water alters the conformation of the side-chains under conditions where that of the chain backbone remains unaltered.[334] On the basis of solubility

[320] G. W. Chang and E. E. Snell, *Biochemistry*, 1968, **7**, 2012.
[321] B. A. McFadden, G. R. Rao, A. L. Cohen, and T. E. Roche, *Biochemistry*, 1968, **7**, 3574.
[322] A. Kotera, M. Yokoyama, M. Yamaguchi, and Y. Miyazawa, *Biopolymers*, 1969, **7**, 99.
[323] J. J. Butzow, *Biochim. Biophys. Acta*, 1969, **168**, 490.
[324] P. de la Llosa and M. Jutisz, *Biochim. Biophys. Acta*, 1969, **181**, 426.
[325] F. J. Castellino and R. Barker, *Biochemistry*, 1969, **8**, 3439.
[326] P. D. Jeffrey, *Biochemistry*, 1968, **7**, 3345.
[327] E. F. Woods, *Biochemistry*, 1969, **8**, 4336.
[328] J. W. Patrick and N. Lee, *J. Biol. Chem.*, 1969, **244**, 4277.
[329] D. L. Correll, E. Steers, K. M. Towe, and W. Shropshire, *Biochim. Biophys. Acta*, 1968, **168**, 46.
[330] A. Scanu, J. Toth, C. Edelstein, S. Koga, and E. Stiller, *Biochemistry*, 1969, **8**, 3309.
[331] H. Mukasa, S. Kajiyama, K. Sugiyama, K. Funakubo, M. Itoh, Y. Nosoh, and T. Sato, *Biochim. Biophys. Acta*, 1969, **168**, 132.
[332] H. Mukasa, Y. Nosoh, and T. Sato, *Biochim. Biophys. Acta*, 1969, **168**, 483.
[333] S. Wilk, A. Meister, and R. H. Haschemeyer, *Biochemistry*, 1969, **8**, 3168.
[334] M. D'Alagni and B. Pispisa, *J. Biol. Chem.*, 1969, **244**, 5843.

studies of amino-acids and peptides in guanidine hydrochloride, it appears that the strongest binding sites in a protein for this substance or its constituent ions are the aromatic side-chains and pairs of adjacent peptide groups.[335]

Other dissociating–denaturing agents which have been used are 1% sodium dodecyl sulphate,[315, 316, 320] methylmercuri-iodide,[336] *p*-chloromercuribenzoate,[331] succinic anhydride,[337] and high or low pH.[315, 328]

Rationalisation of the properties of aqueous solutions of proteins and small molecules in terms of the structure of water, both in the presence and absence of denaturing agents,[338] has been called into question.[339] It is claimed that most of the arguments based on the concept of water structure have been shown to be ambiguous and without meaning. The 'denatured' state of proteins in these dissociating–denaturing solvents and the transition from the native to denatured state have been reviewed by Tanford.[340] The kinetics of denaturation of BSA by urea and urea-type agents has been studied by dilatometry, acrylamide gel electrophoresis, and urea gel perturbation.[341] The process is initiated by a fast reaction producing conformationally altered albumins which, by a series of slower parallel and consecutive reactions, are converted to electrophoretically distinguishable forms of denatured proteins consisting of multiple sub-species. The type and distribution of denatured products depends on the pH. In another study of the denaturation of BSA in the presence of various organic solvents,[342] the effects observed were apparently due to modification of the pK of the free carboxy-groups as the dielectric constant of the medium was lowered. The denaturation of hen egg-white lysozyme by guanidine hydrochloride over a wide range of pH is a two-state process with only the native and fully denatured states contributing significantly to the equilibrium properties. No definite conclusions could be drawn [335] about the mode of action of guanidine hydrochloride on lysozyme although the inherently low stability of the native globular protein was a significant finding: the free-energy of stabilisation of native lysozyme, relative to the denatured state, under physiological conditions in the absence of denaturant, appears to be only 10—12 kcal mole^{-1}.

The interaction of sodium dodecyl sulphate with phycocyanin has been studied by n.m.r. using the fully deuteriated protein to observe the sur-

[335] K. C. Aune and C. Tanford, *Biochemistry*, 1969, **8**, 4579, 4586.
[336] N. R. Lazarus, M. Derechin, and E. A. Barnard, *Biochemistry*, 1968, **7**, 2390.
[337] W. N. Konnings, J. Dijk, T. Wichertjes, E. C. Beuvery, and M. Gruber, *Biochim. Biophys. Acta*, 1969, **188**, 43.
[338] S. Subramanian, D. Balasubramanian, and J. C. Ahluwalia, *J. Phys. Chem.*, 1969, **73**, 266.
[339] A. Holtzer and M. F. Emerson, *J. Phys. Chem.*, 1969, **73**, 26.
[340] C. Tanford, *Adv. Protein Chem.*, 1968, **23**, 122.
[341] S. Katz and J. Denis, *Biochim. Biophys. Acta*, 1969, **188**, 247.
[342] S. F. Sun, *Biochim. Biophys. Acta*, 1969, **181**, 473.

factant resonance without interference.[343] At low ratios of surfactant to protein, the resonance peaks of the surfactant were greatly broadened, which indicated a tightening of the protein structure. At higher ratios, all the resonance peaks of the surfactant were shifted upfield and moderately broadened, indicating that the dodecyl hydrocarbon chain had interacted with the protein and that its bound environment was hydrophobic. The protein had acquired increased freedom of motion at these higher ratios. Other important n.m.r. studies of the structure of proteins under denaturing conditions have been reported.[188, 195] The observation of simultaneous changes in several resonances as denaturant is added enables the assignment of a single-step or multiple-step process.[344] It is apparent that diverse results can be obtained for different proteins in the same denaturant and for one protein in different denaturants. The position of the resonances in the primary sequence appears to be an important factor.

B. Equilibrium Studies.—Apart from two papers reporting work with membrane osmometers,[314, 345] all the molecular weights of proteins in dissociating–denaturing solvents were determined by means of the ultracentrifuge. Although some workers are still using the Archibald technique,[312, 322, 324, 331, 337] the majority are now using full sedimentation equilibrium. Of the various techniques available the high-speed, meniscus-depletion method of Yphantis,[346] combined with use of the Rayleigh interferometer, is becoming very popular. The advantage of this technique in yielding absolute fringe displacements without the need for knowledge of the original concentration must be set against its disadvantages. It is claimed that it is more sensitive to heterogeneity than the lower-speed methods.[347] However, examination of the published graphs of the logarithm of the fringe displacement plotted against the square of the radial co-ordinate show that in most cases the depth of the column of solution over which measurements were actually made was only about 1 mm. This arises because in the top half of a normal 3 mm column the concentration is too low to measure and towards the bottom the spacing of the fringes becomes too small for most optical systems to resolve them. Such short columns inevitably reduce the sensitivity to heterogeneity. In one study[348] a longer column was used (7 mm). Satisfactory use of this technique requires very accurate measurements of the fringe photographs. In another study, the authors[328] claimed to be able to measure fringe displacements of only 0·5 μm, but no justification of this accuracy was

[343] R. M. Rosenberg, H. L. Crespi, and J. J. Katz, *Biochim. Biophys. Acta*, 1969, **175**, 31.
[344] J. H. Bradbury and N. L. R. King, *Nature*, 1969, **223**, 1154.
[345] R. F. Steiner, *Biochemistry*, 1968, **7**, 2201.
[346] D. A. Yphantis, *Biochemistry*, 1964, **3**, 297.
[347] H. E. Sine and L. F. Hass, *J. Biol. Chem.*, 1969, **244**, 430.
[348] D. B. Millar, G. E. Willick, R. F. Steiner, and V. Frattali, *J. Biol. Chem.*, 1969, **244**, 281.

given. In only one paper [318] was sufficient detail given for the reader to judge for himself the reliability of the readings at very low displacements and this author claims to be able to select his base-line to only about 5 μm (Yphantis [346] found a similar limitation). Base-lines can probably not be reproduced to an accuracy as good as this (Yphantis [346] claims an accuracy of 10 μm). Only about half of the papers show the graphs from which the molecular weights were deduced; three of these [312, 349, 350] show linear plots of the logarithm of the fringe displacements against the reading of the comparator on which the photographs were measured. Such plots are only linear when very short columns are used.

Many of the difficulties inherent in the use of the interferometer are eliminated by the use of electronic scanners and absorption optical systems. These yield absolute values of optical density within the cell without the need for ancillary measurements. However, the use of these is not yet common,[315, 351] nor is it yet clear whether or not the scanners can yield results of an accuracy as good as those obtained with the interferometer.

In all sedimentation-equilibrium experiments it is necessary to evaluate the partial specific volume of the solute (or the quantity $d\rho/dc$ for the solution) before the molecular weight can be determined. This quantity was actually measured in only a few cases (*e.g.* 316, 318, 337, 347) while in others (*e.g.* refs. 315, 317, 323, 348) it was calculated from the amino-acid composition. The errors which can arise from failure to determine this quantity, especially in 6M guanidine hydrochloride, have been emphasised.[352] (See also an earlier discussion of related problems.[340])

In the determination of the number of subunits from the relative decrease in molecular weight on dissociation in a suitable solvent, little attention has been paid to the identity or otherwise of the subunits, identity often being concluded from linearity of a sedimentation-equilibrium plot. Some attempts to use the high-speed sedimentation-equilibrium technique to resolve mixtures of two proteins have been made,[353] presumably with this problem in mind. The molecular weights of the individual subunits of collagen have been determined.[354] The results of a number of recent dissociation studies are summarised in Table 1.

Only a few papers have appeared reporting attempts to evaluate the interaction constants of proteins aggregating with themselves.[348] (See Table 2.) These all follow the theoretical work of Adams [356] on association in non-ideal systems. The results are summarised in Table 2.

[349] O. Tarutani and N. Ui, *Biochim. Biophys. Acta*, 1969, **181**, 116.
[350] O. Tarutani and N. Ui, *Biochim. Biophys. Acta*, 1969, **181**, 136.
[351] T. H. Gawronski and E. W. Westhead, *Biochemistry*, 1969, **8**, 4261.
[352] E. Reisler and H. Eisenberg, *Biochemistry*, 1969, **8**, 4572.
[353] P. D. Jeffrey and M. J. Pont, *Biochemistry*, 1969, **8**, 4597.
[354] E. P. Katz, C. J. Francois, and M. J. Glimcher, *Biochemistry*, 1969, **8**, 2609.
[355] M. L. Applebury, and J. E. Coleman, *J. Biol. Chem.*, 1969, **244**, 308,
[356] E. T. Adams, *Biochemistry*, 1965, **4**, 1646; E. T. Adams and D. L. Filmer, *Biochemistry*, 1966, **5**, 2971.

Table 1 *The results of dissociation studies*

Protein	*Number of subunits reported*	*Dissociating agent*	*Ref.*
Pig liver esterase	4	6M GuCl	312
Yeast enolase	2	KCl, KBr	351
Histidine decarboxylase from *Lactobacillus* 30a	10	5M GuCl+1% SDS	320
Yeast hexokinase	2 or 4	MMI	336
Isocitrate lyase from *Pseudomonas indigofera*	4	2—6M GuCl	321
Alkaline phosphatase from *E. coli*	2	pH 2	355
Muscle aldolase	4	pH 12	347
Tryptophan oxygenase from *Pseudomonas acidovorans*	4	6M GuCl pH 12, SDS	315
Liver aldolase	4	6M GuCl, 1% SDS	316
Acetylcholine esterase	4	6M GuCl	317
Rye phytochrome	4	SDS, 8M urea	329
Porcine ceruloplasmin	2	8M urea+PCMB	331
Haemocyanin	20	succinylation	337
Ovine luteinizing hormone	2	2·2M GuCl	324
Hog thyroglobin	2	1 mM SDS	362
Rabbit muscle pyruvate kinase	4	4M urea	360
Arginine decarboxylase	10	—	361
Lactate dehydrogenase	8	6·2M GuCl, pH 8	319
Oyster paramyosin	2	7M GuCl	318
Bovine α-crystallin	—	6M urea, 2% SDS	363

GuCl = guanidine hydrochloride, SDS = sodium dodecyl sulphate, MMI = methylmercuri-iodide, PCMB = *p*-chloromercuribenzoate.

A number of theoretical papers have appeared which are of interest to workers in this field. Adams [367] has extended his work to the case of two different macromolecules combining with one another. There are ten printing errors in this paper which are corrected in a later volume.[368] Steiner [345] has shown how colligative methods can be used to determine association constants in a mixture of two components which are aggregating

[357] T. A. J. Payens, J. A. Brinkhuis, and B. W. van Markwijk, *Biochim. Biophys. Acta*, 1969, **175**, 434.
[358] T. Kotaka and R. L. Baldwin, *Biopolymers*, 1969, **7**, 87.
[359] P. W. Chun and S. J. Kim, *Biochemistry*, 1969, **8**, 1633.
[360] G. L. Cottam, P. F. Hollenberg, and M. J. Coon, *J. Biol. Chem.*, 1969, **244**, 1481; G. S. Johnson, M. S. Kayne, and W. C. Deal, *Biochemistry*, 1969, **8**, 2455.
[361] E. A. Boeker, E. H. Fischer, and E. E. Snell, *J. Biol. Chem.*, 1969, **244**, 5239.
[362] O. Tarutani and N. Ui, *Biochim. Biophys. Acta*, 1969, **181**, 116.
[363] H. J. Hoenders, K. de Groot, J. J. T. Gerding, and H. Bloemendal, *Biochim. Biophys. Acta*, 1969, **188**, 162.
[364] J. C. Swann and G. G. Hammes, *Biochemistry*, 1969, **8**, 1.
[365] D. K. Hancock and J. W. Williams, *Biochemistry*, 1969, **8**, 2598.
[366] P. D. Jeffrey, *Biochemistry*, 1968, **7**, 3352.
[367] E. T. Adams, A. H. Pekar, D. A. Soucek, and L. H. Tang *Biopolymers*, 1969, **7**, 5.
[368] E. T. Adams, A. H. Pekar, D. A. Soucek, and L. H. Tang, *Biopolymers*, 1969, **8**, 157.

Table 2 *The results of association studies*

Protein	*Association*	*Equilibrium constants*	*Ref.*
Glucagon	M-D-H	$K_2 = 3{\cdot}3 \times 10^3 M^{-1}$ $K_6 = 10{\cdot}9 \times 10^{17} M^{-5}$	364
Chymotrypsinogen A	M-D-T	$K_2 = 0{\cdot}908$ dl/g $K_3 = 0{\cdot}84$ (dl/g)^2	365
	Random	$K = 49{\cdot}58$ dl/g	
Liver esterase	M-D	$K_2 = 4 \times 10^{-7} M$	312
A wool protein	M-D-T	$K_2 = 19{\cdot}9$ dl/g $K_3 = 10^3$ (dl/g)^2	366
Soybean proteinase inhibitor	M-D-T	$K_2 = 2{\cdot}85 \times 10^3 M^{-1}$ $K_3 = 1{\cdot}34 \times 10^6 M^{-2}$ $K_{23} = 4{\cdot}71 \times 10^2 M^{-1}$	348

M = monomer, D = dimer, T = trimer, H = hexamer; K_2 is the constant for the reaction $2M \rightleftharpoons D$; K_3 is the constant for the reaction $3M \rightleftharpoons T$; K_6 is the constant for the reaction $6M \rightleftharpoons H$; K_{23} is the constant for the reaction $M + D \rightleftharpoons T$.

both with themselves and with each other. Derechin [369] has extended his multinormal theory to non-ideal, infinitely-associating systems of a single component, and shown how the *z*-average molecular weight can be used with ideal systems.[370] Payens *et al.*[357] have shown how light-scattering and sedimentation measurements can be combined to yield the degree of polymerisation and the second virial coefficient in non-ideal systems. Kotaka and Baldwin [358] have produced a new theory of the approach to equilibrium which they claim enables quicker measurements to be made on compounds of very high molecular weight ($> 10^6$). Chun, Kim *et al.*[359, 371] have shown how equilibrium constants for associating systems may be evaluated from equilibrium partition coefficients determined with gels. Stone and Metzgen [372] have shown how the association constants for macromolecular interactions can be determined from equilibrium molecular sieving.

C. Dynamic Studies.—There has been little published in this field during 1969. The experimental study of associating protein systems using transport boundary analysis has lost ground in favour of the equilibrium methods based on light scattering and sedimentation equilibrium, especially the latter. In practice there are two methods available for studying associating systems using transport boundary analysis: sedimentation velocity and molecular sieve chromatography. The interpretation of sedimentation velocity data has largely depended on the theories developed by Gilbert [373] and colleagues who described the main features that are to be expected in sedimenting boundaries of solutes involved in rapid association–dissociation equilibria. This treatment provides predictions of boundary shapes in monomer–polymer systems so long as the spreading of the

[369] M. Derechin, *Biochemistry*, 1969, **8**, 921.
[370] M. Derechin, *Biochemistry*, 1969, **8**, 927.
[371] P. W. Chun, S. J. Kim, C. A. Stanley, and G. K. Ackers, *Biochemistry*, 1969, **8**, 1625.
[372] M. J. Stone and H. Metzger, *J. Biol. Chem.*, 1968, **243**, 5049.
[373] G. A. Gilbert, *Proc. Roy. Soc.*, 1963, *A*, **276**, 354.

boundaries by diffusion is negligible. Recently Cox [374] has extended the theory to cover both sedimentation and diffusion and has presented the features of the boundary shape for various polymerising systems of the type $nA \rightleftharpoons A_n$ for $n < 6$. A theory which deals with polymerising systems in which there are significant amounts of intermediate species is still lacking. There have been no recent attempts to apply the theories to the quantitative analysis of an experimental system. However, the sedimentation velocity method still remains very popular for following dissociating systems qualitatively and has been used to study protein interactions in several cases.[375, 376]

Molecular sieve chromatography may be used for the quantitative analysis of interacting systems and yields results analogous to those obtained by other methods. It has the advantages of being quick and easy to implement. A procedure has been devised for the evaluation of the weight-fraction of monomer from the experimental weight-average partition coefficients for several cases of finite association and also for the case of indefinite association.[377] The results obtained from this kind of experiment can be combined with molecular weights obtained by an independent method to provide information about the mode of aggregation for the associating system. In the case of glutamate dehydrogenase correlation of the observations in this way indicated a linear, indefinite polymerisation of subunits.[377] Chun and Kim [378] also describe a method for the quantitative evaluation of BM_1 (the non-ideality term) based on a combination of molecular sieve partition coefficient and molecular weights. A simple method for distinguishing between a non-interacting system and one undergoing continual re-equilibration on the basis of a single frontal gel filtration experiment has been described.[379]

At a more empirical level, forty proteins with polypeptide chains of well-characterised molecular weights have been studied [380] by polyacrylamide-gel electrophoresis in the presence of sodium dodecyl sulphate.[381] The results show that the method can be used to determine the molecular weights of only small amounts of protein and of polypeptide chains to an accuracy of about $\pm 10\%$. In a similar study of twenty proteins in the molecular weight range 5×10^3 to $3{\cdot}5 \times 10^5$ the effects of intrinsic molecular charge and conformation have been found to be small, although anomalies can occur.[382] Another new method which has been evaluated [383] is the

[374] D. J. Cox, *Arch. Biochem. Biophys.*, 1969, **129**, 106.
[375] S. Wilk, A. Meister, and R. H. Haschemeyer, *Biochemistry*, 1969, **8**, 3168.
[376] J. J. Irias, M. R. Olmstead, and M. F. Utter, *Biochemistry*, 1969, **8**, 5136.
[377] P. W. Chun, S. J. Kim, C. A. Stanley, and G. K. Ackers, *Biochemistry*, 1969, **8**, 1625.
[378] P. W. Chun and S. J. Kim, *Biochemistry*, 1969, **8**, 1633.
[379] J. A. Ronalds and D. J. Winzor, *Arch. Biochem. Biophys.*, 1969, **129**, 94.
[380] K. Weber and M. Osborn, *J. Biol. Chem.*, 1969, **244**, 4406.
[381] A. L. Shapiro, E. Vinuela, and J. V. Maizel, *Biochem. Biophys. Res. Comm.*, 1967, **28**, 815.
[382] A. K. Dunker and R. R. Rueckert, *J. Biol. Chem.*, 1969, **244**, 5074.
[383] W. W. Fish, K. G. Mann, and C. Tanford, *J. Biol. Chem.*, 1969, **244**, 4989.

estimation of molecular weights by gel-filtration in 6M guanidine hydrochloride:[384] results of similar accuracy to the electrophoretic method were obtained with 6% agarose as a gel-filtration medium. The method should presumably apply to any denaturing medium in which all the proteins are dissociated into their constituent polypeptide chains which then adopt the same conformation, regardless of composition, so that their characteristic dimensions are a function only of molecular weight.

8 Diffusion

contributed by R. Blagrove

The measurement of translational diffusion coefficients (D) from boundary spreading in an ultracentrifuge equipped with absorption optics has been reported.[385] A digital computer is used to perform the complicated statistical calculations required. Measurements from the densitometer tracings of a velocity sedimentation run with a single, though possibly broad and concentration-dependent, boundary are processed to yield the initial concentration, the weight-average sedimentation coefficient (s), the weight-average diffusion coefficient, and a heterogeneity parameter which is the standard deviation of the distribution of sedimentation coefficient among the components present. Kawahara[386] has derived a simple and approximate equation from Fujita's equation[387] which permits the evaluation of D from sedimentation boundary curves. He has shown that the effect of the concentration dependence of the sedimentation coefficient on the evaluation of D decreases with decreasing speeds of rotation and that D can be found with simple procedures under proper conditions. The diffusion coefficients of several typical proteins, both native and denatured, have been determined using both the classical Lamm equation and the equation he proposed. The values of D for native proteins have an uncertainty of less than 10% but the uncertainty is high with denatured proteins.

Theories presented for the calculation of D contain the unrealistic assumption that all the molecules constituting a given mixture have the same diffusion coefficient. These have been extended to cover polydisperse solutes[388] when the product $s \cdot D$ can be regarded as constant (synthetic polymers under theta conditions) and when the product $\sqrt{s} \cdot D$ is constant (globular proteins). Besides applying the rectangular approximation to the sector-shaped cell, the concentration dependence of both s and D must be ignored so that great care must be taken to extrapolate suitably the results to infinite dilution. The problem of how the apparent

[384] P. F. Davison, *Science*, 1968, **161**, 1906.
[385] R. Trautman, S. P. Spragg, and H. B. Halsall, *Analyt. Biochem.*, 1969, **28**, 396.
[386] K. Kawahara, *Biochemistry*, 1969, **8**, 2551.
[387] H. Fujita, 'Mathematical Theory of Sedimentation Analysis', 1962, New York, Academic Press.
[388] H. Fujita, *Biopolymers*, 1969, **7**, 59.

values of the function $q(s)$ and D at non-zero concentrations can be correctly extrapolated to find their true values requires further study. The application to polydisperse systems of diffusion measurements from linewidth studies of scattered laser light has also been discussed.[389]

Two new methods for the measurement of diffusion coefficients by spreading from thin layers have been described.[390] In one, the spreading is measured from a thin layer in a cylindrical tube containing a suitable gel which is rotated horizontally round its long axis to avoid convection. In the other the spreading is from a thin layer under a column of solvent in a Tiselius electrophoresis cell or a micro-cuvette of a spectrophotometer. Both methods only require about 0·5 mg of protein. An improved self-beat spectrometer has been used to study the spectrum of light scattered from solutions of poly-γ-benzyl-L-glutamate.[391] The change in translational diffusion coefficient with solvent composition reflects the well-known helix–coil transition. Further improvement in the speed and accuracy of laser scattering measurements should make this an important technique in the study of the diffusion of macromolecules, in general, including proteins and polypeptides.

[389] R. Pecora and Y. Tagami, *J. Chem. Phys.*, 1969, **51**, 3298.
[390] A. Polson and D. G. Parkyn, *Biopolymers*, 1969, **7**, 107.
[391] N. C. Ford, W. Lee, and F. E. Karasz, *J. Chem. Phys.*, 1969, **50**, 3098.

3
Peptide Synthesis

BY J. H. JONES

1 Introduction

The advances in the subject recorded this year are very varied, although less dramatic than those discussed in the first volume of these Reports. No more total syntheses of material with enzymic activity have yet appeared, but work is in progress on staphylococcal nuclease systems,[1-3] on ribonuclease T_1,[4] and on cytochrome *c*.[5] The first steps have been taken towards a synthesis, by the combination of chemical with biological techniques, of a DNA containing the encoded amino-acid sequence of bovine insulin A chain:[6] when successful, this will open a new dimension in peptide synthesis. In the polypeptide hormone field, the highly potent calcitonin of salmon has been synthesised,[7] and so has the 'connecting peptide' of porcine proinsulin,[8] together with a whole host of analogues of smaller biologically active peptides (see Appendix A). Improvements in the methodology of the art are also legion, with notable progress on racemisation-free alternatives to the azide fragment condensation method, in protective-group strategy, and in the use of rapid repetitive methods which promise to be increasingly useful for the preparation of partially protected fragments for further classical application. In a different tone, serious and difficult problems remain. Methods for the controlled and unambiguous construction of multiple disulphide bridge arrays are not within sight, although recently acquired knowledge of the proinsulin system suggests that in special cases a solution to this problem may be found in a diversion *via* biosynthetic precursors which spontaneously form the cross-links in the proper manner. It is now true that the classical synthesis of sequences of up to, say, twenty

[1] P. Cuatrecasas, H. Taniuchi, and C. B. Anfinsen, *Brookhaven Symposia in Biology*, 1969, **21**, 172.
[2] C. B. Anfinsen, in 'Plenary Lectures presented at the 5th International Symposium on the Chemistry of Natural Products', Butterworths, London, 1968, p. 461.
[3] M. Ohno, A. Eastlake, D. A. Ontjes, and C. B. Anfinsen, *J. Amer. Chem. Soc.*, 1969, **91**, 6842.
[4] N. Yanaihara, C. Yanaihara, G. Dupuis, J. Beacham, R. Camble, and K. Hofmann, *J. Amer. Chem. Soc.*, 1969, **91**, 2184.
[5] S. Sano and M. Kurihara, *Z. physiol. Chem.*, 1969, **350**, 1183.
[6] S. A. Narang and S. K. Dheer, *Biochemistry*, 1969, **8**, 3443.
[7] St. Guttmann, J. Pless, R. L. Huguenin, E. Sandrin, H. Bossert, and K. Zehnder, *Helv. Chim. Acta*, 1969, **52**, 1789.
[8] R. Geiger, G. Jäger, W. König, and A. Volk, *Z. Naturforsch.*, 1969, **24b**, 999.

residues is without fundamental problems and has become largely a routine matter, albeit one requiring much labour, skill, patience, and time. Methods for joining fragments together, however, need much improvement if the full potential of peptide synthesis in elucidating the 'geometric code' which relates amino-acid sequence to structure and function is to be realised. The application of the solid-phase method to peptide chains longer than about ten residues still requires extensive refinement which is not forthcoming at present, although the development of 'functional purification' (see section 2C) may bring certain larger molecules within its scope. All in all, this is a stimulating time to be involved in the subject, since success to date holds the promise of exciting things ahead.

Peptide synthesis has been the subject of two reviews.[9, 10] A revised edition of a well-known laboratory manual on the application of thin-layer chromatography in amino-acid and peptide chemistry has been published.[11]

2 Methods

The year 1932, when Bergmann and Zervas introduced the benzyloxycarbonyl group,[12] was the turning point which marked the beginning of modern peptide synthesis. Since that time, several hundred suggestions for new methods of protection and coupling have appeared, but only a small proportion of these have actually found application. The criteria which any new procedure must satisfy are severely restricting, and there is also a barrier of conservatism to surmount. Such conservatism as exists is understandable. Take, for example, the case of a peptide chemist who is planning a long synthesis and is faced with the choice between an established and a similar but new method of protecting some side-chain functionality. If the new method has been evaluated only in simple low molecular weight examples, only an adventurous man would choose it, since unpredictable side-reactions might accompany its removal from a long peptide, and much effort would then have been expended without reward. Of recent years, however, peptide chemists have set themselves such demanding synthetic objectives that necessity has forced them to turn more and more to new techniques: there has, for example, been a very marked movement, especially in the last two or three years, towards the use of protective groups which are labile towards milder conditions.

In a well-known review, Rydon [13] gave the results of a survey of 91 papers published during 1960—1962 showing the extent to which the main methods of peptide synthesis then available had been used. For comparative purposes a similar survey has been conducted by the Reporter for a sample of

[9] P. G. Katsoyannis and J. Z. Ginos, *Ann. Rev. Biochem.*, 1969, **38**, 881.
[10] E. Schröder and K. Lübke, *Fortschr. Chem. org. Naturstoffe*, 1968, **26**, 48.
[11] G. Patakai, 'Techniques of Thin-layer Chromotagraphy in Amino-acid and Peptide Chemistry', Humphrey Science Publishers, Ann Arbor, 1969.
[12] M. Bergmann and L. Zervas, *Ber.*, 1932, **65**, 1192.
[13] H. N. Rydon, 'R.I.C. Lecture Series', 1962, No. 5, 24.

91 papers published during 1968, and the figures for both periods are given in Table 1. The general trends since 1962 are obvious. t-Butyl-based protective groups are now widely used whereas alkali-labile protecting groups are on the decline, but the popularity of the benzyloxycarbonyl

Table 1 *The popularity of the principal methods of peptide synthesis*

Method	*Date first used*	*Percentage of papers in which the method was used*	
		1968	1960—1962
Z for *N*-protection	1932 [12]	82	90
Boc for *N*-protection	1957 [14]	71	6
ONp for coupling	1955 [15]	50	24
DCCI alone for coupling	1955 [16]	48	68
Azide for coupling	1902 [17]	44	43
OMe for *C*-protection	1906 [18]	39	54
OBu^t for *C*-protection	1959 [19]	37	5
OBzl for *C*-protection	1935 [20]	27	22
The solid-phase method	1963 [21]	24	—
Mixed anhydrides for coupling	1950 [22]	24	42
ONSu for coupling	1963 [23]	20	—
OEt for *C*-protection	1901 [24]	19	22
OTcp for coupling	1963 [25]	14	—
Nps for *N*-protection	1963 [26]	10	—
DCCI-HONSu for coupling	1966 [27, 28]	10	—
$OBzl(NO_2)$ for *C*-protection	1959 [29]	7	11
Tos for *N*-protection	1915 [30]	7	7
OPcp for coupling	1961 [31]	7	5
Reagent K for coupling	1961 [32]	7	5
Trt for *N*-protection	1953 [33]	3	10
Others	—	5	5

[14] F. C. McKay and N. F. Albertson, *J. Amer. Chem. Soc.*, 1957, **79**, 4686; G. W. Anderson and A. C. McGregor, *ibid.*, 1957, **79**, 6180.
[15] M. Bodanszky, M. Szelke, E. Törmörkeny, and E. Weisz, *Chem. and Ind.*, 1955, 1517.
[16] J. C. Sheehan and G. P. Hess, *J. Amer. Chem. Soc.*, 1955, **77**, 1067.
[17] T. Curtius, *Ber.*, 1902, **35**, 3226.
[18] E. Fischer, *Ber.*, 1906, **39**, 453.
[19] R. W. Roeske, *Chem. and Ind.*, 1959, 1121.
[20] M. Bergmann, L. Zervas, and W. F. Ross, *J. Biol. Chem.*, 1935, **111**, 245.
[21] R. B. Merrifield, *J. Amer. Chem. Soc.*, 1963, **85**, 2149.
[22] T. Wieland and R. Schring, *Annalen*, 1950, **569**, 122.
[23] G. W. Anderson, J. E. Zimmerman, and F. M. Callahan, *J. Amer. Chem. Soc.*, 1963, **85**, 3039.
[24] E. Fischer and F. Fourneau, *Ber.*, 1901, **34**, 2868.
[25] J. Pless and R. A. Boissonas, *Helv. Chim. Acta*, 1963, **46**, 1609.
[26] L. Zervas, D. Borovas, and E. Gazis, *J. Amer. Chem. Soc.*, 1963, **85**, 3660,
[27] E. Wünsch and F. Drees, *Chem. Ber.*, 1966, **99**, 119.
[28] F. Weygand, D. Hoffmann, and E. Wünsch, *Z. Naturforsch.*, 1966, **21b**, 426.
[29] H. Schwarz and K. Arrakawa, *J. Amer. Chem. Soc.*, 1959, **81**, 5691; R. Schwyzer and P. Sieber, *Helv. Chim. Acta*, 1959, **42**, 972.
[30] E. Fischer, *Ber.*, 1915, **48**, 93.
[31] G. Kupryszewski and M. Kaczmarek, *Roczniki Chem.*, 1961, **35**, 931.
[32] R. B. Woodward, R. A. Olofson, and H. Mayer, *J. Amer. Chem. Soc.*, 1961, **83**, 1010.
[33] A. Hillmann-Elies, G. Hillman, and H. Jatzkewitz, *Z. Naturforsch.*, 1953, **8b**, 445.

group is only slightly reduced. Activated esters have displaced the mixed anhydride and dicyclohexylcarbodi-imide methods from their position as the most favoured means of peptide bond formation. Several improvements introduced since 1962 have already become established, notably the solid-phase method and new types of active ester. The azide method, now in use for nearly sixty years, was indispensable for fragment condensation until very recently: since this is no longer the case, it will be interesting to examine its popularity again after a few more years have elapsed.

A. Protective Groups.—A review of the reactions of liquid hydrogen fluoride with (*inter alia*) proteins and peptides has been published.[34] The value of this reagent as a means of effecting deprotection in the final stages of a synthesis is now well appreciated by peptide chemists, but relatively few examples of its application have so far been reported (some of those which have appeared, however, rank among the most impressive achievements of the subject: *e.g.* both of the syntheses of material with ribonuclease activity[35, 36]). Perhaps some of the delay in taking full advantage of the selective vigour of hydrogen fluoride is due to the fact that special techniques and apparatus are required, combined with a certain apprehension about the hazards involved. A recently published book[37] contains a very useful and detailed account of practical matters relating to the use of liquid hydrogen fluoride in peptide synthesis.

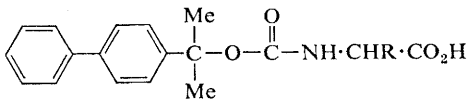
Peptides of biological origin (*e.g.* from proteolytic degradation) are a potentially valuable source of intermediates for peptide synthesis. The selective introduction of protecting groups into polyfunctional peptides is, however, a problem of considerable complexity: preliminary experiments on possible methods of chemical modification have been reported.[38]

Protection of Amino-groups. The principles and practice of amino-group protection have been reviewed.[39]

A useful range of amino-acid derivatives bearing an *N*-2-(*p*-biphenylyl)-isopropoxycarbonyl group (1)[40] have now been described (see Appendix

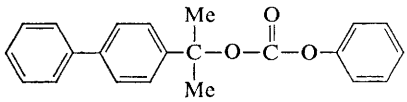
[34] J. Lenard, *Chem. Rev.*, 1969, **69**, 625.
[35] B. Gutte and R. B. Merrifield, *J. Amer. Chem. Soc.*, 1969, **91**, 501.
[36] R. G. Denkewalter, D. F. Veber, F. W. Holley, and R. Hirschmann, *J. Amer. Chem. Soc.*, 1969, **91**, 502; R. G. Strachan, W. J. Paleveda jun., R. F. Nutt, R. A. Vitali, D. F. Veber, M. J. Dickinson, V. Garsky, J. E. Deak, E. Walton, S. R. Jenkins, F. W. Holley, and R. Hirschmann, *ibid.*, p. 503; S. R. Jenkins, R. F. Nutt, R. S. Dewey, D. F. Veber, F. W. Holley, W. J. Paleveda jun., T. Lanza jun., R. G. Strachan, E. F. Schoenewaldt, H. Barkemeyer, M. J. Dickinson, J. Sondey, R. Hirschmann, and E. Walton, *ibid.*, p. 505; D. F. Veber, S. J. Varga, J. D. Milkowski, H. Joshua, J. B. Conn, R. Hirschmann, and R. G. Denkewalter, *ibid.*, p. 506; R. Hirschmann, R. F. Nutt, D. F. Veber, R. A. Vitali, S. L. Varga, T. A. Jacob, F. W. Holley, and R. G. Denkewalter, *ibid.*, p. 507.
[37] J. M. Stewart and J. D. Young, 'Solid Phase Peptide Synthesis', W. H. Freeman and Co., San Fransisco, 1969.
[38] T. A. Backer and R. E. Offord, *Biochem. J.*, 1968, **110**, 3P; R. E. Offord, *Nature*, 1969, **221**, 37.
[39] Y. Wolman in 'Chemistry of the Amino Group', ed. S. Patai, Interscience, London, 1968, p. 669.
[40] P. Sieber and B. Iselin, *Helv. Chim. Acta*, 1968, **51**, 622.

B). This urethane is so acid-sensitive (its acidolysis [40] is several thousand times as fast as that of the t-butoxycarbonyl group) that completely selective



(1)

exposure of α-amino-groups in the presence of t-butyl-based side-chain protection is easily brought about. The use of this means of amino-group protection in solid-phase synthesis is discussed elsewhere (see section 2C). The acid lability of the free acylamino-acids (1), which are generally crystalline, is not so extreme as to render their use inconvenient: they can be kept for months if maintained cold and dry,[41] although isolation and storage as amine salts has been recommended.[40] Reaction of the appropriate amino-acid in dimethylformamide with the carbonate (2) in the



(2)

presence of Triton B [40–42] seems to be the best general means of preparing 2-(*p*-biphenylyl)-isopropoxycarbonylamino-acids, but the corresponding azide [40] gives better results in the cases of tryptophan and tyrosine,[41] and it has been found necessary to substitute triethylamine for Triton B to obtain reasonable yields when dealing with benzyl half esters of dicarboxylic acids. The carbonate reagent (2) is adequately stable only if kept cold and dry;[41, 42] it can be stored in a desiccator at atmospheric pressure and 4 °C for months with negligible detriment, but the decomposition is accelerated if the desiccator is evacuated.[41]

The *N*-*o*-nitrophenylsulphenyl group is very easily cleaved by hydrogen chloride in organic solvents,[26] but the difficulty of achieving selective deprotection in the presence of t-butyl groups, together with the fact that under these acidic conditions undesirable side-reactions can occur (*e.g.* attack on indole rings), has led to investigations of methods such as thiolysis which use non-acidic conditions (see the introductory section of ref. 43 for relevant literature). Thiolytic removal of *o*-nitrophenylsulphenyl groups has recently been used in rather an ingenious fashion (Schemes 1 [44] and 2 [45]) by two groups acting independently. In the procedure shown in

[41] S. S. Wang and R. B. Merrifield, *Internat. J. Protein Res.*, 1969, **1**, 235.
[42] P. Sieber and B. Iselin, *Helv. Chim. Acta*, 1969, **52**, 1525.
[43] K. Poduška and H. M. van den Brink-Zimmermannova, *Coll. Czech. Chem. Comm.*, 1968, **33**, 3769.
[44] J. Šavrda and D. H. Veyrat, *Tetrahedron Letters*, 1968, 6253.
[45] H. Faulstich, *Chimia* (*Switz.*), 1969, **23**, 150.

Scheme 1 enough thiophenol to start the reaction is presumably formed by hydrolysis of the active ester; as soon as thiophenol is present, thiolysis of the protecting group exposes an α-amino-group which is soon acylated with production of more thiophenol, and so on until completion. For the

$$\text{Nps-AA-O}^-\text{Na}^+ + \text{Z-AA'-SPh} \longrightarrow \text{Z-AA-AA'-O}^-\text{Na}^+ + \text{Nps-SPh}$$

Conditions: aq. THF, 70 °C, 6 hr
AA = an amino-acid residue

Scheme 1

$$\text{Nps-AA-OR} + \text{Z-AA'-SPh} \longrightarrow \text{Z-AA-AA'-OR} + \text{Nps-SPh}$$

Conditions: THF, 45 °C, 30 min
AA = an amino-acid residue

Scheme 2

method shown in Scheme 2, some thiophenol (or imidazole, which probably generates thiophenol by reacting with the thiophenyl ester to give an acylimidazole) must be added to initiate reaction. The yields resulting from both of these procedures are good, but the examples so far reported comprise almost exclusively peptides lacking functional side-chains. A variant of the method shown in Scheme 2 has proved useful for cyclisation reactions (see section 2D and Scheme 19). *o*-Nitrophenylsulphenyl groups may also be cleaved (Scheme 3) using the sulphonimide (3), which gives the

$$\underset{(3)}{\text{Nps-NH}\cdot\text{R} + (\text{Ph}\cdot\text{SO}_2)_2\text{NH}} \longrightarrow \underset{(4)}{(\text{Ph}\cdot\text{SO}_2)_2\text{N}^-\text{NH}_3^+\cdot\text{R}} + \text{Nps-N}(\text{SO}_2\cdot\text{Ph})_2$$

Scheme 3

crystalline salts (4) in moderate yield.[43] The original procedure[26] for removal of *o*-nitrophenylsulphenyl groups employed two or more equivalents of hydrogen chloride: only one equivalent of hydrogen chloride is in fact necessary when methanol is the solvent (Scheme 4) since in this case the co-product is (5).[46]

$$\text{Nps-NH}\cdot\text{R} + \text{HCl(1 eq.)} \xrightarrow{\text{MeOH}} \underset{(5)}{\text{Nps-OMe} + \text{Cl}^-\text{NH}_3^+\cdot\text{R}}$$

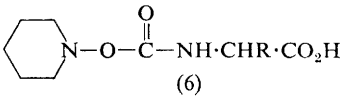
Scheme 4

A method for the direct preparation of t-butoxycarbonyl- and t-amyloxycarbonyl-amino-acids by Schotten–Baumann acylation with the appropriate chloroformates has been published.[47] Conditions [6 hr at room temperature on a column of Zeo-Carb 225 (H^+) in aqueous methanol] for fission of t-butoxycarbonyl groups without disturbing t-butyl esters have been briefly reported.[48]

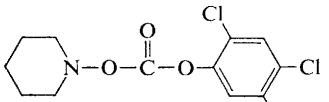
[46] K. Poduška, *Coll. Czech. Chem. Comm.*, 1968, **33**, 3779,
[47] S. Sakakibara, I. Honda, K. Takada, M. Miyoshi, T. Ohnishi, and K. Okamura, *Bull. Chem. Soc. Japan*, 1969, **42**, 809.
[48] C. J. Gray and A. M. Khoujah, *Tetrahedron Letters*, 1969, 2647.

Modified procedures (using *p*-methoxybenzyl chloroformate[49] and *p*-methoxybenzyl 2,4,5-trichlorophenyl carbonate[50] respectively) for the preparation of *p*-methoxybenzyloxycarbonylamino-acids have been described.

N-(Piperidino-oxycarbonyl)-amino-acids (6)[51] can be prepared *via* isocyanato-esters or by use of mixed carbonates such as (7). This protecting



(6)



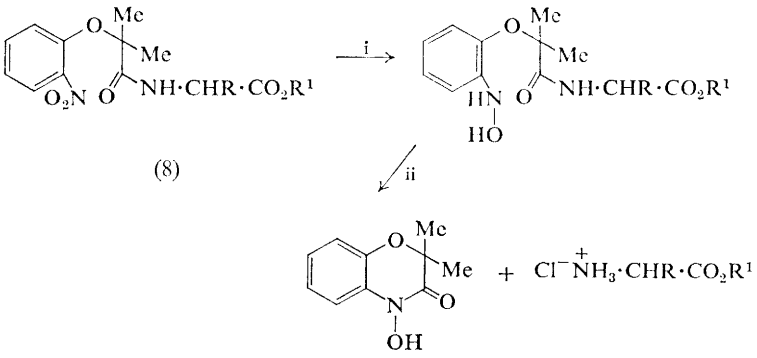
(7)

group is removed under reductive conditions (catalytic hydrogenation, electrolytic reduction, zinc–acetic acid, *etc.*) but resists aqueous alkali and survives severe cold acid treatment (*e.g.* 3 hr in 30% hydrogen bromide in acetic acid) completely unchanged. Since it is a urethane, the group can be expected to prevent racemisation of the residue to which it is attached, and it has been shown that the acylamino-acids (6) do not give oxazolones when activated.[51] In a number of simple examples the derivatives (6) proved very satisfactory and they clearly have some potential as intermediates for peptide synthesis. Because the *N*-(piperidino-oxycarbonyl) group resists acidolysis, benzyl, benzyloxycarbonyl, t-butyl, and t-butoxycarbonyl groups may be removed selectively in its presence. Conversely, it can be cleaved without affecting any of these groups by choice of appropriate reduction conditions. The remarkable acid-insensitivity of the protecting group suggests special usefulness for blocking lysine ε-amino-functions, and this has been confirmed by experiments with lysylglycine as a model.[51] This application may prove particularly valuable in solid-phase synthesis where the cumulative effect of slight loss of ε-benzyloxycarbonyl protection during mild acidolysis of α-amino protection has recently been a cause for some concern (see section 2C).

Further details of the removal of benzyloxycarbonyl groups from methionine-containing peptides by means of catalytic hydrogenation in the presence of boron trifluoride etherate with a palladium catalyst have now appeared.[52] The rates of fission of formyl groups by a series of aromatic amines have been determined.[53] The acid-labile *N*-nitroso-group (previously used in depsipeptide synthesis) has been employed for the *N*-protection of imino-acids,[54] but since the group confers slight optical instability on proline (which is completely optically stable under practically

[49] S. Sofuku, M. Mizumura, and A. Hagitani, *Bull. Chem. Soc. Japan*, 1969, **42**, 278.
[50] E. Klieger, *Annalen*, 1969, **724**, 204.
[51] D. Stevenson and G. T. Young, *J. Chem. Soc.* (*C*), 1969, 2389.
[52] H. Yajima, K. Kawasaki, Y. Kinomura, T. Oshima, S. Kimoto, and M. Okamoto, *Chem. and Pharm. Bull.* (*Japan*), 1968, **16**, 1342.
[53] R. Geiger and W. Siedel, *Chem. Ber.*, 1969, **102**, 2487.
[54] F. H. C. Stewart, *Austral. J. Chem.*, 1969, **22**, 2451.

all circumstances normally encountered in peptide synthesis) it seems unlikely to enjoy much popularity. The new α-methyl-α-(*o*-nitrophenoxy)-propionyl protecting group is very easily removed (in modest yield) by intramolecular cyclisation after reduction of the *o*-nitro-group (Scheme 5).[55] Examples in which amino-acids *N*-protected in this way (8; R^1 = H)

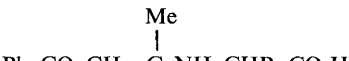


Conditions: i, Zn–NH_4Cl, aq. THF; ii, HCl

Scheme 5

were used in coupling were not described: it can be predicted with confidence based on a knowledge of the behaviour of unsubstituted *N*-acetyl groups that activation of the acylamino-acids (8; R^1 = H) will result in serious racemisation.

The acid-labile protected amino-acids (9) have been recommended for use in solid-phase synthesis.[56a]



(9)

Protection of Carboxy-groups. It frequently happens that the need to activate an acylpeptide followed at some later stage by selective α-amino deprotection makes partially protected derivatives such as (10) or (11)

Boc-[Lys(Z),Glu(OBzl), *etc.*]-OH

(10)

Z-[Lys(Boc),Glu(OBu^t), *etc.*]-OH

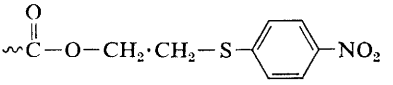
(11)

crucial intermediates in a synthesis; this is the case if azide couplings are to be avoided in a fragment-condensation approach and in the synthesis of

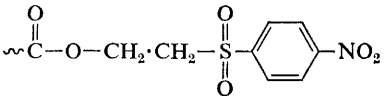
[55] C. A. Panetta, *J. Org. Chem.*, 1969, **34**, 2773.

[56a] G. L. Southard, G. S. Brooke, and J. M. Pettee, *Tetrahedron Letters*, 1969, 3505.

cyclic and sequential polypeptides by the active ester method (see sections 2D and 2E). In the synthesis of compounds of type (10) the use of methyl ester *C*-terminal protection is ruled out by benzyl ester side-chain protection. Acylpeptides of type (11) can in principle be prepared using methyl ester *C*-terminal blocking groups, but the final saponifications are often fraught with difficulty. One approach to this problem is very simple; the *C*-terminal carboxy-group can be left exposed, but this introduces its own problems (*e.g.* solubility). There is therefore a need for new carboxy-protecting groups which can be removed under mild and selective conditions. The β-methylthioethyl group fulfils these requirements (see Volume 1 of these Reports) and so does a recent modification using the β-(*p*-nitrothiophenyl)-ethyl group (12).[56b] The group (12) provides efficient protection during coupling and is easily removed by β-elimination with mild alkali (pH *ca.* 10), after oxidation to the sulphone (13), but the presence of sulphur



(12)



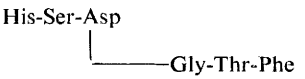
(13)

constitutes a serious restriction on strategy. 4-Picolyl esters [57] (see Volume 1 of these Reports) may also prove to be useful for the selective protection of carboxy-groups, since they are stable to acid but are smoothly removed on catalytic hydrogenation or electrolytic reduction.

The planning of the synthesis of aspartyl peptides requires special care because of the danger of imide formation, which is particularly acute if the aspartylglycine sequence is present. A recent cautionary tale [58] reports attempted syntheses of (14) and (15) by the solid-phase method. Both syntheses gave the same product which was in fact the imide (16). The same

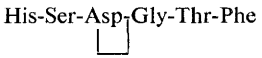
His-Ser-Asp-Gly-Thr-Phe

(14)



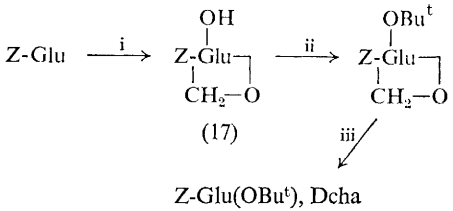
(15)

[56b] M. J. S. A. Amaral, *J. Chem. Soc.* (*C*), 1969, 2495.
[57] R. Camble, R. Garner, and G. T. Young, *J. Chem. Soc.* (*C*), 1969, 1911.
[58] M. A. Ondetti, A. Deer, J. T. Sheehan, J. Pluscec, and O. Kocy, *Biochemistry*, 1968, 7, 4069.



(16)

side-reaction also complicated the successful conventional syntheses [58] of (14) and (15), but only when the carboxy group not intended for peptide bond formation was protected: it may therefore be advisable to leave aspartyl side-chains exposed when possible. Benzyloxycarbonylaspartic and glutamic acids can be condensed with formaldehyde to give 5-oxazolidinones [*e.g.* (17)] which can be used to make ω-half t-butyl esters, as exemplified in Scheme 6 for the glutamic acid case.[59] Unfortunately both of the heterocyclic intermediates in Scheme 6 (overall yield *ca.* 35%) are



Conditions: i, CH_2O–H^+; ii, $(CH_3)_2C{=}CH_2$–H^+; iii, OH^-, then isolate as Dcha salt

Scheme 6

oils. The oxazolidinones (17) cannot be used satisfactorily for the preparation of ω-peptides as the final saponification causes transpeptidation to an α–ω mixture.

Protection of Histidine Side-chains. The most frequently used imidazole-protecting group is the benzyl group, and an alternative preparation of (18) which avoids the use of sodium in ammonia has been reported.[60] However, the main problem with this group is not that sodium in ammonia is required to introduce it but that it is often necessary to resort to this reagent to remove it. The only other method of cleavage (catalytic hydrogenolysis—it even resists liquid hydrogen fluoride) is frequently very slow and in any case is often prevented by circumstances such as the presence of sulphur. Furthermore, the introduction of a benzyl substituent does not impair the basic character of the imidazole side-chain. Weygand's 2,2,2-trifluoro-1-benzyloxycarbonylamino-ethyl group (see Volume 1 of these Reports) does not suffer from these drawbacks but it has recently been pointed out that the differential reactivity to acid of the two protecting groups in, for example, the derivative (19) is marginal,[61] and perhaps not sufficient for use in solid-phase work where the effects of incompletely selective deprotection are cumulative. The crystalline and highly soluble dinitrophenylated

[59] M. Itoh, *Chem. and Pharm. Bull. (Japan)*, 1969, **17**, 1679.
[60] M. A. Tilak and C. S. Hollinden, *Tetrahedron Letters*, 1969, 391.
[61] F. Chillemi and R. B. Merrifield, *Biochemistry*, 1969, **8**, 4344.

Z-His-OH	Boc-His-OH	Boc-His-OH
|	|	|
Bzl	Ztf	Dnp
(18)	(19)	(20)

intermediate (20)[61, 62] may provide an answer to these problems, and its value has been proved in the solid-phase synthesis of a tricosapeptide containing two histidine residues. The 2,4-dinitrophenyl group emerges unchanged from subjection to acidolysis conditions (hydrogen bromide in trifluoroacetic acid) but is smoothly removed by treatment with aqueous mercaptoethanol at pH 8. No account of the use of this group in conventional peptide synthesis has yet been published but the use of catalytic hydrogenation will presumably be prohibited by the nitro-groups.

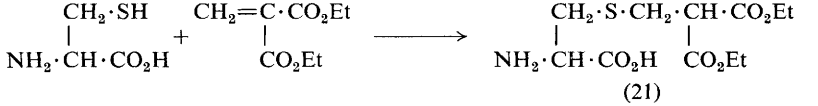
A brief note[63] on the use of the N^{im}-tosyl group has appeared: this group is best removed by brief treatment with hydrogen fluoride, but is also cleaved by the conditions used for ester saponification and (partially) by hydrogen chloride or bromide in organic solvents. The N^{im}-tosyl group is unaffected by trifluoroacetic acid, however, and could therefore be used in conjunction with, for example, t-butoxycarbonyl-α-amino-protection and side-chain protecting groups derived from benzyl alcohol.

Protection of Thiol Groups. The widely used *S*-benzyl group is removable only by use of liquid hydrogen fluoride, which is not quantitative, or with sodium in liquid ammonia. Since in the latter reagent (which is usually troublesome) solvated electrons are probably the effective species, it has been reasoned[64] that electrolytic reduction in liquid ammonia ought to bring about the same cleavage under milder and more easily controlled conditions. Preliminary results[64] are discouraging, but the idea merits fuller examination, although it could be argued that the introduction of the hydrogen fluoride technique has rendered further attempts to improve on reduction in liquid ammonia superfluous, except for cleavage of N^{im}-benzylhistidine.

A large number of alternative methods of *S*-protection have been examined over the years with the objects of avoiding the need for drastic deprotection conditions and of making available a series of protecting groups of differing properties to permit stepwise construction of the cross-linkages of complex disulphide networks. A complete solution to the latter problem still seems distant (see, however, the further details[65] of the tris-cystine peptide synthesis which was discussed in last year's Report), partly because an insufficiently flexible range of selective *S*-deprotection conditions is available. It is therefore of great interest that another base-

[62] M. E. Lombardo, R. Piasio, and J. M. Stewart, quoted in ref. 37, p. 20.
[63] S. Sakakibara and T. Fujii, *Bull. Chem. Soc. Japan*, 1969, **42**, 1466.
[64] D. A. J. Ives, *Canad. J. Chem.*, 1969, **47**, 3697.
[65] R. G. Hiskey, A. M. Thomas, R. L. Smith, and W. C. Jones jun., *J. Amer. Chem. Soc.*, 1969, **91**, 7525.

labile *S*-protecting group (21) has now been devised.[66, 67] *S*-2,2-Diethoxycarbonylethylcysteine (21) is easily prepared as shown in Scheme 7 and can



Conditions: 50% aq. EtOH, 20 °C, 1 hr

Scheme 7

be converted to *N*-alkoxycarbonyl derivatives and used in coupling reactions in the normal way. This method of *S*-protection proved convenient in a new synthesis of glutathione.[66]

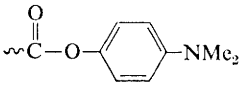
All syntheses of the insulin A chain (which contains four cysteine residues) published until recently used the *S*-benzyl method of protection and sodium–ammonia for cleavage: see ref. 68 for a review of synthetic work on insulin. Recent experiments[69] show that much of the trouble experienced with the sodium–ammonia stage must be attributed to side-reactions (β-elimination?) involving the *S*-benzylcysteine residues. It was found that if natural insulin A-chain tetra-*S*-sulphonate was subjected to reduction with sodium–ammonia, undamaged and pure starting material could be recovered in reasonable yield (55%) after work-up, oxidative sulphitolysis, and chromatography, whereas the yields of the tetra-*S*-sulphonate isolated after similar treatment of synthetic tetra-*S*-benzyl A chain were much lower (10—20%).

Syntheses of insulin A chain using *S*-ethylmercapto[70] and *S*-trityl[71] protection have been reported. In the latter case,[71] trouble with the acidolysis of the *S*-trityl functions was experienced because an equilibrium is set up between free thiol groups, trityl cations, and *S*-trityl groups in the presence of strong acid (see Volume 1 of these Reports). Cleavage by means of mercuric chloride was unsatisfactory because attempted removal of the resulting *S*-chloromercuri-groups caused some desulphurisation. A solution to these problems eventually emerged from a reinvestigation of the effect of strong acid on *S*-trityl groups: it was found that addition of a carbonium ion scavenger (2-ethylphenol) to the acidic medium (hydrogen bromide in trifluoroacetic acid) displaced the equilibrium in favour of the free thiol and brought about smooth deprotection.

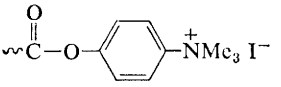
B. Formation of the Peptide Bonds.—*Activated Esters.* Every year which passes produces its crop of new active esters. This year's suggestions include

[66] T. Wieland and A. Sieber, *Annalen*, 1969, **727**, 121.
[67] T. Wieland and A. Sieber, *Annalen*, 1969, **722**, 222.
[68] A. C. Trakatellis and G. P. Schwartz, *Fortschr. Chem. org. Naturstoffe*, 1968, **26**, 120.
[69] Y. Shimonishi, H. Zahn, and W. Puls, *Z. Naturforsch.*, 1969, **24b**, 422.
[70] U. Weber, *Z. physiol. Chem.*, 1969, **350**, 1421.
[71] H. Zahn, W. Dahno, H. Klostermeyer, H. G. Gattner, and J. Repin, *Z. Naturforsch.*, 1969, **24b**, 1127.

esters of *N*-hydroxycarbamates[72] and of ketoximes,[73] but neither of these classes of compound have any outstandingly valuable properties, although they were perfectly satisfactory in simple protected peptide preparations. The reaction of glycine *o*-nitroanilide with active esters can be followed spectrophotometrically, and this may prove useful for evaluating the reactivity of new types of ester.[74] Further examples[75, 76] have been described of the use of 4-(methylsulphonyl)phenyl esters, which can be produced by oxidation of the corresponding unreactive 4-(methylthio)phenyl esters when activation is required (see Volume 1 of these Reports). *p*-Dimethylaminophenyl esters (22) can be used in a similar fashion: in this case quaternisation gives a reactive ester (23).[77] The effects of various catalysts



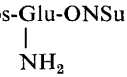
(22)



(23)

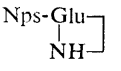
on the aminolysis of benzyloxycarbonylphenylalanine *p*-nitrophenyl ester have been examined.[78]

It has been reported that the succinimido ester (24) undergoes quantita-



(24)

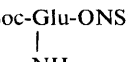
tive conversion to the imide (25) when dissolved in aqueous dimethylformamide [curiously, this was not detected with the asparagine analogue



(25)

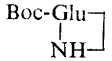
of (24)].[79] This prompted an examination[80] of the behaviour of the ester (26), which had been used[36] in the fragment condensation synthesis of

[72] H. Jeschkeit, *Z. Chem.*, 1969, **9**, 266.
[73] M. Fujino and O. Nishimura, *Chem. and Pharm. Bull. (Japan)*, 1969, **17**, 1937.
[74] D. Merz and H. Determann, *Annalen*, 1969, **728**, 215.
[75] B. J. Johnson and E. G. Trask, *J. Org. Chem.*, 1968, **33**, 4521.
[76] B. J. Johnson, *J. Org. Chem.*, 1969, **34**, 1178.
[77] Yu. V. Mitin and L. B. Nadeshdina, *Zhur. obshchei Khim.*, 1968, **38**, 2627.
[78] N. Nakimozo, *Bull. Chem. Soc. Japan*, 1969, **42**, 1071, 1078.
[79] C. Meyers, R. T. Havran, I. L. Schwarz, and R. Walter, *Chem. and Ind.*, 1969, 136.
[80] R. S. Dewey, H. Barkemeyer, and R. Hirschmann, *Chem. and Ind.*, 1969, 1632.



(26)

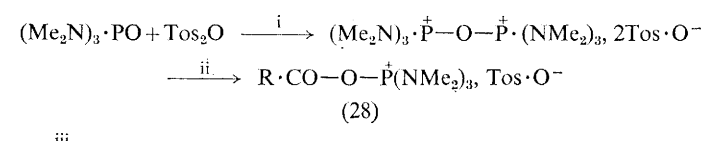
ribonuclease. It appears that although some imide (27) is formed from (26) in aqueous dimethylformamide, the rate of this side-reaction is slow



(27)

compared to intermolecular acylation of α-amino-groups and it does not interfere seriously with coupling. Furthermore, the imide (27) is exceedingly unreactive towards α-amino-esters, so there is no cause for concern about the danger of isoglutamine peptides contaminating the product.[80]

Other Methods. Reports of new methods of peptide synthesis frequently do not command the attention they may deserve because the evaluation has largely been confined to the synthesis of small oligopeptides possessing only those side-chains which do not usually cause trouble. This criticism cannot be levelled at a preliminary account[81] of the use, for coupling, of acyloxyphosphonium salts (28) which can be generated in a number of ways: the preferred procedure is outlined in Scheme 8. Yields were

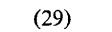


(28)

Conditions: i, Dry, room temperature, 15 min; ii, $R \cdot CO_2^-$, 0 °C, 5—10 min; iii, $R' \cdot NH_2$–Et_3N

Scheme 8

uniformly excellent in more than twenty examples of the use of this method for coupling (including the preparation of several histidine and tryptophan derivatives and chain lengths up to the octapeptide level). The isolation of fully protected peptide products is facilitated by the water solubility of all the co-products. Synthesis of the protected tetrapeptide amide (29) from



(29)

two dipeptides by this method gave material of optical rotation practically equal to that obtained by a stepwise active ester régime. No racemate at all was detected in the Young test, in which the danger of racemisation is magnified to about ten times that which exists in the coupling of acyl-peptides. The freedom of the method from racemisation can be explained

[81] G. Gawne, G. W. Kenner, and R. C. Sheppard, *J. Amer. Chem. Soc.*, 1969, **91**, 5669.

in terms of intramolecular base catalysis providing activation towards aminolysis but not to oxazolone formation, or by the possibility that primary nucleophilic attack occurs at phosphorus. The only serious problems are with histidine (in some cases) and the fact that toluenesulphonic anhydride in hexamethylphosphoramide can cause oxidation of hydroxylic side-chains, but no doubt further examination with more comprehensive protection will provide a solution here. This important and ingenious new method provides a much needed alternative means of directly activating and coupling acylpeptide fragments which terminate in optically active α-amino-acids. The only other racemisation-free procedures available for such critical steps are the azide method and the dicyclohexylcarbodi-imide-*N*-hydroxysuccinimide technique. The last-mentioned method can be accompanied by unexpected side-reactions under some circumstances.[82]

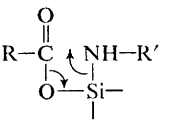
An interesting new reagent, so far applied only for simple amide preparations, is silicon tetrachloride,[83] which may well be worth investigating as a peptide coupling reagent. There is some evidence that the reaction, which is performed as shown in Scheme 9, proceeds *via* intramolecular collapse

$$R\cdot CO_2H + R'\cdot NH_2 + SiCl_4 \longrightarrow R\cdot CO\cdot NH\cdot R' + HCl + SiO_2$$

Conditions: pyridine

Scheme 9

of an adduct (30) rather than by straightforward aminolysis of an acyloxysilane. If this is the case, then the balance between the rates of oxazolone



(30)

formation and coupling might be expected to differ from that observed with coupling methods based on simple activation by electronic effects, although it is difficult to predict *a priori* whether this balance will be more or less favourable. Difficulties arising from incomplete removal of the co-product seem unlikely to arise here, as it is silica.

A review of the chemistry of carbonic mixed anhydrides[84] may be of interest to those engaged in the further development of anhydride coupling methods. Symmetrical anhydrides such as (31) give excellent yields in coupling[85] but only half the acylamino-acid used to make the anhydride

[Boc-Thr(Ztf)]$_2$O (31)

Me·C≡C·NMe$_2$ (32)

[82] F. Weygand, W. Steglich, and N. Chytil, *Z. Naturforsch.*, 1968, **23b**, 1391.
[83] T. H. Chan and L. T. L. Wang, *J. Org. Chem.*, 1969, **34**, 2766.
[84] D. S. Tarbell, *Accounts Chem. Res.*, 1969, **2**, 296.
[85] F. Weygand and C. Di. Bello, *Z. Naturforsch.*, 1969, **24b**, 314.

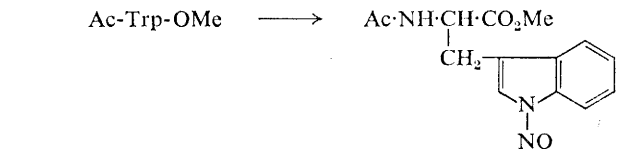
[which is prepared by means of (32) immediately before use] is available for acylation purposes.

It has been found that *N*-nitrosation of indole residues is a serious side-reaction which occurs under the conditions used for generation of acyl-peptide azides from hydrazides.[86] This was detected when the susceptibility of natural ovine/bovine gastrin I (33) to subtilisin was compared to that of

$$\text{H-Glu-Gly-Pro-Trp-Val-Glu}_4\text{-Ala}_2\text{-Tyr-Gly-Trp-Met-Asp-Phe-NH}_2$$

(33)

synthetic material which had been prepared by an azide fragment condensation strategy. The synthetic material, in contrast to the natural hormone, was incompletely degraded at one of the tryptophyl peptide bonds (Trp^4–Val^5), and modification of tryptophan residues under azide-forming conditions seemed a plausible explanation. This was corroborated by the reaction shown in Scheme 10.[86] It has been recognised before[87] that this



Conditions: HCl–$NaNO_2$ in DMF at 0 °C; practically quantitative

Scheme 10

side-reaction can occur but since the cases then examined were easily purified small peptides, it did not appear to be serious at that time. In the present case it was necessary to redesign the synthesis and avoid the problem by obviating the need for azide couplings.

Other relevant papers have included studies of aspects of the chemistry of carbonic anhydrides,[88] the use of succinimido and phthalimido chloroformates for mixed anhydride formation,[89] new coupling reagents (modified isoxazolium salts,[90] phosphonitrilic chloride,[91] and triaryl phosphites[92]), and the synthesis of serine peptides through β-lactone intermediates.[93]

Racemisation. Conditions for the detection on an amino-acid analyser of as little as 0·1% contamination of the tripeptide (34) by its diastereoisomer (35) have been described.[94] This analytical system was used[94] to determine,

[86] K. L. Agarwal, G. W. Kenner, and R. C. Sheppard, *J. Chem. Soc.* (*C*), 1969, 954.
[87] H. Zahn and D. Brandenburg, *Annalen*, 1966, **692**, 220.
[88] H. Determann and I. Kahle, *Annalen*, 1969, **725**, 203.
[89] H. Gross and L. Bilk, *Annalen*, 1969, **725**, 212.
[90] T. R. Govindachari, S. Rajappa, A. S. Akerkar, and V. S. Iyer, *Indian J. Chem.*, 1968, **6**, 557.
[91] K. C. Das, Y. Y. Lin, and B. Weinstein, *Experientia*, 1969, **25**, 1238.
[92] Yu. V. Mitin and O. V. Glinskaya, *Zhur. obshchei Khim.*, 1969, **39**, 1287.
[93] M. Miyoshi, T. Fujii, N. Yoneda, and K. Okamura, *Chem. and Pharm. Bull.* (*Japan*), 1969, **17**, 1617.
[94] N. Izumiya and M. Muraska, *J. Amer. Chem. Soc.*, 1969, **91**, 2391.

Gly-L-Ala-L-Leu
(34)

Gly-D-Ala-L-Leu
(35)

after hydrogenolysis of the protected product, the amount of racemisation arising from the coupling shown in Scheme 11. The results corroborate

Conditions: various coupling methods

Scheme 11

previous determinations by different methods with other models, and it is reassuring to know that 0·0% racemisation was detected in the product from the dicyclohexylcarbodi-imide–*N*-hydroxysuccinimide procedure (conducted in this work with equivalent amounts of the reactants at 0 °C in tetrahydrofuran).

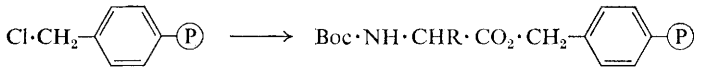
Studies on the following topics have also been published: racemisation in the isoxazolium salt,[95] isocyanate,[96] and isothiocyanate [96] methods and in the so-called oxidation–reduction method;[97] racemisation in the preparation and coupling of active esters;[98] the racemisation of benzyloxythiocarbonylphenylalanine [99] and of phenyl esters of benzyloxycarbonyl-*S*-benzylcysteine;[100] racemisation during the coupling of acetoacetylvaline and its active esters.[101]

C. Repetitive Methods of Peptide Synthesis.—*Solid-phase synthesis.* A comprehensive and critical review of solid-phase peptide synthesis by the inventor is now available,[102] and a book [37] on the subject has been published. This book gives a somewhat unfortunate impression of being a simple guide for do-it-yourself peptide synthesis, but contains all the essential information necessary for a beginner: it emphasises practical matters, whilst skimming over the drawbacks. A good deal of the practical information which is given on laboratory techniques and reagents is also relevant to the wider field of peptide synthesis in general, and is not available in a collected form elsewhere. Anfinsen has reviewed [1, 2] the progress of his multi-disciplinary assault on staphylococcal nuclease (see also section 3D)

[95] R. B. Woodward and D. J. Woodman, *J. Org. Chem.*, 1969, **34**, 2742.
[96] J. H. Jones and R. Fairweather, *Makromol. Chem.*, 1969, **128**, 279.
[97] T. Mukaiyama, R. Matsueda, H. Maruyama, and M. Ueki, *J. Amer. Chem. Soc.*, 1969, **91**, 1554.
[98] J. Kovacs, L. Kisfaludy, M. Q. Ceprini, and R. H. Johnson, *Tetrahedron*, 1969, **25**, 2555.
[99] I. Z. Siemion, D. Konopinska, and A. Dzugaj, *Roczniki Chem.*, 1969, **43**, 989.
[100] V. F. Martynov and M. A. Smartsev, *Zhur. obshchei Khim.*, 1969, **39**, 940.
[101] C. di Bello, F. Filira, V. Giormani, and F. D'Angeli, *J. Chem. Soc.* (*C*), 1969, 350.
[102] R. B. Merrifield, *Adv. Enzymology*, 1969, **32**, 221.

in lectures given in mid-1968. This work includes synthetic approaches using the solid-phase method, and in one of his reviews Anfinsen [2] incorporates a stimulating discussion of the scope and limitations of the Merrifield method together with speculations and preliminary accounts of some experiments which may lead to solutions of some of the problems.

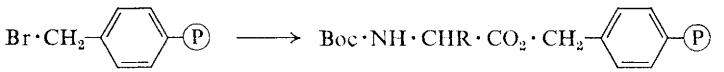
The original procedure [21] for linking the first amino-acid residue to the cross-linked polystyrene resin uses rather drastic conditions (Scheme 12)



Conditions: Boc·NH·CHR·CO_2H, Et_3N–EtOH, 80 °C, 48 hr. Ⓟ = polymer

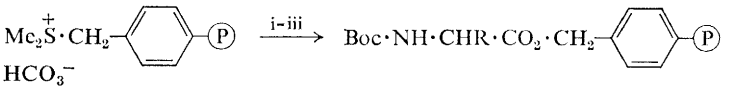
Scheme 12 [21]

and does not give very good yields. Some recently introduced modifications of this step are illustrated in Schemes 13—15. The very mild treatment used



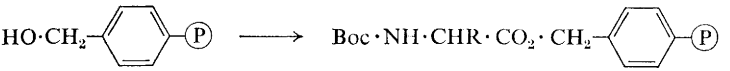
Conditions: Boc·NH·CHR·CO_2H, base, DMF, overnight, room temperature. Ⓟ = polymer

Scheme 13 [103]



Conditions: i, Boc·NH·CHR·CO_2H; ii, dry; iii, 80 °C, *ca.* 48 hr. Ⓟ = polymer

Scheme 14 [104]



Conditions: Boc·NH·CHR·CO_2H, carbonyldi-imidazole. Ⓟ = polymer

Scheme 15 [105]

in the case of the hydroxymethyl polymer (Scheme 15) is compatible with employment of the *o*-nitrophenylsulphenyl group for α-amino-group protection.[106] A synthesis of oxytocin [107] on a hydroxymethylated carrier (excess hydroxy-groups were blocked by acetylation) has been published.

The most commonly used protective group strategy in solid-phase synthesis has hitherto involved α-t-butoxycarbonyl groups in juxtaposition with benzyl-based side-chain protection, but the observation [108] that some

[103] M. A. Tilak, *Tetrahedron Letters*, 1968, 6323.
[104] L. C. Dorman and J. Love, *J. Org. Chem.*, 1969, **34**, 158.
[105] M. Bodanszky and J. T. Sheehan, *Chem. and Ind.*, 1966, 1597.
[106] G. Losse and K. Neubert, *Z. Chem.*, 1968, **8**, 387.
[107] H. C. Beyerman and R. A. in't Veld, *Rec. Trav. chim.*, 1969, **88**, 1019.
[108] A. Yaron and S. F. Schlossmann, *Biochemistry*, 1968, **7**, 2673.

loss of ε-benzyloxycarbonyl groups occurs under the conditions normally used for removal of α-t-butoxycarbonyl groups has undermined existing confidence in the approach. Exposure of side-chain nucleophiles could lead to the formation of branched peptides, and indeed these were detected by one group [108] (but not by another [109]) in the solid-phase synthesis of lysine oligomers using α-t-butoxycarbonyl-ε-benzyloxycarbonyl-lysine for the addition of each residue. The danger of exacerbating the final purification problem by permitting the intervention of chain branching clearly cannot be ignored. A solution to this situation obviously lies in an increase of the differential stability towards acidolysis of the protective groups used for the α-amino- and ε-amino-functions. This can be achieved either by use of more acid-resistant side-chain blocking groups (*e.g.* *p*-nitrobenzyloxycarbonyl [108] or trifluoroacetyl [2]) or by capitalising on recent progress in the development of highly acid-labile protecting groups, *e.g.* by using the *o*-nitrophenylsulphenyl or 2-(*p*-biphenylyl)-isopropoxycarbonyl groups for α-amino-protection. The use of the last protecting group has been further evaluated (Scheme 16):[41] although this example contains no lysine there can

$$\text{Boc-Arg}(NO_2)\text{—O—Ⓟ} \xrightarrow{\text{i}} \text{TFA,Arg}(NO_2)\text{—O—Ⓟ}$$
$$\xrightarrow{\text{ii}} \text{Glu-(OBzl)-Arg}(NO_2)\text{-Val-Glu(OBzl)-Trp-Leu-Arg}(NO_2)\text{—O—Ⓟ}$$
$$\xrightarrow{\text{iii}} \text{Glu-Arg-Val-Glu-Trp-Leu-Arg}$$

Conditions: i, 50% TFA–CH_2Cl_2; ii, wash, neutralise, wash, Bpoc-amino-acid–DCCI, wash, TFA (1·5%)–mercaptoethanol (1%)–CH_2Cl_2, wash, neutralise, then repeated (automated) 5 more times; iii, HF. Ⓟ = polymer

Scheme 16

be no doubt that the very mild deprotection conditions would not disturb any ε-benzyloxycarbonyl groups.

The oxidative degradation of tryptophan residues which can occur under acidic conditions gives rise to special problems in solid-phase synthesis.[110] Thus a recent attempted synthesis of horse heart cytochrome *c* [5] (see also section 3A) had to be abandoned because side-reactions became excessive after the introduction of the tryptophan residue. These side-reactions can be avoided if mercaptoethanol is added to the acidic medium used for deprotection.[110, 41] Fortunately, tryptophan residues suffer no appreciable damage under the conditions used for cleaving the peptide–polymer link with anhydrous hydrofluoric acid [41] (see Scheme 16).

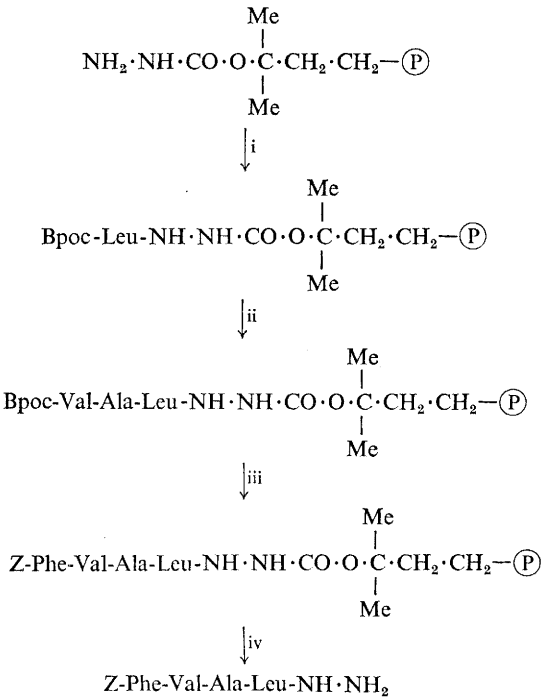
Protection of histidine side-chains with 2,4-dinitrophenyl groups (see section 2A) has been shown to be both convenient and satisfactory.[61, 62]

There have been no substantial contributions to the general and crucial problem of achieving quantitative acylation at every coupling step of a solid-phase synthesis. It has been shown that studies with model compounds in homogeneous solution do not necessarily form a reliable basis for choosing suitable activated derivatives for use in solid-phase synthesis: thus, for

[109] O. Grahl-Nielsen and G. L. Tritsch, *Biochemistry*, 1969, **8**, 187.
[110] G. R. Marshall, *Adv. in Experimental Medicine and Biology*, 1967, **2**, 48.

example, t-butoxycarbonyl-leucine pentachlorophenyl ester is more reactive than the corresponding *o*-nitrophenyl ester towards glycine t-butyl ester in solution but less reactive towards resin-bound glycine.[111]

The rather drastic acidic conditions usually used for removing the peptide from the Merrifield resin (hydrogen bromide in trifluoroacetic acid, or liquid hydrogen fluoride) simultaneously remove the majority of other protecting groups. This is very convenient unless subsequent synthetic steps by conventional methods of peptide synthesis are planned: in this case a partially protected peptide derivative will be required. Hydrazinolysis of the peptide–polymer link is one way of obtaining an intermediate suitable for use in classical procedures; an alternative new route[112] to acylpeptide hydrazides uses a polymeric analogue of the t-butoxycarbonylhydrazide 'safety-catch' method of *C*-terminal protection (Scheme 17).



76% overall yield; crystalline

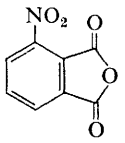
Conditions: i, Bpoc-Leu–DCCI; ii, 0·5% TFA–CH_2Cl_2, neutralise, Bpoc-Ala–DCCI, *etc.*; iii, 0·5% TFA–CH_2Cl_2, neutralise, Z-Phe–DCCI; iv, 50% TFA—CH_2Cl_2. Ⓟ = polymer

Scheme 17

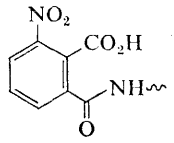
[111] M. Bodanszky and R. J. Bath, *Chem. Comm.*, 1969, 1259.
[112] S. Wang and R. B. Merrifield, *J. Amer. Chem. Soc.*, 1969, **91**, 6488.

Since the purification of solid-phase products with up to, say, ten residues presents in practice no insuperable problem (in several oxytocin syntheses, for example, ammonolytic cleavage from the resin has given crystalline material directly; see Volume 1 of these Reports), rapid synthesis on a polymeric support is a potentially valuable source of intermediates for unambiguous classical synthesis.

If the very reactive acylating agent (36) is used to block residual amino-groups after every coupling step, then the free peptide which is released by the final acidolysis is contaminated only with incomplete sequences carrying an *N*-terminal 2-carboxy-3-nitrobenzoyl group (37). The contaminants



(36)



(37)

are thus labelled by their spectroscopic properties, and the presence of a relatively strongly acidic function (pK_a *ca.* 2—2·5) simplifies their separation by ion-exchange chromatography.[113] This idea has so far only been tried in the case of one pentapeptide, where it was very successful, but it does seem a potentially valuable means of facilitating the purification of low molecular weight solid-phase products. Presumably a recalcitrant amino-group which does not partake in coupling may also be unable to react with (36), so the method may suffer from the very defect it aims to ameliorate. It can also be anticipated that with larger peptides the difference between the desired peptide and the blocked 'error' peptides will not be adequate to assist their separation greatly.

The solid-phase synthesis of a medium-size or large peptide with presently available methods is bound to give a mixture which is unlikely to be successfully resolved by conventional separation techniques because its components are too similar (*i.e.* related by one or more amino-acid deletions). Recent experiments,[2, 114] however, indicate that 'functional purification' may offer some hope here, at least in the special cases to which it is applicable. 'Functional purification' employs the highly specific interactions which occur in systems such as ribonuclease S-protein : S-peptide. A solid-phase preparation of an active fragment (1—15) of ribonuclease S-peptide has been purified very successfully by chromatography on a column of insolubilised ribonuclease S-protein.[114]

It is usually held that solid-phase reaction mixtures are best shaken, not stirred.[37] This is because of the important consideration of preserving the physical form of the resin: if the resin is pulverised by excessive agitation there is a danger that fine particles will block the sinter or be lost through it.

[113] T. Wieland, C. Birr, and H. Wissenbach, *Angew. Chem.*, 1969, **81**, 782.
[114] I. Kata and C. B. Anfinsen, *J. Biol. Chem.*, 1969, **244**, 5849.

A detailed design of a new type of manual reaction vessel (and ancillary glassware) in which the reaction mixture is stirred has been published:[109] scale drawings sufficient to enable reproduction by a skilled glassblower are given.

Other relevant papers include contributions on the following: analytical techniques for following the progress of a solid-phase synthesis;[115, 116] the synthesis of peptides on dextrans [117] (see also Table 1, ref. 102); the use of enamine protection with a benzhydryl resin;[56a] the use of bromoacyl-polystyrene resins [118] (see also Volume 1 of these Reports); the use of the 2,2,2-trifluoro-1-benzyloxycarbonylaminoethyl group for hydroxy-group protection in solid-phase synthesis;[119] and catalysis of alcoholytic cleavage from the resin with anion exchange resins.[120]

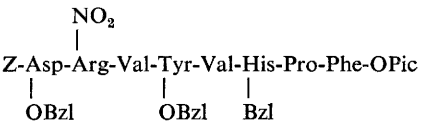
Other Repetitive Methods. Additional details of the use of soluble polystyrene as a polymeric support are now available,[121, 122] and the conditions necessary for successful use of the method have been delineated.[123] One important point seems to be that the excess chloromethyl groups remaining after attachment of the first amino-acid residue must be consumed (by treatment with acetate) before any other synthetic steps are undertaken, or subsequent cross-linking insolubilises the polymer.[123]

Only one further example of the rapid *N*-carboxy-anhydride method has appeared: the two peptides required for the synthesis of the insulin fragment (38) were prepared in this way.[124]

H-Cys-Asn-OH
|
H-Leu-Val-Cys-Gly-Glu-OH
(38)

A detailed paper [57] on the use of 4-picolyl esters (see Volume 1 of these Reports) has been published, and the method has been used in a synthesis of 5-valine-angiotensin II.[125] The overall yield of the protected octapeptide (39) was 34%, based on the protected phenylalanine starting material. Hydrogenation of (39) gave *crude* free hormone with satisfactory elemental and amino-acid analyses and high (73%) biological activity.

[115] F. Weygand and R. Obermeier, *Z. Naturforsch.*, 1968, **23b**, 1390.
[116] K. Esko, S. Karlsson, and J. Porath, *Acta Chem. Scand.*, 1968, **22**, 3342.
[117] G. P. Vlasov and A. Yu. Bilibin, *Izvest, Akad. Nauk S.S.S.R., Ser. khim.*, 1969, 1400.
[118] T. Mizoguchi, K. Shigezane, and N. Takamura, *Chem. and Pharm. Bull. (Japan)*, 1969, **17**, 411.
[119] K. P. Polzhofer, *Tetrahedron*, 1969, **25**, 4127.
[120] W. Pereira, V. A. Close, E. Jellum, W. Patton, and B. Halpern, *Austral. J. Chem.*, 1969, **22**, 1337.
[121] Yu. A. Ovchinnikov, A. A. Kiryushkin, and I. V. Kozhevnikova, *Zhur. obshchei Khim.*, 1968, **38**, 2636.
[122] Yu. A. Ovchinnikov, A. A. Kiryushkin, and I. V. Kozhevnikova, *Zhur. obschhei Khim.*, 1968, **38**, 2631.
[123] B. Green and L. R. Garson, *J. Chem. Soc.* (*C*), 1969, 401.
[124] D. F. Veber, R. Hirschmann, and R. G. Denkewalter, *J. Org. Chem.*, 1969, **34**, 753.
[125] R. Garner and G. T. Young, *Nature*, 1969, **222**, 177.



(39)

D. Synthesis of Homodetic Cyclic Peptides.—Synthesis *via* peptide active esters as generalised in Scheme 18 remains the most frequently employed route: 21 out of 33 cyclopeptide preparations described during 1969 were performed by some variation on this theme. The intermediates (40) are of

$$\underset{(40)}{HX,AA_1\text{-}AA_2\ldots AA_x\text{-}OY} \longrightarrow \textit{cyclo}\text{-}(AA_1\text{-}AA_2\ldots AA_x)+, \text{ in some cases, } \textit{cyclo}\text{-}(AA_1\text{-}AA_2\ldots AA_x)_2$$

Conditions: base, under high dilution conditions. AA = an amino-acid residue; X = anion of a strong acid; Y = an active ester group

Scheme 18

the same general structure as the monomers used for most sequential polypeptide syntheses (see section 2E), but the limitations on permissible strategy which are imposed by optical purity considerations are not so restrictive in peptide cyclisation. This is because small amounts of diastereoisomeric contaminants can usually be separated from cyclic peptides by the standard purification techniques of organic chemistry. In planning the synthesis of a cyclopeptide it is not absolutely essential to effect ring closure by a racemisation-free method or through the carboxy-group of glycine, an imino-acid, or a side-chain, although it is clearly more desirable to *preclude* racemisation if one of these ways offers itself than merely to *minimise* it, with consequent uncertainties about the optical integrity of the product. In point of fact, the majority of the naturally occurring homodetic cyclopeptides so far discovered have at least one link in the ring which could be constructed with no risk whatever of racemisation. The only substantial exception to this generalisation is the polymixin–circulin B group of antibodies which lack imino-acids and glycine, and in which all the peptide bonds in the link involve α-carboxy-groups.

An elegant new method [126] of cyclisation uses *o*-nitrophenylsulphenyl-peptide *p*-nitrothiophenyl esters (41) as shown in Scheme 19. (See also

$$\underset{(41)}{Nps\text{-}AA_1\text{-}AA_2\ldots AA_x\text{-}SNp} \longrightarrow \textit{cyclo}\text{-}(AA_1\text{-}AA_2\ldots AA_x)$$

Conditions: HSNp and/or imidazole under high dilution condition. AA = an amino-acid residue

Scheme 19

[126] H. Faulstich, H. Trischmann, and T. Wieland, *Tetrahedron Letters*, 1969, 4131.

section 2A and Scheme 2.) The *o*-nitrophenylsulphenyl group is removed by thiolysis (*in situ* by *p*-nitrothiophenol, which is preferably deliberately added, with or without imidazole) giving a free amino-group which immediately leads to cyclic material. The yields afforded by this procedure are satisfactory (see *e.g.* the case of antamanide, below) but only preliminary details of three examples have so far been reported.[126]

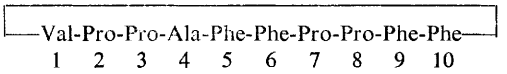
An alternative to the active ester approach is to use 'direct' coupling methods, under conditions of high dilution, on the free peptide: recent applications of such methods include polymerisation *via* carbonic anhydrides,[127] by use of dicyclohexylcarbodi-imide either alone[128, 129] or with added *N*-hydroxysuccinimide,[130, 131] and use of a water-soluble carbodiimide.[132] Of these 'direct' methods, the dicyclohexylcarbodi-imide–*N*-hydroxysuccinimide cyclisation procedure seems especially promising, since it not only gave relatively high yields in the two cases for which it has been described (see below) but is, by analogy with experiments in model acyclic systems,[28] also essentially free from the danger of racemisation.

The yields in peptide cyclisations are never 'good' in the usual sense of that term when used by organic chemists but 30% is generally regarded as a satisfactory outcome; sometimes the yields are very low (in a recent example[133]—the synthesis of *O*-acetyl-evolidine by an active ester cyclisation—the yield was only 1%). It is difficult to generalise about the comparative merits of the various available methods in this respect, as the relative importance of the many other factors which influence the balance between cyclisation, cyclodimerisation, polymerisation, and side-reactions is imponderable at present; among these factors are the number, nature, and configuration of the residues in the peptide starting material, the position at which ring closure is performed, and the concentrations of the reagents. Furthermore, the various yields recorded in the literature no doubt give a distorted picture since they are bound to be greatly influenced by problems of isolation and purification which have nothing to do with the cyclisation *per se*. What is required to put the choice of method on a more rational footing is an exhaustive study of the application of the different techniques to the synthesis of one cyclic peptide. Even this would not be entirely satisfactory, since in view of the subtlety of some of the factors involved, it could well be that the best method for one cyclic peptide might be wholly unsatisfactory for another. With all these reservations in mind, however, it is nevertheless interesting to compare the yields (Table 2) of antamanide (42) obtained by different routes.

[127] T. Wieland, J. Faesal, and W. Konz, *Annalen*, 1969, **722**, 197.
[128] G. Losse and H. Raue, *Tetrahedron*, 1969, **25**, 2677.
[129] K. D. Kopple, M. Ohnishi, and A. Go, *Biochemistry*, 1969, **8**, 4087.
[130] W. König and R. Geiger, *Annalen*, 1969, **727**, 125.
[131] T. Wieland, C. Birr, and F. Flor, *Annalen*, 1969, **727**, 130.
[132] E. C. Jorgensen and W. Patton, *J. Medicin. Chem.*, 1969, **12**, 935.
[133] F. H. C. Stewart, *Austral. J. Chem.*, 1969, **22**, 2663.

Table 2 *Yields of antamanide* (42) *obtained using various cyclisation procedures*

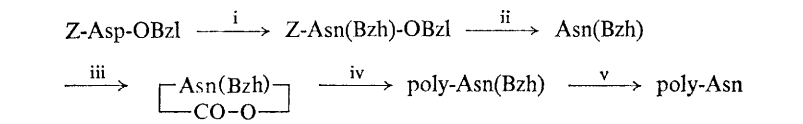
Peptide bond constructed in final step	*ONp method*	*SPh method*	*DCCI–HONSu method*	*Nps-SPh method*	*Carbonic anhydride*	*Ethoxy-acetylene*
Ala[4]-Phe[5]	—	—	36·5% [130]	—	—	—
Phe[5]-Phe[6]	7% [127]	40% [127]	27% [131]	30% [126]	25% [127]	None isolated [127]
Pro[8]-Phe[9]	29% [127]	—	—	—	10·5% [127]	—



(42)

E. Synthesis of Polymeric Models for Studies in Protein Chemistry.—A book (in Russian) on this subject has been published.[134]

Polyamino-acids. Although the *N*-carboxy-anhydrides of asparagine and glutamine have been reported,[135] difficulties in their preparation due to dehydration of the amide side-chain have been encountered,[136] and conversion of these *N*-carboxy-anhydrides to the corresponding polyamino-acids has not been described. An indirect route to polyasparagine and polyglutamine employs ammonolysis of side-chain alkyl esters,[137] but this is not quantitative, and prolonged exposure to ammonia is likely to result in some degradation of the peptide backbone.[138] A new route to poly-asparagine and polyglutamine has now been devised (shown in Scheme 20



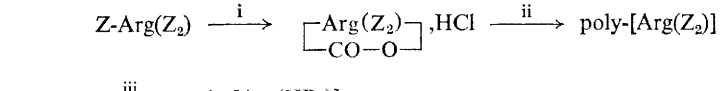
Conditions: Bzh-NH_2, DCCI; ii, H_2–Pd; iii, $COCl_2$–dioxan; iv, Et_3N–dioxan; v, HF

Scheme 20

for the case of asparagine):[139] the diphenylmethyl-protected *N*-carboxy-anhydrides are easily prepared crystalline derivatives, and although the molecular weights of the polymers they have afforded so far are only moderate (*ca.* 5×10^3), no doubt this will be improved by adjustment of the polymerisation conditions.

[134] K. T. Poroshin and V. A. Shibnev, 'Sintez Poliaminokislotti i Regulyarnykh Polipeptidov', Dushanbe, 1968.
[135] R. G. Denkewalter, H. Schwam, R. G. Strachan, T. E. Beesley, D. F. Veber, E. F. Schoenewaldt, H. Barkemeyer, W. J. Palevda jun., T. A. Jacob, and R. Hirschmann, *J. Amer. Chem. Soc.*, 1966, **88**, 3163.
[136] M. Wilchek, S. Ariely, and A. Patchornik, *J. Org. Chem.*, 1968, **33**, 1258.
[137] V. Bruckner, J. Kovacs, and K. Kovacs, *J. Chem. Soc.*, 1953, 1512.
[138] *cf.* A. Kotai, *Acta Chim. Acad. Sci. Hung.*, 1967, **54**, 65.
[139] S. Ariely, M. Fridkin, and A. Patchornik, *Biopolymers*, 1969, **7**, 417.

The preparation of polyarginine also poses considerable problems: the best method hitherto available has been ω-guanylation of polyornithine.[140] It has recently been announced that the crystalline *N*-carboxy-anhydride of *N*-ω,ω′-dibenzyloxycarbonylarginine can be converted to a homopoly-amino-acid (the molecular weight was claimed to be 3×10^4 but this figure must be viewed with caution, as it was based on a carboxy end-group titration) which gives polyarginine hydrobromide on treatment with hydrogen bromide in acetic acid (Scheme 21).[141]



Conditions: i, $SOCl_2$; ii, Ag_2O, then Et_3N–dioxan; iii, HBr–AcOH

Scheme 21

Reasonably satisfactory routes *via* *N*-carboxy-anhydrides are thus now available for the preparation of homopolymers and 'random' copolymers of all the amino-acids which are commonly found in proteins.

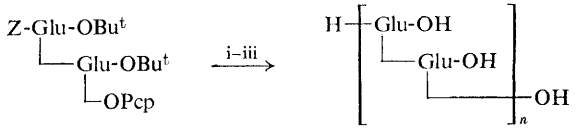
The preparation of poly-α-amino-acids by heating *N*-dithiocarbonyl-ethoxycarbonylamino-acids (43) has been studied in great detail by Higashimura and his co-workers,[142] who concluded that the polymerisation proceeded *via* an isothiocyanato-acid (44). This view now appears to be untenable since recent experiments[96] with model systems show that the

(43)

(44)

formation of peptide bonds by use of isothiocyanates is extremely slow and accompanied by gross racemisation, whereas the thermal polymerisation of the derivatives (43) is fast and gives optically pure poly-α-amino-acids.

An improved synthesis (Scheme 22) of poly-γ-glutamic acid (described for both the all-D- and all-L-polymers) uses t-butyl groups for protection of the α-carboxy-groups:[143] since the final deprotection was performed



Conditions: i, H_2–Pd; ii, Et_3N–DMF; iii, TFA

Scheme 22

[140] E. Katchalski and P. Spitnik, *J. Amer. Chem. Soc.*, 1951, **73**, 3992.
[141] T. Hayakawa, Y. Kondo, H. Yamamoto, and Y. Murakami, *Bull. Chem. Soc. Japan*, 1969, **42**, 479.
[142] H. Kato, T. Higashimura, and S. Okamura, *Makromol. Chem.*, 1967, **109**, 9, and references cited therein.
[143] J. Kovacs, G. N. Schmit, and B. J. Johnson, *Canad. J. Chem.*, 1969, **47**, 3690.

under mildly acidic conditions, transpeptidation, which was a problem with previously reported syntheses in which alkaline deprotection was the last step, was completely avoided.

Sequential Polypeptides. The synthesis of polypeptides with defined repeating sequences continues to be a very active area, although the volume of literature is deceptive, since some workers in the field are given to fragmented and repetitive publication. It is also still true that many papers dealing with sequential polymers fall short of the very reasonable minimum standards of characterisation recommended [144] by a leading authority on the subject in 1967.

Sequential polypeptides are especially valid models for fibrous proteins such as collagen, which contains a high proportion of imino-acids and has glycine at every third position throughout most of its length. A glance at the Appendix will show the reader that many collagen models have been synthesised during 1969 (many had been reported before, however), and these materials have been used in diverse ways to throw light on the chemical and biological problems posed by collagen: *e.g.* as substrates for collagenase [145] and protocollagen hydroxylase;[146] as model antigens in studies of the immunology of collagen;[147] for conformational studies in solution;[148] and for structural investigations in the solid state.[149–151]

Some workers still make use of the tetraethyl pyrophosphite method (see, *e.g.*, ref. 152) of polymerisation (Scheme 23), despite the fact that this

$$\text{H-AA}_1\text{-AA}_2\text{...AA}_x\text{-OH} \longrightarrow \text{H-(AA}_1\text{-AA}_2\text{...AA}_x)_n\text{-OH}$$

Conditions: Tetraethyl pyrophosphite. AA = an amino-acid residue

Scheme 23

route has on several occasions [153] given phosphorus-containing polymers of low molecular weight. In the example just cited, free peptides containing unprotected serine were polymerised by means of tetraethyl pyrophosphite:[152] it seems unlikely that a free hydroxy-group would be completely immune to chemical modification under these conditions, but the authors did not examine their products in any detail, and in particular did not analyse for the possible presence of covalently bound phosphorus. The risk of racemisation concomitant with polymerisation is also very

[144] D. F. DeTar, in 'Peptides', ed. H. C. Beyerman, A. van de Linde, and W. Maassen van den Brink, North-Holland Publishing Co., Amsterdam, 1967, p. 125.
[145] E. Adams, S. Antione, and A. Goldstein, *Biochim. Biophys. Acta*, 1969, **185**, 251.
[146] K. I. Kivirikko and D. J. Prockop, *J. Biol. Chem.*, 1969, **244**, 2755, and references cited therein.
[147] F. Borek, J. Kurtz, and M. Sela, *Biochim. Biophys. Acta*, 1969, **188**, 314.
[148] F. R. Brown tert., J. P. Carver, and E. R. Blout, *J. Mol. Biol.*, 1969, **39**, 307.
[149] D. M. Segal, W. Traub, and A. Yonath, *J. Mol. Biol.*, 1969, **43**, 519.
[150] A. Yonath and W. Traub, *J. Mol. Biol.*, 1969, **43**, 461.
[151] W. Traub, *J. Mol. Biol.*, 1969, **43**, 479.
[152] E. Heidemann and H. W. Nill, *Z. Naturforsch.*, 1969, **24b**, 843.
[153] *e.g.* J. Engel, J. Kurtz, E. Katchalski, and A. Berger, *J. Mol. Biol.*, 1966, **17**, 255.

great with this method if the *C*-terminal amino-acid of the monomer is not glycine or an imino-acid, since model racemisation tests [154] on the use of tetraethyl pyrophosphite for coupling indicated gross racemisation under most conditions. The attraction of the method is obvious—the monomeric free peptides are available by application of the conventional methods of peptide synthesis—but the uncertainties about the chemical and optical purity of the polymers which result from the use of tetraethyl pyrophosphite and similar reagents would appear to render convenience a secondary consideration.

The general consensus of opinion, on the other hand, now seems to be that the best general route is *via* peptide active esters (Scheme 24). The

$$\underset{(45)}{HX,\ AA_1\text{-}AA_2\ldots AA_x\text{-}OY} \longrightarrow H\text{-}(AA_1\text{-}AA_2\ldots AA_x)_n\text{-}OH$$

Conditions: Et_3N, in a polar solvent. AA = an amino-acid residue; X = an anion of a strong acid; Y = an active ester group

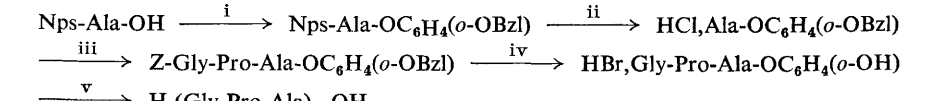
Scheme 24

synthesis of the peptide active ester monomers (45) for this kind of polymerisation often necessitates rather tricky synthetic manœuvres (*e.g.* 'backing-off' with amino-acid active esters as amino-components), but these difficulties can usually be overcome, and the additional effort is well worth while, since in general no side-reactions intervene during the polymerisation, and adequately high molecular weights (average molecular weights usually in the range $5—20 \times 10^3$) are obtained if due care is taken to control the conditions. A number of papers dealing with matters of practical detail in the use of 'backing-off' procedures have appeared recently.[155–159]

Providing that the usual restrictions of permissible synthetic strategy are observed in order to preclude racemisation in the preparation of the peptide active ester monomers (45), sequential polypeptides can be prepared in this way with very little loss of optical purity. Even very small amounts of racemisation, however, can have far-reaching effects on the behaviour of the polypeptides. It is necessary, therefore, to search for new types of active ester which can be used with no risk whatever of losing optical integrity during polymerisation. *o*-Hydroxyphenyl esters are of this type, since they owe their reactivity to intramolecular base catalysis which accelerates

[154] M. W. Williams and G. T. Young, *J. Chem. Soc.*, 1963, 881.
[155] A. Kapoor, L. W. Gerencser, and W. R. Koutnik, *J. Pharm. Sci.*, 1969, **58**, 281.
[156] N. A. Poddubnaya, L. V. Bazaitova, and V. P. Klyagina, *Zhur. obshchei Khim.*, 1969, **39**, 921.
[157] K. T. Poroshin, A. B. Zegel'man, and T. Yusupov, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1969, **12**(1), 20.
[158] A. Kapoor and L. W. Gerencser, *J. Pharm. Sci.*, 1969, **58**, 976.
[159] T. Yusupov, A. V. Zegel'man, and K. T. Poroshin, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1969, **12**(4), 27.

aminolysis without concomitant facilitation of oxazolone formation (see Volume 1 of these Reports). Furthermore, *o*-hydroxyphenyl esters can be introduced in a protected and unreactive form (*i.e.* as an *o*-benzyloxyphenyl ester) which makes the synthesis of the monomers convenient. Poly-(glycylprolylalanine) (of molecular weight $1{\cdot}2 \pm 0{\cdot}1 \times 10^4$ and $[\alpha]_{546}^{20} - 426°$, in water) has been prepared using the *o*-hydroxyphenyl ester route as shown in Scheme 25:[160] the optical rotatory properties of the



Conditions: i, HOC_6H_4(*o*-OBzl), carbonic mixed anhydride method; ii, HCl–Et_2O; iii, Z-Gly-Pro-OH, pivalic mixed anhydride method; iv, HBr–AcOH; v, Et_3N–Me_2SO

Scheme 25

resulting polymer were practically identical to those of poly-(prolylalanylglycine)[161] (which has the same sequence as poly-glycylprolylalanine except at the termini, but which must have been optically pure since the *C*-terminal residue of the monomer was glycine) of similar molecular weight. This comparison serves as a very sensitive racemisation test, since (*a*) the specific rotation of the polymer is large and (*b*) the conformational contribution to the rotation must be critically dependent on complete optical homogeneity, since an earlier preparation (molecular weight $1{\cdot}40 \pm 0{\cdot}05 \times 10^4$) of poly-(glycylprolylalanine) by the *p*-nitrophenyl ester route[162] gave material of very much lower specific rotation ($[\alpha]_{546}^{25} - 208°$, in water), despite the fact that the *p*-nitrophenyl ester product had been shown by enzymic means to contain few, if any, racemised residues.

In the active ester method of sequential polypeptide synthesis, high concentrations of the monomer in an aprotic polar solvent are usually employed in order to ensure profitable competition between polymerisation and cyclisation. The disadvantage of this procedure is that in the majority of cases, polymer quickly separates from the solution, giving lower molecular weights than might otherwise have been obtained. It has recently been reported that dilute solutions can be used with great advantage, providing that a small amount of a partially blocked monomer is added to reduce the amount of cyclisation.[163, 164] In the first applications of this technique [poly-(glutamylalanylglycine)[163] and poly-(lysylalanylglycine)[164]] the molecular weights obtained were of the order of 10^4 (*i.e.* no better than those generally produced by conventional active ester polymerisations),

[160] J. H. Jones, *Chem. Comm.*, 1969, 1436.
[161] F. R. Brown tert., G. P. Lorenzi, and E. R. Blout, to be published.
[162] S. M. Bloom, S. K. Dasgupta, R. P. Pate, and E. R. Blout, *J. Amer. Chem. Soc.*, 1966, **88**, 2035.
[163] B. J. Johnson, *J. Chem. Soc.* (*C*), 1967, 2638.
[164] B. J. Johnson, *J. Chem. Soc.* (*C*), 1968, 3008.

but the approach has been further exemplified by syntheses, in good yield, of poly-(glutamylglutamylalanylglycyl)-glycine ethyl ester (molecular weight reported: *ca.* $2{\cdot}5 \times 10^5$),[165] and of poly-(tyrosylglutamylalanylglycyl)-glycine ethyl ester (molecular weight reported: *ca.* 10^5).[166] If it can be substantiated that the formation of such high molecular weight polymers is general, this method is clearly destined to achieve great popularity. However, a fuller appraisal of the procedure must await a more thorough examination of the molecular weights of the polymers, since considerable difficulties and large errors are often associated with molecular weight determinations in this area.

3 Syntheses Achieved

As in the first volume of these Reports, the ensuing discussion is highly selective, being confined to a few cases of special interest. A comprehensive coverage would be a monumental undertaking and would also be marred by much repetition, since most syntheses employ established methods and strategy. In partial compensation for this selectivity, the list in Appendix A has been made more detailed and as complete as possible.

Many of the materials listed in Appendix A were synthesised for pharmacological study: in this connection a new edition of a standard work[167] on hormone bioassay will be of great value. The proceedings of two symposia[168, 169] which covered aspects of hormonal polypeptide pharmacology (*inter alia*) have been published. Reviews on various aspects of the following peptide hormones have also appeared: calcitonin (thyrocalcitonin),[170, 171] parathyroid hormone,[171] and insulin.[68, 171, 172]

A. Cytochrome *c*.—Cytochrome *c* (Figure 1) can be reconstituted from its apoprotein by treatment with protoporphyrinogen IX followed by iron insertion,[173] and its total synthesis is therefore conceivable, at least in principle. An attempted manual solid-phase synthesis[5] of the apoprotein proved abortive, and could not be pursued far beyond the tryptophan residue in position 59 because of destruction of the indole ring under the acidic α-amino-group deprotection conditions used. The original objective was therefore put aside and a sequence identical to that shown in Figure 1

[165] B. J. Johnson, *J. Chem. Soc.* (*C*), 1969, 1412.
[166] B. J. Johnson and E. J. Trask, *J. Chem. Soc.* (*C*), 1969, 2644.
[167] 'Methods in Hormone Research', ed. R. I. Dorfman, vol. 2, part A, 2nd edn., Academic Press, New York, 1969.
[168] 'Advances in Experimental Medicine and Biology', ed. N. Back, L. Martini, and R. Paoletti, vol. 2, 'Pharmacology of Hormonal Polypeptides and Proteins', Plenum Press, New York, 1968.
[169] 'Prostaglandins, Peptides and Amines', ed. P. Mantegazza and E. W. Horton, Academic Press, New York, 1969.
[170] P. L. Munson, P. F. Hirsch, H. B. Brewer, R. A. Reisfeld, C. W. Cooper, A. B. Wästhed, H. Orimo, and J. T. Potts jun., *Rec. Progr. Hormone Res.*, 1968, **24**, 589.
[171] O. K. Behrens and E. L. Grinnan, *Ann. Rev. Biochem.*, 1969, **38**, 83.
[172] D. F. Steiner, J. L. Clark, C. Nolan, A. H. Rubenstein, E. Margoliash, B. Aten, and P. E. Oyer, *Rec. Progr. Hormone Res.*, 1969, **25**, 207.
[173] S. Sano and K. Tanaka, *J. Biol. Chem.*, 1964, **239**, PC3109.

Ac-Gly-Asp-Val-Glu-Lys-Gly-Lys-Lys-Ile-Phe-Val-Gln-Lys-Cys-Ala-Gln-Cys-
1 **10**
His-Thr-Val-Glu-Lys-Gly-Gly-Lys-His-Lys-Thr-Gly-Pro-Asn-Leu-His-Gly-Leu-
20 **30**
Phe-Gly-Arg-Lys-Thr-Gly-Gln-Ala-Pro-Gly-Phe-Thr-Tyr-Thr-Asp-Ala-Asn-Lys-
40 **50**
Asn-Lys-Gly-Ile-Thr-Trp-Lys-Glu-Glu-Thr-Leu-Met-Glu-Tyr-Leu-Glu-Asn-Pro-
60 **70**
Lys-Lys-Tyr-Ile-Pro-Gly-Thr-Lys-Met-Ile-Phe-Ala-Gly-Ile-Lys-Lys-Lys-Thr-
80
Glu-Arg-Glu-Asp-Leu-Ile-Ala-Tyr-Leu-Lys-Lys-Ala-Thr-Asn-Glu
90 **100** **104**

Figure 1 *Horse heart cytochrome* c

was assembled on the resin, except that the tryptophan residue was replaced by phenylalanine. Since the benzyl group was selected for imidazole (except for His[18], which was apparently introduced in the unprotected form) the nominal product of hydrogen fluoride cleavage from the polymeric support was [His(Bzl)[26], His(Bzl)[33], Phe[59]]-horse heart cytochrome *c* apoprotein. A practical point worth noting is that during the course of this synthesis [5] about 80% of the resin was lost through the sinter because the particle size was progressively reduced by agitation: this problem arose after the first fifty cycles.

The amino-acid analysis of the product disagreed extensively with that expected, in contrast to general experience with solid-phase synthesis of smaller peptides, where the amino-acid analysis often gives figures misleadingly near to the 'correct' values, possibly because residues have been omitted with roughly equal frequency (see Anfinsen's comments [2] on this point). Undaunted by the poor amino-acid analysis, the authors went on to treat their material with protoporphyrinogen IX and eventually obtained, after iron insertion and chromatography, 0·8 mg of 'synthetic cytochrome *c*-like compound'. In the presence of a large amount of free histidine, the specific activity was *ca.* 2% of that of the natural enzyme in the succinic dehydrogenase system: no biological activity was detected in the absence of histidine, which is not required by natural cytochrome *c*. The histidine is presumably fulfilling a haem ligand function which is lacking in the synthetic analogue. The absorption spectrum (λ_{max} 413, 520, 550 nm) of the reduced form of the synthetic material was qualitatively similar to that of natural cytochrome *c*. The spectrum, however, is not compelling evidence, since simple low molecular weight models of the haem region such as the product obtained [174] by treating cysteinyldiglycylcysteine with protoporphyrinogen IX and inserting iron also give spectra of this type.

Many questions are raised by this investigation. Was the low activity due to the well-known deficiencies of the solid-phase method, as implied by

[174] S. Sano, K. Ikeda and S. Sakakibara, *Biochem. Biophys. Res. Comm.*, 1964, **15**, 284.

the disappointing apoprotein amino-acid analysis? Or was it due to the fact that activity might not have been expected in this analogue in any case, because the tryptophan residue has been replaced by phenylalanine and the histidine residues at positions 26 and 33 are benzylated? The complete amino-acid sequences of more than thirty cytochromes *c* from widely separated species (ranging from man through various mammals, birds, fishes, insects to yeast) are known and have been collected together,[175] so it is possible to make a plausible answer to the second question. Out of the hundred or more residues present in all these proteins, only 35 are invariant, and one of these is the tryptophan residue (position 59 in the horse sequence). Considerable variation in the histidine residue which appears at position 33 in the horse sequence is consistent with retention of biological activity, but only one of the cytochromes *c* so far listed [175] does not have histidine at the position corresponding to residue 26 in the horse enzyme. Hence it seems that the synthetic analogue under discussion lacks features which the sequence data imply to be essential. It is therefore obvious that a definite conclusion about the success or failure of the solid-phase synthesis cannot be made on the basis of the low biological activity of the product. This work must be regarded as a necessary preliminary to the total synthesis of material with the natural cytochrome *c* sequence by the solid-phase technique, which, in view of recent progress in the use of tryptophan and histidine in this area, now becomes a feasible proposition.

B. Proinsulin.—The role of proinsulin in the biosynthesis of insulin has been reviewed,[172] and the 'connecting peptide' (Figure 2) of porcine proinsulin has been synthesised.[8] Protected fragments were prepared by stepwise active ester couplings, using mainly *p*-nitrophenyl and succinimido

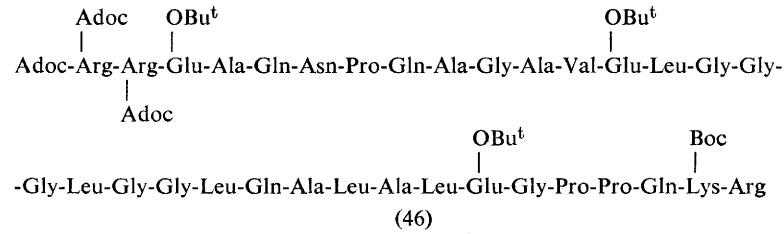
Arg-Arg-Glu-Ala-Gln-Asn-Pro-Gln-Ala-Gly-Ala-Val-Glu-Leu-Gly-Gly-Gly-Leu-
1 **10**
Gly-Gly-Leu-Gln-Ala-Leu-Ala-Leu-Glu-Gly-Pro-Pro-Gln-Lys-Arg
20 **30** **33**

Figure 2 *The porcine proinsulin connecting peptide*

esters with benzyloxycarbonyl α-amino-protection and t-butyl-containing side-chain protection except for arginine, which was blocked by means of a nitro-group in one case early in the synthesis and with bis-adamantyloxycarbonyl substitution in the other two. No practical details of the last method of protecting arginine side-chains are yet forthcoming. Although the molecule contains large amounts of glycine and proline, in the strategy adopted all except one of the six fragment junctions were constructed through sections with *C*-terminal optically active amino-acids. The combined dicyclohexylcarbodi-imide–*N*-hydroxysuccinimide procedure was

[175] 'Atlas of Protein Sequence and Structure', vol. 4, ed. M. O. Dayhoff, National Biomedical Research Foundation, U.S.A., 1969.

used for one '2 + 1' coupling but the other fragment condensation steps were performed with succinimido esters isolated after direct activation of the *C*-terminal residue with dicyclohexylcarbodi-imide. Sometimes the intermediates were inconveniently insoluble in the usual solvents: addition of amino-components dissolved in molten phenol or cresol was then employed. Treatment of the protected final product (46) with trifluoroacetic



(46)

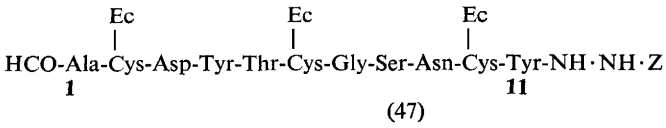
acid gave the connecting peptide. The amino-acid analysis was as expected except in the case of arginine, which gave a somewhat low value: only small amounts of a by-product were detected on electrophoresis.

When natural insulin A and B chains were recombined [176] in the presence of the connecting peptide (C-peptide) the yield of insulin activity was not significantly different from that produced without added C-peptide. Assuming the integrity of the synthetic C-peptide—some examination of its optical purity might be desirable in view of the synthetic strategy used—it seems that covalent bonding between the A, B, and C peptides (or between the C-peptide and either the A or B chain) is necessary to give rise to controlled formation of the disulphide bridges in the correct manner.

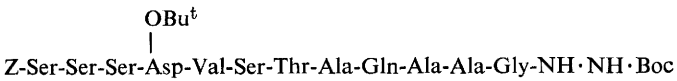
C. Ribonuclease T_1.—Since the syntheses of material with bovine pancreatic ribonuclease A[35] and S'[36] activity were covered in last year's Report, nothing more will be said about these here.

Comparison of the amino-acid content of ribonuclease T_1 with that of bovine pancreatic ribonuclease A (Table 3) suggests that T_1 may be the more hopeful case for unambiguous synthesis: the only adverse feature is the presence of tryptophan in the latter. Furthermore, three of the four half-cysteine residues in T_1 are confined to the *N*-terminal region. Seven protected fragments [(47)—(53)] which together span the entire sequence of T_1 have been synthesised, purified, and characterised.[4] The stage is therefore set for assembly of these intermediates by azide condensations to give the whole enzyme. No doubt these reactions will require much exploratory work on the coupling conditions and purification of the products before pure protected T_1 is obtained: recent experience in the gastrin field with side-reactions in azide condensations of tryptophan-containing

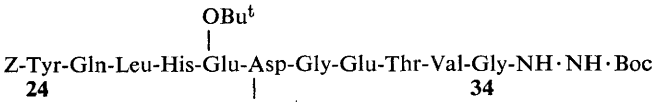
[176] R. Geiger, H. Wissmann, H.-L. Weidenmüller, and H.-G. Schröder, *Z. Naturforsch.*, 1969, **24b**, 1489.



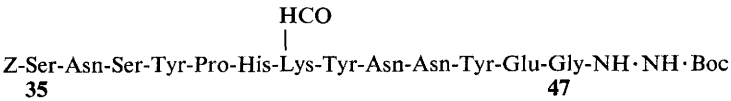
(47)



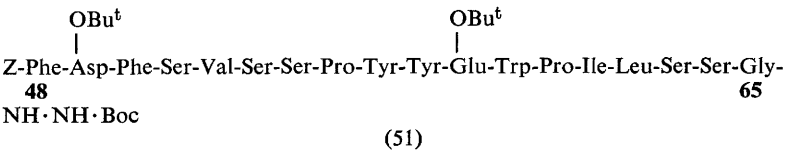
(48)



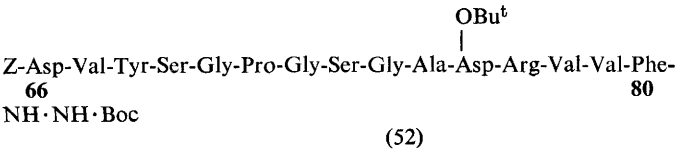
(49)



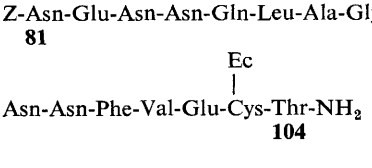
(50)



(51)



(52)



(53)

peptides (see section 2B) must add to the burden here. The only difficulty in the final deprotection which can easily be foreseen is removal of the formyl groups from Ala[1] and Lys[41], but this will probably not be necessary to obtain enzymic activity, since chemical modification of these residues in the native protein does not destroy its catalytic potency. Similarly, the fact that the end-product will be a *C*-terminal amide (native T_1 has a free *C*-terminal carboxy-group) is not expected to be of any great consequence,

Table 3 *The compositions of bovine pancreatic ribonuclease A and ribonuclease T_1*

Amino-acid	*No. of residues in bovine pancreatic ribonuclease A*	*No. of residues in ribonuclease T_1*
Alanine	12	7
Arginine	4	1
Asparagine	10	9
Aspartic acid	5	6
Cystine	4	2
Glutamic acid	5	6
Glutamine	7	3
Glycine	3	12
Histidine	4	3
Isoleucine	3	2
Leucine	2	3
Lysine	10	1
Methionine	4	0
Phenylalanine	3	4
Proline	4	4
Serine	15	15
Threonine	10	6
Tryptophan	0	1
Tyrosine	6	9
Valine	9	8
	Total 124	Total 104

since Thr^{104} can be removed by carboxypeptidase with essentially complete retention of activity

D. Staphylococcal Nuclease.—For an introduction to the illuminating results of concentrated and varied studies (including *X*-ray analysis, chemical modification work, synthetic approaches, conformational investigations, examination of the mechanism, *etc.*) on this protein, the reader is directed to two review lectures [1, 2] and literature cited in ref. 3.

Staphylococcal nuclease (Figure 3) yields three inactive fragments on limited tryptic digestion in the presence of deoxythymidine-3′,5′-diphosphate and calcium ion, which protect the active site and stabilise the conformation. The inactive fragments are designated P_1 (residues 1—5), P_2 (residues 6—48 or 49), and P_3 (residues 49 or 50—149). P_2 and P_3 interact strongly in solution giving a complex ('nuclease-T') which has about 10% of the activity of the intact nuclease. These properties of the nuclease-T system make it especially suitable for studies of structure–activity relationships by synthetic means: both fragments are of a size which places them within reasonable reach of the solid-phase method, and in the case of P_2 (43 residues) even an unambiguous classical synthesis, which is in progress,[1, 2, 3] can be contemplated with equanimity (compare ACTH: 39 residues). Furthermore, the strong interaction between P_2 and P_3 makes 'functional purification' a feasible method for purifying the synthetic products. It seems

Ala-Thr-Ser-Thr-Lys-Lys-Leu-His-Lys-Glu-Pro-Ala-Thr-Leu-Ile-Lys-Ala-Ile-
1 **10**
Asp-Gly-Asp-Thr-Val-Lys-Leu-Met-Tyr-Lys-Gly-Gln-Pro-Met-Thr-Phe-Arg-
20 **30**
Leu-Leu-Leu-Val-Asp-Thr-Pro-Glu-Thr-Lys-His-Pro-Lys-Lys-Gly-Val-Glu-Lys-
40 **50**
Tyr-Gly-Pro-Glu-Ala-Ser-Ala-Phe-Thr-Lys-Lys-Met-Val-Glu-Asn-Ala-Lys-Lys-
60 **70**
Ile-Glu-Val-Glu-Phe-Asn-Lys-Gly-Gln-Arg-Thr-Asp-Lys-Tyr-Gly-Arg-Gly-Leu-
80
Ala-Tyr-Ile-Tyr-Ala-Asp-Gly-Lys-Met-Val-Asn-Glu-Ala-Leu-Val-Arg-Gln-Gly-
90 **100**
Leu-Ala-Lys-Val-Ala-Tyr-Val-Tyr-Lys-Pro-Asn-Asn-Thr-His-Glu-Gln-Leu-Leu-
110 **120**
Arg-Lys-Ser-Glu-Ala-Gln-Ala-Lys-Lys-Glu-Lys-Leu-Asn-Ile-Trp-Ser-Glu-Asn-
130 **140**
Asp-Ala-Asp-Ser-Gly-Gln
149

Figure 3 *The amino-acid sequence of staphylococcal nuclease*

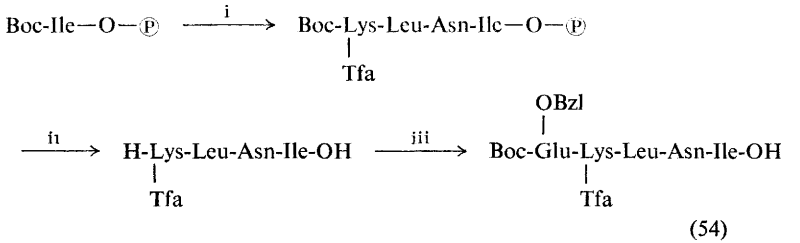
likely that nuclease-T will prove to be a rich source of information on the effect of synthetically controlled amino-acid substitution on an enzyme system. Analogues of P_2 will permit variations in 42 out of the 144 residues of the catalytically active complex. In the only other similar situation which has been investigated in this way (ribonuclease S-protein : S-peptide—see refs. 177—180 and literature cited therein) changes can be rung in only 20 out of 124 residues at present.

The fragment P_2 and several analogues of it have been synthesised by the solid-phase method,[1, 2, 181, 182] but since not all of these publications are available at the time of writing, discussion will be postponed to Volume 3 of these Reports.

It has been indicated[3] that a synthesis of P_3 by fragment condensation has been undertaken, and the preparation of a protected (135—148)-tetradecapeptide from the *C*-terminal section has been described:[3] the use of a rapid procedure (Scheme 26) for making the crystalline partially protected (135—139)-pentapeptide (54), which was obtained analytically pure in an overall yield of 64%, is particularly noteworthy.

An entirely different approach to the preparation of 'unnatural' protien analogues involves use of auxotrophic bacterial strains. If an unnatural

[177] L. Moroder, P. Marchiori, R. Rocchi, A. Fontana, and E. Scoffone, *J. Amer. Chem. Soc.*, 1969, **91**, 3921.
[178] R. Rocchi, F. Marchiori, L. Moroder, G. Borin, and E. Scoffone, *J. Amer. Chem. Soc.*, 1969, **91**, 3927.
[179] R. Rocchi, P. Marchiori, A. Scatturin, L. Moroder, and E. Scoffone, *Gazzetta*, 1968, **98**, 1270.
[180] R. Rocchi, A. Scatturin, L. Moroder, F. Marchiori, A. M. Tamburro, and E. Scoffone, *J. Amer. Chem. Soc.*, 1969, **91**, 492.
[181] D. A. Ontjes and C. B. Anfinsen, *Proc. Nat. Acad. Aci. U.S.A.*, in the press; cited in refs. 3 and 182.
[182] D. A. Ontjes and C. B. Anfinsen, *J. Biol. Chem.*, 1969, **244**, 6316.



(54)

Conditions: i, Standard solid-phase procedure; ii, $HBr–CF_3CO_2H$; iii, Boc-Glu(OBzl)-ONSu and *N*-methylmorpholine in DMF. Ⓟ = Merrifield polymer.

Scheme 26

amino-acid can satisfy the nutritional requirements of the auxotrophic strain completely, then incubation in a medium containing the unnatural amino-acid but lacking the normal requirement leads to incorporation of the unnatural amino-acid into bacterial protein. For example, a tryptophan auxotroph of *E. coli* has recently been used in this way to obtain an active preparation of alkaline phosphatase in which nearly all the seven tryptophan residues were replaced by azatryptophan.[183] In some cases, on the other hand, there are absolute and specific requirements. This means that the natural requirement of an auxotrophic strain cannot be entirely excluded from the incubation medium with the result that if an unnatural amino-acid is added, it is only partially incorporated into bacterial protein, giving inseparable mixtures of very closely related macromolecules. This is the situation with methionine auxotrophs of *S. aureus*, which must have some methionine for central metabolic processes such as transmethylation, but which will reluctantly accept norleucine into their proteins in place of methionine. Thus when a methionine auxotroph of *S. aureus* was grown with limiting methionine in the presence of a large excess of norleucine, only 15% of the resulting nuclease molecules contained exclusively norleucine in place of methionine.[184] However, since cyanogen bromide cleavage is specific for peptide bonds involving methionine, it was possible to degrade selectively all the enzyme molecules containing one or more such residues, and nuclease which contained no methionine was isolated: the unnatural enzyme showed essentially unimpaired catalytic activity.[184] It is a pity that this approach is limited to replacement of amino-acid residues by close analogues. Ideas about obtaining altered proteins by selection of mutants are likely to be subject to similar restrictions, since any mutation which caused a gross change in the protein would have a high chance of proving fatal.

Perhaps some readers will regard the remarks of the previous paragraph as outside the proper syllabus of this chapter. That may be so, but the

[183] S. Schlesinger, *J. Biol. Chem.*, 1968, **243**, 3877.
[184] C. B. Anfinsen and J. G. Corley, *J. Biol. Chem.*, 1969, **244**, 5149.

end-result of the experiments mentioned is the same as one of the main current aims of peptide synthesis. It may be that manipulation of biological systems will eventually displace peptide synthesis as we now know it into a minor role.

4 Appendix A. A List of Syntheses Reported in 1969

A. Natural Acyclic Peptides, Peptide Hormones, Proteins, Analogues, and Partial Sequences.—Analogues and partial sequences are listed under the peptide or protein to which they are related. The syntheses of ribonuclease are included here because they were published in 1969, but discussion of these will be found in Volume 1 of these Reports. No distinction between preliminary communications and full publications has been made: thus some of the entries here refer to papers which give details of syntheses which have been briefly reported before.

Peptide	*Ref.*
Adrenocorticotropins	
[Orn^{8}]-β-corticotropin-(1—24)-tetracosapeptide	185
[Orn^{8,17,18}]-β-corticotropin-(1—24)-tetracosapeptide	185
[Pro^{1}]-β-corticotropin-(1—23)-tricosapeptide amide	186
[β-Ala^{1}]-β-corticotropin-(1—23)-tricosapeptide amide	186
[Leu^{4}]-β-corticotropin-(1—23)-tricosapeptide amide	186
[β-Ala^{1}, Leu^{4}]-β-corticotropin-(1—23)-tricosapeptide amide	186
Human β-corticotropin-(1—28)- and -(1—32)-peptides	187
Angiotensins	
[5-valine,6-β-(pyrazol-3-yl)-alanine]-angiotensin II	188
[Val^{5}]-angiotensin II	125, 189
[Val^{5}]-angiotensin II-(3—8)-*C*-terminal hexapeptide	121
[Gly^{1},Gly^{2},Ile^{5}]-angiotensin II	190
[Ac-Gly^{1},Gly^{2},Ile^{5}]-angiotensin II	190
[Gly^{1},Gly^{2},Ile^{5},His(Bzl)^{6}]-angiotensin II	190
cyclo-(Val-Tyr-Ile-His-Pro-Phe), *i.e.* *cyclo*-([Ile^{5}]-angiotensin II-(3—8)-hexapeptide)	132
Bradykinins	
A large series of partial sequences	191
[Gly^{6}]-bradykinin	192
1-Desamino-bradykinin	193
9-Descarboxy-bradykinin	193
1-Desamino-9-descarboxy-bradykinin	193
Met-Lys-bradykinyl-Ser-Val-Gln (partial sequence of bovine serum kininogen)	194

[185] G. I. Tesser and W. Rittel, *Rec. Trav. chim.*, 1969, **88**, 553.
[186] R. Geiger, H. G. Schröder, and W. Siedel, *Annalen*, 1969, **726**, 177.
[187] L. Kisfaludy and M. Löw, *Acta Chim. Acad. Sci. Hung.*, 1968, **58**, 231.
[188] R. Andreatta and K. Hofmann, *J. Amer Chem. Soc.*, 1968, **90**, 7334.
[189] I. Grava and G. Cipens, *Latv. P.S.R. Zinat. Akad. Vestis, Kim. Ser.*, 1969, 230.
[190] E. C. Jorgensen, G. C. Windridge, W. Patton, and T. C. Lee, *J. Medicin. Chem.*, 1969, **12**, 733.
[191] K. Suzuki, T. Abiko, and N. Endo, *Chem. and Pharm. Bull. (Japan)*, 1969, **17**, 1671.
[192] S. I. Virovets, V. F. Martynov, and M. I. Titov, *Zhur. obshchei Khim.*, 1968, **38**, 2337.
[193] W. D. Johnson, H. D. Law, and R. O. Studer, *Experientia*, 1969, **25**, 573.
[194] E. Schröder, *Experientia*, 1969, **25**, 1126.

Peptide	*Ref.*
Met-Lys-bradykinyl-Ser-Val-Gln-Val (partial sequence of bovine serum kininogen)	194
Calcitonins (thyrocalcitonins)	
The salmon hormone (which is 20—30 times as potent as the other known calcitonins)	7
The porcine hormone	195, 196
Chymotrypsinogen	
Bovine chymotrypsinogen A-(82—88)-protected heptapeptide	75
Cytochrome *c*	
[His(Bzl)^{26,33},Phe^{59}]-horse heart cytochrome *c*	5
Gastrins	
Canine gastrin I	197, 198
Ovine/bovine gastrin I	86
Human gastrin I	198
Tetrazole analogues of the *C*-terminal tetrapeptide amide of gastrin	199
Glutathione	66
Haemoglobin	
Abnormal human haemoglobin α-chain-(65—71)-heptapeptide	200
Human haemoglobin β-chain-(124—146)-tricosapeptide	61
Insulin	
Carba^{A6}-sheep insulin A chain and its combination with natural bovine insulin B chain	201
Horse insulin A chain-(9—14)-protected hexapeptide	202
Sheep insulin A chain and its combination with natural insulin B chain	71
[Glu^{5,15},Ala^{12,18,21},Phe^{19}]-sheep insulin A chain and its combination with natural bovine insulin B chain	203
[Ala^{12}]-sheep insulin A chain and its combination with natural bovine insulin B chain	70
[Glu^{5,15},Ala^{12,14,18,21},Phe^{19}]-sheep insulin A chain and its combination with natural bovine insulin B chain	204
[Glu^{5,15},Ala^{10,12,18,21},Phe^{19}]-sheep insulin A chain and its combination with natural bovine insulin B chain	204
[Glu^{5,15},Ala^{10,12,14,18,21},Phe^{19}]-sheep insulin A chain and its combination with natural bovine insulin B chain	204
[Glu^{5,15},Ala^{12,13,14,18,21},Phe^{19}]-sheep insulin A chain and its combination with natural bovine insulin B chain	204
Insulin-(A20—21; B17—21)-cystine peptide	124

[195] B. Riniker, M. Brugger, B. Kamber, P. Sieber, and W. Rittel, *Helv. Chim. Acta*, 1969, **52**, 1058.
[196] B. Kamber and W. Rittel, *Helv. Chim. Acta*, 1969, **52**, 1074.
[197] K. L. Agarwal, G. W. Kenner, and R. C. Sheppard, *Experientia*, 1969, **25**, 346.
[198] K. L. Agarwal, G. W. Kenner, and R. C. Sheppard, *J. Chem. Soc.* (*C*), 1969, 2213.
[199] J. S. Morley, *J. Chem. Soc.* (*C*), 1969, 809.
[200] B. J. Johnson, *J. Pharm. Sci.*, 1969, **58**, 1042.
[201] K. Jost, J. Rudinger, H. Klostermeyer, and H. Zahn, *Z. Naturforsch.*, 1968, **23b**, 1059.
[202] H. Arold and S. Stibenz, *J. prakt. Chem.*, 1969, **311**, 271.
[203] U. Weber, *Z. physiol. Chem.*, 1968, **349**, 1749.
[204] U. Weber, K.-H. Kerzog, H. Grossmann, S. Hörnle, and G. Weitzel, *Z. physiol. Chem.*, 1969, **350**, 1425.

Peptide	*Ref.*
Human and bovine insulin B chains and their combination with natural bovine insulin A chain	205
Lipotropins	
β-Lipotropin-(44—60)-heptadecapeptide	206
Mellitin	
Mellitin-(1—9)-nonapeptide	207
Mellitin-(25—26)-peptide	208
Melanocyte-stimulating hormone (MSH)	
Human β-MSH	209
Monkey β-MSH	210
Partially protected *C*-terminal pentadecapeptide of monkey and human β-MSH	211
Partial sequences of α-MSH linked through their *C*-terminal carboxy-groups to 5-methoxytryptamine	212
[D-His^{6},D-Phe^{7},D-Arg^{8},D-Trp^{9},Gly^{10}]-α-MSH	213
Oxytocin	107, 214,
[Leu^{9}]-oxytocin	215
Oxytocinoylglycinamide	215
[Ser^{1,6}]-oxytocin	216
[Orn^{4}]-oxytocin	217
[Orn^{5}]-oxytocin	217
[Pro^{4},Ile^{8}]-oxytocin	218
[Leu]^{4}-oxytocin	219
[Ile]^{4}-oxytocin	219
[Leu^{4},Ile^{8}]-oxytocin	218
[Val^{2}]-oxytocin	220
[Leu^{2}]-oxytocin	220
[Nva^{4}]-oxytocin	221
1-Desamino-[Nva^{4}]-oxytocin	221
[Nle^{4}]-oxytocin	221
1-Desamino-[Nle^{4}]-oxytocin	221

[205] T. Okuda and H. Zahn, *Makromol. Chem.*, 1969, **121**, 87.
[206] K. Medzihradszky, S. Bajusz, H. S. Varga, and Z. Lang, *Acta Chim. Acad. Sci. Hung.*, 1969, **59**, 165.
[207] Y. Miura, M. Toyama, and S. Seto, *Sci. Reports Res. Inst., Tôhoku Univ.*, 1968, **20**, *A*, 41.
[208] J. Jentsch, *Z. Naturforsch.*, 1969, **24b**, 597.
[209] H. Yajima, K. Kawasaki, H. Minami, H. Kawatani, N. Mizokami, and Y. Okada, *Biochim. Biophys. Acta*, 1969, **175**, 228.
[210] H. Yajima, Y. Okada, Y. Kinomura, N. Mizokami, and H. Kawatani, *Chem. and Pharm. Bull. (Japan)*, 1969, **17**, 1237.
[211] H. Yajima, H. Kawatani, and N. Mizokami, *Chem. and Pharm. Bull. (Japan)*, 1969, **17**, 1229.
[212] H. Yajima and K. Kawasaki, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 1379.
[213] H. Yajima and K. Kawasaki, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 1387.
[214] F. Vandesande, *Bull. Soc. chim. belges*, 1969, **78**, 395.
[215] H. Aoyagi, M. Kondo, and N. Izimiya, *Bull. Chem. Soc. Japan*, 1968, **41**, 2772.
[216] H. Nesvadba, K. Jost, J. Rudinger, and F. Sorm, *Coll. Czech. Chem. Comm.*, 1968, **33**, 3790.
[217] R. T. Havran, I. L. Schwartz, and R. Walter, *J. Amer. Chem. Soc.*, 1969, **91**, 1836.
[218] J. Rudinger, O. V. Kesarev, K. Poduska, B. T. Pickering, R. E. J. Dyball, D. R. Ferguson, and W. R. Ward, *Experientia*, 1969, **25**, 680.
[219] V. J. Hruby, G. Flouret, and V. du Vigneaud, *J. Biol. Chem.*, 1969, **244**, 3890.
[220] V. J. Hruby and V. du Vigneaud, *J. Medicin. Chem.*, 1969, **12**, 731.
[221] G. Flouret and V. du Vigneaud, *J. Medicin. Chem.*, 1969, **12**, 1035.

Peptide	*Ref.*
[Phe^8]-oxytocin	222
1-Desamino-2-*O*-methyltyrosine-oxytocin	223
Analogues of 1-desamino-oxytocin specifically deuteriated in the 1-β-mercaptopropionic acid residue	224
[Asp(NHMe)^4]-oxytocin	225
[Gln^8]-oxytocin	226
1-Desamino-oxytocinoic acid methylamide and dimethylamide	227
Oxytocinoic acid methylamide	228
Parathyroid hormone	
Bovine parathyroid hormone-(18—24)-heptapeptide	41
Physalaemin	
Physalaemin-(7—11)-pentapeptide amide	52
Phyllocaerulin (a peptide closely related to caerulin which has recently been isolated [229] from the skin of the South American amphibian *Phyllomedusa sauvagei*)	230
Proinsulin	
Porcine proinsulin-(31—63)-tricontatripeptide, *i.e.* the 'connecting peptide'	8
Ribonuclease	
Ribonuclease T_1-(24—30)-protected heptapeptide	231
A set of seven protected fragments (1—11, 12—23, 24—34, 35—47, 48—65, 66—80, and 81—104) which span the entire sequence of ribonuclease T_1	4
Bovine pancreatic ribonuclease A	35
Bovine pancreatic ribonuclease A-(117—121)- pentapeptide	232
Ribonuclease S′	36
Ribonuclease *N*-terminal eicosapeptide ('S-peptide')	233
Des-Lys^1-[Orn^{10}]-S-peptide	177
Des-Ser^{16},Thr^{17},Ser^{18},Ala^{19},Ala^{20},-S-peptide	114
Des-Lys^1,Glu^2,-[Orn^{10}]-S-peptide	177
Des-Lys^1,Glu^2,Thr^3-[Orn^{10}]-S-peptide	177
Des-Lys^7-[Orn^{10}]-S-peptide	178
Des-Phe^8-[Orn^{10}]-S-peptide	178
[Gly^8,Orn^{10}]-S-peptide	179
[Orn^{10},Nle^{13}]-S-peptide	180
Secretin	
Secretin-(1—6)-hexapeptide	58

[222] J. W. M. Baxter, M. Manning, and W. H. Sawyer, *Biochemistry*, 1969, **8**, 3592.
[223] E. Kasafirek, K. Eisler, and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1969, **34**, 2848.
[224] A. T. Blomquist, D. H. Rich, V. J. Hruby, L. L. Nangeroni, P. Glose, and V. du Vigneaud, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 688.
[225] R. T. Havran, I. L. Schwartz, and R. Walter, *Mol. Pharmacol.*, 1969, **5**, 82.
[226] J. W. M. Baxter, T. C. Wuu, M. Manning, and W. H. Sawyer, *Experientia*, 1969, **25**, 1127.
[227] H. Takashima, W. Fraefel, and V. du Vigneaud, *J. Amer. Chem. Soc.*, 1969, **91**, 6182.
[228] H. C. Beyerman and H. M. van den Brink-Zimmermannova, *Rec. Trav. chim.*, 1968, **87**, 1196.
[229] A. Anastasi, G. Bertaccini, J. M. Cei, G. de Caro, V. Erspamer, and M. Impicciatore, *Brit. J. Pharmacol.*, 1969, **37**, 198.
[230] L. Bernardi, G. Bosisio, R. De Castiglione, and O. Goffredo, *Experientia*, 1969, **25**, 7.
[231] T. Kato, N. Mitsuyasu, M. Waki, S. Makisumi, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1968, **41**, 2480.
[232] A. Ito, *Ann. Report, Sankyo Res. Lab.*, 1968, **20**, 59.
[233] K. Hofmann, J. P. Visser, and F. M. Finn, *J. Amer. Chem. Soc.*, 1969, **91**, 4883.

Peptide	*Ref.*
[β-Asp[3]]-secretin-(1—6)-hexapeptide	58
Staphylococcal nuclease	
Staphylococcal nuclease-(6—47)-peptide	1, 2, 181, 182
Staphylococcal nuclease-(135—148)-tetradecapeptide	3
Trypsinogen	
Bovine trypsinogen-(1—9)-nonapeptide	234
Vasopressins	
[β-Ala[9]]-lysine-vasopressin	235
9-desamido-lysine-vasopressin	235
[Hyp[7]]-lysine-vasopressin	235
[D-Gln[4]]-lysine-vasopressin	235
[Ile[2]]-lysine-vasopressin	236
1-Desamino-[Ile[2]]-lysine-vasopressin	236
Specifically tritiated lysine-vasopressin	237
'Acetone-lysine-vasopressin'	238
1-Desamino-dicarba-lysine-vasopressin (strictly speaking, this is a cyclopeptide derivative)	239
1-Desamino-dicarba-arginine-vasopressin (this too is a cyclopeptide derivative)	240
Viral proteins	
Sheath protein of phage fd-(42—49)-*C*-terminal protected octapeptide	241
Tobacco mosaic virus protein-(122—124)-partially protected tripeptide	123
Tobacco mosaic virus protein-(120—124)-partially protected pentapeptide	123

B. Homodetic Cyclic Peptides.—

Antamanide	126, 127, 130, 131
Circulin B	
Protected (linear) peptides related to Circulin B	242
Enniatin B analogues	
cyclo-(MeVal-D-Val-MeVal-D-Val-MeVal-D-Val)	128
Evolidine (only 1% yield in the cyclisation step)	133
Ferrichrome	
cyclo-[Gly_3-(δ-nitro-norvalyl)$_3$]: *i.e.* the cyclohexapeptide intermediate in the ferrichrome synthesis	243
Fungisporin, which was proved by synthesis to be a cyclotetrapeptide, not a cyclo-octapeptide	244
Gramicidin S	
[Gly[5,10]]-gramicidin S	121

[234] J. Savrda and E. Bricas, *Bull. Soc. chim. France*, 1969, 883.
[235] A. S. Dutta, N. Anand, and R. C. Srimal, *Indian J. Chem.*, 1969, **1**, 3.
[236] R. T. Havran and V. du Vigneaud, *J. Amer. Chem. Soc.*, 1969, **91**, 3626.
[237] V. Pliska and N. A. Thorn, *Experientia*, 1969, **25**, 749.
[238] R. T. Havran and V. du Vigneaud, *J. Amer. Chem. Soc.*, 1969, **91**, 2696.
[239] S. Sakakibara and S. Hase, *Bull. Chem. Soc. Japan*, 1968, **41**, 2816.
[240] S. Hase, T. Morikawa, and S. Sakakibara, *Experientia*, 1969, **25**, 1239.
[241] H. Klostermeyer, *Chem. Ber.*, 1969, **102**, 3617.
[242] H. Arold and K. Haller, *J. prakt. Chem.*, 1969, **311**, 3; H. Arold, *J. prakt. Chem.*, 1969, **311**, 278; H. Arold, V. Fischer, and H. Feist, *J. prakt. Chem.*, 1969, **311**, 490; H. Arold, H. Feist, and K. Wilding, *J. prakt. Chem.*, 1969, **311**, 511.
[243] W. Keller-Schierlein and B. Maurer, *Helv. Chim. Acta*, 1969, **52**, 603.
[244] R. O. Studer, *Experientia*, 1969, **25**, 899.

Peptide	*Ref.*
[D-Leu$^{4,4'}$]-gramicidin S (the cyclopentapeptide with half the sequence was also characterised)	245
[D-Val$^{4,4'}$]-gramicidin S (the cyclopentapeptide with half the sequence was also characterised)	245
Iturine	
Protected tripeptides related to iturine	90
Polymixins	
Analogues of (linear) fragments of polymixin M	246
Miscellaneous cyclic peptides	
cyclo-$(Sar)_n$, where $n = 2, 3, 4, 5, 6$, or 8	247
cyclo-(Gly_3-D-Leu)	248
Partially *C*-deuteriated *cyclo*-(Gly_2-His-Gly_2-Tyr)	129
cyclo-(Gly_5-Leu)	129
cyclo-(Gly_5-His)	129
Tyr-Phe-Gln-Asn-Asu-Pro-Lys-Gly-NH_2* (*i.e.* 1-desamino-dicarba-lysine-vasopressin)	239
Tyr-Phe-Gln-Asn-Asu-Pro-Arg-Gly-NH_2* (*i.e.* 1-desamino-dicarba-arginine-vasopressin)	240
cyclo-(Val-Tyr-Ile-His-Pro-Phe) {*i.e. cyclo*-([Ile^5]-angiotensin II-(3—8)-hexapeptide)}	132
cyclo-(Val-Ala-Phe-Ala-Gly_2)	126
cyclo-(Leu-Val-Pro_2-Ala-Phe)	126

* Asu = α-aminosuberic acid.

C. Sequential Polypeptides.—Polymerisation of different monomers can lead to essentially identical polymers: thus poly-(Gly-Pro-Gly), poly-(Gly-Gly-Pro), and poly-(Pro-Gly-Gly) have the same sequence except at the termini. Separate entries are made here, however, to indicate differences in the sequence of the monomers which were used.

Peptide	*Ref.*
Collagen models	
poly-(Ala-Hyp-Hyp)	249
poly-(Ala-Pro-Gly)	249, 250, 251
poly-(Ala-Pro-Pro)	249
poly-(Gly-Ala-Hyp)	249
poly-(Gly-Ala-Pro)	249
poly-(Gly-Gly-Pro)	249

[245] H. Aoyagi, T. Kato, M. Waki, O. Abe, R. Okawa, S. Makisumi, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1969, **42**, 782.

[246] E. A. Morozova, E. S. Oksenoit, and E. N. Gorbacheva, *Zhur. obshchei Khim.*, 1968, **38**, 2647; N. V. Fedoseeva, V. V. Shilin, and A. B. Silaev, *ibid.*, 1969, **39**, 85; E. M. Saad Salem, N. V. Fedoseeva, and A. B. Silaev, *ibid.*, 1969, **39**, 89; E. M. Saad Salem, N. V. Fedoseeva, and A. B. Silaev, *ibid.*, 1968, **38**, 2227.

[247] J. Dale and K. Titlestad, *Chem. Comm.*, 1969, 656.

[248] E. Morozova and E. S. Oksenoit, *Vestnik Moskov. Univ.*, 1968, No. 5, 109.

[249] V. A. Shibnev and A. V. Lazareva, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1969, 398.

[250] V. A. Shibnev, A. V. Lisovenko, T. P. Chuvaeva, and K. T. Poroshin, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1968, 2564.

[251] D. M. Segal and W. Traub, *J. Mol. Biol.*, 1969, **43**, 487.

Peptide	*Ref.*
poly-(Gly-Hyp-Glu)	252
poly-(Gly-Hyp-Hyp)	249, 253
poly-(Gly-Hyp-Pro)	249
poly-(Gly-Hyp-Ser)	252
poly-(Gly-Pro-Ala)	249, 250, 160
poly-(Gly-Pro-Gly)	250, 253
poly-(Gly-Pro-Hyp)	249, 253
poly-(Gly-Pro-Leu)	249
poly-(Gly-Pro-Pro)	249
poly-(Gly-Pro-Ser)	152
poly-(Gly-Pro-Tyr)	249
poly-(Gly-Ser-Pro)	152
poly-(Leu-Gly-Pro)	249
poly-(Pro-Ala-Gly)	254, 161
H-(Pro-Gly-Gly)$_n$-OH, n = 1—8	255
poly-(Gly-Ala-Ala-Gly-Pro-Pro)	256
poly-(Gly-Ala-Pro-Gly-Pro-Ala)	256
poly-(Gly-Ala-Pro-Gly-Pro-Pro)	256
poly-(Gly-Pro-Ala-Gly-Pro-Pro)	256
Other sequential polypeptides	
poly-(Met-Met-Ala)	257
poly-(Gly-Ser-Gly)	258
poly-[Gly-Lys(Tos)-Lys(Tos)]	259
poly-[Gly-Gly-Lys(Tos)]	259
poly-[Gly-Gly-Lys(Tos)-Lys(Tos)]	259
poly-(Ala-Gly-Gly)	260
poly-(Ala-Ala-Gly)	260
poly-(Ala-Ala-Ala-Gly)	260
H-(Ala-Gly)$_n$-OH, n = 5, 10, and 20	261
poly-(Glu-Glu-Ala-Gly)-[1-^{14}C]-Gly-OEt	165
poly-(Tyr-Glu-Ala-Gly)-[1-^{14}C]-Gly-OEt	166
poly-(Sar-Gly)	54

D. Miscellaneous Peptides.—

Val-Asp-Glu-Glu-Glu-DL-Ser-Ile-Ala-Met-Glu-Lys	262
Leu-Ala-Ala-Gly-Lys-Val-Glu-Asp-Ser-Asp	263

[252] Sh. Kh. Khalikov, K. T. Poroshin, and V. A. Shibnev, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1968, **11**(9), 28.
[253] V. A. Shibnev, T. P. Chuvaeva, G. A. Martynova, and K. T. Poroshin, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1969, 637.
[254] E. Heidemann and H. W. Bernhardt, *Nature*, 1968, **220**, 1326.
[255] M. Rothe, R. Theysohn, K.-D. Steffen, H. Schneider, M. Zamani, and M. Kostrzewa, *Angew. Chem. Internat. Edn.*, 1969, **8**, 919.
[256] D. M. Segal, *J. Mol. Biol.*, 1969, **43**, 497.
[257] E. Dellacherie and J. Neel, *Bull. Soc. chim. France*, 1969, 1218.
[258] Sh. Kh. Khalikov, K. T. Poroshin, and V. A. Shibnev, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1968, **11**(7), 29.
[259] K. T. Poroshin and V. K. Burichenko, *Doklady Akad. Nauk Tadzh. S.S.R.*, 1968, **11** (11), 28.
[260] S. Takahashi, *Bull. Chem. Soc. Japan*, 1969, **42**, 521.
[261] S. Ebihara and Y. Kishida, *Nippon Kagaku Zasshi*, 1969, **90**, 819.
[262] K. H. Ney and K. P. Polzhofer, *Tetrahedron*, 1968, **24**, 6619.
[263] K. L. Agarwal, G. W. Kenner, and R. C. Sheppard, *J. Chem. Soc.* (*C*), 1969, 2218.

Peptide	*Ref.*
Tripeptides containing two residues of glycine and one of glutamic acid	264
Dipeptides containing 2-cyclopentylglycine and β-2-thienylalanine	265
Arginine-containing peptides	266
Cysteine-containing peptides	267
DL-β-(2-pyridyl)-alanine-containing peptides	268
β-Lysine-containing peptides	269
β-Alanine-containing peptides	270
Serine-containing peptides	271
Sarcolysine [*i.e.* *p*-di-(2-chloroethyl)aminophenylalanine]-containing peptides	272
Asn-Lys-His-His-Arg	273
β-Alanyltyrosine	274
Tetrapeptides containing D-serine	275
Oligomers of glycine	276
Oligomers of lysine	109
Oligomers of proline	255
γ-Glutamyl-β-cyanoalanine	277
Peptide derivatives containing β-cyanoalanine	278
Asp-Glu-Leu-Thr-Lys	279
Potentially antimitotic peptides	280
A large number of dipeptide derivatives prepared for an investigation of their structure–taste relationships	281
A series of tripeptide derivatives containing Glu, Pro, and His; the acetylation products of one of these (Glu-His-Pro) had thyrotropin-releasing activity	282

264 S. Rolski and R. Paruszewski, *Roczniki Chem.*, 1969, **43**, 983.
265 J. T. Hill and F. W. Dunn, *J. Medicin. Chem.*, 1969, **12**, 737.
266 B. L. Krainova, N. A. Uvarova, V. A. Safonova, and E. S. Chaman, *Zhur. obshchei Khim.*, 1969, **39**, 92.
267 T. A. Lubkova, A. F. Mironov, G. A. Vasil'eva, R. P. Evstigeneeva, and N. A. Dreobrazhenskii, *Zhur. obshchei Khim.*, 1968, **38**, 2247.
268 H. Watanabe, S. Kuwata, K. Nace, and Y. Nichida, *Bull. Chem. Soc. Japan*, 1968, **41**, 1634.
269 L. I. Rostovtseva, P. D. Reshetov, and A. S. Khokhlov, *Zhur. obshchei Khim.*, 1969, **39**, 96.
270 C. Pinelli, M. Portelli, and M. Fioretti, *Farmaco, Ed. Sci.*, 1968, **23**, 859.
271 E. Heideman and H. W. Hill, *Z. Naturforsch.*, 1969, **24b**, 837.
272 O. F. Ginzburg, K. Yu. Mar'yanovska, I. A. Poletaeva, and T. F. Stepanova, *Zhur. obshchei Khim.*, 1969, **39**, 1168; A. A. B. Paulyukonis, K. I. Karpavichus, O. V. Kil'disheva, and I. L. Knunyants, *Izvest. Akad. Nauk S.S.S.R. Ser. khim.*, 1969, 133; A. A. B. Paulyukonis, K. I. Karpavichus, and I. L. Knunyants, *ibid.*, p. 1364; A. A. B. Paulyukonis, K. I. Karpavichus, O. V. Kil'disheva, and I. L. Knunyants, *ibid.*, 1968, 2772.
273 V. K. Naithani, K. B. Mathur, and M. M. Dhar, *Indian J. Biochem.*, 1969, **6**, 10.
274 L. Levenbook, R. P. Bodnaryk, and T. F. Spande, *Biochem. J.*, 1969, **113**, 837.
275 H. Kienhuis, J. P. J. van der Holst, and A. Verweij, *Rec. Trav. chim.*, 1969, **88**, 592.
276 Y. Kanaoka, M. Machida, and M. Machida, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 2160.
277 C. Ressler, S. N. Nigam, and Y.-H. Giza, *J. Amer. Chem. Soc.*, 1969, **91**, 2758.
278 Z. Grzonka and B. Liberek, *Zeszyty Nauk., Mat., Fiz., Chem.*, 1969, **9**, 107.
279 D. E. Nitecki and B. Halpern, *Austral. J. Chem.*, 1969, **22**, 871.
280 A. Jean and J. Anatol, *Compt. rend.*, 1969, **268**, *C*, 852.
281 R. H. Mazur, J. M. Schlatter, and A. H. Goldkamp, *J. Amer. Chem. Soc.*, 1969, **91**, 2684.
282 R. Burgus, T. F. Dunn, D. N. Ward, W. Vale, M. Amoss, and R. Guillemin, *Compt. rend.*, 1969, **268**, *D*, 2116.

Peptide	*Ref.*
A series of tripeptides related to the repeating sequence regions of collagen	283
A large series of oligopeptide derivatives containing *inter alia*, Tyr, His, and Trp: these were prepared for a study of mass spectrometric fragmentation processes	284
Carnosine	285
Gly_2-Leu-Gly	122
Tyr-Leu-Gly-Glu-Phe	113
Ser_2-Thr-Ser-Ala_2	85
Ala-Ser-Phe	119
Ala-Thr-Phe	119
A tris-cystine peptide	65

5 Appendix B. A List of Some Useful New Synthetic Intermediates which were described during 1969

This list is a highly subjective and necessarily selective compilation of new amino-acid derivatives which seem likely to be of general usefulness or interest. The compounds, which are new and crystalline unless otherwise stated, are listed under the amino-acid from which they are derived. Only derivatives of the amino-acids which commonly occur in proteins, and of ornithine, are included.

Compound	*Ref.*
Alanine	
Z(OMe)-Ala-ONp	50
Pipoc-Ala	51
Bpoc-Ala	41, 42
Arginine	
Z(OMe)-Arg(NO_2)	50
Z(OMe)-Arg(NO_2)-OPcp	50
Bpoc-Arg(NO_2),etherate	41
Arg(NO_2)-OPic,2HBr	51
Z-Arg($Adoc_2$)-ONp: no details yet reported	8
Arg(Z_2)*N*-carboxyanhydride hydrochloride	141
Asparagine	
Z-Asn-ONSu	239
Z-Asn(Bzh)-OBzl	139
Asn(Bzh)	139
Asn(Bzh)*N*-carboxyanhydride	139
Tos·OH,Asn-ONp	133
Nps-Asn-ONSu	79
Aspartic acid	
Boc-Asp(OBzl): details of an improved preparation	286
Boc-Asp(OBzl),Dcha	286

[283] V. A. Shibnev, A. V. Lazareva, and M. P. Finogenova, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1969, 392.

[284] E. I. Vinogradova, V. M. Lipkin, Yu. B. Alakhov, M. Yu. Feigina, and Yu. B. Shvetzov, *Zhur. obshchei Khim.*, 1968, **38**, 777; E. I. Vinogradova, V. M. Lipkin, Yu. B. Alakhov, and Yu. B. Shvetzov, *ibid.*, p. 787.

[285] S. Yamashita and N. Ishikawa, *Experientia*, 1968, **24**, 1079.

[286] K. P. Polzhofer, *Tetrahedron Letters*, 1969, 2305.

Compound	*Ref.*
Bpoc-Asp(OBzl)	41
Nps-Asp(OBu^t),Dcha	46
Pipoc-Asp,Dcha$_2$	51
Cysteine	
Cys(Dce)	67
Boc-Cys(Dce): oil	66
Z-Cys(Dce): oil	66
Boc-Cys(H)	287
Boc-Cys(Bzh)	287
Boc-Cys(Trt)	287
Boc-Cys[Bzl(OMe)]	287
Boc-Cys(Btm)	287
Boc-Cys(Thp)	287
Boc-Cys(Ac)	287
Boc-Cys(Ec)	287
Boc-Cys(Z)	287
Z(OMe)-Cys(Bzl)	50
Pipoc-Cys(Bzl),Dcha	51
Pipoc-Cys(Bzl)-ONSu	51
Bpoc-Cys[Bzl(OMe)],Dcha	41
Boc-Cys(SEt): crystallisation difficult; Dcha salt also reported	70
Z(OMe)-Cys(Z)	214
Z(OMe)-Cys(Z)-ONp	214
Cystine	
Z(OMe)-Cys: previously reported as an oil | Z(OMe)-Cys	214
Glutamic acid	
Pht-Glu-OBu^t	288
Bpoc-Glu	41
Bpoc-Glu(OBzl),Cha	41
Pipoc-Glu,Dcha$_2$	51
HCl,D-Glu(OMe)-OBu^t	143
Glutamine	
Bpoc-Gln,Dcha	41, 42
Z-Gln(Bzh)-OBzl	139
Gln(Bzh)	139
Gln(Bzh)*N*-carboxyanhydride	139
Nps-Gln-ONSu	79
Glycine	
Pipoc-Gly	51
Bpoc-Gly	41, 42
Bpoc-Gly-ONp	42
Gly-OPic,2HBr	57
Histidine	
Z-His(Tos),Dcha	63
Aoc-His(Tos)	63
Nps-His(Tos)	63
Bpoc-His(Bzl)	41
Boc-His(Dnp)	61

[287] H. Zahn and K. Hammerström, *Chem. Ber.*, 1969, **102**, 1048.
[288] H. Aquila and T. Wieland, *Annalen*, 1969, **721**, 223.

Compound	*Ref.*
Isoleucine	
Z(OMe)-Ile-ONSu	50
Z(OMe)-Ile-ONp	214
Pipoc-Ile,Dcha	51
Bpoc-Ile	41, 42
Leucine	
Pipoc-Leu,Dcha	51
Leu-OPic,2HBr	57
Lysine	
Lys(Pipoc)	51
Z-Lys(Pipoc)	51
Z-Lys(Pipoc)-OTcp	51
Bpoc-Lys(Z): amorphous	41
Boc-Lys(Z)-ONp: physical constants only have been reported before	205
Boc-Lys(Tos): oil; Dcha salt also reported	205
Z-Lys(Pht)-ONp	239
Methionine	
Z(OMe)-Met: free acid crystallised for the first time	50
Z(OMe)-Met-OTcp	50
Pipoc-Met,Dcha	51
Pipoc-Met-ONSu	51
Bpoc-Met(O),Dcha	41
Met-OPic,2HBr	57
Ornithine	
Z(OMe)-Orn(Z),Dcha	49
Tos-Orn(Tos)	289
Z-Orn(Tos)	289
Z-Orn(Tos)-OPcp	289
Phenylalanine	
Z(OMe)-Phe-OPcp	50
Z(OMe)-Phe-ONp	50
Pipoc-Phe	51
Bpoc-Phe-OPcp	42
Phe-OPic,2HBr	57
Aoc-Phe-ONSu	239
Proline	
Z(OMe)-Pro-ONSu	50
Bpoc-Pro	41
Pro-OPic,2HBr	57
Serine	
Z(OMe)-Ser-ODnp	50
Z(OMe)-Ser(Bzl)-OPcp	50
Bpoc-Ser,Cha	42
Bpoc-Ser(Bzl),Cha	41
Nps-Ser(Ac),Dcha	133
Threonine	
Bpoc-Thr(Bu^t),Cha	42
Boc-Thr(Ztf): reported both oily,[119] and crystalline [85]	
Boc-Thr(Ztf),Dcha	119
Bpoc-Thr(Bzl)	41

[289] V. Gut, J. Rudinger, R. Walter, P. A. Herling, and I. L. Schwarz, *Tetrahedron*, 1968, **24**, 6351.

Compound	*Ref.*
Tryptophan	
Pipoc-Trp	51
Pipoc-Trp-OPcp	51
Bpoc-Trp	41
Tyrosine	
Z-Tyr(Bu^t)OPcp	166
Bpoc-Tyr(Bzl)	41
Z(OMe)-Tyr(Bzl)-ONp	214
Aoc-Tyr(Bu^t)-ONSu	239
Valine	
Z(OMe)-Val-OTcp	50
Z(OMe)-Val-ONSu	50
Bpoc-Val	41
Val-OPic,2HBr	57

4

Peptides with Structural Features Not Typical of Proteins

BY J. S. DAVIES

1 Introduction

The choice of topics reviewed in this chapter follows a similar theme to that defined in Chapter 4 of last year's Report, although this year biosynthetic aspects have been covered more fully. Aspects of the chemistry of the penicillins and cephalosporins have been included for the first time, the coverage being restricted to work of particular interest to workers in the field of peptide chemistry; no attempt has been made, for example, to review in detail the preparation of analogues and derivatives of penicillin recorded in the patent literature.

Fewer new compounds than usual have been characterised this year. Interest has focused increasingly on the finer points of the spatial structure of peptide systems. The application of a number of physico-chemical techniques for determining the conformation of the antibiotics valinomycin and enniatin demonstrates how *X*-ray crystallography, as a method for determining the conformation of molecules, can expect strong competition in future, at least in relatively simple systems. Critics of the method of extrapolating *X*-ray crystallographic results to define the conformation of molecules in solution will surely welcome with interest the recent results obtained from n.m.r. studies in association with other techniques.

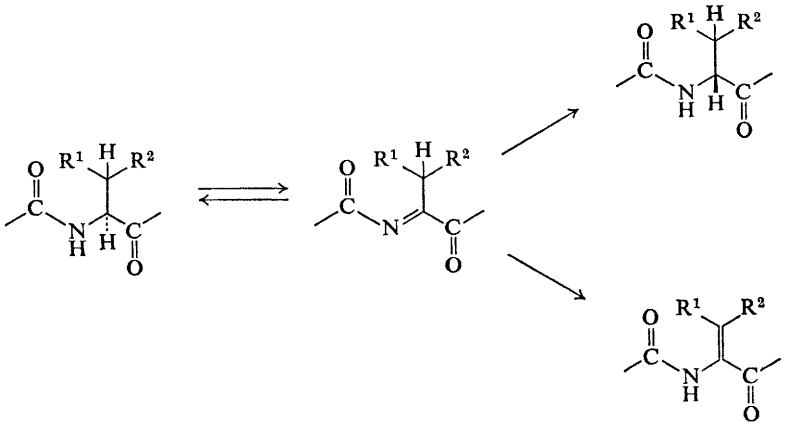
The biosynthesis of microbial peptides and the structures of cell-wall peptidoglycans are other fields of study where certain trends and regular patterns are now emerging as a result of years of research endeavour.

2 Biosynthesis of Peptide Antibiotics and Microbial Peptides

The mechanism of formation of peptide antibiotics and their role in the metabolism of the producing organism continue to attract much attention. The evidence accumulated in two comprehensive reviews [1, 2] of the literature up to 1968 gives strong support to the view that peptide antibiotics are biosynthesised *via* a pathway quite distinct from that of protein biosynthesis. In the absence of a ribosomal template, other enzymatically based syntheses are needed to explain the initiation, assembly, sequencing, and specificity

[1] E. Katz, *Lloydia*, 1968, **31**, 364.
[2] M. Bodanszky and D. Perlman, *Science*, 1969, **163**, 352.

of amino-acids in the synthesis of these antibiotics. Very little is known about the details of these enzyme-based syntheses, but hypotheses have been put forward to explain the origin of certain features in peptide antibiotics. In the examples investigated so far,[1, 2] the D-amino-acids found in microbial peptides appear to be biosynthesised from the L-isomer, and a survey[2] of the occurrence of amino-acids containing two asymmetric centres has shown that while L-isoleucine and D-*allo*-isoleucine were both known in Nature, the other two forms, L-*allo*-isoleucine and D-isoleucine, were not found. It was suggested[2] that L-amino-acids epimerise at the α-carbon atom to give the D-isomer. More recently, however, D-isoleucine has been found[3] in the depsipeptide monamycin, and appears to be the first example contrary to the 'α-epimerisation rule'. In an attempt to explain the occurrence in Nature of both αβ-unsaturated- and D-amino-acids it has been proposed[4] that epimerisation at the α-carbon atom takes place by a sequential dehydrogenation–re-hydrogenation process as shown in Scheme 1, which occurs when the peptide is partially synthesised. This



Scheme 1

explanation, however, only holds if the re-hydrogenation process can be enzymatically controlled to give the correct isomer, and so far this has not been proved. Another theory[5] proposes that amino-acids react to form dioxopiperazines prior to undergoing insertion reactions and ring enlargement.[6] In this way, it is envisaged that epimerisation of the amino-acids to form the D-isomers would take place at the dioxopiperazine stage, a process well known to take place *in vitro*.

[3] K. Bevan, J. S. Davies, C. H. Hassall, and D. A. S. Phillips, *Chem. Comm.*, 1969, 1246.
[4] B. W. Bycroft, *Nature*, 1969, **224**, 595.
[5] A. B. Mauger, *Experientia*, 1968, **24**, 1068.
[6] M. M. Shemyakin and V. K. Antonov, *Pure Appl. Chem.*, 1964, **9**, 75.

Studies on the biosynthesis of individual antibiotics are discussed below in the appropriate sections, but a great deal of work still needs to be done on the origins of microbial peptides.

3 Cyclic Peptides

Cyclic peptides continue to be the subject of numerous investigations on the application of physical methods to the determination of conformation and structure. In addition to the work discussed below for specific compounds, n.m.r. spectroscopy has yielded useful information about the conformation of a number of model cyclic peptides. The shape of the peptide backbone in a *cyclo*-pentaglycyltyrosine derivative has been elucidated[7] from the 100 MHz n.m.r. spectra of the cyclic peptide and that of a partially deuteriated derivative measured in [$^2H_6$]dimethyl sulphoxide, trifluoroacetic

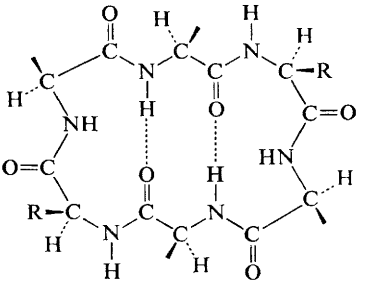
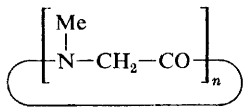


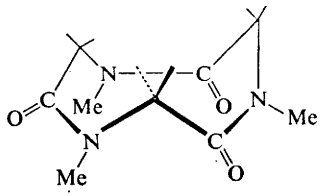
Figure 1 *Conformation of model cyclic hexapeptides*

acid, and water. The fact that two protons are shielded from the solvent and that four are exposed has been interpreted as being consistent with the arrangement shown in Figure 1. Spin decoupling experiments indicate that in dimethyl sulphoxide–water solutions the α-proton of the tyrosyl residue is vicinal to an amide proton at a relatively low field value, and thus they suggest that the tyrosyl side-chain must be attached to a corner position. Studies[8] on *cyclo*-pentaglycyl-L-leucine and *cyclo*-diglycyl-L-histidyl-diglycyl-L-tyrosine also show support for the conformation in Figure 1, the most probable position for the tyrosyl residue in the latter peptide being one of the two positions designated by R in the figure. Six cyclic *N*-methylglycine oligopeptides (1) have been synthesised and their n.m.r. spectra investigated.[9] The dioxopiperazine (1; $n = 2$) has a planar conformation while the cyclic trimer is represented by structure (2). A configuration having four amide bonds in the sequence *cis*, *trans*, *cis*, *trans*, appears to be favoured for the cyclic tetramer, but no clear picture emerges for the higher members of the series. O.r.d. studies on linear and cyclic hexapeptides related to albo-

[7] K. D. Kopple, M. Ohnishi, and A. Go, *J. Amer. Chem. Soc.*, 1969, **91**, 4264.
[8] K. D. Kopple, M. Ohnishi, and A. Go, *Biochemistry*, 1969, **8**, 4087.
[9] J. Dale and K. Titlestad, *Chem. Comm.*, 1969, 656.



(1; $n = 2, 3, 4, 5, 6, 8$)



(2)

mycin [10] and diastereoisomeric cyclohexapeptides [11] containing glycine, leucine, and phenylalanine have been reported.

Recent developments in the application of theoretical calculations to determine the conformation of cyclic peptides have been reviewed.[12] In theoretical studies on the energetically favourable conformers of cyclopentapeptides,[13] stereochemical criteria and energy considerations show that mixed L- and D-residues are not as energetically favourable as all L- or all D-forms in *cyclo*-penta-alanine peptides. Similar calculations have also been made for cyclohexapeptides.[14] Computation [15] of the possible conformations of the cyclic disulphide of cysteinylcysteine shows that the disulphide bridge is only possible if the amide group has a *cis*-configuration. The synthesis of homodetic cyclic peptides is discussed in Chapter 3, section 2D.

A. 2,5-Dioxopiperazines.—A 100 MHz n.m.r. study [16] on a series of dioxopiperazines shows that the preferred conformations of the side-chains of histidine and tryptophan in *cyclo*-(Gly-His) and *cyclo*-(Gly-Trp) have the arylmethyl groups facing the dioxopiperazine ring. Detailed analysis of spin–spin coupling of α-protons and amide protons suggests that in some cases the dioxopiperazine ring is non-planar in dimethyl sulphoxide but reverts to a planar conformation in trifluoroacetic acid. A twist-boat conformation (Figure 2) is in agreement with the coupling constants obtained. The flexibility of the supposedly planar rigid dioxopiperazine ring is further supported by the *X*-ray analysis [17] of *cyclo*-(L-Ala-L-Ala) and *cyclo*-(L-Ala-D-Ala). While the latter *trans*-isomer is completely planar, the ring in the *cis*-isomer deviates from planarity, giving an angle of 26° between the amide group planes. The crystal structure of *NN'*-dimethyldioxopiperazine shows [17a] that the introduction of *N*-methyl groups causes

[10] N. A. Poddubnaya, V. M. Potapov, A. M. El'Naggar, V. M. Demyanovich, and L. G. Makevnina, *Zhur. obshchei Khim.*, 1968, **38**, 2215.
[11] K. Bláha, I. Frič, and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1969, **34**, 3497.
[12] C. M. Venkatchalam and G. N. Ramachandran, *Ann. Rev. Biochem.*, 1969, **38**, 45.
[13] C. Ramakrishnan and K. P. Sarathy, *Internat. J. Protein Res.*, 1969, **1**, 63.
[14] C. Ramakrishnan and K. P. Sarathy, *Internat. J. Protein Res.*, 1969, **1**, 103.
[15] R. Chandrasekaran and R. Balasubramanian, *Biochim. Biophys. Acta*, 1969, **188**, 1.
[16] K. D. Kopple and M. Ohnishi, *J. Amer. Chem. Soc.*, 1969, **91**, 962.
[17] C. Benedetti, P. Corradini, M. Goodman, and C. Pedone, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 650; C. Benedetti, P. Corradini, and C. Pedone, *J. Phys. Chem.*, 1969, **73**, 2891; *Biopolymers*, 1969, **7**, 751.
[17a] P. Groth, *Acta Chem. Scand.*, 1969, **23**, 3155.

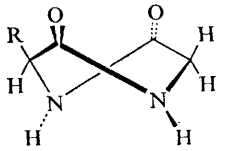


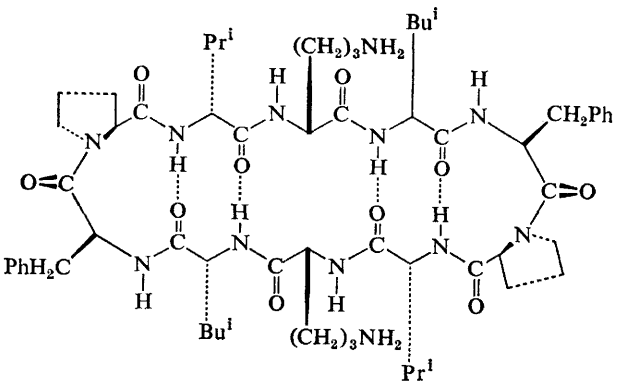
Figure 2 *Twist-boat conformation of dioxopiperazine ring*

the ring system to deviate from planarity to what can be described as a flattened chair form.

The extent to which the presence of urea affects the o.r.d. curves of cyclic dipeptides containing different side-chains has been studied [18] using the dioxopiperazines of serine, alanine, lysine, and valine. The effect of the solvent medium on the o.r.d. results is dependent on the hydrophobic nature of the side-chains. A detailed mass-spectrometric study [19] of a number of desulphurised dioxopiperazines derived from aranotin has revealed a consistent fragmentation pattern within the series. *cyclo*-(L-Ala-L-Leu) has been identified [20] as a metabolite of *Aspergillus niger*.

The rates of hydrolysis of the LL-(*cis*)- and LD-(*trans*)-forms of the dioxopiperazines of leucine and alanine differ significantly.[21] In 0·5M HCl, *cyclo*-(L-Ala-L-Ala) was hydrolysed 1·6 times faster than *cyclo*-(L-Leu-L-Leu), while *cyclo*-(L-Ala-D-Ala) was hydrolysed 3·2 times faster than the leucine analogue. The differences in the steric interactions of the side-chain in the different compounds offers a reasonable explanation for these results.

B. Gramicidins.—Similar conclusions about the shape of the cyclodecapeptide ring system (3) in gramicidin S have been obtained from a



(3)

[18] M. D'Alagni and B. Pispisa, *J. Biol. Chem.*, 1969, **244**, 5843.
[19] R. Nagarajan, J. L. Occolowitz, N. Neuss, and S. M. Nash, *Chem. Comm.*, 1969, 359.
[20] F. Caeser, K. Jansson, and E. Mutschler, *Pharm. Acta Helv.*, 1969, **44**, 676.
[21] O. Grahl-Nielsen, *Tetrahedron Letters*, 1969, 2827.

number of independent investigations [22] using n.m.r. techniques. The overall symmetry of the molecule is consistent with the previously reported *X*-ray data. The structure (3) is based on proton assignments from spectra obtained in different solvents, double-resonance experiments, and variable-temperature techniques. A reassignment [23] of the amide proton shifts previously published for gramicidin S has been made on the strength of the results obtained from the n.m.r. spectra recorded at 220 MHz.

Tritium–hydrogen exchange using a rapid dialysis technique [24] has confirmed the presence in gramicidin S of four rapidly exchanging amide hydrogens and two groups of slowly exchanging hydrogens, as required by structure (3). The structure computed [25] for gramicidin S from energy minimisation considerations is again similar to (3). Current theories relating to circular dichroism have also been tested,[26] using gramicidin S as a model compound.

The search for the necessary topochemical requirements for biological activity in the gramicidins continues. Analogues of gramicidin S containing D-valine or D-leucine in place of D-phenylalanine have been synthesised [27] and both show the same biological activity as natural gramicidin S, proving that the phenylalanyl aromatic side-chains are not required for activity. A review of the topochemical investigations carried out up to 1968 on the gramicidins has been published.[28] Details of the syntheses of [Gly^{5,10}]-gramicidin S on a polymeric carrier have also been reported.[29]

Evidence for the nucleic acid-independent biosynthesis of peptide antibiotics has already been referred to in an earlier section of this chapter. Useful information about the role of enzymes in the biosynthesis has come from the studies [30] carried out on two complementary polyenzyme systems which are capable of forming bound intermediates with the constituent L-amino-acids of gramicidin S in the presence of ATP. The intermediates are envisaged as amino-acyl adenylate intermediates, which are then transferred to another function in the enzyme to induce synthesis of gramicidin S. The biosynthesis of gramicidin S from enzyme-bound thiol ester intermediates has also been investigated.[31] An ornithine-

[22] A. Stern, W. A. Gibbons, and L. C. Craig, *Proc. Nat. Acad. Sci., U.S.A.*, 1968, **61**, 734; R. Schwyzer and U. Ludescher, *Helv. Chim. Acta*, 1969, **52**, 2033; M. Ohnishi and D. W. Urry, *Biochem. Biophys. Res. Comm.*, 1969, **36**, 194.

[23] F. Conti, *Nature*, 1969, **221**, 777.

[24] S. L. Laiken, M. P. Printz, and L. C. Craig, *Biochemistry*, 1969, **8**, 519.

[25] F. A. Momany, G. Vanderkooi, R. W. Tuttle, and H. A. Scheraga, *Biochemistry*, 1969, **8**, 744.

[26] S. L. Laiken, M. P. Printz, and L. C. Craig, *J. Biol. Chem.*, 1969, **244**, 4454.

[27] H. Aoyagi, T. Kato, M. Waki, O. Abe, R. Okawa, S. Makisumi, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1969, **42**, 782.

[28] M. M. Shemyakin, Yu. A. Ovchinnikov, and V. T. Ivanov, *Angew. Chem. Internat. Edn.*, 1969, **8**, 492.

[29] Yu. A. Ovchinnikov, A. A. Kiryushkin, and I. V. Kozhevnikova, *Zhur. obshchei Khim.*, 1968, **38**, 2631, 2636.

[30] H. Kleinkauf, W. Gevers, and F. Lipmann, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 226.

[31] W. Gevers, H. Kleinkauf, and F. Lipmann, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **63**, 1335.

activating enzyme involved in the biosynthesis of gramicidin S has been isolated.[32]

C. Circulin B.—The synthesis of a peptide derivative containing the partial sequence of circulin B has been reported.[33] Peptide derivatives containing a modified partial sequence of circulin B, where threonine is replaced by *allo*-threonine, and $\alpha\gamma$-diaminobutyric acid (Dbu) by lysine or ornithine, have also been synthesised.[34]

D. Polymyxins.—Problems of protection of the side-chain of $\alpha\gamma$-diaminobutyric acid in the synthesis of peptide derivatives used in the formation of polymyxin analogues have been overcome.[35] Selectively removable formyl and benzyloxycarbonyl groups for amino-group protection were found to be useful in these syntheses. Peptide (4) and peptide fragments of a lysine

Pelargonyl-L-Lys-L-Thr-L-Lys-L-Lys-NH·NH_2

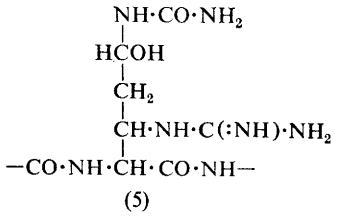
L-Lys-D-Leu-L-Lys-L-Thr-┘

(4)

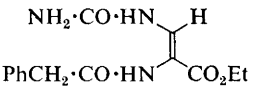
analogue of polymyxin M have also been synthesised.[36] A preliminary note[37] on the isolation of a new antibiotic complex, polymyxin P, from *Bacillus polymyxa* has appeared. The antibiotic contains D-phenylalanine, threonine, and diaminobutyric acid in the ratio 1 : 3 : 6.

E. Viomycin.—The nature of the chromophore and the position of the guanidine residue in viomycin are still the subjects of discussion. The presence in viomycin of the guanidine residue as part of a six-membered ring[38] has been questioned.[39] It is suggested,[39] on the basis of a positive Sakaguchi reaction and the enzymatic removal of an amidine group from viomycin, that the antibiotic contains the partial structure (5). However, this structure has been dismissed[40] as being unlikely for the viomycin chromophore, and further support for the other proposed structure[38] for the chromophore has been obtained from a study of the properties of the synthetic model compound (6). Viomycidine (7), one of the artefacts

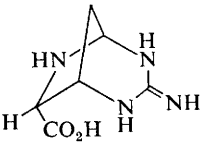
[32] S. Otani, T. Yamanori, and Y. Saito, *J. Biochem.* (*Japan*), 1969, **66**, 445.
[33] H. Arold, *J. prakt. Chem.*, 1969, **311**, 278.
[34] H. Arold and O. Barth, *J. prakt. Chem.*, 1968, **308**, 50; A. Arold and K. Haller, *ibid.*, 1969, **311**, 3; H. Arold, V. Fischer, and H. Feist, *ibid.*, p. 490; H. Arold, H. Feist, and K. Wilding, *ibid.*, p. 511.
[35] E. M. Saad Salem, N. V. Fedoseeva, and A. B. Silaev, *Zhur. obshchei Khim.*, 1968, **38**, 2227; 1969, **39**, 89; N. V. Fedoseeva, V. V. Shilin, and A. B. Silaev, *ibid.*, 1969, **39**, 85.
[36] E. A. Morozova, E. S. Oksenoit, and E. N. Gorbacheva, *Zhur. obshchei Khim.*, 1968, **38**, 2647; E. S. Oksenoit, E. A. Morozova, E. N. Gorbacheva, and E. E. Fokina, *ibid.*, 1969, **39**, 1654.
[37] Y. Kimura, E. Murai, M. Fujisawa, T. Tatsuki, and F. Nobue, *J. Antibiotics* (*Japan*), 1969, **22**, 449.
[38] B. W. Bycroft, D. Cameron, L. R. Croft, A. Hassanali-Walji, A. W. Johnson, and T. Webb, *Tetrahedron Letters*, 1968, 5901.
[39] L. Lechowski, *Tetrahedron Letters*, 1969, 479.
[40] B. W. Bycroft, D. Cameron, A. Hassanali-Walji, and A. W. Johnson, *Tetrahedron Letters*, 1969, 2539.



(5)



(6)



(7)

isolated from viomycin hydrolysates, undergoes hydrogenation [41] to yield 3-guanidinoproline, which in basic solution can be degraded further to 3-aminoproline. Another report of the molecular and crystal structures of (7) has appeared.[42] A previously published structure for dihydroviomycidine has been revised [43] to L-*threo*-β-guanidino-δ-hydroxy-norvaline, in agreement with the results of other workers.[38]

F. Peptides of *Amanita phalloides*.—The preliminary communications published on the isolation and structural elucidation of antamanide, *cyclo*-(Pro-Phe-Phe-Val-Pro-Pro-Ala-Phe-Phe-Pro), were reviewed in last year's Report. Experimental details covering its isolation,[44] and sequence determination [45] using g.l.c. and mass spectrometry, have now been published. The first synthesis [46] of antamanide *via* the cyclisation of a linear decapeptide has been supplemented by three other approaches. In one approach [47] the decapeptide Phe-$(Pro)_2$-$(Phe)_2$-Val-$(Pro)_2$-Ala-Phe, synthesised using the Merrifield technique, has been cyclised using dicyclohexylcarbodi-imide and *N*-hydroxysuccinimide, while in another method [48] the decapeptide $(Phe)_2$-$(Pro)_2$-$(Phe)_2$-Val-$(Pro)_2$-Ala was used for cyclisation. If the Nps-derivative of the latter peptide is converted to its *p*-nitrophenyl thiol ester, cyclisation [49] will take place without prior removal of the Nps-group.

The acyl group in ketophalloidin, the product of periodate oxidation of phalloidin, has been converted [50] to radioactive desmethylphalloin using

[41] C. Gallina, V. Koch, and A. Romeo, *Tetrahedron Letters*, 1969, 3055.
[42] G. Koyama, H. Nakamura, S. Omoto, T. Takita, K. Maeda, and Y. Iitaka, *J. Antibiotics (Japan)*, 1969, **22**, 34.
[43] T. Takita and K. Maeda, *J. Antibiotics (Japan)*, 1969, **22**, 39.
[44] T. Wieland, G. Lüben, H. Ottenheym, and H. Schiefer, *Annalen*, 1969, **722**, 173.
[45] A. Prox, J. Schmid, and H. Ottenheym, *Annalen*, 1969, **722**, 179.
[46] T. Wieland, J. Faesel, and W. Konz, *Annalen*, 1969, **722**, 197.
[47] T. Wieland, C. Birr, and F. Flor, *Annalen*, 1969, **727**, 130.
[48] W. König and R. Geiger, *Annalen*, 1969, **727**, 125.
[49] H. Faulstich, H. Trischmann, and T. Wieland, *Tetrahedron Letters*, 1969, 4131.
[50] H. Puchinger and T. Wieland, *Annalen*, 1969, **725**, 238.

NaB^3H_4. Desmethylphalloin appears to have almost the same toxicity as phalloidin.

G. Tyrocidins.—Preliminary investigations [51] on the application of n.m.r. techniques to elucidate the conformation of the tyrocidins have confirmed that the tyrocidins form high-molecular-weight aggregates in solution. An enzyme system isolated [52] from *Bacillus brevis* ATCC 8185 is capable of synthesising tyrocidins A, B, C, and D. The enzyme system appears to be relatively non-specific, since certain amino-acids, *e.g.* ornithine and leucine, can be replaced by lysine and isoleucine respectively and new tyrocidins produced. Replacement of tyrosine by phenylalanine yielded a new cyclopeptide, tyrocidin E. A cell-free extract from *B. brevis* ATCC 10068 is also able to catalyse the synthesis of the tyrocidins.[53]

H. Other Cyclic Peptides.—The structure of fungisporin, previously believed to be a cyclo-octapeptide, has been modified to account for a M.W. of 482 as determined by mass spectrometry. The structure currently favoured [54] is a cyclotetrapeptide *cyclo*-(L-Phe-D-Val-L-Val-D-Phe), and has been confirmed by synthesis. Increased yields of bacitracin A have been isolated [55] from commercial bacitracin, and further studies have been carried out on the purified material. Ionic zinc appears to bind by co-ordinate bonds to four positions in bacitracin, involving the histidine and amino-terminal thiazoline residue, thus confirming the close proximity between these two residues in bacitracin A. Evolidine, *cyclo*-(Ser-Phe-Leu-Pro-Val-Asn-Leu), has been synthesised [56] using the 2,4,6-trimethylbenzyl group for carboxy-group protection. Investigations of the relationship between structure and antibiotic activity have been carried out on peptides related to colistin.[57] Biosynthetic studies [58] on this peptide antibiotic have also been reported. Malformin A has been isolated [59] from *Aspergillus ficuum*, *A. awamori*, and *A. phoenicus*.

4 Depsipeptides

A. Valinomycin.—Physico-chemical methods have been elegantly combined [60] to choose between plausible molecular models of the conformation of valinomycin (8),* an antibiotic known to enhance K^+ ion permeability in membrane systems. O.r.d., n.m.r., and i.r. studies showed that two main

[51] A. Stern, W. A. Gibbons, and L. C. Craig, *J. Amer. Chem. Soc.*, 1969, **91**, 2794.
[52] H. Fujikawa, Y. Sakamoto, T. Suzuki, and K. Kurahashi, *Biochim. Biophys. Acta*, 1968, **169**, 520.
[53] B. Hodgson and J. S. Walker, *Biochem. J.*, 1969, **114**, 12P.
[54] R. O. Studer, *Experientia*, 1969, **25**, 899.
[55] L. C. Craig, W. F. Phillips, and M. Burachik, *Biochemistry*, 1969, **8**, 2348.
[56] F. H. C. Stewart, *Austral. J. Chem.*, 1969, **22**, 2663.
[57] T. Kurihara, H. Takeda, H. Ito, and K. Sagawa, *J. Pharm. Soc. Japan*, 1969, **89**, 531.
[58] M. Ito, K. Aida, and T. Uemura, *Agric. and Biol. Chem. (Japan)*, 1969, **33**, 262, 949.
[59] S. Iriuchijima and R. W. Curtis, *Phytochemistry*, 1969, **8**, 1397.
[60] V. T. Ivanov, I. A. Laine, N. D. Abdulaev, L. B. Senyavina, E. M. Popov, Yu. A. Ovchinnikov, and M. M. Shemyakin, *Biochem. Biophys. Res. Comm.*, 1969, **34**, 803.

* Hyiv = α-hydroxyisovaleric acid, Lac = lactic acid.

D→Val→L-Lac→L-Val→D-Hyiv→D-Val→L-Lac

D-Hyiv←L-Val←L-Lac←D-Val←D-Hyiv←L-Val

(8)

conformations exist in equilibrium, the one preferred in polar solvents such as $(CD_3)_2SO$, and the other preferred in non-polar solvents (CCl_4 or $CDCl_3$). The spin–spin coupling constants of the protons in —NH—C_αH— moieties were in agreement with the molecule (in non-polar solvents)

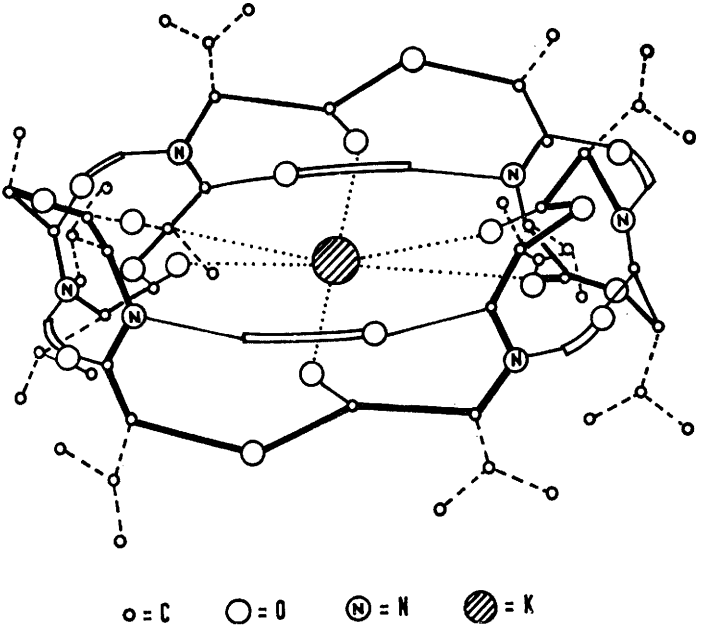


Figure 3 *Conformation of the K^+ complex of valinomycin*
(Reproduced by permission from *Biochem. Biophys. Res. Comm.*, 1969, **34**, 810)

existing as a rigid framework resembling a bracelet approximately 8 Å in diameter and 4 Å high with all amide groups forming six intramolecular hydrogen bonds. This bracelet-type structure appears to be retained in the potassium complex of valinomycin in solution, but the ester carbonyl groups are now involved in ion–dipole interaction with the K^+ ion, as shown in Figure 3. In this conformation the K^+ ion and the system of hydrogen bonds are shielded from solvent action by the hydrophobic side-chains.

In a completely independent investigation,[61] an *X*-ray-crystallographic study on the potassium aurichloride complex of valinomycin has arrived at a very similar structure to that represented in Figure 3.

[61] M. Pinkerton, L. K. Steinrauf, and P. Dawkings, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 512.

The anion–cation transport properties[61] of valinomycin show a high specificity for the transport of potassium ions. Exploratory experiments on the application of n.m.r. to the determination of the rates of complexation in ion transport have also shown[62] that valinomycin functions as a mobile ion carrier in biological membranes. It gives a measurable rate of transfer of the complexed ion in 80% methanol–20% deuteriochloroform.

Valinomycin has been synthesised[63] in 33% overall yield using the Merrifield solid-phase technique. The *N*-butoxycarbonyl-protected didepsipeptides L-Val-D-Hyiv and D-Val-L-Lac were coupled alternately to resin-bound D-Val-L-lactate using dicyclohexylcarbodi-imide, and the final linear depsipeptide was cyclised by the acid chloride method.

B. Enniatins.—The relationship between biological activity and the structure of a number of enniatin analogues has been reviewed.[28] A cyclic depsipeptide structure is a prerequisite for activity, but the activity appears

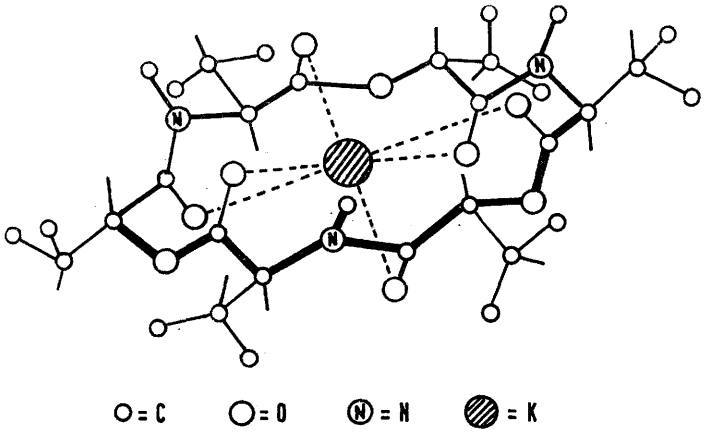


Figure 4 *Conformation of the K^+ complex of enniatin B*
(Reproduced by permission from *Biochem. Biophys. Res. Comm.*, 1969, **37**, 675)

to be more dependent on the nature of the hydroxy-acid than on the amino-acid residues.

Using physico-chemical techniques and an approach similar to that described above for valinomycin, it has been shown[64] that enniatin B and a tri-*N*-desmethyl analogue can each exist as an equilibrium mixture of two conformational forms, whose composition depends on the polarity of the solvent. The K^+ complex of enniatin B is similar to the non-complexed form in polar solvents and can be represented as in Figure 4, with six

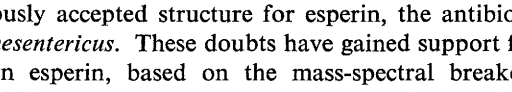
[62] D. Haynes, A. Kowalsky, and B. C. Pressman, *J. Biol. Chem.*, 1969, **244**, 502.
[63] B. F. Gisin, D. C. Tosteson, and R. B. Merrifield, *J. Amer. Chem. Soc.*, 1969, **91**, 2691.
[64] Yu. A. Ovchinnikov, V. T. Ivanov, A. V. Evstratov, V. F. Bystrov, N. D. Abdullaev, E. M. Popov, G. M. Lipkind, S. F. Arkhipova, E. S. Efremov, and M. M. Shemyakin, *Biochem. Biophys. Res. Comm.*, 1969, **37**, 668.

carbonyl groups pointing towards the interior of the molecule. The isopropyl side-chains in this conformer have a pseudo-equatorial orientation, but variable-temperature studies on the *N*-methyl signals in the n.m.r. spectra and the spin–spin coupling constants of the C_α–C_β-protons in non-polar solvents have shown that a conformer with three adjacent isopropyl groups pseudo-axial and the other three pseudo-equatorial can exist. Increasing the size of the central complexing cation gives a corresponding increase in the dihedral angle between H—N—C_α and N—C_α—H planes. The molecule can therefore 'open out' to accommodate the larger cations.

X-Ray-crystallographic studies [65] on the potassium complex of enniatin B reveal a similar structure to that described above, with the six carbonyl oxygen atoms surrounding the K^+ ion and all the isopropyl groups pointing outwards from the centre.

A synthetic enniatin B analogue, in which D-valine replaces all the D-α-hydroxyisovaleric acid residues, and an analogue with L-*N*-methylvaline residues replaced by L-*N*-methylalanine have both been shown [66] to be biologically inactive. A study of the mass-spectrometric breakdown patterns of the enniatin antibiotics has also been reported.[67]

C. Esperin.—Synthetic studies have been reported [68] which cast doubts on the previously accepted structure for esperin, the antibiotic produced by *Bacillus mesentericus*. These doubts have gained support from a structural study [69] on esperin, based on the mass-spectral breakdown of *N*-permethylated derivatives. The fatty-acid residue in esperin is a mixture of three homologues and the antibiotic extract also contains a homologue in which *C*-terminal leucine is replaced by valine. The structure (9) has now been proposed for esperin.

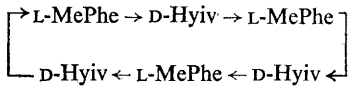


[9, R = $C_{12}H_{25}$ (45%), R = $C_{11}H_{23}$ (35%), R = $C_{10}H_{21}$ (20%)]

D. Stendomycin.—Full details have now been published [70] on the elucidation of the structure of stendomycin, and of stendomycidine, the basic amino-acid found in stendomycin. Details of the structures were discussed in last year's Report.

[65] M. Dobler, J. D. Dunitz, and J. Krajewski, *J. Mol. Biol.*, 1969, **42**, 603.
[66] G. Losse and H. Raue, *Tetrahedron*, 1969, **25**, 2677.
[67] A. A. Kiryushkin, B. V. Rozinov, and Yu. A. Ovchinnikov, *Khim. prirod. Soedinenii*, 1968, 182.
[68] Yu. A. Ovchinnikov, V. T. Ivanov, P. V. Kostetskii, and M. M. Shemyakin, *Khim. prirod. Soedinenii*, 1968, 236.
[69] D. W. Thomas and T. Ito, *Tetrahedron*, 1969, **25**, 1985.
[70] M. Bodanszky, J. Izdebski, and I. Muramatsu, *J. Amer. Chem. Soc.*, 1969, **91**, 2351; M. Bodanszky, G. G. Marconi, and A. Bodanszky, *J. Antibiotics* (*Japan*), 1969, **22**. 40.

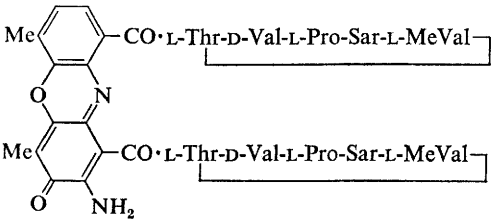
E. Beauvericin.—The structure of a toxic compound isolated from *Beauveria bassiana* has been shown [71] to be a cyclodepsipeptide (10). A



(10)

relatively simple n.m.r. spectrum was characteristic of a symmetrical molecule made up of repeating units of *N*-methylphenylalanine and D-hydroxyisovaleric acid. The exact sequence, however, could only be determined after a detailed mass spectrometric analysis of the product obtained from lithium aluminium deuteride reduction of beauvericin.

F. Actinomycins.—A preliminary report on a study [72] of the 60 MHz n.m.r. spectrum of actinomycin D (11) in deuteriobenzene shows that the



(11)

spin–spin coupling constants for protons in the —NH—C_αH— moiety are in agreement with the proposal that all valyl and threonyl residues have a *trans* conformation. Side-chain groups in the same amino-acid residues situated in different pentapeptide lactone rings appear in most cases to be non-equivalent. The methyl groups in methylvaline (α and β), threonine (β), and sarcosine were not affected by solvent changes and therefore appear to be equivalent.

Oxidative coupling experiments [73] carried out on model compounds such as *O*-acetyl-*N*-[3-amino-2-hydroxy-4-methylbenzoyl]-L-threonine methyl esters have shown that the two products obtained can be separated chromatographically, thus indicating that a possible route to actinomycins might involve an oxidative condensation of *N*-(3-amino-2-hydroxy-4-methylbenzoyl)pentapeptide lactones. Analogues of the actinomycin chromophore have been synthesised.[74]

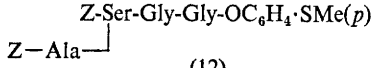
[71] R. L. Hamill, C. E. Higgens, H. E. Boaz, and M. Gorman, *Tetrahedron Letters*, 1969, 4255.
[72] T. A. Victor, F. E. Hruska, C. L. Bell, and S. S. Danyluk, *Tetrahedron Letters*, 1969, 4721.
[73] H. Brockmann and E. Schulze, *Chem. Ber.*, 1969, **102**, 3205.
[74] E. N. Glibin and O. F. Ginzburg, *Khim. geterotsikl. Soedinenii*, 1969, 40.

4-Methyl-3-hydroxyanthranilic acid and certain peptide derivatives have been incorporated [75] biosynthetically into actinomycin, and it therefore appears likely that the enzyme systems are capable of synthesising actinocynil peptides. The extent to which these peptides might be intermediates in the biosynthesis is not known. A review of the work on the mechanism of the biosynthesis of actinomycin has appeared.[76]

G. Synthesis.—Stereoisomers of *N*-lactylvaline have been synthesised [77] by condensing the *O*-t-butyl ether of lactic acid with an ester of valine using the mixed anhydride method. The alkylating properties of *O*-tosyl-glycollates have been utilised in the synthesis of a number of didepsipeptides.[78] *p*-Nitrobenzyl esters of *O*-tosyl-glycollates reacted readily with *N*-acylated amino-acids to give *N*-acyldepsipeptide nitrobenzyl esters.

Two depsipeptide analogues of angiotensin, one in which *p*-hydroxyphenyl-lactic acid replaced tyrosine and another in which L-α-hydroxyisovaleric acid replaced valine, have been synthesised [79] using the Merrifield technique. Monomer tridepsipeptides such as Ala-Ala-Glyc-ONp,HBr* have been synthesised [80] using the pentamethylbenzyl group for carboxy-group protection, and the accelerated active ester method for the formation of the ester bond. Polymerisation of these tridepsipeptides, carried out using a high concentration of *p*-nitrophenyl esters, gave regularly repeating unit sequences.

The problem of removing carboxy-protecting groups in the presence of more labile *O*-peptide ester linkages has been partly solved by using the 4-(methylthio)phenyl group for carboxy-group protection.[81] In the depsipeptide (12) the ester bond was formed by treating *N*-benzyloxycarbonyl-alanine with Z-Ser-Gly-OC_6H_4·SMe(*p*) in the presence of ethyl chloroformate to give the tripeptide. Further extension of the peptide chain to give (12) was carried out by converting the protective 4-(methylthio)-



(12)

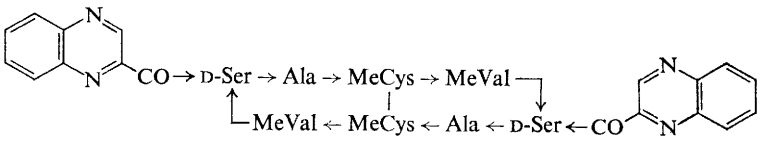
phenyl ester to the 4-(methylsulphonyl)phenyl activated ester using mild oxidation conditions without cleavage of the *O*-depsipeptide link. The method cannot, however, be used for sulphur-containing amino-acid residues.

[75] L. Salzman, H. Wiessbach, and E. Katz, *Arch. Biochem. Biophys.*, 1969, **130**, 536; E. E. Golub, M. A. Ward, and J. S. Nishimura, *J. Bacteriol.*, 1969, **100**, 977.
[76] T. I. Orlova, *Antibiotiki*, 1968, **13**, 939.
[77] C. Wasielewski, *Roczniki Chem.*, 1968, **42**, 1479.
[78] C. Wasielewski, *Roczniki Chem.*, 1969, **43**, 1419.
[79] E. P. Semkin, A. P. Smirnova, and L. A. Shchukina, *Zhur. obshchei Khim.*, 1968, **38**, 2358.
[80] F. H. C. Stewart, *Austral. J. Chem.*, 1969, **22**, 1291.
[81] B. J. Johnson, *J. Org. Chem.*, 1969, **34**, 1178.

* Glyc = glycollic acid.

The synthesis of alkylated derivatives of vernamycin B has been reported in a patent.[82] A cyclic tridepsipeptide, *cyclo*-(DL-Lac-Glyc-Gly)* has been obtained[83] from the base-catalysed rearrangement of *N*-(hydroxyacetyl)-3-methyl-2,5-dioxomorpholine.

H. Quinoxaline Antibiotics.—The biosynthesis of triostin (13) has been shown[84] to be independent of protein synthesis, and all constituent



(13)

residues are derived from L-amino-acids. Tryptophan appears to be the source of the quinoxaline moiety in the molecule and each *N*-methyl group is derived from methionine. A number of *N*-(2-quinoxaloyl) derivatives of amino-acids and dipeptides have been synthesised[85] and tested for anti-tumour activity. The favoured synthetic route for the preparation of these derivatives was one involving initial formation of the peptide bond, followed by quinoxaloylation.

I. Miscellaneous.—Doubt has been cast[86] on the published structure (14) for the linear depsipeptide viscosin, the antibiotic isolated from *Pseudomonas viscosa*. The peptide (14) and a diastereoisomer (with D- replacing

D-$CH_3(CH_2)_6$·CHOH·CH_2·CO-L-Leu-Gly-L-Ser-D-Val-L-Thr-L-Leu

(14)

D-Hyiv-L-MeIle-D-Hyiv-L-MeLeu

(15)

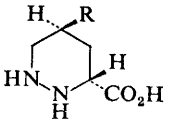
L-valine) have been synthesised, but both show different physical properties from natural viscosin. An *X*-ray diffraction analysis[87] of a crystal of the cyclotetradepsipeptide (15) has shown that the ring structure contains the two hydroxy-acids in a *trans* conformation and the two peptide units in a *cis* conformation. All the carbonyl groups extend in the same direction

[82] M. Bodanszky, U.S. Patent 3,420,816/1969.
[83] R. Kazmierczak and G. Kupryszewski, *Zeszyty Nauk, Mat., Fiz., Chem.*, 1968, **8**, 165.
[84] T. Yoshida and K. Katagiri, *Biochemistry*, 1969, **8**, 2645.
[85] S. Gerchakov and H. P. Schultz, *J. Medicin. Chem.*, 1969, **12**, 141.
[86] M. Hiramoto, K. Okada, S. Nagai, and H. Kawamoto, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 702.
[87] J. Konnert and I. L. Karle, *J. Amer. Chem. Soc.*, 1969, **91**, 4888.

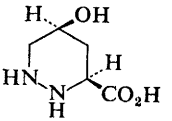
* Glyc = glycollic acid.

almost perpendicular to the average plane of the ring, while the hydrocarbon side-chains extend in the opposite direction.

Preliminary communications have appeared [3, 88] on the structure of novel imino- and amino-acids isolated from the hydrolysates of the monamycins, a new family of cyclic depsipeptide antibiotics. The three piperidazinecarboxylic acids (16), (17), and (18), D-isoleucine, *N*-methyl-D-leucine,



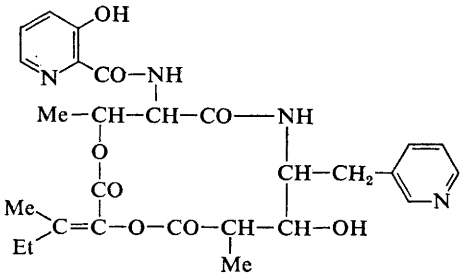
(16, R = H)
(17, R = Cl)



(18)

D-valine, *trans*-4-methyl-L-proline, and a trace of L-proline, together with two hydroxy-acids, have been isolated from the monamycin hydrolysates.

Extensive labelling studies [89] carried out on pyridomycin (19), show that



(19)

the two pyridine nuclei in the cyclic depsipeptide are derived from L-aspartic acid and glycerol or pyruvic acid. L-Isoleucine functions as a precursor of the 2-hydroxy-3-methylpent-2-enoic acid residue.

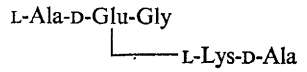
The mass-spectral fragmentation patterns of the depsipeptides staphylomycin S and etamycin show [90] certain of the characteristic breakdown patterns observed in cyclic peptides. The most important characteristic fragmentation is the loss of the elements of the ester link, followed by a sequential loss of amino-acid residues in such a way that the positive charge always remains on the 3-hydroxypicolinic acid residue. The sample of etamycin analysed in this way appeared to contain a lower homologue, probably one in which serine replaced threonine.

[88] C. H. Hassall, R. B. Morton, Y. Ogihara, and W. A. Thomas, *Chem. Comm.*, 1969, 1079.
[89] H. Ogawara, K. Maeda, and H. Umezawa, *Biochemistry*, 1968, **7**, 3296.
[90] I. A. Bogdanova, A. A. Kiryushkin, B. Rozynov, and V. M. Burikov, *Zhur. obshchei Khim.*, 1969, **39**, 891.

5 Peptide–Carbohydrate Linkages

A. Glycopeptides.—The present state of knowledge on the structure of the peptidoglycans in bacterial cell walls has been reviewed.[91] A wide variety of glycopeptides have been found to contain the tetrapeptide L-Ala-γ-D-Glu-L-Lys(or *meso*-diaminopimelic acid, Dpm)-D-Ala as the basic unit linking up glycan polymers which are made up of *N*-acetyl-D-glucosamine (GNAc) and *N*-acetylmuramic acid (MurNAc) units. Other peptide units, for example pentaglycyl (see Figure 5) in the peptidoglycan of *Staphylococcus aureus*,[92] are involved in cross-linking between the ε-amino-group of lysine of the peptide subunit and the *C*-terminal D-alanine of another subunit, to form the three-dimensional structure of peptidoglycans.

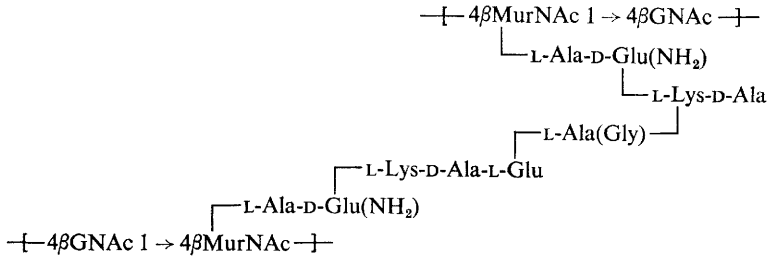
The peptide subunit (20) found in the cell walls of *Micrococcus lysodeikticus*,[91] cross-linked through D-Ala-L-Ala and N^{ε}-(D-Ala)-L-Lys, has also



(20)

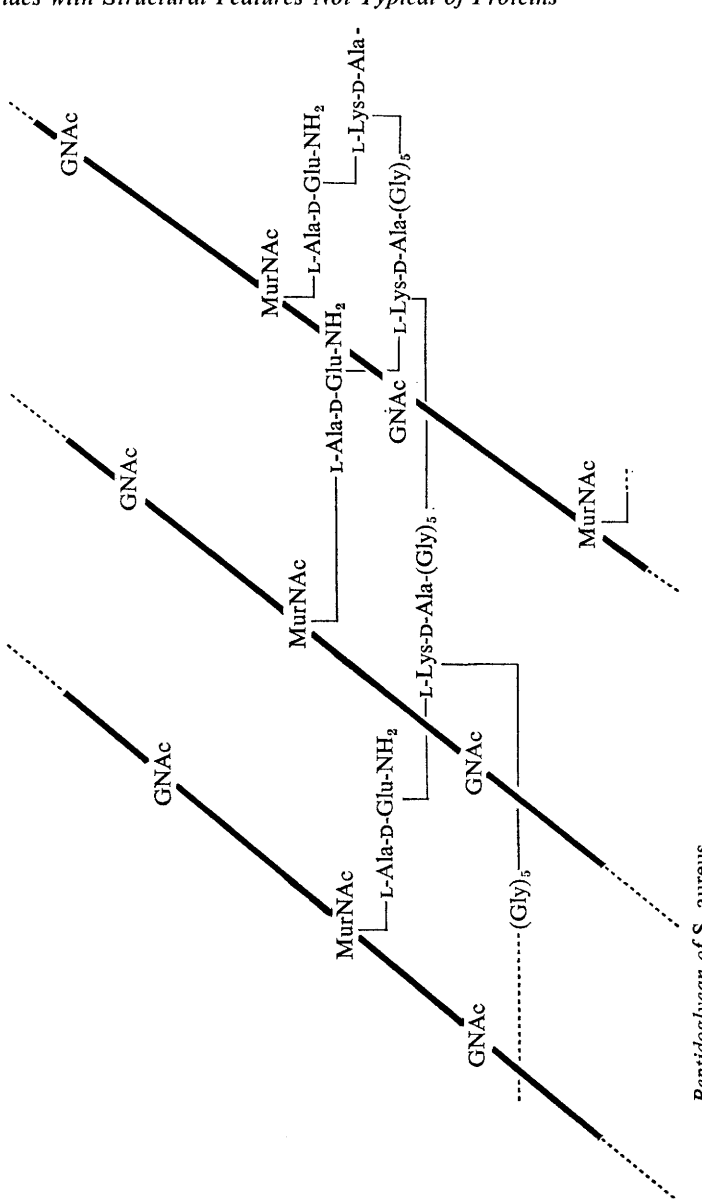
been identified[93] in the walls of *M. flavus*, *M. citreus*, and *Sarcina lutea*. Although, in general, only 1 mole of glutamic acid per mole of peptidoglycan subunit has been found, mureins of *Micrococcus luteus* and *M. freudenreichii* have been shown[94] to contain 2 moles of glutamic acid, the one having the D- and the other having the L-configuration. Investigations using enzymes confirmed that the L-glutamic acid was involved in cross-linking, while the D-isomer was part of the tetrapeptide attached to muramic acid, as shown in (21).

Chemical and enzymatic degradation[95] of the peptidoglycan of *Butyribacterium rettgeri* showed that the peptide subunit was cross-linked by



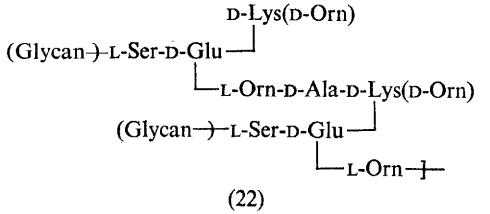
(21)

[91] M. J. Osborn, *Ann. Rev. Biochem.*, 1969, **38**, 501.
[92] J.-M. Ghuysen, *Bacterial Rev.*, 1968, **32**, 425.
[93] J. N. Campbell, M. Leyh-Bouille, and J.-M. Ghuysen, *Biochemistry*, 1969, **8**, 193.
[94] E. Niebler, K. Schleifer, and O. Kandler, *Biochem. Biophys. Res. Comm.*, 1969, **34**, 560.
[95] M. Guinand, J.-M. Ghuysen, K. H. Schleifer, and O. Kandler, *Biochemistry*, 1969, **8**, 200.



Peptidoglycan of S. aureus

means of D-lysine or a D-ornithine residue, as in (22). Degradation of the cell walls using KM endopeptidase showed that the cross-linking was



(22)

exceptional in that the α-carboxy-group of the glutamic acid was linked to the lysine or ornithine residues.

Yet another type of cross-linking occurs [96] between the peptide subunits L-Ala-γ-D-Glu-(L)-*meso*-Dpm-(L)-D-Ala of the peptidoglycans of *Escherichia coli* B and *Bacillus megaterium* KM. Enzymic digestion with KM endopeptidase established that the linkage must involve D-Ala-(D)-*meso*-Dpm, although in the case of *B. megaterium* linkages involving DD-Dpm could also be present. Much of the recent success in the elucidation of bacterial cell-wall structure can be attributed to the availability of enzymes for specific hydrolysis of peptide bonds. Improved techniques have been reported [97] for the preparation of six important peptidases from *Streptomyces* and their specificities have been investigated.

The biosynthetic precursor of the tetrapeptide subunits in peptidoglycans has been known for some time to be a nucleotide-pentapeptide having a *C*-terminal D-Ala-D-Ala group. In the transpeptidation process the terminal D-Ala is lost, and in general peptidoglycan subunits have only 1 mole of D-Ala present. Examples have, however, been found where the pentapeptide attached to the nucleotide has been transferred intact, to a peptidoglycan subunit. Studies carried out [98] on *Chalaropis* B enzyme digest of the cell walls of *S. aureus* and *S. epidermidis* have shown that *C*-terminal peptide subunits carry an intact D-Ala-D-Ala residue. Using Edman degradation techniques it has been confirmed [98] that the major cross-links involve pentaglycyl bridges, but about 6% of the cross-bridge pentapeptides in *S. aureus* are Gly-Ser-$(Gly)_3$, while in *S. epidermidis* serine is non-randomly distributed throughout the pentapeptide cross-links. Digestion of the peptidoglycans of *S. aureus* and *S. epidermidis* with *Myxobacter* AL-1 gave similar conclusions.[99]

Analysis of the glycopeptides obtained after digestion with *Myxobacter* AL-1 enzyme has shown [100] that the structure of the cell-wall peptidoglycan

[96] J. van Heijenoort, L. Elbaz, P. Dezélée, J.-F. Petit, E. Bricas, and J.-M. Ghuysen, *Biochemistry*, 1969, **8**, 207.
[97] J.-M. Ghuysen, L. Dierickx, J. Coyette, M. Leyh-Bouille, M. Guinand, and J. N. Campbell, *Biochemistry*, 1969, **8**, 213.
[98] D. J. Tipper and M. F. Berman, *Biochemistry*, 1969, **8**, 2183.
[99] D. J. Tipper, *Biochemistry*, 1969, **8**, 2192.
[100] K. D. Hungerer, J. Fleck, and D. J. Tipper, *Biochemistry*, 1969, **8**, 3567.

of *Lactobacillus casei* RO94 can be presented as shown in Figure 6. D-Alanine-D-isoasparagine linkages were rapidly hydrolysed by the AL-1

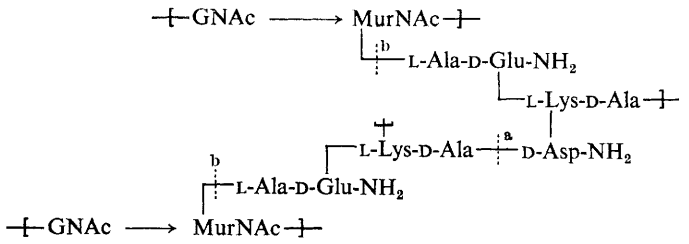
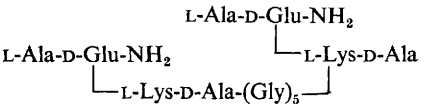


Figure 6 *Structure of peptidoglycan of* L. casei *showing sites of primary attack (a), and secondary attack (b), by* Myxobacter *AL-1 enzyme*

enzyme as well as most of the *N*-acetylmuramyl-L-alanine linkages. The peptide subunits in the vegetative cell wall peptidoglycan of *Bacillus sphaericus* 9602 resemble [101] those of *L. casei* (Figure 6) except that they lack *C*-terminal D-alanine residues and free carboxy-groups replace the amido-group on glutamic acid. *Myxobacter* AL-1 digests of the cell walls of *Mycobacterium smegmatis* [102] and L-3 enzyme digests of *Lactobacillus plantarum* [103] cell wall have yielded tri- and tetra-peptides based on the structure Ala-Glu-Dpm-Ala. Digestion of the cell walls of *L. plantarum* with an enzyme preparation from *Flavobacterium* liberated, amongst other peptides, the tripeptide L-Ala-D-isoglutaminyl-*meso*-Dpm (diamide), a tetrapeptide L-Ala-D-isoglutaminyl-*meso*-Dpm (monoamide)-D-Ala, and a heptapeptide consisting of the residues of the tri- and tetra-peptides joined between the amino-group of diaminopimelic acid and the carboxy-group of D-alanine. Thus the L-3 enzyme must have deamidated the Dpm-amide and hydrolysed the D-Ala-*meso*-Dpm linkage. The identification of a peptidic fraction with *N*-terminal L-alanine in the autolysate fractions of *Listeria monocytogenes* has confirmed [104] that an *N*-acetylmuramyl-L-alanine amidase must be present in the walls themselves. Details of the synthesis of the tridecapeptide (23) from the cell-wall peptidoglycan of *S. aureus* have appeared.[105]



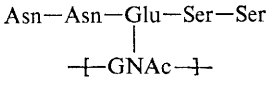
(23)

[101] K. D. Hungerer and D. J. Tipper, *Biochemistry*, 1969, **8**, 3577.
[102] J. F. Petit, A. Adam, J. Wietzerbin-Falszpan, E. Lederer, and J.-M.Ghuysen, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 478.
[103] T. Matsuda, S. Kotani, and K. Kato, *Biken's J.*, 1968, **11**, 111, 127.
[104] R. Tinelli, *Bull. Soc. Chim. biol.*, 1969, **51**, 283.
[105] P. Lefrancier and E. Bricas, *Bull. Soc. chim. France*, 1969, 3561.

Biosynthetic studies carried out to investigate the origins of the pentaglycine bridge in cell walls of *S. aureus* (see Figure 5) have been reported by two research groups.[106, 107] It was found that the pentaglycine bridge was introduced into the peptidoglycan by sequential addition of glycine from glycyl-tRNA to the ε-amino-terminal of lysine residues linked to a lipid precursor.

Glycopeptides isolated from the cell walls of cultures of tomato have been found[108] to be rich in hydroxyproline, the hydroxy-group of which appears to be linked to the sugar moiety. Preliminary studies[109] on two mucopeptides prepared from pronase digests of *Mycobacterium smegmatis* indicate that the peptides show close similarities with mucopeptides from other bacteria. Component amino-acid and sugar analyses indicated that diaminopimelic acid, glutamic acid, alanine, glucosamine, and muramic acid were present in the ratio 1 : 1 : 2 : 1 : 1. Mucopeptides[110] from *Streptococcus bovis* contain alanine, glutamic acid, lysine, and threonine in the ratio 3 : 1 : 1 : 1. The isolation of N^{ε}-Lys-Thr from the hydrolysates suggests that threonine is associated with the cross-bridge linking the ε-amino-group of lysine of one tetrapeptide to the D-alanine of another matrix.

An increasing number of glycopeptides containing *N*-glycosidic amino-acid linkages have been characterised. The glycopeptide (24) and related

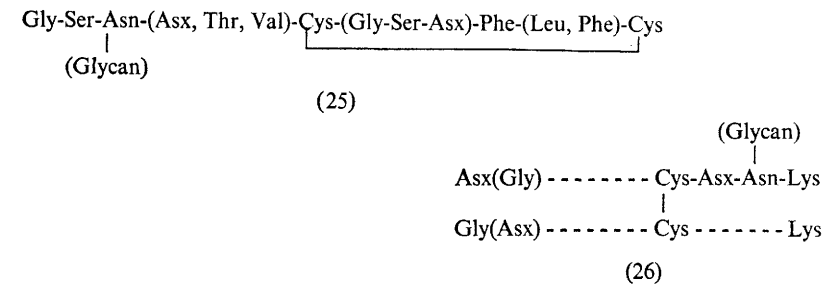


(24)

peptides have been identified[111] in the enzyme digests of pineapple stem bromelain. Glycopeptides from proteolytic digests of yeast invertase[112] and pronase digests of ox fibrin[113] contain linkages between the amide group of asparagine and the oligosaccharide units. All the glycopeptide fractions isolated from hen ovomucoid preparations[114] contained a *N*-(β-aspartyl)-*N*-acetylglucosaminylamine* linkage. Partial sequences (25) and (26) have been proposed for the structure of two glycopeptides obtained from the tryptic hydrolysates of human transferrin.[115]

[106] J. Thorndike and J. T. Park, *Biochem. Biophys. Res. Comm.*, 1969, **35**, 642.
[107] T. Kamiryo and M. Matsuhashi, *Biochem. Biophys. Res. Comm.*, 1969, **36**, 215.
[108] D. T. A. Lamport, *Biochemistry*, 1969, **8**, 1155.
[109] G. Cunto, F. Kanetsuna, and T. Imaeda, *Biochim. Biophys. Acta*, 1969, **192**, 358.
[110] J. A. Kane, H. Lackland, W. Karakawa, and R. M. Karause, *J. Bacteriol.*, 1969, **99**, 175.
[111] N. Takahashi, Y. Yasuda, M. Kuzuya, and T. Murachi, *J. Biochem.* (*Japan*), 1969, **66**, 659.
[112] N. P. Neumann and J. O. Lampen, *Biochemistry*, 1969, **8**, 3552.
[113] M. Meszaros, *Acta Biochim. Biophys. Acad. Sci. Hung.*, 1968, **3**, 421.
[114] M. Monsigny, A. Adam-Chosson, and J. Montreuil, *Bull. Soc. Chim. biol.*, 1968, **50**, 857.
[115] G. Spik and J. Montreuil, *Bull. Soc. Chim. biol.*, 1969, **51**, 1271; P. Charet, M. Monsigny, G. Spik, and J. Montreuil, *Compt. rend.*, 1969, **269**, *D*, 1019.

* *N*-β-aspartyl-2-acetamido-2-deoxyglucosylamine.



The amino-acid content and the nature of the carbohydrate–peptide linkages have been determined for the glycopeptides in enzyme digests [116, 117] of the α_1-acid glycoprotein from human plasma. The isolation [117] of a β-aspartyl-acetylglucosaminylamine fragment provided evidence for the nature of the linkage. Four cystine-containing peptides have been characterised [118] in the pronase digests of reduced α_1-acid glycoprotein from human plasma.

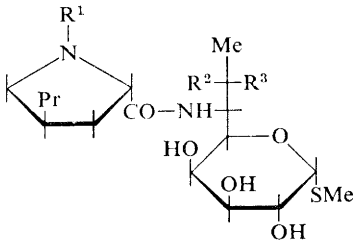
Threonine, proline, and serine have been found [119] to be the predominant amino-acids in the glycopeptides obtained from protease digests of the gastric mucosa of rabbit. The lability of the carbohydrate–peptide linkage towards alkali indicates that the linkage must involve the hydroxy-groups of threonine and serine. Similar linkages appear to be the major type in the sulphated glycopeptides isolated [120] from the intestinal mucosae of rabbit, in the glycopeptide (M.W. 3000) from bovine submaxillary mucin,[121] and in the pronase digests [122] of M and N blood-group glycoproteins.

Proteolytic digests of a 7S protein in soybean globulins have yielded a glycopeptide (M.W. 9870) composed [123] of the amino-acids, Asp_6, Thr_3, Gly_3, Ala_3, Glu, Ser, and the sugar residues $mannose_{39}$ and $glucosamine_{12}$. The partial sequences of tryptic glycopeptides of bovine thyrotropin are similar [124] to the glycopeptides from bovine and ovine luteinising hormone. Typical of the peptides found were Val-Glx-Asx-(Ser Thr_2 Glx Cys_3 His_2)-(Tyr_2 His)-Lys and Asx-Ile-(Ser Thr_2 Glx Cys_2 Ala_2 Val)-Lys. Six glycopeptide fractions [125] from human platelet membranes contain aspartic acid and glucosamine as major constituents. Serine, glycine, aspartic acid, glutamic acid, and alanine have been characterised in the hydrolysates of the first crystalline glycopeptide isolated [126] from human urine.

[116] T. Yamauchi, M. Makino, and I. Yamashina, *J. Biochem.* (*Japan*), 1968, **64**, 683; T. Yamauchi and I. Yamashina, *ibid.*, 1969, **66**, 213.
[117] P. V. Wagh, I. Bornstein, and R. J. Winzler, *J. Biol. Chem.*, 1969, **244**, 658.
[118] M. Ishiguro, A. Yamamoto, and K. Schmid, *Biochim. Biophys. Acta*, 1969, **194**, 91.
[119] T. Nemoto and Z. Yosizawa, *J. Biochem.* (*Japan*), 1969, **66**, 627.
[120] T. Nemoto and Z. Yosizawa, *Biochim. Biophys. Acta*, 1969, **192**, 37.
[121] F. Downs and W. Pigman, *Biochemistry*, 1969, **8**, 1760.
[122] E. Lisowska, *European J. Biochem.*, 1969, **10**, 574.
[123] I. Koshiyama, *Arch. Biochem. Biophys.*, 1969, **130**, 370.
[124] S. M. Howard and J. G. Pierce, *J. Biol. Chem.*, 1969, **244**, 6468.
[125] D. S. Pepper and G. A. Jamieson, *Biochemistry*, 1969, **8**, 3362.
[126] D. Basu, *Biochem. J.*, 1969, **112**, 379.

The results of sequential periodate oxidations have confirmed [127] the sequence Man-Man-GNAc-GNAc-Ser-Asp* for a glycopeptide from Taka-amylase A. The ion-exchange columns of amino-acid analysers have been used [128] for the quantitative estimation of the hydroxylysine-linked carbohydrate units of several collagens. The composition and properties of glycopeptides from κ-casein [129] and rabbit haptoglobin [130] have been reported.

B. Amino-acid–Carbohydrate Antibiotics.—Mass spectrometry, combined with labelling studies, has shown [131] that the *S*-methyl group, *N*-methyl group, and a terminal methyl group on the propyl side-chain of the amino-acid in lincomycin (27), are all derived from C_1 donors such as methionine.



(27, R^1 = Me, R^2 = OH, R^3 = H)
(28, R^1 = H, R^2 = H, R^3 = Cl)

A number of chloro-analogues, *e.g.* (28), of lincomycin have been synthesised [132] as potential antibacterial and antimalarial agents.

6 Peptides and Amino-acids Linked to Nucleosides and Nucleotides

A. Peptidyl-tRNA.—Investigations on the synthesis of acylamino- and dipeptidyl-tRNA have been extended to tri- and tetra-peptidyl-tRNA [133] containing glycine, alanine, valine, and phenylalanine, and to oligopeptidyl-tRNA.[134] The peptidyl-tRNA series were synthesised by treating *N*-hydroxysuccinimide esters of *N*-monomethoxytrityl-protected peptides with amino-acyl-tRNA. The *N*-protecting group was removed under mild conditions (5% dichloroacetic acid for 5 min at 4 °C). Several of the

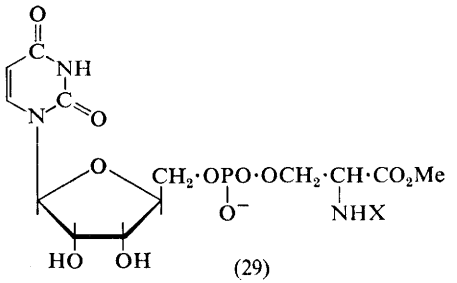
[127] H. Yamaguchi, T. Ikenaka, and Y. Matsushima, *J. Biochem.* (*Japan*), 1969, **65**, 793.
[128] R. G. Spiro, *J. Biol. Chem.*, 1969, **244**, 602.
[129] T. Kuwata, R. Niki, and S. Arima, *J. Agric. Chem. Soc. Japan*, 1969, **43**, 183.
[130] R. I. Cheftel, M. A. Parnaudeau, R. Bourillon, and J. Moretti, *European J. Biochem.*, 1969, **9**, 585.
[131] A. D. Argoudelis, T. E. Eble, J. A. Fox, and D. J. Mason, *Biochemistry*, 1969, **8**, 3408.
[132] B. J. Magerlein and F. Kagan, *J. Medicin. Chem.*, 1969, **12**, 780.
[133] Y. Lapidot, N. De Groot, and S. Rappoport, *Biochim. Biophys. Acta*, 1969, **182**, 105.
[134] Y. Lapidot, N. De Groot, S. Rappoport, and D. Elat, *Biochem. Biophys. Acta*, 1969, **190**, 304.

* Man = Mannose, GNAc = *N*-acetyl-D-glucosamine.

synthetic peptidyl-tRNA compounds have been subjected to enzymic hydrolysis.[135] Peptidyl-tRNA with two peptide bonds was hydrolysed at a considerably faster rate than *N*-acylamino-acyl-tRNA, and Gly_2 Phe-tRNA was hydrolysed faster than Gly-Phe-tRNA. The hydrolysis rate, therefore, depends on the chain length.

B. Nucleopeptides.—Interest continues to be focused on amino-acids bonded to purine or pyrimidine derivatives as simple analogues and potential building blocks for the synthesis of nucleopeptides. N^{ε}-(6-Purinoyl)-L-lysine has been prepared[136] by treatment of L-lysine copper complex or its α-benzyloxycarbonyl derivative with 6-trichloromethylpurine, and β-(5-uracilyl)-DL-alanine has been synthesised[136] from 5-chloromethyluracil and diethyl acetamidomalonate. Studies on (purin-9-yl)-amino-acids have also been reported.[137] Synthetic routes to thymine and cytosine analogues of DL-β-(2,4-dihydroxypyrimidin-1-yl)alanine and DL-β-(6-aminopurin-9-yl)alanine have been published.[138] A brief communication has appeared[139] on the n.m.r. spectrum of DL-*N*-(1,2-dihydro-2-oxo-4-pyrimidinyl)serine.

A great deal of information has been accumulated on the stability and properties of the nucleotide–peptide bond in a number of model compounds. The phosphate linkage in compounds such as uridylyl-(5′ → *O*)-seryl ester (29) is stable between pH 4 and 8 but is hydrolysed in more acid and



(29)

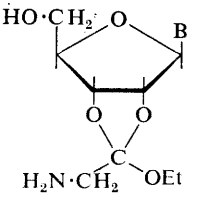
more alkaline media.[140] Provided that the amino-group of the amino-acid is free (*i.e.* X = H in 29), an *O* → *N* migration of nucleotide residue occurs in alkaline media to give uridylyl-(5′ → *N*)-hydroxyamino-acids. The lability of the phosphoramidate bond in a series of model esters[141] such as uridylyl-(5 → *N*)-Ser(or Thr)-Gly·OEt appears to be dependent on the

[135] N. De Groot, Y. Groner, and Y. Lapidot, *Biochim. Biophys. Acta*, 1969, **186**, 286.
[136] A. Vincze, C. Lachman, and S. Cohen, *Israel J. Chem.*, 1968, **6**, 641.
[137] M. Yu. Lidak, Y. A. Schluke, and Y. P. Shavihkin, *Khim. geterotsikl. Soedinenii*, 1968, 955.
[138] M. T. Doel, A. S. Jones, and N. Taylor, *Tetrahedron Letters*, 1969, 2285.
[139] S. Hofmann and E. Muhle, *Z. Chem.*, 1969, **9**, 112.
[140] B. A. Yuodka, E. P. Savel'ev, Z. A. Shabarova, and M. A. Prokof'ev, *Biokhimiya*, 1968, **33**, 907.
[141] B. A. Yuodka, Z. A. Shabarova, and M. A. Prokof'ev, *Vestnik Moskov. Univ.*, 1968, **23**, 104.

nature of the hydroxy-amino-acid residue.[142] In acid media phospho-amide and phospho-ester bonds were hydrolysed at different rates depending upon whether the amino-acid residue linked to the nucleotide was serine or threonine. The involvement of the hydroxy-group of the hydroxy-amino-acid in the mechanism of hydrolysis is implied from the results. Intramolecular catalysis of the cleavage of the phosphoamidate bond by a hydroxy-group also takes place in the β-hydroxyethyl ester of uridylyl-(5′ → *N*)-alanine.[143]

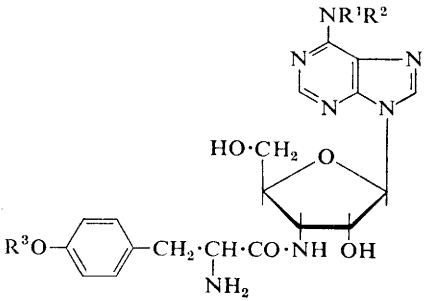
Nucleotidyl-(5′ → *N*)-peptides can be differentiated[144] from nucleotidyl (5′ → *O*)-peptides in their reaction with hydroxylamine or *N*-methyl-hydroxylamine. The former peptides undergo hydrolysis at the phospho-amide bond on treatment with hydroxylamine while the latter (5′ → *O*)-peptides show no cleavage.

Synthetic intermediates such as (30) have been caused to react[145] with appropriate *N*-protected amino-acid active esters to give nucleopeptides



(30)

B = adenine



(31, $R^1 = R^2 = R^3 = H$)
(32, $R^1 = H$, $R^2 = Me$, $R^3 = H$)
(33, $R^1 = R^2 = Me$, $R^3 = H$)
(34, $R^1 = R^2 = R^3 = Me$)

[142] B. A. Yuodka, I. V. Obruchnikov, Z. A. Shabarova, and M. A. Prokof'ev, *Khim. prirod. Soedinenii*, 1968, 323; B. A. Yuodka, I. V. Obruchnikov, V. K. Nedbai, Z. A. Shabarova, and M. A. Prokof'ev, *Biokhimiya*, 1969, **34**, 647.

[143] V. D. Smirnov, N. G. Shinskii, Z. A. Shabarova, and M. A. Prokof'ev, *Khim. prirod. Soedinenii*, 1968, 109.

[144] B. A. Yuodka, V. K. Nedbai, Z. A. Shabarova, and M. A. Prokof'ev, *Biokhimiya*, 1969, **34**, 849.

[145] S. Chladek and J. Zemlicka, *Coll. Czech. Chem. Comm.*, 1968, **33**, 4299.

such as 2′,(3′)-*O*-(L-Val-Gly)-adenosine, simple models of *O*-peptidyl-RNA.

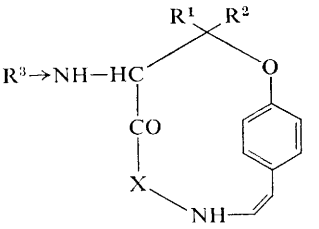
Commercial samples of puromycin (34) from *Streptomyces alboniger* have been found to contain [146] the three desmethyl biosynthetic precursors (31), (32), and (33). Labelling studies favour the sequence (31) → (34) for the biosynthesis of puromycin.

C. Polyoxins.—A complete account of the structure elucidation studies (see Volume 1 of this Report for discussion) and other aspects of the polyoxin family of peptide nucleosides has appeared.[147]

7 Peptide Alkaloids

The highly characteristic fragmentation patterns (see references in Volume 1 of this Report) exhibited by the cyclic peptide alkaloids containing the *p*-alkoxystyryl-amino-residue have been successfully used again in association with chemical degradation, for the identification of a series of new peptide alkaloids.

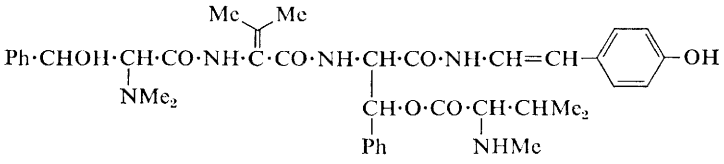
Three of the five peptide alkaloids isolated [148] from the root bark of *Ceanothus americanus* L. have been found to be identical with the alkaloids, frangulanine, adouétine-X, and adouétine-Y isolated from other plants. The two other peptide alkaloids, ceanothine D and ceanothine E, have been shown to have structures (35) and (36) respectively. Aralionine, a peptide



(35, R^1 = Me, R^2 = Et, R^3 = MePro, X = Leu)
(36, R^1 = H, R^2 = Ph, R^3 = Me_2Phe, X = Leu)
(37, R^1 = H, R^2 = Ph, R^3 = Me_2Ile, X = PhCO·CH(NH-)CO-)
(38, R^1 = H, R^2 = -$CHMe_2$, R^3 = MePhe-Pro, X = Leu)

alkaloid from *Araliorhamnus vaginatus*, has been characterised [149] as (37). Although the cyclic nature of lasiodine B (38) isolated [150] from the African Rhamnaceae *Lasiodiscus marmoratus* conforms to the usual type of peptide alkaloid, lasiodine A from the same plant has been shown to have the non-cyclic structure (39).

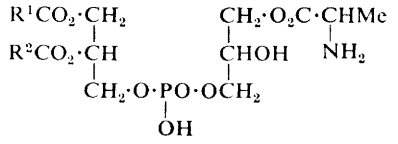
[146] T. N. Pattabiraman and B. M. Pogell, *Biochim. Biophys. Acta*, 1969, **182**, 245.
[147] K. Isono, K. Asahi, and S. Suzuki, *J. Amer. Chem. Soc.*, 1969, **91**, 7490.
[148] R. E. Servis, A. I. Kosak, R. Tschesche, E. Frohberg, and H.-W. Fehlhaber, *J. Amer. Chem. Soc.*, 1969, **91**, 5619.
[149] R. Tschesche, L. Behrendt, and H.-W. Fehlhaber, *Chem. Ber.*, 1969, **102**, 50.
[150] J. Marchand, M. Pais, X. Monseur, and F.-X. Jarreau, *Tetrahedron*, 1969, **25**, 937.



(39)

8 Peptides and Amino-acids Conjugated to Lipids

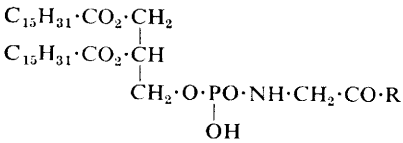
The main components of a mixture of *O*-amino-acid esters of phosphatidylglycerol isolated from *Mycoplasma laidlawii* have been identified[151] as the D- and L-alanylphosphatidylglycerols of general formula (40). Analysis



(40)

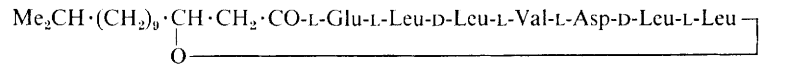
using D- and L-amino-acid oxidases showed that the isomers were present in a non-racemic mixture. It is interesting to note that biosynthetically the L-isomer of (40) was derived from L-Ala-tRNA *via* the usual mechanism for amino-acid 'activation' in protein synthesis. The D-isomer, however, required no activation *via* tRNA but required only the presence of phosphatidylglycerol, ATP, and an amino-acid 'activating' enzyme.

Threonine and hydroxyproline derivatives have been successfully added[152] to the double bond of plasmalogen vinyl ethers, *e.g.* $Me\cdot(CH_2)_{13}\cdot CH{=}CHOEt$ in the presence of toluene-*p*-sulphonic acid to give derivatives such as $Me(CH_2)_{11}\cdot CHOEt\cdot O\cdot CHMe\cdot CH(CO_2Bzl)\cdot NH\cdot Z$. A series of amino-acid and peptide derivatives of the phospho-amide type (41) have been synthesised[153] by condensing benzyl esters of the amino-acid or peptides



(41, R = -OBzl or $-NH\cdot CHMe\cdot CO_2Bzl$)

[151] W. L. Koostra and P. F. Smith, *Biochemistry*, 1969, **8**, 4794.
[152] T. I. Rubtsova, G. A. Serebrennikova, and N. A. Preobrazhenskii, *Zhur. org. Khim.*, 1969, **5**, 167.
[153] M. K. Petrova, V. I. Shvets, V. A. Dachkovskaya, and N. A. Preobrazhenskii, *Zhur. org. Chem.*, 1969, **5**, 1198; *Zhur. Vsesoyuz. Khim. obshch. im. D. I. Mendeleeva*, 1969, **14**, 470.

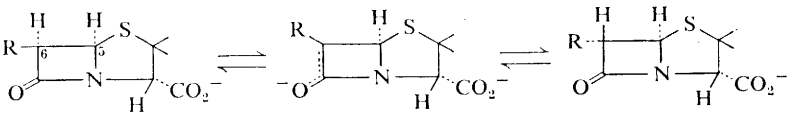


(42)

with the appropriate phosphatidylglycerol. The structure (42) has been proposed [154] for surfactin, the crystalline peptidolipid surfactant produced by *Bacillus subtilis*.

9 Penicillins

The first successful epimerisation [155] at C-6 of an intact penicillin nucleus has spurred on further investigations of the factors affecting epimerisation at the 6-position. Nitrous acid deamination of 6β-aminopenicillanic acid is well known to give inversion at C-6, and the products of such an epimerisation, *e.g.* 6-bromo- and 6-chloro-penicillanic acids, have been subjected [156] to equilibration and deuterium exchange experiments. In D_2O–NaOD solution the doublet at δ 4·82 (H at C-6) in the n.m.r. spectrum of the 6-bromo-compound disappeared slowly, while the doublet at δ 5·41 (H at C-5) sharpened to a singlet. The C-6 proton is, therefore, sufficiently acidic to permit formation of an anion, as shown in Scheme 2. Since the α-

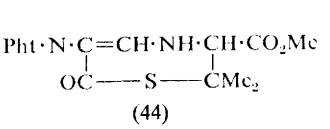


(43)

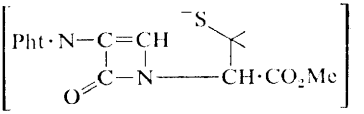
Scheme 2

epimer (43) is recovered unchanged on acidification, it appears that the anion re-protonates exclusively from the *endo*-side. In this study it was also shown that epimerisation is dependent on the electronegativity of the 6-substituent, no epimerisation taking place with substituents such as the acylamido-group. This is contrary to the published work [157] on 6-phthalimidopenicillinate, which has found further support with the isolation of (44) as a product of the reaction [158] between triethylamine and methyl 6-phthalimidopenicillinate. Compound (44) can arise only by nucleophilic attack of the sulphur atom on the β-lactam carbonyl group, which lends support to the suggestion [157] that intermediate (45) is involved in the epimerisation process in 6-phthalimidopenicillinate.

[154] A. Kakinuma, K. Arima, H. Sugino, M. Isono, and G. Tamura, *Agric. and Biol. Chem. (Japan)*, 1969, **33**, 971, 973.
[155] D. A. Johnson, D. Mania, C. A. Panetta, and H. H. Silvestri, *Tetrahedron Letters*, 1968, 1903.
[156] J. P. Clayton, J. H. C. Nayler, R. Southgate, and E. R. Stove, *Chem. Comm.*, 1969, 129.
[157] S. Wolfe and W. S. Lee, *Chem. Comm.*, 1968, 242.
[158] Ö. K. J. Kovacs, B. Ekström, and B. Sjöberg, *Tetrahedron Letters*, 1969, 1863.



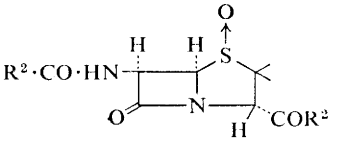
(44)



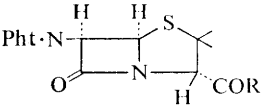
(45)

Epi-6-aminopenicillanic acid (43, R = NH_2) has been prepared[159] from epi-hetacillin by a stepwise removal of the side-chain. Acetylation of the former yielded epi-penicillin G. Nitrous acid deamination[160] of 6-amino-penicillanic acid in the presence of sodium halides gives rise to the 6,6-dihalogeno-derivatives as well as the mono-halogeno-compounds.

The use of penicillin sulphoxides, such as (46), as a starting point for transforming penicillins into cephalosporin analogues[161] has initiated



(46)

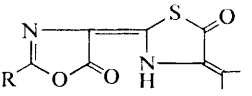


(47, R = $NHBu^t$)
(48, R = OMe)

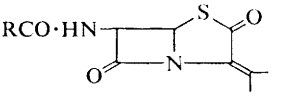
several investigations on the configuration of penicillin sulphoxides. Theoretically two sulphoxide isomers can exist, but reported[162] methods of oxidation have all yielded only one isomer. N.m.r. and *X*-ray studies[162] indicate that the isomer formed is the (*S*)-sulphoxide (46). Intramolecular hydrogen bonding between the amide proton and the sulphoxide oxygen atom and steric control have been suggested as explanations for the predominance of one isomer in the oxidation reaction. However, two independent investigations[163, 164] using n.m.r. techniques have shown that phthalimidopenicillinates (47) and (48), which contain no 6β-amide-proton, still give only one sulphoxide, the (*R*)-isomer, when oxidised with *m*-chloroperbenzoic acid. Thus the oxidation reaction must be subject to steric control. Oxidation of (47) with iodobenzene dichloride, however, gives rise[163] to the two isomeric sulphoxides and is thought to involve a two-step reaction. The sulphoxide (46) having the (*R*)-configuration has been prepared from the (*S*)-isomer under photolytic conditions.[165]

[159] D. A. Johnson and D. Mania, *Tetrahedron Letters*, 1969, 267.
[160] J. P. Clayton, *J. Chem. Soc.* (*C*), 1969, 2123.
[161] R. B. Morin, B. G. Jackson, R. A. Mueller, E. R. Lavagnino, W. B. Scanlon, and S. L. Andrews, *J. Amer. Chem. Soc.*, 1969, **91**, 1401.
[162] R. D. G. Cooper, P. V. De Marco, J. C. Cheng, and N. D. Jones, *J. Amer. Chem. Soc.*, 1969, **91**, 1408.
[163] D. H. R. Barton, F. Comer, and P. G. Sammes, *J. Amer. Chem. Soc.*, 1969, **91**, 1529.
[164] R. D. G. Cooper, P. V. De Marco, and D. O. Spry, *J. Amer. Chem. Soc.*, 1969, **91**, 1528.
[165] R. A. Archer and P. V. De Marco, *J. Amer. Chem. Soc.*, 1969, **91**, 1530.

Tautomeric forms of anhydropenicillene (49) have been proposed [166] as a structure for the product of mercuric acetate oxidation of anhydropenicillin (50). Anhydropenicillin (50, R = $PhOCH_2\cdot$) and an oxazolone



(49)



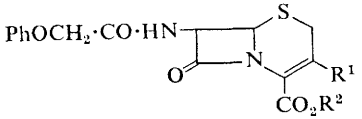
(50)

have been characterised [167] as rearrangement products of penicillin V in triethylamine–methyl chloroformate–DMF. A reverse of this anhydropenicillin rearrangement occurs [168] when the anhydropenicillin [50, R = $C_6H_4(CO)_2N\cdot$] is dissolved in aqueous dimethyl sulphoxide at pH 7·4. The reaction between penicillenic acid [169] and model compounds containing thiol groups suggests that the thiol groups in proteins might be sites for irreversible binding of penicillin to protein.

Penicillins containing δ-lactam rings,[170] 6β-phthalimidohomopenicillanic acid,[171] and α-(5-tetrazolyl)benzylpenicillin [172] have been synthesised. Synthetic penicillins derived from cycloheptatriene- and benzonorcaradiene-carboxylic acids have also been prepared.[173]

10 Cephalosporins

The last stages in the conversion of the penicillin sulphoxide (46) to cephalosporin V (51) have been successfully carried out [174] by allylic bromination (*N*-bromosuccinimide) of the derivative (52) [161] followed by treatment with potassium acetate.

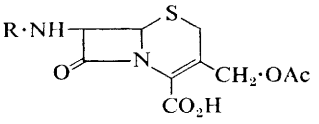


(51, $R^1 = CH_2\cdot OAc$, $R^2 = H$)
[52, $R^1 = Me$, $R^2 = CH_2\cdot C_6H_4\cdot OMe(p)$]

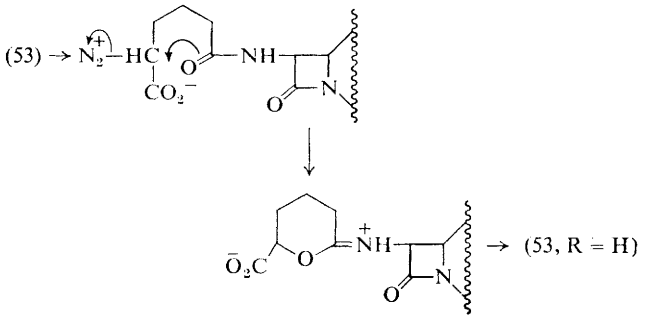
The α-aminoadipyl side-chain of cephalosporin C (53) can be removed [175] from the molecule using nitrosyl chloride in formic acid, giving 7-amino-

[166] S. Wolfe, C. Ferrari, and W. S. Lee, *Tetrahedron Letters*, 1969, 3385.
[167] S. Kukolja, R. D. G. Cooper, and R. B. Morin, *Tetrahedron Letters*, 1969, 3381.
[168] S. Wolfe, R. N. Bassett, S. M. Caldwell, and F. I. Wasson, *J. Amer. Chem. Soc.*, 1969, **91**, 7205.
[169] E. S. Wagner, W. W. Davis, and M. Gorman, *J. Medicin. Chem.*, 1969, **12**, 483.
[170] D. Todd, R. J. Cornell, R. T. Wester, and T. Wittstruck, *J. Chem. Soc.* (*C*), 1969, 408,
[171] B. G. Ramsay and R. J. Stoodley, *J. Chem. Soc.* (*C*), 1969, 1319.
[172] J. M. Essery, *J. Medicin. Chem.*, 1969, **12**, 703.
[173] S. Hanessian and G. Schutze, *J. Medicin. Chem.*, 1969, **12**, 527, 529.
[174] J. A. Webber, E. M. Van Heyningen, and R. T. Vasileff, *J. Amer. Chem. Soc.*, 1969, **91**, 5674.
[175] R. B. Morin, B. G. Jackson, E. H. Flynn, R. W. Roeske, and S. L. Andrews, *J. Amer. Chem. Soc.*, 1969, **91**, 1396.

cephalosporanic acid (53, R = H) in approximately 40% yield. The reaction is envisaged as an intramolecular interaction involving the amide carbonyl group, as shown in Scheme 3.

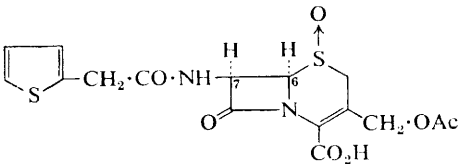


[53, R = $H_3\overset{+}{N}CH(CO_2^-)(CH_2)_3CO\cdot$]



Scheme 3

Decarboxylation of the sulphoxide (54) in basic conditions has given rise to a mixture of two isomers,[176] one of which (80% yield) was the result of



(54)

epimerisation at C-7 of the cephalosporin nucleus. A study of the epimerisation process in triethylamine using n.m.r. and deuterium-labelling techniques revealed no tendency for epimerisation at C-6 of the sulphoxides or in cephalosporins containing no 1-oxide function.

The D-phenylglycine derivative of 7-aminodesacetoxycephalosporin, cephalexin, and other derivatives have been prepared,[177] and an alternative total synthesis of deacetylcephalosporin lactone has been reported.[178]

[176] M. L. Sassiver and R. G. Shepherd, *Tetrahedron Letters*, 1969, 3993.
[177] C. W. Ryan, R. L. Simon, and E. M. Van Heyningen, *J. Medicin. Chem.*, 1969, **12**, 310.
[178] J. E. Dolfini, J. Schwartz, and F. Weisenborn, *J. Org. Chem.*, 1969, **34**, 1582.

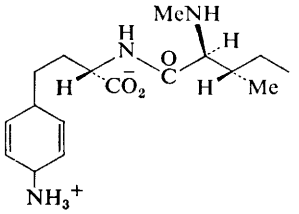
The cepham nucleus and derivatives [179] and certain oxygen analogues [180] have been prepared.

11 Peptide Analogues

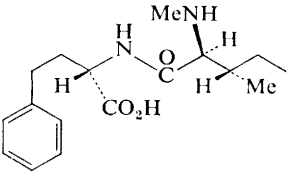
Optically active *N*-hydroxy-peptides of the general formula $R \cdot CHNH_2 \cdot CO \cdot N(OH)CH_2 \cdot CO_2H$ have been synthesised for the first time [181] by treating an acid solution of *N*-hydroxyglycine with the *N*-carboxy-anhydride of the optically active amino-acid. Full experimental details of the synthesis of novel peptides containing aza-asparagine residues (see also Volume 1 of this Report, p. 245) have been published.[182]

12 Miscellaneous

Stravidin, a novel product from *Streptomyces avidinii*, has been shown [183] to have the structure (55) on the basis of degradative evidence, n.m.r., and mass spectrometric studies. Further confirmation for the structure is derived from the synthesis of a transformation product (56) of stravidin from the two component amino-acids.



(55)



(56)

The peptides edeine A and B isolated from *Bacillus brevis* are unusual in that they contain a high percentage of unusual amino-acids. Edeine A contains one residue each of glycine, isoserine, $\alpha\beta$-diaminopropionic acid, β-tyrosine, spermidine, and a diaminocarboxylic acid (57), whose structure

$$HO_2C \cdot CHNH_2 \cdot (CH_2)_3 \cdot CHNH_2 \cdot CHOH \cdot CH_2 \cdot CO_2H$$

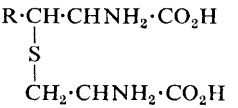
(57)

has been determined using n.m.r. and mass spectrometric techniques.[184] An extra component often found in the hydrolysates has been identified as

[179] A. K. Bose, V. Sudarshanam, B. Anjaneyulu, and M. S. Manhas, *Tetrahedron*, 1969, **25**, 1191.
[180] J. C. Sheehan and M. Dadic, *J. Heterocyclic Chem.*, 1968, **8**, 779.
[181] G. Zvilichovsky and L. Heller, *Tetrahedron Letters*, 1969, 1159.
[182] H. Niedrich, *Chem. Ber.*, 1969, **102**, 1557.
[183] K. H. Baggaley, B. Blessington, C. P. Falshaw, W. D. Ollis, L. Chaiet, and F. J. Wolf, *Chem. Comm.*, 1969, 101.
[184] T. P. Hettinger and L. C. Craig, *Biochemistry*, 1968, **7**, 4147; T. P. Hettinger, Z. Kurylo-Borowska, and L. C. Craig, *ibid.*, p. 4153.

the acrylic acid formed on dehydration of (57). Edeine B differs from edeine A in that *N*-guanyl-*N'*-(3-aminopropyl)-1,4-diaminobutane replaces *N*-(3-aminopropyl)-1,4-diaminobutane as the basic component.

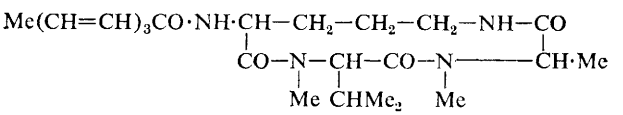
Both subtilin and nisin antibiotics contain an $\alpha\beta$-dehydroalanyllysine residue in the *C*-terminal positions,[185] and therefore can be included in the increasing list of antibiotics which contain $\alpha\beta$-unsaturated amino-acids.[4] It has been suggested [186] that the $\alpha\beta$-unsaturated amino-acids (derived from serine and threonine) in nisin are intermediates in the biosynthesis of lanthionine (58) and β-methyllanthionine (59). Labelling studies [186] show



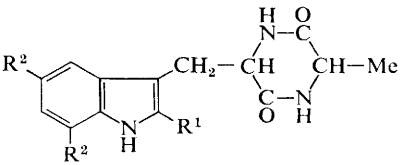
(58, R = H)
(59, R = Me)

that cysteine, serine, and threonine are incorporated into the amino-acids (58) and (59).

A new toxic peptide aspochracin, from *Aspergillus ochraceus* has been assigned [187] the structure (60) on the basis of physical data and the hydrolysis of hexahydroaspochracin to *N*-methyl-L-valine, L-ornithine, *N*-methyl-L-alanine, and caprylic acid. The hexahydroaspochracin has also



(60)



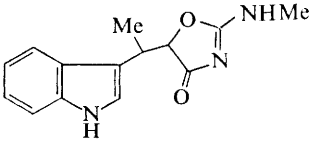
(61, R^1 = ·CMe_2·CH:CH_2, R^2 = ·CH_2·CH:CMe_2)

been synthesised in low yield from the linear peptide analogue. A comparison of the o.r.d. spectrum of echinulin (61) with the spectra of a

[185] E. Gross, J. L. Morell, and L. C. Craig, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 952.
[186] L. C. Ingram, *Biochim. Biophys. Acta*, 1969, **184**, 216.
[187] R. Myokei, A. Sakurai, C.-F. Chang, Y. Kodaira, N. Takahashi, and S. Tamura, *Tetrahedron Letters*, 1969, 695.

series of optically pure *cyclo*-(Ala-Trp) has shown [188] that the configuration of echinulin must be similar to *cyclo*-(L-Ala-L-Trp) and not to *cyclo*-(L-Ala-D-Trp) as previously suggested.

A specific biosynthetic incorporation of a *C*-methyl group from methionine into the tryptophyl side-chain of indolmycin (62) has been



(62)

proved by labelling studies.[189] Two tripeptides, Glu-*O*-phosphoseryl-Leu and Glu-*O*-phosphoseryl-Tyr have been isolated [190] from embryonic bovine enamel. The fluorescence quenching obtained with a series of model flavinyl-aromatic amino-acids [191] could well account for similar changes observed in the binding of flavinyl coenzymes to apoenzymes.

[188] R. Nakashima and G. P. Slater, *Canad. J. Chem.*, 1969, **47**, 2069.
[189] U. Hornemann, L. H. Hurley, M. K. Speedie, H. F. Guenther, and H. G. Floss, *Chem. Comm.*, 1969, 245.
[190] J. Seyer and M. J. Glimcher, *Biochim. Biophys. Acta*, 1969, **181**, 410.
[191] R. E. MacKenzie, W. Föry, and D. B. McCormick, *Biochemistry*, 1969, **8**, 1839.

5

Extracts from the Tentative Rules of the I.U.P.A.C.–I.U.B. Commission on Biochemical Nomenclature

I Tentative Abbreviated Designation of Amino-acid Derivatives and Peptides

(*Reprinted by permission from* I.U.P.A.C. Information Bulletin No. 26)

These Rules are an attempt to achieve a broad systematization of various types of abbreviated notation already in use. They seek 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 of conveying more detailed chemical information in abbreviated form.

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 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^{β} and S^{β} of serine and cysteine, respectively, N^{ε} of lysine, N^{α} 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 in a more clear manner, the hyphen

between the configurational prefix and the symbol may be omitted. (See 5.3.1.1 *et seq.*) (Note: The designation of an amino-acid residue as DL is inappropriate for compounds having another amino-acid residue with an asymmetric centre.)

1.6 Structural formulae of complicated features may be used along with the abbreviated notation wherever necessary for clarity.

2 Abbreviations for Amino-acids

2.1 *Common amino-acids*

alanine	Ala	lysine	Lys
arginine	Arg	methionine	Met
asparagine*	Asn*	ornithine	Orn
aspartic acid	Asp	phenylalanine	Phe
cysteine	Cys	proline	Pro
glutamic acid	Glu	serine	Ser
glutamine*	Gln*	threonine	Thr
glycine	Gly	tryptophan	Trp
histidine	His	tyrosine	Tyr
isoleucine	Ile	valine	Val
leucine	Leu		

* Asparagine and glutamine may also be denoted as Asp(NH_2) or Asp, and Glu(NH_2) or Glu, respectively.

2.2 *Less common amino-acids*

Abbreviations 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 *Hydroxy-amino-acids*

hydroxylysine	Hyl
3-hydroxyproline	3Hyp
4-hydroxyproline	4Hyp

2.2.2 Allo-*amino-acids*

allo-isoleucine	*a*Ile
allo-hydroxylysine	*a*Hyl

2.2.3 '*Nor*' *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.

norvaline	Nva
norleucine	Nle

2.2.4 *Higher unbranched amino-acids*

We suggest the following general rules for guidance in forming abbreviations: The functional prefix 'amino' should be included in the symbol as the letter 'A', diamino as 'D'.
The trivial name of the parent acid should be abbreviated to leave no more than 2 or 3 letters, as convenient and necessary for clarity. The word 'acid' ('-säure', *etc.*) should be omitted from the symbol as carrying no significant information. Unless otherwise indicated (see paragraph below), single groups are in the α position, two amino-groups in the α, ω (monocarboxylic acids) or α, α′ positions (dicarboxylic acids). The location of amino-acids in positions other than α and ω is shown by the appropriate Greek letter prefix.
Examples:

α-aminobutyric acid	Abu
α-aminoadipic acid	Aad
α-aminopimelic acid	Apm
α, γ-diaminobutyric acid	Dbu
α, β-diaminopropionic acid	Dpr
α, α′-diaminopimelic acid	Dpm
β-alanine	βAla
ε-aminocaproic acid	εAcp
β-aminoadipic acid	βAad

2.2.5 *N^{α}-alkylated amino-acids*[p]

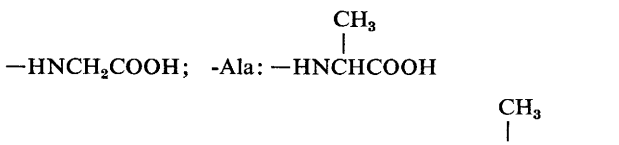
N^{α}-methyl amino-acids are becoming more and more common (*e.g.*, in the large group of depsipeptides). This justifies special symbols:
Examples:

N-methylglycine (sarcosine)	MeGly or Sar
N-methylisoleucine	MeIle
N-methylvaline, *etc.*	MeVal, *etc.*
N-ethylglycine, *etc.*	EtGly, *etc.*

3 Amino-acid Residues

3.1 *Lack of hydrogen on the α-amino-group*

The α-amino-group is understood to be at the left-hand side of the symbol when using hyphens, and—in special cases—at the point of the arrow when using arrows to indicate the direction of the peptide bond (—CO → NH—, —NH ← CO—). *E.g.*,

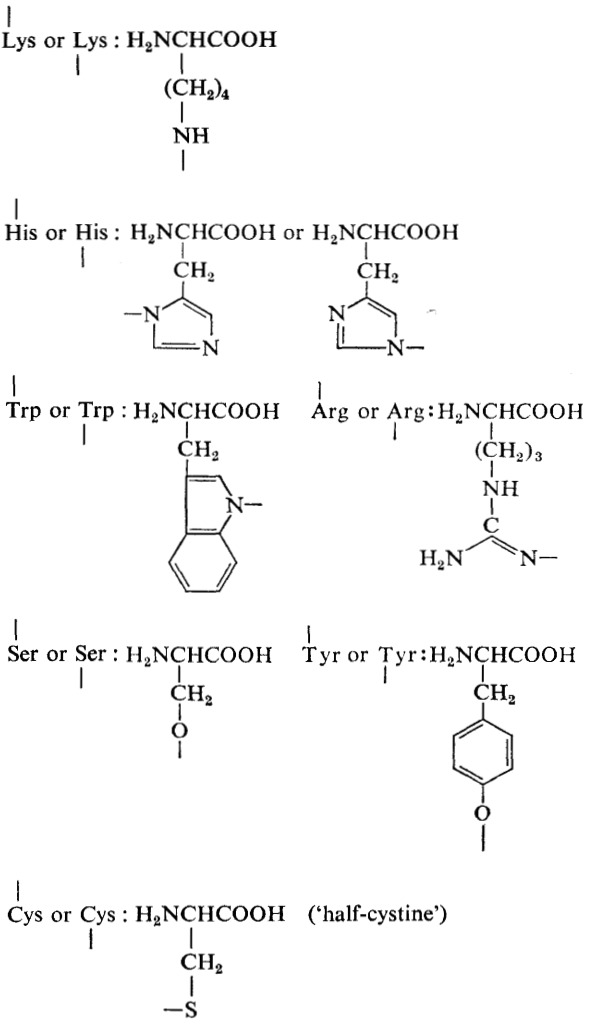


3.2 *Lack of hydroxy on the α-carboxy-group*

The α-carboxy-group is always understood to be on the right-hand side of the symbol when using hyphens, and—in such special cases as 5.3.1.3—at the tail of the arrow when using arrows to indicate the direction of the peptide bond (—CO → NH, —NH ← CO—).

E.g., Gly- : H_2NCH_2CO—

3.3 *Lack of hydrogen on amino, imino, guanidino, hydroxy, and thiol functions in the side chain*



[Cystine would be: Cys–Cys, Cys–Cys or Cys–Cys, *not* Cys-Cys.]

3.4 *Lack of hydroxy on carboxy-groups in the side chain*

Asp or Asp : $H_2NCH(CH_2CO)COOH$; Glu or Glu : $H_2NCH(CH_2CH_2CO)COOH$

4 Substituted Amino-acids

4.1 *Substitution in the α-amino- and α-carboxy-groups*

This follows logically from 3.1 and 3.2. The following examples will make the usage clear.

N-acetylglycine	Ac-Gly
glycine ethyl ester	Gly-OEt
N^{α}-acetyllysine	Ac-Lys
serine methyl ester	Ser-OMe
α-ethyl *N*-acetylglutamate	Ac-Glu-OEt
isoglutamine	Glu-NH_2
N-ethyl-*N*-methylglycine	EtMeGly, $\begin{matrix}\text{Et}\\\text{Me}\end{matrix}$>Gly, $\begin{matrix}\text{Me}\\\text{Et}\end{matrix}$—Gly

4.2 *Substitution in the side chain*

Side chain substituents may be portrayed above or below the amino-acid symbol, 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. Where a more complex substituent is involved, it is recommended that the vertical stroke and a two-line abbreviation be used.

aspartic acid β-methyl ester	$\begin{matrix}\text{OMe}\\\text{Asp or Asp or Asp(OMe)}\\\text{OMe}\end{matrix}$
N^{ε}-acetyllysine	$\begin{matrix}\text{Ac}\\\text{Lys or Lys or Lys(Ac)}\\\text{Ac}\end{matrix}$

O-acetylserine : Ser or Ser or Ser(Ac) (Ac)

O-methyltyrosine : Tyr or Tyr or Tyr(Me) (Me)

S-ethylcysteine : Cys or Cys or Cys(Et) (Et)

5 Polypeptides

5.1 *Polypeptide chains*

Polypeptides may be dealt with in the same manner as substituted amino-acids, *e.g.*,

glycylglycine Gly-Gly

α-glutamylglycine Glu-Gly

γ-glutamylglycine Glu(—Gly) or Glu(—Gly)

glutathione Glu(—Cys-Gly) or Glu(γ—Cys-Gly)

(Note that Glu(—Cys-Gly) would represent the corresponding thiolester with a bond between the γ-carboxy of glutamic acid and the thiol group of cysteine.)

N^{ε}-α-glutamyllysine Glu(—Lys)

N^{ε}-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-Gly-OH

its di-hydrochloride $^{+}H_2$-Gly-Lys-Gly-OH, 2 Cl^{-} ($^{+}H_2$)

its sodium salt Gly-Lys-Gly-O^-Na^+

its N^ε-formyl derivative Gly-Lys(CHO)-Gly or Gly-Lys(CHO)-Gly

etc.

5.2 *Peptides substituted at* N^α

Examples:

glycylnitrosoglycine Gly(NO)-Gly

glycylsarcosine Gly(Me)-Gly or Gly-MeGly or Gly-Sar

N^α-glycyl-N^α-acetylglycine Gly(Ac)-Gly

N^α-glycyl-N^α-glycylglycine Gly(Gly-)-Gly

etc.

5.3 *Cyclic polypeptides*

5.3.1 *Homodetic Cyclic Polypeptides* (the ring consists of amino-acid residues in peptide linkage only).

Three representations are possible:

5.3.1.1 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-)

5.3.1.2 The terminal residues may be written on one line, as in 5.3.1.1, but joined by a lengthened bond. Using the same example in the two forms (see 1.5):

—Val-Orn-Leu-D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro—

or

—Val-Orn-Leu-DPhe-Pro-Val-Orn-Leu-DPhe-Pro—

5.3.1.3 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):

→Val → Orn → Leu → DPhe → Pro—
—Pro ← DPhe ← Leu ← Orn ← Val←

5.3.2 *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.

Example:

oxytocin Cys-Tyr-Ile-Asn-Gln-Cys-Pro-Leu-Gly-NH_2
etc.
cyclic ester of threonyl-glycyl-glycyl-glycine

Thr-Gly-Gly-Gly— or

H-Thr-Gly-Gly-Gly—

etc.

6 Abbreviations for Substituents

Groups substituted for hydrogen or for hydroxy may be indicated either by their structural formulae or by (accepted) abbreviations, *e.g.*,
benzoylglycine (hippuric acid):


—CO-Gly or C_6H_5CO-Gly or Bz-Gly[1]

glycine methyl ester:

$Gly\text{-}OCH_3$ or Gly-OMe

Suggestions for the abbreviations of protecting groups common in polypeptide chemistry follow. All such symbols (except those allowed by individual journals, *e.g.*, Bz, Ac, Ph, Me, Et, *etc.*) *should be defined in each paper.* Although symbolization by the use of capital letters throughout would be useful for distinguishing these symbols from those of the amino-acids, we propose the use of one capital letter followed by lower-case letters in order not to increase the flood of capital-letter abbreviations in biological chemistry.

6.1 *N-protecting groups of the urethane type*

benzyloxycarbonyl	Z-
p-nitrobenzyloxycarbonyl	$Z(NO_2)$-
p-bromobenzyloxycarbonyl	Z(Br)-
p-methoxybenzyloxycarbonyl	Z(OMe)-
p-methoxyphenylazobenzyloxycarbonyl	Mz-
p-phenylazobenzyloxycarbonyl	Pz-
t-butyloxycarbonyl	Boc-
cyclopentyloxycarbonyl	Poc-

[1] Bz- is the symbol generally used for *benzoyl* in organic chemistry. Its use for *benzyl* (which has become rather common in polypeptide chemistry) should be discouraged. We propose *Bzl-* for *benzyl.*

6.2 *Other N-protecting groups*

acetyl	Ac-
benzoyl	Bz-
tosyl	Tos-
trifluoroacetyl	Tfa-
phthalyl	Pht-
benzyl	Bzl-
trityl	Trt-
tetrahydropyranyl	Thp-
dinitrophenyl	Dnp-
benzylthiomethyl	Btm-
o-nitrophenylsulfenyl	Nps-

6.3 *Carboxy-protecting groups*

methoxy (methyl ester)	-OMe
ethoxy (ethyl ester)	-OEt
tertiary butoxy (t-butyl ester)	-OBu^t
benzyloxy (benzyl ester)	-OBzl
diphenylmethoxy (benzhydryl ester)	-OBzh
p-nitrophenoxy (*p*-nitrophenyl ester)	-ONp
phenylthio (phenyl thiolester)	-SPh
p-nitrophenylthio	-SNp
cyanomethoxy	-OCH$_2$CN

Note: Contrary to the symbols for amino-acid residues, the position of the hyphens in the symbols for substituents carries no significant information.

II Tentative Rules for Naming Synthetic Modifications of Natural Peptides

(*Reprinted by permission from* I.U.P.A.C. Information Bulletin No. 27)

Introduction

During the last few years, chemists have made many compounds that are variants of naturally-occurring peptides (or proteins) having trivial names. Therefore, the need has arisen for 'semi-trivial' names to designate these variants without the necessity of designating every residue in the chain.

These proposals are *not* suitable for application to 'abnormal' links in a peptide sequence—*e.g.* to disulphide links or γ-peptide links. They are *only* suitable for modifications involving normal α-peptide links.

Rules

1. *Replacement*

In a polypeptide of trivial name X, if the qth amino-acid residue (starting from the *N*-terminal end of the chain) is *replaced* by the amino-acid residue Abc, the semi-trivial name of the modified polypeptide is [q-amino-acid]-X and the abbreviated form, chiefly for use in tables, is [Abc^q]-X.

Examples

[8-Citrulline]-vasopressin; [Cit^{8}]-vasopressin
(BODANSZKY and BIRKHIMER, *J. Amer. Chem. Soc.*, 1962, **84**, 4963).
[5-Isoleucine,7-alanine]-hypertensin II;
[Ile^{5},Ala^{7}]-hypertensin II;
(SEU, SMEBY, and BUMPUS, *J. Amer. Chem. Soc.*, 1962, **84**, 3883).

Comments

(a) In the full name, the replacement amino-acid is designated by its its *own full name*, *not* the name of its radical (*cf.* 4 below). This name, and the position of replacement, are given in square brackets [ ], as for isotopic replacement.

(b) In the abbreviated form, the amino-acid residues are designated by the standard three-letter symbols (Section I), the first letter *only* being a capital, in square brackets [ ].

(c) In the abbreviated form, the *position* of substitution is indicated in a special fashion, *i.e.*, by a superior numeral ^{q}, to indicate that it is a *residue*, not an individual atom, that is being replaced—and also for the reason indicated in comment (d).

(d) The nature of the residue replaced is *not* designated in either the full or the abbreviated name. This is contrary to a general principle of organic nomenclature requiring that an atom (or group) that is replaced, should (unless it is hydrogen) be clearly designated, as in 2-amino-2-deoxy-D-glucose. It has been decided *not* to insist on the designation of the residue replaced in these semi-trivial names in order to keep the names as short as possible, and because the form of nomenclature in Rule 1 clearly differs from ordinary substitution nomenclature.

(e) A partial analogy may be drawn with the form used for isotopic replacement, where the isotope symbol is indicated in square brackets before the name.

(f) The replacement of one amino-acid residue by its enantiomer may be shown logically by the application of this rule as follows: the replacement in X of L-alanine at position 7 by D-alanine results in [7-D-alanine]-X with the abbreviation [D-Ala^{7}]-X. An example may be found in BOISSONNAS, GUTTMAN, and PLESS, *Experientia*, 1966, **22**, 526, dealing with the D-Ser^{1} ... derivative of β-Corticotropin; the natural compound has L-serine in position 1. Another example is the [α-D-Asp^{1}]-hypertensin II of RINIKER and SCHWYZER (*Helv. Chim. Acta*, 1964, **47**, 2357).

2. *Extension*

The compounds obtained by the extension of polypeptide X at either (a) the *N*-terminal end or (b) the *C*-terminal end are designated by the kinds of names and abbreviations shown below; these are in accordance with the general principles of polypeptide nomenclature (Section I).

Examples

(a) Extension at *N*-terminal end:

	Aminoacyl-X	Abc-X
e.g.	Valyl-X	Val-X
or	Valylglycyl-X	Val-Gly-X (for extension by two residues)

(b) Extension at *C*-terminal end:

	X-yl-amino-acid	X-yl-Abc
e.g.	X-yl-leucine	X-yl-Leu

(where X-yl is the trivial name of polypeptide X with the ending -yl).

Comment

This rule is not applicable to the extension at the *C*-terminal of natural peptides having a terminal α-carboxamido group, as in the case of oxytocin or α-MSH. It has been suggested that new names be given to the peptides having a free terminal α-carboxy-group (*e.g.* oxytocinoic acid) and that extension at the *C*-terminal end be denoted as in the example given above (*e.g.* oxytocinoyl-Abc).

3. *Insertion*

The compound obtained by the *insertion* of an additional amino-acid residue Abc in the position between the qth and (q + 1)th residues of a polypeptide X is named endo-qa-amino-acid-X (abbreviated form, endo-Abc^{qa}-X).

Example

Endo-4a-tyrosine-hypertensin II: Endo-Tyr^{4a}-hypertensin II.

Comments

(a) This form has analogies in other fields where '*endo*' implies the insertion of something into a structure (*e.g.*, *endo*-methylene). The prefix or index qa is based on analogies with the steroids where the atoms inserted in a ring after atom no. q are designated qa, qb, *etc.*

(b) The prefix homo is *not* suitable for designating the insertion of a whole residue, since it is commonly used to modify the names of *individual* amino-acids, *e.g.* homoserine.

(c) Multiple insertions, and insertion of two or more residues together in the same place in the chain, are shown by a logical extension of this rule. For example the insertion into the polypeptide X of threonine between residues 4 and 5, and of valine and glycine (*in that order*) between residues 6 and 7, is shown by the name 'Endo-4a-threonine,6a-valine,6b-glycine-X' and the abbreviation 'Endo-Thr^{4a}, (Val^{6a}-Gly^{6b})-X'.

4. *Removal*

The compound obtained by the formal *removal* of an amino-acid residue from a polypeptide X in position q, is designated by the name des-q-amino-acid-X, abbreviated des-Abc^{q}-X.

Example

des-7-Proline-oxytocin; des-Pro^{7}-oxytocin (JACQUENOUD and BOISSONNAS, *Helv. Chim. Acta*, 1962, **45**, 1462).

Comment

(a) Removal of a whole residue is indicated as is the removal of a ring in steroids, *e.g.*, des-A-androstane.

(b) 'de' is *not* suitable as a prefix because it is easily confused, in speaking, with D (for configuration).

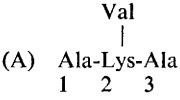
5. *Substitution forming a side-chain*

The compound formed by the substitution of an additional amino-acid residue as a side-chain into a polypeptide X is named by applying the ordinary rules of nomenclature to the trivial name.

(a) If the substitution is on a side-chain *amino*-group of polypeptide X, the name of the additional amino-acid *residue* is written (with the termination '-yl') and prefixed by symbols indicating the position of substitution (residue number and atom).

Example

An imaginary compound (A)

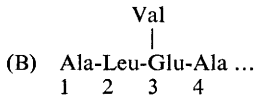


in which a valyl group is substituted at the ε-amino-group of lysine at position 2 of the chain of a peptide X is named $N^{\varepsilon 2}$-valyl-X (abbreviated $N^{\varepsilon 2}$-Val-X).

(b) If the substitution is on a side-chain *carboxy*-group of polypeptide X, the additional amino-acid having a free α-carboxy-group, the substituted derivative is named by specifying the position of substitution (residue number, and atom) and is given the designation 'X-yl-amino-acid'.

Example

An imaginary compound (B)



in which a valine residue is substituted into the δ-carboxy-group of glutamic acid in position 3 of the chain of a peptide X would be named $C^{\delta 3}$-X-yl-valine (abbreviated $C^{\delta 3}$-X-yl-Val).

Comment

Note the importance of clear distinction from *replacement* as indicated in Rule 1.

6. *Partial sequences (fragments)*

Polypeptide sequences that form fragments of a longer sequence that already has a trivial name may be designated as follows. The *trivial name* is followed by numbers giving the positions of the first and last amino-acids, and then the usual *Greek* designation giving the number of amino-acid units in the fragment: thus

Trivial name (-X-Y-)... peptide.

Example

From α-MSH (α-melanophore-stimulating hormone)

Ac-Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val-NH_2
1 2 3 4 5 6 7 8 9 10 11 12 13
α-MSH

we may have

Met-Glu-His-Phe-Arg-Trp-Gly — α-MSH-(4-10)-heptapeptide
4 10

and

His-Phe-Arg-Lys-Pro-Val-NH_2 — α-MSH-(6-8)-(11-13)-hexapeptide amide
6 8 11 13

The last example illustrates the nomenclature for a composite sequence of two fragments, and also for an amide terminal group.

Summary with Examples

The systematic application of these principles to the name of an imaginary pentapeptide 'Iupaciubin' may illustrate the symbolism.

Rule	*Operation*	*Short name*	*Structure*
	(Fundamental name)	Iupaciubin	1 2 3 4 5 Ala-Lys-Glu-Tyr-Leu
1.	Replacement	[Phe^4]-iupaciubin*	4 Ala-Lys-Glu-Phe-Leu
2(a)	Extension, (*N*-terminal)	Arginyl-iupaciubin, Arg-iupaciubin	1 5 Arg-Ala-Lys-Glu-Tyr-Leu
2(b)	Extension, (*C*-terminal)	Iupaciubyl-methionine, Iupaciubyl-Met	1 5 Ala-Lys-Glu-Tyr-Leu-Met
3.	Insertion	Endo-Thr^{2a}-iupaciubin	2 2a 3 Ala-Lys-Thr-Glu-Tyr-Leu

* Note that only for *replacement* are square brackets required.

Rule	*Operation*	*Short name*	*Structure*
4.	Removal	Des-Glu^{3}-iupaciubin	2 4 Ala-Lys-Tyr-Leu
5(a)	Side-chain substitution on amino-group	$N^{\varepsilon 2}$-Val-iupaciubin	Val |$^{\varepsilon}$ Ala-Lys-Glu-Tyr-Leu 2
5(b)	Side-chain subsitution on carboxy-group	C-$^{\delta 3}$-Iupaciubyl-valine	Val |$^{\delta}$ Ala-Lys-Glu-Tyr-Leu 3
6.	Partial sequence	Iupaciubin-(2-4)-tri-peptide	2 3 4 Lys-Glu-Tyr

III Tentative Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino-acids)

(*Reprinted by permission from* I.U.P.A.C. Information Bulletin No. 30)

The numerous studies on the physical, chemical, and biological properties of synthetic polypeptides have brought with them different ways of describing, in abbreviated form, these products, whose structures are often incompletely known. The use of a variety of nomenclatures complicates the literature; hence, a consistent and clearly defined system for naming such polypeptides is desirable. The proposals set forth here, which represent the consensus of many discussions and suggestions, should aid in systematizing the nomenclature of a wide variety of synthetic polypeptides.

The abbreviations of the amino-acid residues and their derivatives or modifications are those indicated in the Tentative Rules (Section I) or formulated according to the principles set out in them. Hyphens or commas between the symbols for residues or groups of residues mean known or unknown sequence, respectively, and involve only the α-NH_2 and α-COOH groups (the peptide link). [Commas may be omitted when other symbols (*e.g.*, subscripts or superscripts) separate symbols in unknown sequences.] Vertical strokes indicate covalent bonds involving functional groups or the remaining H-atom of the peptide bond, depending upon their placement (Section I) L-Amino-acids and α-peptide links, read from left (NH_2-terminus) to right (COOH-terminus), are assumed unless indicated otherwise (Section I).

Definitions

1. Linear polymer: all amino-acids are in an unbranched chain.
2. Graft polymer: one or more polymeric segments are linked to the functional groups of a linear chain, thus creating a branch or branches. (Functional groups include ε-NH_2, β- or γ-COOH, *etc.*, and the remaining H-atom of an α-peptide link.)

3. Block polymer: two or more linear or graft polymeric segments are linked to form a larger polymer.
4. Polymeric segment: a polymer that forms a distinct part of a larger polymer (*e.g.*, a block or graft polymer may contain several polymeric segments).

Rules

1. *Polymeric segments* that contain more than one amino-acid symbol are enclosed in parentheses or brackets. A superscript outside of the parentheses indicates the number of repeating sequences per 100 residues of polymer, and it is given to the first decimal place.
2. *The molar percentage* of a single type of amino-acid residue within a copolymer, derived from the amino-acid analysis and assuming copolymerization, is indicated by a superscript attached to the symbol of the residue. The molar percentages are given in whole numbers and should total 99—101%.
3. *Designation of polymeric segments or linear polymers.* The prefix 'poly' or the subscript '*n*' indicates 'polymer of'. It is attached to each main chain and is repeated in each polymeric segment within a larger polymer unless there is sufficient indication of size and of structure to make this repetition unnecessary. For example, poly Glu or $(Glu)_n$ for polyglutamic acid, and $(Glu)_{10}$ for a decapeptide of glutamic acid.

   Thus *n* is chosen in place of 'p' in order to avoid confusion with the 'p' used for a phosphoric acid residue in the latter scheme. The '*n*' may be replaced by a definite number (*e.g.*, 10 above), an average (*e.g.*, 10), or a range (*e.g.*, 8—12), as appropriate.
4. *Designation of branches and branch points.* Branches (side chains) connected to the main chain can be designated in one of three ways: by a vertical line joining the main chain and the branch (side chain); by an extended bond joining the appropriate residues with the main chain written first; or by a horizontal double dash (not preferred). *The branch points* are indicated by the origin and terminus of the vertical line. If the origin is *unknown*, the line originates at the 'p' in 'poly', if 'poly' is used, or at the first parenthesis (bracket), if the subscript '*n*' is used (see Rule 1). If the origin is *known*, the line originates (a) vertically at the initial letter of the appropriate symbol, if functional groups other than α-NH_2 or α-COOH residues are involved; (b) vertically at the position of the appropriate link, if substitution for the remaining H-atom of a peptide link is involved; or (c) horizontally to the left or right of the symbol, respectively, if α-NH_2 or α-COOH groups are involved. The same rules apply to the termination of the line. Thus, the linkage between a side chain functional group and an α-NH_2 or α-COOH group in the main chain is indicated by two perpendicular lines with the vertical line originating in the functional group and the horizontal line in the α-NH_2

or α-COOH group. A number in parentheses lying beside the line indicates the number of such links per 100 residues of polymer, if known.

Comment: A limitation of the double dash as a connecting link lies in its inability to originate or to terminate definitively in a specific residue. Either the arrangement of the symbols must be such that connected ones are adjacent, or the information must be given independently.

Examples

1. Simple homopolymer:

   poly Ala or $(Ala)_n$

2. Linear copolymer, random sequence, composition unknown:

   poly DLAla,Lys or $(DLAla,Lys)_n$

3. Linear copolymer, alternating sequence, composition unknown:

   poly DLAla-Lys or $(DLAla\text{-}Lys)_n$

4. Linear sequence of unknown order [composition: 56% Glu, 38% Lys, and 6% Tyr (Σ = 100%)]:

(a) poly $Glu^{56}Lys^{38}Tyr^{6}$ or $(Glu^{56}Lys^{38}Tyr^{6})_n$ (all L)

(b) poly $DGlu^{56}DLys^{38}Tyr^{6}$ (only Tyr is L)

(c) poly $DLGlu^{56}Lys^{38}DTyr^{6}$ (Glu is DL, Tyr is D)

5. Block polymer of poly Glu combined through the α-COOH terminus to the α-NH_2 terminus of poly Lys [composition: 56% Glu, 44% Lys (Σ = 100%)]:

   poly Glu^{56}-poly Lys^{44} or $(Glu^{56})_n$-$(Lys^{44})_n$

6. (a) Known, repeating sequence (a polymer of Glu-Lys-Lys-Tyr):

   poly Glu-Lys_2-Tyr or $(Glu\text{-}Lys_2\text{-}Tyr)_n$

   (b) Known, repeating sequences within each of two constituent blocks of a linear polymer [composition: 37·5% Glu, 25% Lys, 25% Tyr, 12·5% Ala (Σ = 100%)]:

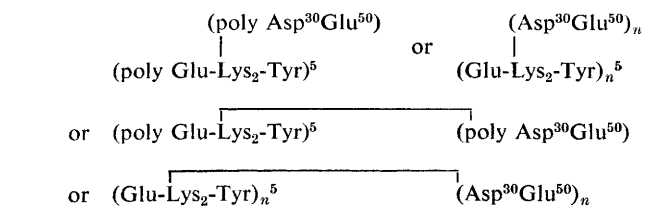
   $(poly\ Glu\text{-}Lys)^{25}$-$(poly\ Ala\text{-}Tyr_2\text{-}Glu)^{12{\cdot}5}$

   or $(Glu\text{-}Lys)_n{}^{25}$-$(Ala\text{-}Tyr_2\text{-}Glu)_n{}^{12{\cdot}5}$

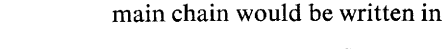
   [The connection between the polymeric segments shown here is from the α-COOH of Lys to the α-NH_2 of Ala. Origin or termination in any other residue or functional group can be shown by rearranging the order of residues and by the orientation of the connecting line at its origin and terminus (see Examples 7, 8, 9).]

   (c) Known, repeating sequence in the main chain connected by the ε-NH_2 of a lysine (which of the two is not known) to an unknown

point in an unknown sequence in the side chain [composition: 30% Asp, 55% Glu, 10% Lys, 5% Tyr (Σ = 100%)]:

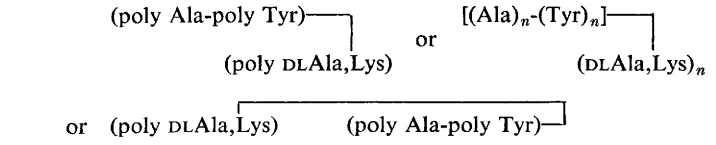


(Note: The double hyphen system is not applicable here.)
If the lysine residue connected to the side chain were known, the main chain would be written in the form, *e.g.*,



as appropriate.

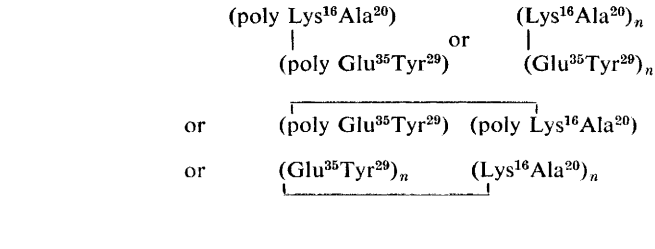
7. Graft polymer with the main chain of DL-alanine and L-lysine connected through the ε-NH_2 group of lysine to the α-COOH group of L-tyrosine in the side chain, which consists of a block polymer of L-tyrosine and L-alanine (no analytical data for the main chain):



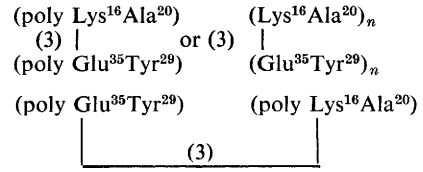
(Note: The points of attachment of Lys and Tyr cannot be specified in the last example.)

8. Graft polymer with an unknown sequence in the main chain and in the side chain [composition: 16% Lys, 20% Ala, 35% Glu, 29% Tyr (Σ = 100%)]:

(a) Number and position of the points of attachment in the main chain unknown, but terminating in the lysine residues of the side chain:

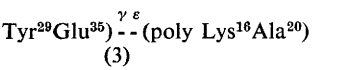


(b) Same, but attachments are 3 in number and connect the ε-NH_2 groups of the lysine residues in the side chain and the γ-COOH groups of the glutamic acid residues in the main chain:

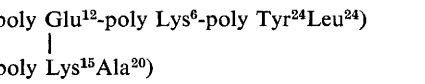


or

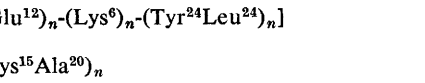
or



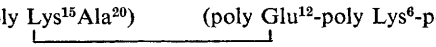
9. Graft polymer with a block polymer and an unknown sequence in the side chain (upper) attached to an unknown sequence in the main chain (lower); the points of attachment are between the γ-COOH groups of glutamic acid in the side chain and the ε-NH_2 groups of lysine in the main chain [composition: Glu 12%, Lys 21%, Tyr 24%, Leu 24%, Ala 20% (Σ = 101%)]:



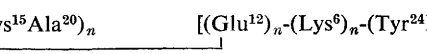
or



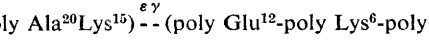
or



or



or



Author Index