Amino-acids, Peptides, and Proteins—Volume 1

Specialist Periodical Reports

The Chemical Society

SPECIALIST PERIODICAL REPORTS

At the end of 1968, The Chemical Society introduced the first two titles in this series of comprehensive, in depth accounts by experts, of progress in well defined though limited areas of Chemistry. These were:—

1 **Carbohydrate Chemistry**—Volume 1. 282 pages.
(*Senior Reporter: R. D. Guthrie*)
Price £3.10.0 (Fellows price 40/–)

2 **Spectroscopic Properties of Inorganic and Organometallic Compounds**—Volume 1. 372 pages.
(*Senior Reporter: N. N. Greenwood*)
Price £5.0.0 (Fellows price 70/–)

The second volumes of the first two titles have now been published as under:—

1 **Carbohydrate Chemistry**—Volume 2. 298 pages.
(*Senior Reporter: R. D. Guthrie*)
Price £3.10.0 (Fellows price 40/–)

Published in September 1969

2 **Spectroscopic Properties of Inorganic and Organometallic Compounds**—Volume 2. 512 pages.
(*Senior Reporter: N. N. Greenwood*)
Price £5.0.0 (Fellows price 70/–)

Published in September 1969

In addition to the present title, the following further volumes and titles will become available as indicated below:—

1 **Photochemistry**—Volume 1.
(*Senior Reporter: D. Bryce-Smith*)
To be published in February 1970

2 **Organophosphorus Chemistry**—Volume 1.
(*Senior Reporter: S. Trippett*)
To be published in March 1970

3. **Electrochemistry**—Volume 1.
(*Senior Reporter: G. J. Hills*)
To be published in the Spring 1970

This volume: SBN 85186 004 4

A Specialist Periodical Report

Amino-acids, Peptides, and Proteins

Volume 1

A Review of the Literature Published during 1968

Senior Reporter

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

Reporters

C. C. F. Blake, *University of Oxford*
J. S. Davies, *University College of Swansea, University of Wales*
R. D. Gillard, *University of Kent*
P. M. Hardy, *University of Exeter*
J. H. Jones, *University of Oxford*
S. Laurie, *University of Kent*
R. N. Perham, *Cambridge University*
D. G. Smyth, *National Institute for Medical Research, London*

SBN: 85186 004 4

The Chemical Society
Burlington House, London, W1V 0BN

Preface

The introduction of Specialist Periodical Reports, giving comprehensive accounts of progress in specialised, limited areas of chemistry, is an important development in the publications policy of the Chemical Society. These volumes are designed to assist the research worker in his own field, and their appearance allows the long-established *Annual Reports on the Progress of Chemistry* to return to their original aim, to keep the non-specialist informed of the highly significant advances in the major fields of pure chemistry.

I was asked by the Reports and Reviews Committee of the Society to invite experts to collaborate in the production of a Report on Amino-acids, Peptides, and Proteins. It seemed to me that it would be unreasonable to ask active workers to extend their survey much beyond the area of their special interest, and this volume is therefore the result of the co-operation of eight authors. I must make it clear that my contribution goes no further than the overall editorial supervision, and I record here my gratitude to each author for his willing help. I believe that those working or intending to work in this field will be grateful also.

The scope of the Report is essentially limited to the chemistry of amino-acids, peptides, and proteins, but Chapter 5 discusses some aspects of the relationship between the structure and biological activity of selected peptides and proteins; the *Annual Reports* for 1968 include a section on Enzyme Mechanisms. The intention has been to review all relevant papers listed in *Current Chemical Papers* during 1968, together with those appearing in the main journals that year even if not so listed, but the late arrival in this country of some December issues of U.S. journals prevented their inclusion in this survey. Some papers published early in 1969 have been noted. The delay which would be involved in forming a subject index was considered to be unacceptable, and it is hoped that the extended list of contents will enable readers to find material without difficulty; an Author Index is included. It should be noted that reference numbers apply to the Chapter in which they occur, or, if the Chapter is divided into Parts, to the Part concerned.

We have adopted the abbreviated designations of amino-acid derivatives and polypeptides recommended in *I.U.P.A.C. Information Bulletin* No. 25 (1966). There are still examples in the literature of formulations, particularly of branched-chain and cyclic polypeptides and depsipeptides, which are ambiguous because these Rules are not followed; for

example, Glu must be understood to mean *O*-(γ-glutamyl)-threonine, and
|
Thr
N-(α-glutamyl)-threonine is formulated as Glu—Thr. Rearrangement of structures to give the correct alignment may at times be awkward, but an ambiguous formulation is still more undesirable, and we appeal to authors and editors to adopt the conventions which have been agreed internationally.

In a survey of this kind, overlap between sections is bound to occur and, within limits, is desirable, for the convenience of the reader. It was felt that it would be helpful in certain sections to give in this first issue a brief introductory account of earlier work; this will not of course be appropriate in subsequent years. It is intended that these Reports should appear annually, and the second volume will survey literature published in 1969, together with any important papers which escaped our search this time; we shall be grateful if our attention is drawn to such omissions. Authors can assist us here, as well as themselves, by ensuring that the titles of their papers indicate the contents as clearly as possible. Some modification of treatment will be made in the next volume, and we shall welcome comments and criticisms, but we hope that the reader will find the arrangement of the material convenient on the whole.

Among the important advances recorded in this volume are the three-dimensional structure of yet another enzyme, papain, the complete amino-acid sequence of several other enzymes, and the chemical synthesis of material having the enzymic activity of ribonuclease. The increasing rate of progress is likely to make future volumes still more difficult to confine within a reasonable size, but one could hardly ask for a more exciting time at which to review the field.

G. T. YOUNG

Contents

Chapter 3 Peptide Synthesis

By J. H. Jones

Abbreviations

Abbreviations for amino-acids and their use in the formulation of derivatives follow the recommendations in *I.U.P.A.C. Information Bulletin* No. 25 (Butterworths, 1966), and reprinted in *J. Biol. Chem.*, 1966, **241**, 2491, and in *Biochim. Biophys. Acta*, 1966, **121**, 1.

Other abbreviations which have been used without definition are:

Adoc	adamantyloxycarbonyl
ADP	adenosine 5′-diphosphate
AMP	adenosine 5′-phosphate
Aoc	amyloxycarbonyl
ATP	adenosine 5′-triphosphate
Asx	aspartic acid *or* asparagine (not yet determined)
c.d.	circular dichroism
Cm	carboxymethyl
DNA	deoxyribonucleic acid
DNP	2,4-dinitrophenyl
DNS	1-dimethylaminonaphthalene-5-sulphonyl (dansyl)
DP	degree of polymerisation
e.s.r.	electron spin resonance
g.l.c.	gas–liquid chromatography
Glp or Pca	pyrrolid-2-one-5-carboxylic acid
Glx	glutamic acid *or* glutamine (not yet determined)
i.r.	infrared
n.m.r.	nuclear magnetic resonance
o.r.d.	optical rotatory dispersion
PTH-Gly	the phenylthiohydantoin derived from glycine, *etc.*
RNA	ribonucleic acid
t.l.c.	thin-layer chromatography
u.v.	ultraviolet

Note. Amino-acids are assumed to be of the L-configuration unless stated otherwise.

1
Amino-acids

BY J. H. JONES

In one way or another, the whole of this Report is about the chemistry of amino-acid derivatives, but here we are concerned only with the amino-acids themselves. Emphasis is on α-amino-acids, and it is only for these that comprehensive coverage has been attempted. Furthermore, although many may think it retrogressive to draw demarcation lines between disciplines in these enlightened times, amino-acid biochemistry has been excluded. Amino-acid chemistry is under continuous scrutiny from every conceivable angle, and so papers on these compounds span a very diverse range of interests. The Reporter therefore makes no apology for the fact that this chapter is something of a miscellany.

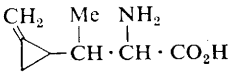
1 Naturally Occurring Amino-acids

A. Occurrence of Known Amino-acids.—Pipecolic acid (of unstated configuration, but presumably L) has been isolated from azuki beans (*Phaseolus angularis*) in a yield of about 0·05%.[1] *N*-Methyl-L-alanine, which has not been found in higher plants before, has been isolated from the leaves of *Dichapetalum cymosum*.[2] This plant is toxic and causes considerable cattle loss in southern Africa because it produces fluoroacetate. Young leaves contain remarkably large amounts of *N*-methyl-L-alanine (up to 5·6% by weight on a dry-weight basis) and it was suggested that the metabolism of the amino-acid and the toxin might be linked. *Erythro*-γ-methyl-L-glutamic acid has been isolated from seeds of *Lathyrus maritimus* and distinguished from the other possible methylglutamic acids by nuclear magnetic resonance (n.m.r.) spectrometry: chromatographic evidence indicates that this amino-acid also occurs in other *Lathyrus* species.[3] N^{ϵ}-Trimethyl-L-lysine has been isolated from seeds of *Reseda luteola*[4] and from chicken erythrocyte histones, where it occurs together with N^{ϵ}-methyl- and N^{ϵ}-dimethyl-L-lysine.[5,6] 6-Hydroxykynurenic acid has been obtained from tobacco leaves,[7] thus providing the first example of the occurrence of a kynurenic

[1] S. Hatanaka, *Sci. Papers Coll. Gen. Educ., Univ. Tokyo*, 1968, **17**, 219.
[2] J. N. Eloff and N. Grobbelaar, *J. S. African. Chem. Inst.*, 1967, **20**, 190.
[3] J. Przybylska and F. M. Strong, *Phytochemistry*, 1968, **7**, 471.
[4] P. O. Larsen, *Acta Chem. Scand.*, 1968, **22**, 1369.
[5] K. Hempel, H. W. Lange, and L. Birkhofer, *Z. Physiol. Chem.*, 1968, **349**, 603.
[6] K. Hempel, H. W. Lange, and L. Birkhofer, *Naturwiss.*, 1968, **55**, 37.
[7] P. K. Macnicol, *Biochem. J.*, 1968, **107**, 473.

acid derivative in plants. A survey of the distribution of fifteen non-protein amino-acids in about forty species of the genus *Acacia* shows that members of the Gummiferae series can be distinguished by their amino-acid content: they alone contain *N*-acetyldjenkolic acid, and they lack several amino-acids which are widely distributed among other *Acacia* species.[8] The possibility that the toxicity of various species of *Crotalaria* may not be due solely to the presence of pyrrolizidine alkaloids in these plants has been demonstrated by the isolation of the neurotoxin α-amino-β-oxalylaminopropionic acid from seeds of *C. incana* and *C. mucronata*.[9] *Trans*-3-hydroxy-L-proline has been isolated from seeds and vegetative tissue of *Delonix regia*, where it is a major component of the free amino-acid pool.[10]

B. New Naturally Occurring Amino-acids.—A number of new natural amino-acids have been characterised during the year: these are listed at the end of this section, together with their sources. Those whose structure has been confirmed by synthesis are included in the list of newly synthesised amino-acids in section 2. The increasing usefulness of physical methods for the characterisation of amino-acids is apparent: n.m.r. and mass spectrometry (applications of mass spectrometry in amino-acid and peptide chemistry have been reviewed[11]) have been particularly valuable. Especially noteworthy is the characterisation of the novel amino-acid (1) by spectroscopic methods.[12]



(1)

C. A List of New Naturally Occurring Amino-acids.—

Amino-acid	*Source*	*Ref.*
L-*N*-(3-amino-3-carboxypropyl)-β-carboxypyridinium betaine. (Nicotianine)	Tobacco leaves	13, 14
L-α-amino-ε-amidinocaproic acid. (Indospicine)	*Indigofera spicata*	15
β-(4-hydroxybenzothiazol-6-yl)alanine	Gallopheomelanins from chicken feathers	16
β-(2-methyl-4-hydroxybenzothiazol-6-yl)-alanine	Gallopheomelanins from chicken feathers	16
mixed disulphide of β-mercaptolactic acid and cysteine	Urine of mentally defective patient	17

[8] A. S. Seneviratne and L. Fowden, *Phytochemistry*, 1968, **7**, 1039.
[9] E. A. Bell, *Nature*, 1968, **218**, 197.
[10] M. L. Sung and L. Fowden, *Phytochemistry*, 1968, **7**, 2061.
[11] J. H. Jones, *Quart. Rev.*, 1968, **22**, 302.
[12] D. S. Millington and R. C. Sheppard, *Phytochemistry*, 1968, **7**, 1027.
[13] M. Noguchi, H. Sakuma, and E. Tamaki, *Arch. Biochem. Biophys.*, 1968, **125**, 1017.
[14] M. Noguchi, H. Sakuma, and E. Tamaki, *Phytochemistry*, 1968, **7**, 1861.
[15] M. P. Hegarty and A. W. Pound, *Nature*, 1968, **217**, 354.
[16] L. Minale, E. Fattorusso, G. Cimino, S. de Stefano, and R. A. Nicolaus, *Gazzetta*, 1967, **97**, 1636.
[17] M. Ampola, E. M. Bixby, J. C. Crawhall, M. L. Efron, R. Parker, W. Sneddon, and E. P. Young, *Biochem. J.*, 1968, **107**, 16P.

Amino-acid	*Source*	*Ref.*
N^{ϵ}-(indole-3-acetyl)-L-lysine	*Pseudomonas savastanoi*	18
p-hydroxymethyl-L-phenylalanine	*Escherichia coli*	19
O-ethyl-L-homoserine	*Corynebacterium ethanolaminophilum*	20
O-n-propyl-L-homoserine	*Corynebacterium ethanolaminophilum*	20
O-n-butyl-L-homoserine	*Corynebacterium ethanolaminophilum*	20
L-5-methyl-2-aminohex-4-enoic acid	*Leucocortinarius bulbiger*	21
2-amino-4-methylhex-4-enoic acid	*Aesculus californicus*	22
2-amino-4-methylhexanoic acid	*Aesculus californicus*	22
2-amino-4-methyl-6-hydroxyhex-4-enoic acid	*Aesculus californicus*	12, 22
β-(methylenecyclopropyl)-β-methylalanine	*Aesculus californicus*	12, 22
β-acetamido-L-alanine	*Acacia armata*	8
α-(3-hydroxyphenyl)glycine	*Euphorbia helioscopa*	23
α-(3,5-dihydroxyphenyl)glycine	*Euphorbia helioscopa*	23
1-methyl-6-hydroxy-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid	*Euphorbia myrsinites*	24
L-*cis*(?)-2-amino-3-hydroxymethylpent-3-enoic acid	*Bankera fulgineoalba*	25
L-*cis*(?)-2-amino-3-formylpent-3-enoic acid	*Bankera fulgineoalba*	25

2 Chemical Synthesis and Resolution of Amino-acids

The majority of new syntheses reported this year 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 in the following outline, the remainder being merely mentioned or incorporated into the appendix to this section (see p. 13).

A. Protein Amino-acids.—New syntheses have been reported for DL-lysine,[26] DL-histidine,[27] DL-cystine,[28] and DL-tryptophan,[29] and reactions for the conversion of serine to DL-cystine,[30] L-tyrosine to L-phenylalanine,[31] and L-ornithine to L-proline [32] have been described. The resolution of racemic

[18] O. Hutzinger and T. Kosuge, *Biochemistry*, 1968, 7, 601.
[19] N. H. Sloane and S. C. Smith, *Biochim. Biophys. Acta*, 1968, **158**, 394.
[20] Y. Murooka and T. Harada, *Agric. and Biol. Chem. (Japan)*, 1967, **31**, 1035.
[21] G. Dardenne, J. Casimar, and J. Jadot, *Phytochemistry*, 1968, 7, 1401.
[22] L. Fowden and A. Smith, *Phytochemistry*, 1968, 7, 809.
[23] P. Müller and H. R. Schütte, *Z. Naturforsch.*, 1968, **23b**, 659.
[24] P. Müller and H. R. Schütte, *Z. Naturforsch.*, 1968, **23b**, 491.
[25] R. R. Doyle and B. Levenberg, *Biochemistry*, 1968, 7, 2457.
[26] S. Motoki, S. Satsumabayashi, and F. Minemura, *J. Org. Chem.*, 1968, **33**, 3667.
[27] J. Fernandez-Bolanos and D. Martinez-Ruiz, *Anales real Soc. espaň. Fis. Quim.*, 1968, **64**, 423.
[28] P. Rambacher, *Chem. Ber.*, 1968, **101**, 3433.
[29] N. I. Aboskalova, A. S. Polyanskala, and V. V. Perekalin, *Doklady Akad. Nauk S.S.S.R.*, 1967, **176**, 829.
[30] P. Rambacher, *Chem. Ber.*, 1968, **101**, 2595.
[31] T. Kishi, Y. Kato, and M. Tanaka, *J. Agric. Chem. Soc. Japan*, 1968, **42**, 238.
[32] S. Ohshiro, K. Kuroda, and T. Fujita, *J. Pharm. Soc. Japan.*, 1967, **87**, 1184.

glutamic acid [33-39] and alanine [40] by preferential crystallisation has been studied in detail by Japanese workers.

B. Other α-Amino-acids.—A very brief preliminary description of a new general route to α-amino-acids has appeared:[41] treatment of an *NN*-bis(trimethylsilyl)glycine ester (2) with base followed by reaction with an alkyl halide results in alkylation of the α-carbon atom giving (3) which is very easily hydrolysed to an α-amino-ester with dilute acid.

$$[(Me)_3Si]_2N \cdot CH_2 \cdot CO_2R^1 \qquad [(Me)_3Si]_2N \cdot \underset{\displaystyle R^2}{\underset{|}{CH}} \cdot CO_2R^1$$

(2) (3)

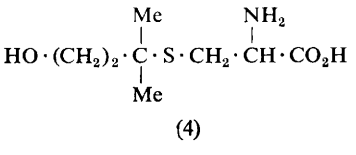
A convenient synthesis of DL-α-aminosuberic acid (by the diethyl acetamidomalonate route) has been reported:[42] resolution can be achieved enzymically or, preferably, by the method of Vogler *et al.*[43] using optically active tyrosine hydrazide. An improved method for the conversion of tyrosine (D and L) to 3,5-dichlorotyrosine has been described.[44] Synthesis of α,α′-diaminopimelic acid by the method of Work *et al.*[45] gives an equimolecular mixture of racemic and *meso* diaminodiacids: a simpler method for separating the isomers and resolving the racemate has been published.[46] DL-α-Acetamido-β-methylaminopropionic acid has been obtained by an improved procedure.[47] Stereospecific enzymic deacylation gave L-α-amino-β-methylaminopropionic acid which had the same specific rotation as a sample of the same amino-acid isolated in 1967 from *Cycas circinalis*, thus confirming the L configuration of the latter.

L-Felinine (4) can be prepared by acid-catalysed *S*-alkylation of L-cysteine with 2-methylbut-1-ene-4-ol or 2-methylbut-2-ene-4-ol.[48]

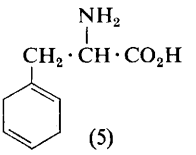
Birch reduction of phenylalanine gives 3,6-dihydrophenylalanine (5), which can be partially hydrogenated to 3,4,5,6-tetrahydrophenylalanine (6). The position of the double bond in (6) was confirmed by isolation of the

[33] T. Watanabe and G. Noyori, *J. Chem. Soc. Japan, Ind. Chem. Sect.*, 1967, **70**, 2164.
[34] T. Watanabe and G. Noyori, *J. Chem. Soc. Japan, Ind. Chem. Sect.*, 1967, **70**, 2167.
[35] T. Watanabe, H. Kurokawa, T. Koga, Y. Kawauchi, and G. Noyori, *J. Chem. Soc. Japan, Ind. Chem. Sect.*, 1967, **70**, 2170.
[36] T. Watanabe and G. Noyori, *J. Chem. Soc. Japan, Ind. Chem. Sect.*, 1967, **70**, 2174.
[37] T. Watanabe and G. Noyori, *J. Chem. Soc. Japan, Ind. Chem. Sect.*, 1968, **71**, 676.
[38] N. Mizoguchi, *J. Agric. Chem. Soc. Japan.* 1967, **41**, 607.
[39] N. Mizoguchi, *J. Agric. Chem. Soc. Japan*, 1967, **41**, 616.
[40] I. Chibata, S. Yamada, M. Yamamoto, and M. Wada, *Experientia*, 1968, **24**, 638.
[41] K. Rühlmann and G. Kuhrt, *Angew. Chem. Internat. Edn.*, 1968, **7**, 809.
[42] S. Hase, R. Kivoi, and S. Sakakibara, *Bull. Chem. Soc. Japan*, 1968, **41**, 1266.
[43] K. Vogler and P. Lanz, *Helv. Chim. Acta*, 1966, **49**, 1348.
[44] K. R. Brody and R. P. Spencer, *J. Org. Chem.*, 1968, **33**, 1665.
[45] E. Work, S. M. Birnbaum, M. Winitz, R. J. Koegel, and J. P. Greenstein, *J. Amer Chem. Soc.*, 1957, **79**, 648.
[46] J. van Heijenoort and E. Bricas, *Bull. Soc. chim. France*, 1968, 2828.
[47] A. Vega, E. A. Bell, and P. B. Nunn, *Phytochemistry*, 1968, **7**, 1885.
[48] A. Schoeberl, J. Borchers, and D. Hantzsch, *Chem. Ber.*, 1968, **101**, 373.

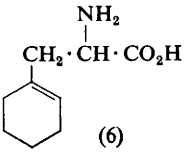
lactone (7) after treatment with hydrochloric acid, and hydrolysis of (7) gave the new amino-acid (8).[49]



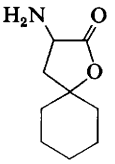
(4)



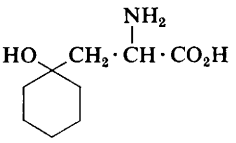
(5)



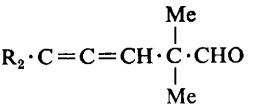
(6)



(7)

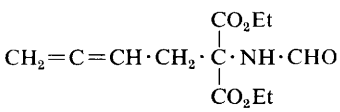


(8)

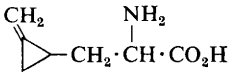


(9)

Allenic aldehydes such as (9) which are fully substituted at the 2-position give good yields in the Strecker synthesis, but poor yields of allenic amino-acids are obtained if this position is unsubstituted.[50a] The reaction of allenic bromides with diethyl formamidomalonate is a more general route to such amino-acids:[50b] the reaction is applicable to 1-bromoalka-1,2-, -2,3-, and -3,4-dienes. This reaction was used for the preparation of (10), which was treated with di-iodomethane and a zinc–copper couple followed



(10)



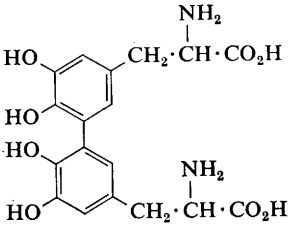
(11)

by hydrolysis and decarboxylation to yield (±)hypoglycin A (11), a naturally occurring hypoglycaemic amino-acid.[51] The final decarboxylation in this synthesis was stereoselective, and only one of the two racemates of (11) was obtained: this proved to be the required (±)amino-acid. The stereoselectivity was attributed to thermodynamic control, and models showed that the most stable racemate should consist of the (2*S* : 4*S*) and (2*R* : 4*R*) diastereoisomers. Since the susceptibility of natural hypoglycin A to enzymic oxidation had already established the configuration at the α-carbon as (*S*) it was predicted that the natural amino-acid would be found to have the absolute stereochemistry (2*S* : 4*S*), and this was confirmed by chemical correlation (see p. 15). The formamidomalonate route is also suitable for the preparation of aminoenynoic acids: thus alkylation of

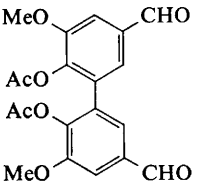
[49] M. L. Snow, C. Lauinger, and C. Ressler, *J. Org. Chem.*, 1968, **33**, 1774.
[50a] D. K. Black and S. R. Landor, *J. Chem. Soc.* (*C*), 1968, 281. [b] D. K. Black and S. R. Landor, *J. Chem. Soc.* (*C*), 1968, 283.
[51] D. K. Black and S. R. Landor, *J. Chem. Soc.* (*C*), 1968, 288.

diethyl formamidomalonate with 5-bromopent-3-en-1-yne followed by hydrolysis and decarboxylation gave 2-aminohept-4-en-6-ynoic acid in good yield.[50b]

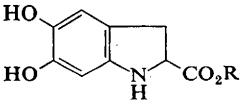
The diaminodiacid (12), which is one of the possible products of oxidative dimerization of 3,5-dihydroxyphenylalanine (DOPA), has been synthesised by the oxazolone route from the dialdehyde (13).[52] This compound ('DOPA dimer') was shown to polymerise in the presence of tyrosinase and oxygen at the same rate as DOPA itself, which is consistent with the possibility of (12) as an intermediate in the melanogenesis of DOPA.



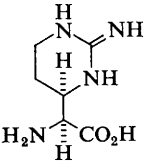
(12)



(13)



(14) R = Me
(15) R = H



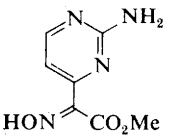
(16)

Oxidative cyclisation of DOPA methyl ester with ferricyanide followed by dithionite reduction gives (14) which can be converted to 'cycloDOPA' (15) by anaerobic hydrolysis after acetylation.[53]

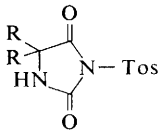
DL-Capreomycidine (16), a guanidino-amino-acid obtained from acid hydrolysates of antibiotics of the capreomycin group, has been synthesised by catalytic reduction of the oxime (17) followed by saponification and separation of the mixture of diastereoisomers.[54]

C. α-Dialkyl-α-amino-acids.—The hydantoin route is in frequent use for the synthesis of α-dialkyl-α-amino-acids but difficulties due to the resistance of 5,5-disubstituted hydantoins to hydrolysis are sometimes encountered. It has been reported [55] that this difficulty can be circumvented by conversion of such hydantoins to their 3-tosyl derivatives (18). Alkaline hydrolysis (dilute sodium hydroxide) of (18) gives hydantoic acid derivatives (19) which

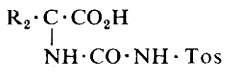
[52] Y. Omote, Y. Fujinuma, and N. Sugiyama, *Chem. Comm.*, 1968, 190.
[53] H. Wyler and J. Chiovini, *Helv. Chim. Acta.*, 1968, **51**, 1476.
[54] B. W. Bycroft, D. Cameron, L. R. Croft, and A. W. Johnson, *Chem. Comm.*, 1968, 1301.
[55] K. Hiroi, K. Achiwa, and S. Yamada, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 444.



(17)



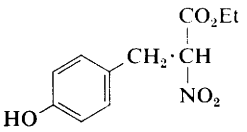
(18)



(19)

can be hydrolysed further to amino-acids with dilute hydrochloric acid. Although developed primarily for α-dialkyl-α-amino-acids, this method of hydrolysing hydantoins under mild conditions may find application in the synthesis of amino-acids which would not survive the usual vigorous hydrolysis conditions.

A new synthesis of DL-α-methyltyrosine has been described:[56] *NN*-dimethyl-*p*-hydroxybenzylamine reacts with ethyl α-nitropropionate in the presence of a catalytic amount of sodium hydride *via* a quinone methide intermediate to give (20), which can be reduced and hydrolysed to DL-α-



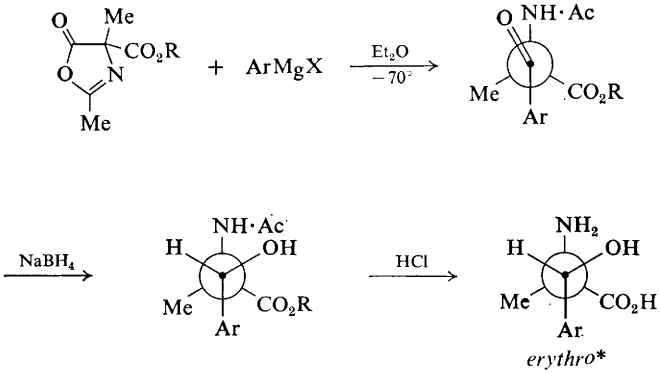
(20)

methyltyrosine. The resolution of α-methyl-α-amino-acids is a wasteful process if only one of the isomers is required, because the lack of an α-hydrogen atom prohibits racemisation and recycling of the isomer which is not required. This difficulty is avoided if resolution is performed at an earlier stage in the synthesis, and an example of this approach is provided by a new route to L-α-methylDOPA (α-methyl-3,4,-dihydroxy-L-phenylalanine).[57] The synthesis was performed by the Strecker method, with resolution at the aminonitrile stage: as the aminonitrile is formed in an equilibrium reaction, it is easily racemised, and it was therefore possible to recycle the D-aminonitrile.

The preparation of some α-methylserines is discussed in the next section.

D. Amino-acids with Aliphatic Hydroxyl Groups in the Side Chain.—A new synthesis[58] of β-aryl-α-methylserines is shown in Scheme 1. The synthesis comprises reaction of an aryl Grignard reagent with a 4-carboalkoxy-2,4-dimethyloxazol-5-one at low temperatures, followed by reduction with sodium borohydride and hydrolysis. This gives a mixture of diastereoisomers in which the *erythro* form predominates. The interconversion of the *erythro* and *threo* series is easily accomplished with thionyl chloride.[59]

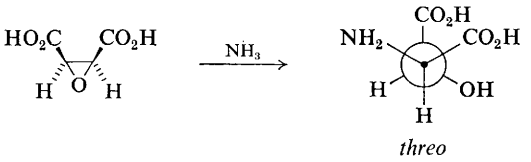
[56] W. S. Saari, *J. Org. Chem.*, 1967, **32**, 4074.
[57] D. F. Reinhold, R. A. Firestone, W. A. Gaines, J. M. Chemerda, and M. Sletzinger, *J. Org. Chem.*, 1968, **33**, 1209.
[58] S. H. Pines, S. Karady, and M. Sletzinger, *J. Org. Chem.*, 1968, **33**, 1758.
[59] S. H. Pines, S. Karady, M. A. Kozlowski, and M. Sletzinger, *J. Org. Chem.*, 1968, **33**, 1762.



* Some *threo* amino-acid is also formed.

Scheme 1

An attempt [60] to synthesise γ-hydroxyarginine by a multistage alternative to the direct preparation [61] from γ-hydroxyornithine has so far met with little success. Ammonolysis of *cis*-epoxysuccinic acid gives exclusively *threo*-β-hydroxy-DL-aspartic acid [62] (Scheme 2) but the *trans* epoxide was



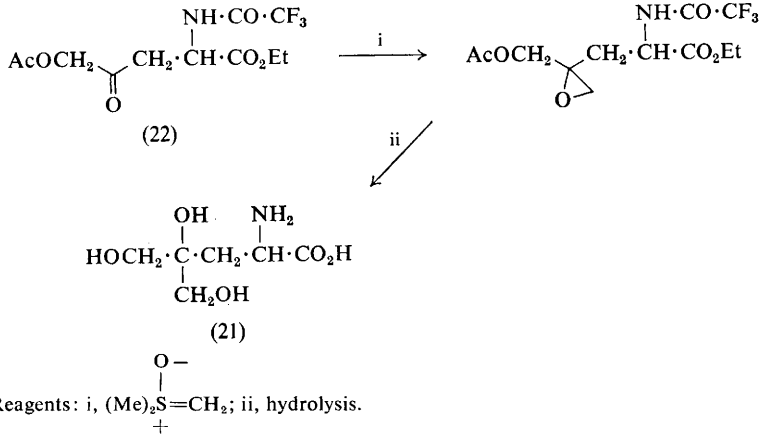
Scheme 2

found to give the *threo* and *erythro* isomers in about 1 : 2 proportions.[62] This last finding is surprising in view of the fact that the reaction of benzylamine with *trans*-epoxysuccinic acid is stereospecific,[63, 64] but it was admitted that the *trans*-epoxide used in the ammonolysis studies may have contained some *cis* isomer. *N*-Benzyl-*threo*-β-hydroxy-DL-aspartic acid is very conveniently (and almost quantitatively) resolved with ephedrine,[65] and hydrogenolysis after resolution gave L- and D-hydroxy-amino-acids of specific rotation slightly greater than previously reported. Optically pure *erythro*-β-hydroxy-D-aspartic acid has been obtained from optically pure *trans*-L-epoxysuccinic acid.[64] *Threo*- and *erythro*-β-hydroxy-DL-isoleucine

[60] B. A. Santa Rossa and T. Viswanatha, *Canad. J. Biochem.*, 1968, **46**, 725.
[61] Y. Fujita, *Bull. Chem. Soc. Japan.*, 1959, **32**, 439.
[62] H. Okai, N. Imamura, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1967, **40**, 2154.
[63] Y. Liwschitz, Y. Rabinsohn, and A. Haber, *J. Chem. Soc.*, 1962, 3589.
[64] J. Oh-Hashi and K. Harada, *Bull. Chem. Soc. Japan*, 1967, **40**, 2977.
[65] Y. Liwschitz, Y. Edlitz-Pfeffermann, and A. Singermann, *J. Chem. Soc.* (*C*), 1967, 2104.

have been synthesised by stereospecific routes from *cis*- and *trans*-3-methylpent-2-enoic acids respectively: the diastereoisomers can be distinguished by n.m.r.[66]

The recent isolation[67] of γ,δ,δ'-trihydroxyleucine (21) has prompted the synthesis of the L-isomer[68] by epoxidation of (22) (which had earlier been synthesised[69] starting with L-asparagine), followed by hydrolysis (Scheme 3).



(22)

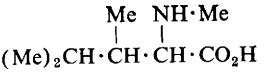
(21)

Reagents: i, $(Me)_2S{=}CH_2$; ii, hydrolysis.

Scheme 3

Dihydroxyprolines are discussed in the next section.

E. *N*-Substituted Amino-acids.—N^{α}-Methyl-L-histidine can be prepared[70] from L-histidine by a modification of the procedure of Quitt *et al.*[71] (formation of the N^{α}-benzyl derivative by reduction of the Schiff base, N^{α}-methylation by means of a Leuckart reaction, and finally removal of the N^{α}-benzyl group by hydrogenolysis). $N^{\alpha}N^{\alpha}$-Dimethyl-L-histidine has been prepared[70] for the first time by subjection of L-histidine to catalytic hydrogenation conditions in the presence of formic acid. A new synthesis[72] of *N*,β-dimethylleucine (23) has made use of the *N*-methylation procedure

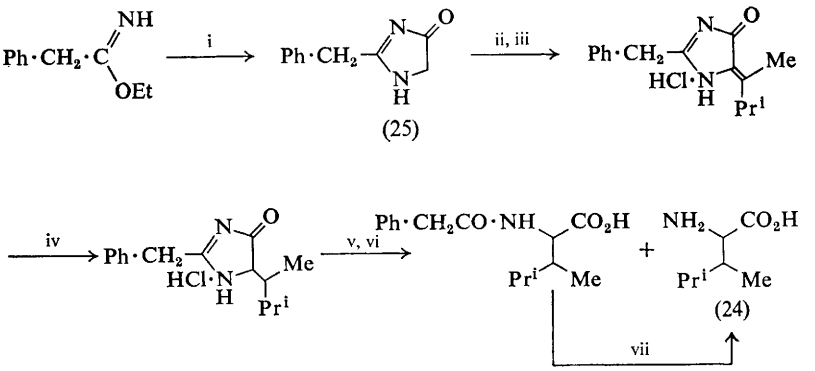


(23)

of Quitt *et al.*[71] The synthesis of β-methylleucine (24) from the imidazolone

[66] T. A. Dobson and L. C. Vining, *Canad. J. Chem.*, 1968, **46**, 3007.
[67] U. Gebert, T. Wieland, and H. Boehringer, *Annalen*, 1967, **705**, 227.
[68] F. Weygand and F. Mayer, *Chem. Ber.*, 1968, **101**, 2065.
[69] F. Weygand, P. Klinke, and I. Eigen, *Chem. Ber.*, 1957, **90**, 1896.
[70] V. N. Reinhold, Y. Ishikawa, and D. B. Melville, *J. Medicin. Chem.*, 1968, **11**, 258.
[71] P. Quitt, J. Hellerbach, and K. Vogler, *Helv. Chim. Acta*, 1963, **46**, 327.
[72] H. Kotake, T. Saito, and K. Okubo, *Tetrahedron Letters*, 1968, **24**, 2015.

(25) (which is readily available [73]) shown in Scheme 4 gave both diastereoisomers, which were separated by fractional crystallisation and *N*-methylated.

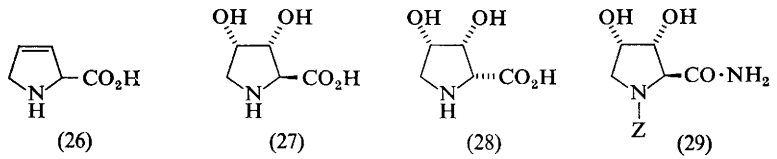


Reagents: i, Gly·OEt; ii, Me·CO·Pr^i; iii, HCl—EtOH; iv, H_2—PtO_2; v, $Ba(OH)_2$; vi, Et_3N; vii, HCl

Scheme 4

These two diastereoisomers of (23) were identified as *N*,γ-dimethylisoleucine and *N*,γ-dimethylalloisoleucine by spectroscopic examination, and the latter was resolved by way of its benzyloxycarbonyl derivative with ephedrine.[72]

Treatment of 3,4-dehydro-DL-proline (26) with alkaline permanganate gives an equimolecular mixture of 2,3-*trans*-3,4-*cis*-3,4-dihydroxy-DL-proline (27) and the 2,3-*cis*-isomer (28).[74] Separation of these isomers was



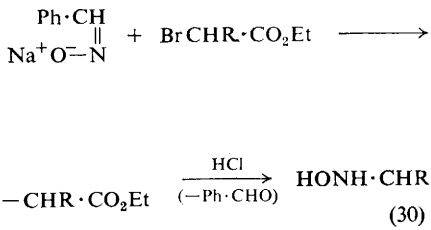
(26) (27) (28) (29)

difficult but was achieved by fractional crystallisation of their copper salts, and the two new amino-acids were characterised. When *N*-substituted 3,4-dehydroprolines are treated with osmium tetroxide, however, glycolation occurs exclusively from the less hindered side: *e.g.* benzyloxycarbonyl-3,4-dehydroprolineamide gave a single glycol (29), which could be converted to (27) by hydrogenation and hydrolysis, also in good yield.

N-Hydroxy-α-amino-acids (30) can be synthesised by a number of routes which were summarised in 1967,[75] but the most general route is *via*

[73] H. Lehr, S. Karlan, and M. W. Goldberg, *J. Amer. Chem. Soc.*, 1953, **75**, 3640.
[74] C. B. Hudson, A. V. Robertson, and W. R. J. Simpson, *Austral. J. Chem.*, 1968, **21**, 769.
[75] E. Buehler and G. B. Brown, *J. Org. Chem.*, 1967, **32**, 265.

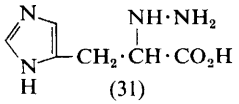
N-alkylation of the *anti*-benzaldoxime anion with an α-bromoester followed by hydrolysis of the resulting nitrone,[75, 76] as shown in Scheme 5. The



Scheme 5

syn-benzaldoxime anion undergoes alkylation on oxygen.[75] Hydrazine can also be used for cleavage of the nitrone.[76] This general route has been used for the synthesis of N^{α}-hydroxy-DL-asparagine,[77] *N*-hydroxy-[α-^{14}C]-DL-phenylalanine,[78] and also some *N*-hydroxy-β-amino-acids.[79]

The L, D, and DL forms of the *N*-amino derivative (31) of histidine have



(31)

been prepared by reaction of hydrazine with the chloroacids obtained from D, L, and DL-histidine respectively by deamination in concentrated hydrochloric acid.[80]

F. β-Amino-acids.—A large number of *N*-benzoyl-β-amino-acids have been prepared by means of a Ritter reaction (treatment with acetonitrile and sulphuric acid) starting with 3-hydroxypropanoic esters.[81] Aminopivalic acid is produced in poor yield when hydroxypivalic acid is catalytically hydrogenated in aqueous ammonia at high temperature and pressure.[82]

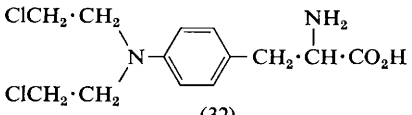
G. Labelled Amino-acids.—A short route to [1-^{14}C]-DL-arginine involving a Strecker synthesis with γ-guanidinobutyraldehyde has been reported: the synthesis uses radioactive intermediates only in the last two stages.[83] Alternative syntheses of [γ-^{14}C]-DL-aspartic and [δ-^{14}C]-DL-glutamic acids [84] have been claimed to be more convenient and less expensive than

[76] E. Bellasio, F. Parravicini, T. La Noce, and E. Testa, *Ann. Chim.* (*Italy*), 1968, **58**, 407.
[77] E. Falco and G. B. Brown, *J. Medicin. Chem.*, 1968, **11**, 142.
[78] H. Kindl and E. W. Underhill, *Phytochemistry*, 1968, **7**, 745.
[79] E. Ballasio, F. Parravicini, A. Vigevani, and E. Testa, *Gazzetta*, 1968, **98**, 1014.
[80] M. Sletzinger, R. A. Firestone, D. F. Reinhold, C. S. Rooney, and W. H. Nicholson, *J. Medicin. Chem.*, 1968, **11**, 261.
[81] C. Ivanov and A. Dobrev, *Monatsh.*, 1967, **98**, 2001.
[82] D. H. Johnson, *J. Chem. Soc.* (*C*), 1968, 126.
[83] L. Pichat, J. -P. Guermont, and P. N. Liem, *J. Labelled Compounds*, 1968, **4**, 251.
[84] R. J. Correla, C. P. Arciprete, and A. E. A. Mitta, *Anales Asoc. quim. argentina*, 1967, **55**, 173.

previous methods. Synthesis of isoleucine from [$^{14}C_4$]-2-bromobutane by the acetamidomalonate route gave [$\beta,\gamma,\gamma',\delta$-^{14}C]-DL-isoleucine with a completely labelled side-chain, and this was obtained free of the allo-isomer by preparative paper chromatography.[85] The methods previously described for the 'cold' amino-acids have been used for the preparation of [α-^{14}C]-β-(1-pyrazolyl)-DL-alanine, [α-^{14}C]-β-(3-pyrazolyl)-DL-alanine, and [α-^{14}C]-β-(2-furyl)-DL-alanine.[86] A rapid and simple enzymic method for the resolution of [ε-^{14}C]-α-aminoadipic acid has been briefly described.[87] Syntheses of [α-^{14}C]- and [^{15}N]-DL-α-allylglycine and also [α-^{14}C]- and [^{15}N]-DL-homomethionine have been reported.[88] The preparation of [γ-^{14}C]-γ-aminobutyric acid (starting with radioactive cyanide) and its conversion to [4-^{14}C]-DL-azetidine-2-carboxylic acid have been published.[89]

Iodine monochloride (liberated *in situ* by the action of chloramine-T on radioactive iodide solutions) has been used in the preparation of [^{131}I]-[90, 91] and [^{125}I]-[90] 3-iodo-L-tyrosine and 3,5-di-iodo-L-tyrosine. The iodinated tyrosines were obtained with high specific activities: *e.g.* [^{131}I]-3-iodo-L-tyrosine prepared by this method had a specific activity of *ca.* 2C/μmole.[90] The chemical stability of radioactive iodotyrosines has been studied under a variety of experimental conditions under which they might be used as tracers:[91] they are stable only in the dark or in red light, and in the presence of traces of cupric ions (which might, for example, be present in chromatographic materials) they undergo rapid decomposition, even in the dark. Details of simplified small-scale preparations of [^{131}I]-labelled 3,3′,5-tri-iodo-L-thyronine [92, 93] and L-thyroxine [92] have been published, and the optimal conditions for the labelling of 3,3′,5-tri- and 3,3′,5,5′-tetra-iodo-L-thyronine by exchange have been studied.[93]

[$^2H_5$]-L-Phenyl-[α,β,β-$^2H_3$]-alanine has been synthesised in 18% overall yield from [$^2H_6$]-benzene by the oxazolone route: resolution was achieved *via* the *N*-acetyl derivative using renal acylase.[94] Melphalan (32) with



(32)

tritium in the aromatic ring has been obtained by a synthesis involving iodination and catalytic reduction with tritium.[95] Incubation of glycine

[85] G. Pascal, L. Pichat, and C. Baret, *Bull. Soc. chim. France*, 1968, 1481.
[86] V. Tolman, J. Hanuš, and K. Vereš, *J. Labelled Compounds*, 1968, **4**, 243.
[87] J. Mizon, *J. Labelled Compounds*, 1968, **4**, 278.
[88] M. Matsuo, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 1030.
[89] L. Pichat, P. N. Liem, and J.-P. Guermont, *Bull. Soc. chim. France*, 1968, 4079.
[90] B. L. Brown and W. S. Reith, *Biochim. Biophys. Acta*, 1967, **148**, 423.
[91] A. E. A. Mitta and G. B. de Salas, *Anales Asoc. quim. argentina*, 1967, **55**, 85.
[92] Å. Høye, *Acta Chem. Scand.*, 1968, **22**, 695.
[93] G. N. B. de Salas and A. E. A. Mitta, *Anales Asoc. quim. argentina*, 1967, **55**, 89.
[94] A. T. Blomquist and R. J. Cedergreen, *Canad. J. Chem.*, 1968, **46**, 1053.
[95] R. Wade and T. S. Murthy, *J. Chem. Soc.* (*C*), 1968, 2564.

with serine hydroxymethylase in a system containing tetrahydrofolate and pyridoxal phosphate (but lacking formaldehyde, which is required for the normal reaction of the enzyme) brings about rapid stereospecific exchange of one of the two α-hydrogen atoms of the glycine. Treatment of [α-$^3H_2$]-glycine under these conditions thus gives one enantiomer of [α-$^3H_1$]-glycine, whereas exchange of ordinary glycine with tritiated water gives the other.[96]

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

Compound	*Ref.*
L-*N*-(3-amino-3-carboxypropyl)-β-carboxypyridinium betaine. (Nicotianine)	13, 14
γ,δ,δ′-trihydroxy-L-leucine	68
N^{ϵ}-(indole-3-acetyl)-L-lysine	18
p-hydroxymethylphenylalanine (L, D and DL)	97
O-ethylhomoserine (L and DL)	98
DL-5-methyl-2-aminohex-4-enoic acid	21
DL-α-(3-hydroxyphenyl)glycine	23
DL-α-(3,5-dihydroxyphenyl)glycine	23
β-(cyclohexa-1,4-dienyl)alanine (*i.e.* 3,6-dihydrophenylalanine: L and DL)	49
β-(cyclohex-1-enyl)-DL-alanine (*i.e.* 3,4,5,6-tetrahydrophenylalanine)	49
β-(1-hydroxycyclohexyl)-DL-alanine	49
β-(cyclopent-2-enyl)-DL-alanine	99
β-(cyclopent-3-enyl)-DL-alanine	99
β-(cyclohex-2-enyl)-DL-alanine	99
β-(cycloheptl-1-enyl)-DL-alanine	99
DOPA dimer (12)	52
cycloDOPA (15)	53
4-bromoacetyl-DL-phenylalanine	100
4-bromoacetamido-DL-phenylalanine	100
3-chloroacetamido-DL-phenylalanine	100
4-fluoro-3-chloroacetamido-DL-phenylalanine	100
3,4,5-tri-iodo-DL-phenylalanine	101
3,5-di-isopropyl-3′-iodo-DL-thyronine	102
β-(3-chloro-4-methoxy-1-naphthyl)-DL-alanine	103
β-)3-chloro-4-hydroxy-1-naphthyl)-DL-alanine	103
β-(4-methoxy-1-naphthyl)-DL-alanine	103
β-(4-hydroxy-1-naphthyl)-DL-alanine	103
DL-β-(4-methoxy-1-naphthyl)-α-methylalanine	103
DL-β-(4-hydroxy-1-naphthyl)-α-methylalanine	103
DL-α-(2-indanyl)glycine	104
β-trimethylsilyl-DL-alanine	104

[96] M. Akhtar and P. M. Jordan, *Chem. Comm.*, 1968, 1691.
[97] S. C. Smith and N. H. Sloane, *Biochim. Biophys. Acta*, 1967, **148**, 414.
[98] Y. Murooka, T. Harada, and Y. Izumi, *Bull. Chem. Soc. Japan*, 1968, **41**, 633.
[99] T. H. Porter, R. M. Gipson, and W. Shive, *J. Medicin. Chem.*, 1968, **11**, 263.
[100] J. I. DeGraw, M. Cory, W. A. Skinner, M. C. Theisen, and C. Mitoma, *J. Medicin. Chem.*, 1968, **11**, 225.
[101] V. B. Schatz, B. C. O'Brien, and W. R. Sandusky, *J. Medicin. Chem.*, 1968, **11**, 140.
[102] T. Matsuura, T. Nagamachi, K. Matsuo, and A. Nichinaga, *J. Medicin. Chem.*, 1968, **11**, 899.
[103] A. J. Ablewhite and K. R. H. Wooldridge, *J. Chem. Soc.* (*C*), 1967, 2488.
[104] T. H. Porter and W. Shive, *J. Medicin. Chem.*, 1968, **11**, 402.

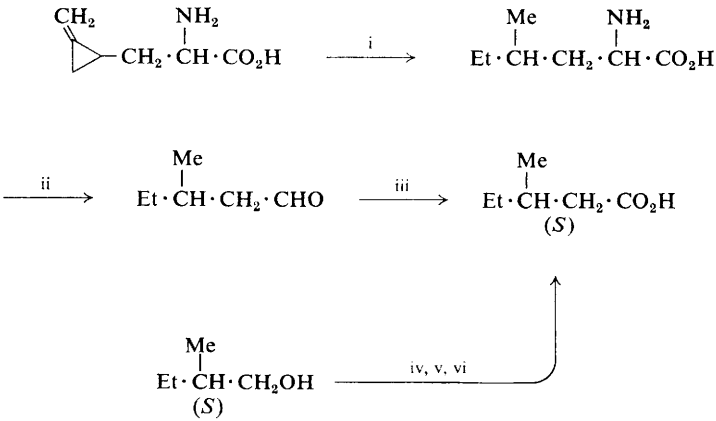
Compound	*Ref.*
DL-2-amino-3,3-dimethylhex-5-enoic acid	50[a]
DL-2-aminohexa-4,5-dienoic acid	50[a], 50[b]
DL-2-amino-3,3-dimethylhexa-4,5-dienoic acid	50[a]
DL-2-aminohepta-4,5-dienoic acid	50[a], 50[b]
DL-2-amino-3,3-dimethylhepta-4,5-dienoic acid	50[a]
DL-2-amino-3,3-dimethylnona-4,5-dienoic acid	50[a]
DL-2-aminohepta-5,6-dienoic acid	50[b]
DL-2-amino-3-methylhepta-5,6-dienoic acid	50[b]
DL-2-amino-5-*t*-butyl-6,6-dimethylhepta-3,4-dienoic acid	50[b]
DL-2-amino-5-methylhepta-3,4-dienoic acid	50[b]
DL-2-aminohept-4-en-6-ynoic acid	50[b]
ε-hydroxy-β-carboxy-DL-norleucine (diastereoisomers separated but not distinguished)	105[a]
β-carboxy-DL-lysine (diastereoisomers separated but not distinguished)	105[b]
β-(3,4-dihydroxyphenyl)-α-methyl-DL-serine (*threo* and *erythro*)	58
S-benzyl-β,γ-dimethyl-DL-homocysteine	106
S-benzyl-α,γ,γ-trimethyl-DL-homocysteine	106
Se-benzyl-β,γ-dimethyl-DL-selenohomocysteine	106
Se-benzyl-α,γ,γ-trimethyl-DL-selenohomocysteine	106
β-methyl-DL-methionine	107
α-methyl-DL-selenomethionine	107
β-methyl-DL-selenomethionine	107
γ-methyl-DL-selenomethionine	107
Se-benzylselenohomocysteine (L and D)	108
Selenomethionine (L and D)	108
Se-benzylselenocysteine (L and D)	108
Selenolanthionine (L, D, and *meso*)	108
DL-5,5,5-trifluoro-4-trifluoromethyl-2-aminovaleric acid	109
γ,γ′-difluoro-DL-valine	110
δ,δ′-difluoro-DL-leucine	110
γ-fluoro-DL-allothreonine	110
β-(5-nitro-3,4,-dihydro-2,4-dioxo-1(2*H*)-pyrimidinyl)-DL-alanine (5-nitro-DL-Willardiine)	111
β-(5-bis[2-chloroethyl]amino-3,4-dihydro-2,4-dioxo-1(2*H*)-pyrimidinyl)-DL-alanine (DL-Willardiine mustard)	111
DL-α-amino-α-methyl-γ-(1-oxo-2(2*H*)-isoquinolyl)butyric acid	112
DL-α-amino-α-methyl-γ-(2-oxo-1(1*H*)-quinolyl)butyric acid	112
β-hydroxy-L-asparagine (*threo* and *erythro*)	113
β-hydroxy-DL-isoleucine (*threo* and *erythro*)	66
β-methoxy-DL-isoleucine (*threo* and *erythro*)	66
DL-α-amino-γ-(methylamino)butyric acid (5-azanorleucine)	114
DL-α-amino-β-(ethylamino)propionic acid (4-azanorleucine)	114

[105a] E. Kuss, *Z. physiol. Chem.*, 1967, 348, 1589. [b] E. Kuss, *Z. physiol. Chem.*, 1967, 348, 1596.
[106] G. Zdansky, *Arkiv Kemi*, 1968, **29**, 47.
[107] G. Zdansky, *Arkiv Kemi*, 1967, **27**, 447.
[108] G. Zdansky, *Arkiv Kemi*, 1968, **29**, 437 and 443.
[109] J. Lazar and W. A. Sheppard, *J. Medicin. Chem.*, 1968, **11**, 138.
[110] H. Lettré and U. Wölcke, *Annalen*, 1967, **708**, 75.
[111] A. P. Martinez, W. W. Lee, and L. Goodman, *J. Medicin. Chem.*, 1968, **11**, 60.
[112] S. Kamiya and K. Koshinuma, *Chem. and Pharm. Bull. (Japan)*, 1967, **15**, 1985.
[113] A. Singerman and Y. Liwschitz, *Tetrahedron Letters*, 1968, 4733.
[114] T. J. McCord, L. D. Booth, and A. L. Davies, *J. Medicin. Chem.*, 1968, **11**, 1077.

Compound	*Ref.*
DL-α-amino-β-(methylamino)propionic acid (4-azanorvaline)	114
L- and D-α-amino-β-(methylamino)propionic acid (4-azanorvaline)	47
$N^{\epsilon}N^{\epsilon}$-bis(2-cyanoethyl)-L-lysine	115
DL-Capreomycidine (16)	54
DL-α,γ-dimethylnorleucine	116
DL-α-methyl-$N^{\delta}N^{\delta}$-diethylornithine	116
DL-α-(1-methylcyclopropyl)alanine	116
α-ethyl-3,4-dimethoxy-DL-phenylalanine	116
α-methyl-4-morpholino-DL-phenylalanine	116
β-(2-amino-4-pyrimidinyl)alanine. (Laterine: DL and L)	117
DL-1-methyl-6-hydroxy-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid	24
β-(2-amino-4-pyrimidinyl)alanine. (Laterine: DL and L)	117
2,3-*trans*-3,4-*cis*-3,4-dihydroxy-DL-proline	74
N^{α},N^{α}-dimethyl-L-histidine	70

3 Physical and Stereochemical Studies of Amino-acids

A. Determination of Absolute Configuration.—Hypoglycin A (11) had been predicted [51] to have the configuration (2*S* : 4*S*). Degradation to 3-methylpentanoic acid and comparison with material of known (*S*) configuration (Scheme 6) established the configuration at C-4 as (*S*) and confirmed the



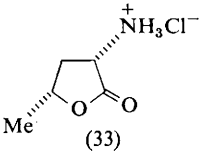
Reagents: PtO_2—H_2; ii, NBS; iii, Ag_2O; iv, HBr; v, Mg; vi, CO_2

Scheme 6

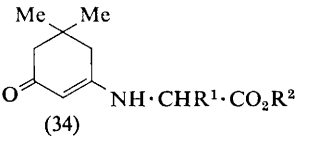
prediction, since the configuration at C-2 was known to be (*S*) from enzymic studies. Chemical correlation with (*R*)-isovaline has shown that both (+)-α-methylphenylglycine [118] and (+)-α-methylserine [119] belong to the

[115] J. F. Cavins and M. Friedman, *Biochemistry*, 1967, **6**, 3766.
[116] C. J. Abshire, *Experientia*, 1968, **24**, 778.
[117] Yu. P. Shvachkin, G. A. Korshunova, N. A. Bashkirova, and M. A. Prokof'ev, *Doklady Akad. Nauk S.S.S.R.*, 1968, **179**, 1127.
[118] H. Mizuno, S. Terashima, K. Achiwa, and S. Yamada, *Chem. and Pharm. Bull. (Japan)*, 1967, **15**, 1749.
[119] N. Takamura, S. Terashima, K. Achiwa, and S. Yamada, *Chem. and Pharm. Bull. (Japan)*, 1967, **15**, 1776.

(*S*) series. N.m.r. studies on the corresponding lactone hydrochlorides have enabled the absolute configurations at C-4 of some γ-hydroxy-L-amino-acids obtained from toxins of *Amanita phalloides* to be determined.[120] Similarly, n.m.r. was used to assign the (2*S*:4*R*) configuration to the lactone hydrochloride (33), which was obtained by photochlorination of L-norvaline in concentrated hydrochloric acid, followed by lactonisation and separation from the resulting mixture of diastereoisomers. Comparison of the chromatographic mobility of the (2*S*:4*R*)-γ-hydroxynorleucine obtained by hydrolysis of (33) with that of γ-hydroxy-L-norleucine isolated[121] in 1966 from *Lathyrus odoratus* indicates that the absolute configuration of the latter is (2*S*:4*S*).



(33)



(34)

The signs of the Cotton effects shown by cobalt complexes of amino-acids,[122] *N*-acryloyl derivatives of amino-acids,[123] osmate esters of *N*-acryloyl derivatives of amino-acids,[125] *N*-(*N*-oxido-2-pyridyl)-amino-acids,[125] and dimedonyl derivatives of amino-acids and their esters (34)[126] have been correlated with absolute configuration. The compounds (34) obtained by condensing dimedone (the analogous derivatives of dihydroresorcinol were also examined) with amino-acid alkyl esters were investigated particularly thoroughly and offer some promise as convenient chromophoric derivatives for the assignment of absolute configuration from o.r.d. measurements.[126]

B. Crystal Structures of Amino-acids.—The crystal structures of L-cysteine,[127] L-cysteic acid,[128] L-ornithine hydrochloride,[129] sarcosine hydrochloride,[130] diiodo-L-tyrosine dihydrate,[131] and 1-aminocyclopentane carboxylic acid[132] have been described, and further refinements of the structure of L-cysteine hydrochloride[133] have been published. The mono-

[120] T. Wieland, M. Hasan, and P. Pfaender, *Annalen*, 1968, **717**, 205.
[121] L. Fowden, *Nature*, 1966, **209**, 807.
[122] C. J. Hawkins and P. J. Lawson, *Chem. Commun.*, 1968, 177.
[123] N. Sakota, *J. Chem. Soc. Japan*, 1968, **89**, 425.
[124] N. Sakota and N. Koine, *J. Chem. Soc. Japan*, 1967, **88**, 1087.
[125] V. Tortorella and G. Bettoni, *Gazzetta*, 1968, **98**, 316.
[126] P. Crabbe, B. Halpern, and E. Santos, *Tetrahedron*, 1968, **24**, 4315.
[127] M. M. Harding and H. A. Long, *Acta Cryst.*, 1968, **24**, 1096.
[128] H. Konishi, T. Ashida, and M. Kakudo, *Bull. Chem. Soc. Japan*, 1968, **41**, 2305.
[129] N. N. Saha, S. K. Mazumdar, and S. Guha, *Science and Culture*, 1968, **34**, 72.
[130] N. N. Saha, S. K. Mazumdar, and S. C. Bhattacharyya, *Science and Culture*, 1968, **34**, 47.
[131] J. A. Hamilton and L. K. Steinrauf, *Acta Cryst.*, 1967, **23**, 817.
[132] R. Chandrasekharan, M. Mallikarjunan, G. Chandrasekharan, and R. Zand, *Current Sci.*, 1968, **37**, 91.
[133] R. Ramachandra Ayyar, *Z. Krist.*, 1968, **126**, 227.

clinic crystalline modification of DL-α-amino-*n*-butyric acid contains molecules in three different conformations, but only one conformer is found in the tetragonal form.[134]

C. Optical Rotatory Dispersion (O.r.d.) and Circular Dichroism (C.d.).—Details of an improved dichrograph for use with amino-acids and proteins have been reported.[135] The improved technique uses an average transient computer to achieve a better signal : noise ratio, and the c.d. spectrum of L-tryptophan in water in the range 310–190 nm was determined in this way. C.d. measurements on a number of L-amino-acids have revealed a negative dichroic band near 250 nm in addition to the strong band previously found at shorter wavelengths.[136]

There is considerable current interest in the o.r.d. of L-cystine and related compounds,[137–139] but as these studies are concerned with contributions from disulphide bonds to the o.r.d. of proteins, they will not be discussed here. Changes in the optical rotatory properties of amino-acids are induced by 6M-urea: these effects cannot be ascribed to changes in the ionisation state or to refractive index changes.[140] Papers on the assignment of configuration to amino-acids by examination of the o.r.d. of suitable derivatives have been listed above (see p. 15).

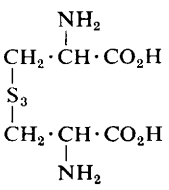
D. Nuclear Magnetic Resonance (N.m.r.) Spectra.—The interpretation of the ^1H n.m.r. spectra of DL-threonine and DL-valine in terms of 'rotational isomers' by Aruldas[141] has been under fire from several quarters.[142, 143a] It was pointed out that the barriers to rotation about the C(α)—C(β) bonds were unlikely to be great enough to enable observation of distinct rotamers at room temperature,[142] and it was also shown that the DL-threonine examined by Aruldas must have been grossly contaminated with DL-allothreonine.[142, 143a] It is, of course, unnecessary to postulate restricted rotation about the C(α)—C(β) bond in order to account for the non-equivalence of the methyl groups which is always observed in the n.m.r. spectra of valine and its derivatives—the methyl groups would be non-equivalent even if completely free rotation were possible, because of the adjacent asymmetric centre. Further studies on the 'anomalous' vicinal coupling constants of the aliphatic protons in the phenylalanine anion in aqueous solution are consistent with the suggestion that the two less

[134] T. Ichikawa, Y. Ittaka, and M. Tsuboi, *Bull. Chem. Soc. Japan*, 1968, **41**, 1027.
[135] Y. P. Myer and L. H. MacDonald, *J. Amer. Chem. Soc.*, 1967, **89**, 7142.
[136] R. D. Anand and M. K. Hargreaves, *Chem. and Ind.*, 1968, 880.
[137] D. L. Coleman and E. R. Blout, *J. Amer. Chem. Soc.*, 1968, **90**, 2405.
[138] M. Carmack and L. A. Neubert, *J. Amer. Chem. Soc.*, 1967, **89**, 7134.
[139] P. C. Kahn and S. Beychok, *J. Amer. Chem. Soc.*, 1968, **90**, 4168.
[140] L. I. Katzin and G. C. Kresheck, *Arch. Biochem. Biophys.*, 1968, **126**, 418.
[141] G. Aruldas, *Spectrochim. Acta*, 1967, **23**, *A*, 1345.
[142] J. F. Newmark and R. A. Newmark, *Spectrochim. Acta*, 1968, **24**, *A*, 952.
[143a] B. Bak and F. Nicolaisen, *Acta Chem. Scand.*, 1967, **21**, 1980. [b] J. R. Cavanaugh, *J. Amer. Chem. Soc.*, 1968, **90**, 4533. [c] B. Bak, C. Dambmann, F. Nicolaisen, E. J. Pedersen, and N. S. Bhacca, *J. Mol. Spectroscopy*, 1968, **26**, 78. [d] B. Bak, C. Dambmann, and F. Nicolaisen, *Acta Chem. Scand.*, 1967, **21**, 1674.

favourable staggered conformations have enhanced stability at low temperatures and concentrations, possibly due to specific solute–solvent interactions.[143b]

The ^1H n.m.r. spectra of all the protein amino-acids have been determined in trifluoroacetic acid and deuteriotrifluoroacetic acid at 220 MHz.[143c] The spectrum of tryptophan in deuteriotrifluoroacetic acid indicated exchange of hydrogen for deuterium in both rings of the indole system.[143c, 143d] Since this exchange only proceeds at a significant rate under conditions of very high acidity, it was suggested that it may be possible to use acid-catalysed exchange for selective isotopic labelling of tryptophan residues.[143d] The protons at the 2- and 6-positions of the indole ring undergo exchange most rapidly, and it is of interest in this connection to note that the dissociation constant for the protonation of the side chain of tryptophan in strong sulphuric acid solutions has recently been determined.[144]

^1H N.m.r. data at 60 MHz for 4-oxoprolines and *cis*- and *trans*-4-substituted prolines have been published, and correlations permitting assignment of configuration at the 4-position from the spectra have been deduced.[145] The ^1H n.m.r. spectra of a number of sulphur-containing amino-acids (including L-cystine, L-cysteine, L-cysteic acid, DL-lanthionine, and L-djenkolic acid) have been determined at 60 MHz in deuterium oxide as part of an n.m.r. study of bis(2-amino-2-carboxyethyl)-trisulphide (35),



(35)

which is found in acid hydrolysates of wool.[146] N.m.r. is now being used routinely for the characterisation of new amino-acids when sufficient material is available and the majority of references in the appendix to the section devoted to naturally occurring amino-acids report n.m.r. data. A particularly important application is in differentiation between diastereoisomers of amino-acids with two asymmetric centres—see, *e.g.*, refs. 3, 66, and 145.

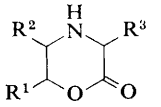
The ^{13}C n.m.r. spectra of glycine and alanine have been examined using a double-resonance technique as a preliminary part of a programme to investigate the potential of ^{13}C n.m.r. spectrometry as a tool for conformational work with macromolecules.[147]

[144] R. C. Armstrong, *Biochem. Biophys. Acta*, 1968, **158**, 174.
[145] R. H. Andreatta, V. Nair, and A. V. Robertson, *Austral. J. Chem.*, 1967, **20**, 2701.
[146] K. D. Bartle, J. C. Fletcher, D. W. Jones, and R. L'Amie, *Biochim. Biophys. Acta*, 1968, **160**, 106.
[147] W. J. Horsley and H. Sternlicht, *J. Amer. Chem. Soc.*, 1968, **90**, 3738.

4 Chemical Studies of Amino-acids

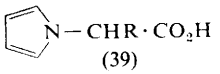
A. General Reactions.—The reactions of fluoronitropyridines with amino-acids have been described,[148, 149] and the *N*-(5-nitro-6-methyl-2-pyridyl) derivatives of most of the common α-amino-acids have been characterised.[149] The reaction of glycine (as its sodium salt or ethyl ester) with epoxides gives *N*-(2-hydroxyalkyl) derivatives (36) from which lactones (37) can be

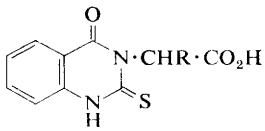
$$HOCHR^1 \cdot CHR^2 \cdot NH \cdot CHR^3 \cdot CO_2R^4$$
(36)


(37)

$$HOCH_2 \cdot CH_2 \cdot CO \cdot NH \cdot CHR \cdot CO_2H$$
(38)

obtained: 2,5-dioxopiperazine and the products of its reaction with the epoxide are also formed.[150] β-Propiolactone reacts rapidly with all the protein amino-acids at pH 9 and room temperature.[151] It was inferred from earlier work on the reaction of β-propiolactone with simple amines that acylation of the α-amino group giving (38) occurs in the absence of other nucleophilic functionalities, but this was not confirmed by isolation of the products. At low pH values only sulphur-containing amino-acids react, and the reaction is specific at pH 3·0 and 0° for methionine, which presumably gives an acylsulphonium derivative.[151] α-Amino-acids react with 2,5-diethoxytetrahydrofuran in acetic acid in the presence of sodium acetate giving α-pyrrolo-acids (39), which are of interest as amino-acid analogues.[152] Reaction of *o*-(methoxycarbonyl)phenyl isothiocyanate with α-amino-acids in alkaline solution gives hydroquinazolones (40) which may


(39)

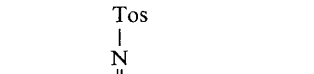

(40)

undergo further cyclisation if functional groups are present in the side chain.[153] A detailed study of the acylation of α,ω-diamino-acids using phenyl esters has been reported:[154] *p*-nitrophenyl acetate acylates the unprotected diamino-acids homolysine, lysine, and ornithine exclusively at the ω-amino group at pH 11, but the selectivity is less at lower pH values. Selective ω-acylation is not possible at any alkaline pH in the cases of

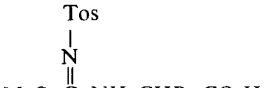
[148] T. Talik and Z. Talik, *Bull. Acad. polon. Sci., Ser. Sci. chim.*, 1968, **16**, 13.
[149] Z. Talik and B. Brekiesz-Lewandowska, *Roczniki Chem.*, 1967, **41**, 2095.
[150] K. Jankowski and C. Berse, *Canad. J. Chem.*, 1967, **45**, 2865.
[151] M. A. Taubman and M. Z. Atassi, *Biochem. J.*, 1968, **106**, 829.
[152] J. Gloede, K. Poduska, H. Gross, and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1968, **33**, 1307.
[153] E. Cherbuliez, O. Espejo, B. Willhalm, and J. Rabinowitz, *Helv. Chim. Acta*, 1968, **51**, 241.
[154] J. Leclerc and L. Benoiton, *Canad. J. Chem.*, 1968, **46**, 1047.

α,γ-diaminobutyric acid and α,β-diaminopropionic acid, however. Treatment of amino-acid salts with *S*-methylisothiosemicarbazide hydroiodide in boiling water or ethanol gives *N*-(aminoguanyl)-amino-acids (41), which are rather bizarre dipeptide analogues.[155] The related derivatives (42) have also been obtained, by hydrazinolysis of the products (43) of the reaction of amino-acid salts with (44).[155]

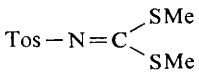
(41)



(42)

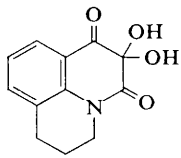


(43)



(44)

Thermal decarboxylation of α-amino-acids in the presence of ketones gives, after hydrolysis of the intermediate decarboxylated Schiff base, either the normal decarboxylation product or the amine corresponding to the ketone (*i.e.* the transamination product) or both, depending on the structures of the reactants.[156] Studies [157, 158] of the reaction of the ninhydrin analogue (45) with α-amino-acids have provided, by analogy, further corroboration of McCaldin's mechanism [159] for the ninhydrin reaction.



(45)

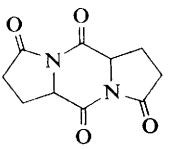
B. Other Reactions.—The volatile products obtained by pyrolysis of phenylalanine, tyrosine, tryptophan and histidine have been examined by gas–liquid chromatography (g.l.c.) and mass spectrometry.[160] Among the volatile degradation products obtained from phenylalanine were toluene, styrene, benzene, ethylbenzene, benzonitrile, phenylethylamine, water and carbon dioxide. The complexity of the pyrolytic pathway and the multiplicity of products obtained from this one amino-acid alone must undermine

[155] J. Gante, *Chem. Ber.*, 1968, **101**, 1195.
[156] A. F. Al-Sayyab and A. Lawson, *J. Chem. Soc.* (*C*), 1968, 406.
[157] H. Wittmann, W. Dreveny, and E. Ziegler, *Monatsh.*, 1968, **99**, 1205.
[158] H. Wittmann, W. Dreveny, and E. Ziegler, *Monatsh.*, 1968, **99**, 1543.
[159] D. J. McCaldin, *Chem. Rev.*, 1960, **60**, 39.
[160] G. P. Shulman and P. G. Simmonds, *Chem. Comm.*, 1968, 1040.

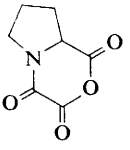
any hopes [161] that identification of characteristic pyrolysis products might be useful for qualitative amino-acid analysis.

Amino-acids with thioether side chains are oxidised to the corresponding sulphoxides in high yield by means of diethyl azodicarboxylate.[162] This reagent converts cysteine to cystine (cysteic acid if a very large excess of oxidant is used) and also reacts with tyrosine, but in this case the nature of the products has not yet been determined.

Dehydration of glutamic acid with acetic anhydride–pyridine gives a high melting substance to which the structure (46) was originally assigned.[163] This structure was recently criticised [164] and a different structure proposed, but the original work [163] has now been repeated: [165] the formulation (46) was confirmed and it was shown that the criticisms of it are invalid because they are based on an erroneous interpretation of the i.r. data. A convenient method for the dehydration of asparagine and glutamine (discovered during attempted *N*-carboxyanhydride preparations) uses phosgene in dioxan:[166] yields of the benzyloxycarbonyl derivatives of β-cyano-L-alanine and γ-cyano-L-α-aminobutyric acid prepared in this way were excellent. This alternative to the usual dehydration using dicyclohexylcarbodi-imide [167] seems to be a great improvement.



(46)



(47)

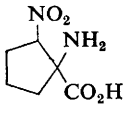
When α-amino-acids with primary amino-groups are treated with oxalyl chloride, intractable tarry substances are produced under all conditions.[168] These tarry materials probably arise *via* oxazolones, and this suggestion is borne out by the fact that proline, which cannot form an oxazolone, gives a crystalline *N*-oxalic anhydride (47) when heated with oxalyl chloride in dioxan.[168] The cyclisation of *N*-oxalyl-α-dialkyl-α-amino-acids has also been investigated.[169]

α-Sulpho-β-alanine, previously not easily available in quantity, can be obtained from the reaction of acrylonitrile with oleum in poor yield,[170] but in almost quantitative yield by treatment of β-alanine with a large

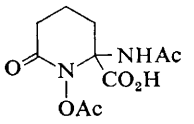
[161] C. Merritt and D. H. Robertson, *J. Gas Chromatog.*, 1967, **5**, 96.
[162] R. Axen, M. Chaykovsky, and B. Witkop, *J. Org. Chem.*, 1967, **32**, 4117.
[163] J. A. King and F. H. McMillan, *J. Amer. Chem. Soc.*, 1952, **74**, 2859.
[164] S. El-Zanfally, M. Khalifa, and Y. M. Abou-Zeid, *Tetrahedron*, 1966, **22**, 2307.
[165] M. R. Harnden, *J. Heterocyclic Chem.*, 1968, **5**, 307.
[166] M. Wilchek, S. Ariely, and A. Patchornik, *J. Org. Chem.*, 1968, **33**, 1258.
[167] C. Ressler and H. Ratzkin, *J. Org. Chem.*, 1961, **26**, 3356.
[168] W. R. Hearn and R. E. Worthington, *J. Org. Chem.*, 1967, **32**, 4072.
[169] W. R. Hearn and J. Medina-Castro, *J. Org. Chem.*, 1968, **33**, 3980.
[170] A. Zilkha, I. Barzilay, J. Naiman, and B.-A. Feit, *J. Org. Chem.*, 1968, **33**, 1686.

excess of oleum at room temperature.[171] Preliminary experiments indicate that this reaction may be general for β-amino-acids.

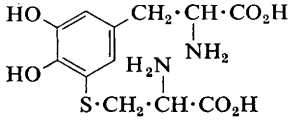
Unless the acetylation of 1-amino-2-nitrocyclopentane carboxylic acid (48) is performed under mild conditions, further reaction occurs with rearrangement, giving an *N*-acetoxypiperidone (49).[172] The mechanism was not examined in very great detail, but is presumably related to similar rearrangements of 2-nitrocyclopentanones, which yield *N*-acetoxy-imides when heated with acetic anhydride.[173]



(48)



(49)



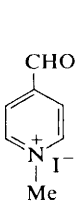
(50)

C. Non-enzymic Models of Biochemical Processes Involving Amino-acids.— The reactions of cysteine with *o*-quinones have been studied [174, 175] in order to obtain background information on the reactions which are probably involved in the biosynthesis of phaeomelanins. Thus cysteine and *N*-acetylDOPAquinone ethyl ester give, after removal of the protecting groups, 5-*S*-cysteinyl-DOPA (50). Since (50) gives on oxidation two pigments which are spectroscopically identical to pigments obtained from chicken feathers, it may be implicated in the biogenesis of phaeomelanins.[175]

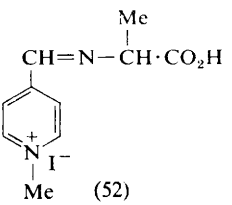
In non-enzymic transamination reactions catalysed by pyridine-4-aldehydes in the absence of metal ions, a 3-hydroxy group is generally required to (*a*) catalyse aldimine formation and (*b*) assist the aldimine-ketimine tautomeric change. If the pyridine nitrogen carries a positive charge, however, this requirement lapses because the positive charge provides an inductive effect strong enough to bring about the necessary changes without additional assistance. Thus *N*-methyl-4-formyl-pyridinium iodide (51) rapidly forms an aldimine (52) with alanine, and a slower subsequent prototropic shift gives the ketimine (53) and finally transamination products after hydrolysis. The conversion of (52) to (53) was shown to be a general base catalysed process with the dihydropyridine species (54) as a sufficiently long-lived intermediate to be detected spectroscopically.[176]

The oxidative coupling of 4-hydroxy-3,5-di-iodophenylpyruvic acid (55) and 3,5-di-iodophenylalanine gives fair yields of thyroxine (56), and is therefore an attractive model for its biosynthesis.[177, 178] Evidence has

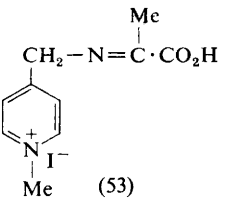
[171] D. Wagner, D. Gertner, and A. Zilkha, *Tetrahedron Letters*, 1968, 4479.
[172] W. B. Turner, *J. Chem. Soc.* (*C*), 1967, 2225.
[173] A. Hassner and J. Larkin, *J. Amer. Chem. Soc.*, 1963, **85**, 2181.
[174] G. Prota, G. Scherillo, F. Napolano, and R. A. Nicolaus, *Gazzetta*, 1967, **97**, 1451
[175] G. Prota, G. Scherillo, and R. A. Nicolaus, *Gazzetta*, 1968, **98**, 495.
[176] J. R. Maley and T. C. Bruice, *J. Amer. Chem. Soc.*, 1968, **90**, 2843.
[177] G. Hillman, *Z. Naturforsch.*, 1956, **11b**, 424.
[178] R. I. Meltzer and R. J. Stanaback, *J. Org. Chem.*, 1961, **26**, 1977.



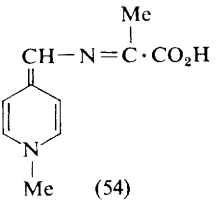
(51)



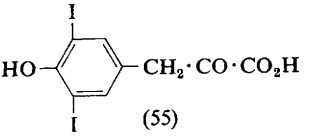
(52)



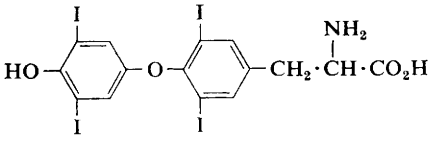
(53)



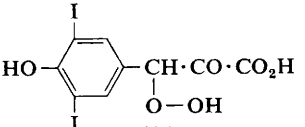
(54)



(55)



(56)



(57)

recently been presented to show that hydroperoxide (57) is first formed, followed by an anaerobic reaction with 3,5-di-iodotyrosine, but the mechanism of the final coupling has not yet been clarified.[179, 180]

The non-enzymic hydroxylating system of Udenfriend (ferrous ion, ascorbic acid, ethylenediamine tetra-acetic acid and oxygen) causes hydroxylation of proline in a manner which does not involve the production of hydrogen peroxide or hydroxyl radicals:[181] the mechanism probably conforms to the general mechanism which has been proposed[182] for oxidations by this system.

The oxidative decarboxylation reactions of primary, secondary, and tertiary amino-acids which are induced by hypohalite have been studied in very great detail as models for some stages in alkaloid biosynthesis.[183]

D. Effects of Electromagnetic Radiation on Amino-acids.—There continues to be much activity in research on the effects of ionising radiation on amino-acids. Several electron spin resonance (e.s.r.) studies of irradiated

[179] A. Nishinaga, H. J. Cahnmann, H. Kon, and T. Matsuura, *Biochemistry*, 1968, **7**, 388.
[180] F. Blasi, F. Fragomele, and I. Covelli, *European J. Biochem.*, 1968, **5**, 215.
[181] M. Bade and B. S. Gould, *Biochim. Biophys. Acta*, 1968, **156**, 425.
[182] G. A. Hamilton, *J. Amer. Chem. Soc.*, 1964, **86**, 3391.
[183] E. E. van Tamelen, V. B. Haarstad, and R. L. Orvis, *Tetrahedron*, 1968, **24**, 687.

amino-acids (in crystals or glasses) have been reported.[184-189] The radicals formed by irradiation of amino-acid crystals are normally lost on dissolution by reaction with the solvent, but if the irradiated crystals are dissolved in an aqueous solution of t-nitrosobutane the radicals (R·) which were formerly held in the crystalline matrix are trapped as relatively stable nitroxide radicals (Scheme 7), and these can then be examined by e.s.r.[190] There have

$$R^{\bullet} + Bu^{t}-N=O \longrightarrow Bu^{t}-\underset{\underset{O^{\bullet}}{|}}{N}-R$$

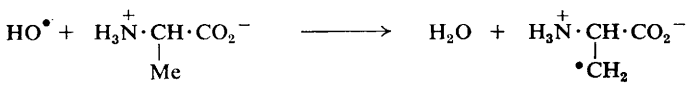
Scheme 7

also been a number of papers dealing with the radiolysis of amino-acids in aqueous solution: interest has been concentrated on aromatic[191-193] and sulphur-containing[194-197] amino-acids, but the effects of γ-radiation on proline and hydroxyproline have also been briefly described.[198]

In aerated aqueous solutions, most of the radiation damage to amino-acids is caused by the oxidising species (HO·), as most of the reducing species (solvated electrons and hydrogen atoms) are scavenged by oxygen giving hydroperoxyl radicals, which are not very reactive towards most amino-acids. The transient primary free radicals formed by radiolysis of amino-acids in aerated solutions cannot be observed directly by e.s.r., as it would require inconveniently vigorous irradiation to achieve a sufficient concentration. Chemically produced hydroxyl radicals have therefore been used to simulate the effects of radiation. The generation of hydroxyl radicals (by reaction of titanium trichloride with hydrogen peroxide) in the presence of amino-acids thus leads to the same free radicals as are formed by radiolysis, and these have been studied by e.s.r.[199-201] Except in the case of cysteine and cystine, when sulphur radicals were

[184] P. B. Ayscough and A. K. Roy, *Trans. Faraday Soc.*, 1968, **64**, 582.
[185] P. B. Ayscough, K. Mach, J. P. Oversby, and A. K. Roy, *Chem. Comm.*, 1967, 1084.
[186] M. Fujimoto, W. A. Seddon, and D. R. Smith, *J. Chem. Phys.*, 1968, **48**, 3345.
[187] D. G. Cadena and J. R. Rowlands, *J. Chem. Soc.* (*B*), 1968, 488.
[188] F. G. Liming jun. and W. Gordy, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 794.
[189] A. Hedberg and A. Ehrenberg, *J. Chem. Phys.*, 1968, **48**, 4822.
[190] C. Lagercrantz and S. Forshult, *Nature*, 1968, **218**, 1247.
[191] G. A. Brodskaya and V. A. Sharpati, *Khim. vysok. Energii*, 1968, **2**, 254.
[192] G. A. Brodskaya and V. A. Sharpati, *Zhur. fiz. Khim.*, 1967, **41**, 2850.
[193] J. Chrysochoos, *Radiation Res.*, 1968, **33**, 465.
[194] J. E. Packer and R. V. Winchester, *Chem. Comm.*, 1968, 826.
[195] T. C. Owen, M. Rodriguez, B. G. Johnson, and J. A. G. Roach, *J. Amer. Chem. Soc.*, 1968, **90**, 196.
[196] V. G. Wilkening, M. Lal, M. Arends, and D. A. Armstrong, *J. Phys. Chem.*, 1968, **72**, 185.
[197] A. Al-Thannon, R. M. Peterson, and C. N. Trumbore, *J. Phys. Chem.*, 1968, **72**, 2395.
[198] K. S. Korgaonkar, S. V. Marathe, and K. A. Chaubal, *Science and Culture*, 1968, **34**, 100.
[199] W. A. Armstrong and W. G. Humphreys, *Canad. J. Chem.*, 1967, **45**, 2589.
[200] H. Taniguchi, K. Fukui, S. Ohnishi, H. Hatano, H. Hasegawa, and T. Maruyama, *J. Phys. Chem.*, 1968, **72**, 1926.
[201] R. Poupko, B. L. Silver, and A. Lowenstein, *Chem. Comm.*, 1968, 453.

produced, the radicals observed were mostly formed by hydrogen abstraction from the β-carbon atom (*e.g.* Scheme 8) because hydrogen on the



Scheme 8

α-carbon atom is deactivated towards the electrophilic hydroxyl radical by the protonated amino-group. Some amino-acids also suffered hydrogen abstraction from other positions in the side-chain, but only glycine (which has little alternative) gave a radical formed by loss of a hydrogen radical from the α-carbon atom. Amino-acid analysis of the ninhydrin-positive products of these reactions showed that the products were identical with those produced by γ-irradiation, confirming the validity of the model. These results suggest that ideas about the course of radiolysis of aerated aqueous amino-acid solutions may have to be revised, as earlier work (based largely on product studies) gave rise to the conclusion [202] that abstraction of hydrogen from the α-carbon atom (*e.g.* Scheme 9) was the first step in the main pathway to products.

Scheme 9

The rates of reaction of a series of amino-acids with radiolytically produced hydrogen atoms in degassed aqueous solution have been determined.[203]

An e.s.r. study (covering most of the protein amino-acids) of the free radicals produced by irradiation of polycrystalline amino-acid samples with light of wavelength 253·7 nm has been published.[204] The signals which were observed were frequently different from those obtained after exposure to ionising radiation: most amino-acids gave radicals formed by loss of a hydrogen atom, but the spectrum obtained from phenylalanine showed evidence of the addition of a hydrogen atom giving a cyclohexadienyl radical. Radicals produced by the addition of hydrogen atoms to the aromatic rings of phenylalanine, tyrosine, and tryptophan have also been detected after γ-irradition or bombardment with thermal hydrogen atoms.[188]

It is now generally accepted that aromatic amino-acid residues, particularly tryptophan and to a lesser extent tyrosine and the other aromatic amino-acids, play a very important role in the photochemistry of proteins, and

[202] B. M. Weeks, S. A. Cole, and W. M. Garrison, *J. Phys. Chem.*, 1965, **69**, 4131.
[203] W. A. Volkert and R. R. Kuntz, *J. Phys. Chem.*, 1968, **72**, 3394.
[204] W. F. Forbes and P. D. Sullivan, *Canad. J. Biochem.*, 1967, **45**, 1831.

this has stimulated considerable interest in the photochemical behaviour of tyrosine and tryptophan.[205-212] It has recently been shown that the photochemical degradation of cystine is sensitised by both tryptophan and tyrosine in aqueous solution, providing a model for the inactivation of cystine-containing enzymes by u.v. light,[213] since there is evidence that quanta primarily absorbed by neighbouring aromatic chromophores contribute to the destruction of cystine residues.

5 Analytical Methods

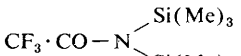
At the present time, roughly a quarter of the papers on amino-acids which appear are devoted to analytical methods. A proportional amount of space will not be allocated to these studies here, however, as most of them are either concerned with situations which are unlikely to interest many readers (*e.g.* amino-acids in biological fluids) or describe minor modifications of established techniques. In any case, the essence of a new or improved analytical procedure is embodied in the practical detail, and it would be inappropriate to present this here. In the ensuing outline, therefore, the majority of publications in this area will be cited without discussion, and only a few advances of general interest will be dealt with more fully.

A book on analytical methods in amino-acid chemistry has been published:[214] the theory and practice of most of the important quantitative methods are covered.

A. Gas–Liquid Chromatography.—The preparation of suitable derivatives of amino-acids and their analysis by g.l.c. has been reviewed recently.[215] Further studies on the use of *N*-trifluoroacetylamino-esters have been described[216-218] and a g.l.c. system (using *N*-trifluoroacetylamino-acid n-butyl esters) which permits quantitative analysis (in 55 min.) of mixtures containing all the protein amino-acids has been developed.[218] This promising new method was tested by analysis of a ribonuclease hydrolysate, which gave results in excellent agreement with those obtained by ion-exchange chromatography.

[205] Yu. A. Vladimirov and E. E. Fesenko, *Photochem. and Photobiol.*, 1968, **8**, 209.
[206] B. Rabinovitch, *Arch. Biochem. Biophys.*, 1968, **124**, 258.
[207] E. E. Fesenko and D. I. Roshchupkin, *Zhur. priklad. Spectroskopii*, 1968, **8**, 834.
[208] H. Hase, *J. Phys. Soc. Japan*, 1968, **24**, 223.
[209] R. F. Chen, *Analyt. Letters*, 1967, **1**, 35.
[210] I. I. Sapezhinskii, *Biofizika*, 1968, **13**, 517.
[211] R. Santus, C. Helene, and M. Ptak, *Photochem. and Photobiol.*, 1968, **7**, 341.
[212] H. B. Steen, *Photochem. and Photobiol.*, 1968, **8**, 47.
[213] K. Dose, *Photochem. and Photobiol.*, 1968, **8**, 331.
[214] S. Blackburn, 'Amino-acid Determination—Methods and Techniques', Edward Arnold, London, 1968.
[215] C. W. Gehrke and D. L. Stalling, *Separation Sci.*, 1968, **2**, 101; K. Blau, in 'Biomedical Applications of Gas Chromatography', vol. 2, ed. H. A. Szymanski, Plenum Press, 1968, p. 1.
[216] C. Landault and G. Guichon, *Bull. Soc. chim. France*, 1967, 3985.
[217] A. Darbre and A. Islam, *Biochem. J.*, 1968, **106**, 923.
[218] C. W. Gehrke, R. W. Zumwalt, and L. L. Wall, *J. Chromatog.*, 1968, **37**, 398.

Pertrimethylsilyl derivatives have been employed for g.l.c. analysis of iodo-amino-acids,[219-221] sulphur-containing amino-acids,[222, 223] and seleno-amino-acids.[223] The new silylating reagent (58) has been prepared and used



(58)

for the chemical modification of amino-acids for g.l.c. purposes:[224] the co-products which arise from use of this reagent interfere with the g.l.c. analysis less than those from more familiar reagents such as *N,O*-bis(trimethylsilyl)acetamide.

Condensation of the guanidine side-chain of arginine with acetylacetone gives an N^{δ}-(2-pyrimidinyl)ornithine derivative which is suitable for g.l.c. and mass spectrometric work.[225] Other derivatives which have been recommended for g.l.c. work include *N*-acetylamino-acid n-propyl esters,[226] α-isothiocyanato-esters [227] [(59): these are easily obtained from the corresponding α-amino-esters by successive treatment with carbon disulphide and methyl chloroformate] and *N*-alkylamino-esters.[228] Derivatives suitable for the separation of the optical isomers of some protein amino-acids with functional side-chains have been reported.[229]

(59) (60)

The use of phosphorus-containing modifying groups enables very small amounts of amino-acids to be detected by g.l.c., using a modified alkali flame detector.[230] The minimum detectable limit (response : noise ratio of 2 : 1) for the alanine derivative (60) was *ca.* 5×10^{-12} g.

B. Ion-exchange Chromatography.—2,4,6-Trinitrobenzenesulphonic acid can be used with automatic amino-acid analysers for quantitative determinations,[231] and the same reagent can be used for simultaneous identification and determination of amino-acids and hexosamines, using the fact

[219] N. M. Alexander and R. Scheig, *Analyt. Biochem.*, 1968, **22**, 187.
[220] L. B. Hansen, *Analyt. Chem.*, 1968, **40**, 1587.
[221] E. T. Backer and V. G. Pileggi, *J. Chromatog.*, 1968, **36**, 351.
[222] F. Shahrokhi and C. W. Gehrke, *J. Chromatog.*, 1968, **36**, 31.
[223] K. A. Caldwell and A. L. Tappel, *J. Chromatog.*, 1968, **32**, 635.
[224] D. L. Stalling, C. W. Gehrke, and R. W. Zumwalt, *Biochem. Biophys. Res. Comm.*, 1968, **31**, 616.
[225] H. Vetter-Diechtl, W. Vetter, W. Richter, and K. Biemann, *Experientia*, 1968, **24**, 340.
[226] J. R. Coulter and C. S. Hahn, *J. Chromatog.*, 1968, **36**, 42.
[227] B. Halpern, V. A. Close, A. Wegmann, and J. W. Westley, *Tetrahedron Letters*, 1968, 3119.
[228] J. W. Davies jun. and A. Furst, *Analyt. Chem.*, 1968, **40**, 1910.
[229] G. E. Pollock and A. H. Kawauchi, *Analyt. Chem.*, 1968, **40**, 1356.
[230] G. Ertingshausen, C. W. Gehrke, and W. A. Aue, *Separation Sci.*, 1967, **2**, 681.
[231] J. Harmeyer, H.-P. Sallmann, and L. Ayoub, *J. Chromatog.*, 1968, **32**, 258.

that the ratio of the optical density at 355 nm : 475 nm is >2 for trinitrophenylamino-acids but <0·5 for trinitrophenylhexosamines.[232] A system has been developed for the use of Rosen's ninhydrin solution with the Beckman-Spinco amino-acid analyser.[233] When a problem calls for amino-acid analysis and determination of ^{14}C distribution among the amino-acids the effluent containing the ninhydrin reaction products can be counted.[234] A modified procedure for the resolution of mixtures containing an unusual complement of basic amino-acids (*e.g.* elastin hydrolysates) on an amino-acid analyser has been described.[235] Interest in the use of lithium buffers with automatic amino-acid analysers continues, with particular emphasis on the possibilities for obtaining improved separations in analyses of physiological fluids.[236, 237] The effects of the presence of heavy metal cations (Fe^{3+}, Cu^{2+}, Ni^{2+}) on the results of amino-acid analysis by ion-exchange chromatography have been investigated: iron affects the extinction coefficients of the ninhydrin reaction products of several amino-acids, and interferes seriously with the determination of cysteic acid.[238] If methanol-containing buffers are used in ion-exchange chromatography to improve the resolution of threonine and serine, there is a danger of loss of aspartic and glutamic acids by esterification.[239, 240] The substitution of t-butanol for methanol avoids this difficulty and causes no decrease in the resolution of threonine and serine.[240] Preliminary results on the use of an amphoteric modified cellulose for ion-exchange chromatography have been reported.[241]

C. Thin-layer Chromatography.—The t.l.c. of amino-acids has been reviewed by Pataki:[242] the invaluable handbook [243] by the same author on the subject is no doubt well known to most readers.

A new method for the detection of amino-acid derivatives with aromatic groups on thin-layer chromatograms uses the fact that such aromatic groups can act as electron donors and form coloured charge-transfer complexes with acceptors such as chloranil and 2,4,7-trinitrofluorenone. An important point about this method of detection is its non-destructive nature: the components originally applied to the plate can be recovered by extraction with an aqueous solvent in which the charge-transfer acceptor is insoluble.[244] Tyrosine derivatives can be detected on t.l.c. plates by

[232] P. C. Kelleher and C. J. Smith, *J. Chromatog.*, 1968, **34**, 7.
[233] W. S. Knight, *Analyt. Biochem.*, 1968, **22**, 539.
[234] A. C. Olson, L. M. White, and A. T. Noma, *Analyt. Biochem.*, 1968, **24**, 120.
[235] B. C. Starcher, *J. Chromatog.*, 1968, **38**, 293.
[236] J. H. Peters, B. J. Berridge jun., J. G. Cummings, and S. C. Lin, *Analyt. Biochem.*, 1968, **23**, 459.
[237] T. L. Perry, D. Stedman, and S. Hansen, *J. Chromatog.*, 1968, **38**, 460.
[238] A. Rudolph, *J. Chromatog.*, 1967, **31**, 479.
[239] L. W. Nauman and W. Galster, *J. Chromatog.*, 1968, **34**, 102.
[240] L. W. Nauman, *J. Chromatog.*, 1968, **36**, 398.
[241] G. Manecke and P. Gergs, *J. Chromatog.*, 1968, **34**, 125.
[242] G. Patakai, *Chromatog. Rev.*, 1967, **9**, 23.
[243] G. Patakai, 'Dünnschichtchromatographie in der Aminosäure- und Peptid-Chemie', Walter de Gruyter, Berlin, 1966.
[244] Y. Burstein, M. Fridkin, and M. Shinitzky, *Biochim. Biophys. Acta*, 1968, **160**, 141.

irradiation with u.v. light, which causes formation of brown pigments.[245] Other papers on the t.l.c. of amino-acids have described studies on the separation of sulphur-containing amino acids,[246] detection with variamine blue and cupric acetate,[247] applications of circular t.l.c.,[248] quantitative determination of amino-acids based on spot size in two-dimensional t.l.c.,[249] the comparative efficiency of different kinds of cellulose for t.l.c. separations,[250] analysis of amino-acid mixtures containing large amounts of glycerol,[251] and the t.l.c. analysis of amino-acids in physiological fluids.[252-254]

D. Other Methods.—Two procedures for the determination of D-amino-acids using a D-amino-acid acetyltransferase (AAT) from yeast have been described.[255] In one method (which was investigated for amounts of D-amino-acids of the order 1 μmole) the D-amino-acid is incubated with AAT, coenzyme A (CoA), excess acetyl phosphate, and phosphotransacetylase (PTA) at pH 7·6 and 30°, when the reactions shown in Scheme 10

$$\text{CoA} + \text{Acetyl phosphate} \xrightarrow{\text{PTA}} \text{Ac·CoA} + \text{Phosphate}$$

$$\text{Ac·CoA} + \text{D-Amino-acid} \xrightarrow{\text{AAT}} \text{Ac-D-Amino-acid} + \text{CoA}$$

Scheme 10

occur. The extent of acetylation, and hence the amount of D-amino-acid, is determined by colorimetric estimation (hydroxamic acid reaction) of the acetyl phosphate which is consumed. The errors of the method are about $\pm 1\%$. The second procedure involves enzymic acetylation with radioactive acetyl-CoA. AAT is completely stereospecific and has a very broad substrate specificity; it catalysed the acetylation of 44 of the 47 D-amino-acids tested.

Many other topics in the analytical chemistry of amino-acids have been discussed, including: electrophoresis on paper [256, 257] and thin layers,[258, 259] the use of papers impregnated with liquid ion-exchangers,[260] the comparative efficiency of several chromatographic papers for amino-acid separation and determination of basic amino-acids with mercuric salt

[245] A. Ber and L. Wasserman, *Experientia*, 1968, **24**, 224.
[246] V. R. Villaneuva and M. Barbier, *Bull. Soc. chim. France*, 1967, 3992.
[247] B. Duk, Z. Kwapniewski, and J. Sliwiok, *Chem. analit.*, 1968, **13**, 193.
[248] M. M. Hashmi, A. S. Adil, N. A. Chughtai, F. R. Chugtai, and M. A. Shahid, *Mikrochim. Acta*, 1968, 291.
[249] E. Bancher, J. Washüttl, and M. D. Olfat, *Mikrochim. Acta*, 1968, 773.
[250] C. L. de Ligny and E. C. M. Kok, *J. Chromatog.*, 1968, **38**, 224.
[251] E. J. Shellard and G. H. Jolliffe, *J. Chromatog.*, 1967, **31**, 82.
[252] R. B. Meffered jun., R. M. Summers, and J. G. Fernandez, *Analyt. Letters*, 1968, **1**, 279.
[253] H. H. White, *Clinica Chim. Acta*, 1968, **21**, 297.
[254] E. Plöchl, *Clinica Chim. Acta*, 1968, **21**, 271.
[255] J. H. Schmitt and M. H. Zenk, *Analyt. Biochem.*, 1968, **23**, 433.
[256] J. L. Frahn and J. A. Mills, *Analyt. Biochem.*, 1968, **23**, 546.
[257] P. J. Peterson, *J. Chromatog.*, 1968, **38**, 301.
[258] R. L. Munier, C. Thommegay, and G. Sarrazin, *Bull. Soc. chim. France*, 1967, 397.
[259] J. Chudzik and A. Klein, *J. Chromatog.*, 1968, **36**, 262.
[260] E. Soczewinski and M. Rojowska, *J. Chromatog.*, 1968, **32**, 364.

precipitants,[261, 262] fluorimetric determination using butan-2,4-dione,[263] potentiometric determination of sulphur-containing amino-acids in the presence of ascorbic acid,[264] polarographic determination,[265] and titrimetric estimation in non-aqueous media.[266]

E. Determination of Specific Amino-acids.—Papers on the determination of the following amino-acids have appeared: glycine,[267] L-leucine and DL-valine,[268] ornithine,[269] γ-aminobutyric acid,[269, 270] δ-aminolaevulinic acid,[271] available lysine in foods,[272, 273] ε-aminocaproic acid,[274] tryptophan,[275–277] cystine,[278, 279] lanthionine,[280, 281] lysinoalanine,[281] hydroxyproline,[282] phenylalanine,[283, 284] and tyrosine.[284, 285]

[261] F. Kai, *Bull. Chem. Soc. Japan*, 1968, **41**, 875.
[262] F. Kai, *Bull. Chem. Soc. Japan*, 1967, **40**, 2297.
[263] E. Sawicki and R. A. Carnes, *Analyt. Chim. Acta*, 1968, **41**, 178.
[264] N. Santi and E. Peillon, *Ann. pharm. franç.*, 1968, **26**, 177.
[265] B. P. Zhantalai and Ya. I. Tur'yan, *Zhur. analit. Khim.*, 1968, **23**, 282.
[266] G. M. Gal'pern, V. A. Il'ina, L. P. Petrova, and F. D. Sidel'kovskaya, *Zavodskaya Lab.*, 1968, **34**, 416.
[267] A. Masood and O. C. Saxena, *Microchem. J.*, 1968, **13**, 178.
[268] O. C. Saxena, *Microchem. J.*, 1968, **13**, 321.
[269] K. Molitoris, *J. Chromatog.*, 1968, **34**, 399.
[270] N. Seiler and M. Wiechmann, *Z. physiol. Chem.*, 1968, **349**, 588.
[271] J. J. Chisolm jun., *Analyt. Biochem.*, 1968, **22**, 54.
[272] L. Blom, P. Hendricks, and J. Caris, *Analyt. Biochem.*, 1967, **21**, 382.
[273] P. Hocquelet, *Ann. Fals. et Expertise chim.*, 1968, **61**, 155.
[274] S. Simard-Savoie, L. M. Breton, and M. Beaulieu, *J. Chromatog.*, 1968, **38**, 143.
[275] J. Cegarra and J. Gacén, *J. Soc. Dyers and Colourists*, 1968, **84**, 216.
[276] M. E. Rio and J. C. Sanahuja, *Clinical Chem.*, 1968, **14**, 429.
[277] G. Sternkopf, *Nahrung*, 1968, **12**, 75.
[278] J. A. Schneider, K. H. Bradley, and J. E. Seegmiller, *Analyt. Biochem.*, 1968, **23**, 129.
[279] M. Wronski and W. Goworek, *Chem. analit.*, 1968, **13**, 197.
[280] M. Marzona and G. Di Modica, *J. Chromatog.*, 1968, **32**, 755.
[281] A. Robson, M. J. Williams, and J. M. Woodhouse, *J. Chromatog.*, 1967, **31**, 284.
[282] H. Stegemann and K.-H. Stalder, *Clinica Chim. Acta.*, 1967, **18**, 267.
[283] H. K. Berry, *Clinica Chim. Acta.*, 1968, **20**, 299.
[284] K. Uchiyama, H. Yamada, T. Tochikura, and K. Ogata, *Agric. and Biol. Chem. (Japan)*, 1968, **32**, 764.
[285] B. G. Searle, M. Li, J. Briggs, P. Segall, D. Widelock, and B. Davidow, *Clinical Chem.*, 1968, **14**, 623.

2

Structural Investigation of Peptides and Proteins

BY R. N. PERHAM, P. M. HARDY, C. C. F. BLAKE

PART I: Primary Structures and Chemical Modification

by R. N. Perham

1 Introduction

The purpose of the present article is to review advances in the determination of the primary structure of proteins and the knowledge gained therefrom. Most of the results discussed have been reported during approximately the past 18 months, but, in order that these results may be seen within their proper context, some earlier literature has been included. Where recent authoritative reviews of particular subjects exist, these have been made the starting-point for the discussion provided here, reiterating detail only where it is necessary for coherence. However, the opportunity has been taken to collect together some recently published amino-acid sequences so that they are readily to hand. Many more are now available in a most enterprising annual volume which is intended to be a repository of the cumulative knowledge of the primary structure of proteins and nucleic acids.[1] Some selection has had to be made from among the growing flood of sequence information from so many laboratories, but it is hoped that the choice of topics is sufficiently eclectic to illustrate the major growing points of the subject. Given this need for some selection, it is further hoped that protein chemists who find significant work of theirs omitted will be free in their forgiveness; that they will feel, with Junius, 'the injustice done to an individual is sometimes of service to the public'.[2]

For convenience in presentation it has been found useful to separate a discussion of the methodology of sequence determination from a consideration of the results obtained. Inevitably this distinction cannot wholly be adhered to, since desire to acquire certain structural information about particular proteins frequently leads to the development of some special technique to obtain it. This technique, in turn, then finds more generalised applications and the results so obtained are perhaps best considered as examples of the technique.

It will be observed that the methods of amino-acid sequence determination are currently undergoing rapid development and that larger and more

[1] M. O. Dayhoff and R. V. Eck, 'Atlas of Protein Sequence and Structure,' National Biomedical Research Foundation, Maryland, 1969.

[2] 'Letters of Junius', Letter 41, November 14, 1770.

complicated proteins are coming in for attention at the level of primary structure. If the word 'automation' occurs at frequent intervals, that only reflects the general tenor of the subject.

2 The Methodology of Amino-acid Sequence Determination

In 1968 the great majority of amino-acid sequence investigations were still being conducted along the lines pioneered 15 years before by Sanger in his classic study of insulin; that is, partial hydrolysis of the polypeptide chain by various methods, enzymic and chemical, and characterisation of the resulting peptides. This in itself need not surprise us since those methods themselves represented a major breakthrough in thinking about amino-acid sequence determination and, indeed, are currently demonstrating their worth in nucleic acid chemistry.[3] The past few years have, however, seen the introduction of several new techniques that may eventually alter significantly the protein chemist's approach. Among these we may number automated sequential degradation of the polypeptide chain and mass spectrometry. In both cases it is the prospect of automated sequence determination that is most appealing; with the courage to tackle larger and larger proteins has come the realisation of the enormous investment of the time of skilled research workers that this requires. Both techniques have certain technical limitations that may or may not be overcome but the auguries, at least, are encouraging.

The other great aim of amino-acid sequence determination is to increase the sensitivity of the methods to enable useful work to be undertaken on material that can be obtained only in small amounts. The situation here is reasonably encouraging also, with steady progress in both amino-acid and sequence analysis at high sensitivity. Improvements have also been made in what might be called specialised techniques, for example methods for studying amino-acid sequence in selected regions of a protein molecule, which have had some considerable impact on the study of amino-acid sequence homology in related proteins.

Most of the methods of sequence determination have recently been described *in extenso* [4] and comment here will largely be confined to developments since the compilation of that invaluable volume, which is likely to serve as a definitive source of detail for some time to come. A less comprehensive but none the less useful book has also been brought more up to date with the publication of its second edition.[5]

A. Amino-acid Analysis.—The usual preliminary to most conventional (and projected) approaches to amino-acid sequence determination is

[3] G. G. Brownlee, F. Sanger, and B. G. Barrell, *J. Mol. Biol.*, 1968, **34**, 379.
[4] 'Methods in Enzymology,' ed. S. P. Colowick and N. O. Kaplan; vol. 11, ed. C. H. W. Hirs, Academic Press, New York, 1967.
[5] J. Legget Bailey, 'Techniques in Protein Chemistry,' 2nd edn., Elsevier, Amsterdam and London, 1967.

amino-acid analysis. Methods, both chemical and enzymic, for achieving the hydrolysis of a protein into its constituent amino-acids have been covered in a recent review [6] and need not be repeated here. Humin, which frequently forms during the acid hydrolysis of proteins contaminated with plant material, may be removed by passing the hydrolysate through a small ion-exchange column which retains only the humin.[7] Many methods are available for the analysis of amino-acid mixtures and these will be considered in turn.

Ion-exchange Chromatography. The introduction of automatic amino-acid analysis on ion-exchange columns [8] was one of the milestones of protein chemistry. Since that time, the machinery has been much improved, both with respect to speed of operation and to sensitivity.[6, 9] Accurate analysis of 5 nmoles or less of amino-acid is now possible in some 2–3 h., with even shorter times in prospect. The availability of such rapid analysis, together with the desire to operate expensive machinery round the clock, has led to the development of several devices that enable samples for analysis to be applied automatically to ion-exchange columns which are themselves regenerated automatically.[10]

The logical extension of rapid, automatic production of data is, of course, its automatic computation. Two approaches are possible here. In the first, a suitable integrating device detects the emergence of a peak from the analyser and computes its area. The calculation of the amount of any given amino-acid from the peak area is then done by hand. Such integrators are made by Infotronics Corp., Houston, Texas, and by Beckman/Spinco Division, Palo Alto, California. In the second approach, the output of the analyser recorder is converted into a form suitable for processing in a high-speed digital computer. Details of several such systems have been published.[11] These efforts have also been extended to the provision of programmes for the direct calculation of the amino-acid composition of a peptide or protein.[11b, 12]

An improved ninhydrin reagent, incorporating cyanide, has been suggested.[13] It is claimed to result in improved stability of the ninhydrin reagent towards light and air. Moore, however, mindful of the health hazard posed by the methyl cellosolve conventionally used to dissolve the ninhydrin, has proposed a new ninhydrin reagent in which the methyl

[6] R. L. Hill, *Adv. Protein Chem.*, 1965, **20**, 37.
[7] I. Larsen and J. V. Mortensen, *Analyt. Biochem.*, 1967, **21**, 466.
[8] D. H. Spackman, W. H. Stein, and S. Moore, *Analyt. Chem.*, 1958, **30**, 1190.
[9] D. H. Spackman, *Methods in Enzymology*, 1967, **11**, 3.
[10a] J. W. Eveleigh and A. R. Thomson, *Biochem. J.*, 1966, **99**, 49P. [b] K. Dus, S. Lindroth, R. Pabst, and R. M. Smith, *Analyt. Biochem.*, 1967, **18**, 532. [c] P. Slump and A. M. Verbeek, *J. Chromatog.*, 1968, **34**, 401.
[11a] A. Yonda, D. L. Filmer, H. Pate, N. Alonzo, and C. H. W. Hirs, *Analyt. Biochem.*, 1965, **10**, 53. [b] W. C. Starbuck, C. M. Mauritzen, C. McClimans, and H. Busch, *ibid.*, 1967, **20**, 439.
[12] K. Ozawa and S. Tanaka, *Analyt. Biochem.*, 1967, **24**, 270.
[13] W. S. Knight, *Analyt. Biochem.*, 1968, **22**, 539.

cellosolve is replaced by dimethyl sulphoxide.[14] It seems that when this is done the sodium acetate buffer also used in the ninhydrin reagent should be replaced by a lithium acetate buffer. The resulting reagent is more stable and less toxic. As an alternative to ninhydrin entirely, trinitrobenzenesulphonic acid has been used to detect amino-acids in the effluent of ion-exchange columns.[15] The yellow colour is developed at room temperature and while the sensitivity is only about half that of ninhydrin, the reproducibility is as good if not better. A technique has also been described [16] for the scintillation counting of radioactive amino-acids in the analyser effluent after the ninhydrin reaction. Because of the loss of the amino-acid carboxyl group in that reaction, the radioactive label clearly must be elsewhere in the molecule.

High-voltage Paper Electrophoresis and Thin-layer Chromatography. Separation of amino-acids by high-voltage paper electrophoresis has been dealt with elsewhere.[5, 17] The method has the advantages of cheapness and simplicity, and rapid qualitative analysis can be achieved with as little as 1 nmole of amino-acid. A new cooled-plate apparatus for two-dimensional electrophoresis has been reported.[18] Thin-layer chromatography has also been the subject of extensive recent reviews.[19] Two-dimensional separation of amino-acids by electrophoresis and chromatography on thin-layer plates has been described in detail.[20]

Neither paper electrophoresis nor thin-layer chromatography can challenge ion-exchange chromatography for precision analysis: they are, none the less, extraordinarily useful for qualitative, or even semi-quantitative, analysis. The facility with which autoradiographs of the separation can be prepared by placing in contact with *X*-ray film makes them especially useful for the identification of radioactive amino-acids in a mixture.

Special Techniques for Tryptophan. Because of its destruction during acid hydrolysis, determination of the tryptophan content of proteins has long been a source of problems. Many techniques, spectrophotometric and chemical, have been developed for the purpose.[5] A new spectrophotometric method, based on the absorbance of the protein at 280 and 288 nm in buffers containing 6M-guanidine hydrochloride, has been reported.[21] The contribution from cystine, if any, can be allowed for. The spectrophotometric determination of tryptophan following its oxidation with *N*-bromosuccinimide (NBS) has been described in detail by Spande and Witkop.[22]

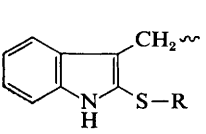
[14] S. Moore, *J. Biol. Chem.*, 1968, **243**, 6281.
[15] J. Harmeyer, H.-P. Sallmann, and L. Ayoub, *J. Chromatog.*, 1968, **32**, 258.
[16] A. C. Olson, L. M. White, and A. T. Noma, *Analyt. Biochem.*, 1968, **24**, 120.
[17a] W. J. Dreyer and E. Bynum, *Methods in Enzymology*, 1967, **11**, 32. [b] C. J. O. R. Morris and P. Morris, 'Separation Methods in Biochemistry,' Pitman, London, 1964.
[18] J. H. Buchanan and M. C. Corfield, *J. Chromatog.*, 1967, **31**, 274.
[19a] M. Brenner and A. Niederwieser, *Methods in Enzymology*, 1967, **11**, 39. [b] G. Pataki, *Chromatog. Rev.*, 1967, **9**, 23.
[20] R. L. Munier, C. Thommegay, and G. Sarrazin, *Bull. Soc. chim. France*, 1967, 3935.
[21] H. Edelhoch, *Biochemistry*, 1967, **6**, 1948.
[22] T. F. Spande and B. Witkop, *Methods in Enzymology*, 1967, **11**, 498.

A study of the reaction of NBS with several proteins [23] indicated that partial reaction of the tryptophan was occurring owing to the fact that some tryptophan residues are buried and some exposed in the native protein. A recent report [24] where the oxidation of lysozyme was evaluated by chemical, as well as spectral, analysis has suggested that it might not be quite as simple as that. Extensive destruction of tyrosine and histidine is found at reagent concentrations where spectral assay indicates incomplete reaction with tryptophan. After alkaline hydrolysis, tryptophan analysis shows higher levels of tryptophan oxidation than calculated from the spectral data. Further, no pH dependence for oxidation is evident when assayed this way. It is suggested that the apparent differential reactivity of tryptophan residues in lysozyme may be an artefact caused by concomitant oxidation of tyrosine confusing the spectral assay. In any event, more than the usual care would seem necessary in studying the oxidation of tryptophan by *N*-bromosuccinimide. The determination of tryptophan and its photo-oxidation products in lysozyme has also been discussed.[25] The reaction of NBS with *N*-terminal tyrosine residues in peptides to form the chromophore of 5,7-dibromo-6-hydroxyindole-2-carboxamide has been suggested as a spectral method to estimate *N*-terminal tyrosine.[26]

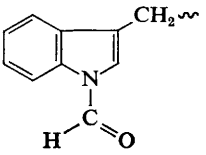
Various other chemical modifications of tryptophan that can be used for spectral analysis have been reported. 2-Hydroxy-5-nitrobenzyl bromide reacts with tryptophan to yield several isomeric monosubstituted derivatives of the indole ring.[27] The reaction can be followed at 410 nm and the yellow colour imparted to the modified tryptophan residue is an aid in isolating tryptophan-containing peptides from protein digests. A study of the tryptophan residues in porcine pepsin has been made by this method.[28] Approximately two tryptophans are found to react in native pepsin, to yield a derivative which is still enzymically active, indicating the non-essential nature of the two side-chains modified. On the other hand, a total of four tryptophans react in reduced and carboxymethylated pepsin and pepsinogen, in accord with the total tryptophan content by other methods (with the exception of spectral analysis which may be a little high) and the four unique sequences containing modified tryptophan residues isolated from proteolytic digests of the reacted enzyme. The use of sulphenyl halides as modifying agents for tryptophan residues has also been commended.[29] Both 2-nitro- and 2,4-dinitrophenylsulphenyl chloride are described. The reaction is carried out in 30–50% acetic acid, under which conditions tryptophan is converted into a derivative with a thioether

[23] T. F. Spande and B. Witkop, *Methods in Enzymology*, 1967, **11**, 528.
[24] M. J. Kronman, F. M. Robbins, and R. E. Andreotti, *Biochim. Biophys. Acta*, 1967, **147**, 462.
[25] V. K. Lapuk, L. A. Chistyakova, and N. A. Kravchenko, *Analyt. Biochem.*, 1968, **24**, 80.
[26] M. Wilchek, T. Spande, and B. Witkop, *Biochemistry*, 1968, **7**, 1787.
[27] T. E. Barman and D. E. Koshland jun., *J. Biol. Chem.*, 1967, **242**, 5771.
[28] T. A. A. Dopheide and W. M. Jones, *J. Biol. Chem.*, 1968, **243**, 3906.
[29] E. Scoffone, A. Fontana, and R. Rocchi, *Biochemistry*, 1968, **7**, 971.

function at the 2-position of the indole nucleus (1); cysteine forms a mixed disulphide which can subsequently be cleaved with 2-mercaptoethanol. The reaction can be followed quantitatively at 365 nm ($\epsilon = 4000$). Bovine ribonuclease, which contains no tryptophan or cysteine, is found, appropriately, not to react with these compounds.



(1)



(2)

To complement these irreversible modifications of the tryptophan residue, a reversible modification of tryptophan has been reported.[30] The protein is allowed to react with anhydrous formic acid saturated with hydrogen chloride for up to 40 min. at room temperature. During this time, tryptophan is converted to 1-formyltryptophan (2) and the reaction can be measured by the increase in absorbance at 298 nm. The hydroxyl groups of serine, threonine, and tyrosine do not react; neither do the protein amino-groups, possibly because they are protected by protonation. Lysozyme, chymotrypsinogen A, and the α- and β-chains of haemoglobin are all reported to react specifically in better than 90% yield. At pH values >9 the tryptophan is rapidly and quantitatively regenerated.

Despite these improvements, tryptophan is still not an easy amino-acid to estimate in proteins. In view of its possible role in the catalytic site of some enzymes, *e.g.* dehydrogenases,[31] it is often crucial that its analysis should be entirely reliable.[32] In common with most other residues in proteins, it may well be that the only sure way of achieving this is to establish the amino-acid sequence of the protein concerned.

Gas–Liquid Chromatography. Analysis of amino-acids by gas–liquid chromatography has been covered recently in several reviews.[33] Weinstein [33a] concluded 'the qualitative goal of gas–liquid chromatography of amino-acids has been partially achieved; a truly quantitative method using this technique remains to be developed over the next decade'. Much effort has been put into that development and the pace is increasing.

In its favour, g.l.c. is rapid and very sensitive; an analysis might take *ca.* 20 min. instead of *ca.* 2 hr. currently required by ion-exchange chromatography and can be done on 10^{-11} to 10^{-12} mole of amino-acid. The apparatus

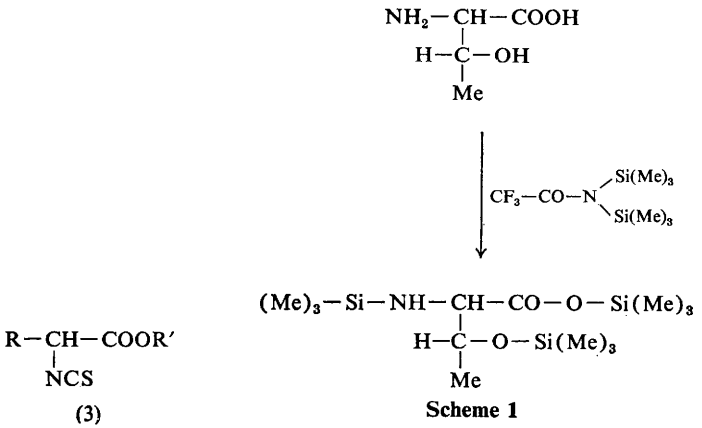
[30] A. Previero, M.-A. Coletti-Previero, and J.-C. Cavadore, *Biochim. Biophys. Acta*, 1967, **147**, 453.
[31] K. A. Schellenberg, *J. Biol. Chem.*, 1965, **240**, 1165.
[32] G. B. Kitto and N. O. Kaplan, *Biochemistry*, 1966, **5**, 3966.
[33a] B. Weinstein, *Methods Biochem. Analysis*, 1966, **14**, 203. [b] C. Landault and G. Guichon, *Bull. Soc. chim. France*, 1967, 3985.

is simple, with a minimum of moving parts, and the resolution is such that even optical isomers can be separated on an asymmetric support. Set against this is the unfortunate fact that amino-acids are not volatile substances and must therefore be converted into some suitable volatile derivative. Clearly, these derivatives must be capable of simple and quantitative preparation with a minimum of manipulation and it is on this aspect of the problem that much research is currently centred.

Among the most widely used are the *N*-trifluoroacetyl derivatives of the amino-acid n-butyl esters.[34] The methyl esters are potentially better because they are more easily prepared and more volatile, which means chromatography may be effected at a lower temperature. On the other hand, this increased volatility can cause associated losses and for this reason most people have chosen to work with higher esters. However, successful chromatography of the methyl esters prepared by a modified method has recently been described.[35] Temperature programming was advocated. A study has also been made of the *N*-trifluoroacetyl methyl esters of methionine, methionine sulphoxide, and methionine sulphone, with the column eluant passed directly into the ion source of a mass spectrometer.[36] It is worth noting that the *N*-heptafluorobutyryl derivatives of amino-acids have significantly shorter gas-chromatographic retention times than the trifluoroacetyl derivatives.[37]

A simple two-step reaction procedure to convert all amino-acids except arginine and histidine into the corresponding *N*-acetyl propyl esters has been outlined.[38] The procedures devised to deal with arginine and histidines repay study: they typify the ingenuity that needs to be brought to bear in these matters. Arginine is converted into ornithine with the enzyme arginase, and ornithine may be treated in the same way as the other amino-acids. Histidine can be converted into aspartic acid by ozonolysis. Analysis of the aspartic acid before and after ozonolysis therefore gives the histidine content. A preliminary account of the use of alkylidene and alkyl amino-acid esters has been given,[39] which may offer some advantages with further development. Chromatography of *N*-thiocarbonyl ester derivatives (3) would also appear to be promising.[40] The molecular weight of the amino-acid is increased less in forming this derivative than with the *N*-trifluoroacetyl butyl esters and the product is claimed to be more stable and more volatile. Unfortunately, there are problems with arginine and the reaction scheme for making the derivatives involves several steps.

[34] C. W. Gehrke, W. M. Lamkin, D. L. Stalling, and F. Shahrokhi, *Biochem. Biophys. Res. Comm.*, 1965, **19**, 328.
[35] A. Darbre and A. Islam, *Biochem. J.*, 1968, **106**, 923.
[36] S. R. Tannenbaum, W. G. Thilly, and P. Issenberg, *Analyt. Chem.*, 1968, **40**, 1723.
[37] G. E. Pollock, *Analyt. Chem.*, 1967, **39**, 1194.
[38] J. R. Coulter and C. S. Hann, *J. Chromatog.*, 1968, **36**, 42.
[39] J. W. Davis jun. and A. Furst, *Analyt. Chem.*, 1968, **40**, 1910.
[40] B. Halpern, V. A. Close, A. Wegmann, and J. W. Westley, *Tetrahedron Letters*, 1968, 3119.



Scheme 1

Increasing use is being made of trimethylsilylated amino-acids for g.l.c. Chromatography of trimethylsilylated amino-acids was reported [41] where the trimethylsilylating agent used was bis(trimethylsilyl) acetamide. This procedure has been improved [42] by the introduction of bis(trimethylsilyl) trifluoroacetamide to form the derivatives (Scheme 1). An obvious advantage is that the derivative is formed in one step; a disadvantage is that arginine once again resists analysis. The method has been extended to cover several iodine- [43] and sulphur-containing [44] amino-acids.

So far, the number of papers describing methods far exceeds the number describing applications in amino-acid sequence work. This may be due to a number of reasons. First, no g.l.c. method yet described is perfect. Secondly, before g.l.c. analysis the amino-acids must be converted into some form of volatile derivative, with the attendant risk of adventitious losses. Thirdly, the continual development of ion-exchange chromatography towards more rapid and sensitive analysis has kept pace with the improvements in the other techniques of conventional protein chemistry, *e.g.* peptide fractionation and end-group analysis. It is entirely possible that this situation will change in the near future.

B. End-group Analysis.—*N-Terminal Methods.* Most of these techniques have already been given in detail.[4, 5] Improved separations of DNP- [45] and PTH- [46] amino-acids on polyamide thin-layer sheets have been reported and procedures for the repetitive use and cleaning of these sheets have been

[41] J. F. Klebe, H. Finkbeiner, and D. M. White, *J. Amer. Chem. Soc.*, 1966, **88**, 3390.
[42] D. L. Stalling, C. W. Gehrke, and R. W. Zumwalt, *Biochem. Biophys. Res. Comm.*, 1968, **31**, 616.
[43] F. Shahrokhi and C. W. Gehrke, *Analyt. Biochem.*, 1968, **24**, 281.
[44] F. Shahrokhi and C. W. Gehrke, *J. Chromatog.*, 1968, **36**, 31.
[45] K.-T. Wang and I. S. Y. Wang, *J. Chromatog.*, 1967, **27**, 318.
[46] K.-T. Wang, I. S. Y. Wang, A. L. Lin, and C.-S. Wang, *J. Chromatog.*, 1967, **26**, 323.

described.[47] PTH-Amino-acids may be located direct on paper chromatograms with the aid of u.v. light,[48] and i.r. spectra have been used [49] to help to identify PTH-amino-acids separated by t.l.c. on silica gel.[50] As an alternative to direct identification of the PTH-amino-acid, the corresponding amino-acid may be regenerated for subsequent analysis. The hydrolysis conditions required for this have been the subject of a recent careful study.[51] Hydrolysis is best carried out in 0·1N-NaOH for 12 hr. at 120°. Poor recoveries are encountered unless great care is taken to exclude oxygen. Mass spectrometry has been employed for the identification of PTH- [52] and DNP-amino-acids,[53] and g.l.c. has been successfully used for the trimethylsilyl derivatives of PTH-amino-acids (except arginine).[54]

An interesting observation has been made concerning DNP-proteins. It has been shown [55] that the DNP-group may be removed quantitatively from the imidazole, thiol, and phenolic hydroxyl groups of DNP-proteins by treatment at pH 8·0 and 22° with an excess of 2-mercaptoethanol; other thiols can also be used. Imidazole-DNP-histidine appears to be the most labile and under these conditions there is no effect on α- and ε-DNP-amino-groups. As Shaltiel [55] pointed out, these results provide a method for the reversible modification of functional groups and should also be borne in mind when treating DNP-proteins with thiol reagents, *e.g.* to reduce disulphide bridges.

A variant of the Edman method for quantitative *N*-terminal analysis has also been suggested [56] in which the principle of isotope dilution is employed. In this technique, ^{35}S-labelled phenylisothiocyanate is used to form a radioactive phenylthiohydantoin to which the appropriate unlabelled carrier phenylthiohydantoin is added before extraction. The extent of isotope dilution can then be used to calculate the recovery of *N*-terminal group irrespective of losses during the isolation procedure. The method applies only to a protein consisting of a single polypeptide chain or of an aggregate of identical chains.

The dansyl chloride procedure, which has been described in detail,[57, 58] continues to be developed. Several new methods [59, 60] for the separation of

[47] K.-T. Wang and P.-H. Wu, *J. Chromatog.*, 1968, **37**, 353.
[48] H. Pirkle, *Analyt. Biochem.*, 1967, **21**, 472.
[49] M. Murray and G. F. Smith, *Analyt. Chem.*, 1968, **40**, 440.
[50] G. F. Smith and M. Murray, *Analyt. Biochem.*, 1968, **23**, 183.
[51] B. Africa and F. H. Carpenter, *Biochem. Biophys. Res. Comm.*, 1966, **24**, 113.
[52] N. S. Wulfson, V. M. Stepanov, V. A. Puchkov, and M. A. Zyakoon, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1963, 1524.
[53] T. J. Prenders, H. Copier, W. Heerma, G. Dijkstra, and J. F. Arens, *Rec. Trav. chim.*, 1966, **85**, 216.
[54] R. E. Harman, J. L. Patterson, and W. J. A. Vandenheuvel, *Analyt. Biochem.*, 1968, **25**, 452.
[55] S. Shaltiel, *Biochem. Biophys. Res. Comm.*, 1967, **29**, 178.
[56] G. L. Callewaert and C. A. Vernon, *Biochem. J.*, 1968, **107**, 728.
[57] W. R. Gray, *Methods in Enzymology*, 1967, **11**, 139.
[58] C. Gros and B. Labouesse, *European J. Biochem.*, 1969, **7**, 463.
[59] D. Morse and B. L. Horecker, *Analyt. Biochem.*, 1966, **14**, 429.
[60] C. Gros, *Bull. Soc. chim. France*, 1967, 3952.

DNS-amino-acids on thin-layer silica gel have been described, but Gray [57] warns that in his experiments with that technique there was a rapid destruction of dansyl amino-acids adsorbed on to silica. A two-dimensional separation system involving chromatography and electrophoresis on thin-layer plates of cellulose powder has also been reported.[61] Probably the best, and most sensitive, separation is currently achieved on polyamide layer sheets:[62] 10^{-10} mole of DNS-amino-acid can be detected with ease. The mass spectrometer can also be used for the highly sensitive (10^{-11} mole) identification of DNS-amino-acids.[63] They are first converted into the methyl esters to raise volatility and are easily recognised through their nominal molecular weights and the characteristic pattern imposed by the natural abundance of sulphur isotopes. Fragmentation of DNS-isoleucine and DNS-leucine is sufficiently different for identification to be made.

A most promising improvement [64] on the classical 2,4-dinitrofluorobenzene method of Sanger has not attracted much attention. The reagent suggested was 2-chloro-3,5-dinitropyridine, which resembles 2,4-dinitrofluorobenzene in its reaction with proteins. However, during the subsequent acid hydrolysis the presence of the pyridine nitrogen facilitates cleavage of the first peptide bond, enabling shorter times of hydrolysis to be employed and thereby diminishing one of the principal causes of poor end-group recovery in the earlier method. In fact, hydrolysis in 6N-HCl containing 30% formic acid can be achieved in 8–10 hr. at 60°, with reported yields of over 90%. In all other respects, spectrophotometric quantitation of the yellow derivatives *etc.*, the methods are entirely analogous. The 2-chloro-3,5-dinitropyridyl derivatives can also be converted into the corresponding free amino-acids in good yield [65] by hydrolysis with 2N-NaOH at 100°.

The acid hydrolysis of 3- and 5-nitro-2-pyridyl peptides is also assisted by neighbouring group participation [66] involving the heterocyclic nitrogen atom, the ease of hydrolysis paralleling the pK of that atom, which in this case is in the range 1–2. As expected, therefore, the mononitropyridyl peptides are also readily cleaved in dilute acid and this has been made the basis [67] of an even milder assay for *N*-terminal residues. The reagents used are 2-fluoro-3-nitropyridine or 2-fluoro-5-nitropyridine. These compounds are not as reactive as 2-chloro-3,5-dinitropyridine but complete reaction with a protein is reported within 24 hr. at 40°. Release of the *N*-terminal residue is achieved at 100° with 0·1–0·2N-HCl containing 20% formic acid, 30 min. hydrolysis being sufficient for the 3-nitro-derivatives whereas *ca.* 24 hr. are required for the 5-nitro-isomers. A t.l.c. system for the

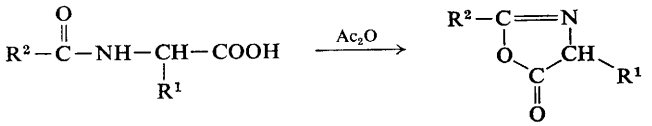
[61] M. S. Arnott and D. N. Ward, *Analyt. Biochem.*, 1967, **21**, 50.
[62] K. R. Woods and K.-T. Wang, *Biochim. Biophys. Acta*, 1967, **133**, 369.
[63] G. Marino and V. Buonocore, *Biochem. J.*, 1968, **110**, 603.
[64] A. Signor, A. Previero, and M. Terbojevich, *Nature*, 1965, **205**, 596.
[65] A. Signor and L. Biondi, *Ricerca sci.*, 1964, **4 (II-A)**, 165.
[66] A. Signor, E. Bordignon, and G. Vidali, *J. Org. Chem.*, 1967, **32**, 1135.
[67] A. Signor, L. Biondi, A. M. Tamburro, and E. Bordignon, *European J. Biochem.*, 1969, **7**, 328.

separation and estimation of the mononitropyridyl amino-acids is available [68] and these methods may find more widespread use for *N*-terminal analysis in the future. The analogous application of 3-fluoro-4-nitropyridine *N*-oxide to the determination of *N*-terminal amino-acids has also been reported.[69]

C-Terminal Methods. Extensive reviews of hydrazinolysis [70] and enzymic hydrolysis with carboxypeptidases [71] have recently appeared. The unfortunate loss of *C*-terminal amino-acid which, sadly, often occurs during hydrazinolysis has now been ascribed [72] predominantly to the formation of the amino-acid hydrazide during the reaction. Catalytic hydrazinolysis at 80°, in which the hydrogen form of ion-exchange resins is added as a source of protons to promote nucleophilic attack on the peptide bonds, may well be much better in this connection.[73] The highly sensitive dansylation procedure has been used to identify the *C*-terminal residue released by hydrazinolysis.[74]

It has been known for many years that peptides react with ammonium thiocyanate and acetic anhydride under relatively mild conditions to form peptidylthiohydantoins at the *C*-terminus. The possibility of using this system for sequential degradation of peptides from the *C*-terminus has recently been re-investigated.[75] Treatment with acetohydroxamate cleaves the peptidylthiohydantoin to yield the thiohydantoin of the *C*-terminal residue and an acetylated peptide with a new *C*-terminal residue. A best degradation of six residues was achieved; more commonly, only *ca.* three steps were possible. The method was found to be limited in that aspartic acid and proline are not removed and incomplete cleavage rapidly caused equivocal results. For identification of the thiohydantoin cleaved off, subtractive analysis of the residual peptide was preferred because the unfortunate instability of the thiohydantoins prevented chromatographic analysis.

The *C*-terminal residue in polypeptides can be cyclised to the oxazalone in the presence of acetic anhydride (Scheme 2). The active hydrogen in the



Scheme 2

[68] E. Celon, L. Biondi, and E. Bordignon, *J. Chromatog.*, 1968, **35**, 47.
[69] T. Skraba-Blotnicka, *Chem. analit.*, 1968, **13**, 587.
[70] H. Fraenkel-Conrat and C. M. Tsung, *Methods in Enzymology*, 1967, **11**, 151.
[71] R. P. Ambler, *Methods in Enzymology*, 1967, **11**, 155.
[72a] D. M. P. Phillips, *J. Chromatog.*, 1968, **37**, 132. [b] H. Wojciechowska and E. Borowski, *Roczniki Chem.*, 1967, **41**, 1727.
[73] V. Braun and W. A. Schroeder, *Arch. Biochem. Biophys.*, 1967, **118**, 241.
[74] B. Mesrob and V. Holeysovsky, *Coll. Czech. Chem. Comm.*, 1967, **32**, 1976.
[75] G. R. Stark, *Biochemistry*, 1968, **7**, 1796.

oxazalone can be replaced with tritium by treatment with ^3H_2O in pyridine, and after hydrolysis the tritiated amino-acid is identified as the *C*-terminal residue.[76] A critical analysis of methods for carrying out this reaction has recently appeared.[77a] Selective tritiation was attempted in a number of solvents but a mixture of acetic anhydride and pyridine in ^3H_2O was found to give the best results. Proline does not react under these conditions and some side-reaction occurs owing to protonation of the basic groups. This non-selective tritiation generally constituted some 10% of the label found in the *C*-terminal amino-acid after hydrolysis and need not be confusing, therefore, unless the *C*-terminus does not react, *e.g.* because it is an α-carboxyamide. Correct assignment of serine and aspartic acid as the *C*-terminal residues in ovine luteinising hormone was made. In another study,[77b] the *C*-terminal residue of neocarzinostatin, an acidic protein (*M* 9000) which inhibits bacterial DNA synthesis, was investigated. Carboxypeptidase released asparagine and by use of a selective tritiation technique aspartic acid was found to be the labelled amino-acid after hydrolysis, in accord with asparagine being *C*-terminal.

C. Sequential Degradation.—The advantages of sequential degradations for the determination of amino-acid sequence are clear for all to see and need no reiteration here. However, one advantage is only just beginning to acquire its full significance: that, of course, is the prospect of automation. No doubt this will become more and more dominant as time goes on.

The Edman Method. Numerous prescriptions [4, 5, 78] have been described for carrying out the Edman degradation of proteins and peptides sequentially from the *N*-terminus. In essence, the polypeptide is treated with phenylisothiocyanate and, under conditions of anhydrous acid, the modified *N*-terminal residue is cleaved off as a 2-anilinothiazolin-5-one, which rearranges in aqueous acid to the familiar 3-phenyl-2-thiohydantoin. The residual peptide is then subjected to another round of the degradation. In order not to expose the residual peptide to aqueous acid the thiazolinone is extracted after cleavage and converted into the thiohydantoin in a separate step. In this way, one might hope to achieve the maximum possible number of rounds of degradation.

Clearly, in the absence of side-reactions, that number will be strongly dependent on the efficiency of coupling and cleavage at each step; yields in excess of, say, 98% are required if the reaction scheme is to form the basis of a procedure that offers information for, say, 40 or 50 steps. An apparatus, the 'protein sequenator', which achieves just this in a series of automated reactions has now been described.[79] *Ca.* 0·25 μmole of protein are required

[76] H. Matsuo, Y. Fujimoto, and T. Tatsuno, *Biochem. Biophys. Res. Comm.*, 1966, **22**, 69.
[77a] G. N. Holcomb, S. A. James, and D. N. Ward, *Biochemistry*, 1968, **7**, 1291. [b] H. Maeda, T. Koyanagi, and N. Ishida, *Biochim. Biophys. Acta*, 1968, **160**, 249.
[78] R. F. Doolittle, *Biochem. J.*, 1965, **94**, 742; D. G. Smyth and D. F. Elliott *Analyst*, 1964, **89**, 81.
[79] P. Edman and G. Begg, *European J. Biochem.*, 1967, **1**, 80.

and 15 cycles can be completed in 24 hr. The design of the apparatus is extremely ingenious; the reactions are carried out in a rotating vessel so that the reactants are spread out in a thin film on the walls. This enables the various additions of fresh reagents and the various extractions to be carried out with high efficiency. Much attention is paid to the purity of the solvents and reagents: in particular, the presence of aldehydes could cause disastrous blocking of exposed α-amino-groups. Air is excluded to prevent oxidative loss of sulphur from the phenylthiocarbamyl protein. As a test of the apparatus, the sequence of the first 60 residues in humpback whale apomyoglobin was successfully determined. This is close to the theoretical limit if a 2% loss occurs at each step. No doubt the system can be improved still further, with corresponding gain in the length of chain analysed.

One obvious drawback to the technique exists. Much depends upon the efficiency with which the thiazolinone can be extracted away from the residual peptide and, as the residual peptide becomes progressively shorter, it may be expected that this efficiency will begin to decrease. Thus it is clear that the automated method may be difficult to apply to small peptides as it stands. It is possible that this problem can be circumvented by attaching the peptide to some form of insoluble resin and some preliminary studies of this type have been reported.[80, 81] Further progress with the sequenator and accounts of other amino-acid sequences analysed with it [82, 83] are awaited with much interest.

For the sequence analysis of small peptides the dansyl–Edman procedure [84] is clearly the current system of choice. It enjoys the benefits of sequential degradation by the Edman reaction, together with the privileges of high-sensitivity identification of the fresh *N*-terminal group exposed in each cycle. Ten nanomoles of a small peptide is frequently sufficient to establish its entire sequence. The advent of dansyl amino-acid separation by chromatography on polyamide thin-layer sheets [62] has brought about considerable improvement in the identification of these derivatives.

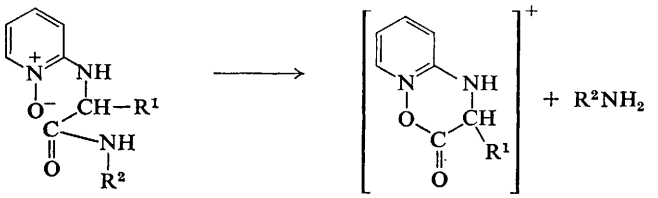
Other Chemical Methods. Several variations on the basic Edman theme are, of course, possible. One of the most interesting appears to be the suggested use [85] of fluorescein isothiocyanate as an alternative to phenyl isothiocyanate. While the degradation retains its sequential character, the thiohydantoins produced are now themselves fluorescent and it should therefore be possible to identify them with the same sort of sensitivity that the dansyl–Edman procedure provides.

Treatment of peptides with sodium thiobenzoylthioglycollate results in the formation of the *N*-thiobenzoyl peptide which, for the purposes of

[80] R. A. Laursen, *J. Amer. Chem. Soc.*, 1966, **88**, 5344.
[81] A. Schellenberger, H. Jeschkeit, R. Henkel, and H. Lehmann, *Z. Chem.*, 1967, **7**, 191.
[82] H. D. Niall and P. Edman, *Nature*, 1967, **216**, 262.
[83] P. Edman and A. G. Cooper, *F.E.B.S. Letters*, 1968, **2**, 33.
[84] W. R. Gray, *Methods in Enzymology*, 1967, **11**, 469.
[85] H. Maeda and H. Kawauchi, *Biochem. Biophys. Res. Comm.*, 1968, **31**, 188.

sequential degradation, may be cleaved with trifluoroacetic acid.[86] The *N*-thiobenzoyl derivative released is identified following conversion into its anilide. Useful information is now available on the c.d. properties of *N*-thiobenzoyl amino-acids[87] and peptides,[88] and accounts of further experience with this technique will be welcome.

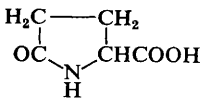
A stepwise degradation of peptides using 2-fluoropyridine-*N*-oxide has also been described.[89] In acidic conditions the *N*-terminal peptide bond is labilised with respect to other peptide bonds, presumably because of *N*-oxide group participation to form the cyclic intermediate (Scheme 3).



Scheme 3

Di- and tri-peptides were successfully cleaved this way in a sequential fashion. The system clearly has affinities with some end-group methods discussed earlier (p. 40), but at the moment it is hard to see it offering any real advantage over the well-characterised methods already described.

D. Blocked End-groups.—Protein chemists have long grown accustomed to the presence of blocked *N*-terminal residues in proteins, although their chemical diversity continues to surprise. The acetyl group, first detected in tobacco mosaic virus protein,[90] is almost *de rigueur* in the mammalian cytochromes-*c*.[91] Pyrrolidone carboxylic acid (4) exists as the *N*-terminal



(4)

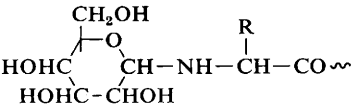
residue of many peptides and proteins, including gastrin, fibrinopeptides, and immunoglobulins. A great problem with this residue is that, because it is known that *N*-terminal glutamine can easily cyclise to the pyrrolidone carboxylic acid,[92] it is difficult to decide whether the cyclised form exists

[86] G. C. Barrett, *Chem. Comm.*, 1967, 487.
[87] G. C. Barrett, *J. Chem. Soc.* (*C*), 1967, 1.
[88] G. C. Barrett, *Chem. Comm.*, 1968, 40.
[89] V. Tortorella and G. Tarzia, *Gazzetta*, 1968, **97**, 1479.
[90] K. Narita, *Biochim. Biophys. Acta*, 1958, 28, 184.
[91] E. Margoliash and A. Schejter, *Adv. Protein Chem.*, 1966, **21**, 113.
[92] B. Blombäck, *Methods in Enzymology*, 1967, **11**, 398.

in vivo or is an artefact of the protein purification procedure. This cyclisation can certainly happen during sequence studies with *N*-terminal glutaminyl peptides. In any event its existence is an effective bar to sequential degradation from the *N*-terminus. Fortunately, an enzyme has recently been discovered [93] that selectively cleaves pyrrolidone carboxylic acid from its terminal position. The enzyme, pyrrolidonyl peptidase, has been partially purified from *Pseudomonas fluorescens*. Although its physiological role remains obscure, no doubt it will find extensive use for rescuing afflicted Edman degradations.

The *N*-formyl group is present in the small peptide antibiotic valinegramicidin A.[94] The use of *N*-formylmethionine in the initiation of polypeptide chain synthesis in bacteria [95] means that this residue must exist at the *N*-terminus of the growing protein chain. An enzyme that removes the *N*-formyl groups has been partially purified from *E. coli* extracts.[96] As expected, the enzyme removes the formyl group from *N*-formylmethionine only after the formation of the first peptide bond; this is essential if the *N*-formylmethionyl-transfer RNA is to serve as the initiator of protein biosynthesis. After deformylation, another amino-peptidase presumably can remove the *N*-terminal methionine to account for the fact [97] that the *N*-terminal amino-acids of soluble *E. coli* proteins consist mainly of methionine, alanine, and serine.

An unusual blocking group has been detected [98] in haemoglobin A_{Ic}, the largest (5–7%) of several minor components in normal human adult blood. It transpires from further work [99] that the blocking group is probably a hexose and that it is attached to both β-chains of the haemoglobin molecule. The acid lability of the blocking group, its reduction by borohydride, and its stability at neutral pH might be explained by a structure of the type shown (5). Possible functions for Hb A_{Ic} have been discussed in detail.[100]



(5)

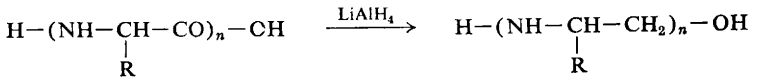
E. Amino-acid Sequence by Mass Spectrometry.—Mass spectrometry is obviously potentially very useful for studying the structure of peptides.

[93] R. F. Doolittle and R. W. Armentrout, *Biochemistry*, 1968, **7**, 516.
[94] R. Sarges and B. Witkop, *J. Amer. Chem. Soc.*, 1965, **87**, 2011.
[95] K. A. Marcker, B. F. C. Clark, and J. S. Anderson, *Cold Spring Harbour Symp. Quant. Biol.*, 1966, **31**, 279.
[96] D. M. Livingston and P. Leder, *Biochemistry*, 1969, **8**, 435.
[97] J.-P. Waller, *J. Mol. Biol.*, 1963, **7**, 483.
[98] W. R. Holmquist and W. A. Schroeder, *Biochemistry*, 1966, **5**, 2489.
[99] R. M. Bookchin and P. M. Gallop, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 86.
[100] W. A. Schroeder and W. R. Holmquist, 'Structural Chemistry and Molecular Biology,' ed. A. Rich and N. Davidson, Freeman, San Francisco, 1968, p. 238.

Recently it has become clear that, in addition, mass spectrometry might also fruitfully be used for the determination of amino-acid sequence. Excellent comprehensive reviews have been published.[101-103]

As with g.l.c. one has to render the peptides volatile before this method of analysis is applicable and, clearly, with a large peptide that might be very difficult. A very real attraction of the method is its high sensitivity (1 μg of a small peptide could well be enough to establish its entire sequence) and the possibility of automation.

Early efforts[104] involved the reduction of the peptide or acetylpeptide with lithium aluminium hydride to give polyamino-alcohols (Scheme 4)



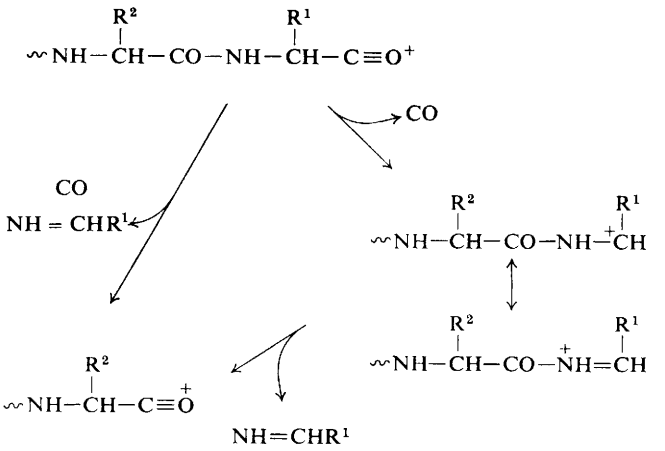
Scheme 4

which gave reasonable mass spectra. Volatile derivatives are available, however, in the *N*-acyl derivatives of peptide esters, which have been the most widely studied derivatives to date. Methyl,[105] ethyl,[106] and t-butyl[107] esters have all been used and *N*-protection has been achieved by acetylation[106] and trifluoroacetylation.[105, 108] Methods for conducting these modifications on a sufficiently small scale have been described.[109, 110]

Fragmentation of Peptide Derivatives. The possibility of using the mass spectrometer for deriving amino-acid sequences stems from the observation that in the mass spectra of *N*-acylated peptide derivatives peaks occur which correspond to the rupture of the peptide bond in two modes. The principal cleavage is that of the CO—N bond to give acylium ions, which can then lose the next amino-acid residue either by loss of carbon monoxide and a neutral imine fragment or by intermediate conversion to the acyliminium ion (Scheme 5). This splitting appears to be a sequential process proceeding towards the *N*-terminus,[111] and is more pronounced in higher peptides

[101] E. Lederer and B. C. Das, 'Peptides,' ed. H. C. Beyerman, A. van de Linde, and W. Maassen van den Brink, North-Holland Publishing Co., Amsterdam, 1967, p. 131.
[102] Yu. A. Ovchinnikov, A. A. Kiryushkin, E. I. Vinograd, B. V. Rosinov, and M. M. Shemyakin, *Biokhimiya*, 1967, **32**, 427.
[103] J. H. Jones, *Quart. Rev.*, 1968, **22**, 302.
[104] K. Biemann and W. Vetter, *Biochem. Biophys. Res. Comm.*, 1960, **3**, 578.
[105] E. Stenhagen, *Z. analyt. Chem.*, 1961, **181**, 462.
[106] V. G. Manusadzhyan, A. M. Zyakoon, A. V. Chuvilin, and Ya. M. Varshavskii, *Izvest. Akad. Nauk Arm. S.S.S.R., Ser. khim.*, 1964, **17**, 143.
[107] M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kiryushkin, E. I. Vinogradova, A. I. Miroshnikov, Yu. B. Alakhov, V. M. Lipkin, Yu. B. Shvetsov, N. S. Wulfson, B. V. Rosinov, V. N. Bocharev, and V. M. Burikov, *Nature*, 1966, **211**, 361.
[108] F. Weygand, A. Prox, H. H. Fessel, and K. K. Sun, *Z. Naturforsch.*, 1965, **20b**, 1169.
[109] M. Senn, R. Venkataraghavan, and F. W. McLafferty, *J. Amer. Chem. Soc.*, 1966, **88**, 5593.
[110] A. A. Kiryushkin, Yu. A. Ovchinnikov, M. M. Shemyakin, V. N. Bocharev, B. V. Rosinov, and N. S. Wulfson, *Tetrahedron Letters*, 1966, 33.
[111] M. Barber, P. Jollès, E. Vilkas, and E. Lederer, *Biochem. Biophys. Res. Comm.*, 1965, **18**, 469.

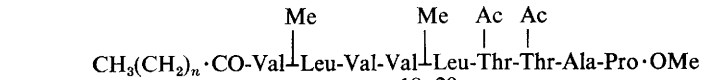
than in di- and tri-peptides. The alternative cleavage, that of the C—CO bond, often accompanies the CO—N cleavage, with charge retention by either fragment. Fortunately, it it usually less pronounced.



Scheme 5

The mass spectra of peptides always contain, in addition to the peaks corresponding to peptide bond cleavage, peaks deriving from side-chain fragmentation. The intensity of these varies a good deal and they are characteristic of the individual side-chains.[101-103] However, it should be clear that if the ions (the 'sequence ions') corresponding to the sequential loss of amino-acids from the *C*-terminus of the molecular ion can be distinguished, then the amino-acid sequence of the peptide can be deduced.

Techniques for Sequence Determination. Recognition of the sequence ions can be aided in a number of ways. If the *N*-terminus of the peptide is protected with an equimolar mixture of acetyl and trideuterioacetyl groups, the sequence ions appear as pairs of peaks of equal intensity separated by three mass units.[112] Work with the naturally occurring peptide fortuitine (6) suggested that it might be possible to use an equimolar mixture of heptadecanoic and octadecanoic acids in a similar way,[113] with the added



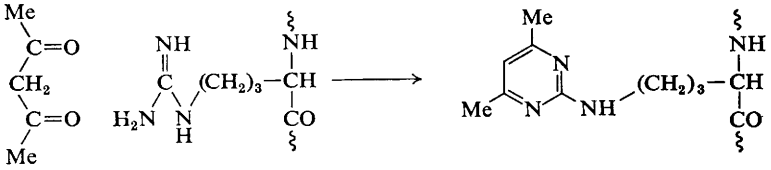
(6)

[112] J. van Heijenoort, E. Bricas, B. C. Das, E. Lederer, and W. A. Wolstenholme, *Tetrahedron*, 1967, **23**, 3403.

[113] E. Bricas, J. van Heijenoort, M. Barber, W. A. Wolstenholme, B. C. Das, and E. Lederer, *Biochemistry*, 1965, **4**, 2254.

advantage that the sequence ions are brought into the high mass range of the spectrum away from overlapping peaks.

Accumulating experience with the method has indicated that its success has been somewhat limited by the poor volatility of the higher oligopeptides. In particular, the presence of arginine in a peptide formerly precluded its analysis entirely. This has now been overcome either by converting the arginine into ornithine by refluxing with hydrazine [114] or by blocking the arginine side-chain by reaction with 1,3-diketones [114–116] (Scheme 6). Both of these procedures lead to satisfactory mass spectra.[114, 116]



Scheme 6

Peptides containing cystine have also been examined.[117] It appears that the mass spectra of esters of cystine-containing *N*-acylpeptides are actually those of the corresponding cysteine derivatives with unprotected thiol groups owing to disulphide bond rupture in the mass spectrometer. The *S*-carboxymethyl derivatives fortunately are suitable for analysis: fortunately, because these are the most common derivatives of cysteine met with in amino-acid sequence determinations.

Consideration of the successful elucidation of the structure of fortuitine (6) [111] has led to the suggestion [118] that one of the major limiting factors in the volatility of oligopeptides might be inter-chain hydrogen bonding between the CO—NH groups. Fortuitine has three tertiary amide bonds and its diminished hydrogen bonding might therefore account, in part, for its satisfactory volatility. Permethylation of the peptide bonds of *N*-acyl peptide methyl esters by treatment in dimethylformamide with an excess of methyl iodide in the presence of silver oxide resulted in a considerable increase in the volatility of the peptide derivative. It was pointed out that if the peptide already contains *N*-methyl groups, methylation with trideuteriomethyl iodide can be used to distinguish the natural from the

[114] M. M. Shemyakin, Yu. A. Ovchinnikov, E. I. Vinogradova, M. Yu. Feigina, A. A. Kiryushkin, N. A. Aldanova, Yu. B. Alakhov, V. M. Lipkin, and B. V. Rosinov, *Experientia*, 1967, **23**, 428.

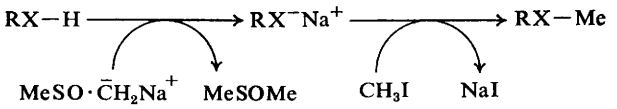
[115] T. P. King, *Biochemistry*, 1966, **5**, 3454.

[116] H. Vetter-Diechtl, W. Vetter, W. Richter, and K. Biemann, *Experientia*, 1968, **24**, 340.

[117] A. A. Kiryushkin, V. A. Gorlenko, Ts. E. Agadzhanyan, B. V. Rosinov, Yu. A. Ovchinnikov, and M. M. Shemyakin, *Experientia*, 1968, **24**, 883.

[118] B. C. Das, S. D. Gero, and E. Lederer, *Biochem. Biophys. Res. Comm.*, 1967, **29**, 211.

artificially introduced methyl groups. A number of papers [119-122] extending these observations have since been published. The sequences of several decapeptides were successfully established [119] where it was reported that the sequence determination was made easier because the spectra arose principally from $CO—NCH_3$ bond cleavage. Some unusual results were obtained if histidine, arginine, methionine, or aspartic acid were present and other limitations included the fact that glutamic acid, phenylalanine, tyrosine, and tryptophan became methylated. Remedies for some of these problems have been suggested.[120] Free amino-groups which would give a quaternised methyl iodide salt with very low volatility should be acetylated before permethylation is attempted. Arginine may be converted to ornithine by a hydrazine treatment:[114] after acetylation of the resulting amino-group permethylation of the peptide is carried out as before. With peptides containing methionine there is a somewhat different problem. Prominent peaks in the mass spectrum delineate the sequence up to, but not including nor beyond, the methionine residue in the permethylated peptide, possibly because of formation of the polar sulphonium iodide. Desulphurisation of the methionine with Raney nickel is suggested as a possible remedy. A word of warning has also been sounded. In an independent account of the permethylation technique [121] it has been reported that chain cleavage can occur during the permethylation of peptides containing methionine or glutamic acid and that aspartic acid also undergoes complex transformations during the reaction, although without chain cleavage. However, a change of permethylation reaction conditions dispensing with silver oxide is found [122] to cause aspartic acid, glutamic acid, and tryptophan to behave normally in the peptides tested and to give satisfactory, interpretable mass spectra. In these latest studies methylation is achieved by using methyl iodide with a methylsulphinyl carbanion in dimethylsulphoxide as described for glycolipids [123] (Scheme 7). This method has clear potential merit.



Scheme 7

The mass spectral behaviour of the benzyloxycarbonyl derivatives of amino-acid and dipeptide phenylthioesters has also been examined.[124]

[119] D. W. Thomas, B. C. Das, S. D. Géro, and E. Lederer, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 199.
[120] D. W. Thomas, B. C. Das, S. D. Géro, and E. Lederer, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 519.
[121] K. L. Agarwal, R. A. W. Johnstone, G. W. Kenner, D. S. Millington, and R. C. Sheppard, *Nature*, 1968, **219**, 498.
[122] D. W. Thomas, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 483.
[123] S. I. Hakamori, *J. Biochem. (Japan)*, 1964, **55**, 205.
[124] R. T. Aplin, J. H. Jones, and B. Liberek, *J. Chem. Soc.* (*C*), 1968, 1011.

The peak of highest *m/e* in the spectra of these compounds corresponds to the acylium ion formed by loss of a phenylthio-radical. If the phenylthioesters act to 'trigger' the stepwise elimination of amino-acid residues from the *C*-terminus they could clearly be very useful in sequence studies by promoting the formation of the sequence ions.[124]

Use of Computers for Sequence Determination. Automatic sequence determination of peptides can be obtained from computer examination of high-resolution mass spectra in several ways. In the first,[125] a search is made for the *N*-terminal amino-acid by checking the sum of the exact mass of the *N*-protecting group plus all possible amino-acid residues against the masses of ions observed in the spectrum. When a unique match is made, thereby identifying the *N*-terminal residue, subsequent amino-acids in the sequence are identified in a similar way up to the molecular ion. An analogous programme in which it is the masses of the possible amino-acid side-chains that are used for checking against the observed spectrum has also been proposed.[126] Both of these approaches eliminate entirely the need for previous chemical information about the peptide, identifying not only the sequence but also the nature of the amino-acids in the peptide. If one is prepared to use preliminary high-sensitivity amino-acid analysis to obtain the latter information, the problem can obviously be simplified. Such a programme is available:[127] it works backwards, so to speak, from the *C*-terminus by first identifying the molecular ion and then proceeding to identify the sequence ions formed by cleavage from the *C*-terminus. Any available chemical information can be accommodated.

Prospects. Sufficient success has now been recorded with the mass spectrometric analysis of peptides for its proponents to speak hopefully of its becoming a routine tool for sequence analysis. Until recently, the inability to handle certain amino-acid residues, *e.g.* arginine, and the limited volatility of oligopeptides seemed powerful arguments against its early introduction. However, there are now clear indications that these problems are being overcome. A strong combination of mass spectral analysis of the eluants of g.l.c. columns is often available for some compounds. Perhaps we may yet see the separation of suitably modified peptides on g.l.c. and their subsequent analysis in the mass spectrometer. In this connection, the observation[128] that the *N*-heptafluorobutyryl derivatives of simple peptides have their g.l.c. retention times shortened by up to 50% compared with the corresponding trifluoroacetyl derivatives, and that a lower temperature is needed for their vaporisation in the direct inlet system of the mass spectrometer, may well be of some interest.

[125] M. Senn and F. W. McLafferty, *Biochem. Biophys. Res. Comm.*, 1966, **23**, 381.
[126] K. Biemann, C. Cone, B. R. Webster, and G. P. Arsenault, *J. Amer. Chem. Soc.*, 1966, **88**, 5598.
[127] M. Barber, P. Powers, M. J. Wallington, and W. A. Wolstenholme, *Nature*, 1966, **212**, 784.
[128] B. A. Anderson, *Acta Chem. Scand.*, 1967, **21**, 2906.

F. Peptide Chain Cleavage and Peptide Fractionation.—The methods of polypeptide chain cleavage, enzymic and chemical, have been covered in great detail elsewhere.[4-6] However, there have been come recent developments in the subject that deserve comment.

Proteolytic Enzymes. The purification and properties of a protease from *Crotalus atrox* snake venom have been described.[129] The enzyme, prepared by column chromatography of crude venom on DEAE-Sephadex, hydrolyses bonds of the amino-groups of aliphatic hydrophobic residues, preferably at leucine, isoleucine, and valine. Peptides smaller than five residues are not hydrolysed by the enzyme, leading to the production of overlapping sequences in high yields. Thus, whereas tryptic digestion of lactic dehydrogenase gave the theoretical number of *ca.* 30–35 peptides, digestion with α-protease gave only *ca.* 10, in contrast with the theoretical number of *ca.* 100. Additional causes for this are suggested to be the lack of hydrolysis of sequences such as Val-Val-X-Y- *etc.* owing, perhaps, to the inhibition of hydrolysis by adjacent charged groups, *e.g.* the α-amino-groups generated in the digestion. It is clear that the specificity of α-protease is similar to that of thermolysin [130] (from *B. thermoproteolyticus*) which may indicate a common mode of action of a large group of proteolytic enzymes of lower organisms. The use of thermolysin in amino-acid sequence determination has been the subject of a recent study [131] of the digestion products of the copper protein azurin from *Pseudomonas fluorescens.* Its potential use in the secondary cleavage of larger tryptic and chymotryptic peptides is reported.

The purification and properties of a collagenolytic enzyme from *Aspergillus oryzae* have been described.[132] The enzyme has a molecular weight of *ca.* 20,000, is inhibited by di-isopropylphosphofluoridate (DFP) and related inhibitors, and has a pH optimum of 9–10. Preferential splitting is observed at acidic side-chains and the enzyme will digest native collagen. It may be distinguished from collagenase in that bonds involving hydroxyproline are resistant and that collagenase digests only collagen, and it is suggested that the enzyme should aid structural studies on collagen by enabling longer sequences to be derived from the apolar region.

Some interesting properties of the specificity of human plasmin have been reported.[133] Human plasmin resembles trypsin in splitting the insulin B-chain and it cleaves only at lysine and arginine residues in *S*-sulphofibrinogen.[134] However, whereas trypsin cleaves 80% of the available bonds

[129] K. Mella, M. Volz, and G. Pfleiderer, *Analyt. Biochem.*, 1967, **21**, 219.
[130] H. Matsabura, R. Sasaki, A. Singer, and T. H. Jukes, *Arch. Biochem. Biophys.*, 1966, **115**, 324.
[131] R. P. Ambler and R. J. Meadway, *Biochem. J.*, 1968, **108**, 893.
[132] A. Nordwig and W. F. Jahn, *European J. Biochem.*, 1968, **3**, 519.
[133] W. R. Groskopf, B. Hsieh, L. Summaria, and K. C. Robbins, *Biochim. Biophys. Acta*, 1968, **168**, 376.
[134] S. Iwanaga, P. Wallén, N. J. Gröndahl, A. Henschen, and B. Blombäck, *European J. Biochem.*, 1968, **8**, 189.

in *S*-sulphofibrinogen, plasmin cleaves only 50%.[135] Thus, although plasmin resembles trypsin its specificity seems more limited. Perhaps plasmin can find some use in establishing tryptic peptide overlaps.

It is well known that most proteins in the native state are attacked only slowly by proteolytic enzymes. One of the best examples of the use to which this knowledge can be put is the limited subtilisin digestion of pancreatic ribonuclease [136] to yield a single bond cleavage. An improved separation of the S-protein and S-peptide using gel filtration on columns of Sephadex G-75 has recently been described.[137] Similar experiments have now been reported concerning the conversion of ovalbumin to plakalbumin by limited digestion with subtilisin.[138] During the conversion one peptide bond only is split, creating a new *C*-terminal residue, alanine, and a new *N*-terminal residue, serine (the *N*-terminal residue of ovalbumin is acetylated). After *S*-carboxymethylation in 7M-urea, two components can be separated from plakalbumin by gel chromatography in 6M-urea. The smaller component contains *ca.* 33 residues, including two residues of *S*-carboxymethylcysteine, has *N*-terminal serine, and presumably derives from the *C*-terminal end of the ovalbumin molecule. (This contrasts nicely with the experiments on ribonuclease, where the limited cleavage is towards the *N*-terminus of the molecule.) As with ribonuclease-S, too, the peptide fragment in plakalbumin appears not to be linked to the remainder of the protein by covalent bonds, because it comes off in urea without reduction.

Another good example of the technique is afforded by the preparation of active derivatives of staphylococcal nuclease during trypsin digestion.[139] The work was done on the Foggi strain, which is very similar to the V8 strain (see p. 71). Trypsin cleavage carried out in the presence of deoxythymidine-3′,5′-diphosphate and calcium ions to stabilise the nuclease gives fragments P1 (residues 1–5), P2 (residues 6–48), P3 (residues 49–149 = P3a; residues 50–149 = P3b), and P4 (free lysine, residue 49). None is individually active. However, P2 and P3 (either P3a or P3b) can associate to form an active, non-covalently bonded derivative, nuclease-T. Specific cleavage of the bonds 5–6, 48–49, and 49–50 suggests that these regions protrude from the body of the protein in an analogous way to the sensitive loop of pancreatic ribonuclease.[136b]

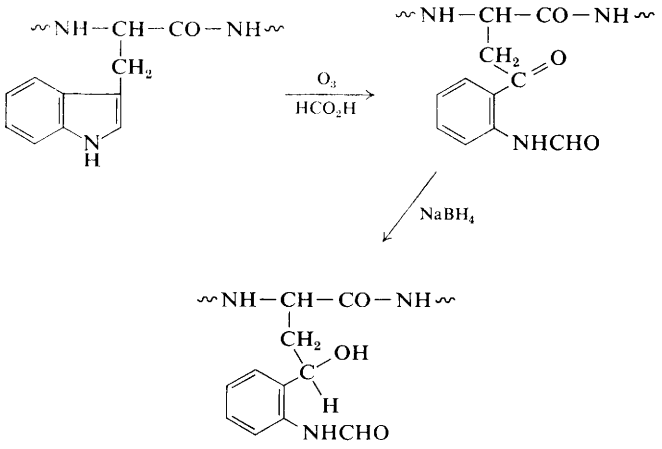
Chemical Methods of Chain Cleavage. Very few good methods for peptide chain cleavage have been invented, specific cleavage at the methionine residues with cyanogen bromide [140] being a notable exception. It is worth remarking that cleavage with cyanogen bromide can sometimes be limited

[135] P. Wallén and S. Iwanaga, *Biochim. Biophys. Acta*, 1968, **154**, 414.
[136a] F. M. Richards and P. J. Vithayathil, *J. Biol. Chem.*, 1959, **234**, 1459. [b] H. W. Wyckoff, K. D. Hardman, N. M. Allewell, T. Inagami, L. N. Johnson, and F. M. Richards, *J. Biol. Chem.*, 1967, **242**, 3984.
[137] M. S. Doscher and C. H. W. Hirs, *Biochemistry*, 1967, **6**, 304.
[138] M. B. Smith, *Biochim. Biophys. Acta*, 1968, **154**, 263.
[139] H. Taniuchi and C. B. Anfinsen, *J. Biol. Chem.*, 1968, **243**, 4778.
[140] E. Gross, *Methods in Enzymology*, 1967, **11**, 238.

to specific methionine residues. For example, it is possible to prepare biologically active fragments of immunoglobulin molecules by cleavage under mild conditions.[141]

A detailed study of the reductive cleavage of acylproline peptide bonds has been reported.[142] Cleavage is induced by sodium metal in liquid ammonia. Variation in the lability of different acylproline peptides is observed and the maximal cleavage obtained is in the range 56–90%. Non-specific cleavage of other peptide bonds, particularly that at the *N*-terminus, can be a problem and the nature of the new *C*-terminus produced is in doubt. It is concluded that general applications will be fairly limited but that the technique could be useful in cleaving Arg–Pro and Lys–Pro bonds which ordinarily resist the action of trypsin.

Tryptophan can be converted into *N*-formylkynurenine by ozonisation in formic acid.[143] (Methionine may be converted into its sulphoxide during the reaction but that oxidation can be reversed with mercaptoethanol.) Cleavage at the *N*-formylkynurenine residues can then be effected by heating the peptide at 100° in 0·5M-$NaHCO_3$ either before [144] or after [145] reduction of the kynurenine residue with sodium borohydride, presumably by intramolecular participation of the γ-keto- or -hydroxy-group (Scheme 8). The modified tryptophan residue actually is excised if the reduction step is

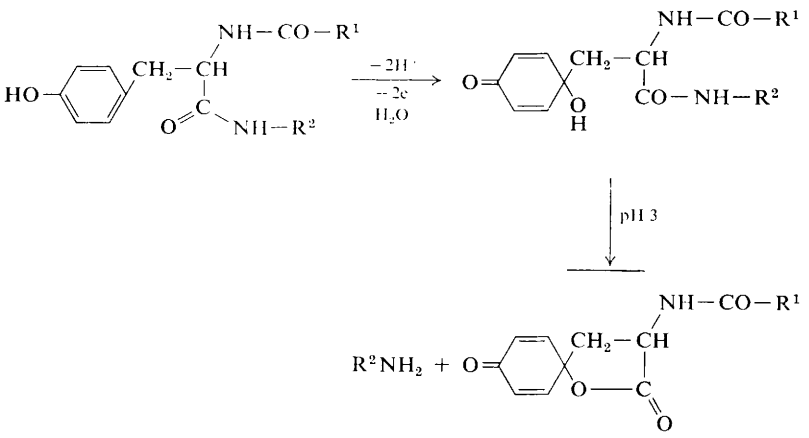


Scheme 8

[141] H. J. Cahnmann, R. Arnon, and M. Sela, *J. Biol. Chem.*, 1965, **240**, PC 2762.
[142] W. F. Benisek, M. A. Raftery, and R. D. Cole, *Biochemistry*, 1967, **6**, 3780.
[143] A. Previero and E. Bordignon, *Gazzetta*, 1964, **94**, 630; A. Previero, M.-A. Coletti-Previero, and P. Jollès, *J. Mol. Biol.*, 1967, **24**, 261.
[144] A. Previero, M.-A. Coletti-Previero, and P. Jollès, *Biochem. Biophys. Res. Comm.*, 1966, **22**, 17.
[145] A. Previero, M.-A. Coletti-Previero, and P. Jollès, *Biochim. Biophys. Acta*, 1966, **124**, 400.

included in the treatment.[145] The conversion of tryptophan to *N*-formylkynurenine by ozonisation can also be used to prevent peptide bond cleavage at tryptophan residues during treatment of the peptide chain with *N*-bromosuccinimide.[146] In general, *N*-bromosuccinimide treatment can be used to promote cleavage at tyrosine, tryptophan, and histidine residues.[147] Tryptophyl bonds are cleaved most rapidly and tyrosyl bonds are split only with larger excesses of the reagent. The more vigorous conditions required for cleavage at histidine give concomitant cleavage at tyrosine and tryptophan, although the latter two types of bond can be cleaved with little effect on histidyl bonds. Thus the prevention of cleavage at tryptophan by ozonisation before *N*-bromosuccinimide treatment can be used to provide specific splitting at tyrosine residues, and in tests of the method with the α-chain of human haemoglobin[146] useful recoveries (10–50%) of the expected peptides are reported. Previous reaction of the tryptophan residues with 2-hydroxy-5-nitrobenzyl bromide can also be used to prevent cleavage of tryptophyl bonds by *N*-bromosuccinimide.[148]

Earlier work on the cleavage of tyrosyl peptide bonds by means of electrolytic oxidation[149] has been followed up. A continuous flow electrolysis apparatus in which to carry out the reaction has been devised[150] so that the time the substrate spends in contact with the electrode is limited. Electrolytic oxidation of tyrosine in proteins converts the tyrosine residue to a *p*-hydroxydienone without chain breakage. The chain is then cleaved at pH 3 (Scheme 9). In tests with ribonuclease, 85% modification of tyrosine



Scheme 9

[146] A. Previero, M.-A. Coletti-Previero, and C. Axelrud-Cavadore, *Arch. Biochem, Biophys.*, 1967, **122**, 434.
[147] L. K. Ramachandran and B. Witkop, *Methods in Enzymology*, 1967, **11**, 283.
[148] M. Wilchek and B. Witkop, *Biochem. Biophys. Res. Comm.*, 1967, **26**, 296.
[149] L. A. Cohen and L. Farber, *Methods in Enzymology*, 1967, **11**, 299.
[150] S. Isoe and L. A. Cohen, *Arch. Biochem. Biophys.*, 1968, **127**, 522.

was achieved without attack at other functional groups. Subsequent hydrolysis cleaved all six tyrosyl bonds specifically but to varying degrees because of incomplete hydrolysis of the hydroxydienones and competing dienonephenol rearrangements. Obviously there is room for improvement, but the method is potentially very powerful for specific chain cleavage and for examining the three-dimensional disposition of tyrosine residues.

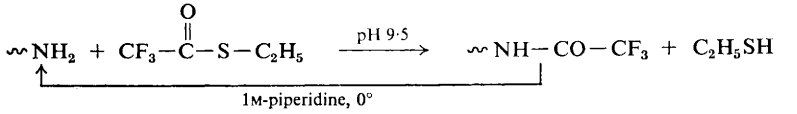
It is sad but true that few good chemical methods exist for the specific cleavage of polypeptide chains. When a good one is devised, as with the cyanogen bromide method, its impact on amino-acid sequence determination is very great. A glance at the current literature is a simple reminder of how widely cleavage at methionine residues is employed. To be useful a method must be specific and, therefore, it usually involves the side-chains of the potentially reactive, rarer amino-acids such as tryptophan, histidine, tyrosine, or methionine. A secondary advantage implicit here is that owing to the rarity of the particular amino-acids concerned fewer and larger fragments result from the reaction than is generally the case with proteolytic digestion.

It is also sad but true that, too often, many of the chemical methods reported proceed in dismal yield, and then only with simple peptides rather than the more demanding proteins. Hopefully, further development can improve the potentially better methods and they, no doubt, will both get and deserve it. However, for those methods that abound in the literature and disappear unsung and unlamented the words of St. Matthew are still appropriate.[151]

Chemical Modification and Enzymic Cleavage. The enzyme trypsin cleaves polypeptide chains on the *C*-terminal side of arginine and lysine residues only. It is clear that specific chemical modification of either side-chain such that tryptic cleavage is inhibited at that residue would greatly extend the usefulness of the enzyme to the protein chemist: modification of lysine residues would permit cleavage only at arginine residues and *vice versa*. The design of reagents for this purpose has recently become something of a minor industry. The fact that the lysine side-chain carries an ε-amino-group compared with the guanidino-group of arginine has meant that more successful modifications have been devised for lysine than for arginine.

(*i*) *Modifications of lysine.* One of the most successful techniques for blocking lysine residues is trifluoroacetylation of the amino-groups with *S*-ethyl trifluorothioacetate.[152] The modification has the added attraction that after cleavage of the arginyl bonds with trypsin, the peptides may be separated and the blocking groups removed in 1M-piperidine at 0° (Scheme 10). The lysyl bonds may now be cleaved by a second exposure to trypsin and the resulting tryptic peptides identified to obtain the desired sequence information. Care must be taken never to submit the trifluoroacetylated

[151] St. Matthew, ch. 22, v. 14, 'For many are called, but few are chosen.'
[152] R. F. Goldberger and C. B. Anfinsen, *Biochemistry*, 1962, **1**, 401.



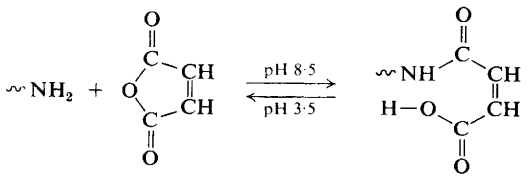
Scheme 10

protein to too high a pH (>*ca.* 9·5) owing to the alkali lability of the blocking groups. The trifluoroacetylated protein is generally insoluble below neutral pH; so, indeed, are some of the larger trifluoroacetylated peptides. If the protein becomes insoluble during the trifluoroacetylation, the reaction can be done successfully in urea.[153] Trifluoroacetylation appears to be entirely specific for the α- and ε-amino-groups of a protein and has achieved wide currency.[154] A possible disadvantage of the method for some purposes is that one product of the reaction is ethyl mercaptan which could cause reduction of disulphide bridges in proteins containing them.

The reaction of succinic anhydride with the amino-groups of proteins was described some years ago.[155] Although it was then thought that this reaction was specific for amino-groups (with the possible exception of thiol groups), it is now clear that this is not so. Succinylation of pepsinogen [156] gives reaction with all ten ε-amino-groups together with reaction of the hydroxy-groups of 16 tyrosine residues and *ca.* 10–13 serine and threonine residues. While *O*-succinyltyrosine decomposes spontaneously at alkaline pH, the *O*-succinyl derivatives of the hydroxyamino-acids are comparatively stable. They can, however, be decomposed at high pH by treatment with the strong nucleophile hydroxylamine. Further, while the conversion of the charge at the lysine side-chain from positive to negative effectively inhibits tryptic action, it is impossible subsequently to remove the blocking group.

To circumvent these drawbacks maleic anhydride [157] has been recommended as a blocking group for protein amino-groups. Maleylation has the great advantage that it is thought to be specific for amino-groups and, further, the maleyl group can be removed because the protonated form of the free carboxyl group catalyses the hydrolysis of the amide bond, presumably by intramolecular general-acid catalysis (Scheme 11). Overnight incubation at pH 3·5 and 60° is suitable for the removal of the maleyl group, which shows a half-life of 11 hr. at pH 3·5 and 37°. Fortunately, these rather stringent conditions can be alleviated by introducing methyl groups into the molecule of maleic anhydride.[158] 2,3-Dimethylmaleyl groups are

[153] R. N. Perham and F. M. Richards, *J. Mol. Biol.*, 1968, **33**, 795.
[154] R. F. Goldberger, *Methods in Enzymology*, 1967, **11**, 317.
[155] A. F. S. A. Habeeb, H. G. Cassidy, and S. J. Singer, *Biochim. Biophys. Acta*, 1958, **29**, 587.
[156] A. D. Gounaris and G. E. Perlmann, *J. Biol. Chem.*, 1967, **242**, 2739.
[157] P. J. G. Butler, J. I. Harris, B. S. Hartley, and R. Leberman, *Biochem. J.*, 1967, **103**, 78P.
[158] H. B. F. Dixon and R. N. Perham, *Biochem. J.*, 1968, **109**, 312.



Scheme 11

removed within 5 min. at pH 3·5 and 20°, indicating extremely rapid breakdown of the protonated form of the product. Even at neutral pH the dimethylmaleyl group is insufficiently stable for the more protracted procedures involved in the isolation of blocked tryptic peptides. 2-Methylmaleic anhydride (citraconic anhydride), however, has all the properties required of a reversible blocking reagent for protein amino-groups.[158] It has a theoretical disadvantage compared with the previous reagents that two products may be formed, according to which carbonyl group of the reagent is attacked (7). However, the difference in stability is not very great

(7)

and a half-life of *ca.* 1·5 hr. at pH 2 and 20° may be assumed. Overnight incubation at pH 3·5 and 20° or 6–8 hr. at pH 2 and 20° should suffice to unblock all amino-groups in a protein.

Other suggested reversible modifications of the ε-amino-group include tetrafluorosuccinylation[159] and acetoacetylation.[160] The reaction with tetrafluorosuccinic anhydride is carried out at pH 7 and 0° in 7M-urea and the blocking group may be removed at pH 9·5 and 0°. Further details of the technique are awaited. Acetoacetylation is achieved by reaction of the protein with diketen at pH 8·5. With high ratios of diketen to protein, acetoacetylation of the hydroxy-groups of serine and tyrosine residues occurs but these may be selectively unblocked by treatment with carbonate–bicarbonate buffer at pH 9·5. The acetoacetyl groups are removed from amino-groups by treatment with hydroxylamine at neutral pH and room temperature. Further details of this technique will also be of interest.

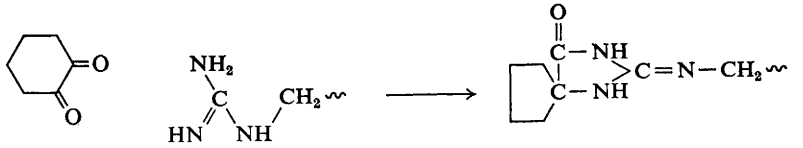
In general it seems likely that for the purposes of amino-acid sequence determination the reactions with anhydrides offer advantages over the other methods. The anhydride derivatives are highly negatively charged which promotes their solubility and helps prevent aggregation of peptides. On

[159] G. Braunitzer, K. Beyreuther, H. Fujiki, and B. Schrank, *Z. physiol. Chem.*, 1968, **349**, 265.

[160] A. Marzotto, P. Pajetta, L. Galzigna, and E. Scoffone, *Biochim. Biophys. Acta*, 1968, **154**, 450.

the other hand, this charge renders them difficult to purify by ion-exchange chromatography. Easy reversibility is a most desirable property of any such reagent, which explains why the imidoesters, highly specific for reaction with amino-groups,[161] have not been widely used for this purpose despite their great usefulness in other types of study involving amino-group modification.[153]

(*ii*) *Modifications of arginine.* If modifications of the arginine side-chain once lagged behind those for amino-groups, they are now coming thick and fast. Reaction of the guanidino-group with 1,3-diketones [114–116] has already been mentioned (Scheme 6, p. 48). Reactions with 1,2-diketones in 0·2N-sodium hydroxide have also been reported.[162] The modification with 1,2-cyclohexanedione is illustrated in Scheme 12. Both these approaches



Scheme 12

require extremes of pH to effect reaction, since only under those conditions will the normally unreactive guanidino-group be even slightly in a reactive form. Reactions that can be done at neutral pH would be preferable. Recent work has shown [163] that diacetyl trimerises at neutral pH to give a crystalline solid. This product reacts at neutral pH predominantly with arginine residues. Another reagent that shows great promise is phenylglyoxal.[164] Reaction with arginine occurs at pH 7–8 and 25°. The derivative degrades slowly at neutral pH (the reagent can therefore be used as a reversible blocking agent) but is stable below pH 4 and survives performic acid oxidation. Two phenylglyoxal moieties are bound per guanidino-group and in all peptides examined any available α-amino-group is deaminated. An excess of the reagent will also react with ε-amino-groups over a long period of time.

At the moment it is difficult to evaluate the relative merits of these blocking reactions for arginine. It might be expected that coupling reactions proceeding at neutral pH would offer some advantages, *e.g.* with respect to random peptide bond cleavage. Time will tell.

(*iii*) *The use of aminoethylation.* Reduction and *S*-aminoethylation of cysteine and cystine residues in proteins provides a means of promoting

[161] M. L. Ludwig and M. J. Hunter, *Methods in Enzymology*, 1967, **11**, 595.
[162] K. Toi, E. Bynum, E. Norris, and H. A. Itano, *J. Biol. Chem.*, 1967, **242**, 1036.
[163] J. A. Yankeelov jun., C. D. Mitchell, and T. H. Crawford, *J. Amer. Chem. Soc.*, 1968, **90**, 1664.
[164] K. Takahashi, *J. Biol. Chem.*, 1968, **243**, 6171.

tryptic cleavage at those residues during structural investigations.[165] The additional bonds split can be very useful in reducing the amount of a polypeptide chain that is resistant to tryptic attack (the 'core' problem). Aminoethylation has proved very valuable in eliminating the β-chain core in studies of haemoglobin variants.[166] A procedure for finger-printing the α-chain core after chymotryptic digestion has now been given,[167] and the use of aminoethylation in distinguishing the wheat proteins gliadin and glutenin has been described.[168] A modification of methionine residues during aminoethylation can occur if the reaction mixture is acidified for long periods.[169] The product, probably a sulphonium salt, interferes with subsequent attempts to cleave at the methionine residues by means of the cyanogen bromide method.

A combination of many blocking methods has been used to enable immunoglobulin light chains to be cleaved at the cysteinyl residues only.[170] The lysines are modified by succinylation, after which the disulphide bridges are reduced and aminoethylated. The arginine residues are next blocked by reaction with 1,2-cyclohexanedione and tryptic cleavage can now occur only at the modified cystine residues, providing a very neat example of the usefulness of limited enzyme digestion after chemical modification of the polypeptide chain. Such large peptide fragments are potentially very valuable in studies of amino-acid sequence homology.

Peptide Fractionation. Detailed descriptions of the general methods of peptide fractionation are available in many recent books and reviews.[4, 5, 171] Diagonal paper electrophoresis will be considered later (p. 61). In general, column methods can handle larger quantities of material and are capable of automation [172, 173] but paper methods are somewhat simpler, faster, and capable of higher sensitivity. No one method is usually sufficient to fractionate the components of a complex mixture and an adroit combination of techniques is normally required. In this connection it is clearly advantageous to use some form of column chromatography as a first step because of its high capacity. Individual fractions from the column chromatogram can then be transferred to paper for further purification.

The advent of gel filtration of peptides on columns of synthetic cross-linked dextrans has added a new dimension to peptide separation. As fractionation depends on peptide size, it has been widely used to separate the products of the cyanogen bromide cleavage of proteins. Depending on

[165] M. A. Raftery and R. D. Cole, *J. Biol. Chem.*, 1966, **241**, 3457.
[166] J. B. Clegg, M. A. Naughton, and D. J. Weatherall, *J. Mol. Biol.*, 1966, **19**, 91.
[167] R. W. Carrell and D. Irvine, *Biochim. Biophys. Acta*, 1968, **154**, 78.
[168] J. A. Rothfus and M. J. A. Crow, *Biochim. Biophys. Acta*, 1968, **160**, 404.
[169] W. A. Schroeder, J. R. Shelton, and B. Robberson, *Biochim. Biophys. Acta*, 1967, **147**, 590.
[170] L. I. Slobin and S. J. Singer, *J. Biol. Chem.*, 1968, **243**, 1777.
[171] H. A. Sober, R. W. Hartley jun., W. R. Carroll, and E. A. Peterson, 'The Proteins,' ed. H. Neurath, 2nd edn., Academic Press, New York, 1965, **3**, 1.
[172] T. Okuyama, K. Takio, and K. Narita, *J. Biochem. (Japan)*, 1967, **62**, 624.
[173] R. L. Hill and R. Delaney, *Methods in Enzymology*, 1967, **11**, 339.

the distribution of the (usually) few methionine residues, these products are often separable entirely by judicious choice of gel. Further, gel filtration is not affected by the fact that the new *C*-terminal homoserine residues generated by the cyanogen bromide reaction can exist either as homoserine or its lactone. Separation methods that depend on charge may spuriously indicate heterogeneity under such circumstances. The interconversion of homoserine and its lactone is possible,[174] but as homoserine and its lactone rapidly form an equilibrium mixture this interconversion is probably best reserved for amino-acid analysis of homoserine-containing peptides. The curious retardation of aromatic compounds on Sephadex columns can occasionally be turned to good account. Thus, di-, tri-, and tetra-phenylalanine synthesised by *E. coli* extracts in the presence of polyuridylate can be cleanly separated in this way,[175] emerging from the column after the salt. This serves as a timely reminder that the column effluent in gel filtration should be checked for some distance after the emergence of salt.

Gel filtration in strong solutions of urea is frequently employed to separate peptide chains which would otherwise aggregate. Since in aqueous solutions of urea an equilibrium is established between urea and ammonium cyanate, the danger of carbamylation [175] is always present. Long exposure to old, and, in particular, heated, solutions of urea should be avoided. Heterogeneity of globin chains on column chromatography deriving principally from such carbamylation of the *N*-terminal valine residues has been noted.[176] It may well be that in some instances reversible reaction of the protein amino-groups with the anhydrides described above will be preferable for disaggregating peptide chains before gel filtration (value in chromatography would be much smaller because of the high negative charge on all peptides). To dispense with messy urea solutions would certainly be of great benefit!

One of the most important applications of paper methods is the preparation of peptide maps. It has been reported [177] that finger-prints of haemoglobin are much improved by the repetition of the chromatographic step; that is, after the initial electrophoresis and subsequent chromatography, the paper is dried and the same chromatographic step then re-run. The re-cycled finger-prints are now more clearly resolved. A useful sequential staining procedure for various residues has been described:[178] tests for sulphur, arginine, histidine, α-amino-groups (ninhydrin), and tyrosine are all done on the one paper, conserving time and materials. It is worth noting here that it is common experience that peptides with *N*-terminal valine or isoleucine give poor colour yield with ninhydrin. An extension of the finger-printing procedure which enables peptide maps to be obtained from

[174] R. P. Ambler, *Biochem. J.*, 1965, **96**, 32P.
[175] R. K. Bretthauer and A. M. Golichowski, *Biochim. Biophys. Acta*, 1968, **155**, 549.
[176] J. Čejka, Z. Vodrážka, and J. Salák, *Biochim. Biophys. Acta*, 1968, **154**, 589.
[177] K. C. Hoerman and K. Kamel, *Analyt. Biochem.*, 1967, **21**, 107.
[178] C. W. Easley, B. J. M. Zegers, and M. de Vijlder, *Biochim. Biophys. Acta*, 1969, **175**, 211.

very small amounts of protein has been developed further.[179] The tryptic digest of the protein is treated with dansyl chloride and then finger-printed on thin-layer plates. The peptides may be located under u.v. light and then eluted and hydrolysed to obtain the *N*-terminal residues. As little as 0·5 nmole is required.

A recurrent problem is the assignment of amide groups to residues of aspartic acid and glutamic acid. While this can be done by identifying the residues of asparagine and glutamine in each peptide by digestion with amino- or carboxy-peptidases, useful information can be derived from the electrophoretic mobility of a peptide. This has now been put on to a more quantitative basis.[180] If the electrophoretic mobility of a peptide at pH 6·5 is measured and its molecular weight is calculated from the amino-acid analysis, it is possible to deduce how many charged groups there are in the molecule. In view of the fact that many peptides are purified by paper electrophoresis this method is particularly valuable in making full use of the data obtained.

G. Diagonal Paper Electrophoresis.—No account of current methods of amino-acid sequence determination would be complete without a consideration of the applications of diagonal paper electrophoresis. While it is true that diagonal electrophoretic techniques have a comparatively long history it is only in the last few years that 'diagonal thinking', as Hartley called it, has achieved such a prominent and respectable position. This can be attributed in large part to the success of the method for locating disulphide bridges devised by Brown and Hartley.[181]

General Principles. The general principles of diagonal paper electrophoresis as enunciated by Brown and Hartley may be summarised as follows:

1. Separation of a peptide mixture by electrophoresis.
2. Treatment of the electrophoretogram so that specific modification of certain residues occurs. This modification must induce a change in charge or molecular weight.
3. Second electrophoresis at right angles to the first.

Peptides unaffected by the chemical modification will retain their original electrophoretic mobility and will therefore form a 45° diagonal line across the electrophoresis paper. Peptides with modified charges or modified molecular weights, however, will have altered electrophoretic mobilities and should lie off the diagonal. Diagonal procedures can therefore be used to purify selectively certain peptides from a very complex mixture and, as with most paper methods, are simple and require small amounts of starting protein.

A large number of diagonal techniques have been described in the past few years. These will be now considered in turn.

[179] R. S. Atherton and A. R. Thomson, *Biochem. J.*, 1969, **111**, 797.
[180] R. E. Offord, *Nature*, 1966, **211**, 591.
[181] J. R. Brown and B. S. Hartley, *Biochem. J.*, 1966, **101**, 214.

Diagonal Method for Cystine Peptides. This method and its widespread applications to the location of the disulphide bridges of proteins have been covered in detail.[181] The protein is digested so that its disulphide bridges remain intact. Pepsin is the most useful enzyme for this purpose because its acidic pH optimum ensures that no disulphide interchange can occur during the digestion. (This could be a problem with other proteolytic enzymes if the digestion is carried out above neutral pH.) The peptide mixture is subjected to paper electrophoresis at pH 6·5 and, after drying, is exposed to the vapour of performic acid in a closed vessel. This cleaves the disulphide bridges with the formation of two residues of charged cysteic acid from

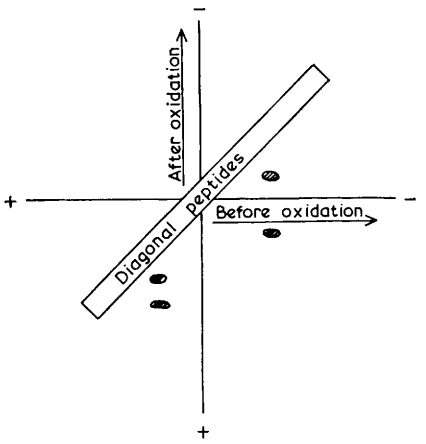


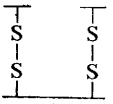
Figure 1 *Diagonal peptide map for protein with two disulphide bridges*

each residue of cystine. The oxidised paper is now dried, stitched on to a fresh sheet of paper, and submitted to a second electrophoresis at pH 6·5 at right angles to the first. On staining with ninhydrin, pairs of peptides derived from each disulphide bridge peptide in the original digest should now be seen lying off the diagonal. A hypothetical picture for a protein with two disulphide bridges is shown (Figure 1). The off-diagonal peptides can now be isolated for analysis and identification. The fact that the peptides appear in pairs identifies them as being the components of an original disulphide bridge.

Naturally, there are certain precautions that must be taken to ensure correct interpretation of the diagonal peptide maps. The neutral band at pH 6·5 may well contain a mixture of cystine peptides and it will be impossible in a two-dimensional map at pH 6·5 to assign peptides arising from this region to particular disulphide bridges. This problem can be overcome by cutting out the neutral band before oxidation, stitching it on to a fresh piece of paper, and subjecting the neutral peptides to a further electrophoretic separation at, say, pH 2 or pH 3·5. After oxidation these

peptides are re-run at the same pH to produce pH 2 or pH 3·5 diagonals from which the cysteic acid-containing peptides will separate as pairs. An alternative procedure suggested is to re-run the oxidised paper strip from the pH 2 or pH 3·5 electrophoresis at right angles at pH 6·5; neutral peptides are left at the origin and the pairs of cysteic acid peptides move to the anode.

Other complexities can arise as, for example, when pepsin fails to split the peptide chain between two half-cystine residues (8). In such a case three



(8)

peptides will appear off the diagonal, one of them containing two cysteic acid residues. In these circumstances it is necessary to elute the appropriate cystine peptide from the original pH 6·5 electrophoretogram and digest it with another proteolytic enzyme so that cleavage is effected between the two cystine residues. The disulphide bridges can now be unequivocally allocated by repetition of the standard diagonal procedure on this digest.

All these problems were met with and overcome in the original application of the method to locating the five disulphide bridges of bovine chymotrypsinogen A. Since that time the method has been successfully used to examine the amino-acid sequences around the disulphide bridges of many other proteins,[181] including lysozyme, bovine trypsinogen and chymotrypsinogen B, porcine elastase, and immunoglobulins. The only minor disadvantage is the destruction of tryptophan during the performic acid oxidation, for which a watch must be kept.

Lysine Diagonals. It is clear that the range of diagonal procedures could be extended by modifying the protein with a suitable reversible blocking group before committing it to enzymic digestion. After the initial electrophoresis the blocking groups are removed to regenerate the original amino-acid residue, thereby causing specific peptides to move off the diagonal. This idea [182] has been made the basis of a method for the rapid ordering of the lysine-containing tryptic peptides of proteins.

The protein is treated with *S*-ethyl trifluorothioacetate (p. 55) to trifluoroacetylate all the amino-groups (α- and ε-), digested enzymically, and the resulting peptides are separated by paper electrophoresis. After exposure of the peptides on the paper to ammonia vapour the electrophoresis is repeated and peptides from which a trifluoroacetyl group is removed by the ammonia treatment will vacate the 45° diagonal owing to the uncovering of an additional amino-group and consequent increased

[182] R. N. Perham and G. M. T. Jones, *European J. Biochem.*, 1967, **2**, 84.

electrophoretic mobility towards the cathode. Study of the amino-acid sequence of these peptides enables lysine-containing tryptic peptides to be placed in order. A peptide derived from the *N*-terminus of the molecule should also be found off the diagonal. In so far as many proteins contain a considerable number of lysine residues, with a commensurate number of peptides therefore moving off the diagonal, it is easier to apply this technique to the blocked peptides obtained from the tryptic digest of trifluoroacetylated proteins.

The method can be extended for studying amino-acid sequences around cysteine and cystine residues. In this application [183] these residues are first converted into *S*-aminoethylcysteine before the diagonal procedure is carried out, in which case aminoethylcysteine sequences will also be found lying off the diagonal. Various ways for distinguishing peptides containing lysine from those containing aminoethylcysteine are possible. ^{14}C-Labelled ethyleneimine can be used for the aminoethylation reaction. Off-diagonal peptides containing aminoethylcysteine would then be radioactive and could easily be resolved from lysine peptides by autoradiography of the peptide maps. This technique would have the additional advantage of making it less easy to overlook an off-diagonal peptide as a result of the complexity of the peptide map. An alternative method would be to block the protein amino-groups before conversion of the cysteine or cystine residues into aminoethylcysteine. Providing the blocking group chosen is not removed by the ammonia treatment in the diagonal procedure, the only peptides lying off the diagonal now should be related to the aminoethylcysteine residues. The method cannot be used for locating disulphide bridges; on the other hand, the loss of tryptophan associated with performic acid oxidation is avoided.

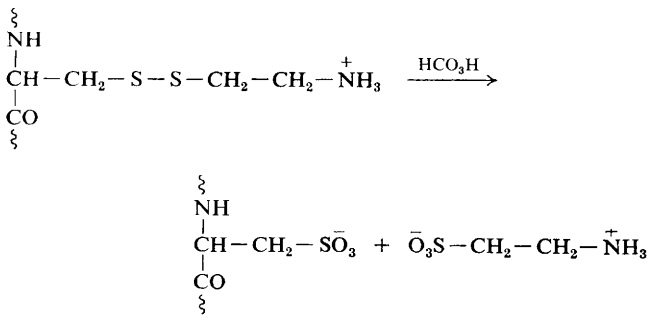
Other reversible blocking-groups for amino-groups should be suitable candidates for preparing lysine diagonals, in particular, maleic [157] and citraconic [158] anhydrides. Lysine diagonals have been successfully used to order the tryptic peptides of pepsin,[182] in helping to establish the amino-acid sequence of glyceraldehyde 3-phosphate dehydrogenases,[184, 185] and in studying the modification of protein amino-groups.[153] In view of the development of techniques for the specific tryptic cleavage of peptide chains at aminoethylcysteine residues only,[170] methods for the study of amino-acid sequence around these residues could prove valuable in the future.

Diagonal Methods for Cysteine Residues. A possible extension of the lysine diagonal technique for the study of amino-acid sequence around cysteine and cystine residues has already been mentioned.[183] An alternative method for the selective purification of cysteine peptides has also been described.[186]

[183] R. N. Perham, *Biochem. J.*, 1967, **105**, 1203.
[184] B. E. Davidson, M. Sajgo, H. F. Noller, and J. I. Harris, *Nature*, 1967, **216**, 1181.
[185] J. I. Harris and R. N. Perham, *Nature*, 1968, **219**, 1025.
[186] A. G. Weeds and B. S. Hartley, *Biochem. J.*, 1968, **107**, 531.

The thiol groups of a protein are treated with cystine by disulphide–thiol interchange, which converts all cysteine residues of the protein into cystine. After enzymic digestion, the cystine-containing peptides can be separated by the performic acid diagonal technique. If radioactive cystine is used for the exchange reaction, it is less easy to overlook potential off-diagonal peptides in any preliminary fractionation of the peptides from complicated proteins, *e.g.* myosin, although the label is, of course, lost as cysteic acid during the performic acid oxidation. This approach to isolating the cysteine peptides of proteins has been used to characterise the thiol groups of rabbit muscle phosphorylase.[187] Whereas amino-acid analysis suggests a total of 18 cysteic acid residues per mole of phosphorylase *b* (*M* 185,000), only 8–9 unique cysteine sequences can be established by the diagonal method with radioactive cystine, indicating that the phosphorylase *b* molecule consists of two more or less identical subunits.

A further variation of the method is also available. If the thiol groups are labelled by exchange with cystamine there is a charge change of two units per thiol group reacted when the susceptible peptides are cleaved with performic acid (Scheme 13), which means that they move further and



Scheme 13

more cleanly away from the diagonal position. When this technique is applied to rabbit muscle myokinase (ATP–AMP phosphotransferase) and the labelled protein is digested with trypsin, two cysteine residues react and are found to reside in the following sequences:[188]

(*a*) Val-Asn-Ala-Glu-Gly-Ser-Val-Asp-Asn-Val-Phe-Ser-Gln-Val-Cys-Lys

(*b*) Gly-Thr-Gln-Cys-Glu-Lys

There is some evidence that bonds involving the cystamine residues, like those of aminoethylcysteine, are susceptible to tryptic cleavage.

[187] C. G. Zarkadas, L. B. Smillie, and N. B. Madsen, *J. Mol. Biol.*, 1968, **38**, 245.
[188] A. G. Weeds and L. Noda, *Biochem. J.*, 1968, **107**, 311.

Other Diagonals. One of the earliest diagonal techniques involved the change in electrophoretic mobility of histidine peptides after photo-oxidation[189] and another[190] depends on the formation of the positively charged sulphonium derivative when the neutral side-chain of methionine is reacted with iodoacetamide.

Enzymes can also be used to effect the chemical modification of peptides on paper. Thus, if carboxypeptidase B is used to digest tryptic peptides on paper between the successive electrophoreses of a diagonal procedure, the only peptide remaining on the diagonal will be the *C*-terminal peptide.[191] This is because carboxypeptidase B is specific for *C*-terminal lysine and arginine residues and the only peptide in a tryptic digest that does not terminate with a basic residue should be the one derived from the *C*-terminus of the molecule (supposing the protein itself does not terminate in either lysine or arginine). Dephosphorylation with alkaline phosphatase can be used to isolate peptides deriving from serine *O*-phosphate residues in proteins.[192] In an application to the serine phosphate peptides of ovalbumin, it has been found that there are two distinct sequences, (*a*) and (*b*):

(*a*) -Ala-Gly-Arg-Glu-Val-Val-Gly-Ser(P)-Ala-Glu-Ala-Gly-Asp-Val-Ala-Ala-Ser-(Val, Glx, Glx, Ser, Phe)-Arg-

(*b*) -Asp-Lys-Leu-Pro-Gly-Phe-Gly-Asp-Ser(P)-Ile-Glx-Ala-Glx-Cya-Gly-(Thr, Ser, Val)-(Asp, His, Val)-

[Ser(P) = *O*-phosphorylserine, Cya = cysteic acid]

Whereas sequence (*a*) is susceptible to prostatic phosphatase, sequence (*b*) is resistant.

There are clearly many occasions when a diagonal electrophoretic procedure can be used to good account for a particular purpose. One of the most interesting examples of such use is in the study of the reaction of pepsin with the active-site-directed irreversible inhibitor *N*-diazo-[1-^{14}C]acetyl-L-phenylalanine methyl ester.[193] This reagent reacts with a single aspartic acid residue at the active site of pepsin. After peptic digestion of the labelled enzyme, the peptides are separated by electrophoresis at pH 3·5, treated with triethylamine, and re-run at pH 3·5. This treatment unmasks the carboxyl group of the blocked aspartic acid residue and the appropriate peptide therefore moves off the diagonal; the attachment of the inhibitor is almost certainly an ester bond to the β-carboxyl group of the aspartic acid. The amino-acid sequence around the aspartic acid is now known to be:[194]

-Ile-Val-Asp-Thr-Gly-Thr-Ser-

[189] C. Milstein and F. Sanger, *Biochem. J.*, 1961, **79**, 456.
[190] J. Tang and B. S. Hartley, *Biochem. J.*, 1967, **102**, 593.
[191] M. A. Naughton and H. Hagopian, *Analyt. Biochem.*, 1962, **3**, 276.
[192] C. P. Milstein, *Biochem. J.*, 1968, **110**, 127.
[193] R. S. Bayliss and J. R. Knowles, *Chem. Comm.*, 1968, 196.
[194] J. R. Knowles and G. B. Wybrandt, *F.E.B.S. Letters*, 1968, **1**, 211.

H. General Conclusions.—The conventional methods of amino-acid and sequence analysis have witnessed steady development until it is now possible to undertake amino-acid sequence determinations on what only a few years ago would have been regarded as vanishingly small quantities of material. Some of these methods are potentially capable of even further development and no doubt further improvements will be made. In this connection, the dansyl–Edman procedure clearly occupies a central position and the perfection of g.l.c. analysis of amino-acids would also be of benefit.

There is no obvious prospect of mass spectrometry being applied to the amino-acid sequence analysis of large peptides but, clearly, it could make a big contribution to work with peptides arising from enzymic or chemical digestion of proteins. In this sense it is complementary to sequential degradation in the protein sequenator, which has only been applied successfully to large polypeptide chains. In any event, all attempts to provide automated sequence determination are welcome and the discovery of an enzyme, pyrrolidone peptidase, which acts to remove pyrrolidone carboxylic acid from the *N*-terminal position in peptides will be of great value in enabling Edman degradations to proceed in peptides and proteins previously inaccessible to the technique.

Not all investigations of primary structure are aimed at establishing the complete amino-acid sequences of proteins. Thus, studies of the active sites of enzymes, of amino-acid sequence homology between related proteins, and of the effect of chemical modification, frequently demand the isolation of specific peptides, often from complex mixtures. It is clear that diagonal electrophoretic procedures have a particularly important part to play here and, indeed, have already made major contributions to the subject.

It is also worth remarking that establishing the amino-acid sequence of a peptide obtainable even in only very small amounts is no longer a major problem. It is the fractionation of complex mixtures of peptides that continues to pose difficulties as the study of primary structure moves to more and more complicated proteins. Methods for the sequential degradation of proteins and improved limited cleavage of large peptide chains go some way to combat this problem but it may be that radical solutions will be required eventually. One method that would eliminate entirely the need for peptide separation has recently been discussed.[195] Sequential degradation is applied to a mixture of peptides deriving from the digestion of a protein and the results are compared with those obtained from a different mixture produced by a different digestion procedure. If enough mixtures are examined, it should be possible, by knowing which amino-acids can occur next to one another, to arrive at a unique amino-acid sequence for the protein. Certain contingencies might arise—'inaccessible' sequences could occur because cleavage in them was not secured by any of

[195] W. R. Gray, *Nature*, 1968, **220**, 1300.

the digestion procedures used. Presumably the use of alternative digestion techniques or the extension of the ability to degrade sequentially should remedy this. Computer comparison of the degradation results is clearly called for with large proteins. Theoretical analysis of the established amino-acid sequence of several proteins and their cleavage products with proteolytic enzymes suggests that the proposal is feasible but a practical demonstration is awaited.

The move to automated amino-acid sequence determination is definitely with us and the next few years will doubtless see this movement gather momentum. Given the almost infinite variety of which proteins are capable and the widely differing reactivities and stabilities of the amino-acid side-chains, the development of automated sequence analysis is an awesome task. Proteins have so far managed, on occasion, to confound those who would rely on a single technique for the determination of amino-acid sequence. How long they will continue to resist in this fashion is just one of the interesting questions facing the protein chemist at this time.

3 Structural Proteins

The term 'structural proteins' is a convenient title under which to include the fibrous proteins and those globular proteins that serve structural roles, *e.g.* actin. Histones are also included here for convenience. Studies on the fibrous proteins are often complicated by the wide variety of unusual cross-links that may join subunits either to one another or, perhaps, to another tissue component. Such cross-links, *e.g.* in elastin, sometimes make it impossible to recognise and isolate the precursor subunits, which makes study of primary structure even more difficult.

In terms of amino-acid sequence, it has become abundantly clear that structural proteins are no exception to the general rule[196] that primary structure governs three-dimensional structure. For example, every third residue in the collagen chain must be glycine if the desired triple helix is to form.[197] Because of the large size of the polypeptide chains in many structural proteins, it is no surprise that cyanogen bromide cleavage at (the rare) methionine residues or studies of amino-acid sequence around selected residues, *e.g.* cysteine, are in such vogue at present.

Structural proteins have been the subject of an excellent detailed review.[198]

A. The Structural Proteins of Muscle.—It is evident that in striated muscle contraction occurs by a sliding mechanism in which actin filaments move inwards with respect to myosin filaments in the myofibril, resulting in a

[196] C. J. Epstein, R. F. Goldberger, and C. B. Anfinsen, *Cold Spring Harbour Symp. Quant. Biol.*, 1963, **28**, 439.

[197] R. E. Dickerson, 'The Proteins,' ed. H. Neurath, 2nd edn., Academic Press, New York, 1964, p. 603.

[198] S. Seifter and P. M. Gallop, 'The Proteins,' ed. H. Neurath, 2nd edn., Academic Press, New York, 1966, p. 153.

shortening of fibril length. Much energy is currently being devoted to establishing the structure of these proteins in order to account for this remarkable process.[199] Discussion here will be limited to actin and myosin which together comprise the bulk of the structural protein of the myofibril. All the work presented concerns the rabbit proteins unless otherwise stated.

Actin. Actin can be isolated from muscle in a globular form, G-actin, which in the presence of dilute aqueous salt solutions is transformed to a polymeric form, F-actin. Each molecule of G-actin contains a bound molecule of ATP which is dephosphorylated to ADP during the polymerisation. The molecular weight of G-actin has previously been estimated as *ca.* 60,000, but estimations by physical methods are complicated by the tendency of the monomer to polymerise. Several investigations have now indicated that the G-actin molecule is a single polypeptide chain with a molecular weight of *ca.* 45,000.

Examination of reduced, *S*-carboxymethylated G-actin by centrifugation in concentrated solutions of guanidine hydrochloride gives a molecular weight of *ca.* 47,000, in good agreement with the molecular weight from estimations of protein-bound nucleotide.[200] The presence of only one band on gel electrophoresis in 8M-urea suggests that only one peptide chain is present. These results are in good accord with chemical studies [201] on actin. Treatment of actin with carboxypeptidase A in the presence of 6M-urea liberates 1 mole of *C*-terminal phenylalanine per 43,000 g. of protein. Three peptides containing carboxymethylcysteine were also isolated from tryptic digests of carboxymethylated actin.[201] One, which exhibits no free α-amino-group, contains 17–18 residues and is tentatively identified with the *N*-terminal sequence of the actin molecule, which is thought [202] to have a terminal *N*-acetylaspartic acid residue. Another, Cys(Cm)-Phe is assigned to the *C*-terminus of the molecule. As amino-acid analysis based on a molecular weight of 45,000 indicates the presence of five cysteine residues in the actin subunit, two more carboxymethylcysteine-containing peptides remain to be isolated. Similar conclusions based on finger-printing and amino-acid analysis of bovine carotid actin have also been presented.[203] The actin filament is believed to consist of two regularly intertwining strands composed of these subunits.[199]

In addition to the usual amino-acids, both bovine carotid [203] and rabbit [204] actin contain 3-methylhistidine. A little over 10% of the histidine residues

[199] S. V. Perry, *Progr. Biophys. Mol. Biol.*, 1967, **17**, 325; H. E. Huxley, *Sci. American*, 1965, **213**, pt. 6, 18.
[200] M. K. Rees and M. Young, *J. Biol. Chem.*, 1967, **242**, 4449.
[201] P. Johnson and S. V. Perry, *Biochem. J.*, 1968, **110**, 207.
[202] E. Gaetjens and M. Barany, *Biochim. Biophys. Acta*, 1966, **117**, 176; R. E. Alving and K. Laki, *Fed. Proc.*, 1966, **25**, 223.
[203] C. Gosselin-Rey, C. Gerday, A. Caspar-Godfroid, and M. E. Carsten, 1968, in the press.
[204] P. Johnson, C. I. Harris, and S. V. Perry, *Biochem. J.*, 1967, **105**, 361.

are present in this form. 3-Methylhistidine is also found in myosin to the extent of *ca.* 2 residues per mole,[204] but has not so far been reported in any other protein. Its role in the stuctural proteins of muscle remains obscure.

As a nice example of how similar proteins seem to fulfil different functions in different circumstances, considerable resemblance has been reported [205] between actin, the outer-fibre protein of cilia, and the colchicine-binding protein associated with cytoplasmic microtubules. It is of note that whereas actin contains bound ADP, the latter proteins contain a bound guanine nucleotide.

Myosin (see also p. 172). There has been some disagreement until recently about the number of polypeptide chains comprising the myosin molecule.[198] Most of the disagreement has concerned the molecular weight of myosin in and out of strong solutions of guanidine hydrochloride. This type of measurement is notoriously difficult to make and the argument has revolved round a triple α-helical coiled-coil with a molecular weight of 600,000 and a double α-helical coiled-coil with a molecular weight of *ca.* 470,000. This argument now seems to have been resolved in favour of the latter. Several pieces of evidence contribute to the conclusion.

To establish the complete amino-acid sequence would brook no uncertainty in calculating the molecular weight of the polypeptide chain. Equally, in view of the putative molecular weight of over 200,000, this would be a monumental labour. However, valid conclusions can be reached by establishing the number of unique amino-acid sequences containing a particular type of residue and then calculating the molecular weight from an amino-acid composition based on that residue. This has now been done for the cysteine residues either by labelling with [^{14}C]iodoacetic acid or by using the thiol-disulphide exchange reaction and cystine diagonal technique [186] (p. 6). Taken together, these results indicate [186] that some 22 unique thiol sequences can be detected, which clearly supports a two-chain structure for myosin in which there are *ca.* 42 thiol groups for a corresponding molecular weight of *ca.* 500,000. These conclusions are in much better agreement also with the results of electron microscopy of the myosin molecule. It is generally agreed that the myosin molecule is composed of a globular head *ca.* 200 Å × 50 Å attached to a rod-like tail 1300–1400 Å long and *ca.* 20 Å wide. Beautiful electron micrographs obtained by a rotating shadowing technique have recently been published [206] in which the globular head is seen to comprise two subunits, a result obviously more compatible with a two- rather than a three-chain model for myosin.

In fact, certain refinements have to be made to this basic two-chain model. The presence of four *C*-terminal isoleucine residues in 1 mole

[205a] F. L. Renaud, A. J. Rowe, and I. R. Gibbons, *J. Cell. Biol.*, 1968, **36**, 79. [b] R. E. Stephens, *Quart. Rev. Biophysics*, 1969, **1**, 377.
[206] H. S. Slayter and S. Lowey, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 1611.

(530,000 g.) of myosin has been reported.[207] It now seems clear that the myosin molecule consists of two long polypeptide chains (*M ca.* 220,000) and two or three short chains (*M ca.* 20,000) which can be dissociated in guanidine hydrochloride or alkali. The exact standing of these small chains is still a matter for some conjecture and is currently arousing much interest.[208]

Controlled proteolytic digestion can separate the globular head from the rod-like coiled-coil tail [209] providing a useful preliminary step in studies of the primary structure of myosin. No free *N*-terminal groups can be detected in myosin, in accord with the isolation of acetylated peptides derived from the *N*-terminus of the long chains.[210] These results unexpectedly show that there are two *N*-terminal sequences:

N-acetyl-Ser-Ser-Asp-Ala-Asp-Met-Ala-*Ile*-Phe- and
N-acetyl-Ser-Ser-Asp-Ala-Asp-Met-Ala-*Val*-Phe

The origin of this heterogeneity is unknown but could reflect slight differences in the amino-acid sequences of the heavy chains. The nature of the blocking at the *N*-termini of the light chains remains obscure. Since acetyl peptides can be isolated from the globular part of the myosin molecule but not from the rod part, it is concluded that the constituent polypeptide chains run parallel in the rod-like region with their *N*-terminal ends in the globular head.

B. Collagen.—The structure of collagen has been considered in detail in many reviews [197, 198] and books.[211] Since that time big strides have been made in examining the primary structure of collagen and the nature of the cross-links which bind its polypeptide chains.

It is commonly accepted that the fundamental unit of the collagen fibril is a molecule, tropocollagen, the dimensions of which are *ca.* 3000 Å long and 15 Å across, with a molecular weight of *ca.* 300,000. End-to-end and side-by-side aggregation of these soluble subunits forms the collagen fibril found in connective tissue. Since newly formed collagen becomes increasingly insoluble with the passage of time, various cross-linking reactions within and between the tropocollagen units have been postulated to occur. There are generally two kinds of peptide chain in the tropocollagen molecule, designated $\alpha 1$ and $\alpha 2$, and the intact molecule is assigned the structure $(\alpha 1)_2(\alpha 2)_1$. Intramolecular cross-linking leads to the formation of covalently linked two-chain components, designated β_{12} ($\alpha 1$–$\alpha 2$) and β_{11} ($\alpha 1$–$\alpha 1$), which appear to varying degrees in different preparations of denatured collagens. Cross-linked three-chain (γ) components are also

[207] J. Sarno, A. Tarendash, and A. Stracher, *Arch. Biochem. Biophys.*, 1965, **112**, 378.
[208] P. P. Trotta, P. Dreizen, and A. Stracher, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 659.
[209] S. Lowey, H. S. Slayter, A. G. Weeds, and H. Baker, *J. Mol. Biol.*, 1969, **42**, 1.
[210] G. W. Offer and R. L. Starr, results presented at I.U.P.A.B. Meeting on 'Interactions between Subunits of Biological Macromolecules,' Cambridge, June, 1968.
[211] 'Treatise on Collagen,' ed. G. N. Ramachandran, Academic Press, London, 1967.

known. In rare cases, *e.g.* codfish skin collagen, it is thought that a third α-component is present, giving a tropocollagen structure $(\alpha 1)_1(\alpha 2)_1(\alpha 3)_1$.

The Primary Structure of Collagen. It is clearly important to know what distinguishes the α1-chain from the α2-chain. Although much work has been done in the past with proteolytic enzymes, the size (*ca.* 1000 residues per chain) and the repetitive nature of much of the sequence have made such studies very difficult. However, it is clear that glycine occurs as every third residue throughout most of the collagen molecule, Gly-Pro-Hyp-Gly and Gly-Pro-Ala being particularly predominant sequences. The proline and hydroxyproline residues contribute much to the stability of the molecule. The hydroxyproline (3- and 4-) is, of course, formed by enzymic hydroxylation of proline residues already in peptide linkage.

Perhaps the biggest step forward in recent work has been the application of the cyanogen bromide cleavage reaction to a study of the separated α1- and α2-chains.[212, 213] With the α1-chain of rat skin collagen, amino-acid analysis based on a molecular weight of 95,000 indicates the presence of seven methionine residues. Eight unique peptides, which together add up to the starting composition, have now been isolated by gel filtration and ion-exchange chromatography,[213a] and their order studied by electron microscopy.[213b] The *N*-terminal peptides from the α1- and α2-chains of rat skin collagen have been identified and compared [212] as follows:

(α1) Gly-**Tyr**-Asp-Glu-**Lys**-Ser-Ala-Gly-Val-Ser-Val-**Pro-Gly-Pro-Met-**
1 **10**

(α2) Glp-**Tyr**-Ser-Asp -**Lys**-Gly-Val-Ser-Ala-Gly -----**Pro-Gly-Pro-Met-**
1 **10**

and the amino-acid sequence of the succeeding 36 residues of the α1-chain has also been described:[214]

↓ ↓ ↓ ↓ ↓
Gly-Pro-Ser-Gly-Pro-Arg-Gly-Leu-Hyp-Gly-Pro-Hyp-Gly-Ala-Hyp-
20 **30**

↓
Gly-Pro-Gln-Gly-Phe-Gln-Gly-Pro-Hyp-Gly-Glu-Hyp-Gly-Glu-Hyp-
40

↓
Gly-Ala-Ser-Gly-Pro-Met-
50

The arrows mark the points of cleavage by collagenase. Since the *N*-terminal sequence lacks hydroxyproline and does not contain every third residue as glycine, it may be assumed that its structure is not that of the collagen triple helix. On the other hand, it is apparent that the amino-acid sequence necessary to the helix starts within a few residues of the

[212] A. H. Kang, P. Bornstein, and K. Piez, *Biochemistry*, 1967, **6**, 788.
[213a] W. T. Butler, K. A. Piez, and P. Bornstein, *Biochemistry*, 1967, **6**, 3771. [b] J. Rauterberg and K. Kühn, *F.E.B.S. Letters*, 1968, **1**, 230.
[214] P. Bornstein, *J. Biol. Chem.*, 1967, **242**, 2572.

N-terminus of the molecule. On the basis of these experiments it seems that the α1-chains must be identical, or at least very similar, and that there is a unique sequence for the entire length of the thousand or so amino-acids, with the exception of partial hydroxylation of certain proline and lysine residues.[213a, 214, 215] Further, there are clear differences in amino-acid sequence between the α1- and α2-chains.

The cyanogen bromide cleavage technique is also being applied to other collagens. Thus, when the corresponding fragments of rat skin and human skin collagens are compared [216] by amino-acid analysis and finger-printing, it is found that the fragments derived from the helical regions are very similar, suggesting a strong conservation of amino-acid sequence. Conversely, the *N*-terminal fragments may vary quite widely in sequence. An interesting comparison is that between collagens taken from different tissues of the same animal. Differences have been described between the collagen of rat skin and of rat tendon.[217] In addition to minor differences in proline hydroxylation, it appears that the α1-chain from rat skin collagen lacks the sequence Glx-Met-Ser-Tyr- present at the *N*-terminus of the α1-chain of collagen from rat tendon. Although this may be a genuine tissue difference, it has also been suggested [217] that an enzyme might exist to cleave off specifically these extra four residues. In this connection, it is also of interest that a similar addition of four residues has been reported [218] at the *N*-terminus of the α1-chain of chick bone collagen.

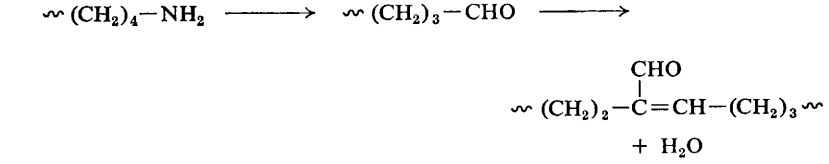
Given these advances, there now seems to be no reason why a complete derivation of the amino-acid sequence of the component chains of collagen should not be possible.

The Nature of the Collagen Cross-links. This thorny problem has been the subject of a recent review by Piez.[219] Many bonds have been considered in the past, including esters, carbohydrate bridges, and bonds involving the ε-amino-groups of lysine and the ω-carboxyl groups of aspartic and glutamic acids. Few of these still seem likely.

One inter-chain bond which has been substantiated is that involving the δ-semialdehyde of α-aminoadipic acid. It was argued [220] that a comparison of proteolytic digests of single-chain (α) and double-chain (β) components of rat skin collagen should show differences in the fragments produced that could be used to identify the site and nature of covalent cross-links. Examination of the *N*-terminal cyanogen bromide cleavage fragments of the α1-, α2-, and β_{12}-components showed that the cross-linking was restricted to the non-helical *N*-terminal fragments of the chains. In digests of older, more highly cross-linked collagen, the *N*-terminal peptides of the

[215] W. T. Butler, *Science*, 1968, **161**, 796.
[216] P. Bornstein, *Science*, 1968, **161**, 592.
[217] P. Bornstein, *Biochemistry*, 1969, **8**, 63.
[218] E. J. Miller, J. M. Lane, and K. A. Piez, *Biochemistry*, 1969, **8**, 30.
[219] K. A. Piez, *Ann. Rev. Biochem.*, 1968, **37**, 547.
[220] P. Bornstein and K. A. Piez, *Biochemistry*, 1966, **5**, 3460.

α1- and α2-chains are found to exist partly in a form where the single lysine residue is converted to the δ-semialdehyde of α-aminoadipic acid and it is therefore suggested [212, 220] that the cross-link involves an aldol condensation and subsequent dehydration between two lysine-derived δ-semialdehydes on adjacent chains (Scheme 14). In view of the established



Scheme 14

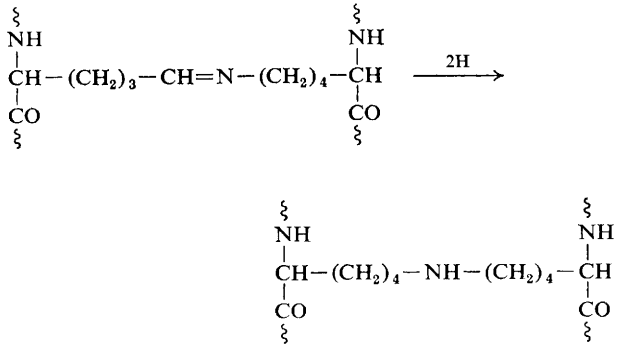
N-terminal sequences of the α1- and α2-chains,[212] it is clear that it must be the single lysine residue at the 5-position which is concerned. Tracer experiments involving radioactive lysine to follow the biosynthesis of the intramolecular cross-link [221] are also in accord with the participation of two lysine-derived aldehydes in one aldol-type link. Partial hydroxylation of the corresponding lysine residue in chick bone collagen has been noted.[217] The significance of this for cross-linking is obscure.

Calf and rat skin collagens probably differ in the distribution of cross-links. While the links just described are found to be at the *N*-terminal (A) end of rat skin tropocollagen, it has been reported,[222] on the basis of electron microscopic examination of the fragments remaining after enzymic and chemical proteolysis, that in calf skin tropocollagen cross-links can occur at both ends of the molecule. Thus it would seem dangerous to extrapolate too much from one type of collagen to another.

While the intramolecular cross-links detailed above are obviously very important, they cannot contribute much to the mechanical strength of the collagen fibre. Intermolecular cross-links must exist to provide this strength and to account for the existence of three-chain components such as γ_{111} and γ_{222} in preparations of denatured aged collagen. There is no direct evidence as to the nature of these links, for when they occur the resultant collagen becomes very insoluble and difficult to work with. However, in animals suffering from lathyrism there is a loss in the intra- and inter-molecular cross-links. Lathyrism may be produced experimentally in animals by feeding β-aminopropionitrile, and since this affects both intra- and inter-molecular cross-linking processes it would seem likely that they proceed by related mechanisms. Lathyrogens are thought to prevent the formation of intramolecular links by inhibiting the conversion of the susceptible lysine residue into the corresponding aldehyde [220] and could

[221] M. Rojkind, L. Rhi, and M. Aguirre, *J. Biol. Chem.*, 1968, **243**, 2266.
[222] M. P. Drake and P. F. Davison, *J. Biol. Chem.*, 1968, **243**, 2890.

presumably affect intermolecular bonds in the same way. There is some evidence that these intermolecular bonds involve intermediate Schiff base formation. For example, in a study of the changes in fibre strength which accompany chemical treatment,[223] it has been found that prior reduction completely inhibits cleavage of the intermolecular bonds by acid, alkali, and β-aminothiols. It is concluded that reversible Schiff base intermediates occur which are gradually stabilised with age. Since it appears that lysine-derived aldehydes occur only near the *N*-terminus of the tropocollagen molecule,[220] it is likely that Schiff base formation occurs between an aldehyde at the *N*-terminal region of one molecule and some specific ε-amino-group of another lying alongside in the collagen fibril. If the Schiff base were subsequently reduced, a residue of lysinonorleucine would result (Scheme 15) but there is no direct evidence yet for this.[219]



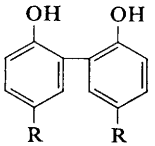
Scheme 15

Other possible cross-links have also come in for further recent consideration. Much of the carbohydrate of soluble collagen occurs as a disaccharide in *O*-glycosidic linkage with a hydroxylysine residue,[224] but there is no evidence that this prosthetic group contributes to the formation of insoluble collagen. Dityrosine, possibly formed by the action of peroxidases, has been reported in soluble collagen [225] and α-aminoaldehydes of glycine, alanine, aspartic acid, lysine, serine, and threonine have been found in various collagens,[226] although the role of aldehydes other than that of lysine is unclear.

Thus, while the chemistry of the intramolecular cross-links is now comparatively well defined, much work remains to be done on the nature of the intermolecular links and the biochemical changes involved in the formation of the highly cross-linked insoluble collagen.

[223] A. J. Bailey, *Biochim. Biophys. Acta*, 1968, **160**, 447.
[224] L. W. Cunningham and J. D. Ford, *J. Biol. Chem.*, 1968, **243**, 2390.
[225] F. LaBella, P. Waykole, and G. Queen, *Biochem. Biophys. Res. Comm.*, 1968, **30**, 333.
[226] P. M. Gallop, O. O. Blumenfeld, E. Henson, and A. L. Schneider, *Biochemistry*, 1968, **7**, 2409.

Some Comparisons with Other Structural Proteins. Despite the fact that the function of collagen demands that it should be inextensible, there are several interesting similarities between collagen and the rubber-like proteins of connective tissue, resilin, and elastin.[198, 219] Elastin, which is found only in vertebrates, resembles collagen in having *ca.* $\frac{1}{3}$ of its residues as glycine and in containing hydroxyproline, although there is no information about the amino-acid sequence. On the other hand, the nature of the cross-links is known in some detail. One of the best characterised cross-links is that involving the condensation of δ-semialdehydes of lysine residues to form desmosine, isodesmosine, and a probable intermediate, merodesmosine.[227] The obvious connection between the formation of cross-links in collagen and in elastin through reactions of the δ-semialdehyde of α-aminoadipic acid is strengthened by the fact that lathyrogens inhibit cross-linking in both elastin and collagen, presumably by the same mechanism.[219] The formation of lysinonorleucine in elastin is also recorded [228] and postulated to occur *via* the same δ-semialdehyde (Scheme 15), although this has not yet been found in collagen.[219] Another possible cross-link in common is that of dityrosine (9) discovered in insect resilin,[229] and since reported for vertebrate elastin [230]



(9)

and collagen.[225] Thoughts such as these have led to the suggestion [219] that a collagen may have been an evolutionary precursor of elastin, which, since it is confined to vertebrates, probably arose late in evolution.

In contrast with collagen, the keratins [198] present an entirely different picture. Examination of component 7, a low-sulphur protein derived from the protofibrillar portion of wool, has indicated that it is probably a family of related proteins [231] with little hope of resolution to unique polypeptide chains. It is suggested that a pattern of amino-acid replacement in non-critical regions may mask a fundamental similarity, a type of problem already met with in the immunoglobulins (p. 104). A recent investigation of the methionine sequences in the low-sulphur proteins [232] is also in accord with the presence of proteins of *M ca.* 45,000 with acetylated *N*-terminal

[227] S. M. Partridge, *Fed. Proc.*, 1966, **25**, 1023; B. C. Starcher, S. M. Partridge, and D. F. Elsden, *Biochemistry*, 1967, **6**, 2425.
[228] C. Franzbleau, F. Marrot Sinex, B. Faris, and R. Lampidis, *Biochem. Biophys. Res. Comm.*, 1965, **21**, 575.
[229] S. O. Andersen, *Biochim. Biophys. Acta*, 1964, **93**, 213.
[230] F. LaBella, F. Keeley, S. Vivian, and D. Thornhill, *Biochem. Biophys. Res. Comm.*, 1967, **26**, 748.
[231] E. O. P. Thompson and I. J. O'Donnell, *Austral. J. Biol. Sci.*, 1967, **20**, 1001.
[232] R. Hosken, B. A. Moss, I. J. O'Donnell, and E. O. P. Thompson, *Austral. J. Biol. Sci.*, 1968, **21**, 593.

residues, contrary to the protein of *M* 75,000 containing high- and low-sulphur moieties previously thought to exist. A large number of amino-acid sequences derived from performic acid-oxidised wool have been established.[233] Hosken and his collaborators [232] indicate that one of the two methionine sequences reported in that study [233] should in fact be:

-Glu-Thr-Met-Gln-Phe-Leu-Asp-Asp-Arg-

C. Fibrinogen.—The fibrinogen molecule has a molecular weight of *ca.* 340,000 and contains three different kinds of polypeptide chains, A, B, and C. Each chain occurs twice in the fibrinogen molecule to form a structure $A_2B_2C_2$. The enzyme thrombin cleaves off fibrinopeptides A and B from the *N*-terminus of the A- and B-chains respectively and polymerisation to the fibrin clot ensues.[234] The thrombin cleavage occurs at a specific Arg-Gly bond in the A- and B-chains, and the amino-acid sequences of fibrinopeptides from many different animals have been determined.[1] The removal of fibrinopeptide A is essential for the early steps of polymerisation; that of B is not. It is interesting that more variation in amino-acid sequence has occurred in peptide B than A during evolution [235] and that the tyrosine residue at position B-6 of some species is found as the *O*-sulphate derivative.[1, 236]

Two approaches to further study of the primary structure of fibrinogen have been profitable. Digestion with cyanogen bromide has enabled a fragment containing the *N*-terminal sequences of the A-, B-, and C-chains to be isolated.[237, 238] Analysis of this fraction shows it to be a 'disulphide-knot' consisting of the *N*-terminal fragments of one A-, one B-, and one C-chain held together in multiple disulphide linkage. Clearly, there must be two such 'knots' per molecule of fibrinogen. The 'knot' has a molecular weight of 26,000 and still retains its susceptibility to thrombin, releasing the fibrinopeptides: it should prove very useful, therefore, in further investigation of the thrombin-catalysed fibrinogen–fibrin polymerisation.[238] An alternative procedure [239] has been to digest *S*-sulphofibrinogen with plasmin, an enzyme that somewhat resembles trypsin in its specificity but cleaves fewer bonds (p. 51). Long *N*-terminal sequences for the A-,[237, 239] B-,[237, 239] and C-[240] chains of human fibrinogen have now been derived by these techniques and are shown in Figure 2. The serine residue at position-3 in the A-chain can occur as the *O*-phosphate and carbohydrate is probably attached to Asn-52 in the C-chain. Independent evidence of an identical

[233] M. C. Corfield, J. C. Fletcher, and A. Robson, *Biochem. J.*, 1967, **102**, 801.
[234] L. Lorand, *Fed. Proc.*, 1965, **24**, 784.
[235] R. F. Doolittle and B. Blombäck, *Nature*, 1964, **202**, 147.
[236] T. Krajewski and B. Blombäck, *Acta Chem. Scand.*, 1968, **22**, 1339.
[237] B. Blombäck, M. Blombäck, B. Hessel, and S. Iwanaga, *Nature*, 1967, **215**, 1445.
[238] B. Blombäck, M. Blombäck, A. Henschen, B. Hessel, S. Iwanaga, and K. R. Woods, *Nature*, 1968, **218**, 130.
[239] S. Iwanaga, P. Wallén, N. J. Gröndahl, A. Henschen, and B. Blombäck, *European J. Biochem.*, 1969, **8**, 189.
[240] S. Iwanaga, B. Blombäck, N. J. Gröndahl, B. Hessel, and P. Wallén, *Biochim. Biophys. Acta*, 1968, **160**, 280.

A-*chain*

↓

Ala-Asp-Ser-Gly-Glu-Gly-Asp-Phe-Leu-Ala-Glu-Gly-Gly-Gly-Val-Arg-Gly-Pro-Arg-Val-Val-Glu-Arg-His-Gln-Ser-Ala-Cys-Lys-Asp-

1 **10** **20** **30**

Ser-Asp-Trp-Pro-Phe-(Cys, Ser, Asp, Glu, Trp, Asn, Tyr)-Lys-Cys-Pro-Ser-Gly-Cys-Arg-Met-

40 **50**

B-*chain*

↓

Glp-Gly-Val-Asn-Asp-Asn-Glu-Glu-Gly-Phe-Phe-Ser-Ala-Arg-Gly-His-Arg-Pro-Leu-Asp-Lys-

1 **10** **20**

C-*chain*

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

1 **10** **20** **30**

Thr-Tyr-Gln-Thr-Lys-Val-Asp-Lys-Asp-Leu-Gln-Ser-Leu-Glu-Asp-Ile-Leu-His-Gln-Val-Glu-Asn-Lys-Thr-Ser-Glu-Val-Lys-Gln-Leu-

40 **50** **60**

Ile-Lys-Ala-Ile-Gln-Leu-Thr-Tyr-Asn-Pro-Asp-Glu-Ser-Ser-(Asp, Lys, Pro)-Met-

70 **78**

Figure 2 N-*terminal sequences of the* A-, B-, *and* C-*chains of human fibrinogen.*[237, 239, 240] ***Arrows indicate the point of cleavage with thrombin in the*** A- *and* B-*chains*

sequence for the *N*-terminal seven residues of the C-chain of bovine fibrinogen has also been reported.[241]

Human congenital dysfibrinogenaemia has occasionally been described. In one particular case, a possible molecular defect in the 'disulphide knot' may be involved.[242] The abnormal fibrinogen had low reactivity towards thrombin and sequence studies showed that the arginine residue at position-19 in the A-chain had been replaced, probably by serine. While thrombin still cleaves at Arg-16, releasing a fibrinopeptide, it is postulated that the Arg-Ser substitution nearby could affect the subsequent polymerisation step, resulting in a severe bleeding disorder. It is most encouraging that it may yet be possible to relate pathological conditions to molecular defects in proteins as large as fibrinogen, comparable to the classic studies on the haemoglobins.

Covalent cross-linking of fibrin is the terminal step in blood clotting. It is catalysed by an enzyme commonly known either as fibrin-stabilising-factor (FSF) or as plasma transglutaminase. It was originally thought that the *N*-terminal glycine residues generated by thrombin cleavage of fibrinogen were involved as amino-group donors in a transpeptidation with glutamine acceptors. Several recent pieces of evidence disprove this. It has been found[243] that fibrinogen in which all the ε-amino-groups are amidinated can be converted into fibrin by thrombin but that the fibrin will not clot under the action of FSF. In all other respects, including glycyl end-groups, the fibrin is normal. The participation of ε- rather than α-amino-groups was therefore suggested.[243] A direct demonstration of this has now been achieved[244] by the isolation of γ-glutamyl-ε-lysine from enzymic digests of cross-linked fibrin. Two to three such cross-links per fibrin monomer of molecular weight 320,000 were reported.[244b] Naturally, other types of cross-link are not eliminated by these experiments. It is worth noting that the mureins of bacterial cell walls also contain lysine residues, the ε-amino-groups of which are involved in peptide linkage.[245]

D. The Histones.—The chromosomal proteins are obviously of interest to the biochemist because of their intimate connection with DNA. Their chemistry and functions have recently been extensively reviewed.[246]

The histones contain several modifications of the lysine residue. In addition to the well-known ε-*N*-methyl-lysine,[247] the presence of

[241] H. Pirkle and A. Henschen, *Biochemistry*, 1968, **7**, 1362.
[242] M. Blombäck, B. Blombäck, E. F. Mammen, and A. S. Prasad, *Nature*, 1968, **218**, 134.
[243] G. M. Fuller and R. F. Doolittle, *Biochem. Biophys. Res. Comm.*, 1966, **25**, 694.
[244a] J. J. Pisano, J. S. Finlayson, and M. J. Peyton, *Science*, 1968, **160**, 892. [b] S. Matačić and A. G. Loewy, *Biochem. Biophys. Res. Comm.*, 1968, **30**, 356.
[245] R. Plapp and O. Kandler, *Z. Naturforsch.*, 1967, **22b**, 1062; K. H. Schliefer, R. Plapp, and O. Kandler, *Biochim. Biophys. Acta*, 1968, **154**, 573.
[246] J. A. V. Butler, E. W. Johns, and D. M. P. Phillips, *Progr. Biophys. Mol. Biol.*, 1968, **18**, 209.
[247] K. Murray, *Biochemistry*, 1964, **3**, 10; R. P. Ambler and M. W. Rees, *Nature*, 1959, **184**, 56.

ε-*N*-dimethyl-lysine has been noted.[248] More recently, ε-*N*-trimethyl-lysine has also been found in histones from calf thymus [249] and chicken erythrocytes:[250] the concentration of methylated lysines is higher in the arginine-rich than the lysine-rich fractions.[250] The occurrence [251] of ε-*N*-acetyl-lysine in calf thymus histone provides yet another possible modification for consideration. Of course, the *N*-terminal residue is also often acetylated, *e.g.* the *N*-terminal sequence *N*-Ac-Ser-Gly-Arg determined for two calf thymus histones.[251] Acetylation has been implicated in the regulation of chromatin structure and nuclear RNA synthesis.[246]

As an aid to purification of the arginine-rich histones, recycling chromatography on Sephadex G-50 has been commended.[252] However, one of the most significant observations of recent histone chemistry is that of their limited heterogeneity. Only *ca.* eight distinct molecular species are recognised in both pea-bud and vertebrate histones.[253] Moreover, the corresponding pea-bud and vertebrate histones closely resemble one another. For example, in a comparative study [254] of the arginine-rich histones III and IV from calf thymus and pea, it has been found that there are few differences in terms of amino-acid composition, finger-prints, and so on, between the corresponding proteins from the two sources. Thus histone III from both calf and pea has *N*-terminal alanine and *C*-terminal alanine, whereas histone IV from both sources has a blocked *N*-terminal (possibly acetylated) residue and *C*-terminal glycine. On the other hand, calf thymus histone III contains two cysteine residues per molecule and forms oligomers by disulphide bridge formation, unlike pea histone III which contains only one cysteine residue and can therefore only form dimers. The biological significance, if any, of this potential disulphide bridge formation is unclear.

This remarkable similarity in structure between corresponding proteins from plant and animal sources has been further defined by amino-acid sequence studies on histone IV. The complete amino-acid sequence of calf thymus histone IV [255] is shown in Figure 3. That of pea histone IV is thought to be very similar; certainly, the *C*-terminal nineteen residues are in an identical sequence.[256] This extent of homology is even more remarkable when it is recalled [256] that in a comparison of the sequences of the cytochromes-*c* from wheat germ and ox, 37 out of 104 residues differ.[1]

[248] K. W. Paik and S. Kim, *Biochem. Biophys. Res. Comm.*, 1967, **27**, 479.
[249] K. Hempel, H. W. Lange, and L. Birkofer, *Naturwiss.*, 1968, **55**, 37.
[250] K. Hempel, H. W. Lange, and L. Birkofer, *Z. physiol. Chem.*, 1968, **349**, 603.
[251] D. M. P. Phillips, *Biochem. J.*, 1968, **107**, 135.
[252] G. Biserte, M. Bonte, P. Sautierre, A. Martinage, Y. Moschetto, and P. Boulanger, *J. Chromatog.*, 1968, **35**, 168.
[253] D. M. Fambrough and J. Bonner, *Biochim. Biophys. Acta*, 1969, **175**, 113.
[254] D. M. Fambrough and J. Bonner, *J. Biol. Chem.*, 1968, **243**, 4434.
[255] R. J. DeLange, D. M. Fambrough, E. L. Smith, and J. Bonner, *J. Biol. Chem.*, 1969, **244**, 319.
[256] R. J. DeLange, D. M. Fambrough, E. L. Smith, and J. Bonner, *J. Biol. Chem.*, 1968, **243**, 5906.

Ac-Ser-Gly-Arg-Gly-Lys-Gly-Gly-Lys-Gly-Leu-Gly-Lys-Gly-Gly-Ala-Lys(Ac)-Arg-His-Arg-Lys(Me)-Val-Leu-Arg-Asp-Asn-Ile-Gln-
1 **10** **20**
Gly-Ile-Thr-Lys-Pro-Ala-Ile-Arg-Arg-Leu-Ala-Arg-Gly-Gly-Val-Lys-Arg-Arg-Ile-Ser-Gly-Leu-Ile-Tyr-Glu-Glu-Thr-Arg-Gly-Val-
30 **40** **50**
Leu-Lys-Val-Phe-Leu-Glu-Asn-Val-Ile-Arg-Asp-Ala-Val-Thr-Tyr-Thr-Glu-His-Ala-Lys-Arg-Lys-Thr-Val-Thr-Ala-Met-Asp-Val-Val-
60 **70** **80**
Tyr-Ala-Leu-Lys-Arg-Gln-Gly-Arg-Thr-Leu-Tyr-Gly-Phe-Gly-Gly·
90 **100**

Figure 3 *Amino-acid sequence of calf thymus histone* IV [255]
(Reproduced by permission from *J. Biol. Chem.*, 1969, **244**, 319)

Several features of the sequence in particular are worth noting. The distribution of basic amino-acids is uneven, with considerably more in the *N*-terminal than the *C*-terminal half of the molecule, which may be significant in the interaction with nucleic acid. The two modified lysine residues (ε-*N*-acetyl and ε-*N*-methyl) occur close together in a cluster of basic amino-acids (residues 16–20), and the particular lysine residue is only *ca.* 50% acetylated. It has been suggested [257] on the basis of recurrent stretches of amino-acid sequence in the analogous clupeines [1] that they are the products of genes that arose by a process of partial duplication from an ancestral gene that coded for the pentapeptide Ala-Arg-Arg-Arg-Arg. There is no firm evidence for this in the sequence of calf histone IV although, as DeLange and co-workers point out,[255] there are several similar regions in the molecule, *e.g.*:

Leu-Gly-Lys-Gly-Gly-Ala-Lys-Arg
10 **17**
Leu-Ala-Arg-Gly-Gly-Val-Lys-Arg
37 **44**

In any event, it is clear that the limited heterogeneity and remarkable evolutionary conservation of histone sequence must be taken into account when considering roles for histones in chromatin structure and gene action. In this connection, it is of interest that it has been reported that chromosomal RNA is required for the specific reconstitution of the original chromatin from the separated DNA and chromosomal proteins.[258] Even if this is so, one is still left with the problem of chromosomal RNA, rather than DNA, having to recognise the appropriate histone. Further developments in this field are obviously awaited with interest.

Another fascinating correlation that needs further detail before the obvious inferences are drawn concerns the coat proteins of the adenoviruses. First reports [259] indicate that they are very similar to the arginine-rich histones (*ca.* 15% arginine) with alanine as the predominant *N*-terminal residue.

4 Peptides and Hormones

A. Growth Hormones.—The almost complete amino-acid sequence for human growth hormone was published some time ago.[260] This sequence has now been corrected and completed [261] and reads as follows in the region concerned:

-Ser-Glu-Ser-Ile-Pro-Thr-Pro-Ser-Asn-Arg-
70 **75** **79**

[257] J. A. Black and G. H. Dixon, *Nature*, 1967, **216**, 152.
[258] I. Bekhor, G. M. Kung, and J. Bonner, *J. Mol. Biol.*, 1969, **39**, 351.
[259] W. C. Russell, W. G. Laver, and P. J. Sanderson, *Nature*, 1968, **219**, 1127.
[260] C. H. Li, W.-K. Liu, and J. S. Dixon, *J. Amer. Chem. Soc.*, 1966, **88**, 2050.
[261] C. H. Li, J. S. Dixon, and D. Chung, *Biochim. Biophys. Acta*, 1968, **160**, 472.

Amino-acid sequences around the cystine residues of horse growth hormone have been studied [262] by the cystine diagonal technique. Comparison of growth hormone from ox, human, and horse reveals close homology at the *C*-terminal end of the molecules:[262]

Horse -Arg-Val-Met-Lys-**Cys-Arg**-Arg-Phe-Val-**Glu**-Ser-**Ser-Cys**-Ala-**Phe**

Ox -Met-Lys-**Cys-Arg**-Arg-Phe-Gly-**Glu**-Ala-**Ser-Cys**-Ala-**Phe**

Human -Arg-Ile -Val-Gln -**Cys-Arg**-------Ser-Val-**Glu**-Gly-**Ser-Cys**-Gly-**Phe**

On the other hand, whereas the second disulphide bridge in human growth hormone links positions 68 and 162,[260] the second disulphide bridge of the horse hormone is in the sequence:[262]

-Leu-Lys-Cys-Arg-Ile-Val-Tyr-Asp-Asn-Cys-Ser-

and there is no obvious similarity with either half-cystine residue of the human protein. It would be interesting to know how this prominent difference in primary structure is reflected in the three-dimensional structure of the hormones. It is doubtless significant that immunological studies suggest that ox and horse growth hormones differ considerably in structure from human growth hormone.[263]

B. Insulin.—See also chapter 5, section 3, and chapter 2, part II, section 15. The existence of a single-chain precursor of the two-chain insulin molecule has been suspected for several years [264] but, until recently, all attempts to isolate it have proved abortive. The purification from a number of sources has now been achieved and the unravelling of its chemistry has provided a fascinating story.[265, 266]

The complete amino-acid sequence of proinsulin from pig has been described [265] and that from ox should be available soon [267] for what ought to be a most interesting comparison. The pig proinsulin sequence is shown in Figure 4. The *N*-terminal B-chain is seen to be separated from the *C*-terminal A-chain by a connecting peptide of 33 residues. Before the secretion of insulin, this peptide must be excised, presumably by enzyme(s). This activation from inactive precursor to biologically active hormone is analogous to zymogen activation of the proteolytic enzymes, and may be imitated *in vitro* by incubation of the proinsulin with trypsin. This results

[262] L. Oliver and A. S. Hartree, *Biochem. J.*, 1968, **109**, 19.
[263] A. S. Hartree, J. B. Mills, R. A. S. Welch, and M. Thomas, *J. Reprod. Fertil.*, 1968, **17**, 291.
[264] D. Givol, F. De Lorenzo, R. F. Goldberger, and C. B. Anfinsen, *Proc. Nat. Acad. Sci. U.S.A.*, 1965, **53**, 676.
[265] R. E. Chance, R. M. Ellis, and W. W. Bromer, *Science*, 1968, **161**, 165.
[266] D. F. Steiner, O. Hallund, A. Rubenstein, S. Cho, and C. Bayliss, *Diabetes*, 1968, **17**, 725.
[267] C. Nolan, E. Margoliash, O. Hallund, and D. F. Steiner, quoted in ref. 266.

in an insulin molecule in which the B-chain terminates at Lys-29 instead of the naturally occurring Ala-30, but which in all other respects is normal. This is a clue to the mechanism *in vivo*, suggesting that a trypsin-like enzyme is probably involved, with some nuance of specificity that prevents cleavage of the lysyl bond at position-29, followed by a carboxypeptidase B action to remove *C*-terminal arginine from the incipient B-chain. A two-chain

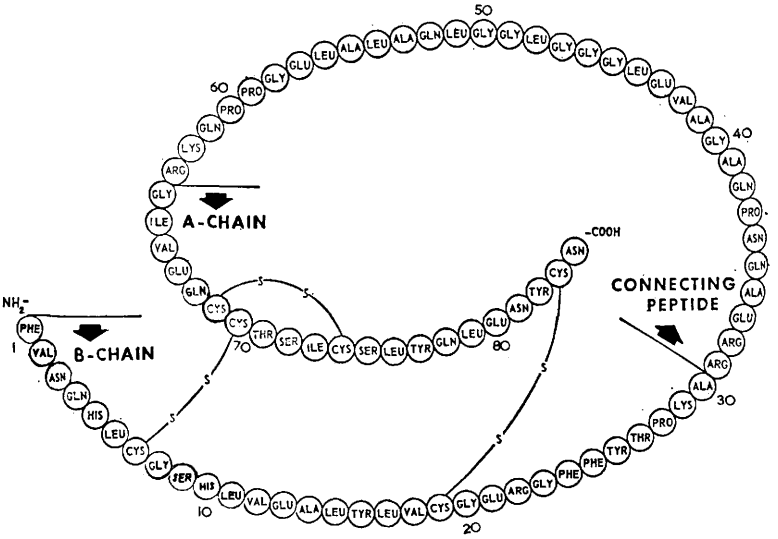


Figure 4 *Amino-acid sequence of pig proinsulin* [265]
(Reproduced by permission from *Science*, 1968, **161**, 165)

protein that resembles insulin has also been reported in preparations of crystalline bovine insulin.[266, 268] Sulphite cleavage of this protein gives an A-chain of about the right size but a B-chain that is some 22 residues too long. As this chain has the proper *N*-terminal phenylalanine residue, it is clear that the additional sequence must be attached to its *C*-terminal end. Until this protein is characterised further, it is difficult to define its exact relation to the single-chain proinsulin. Possibly it is some intermediate in the activation process.

It is worth noting that in cod insulin [269] the B-chain terminates at Lys-29 but the normal length of 30 residues is maintained by the presence of an additional methionine residue in the *N*-terminal position. This type of alteration seems generally true for the fish insulins (references given in ref. 269) and it would be interesting to know something of the chemistry of the fish proinsulins. The sequence -Cys-Cys- which had previously been

[268] C. C. Yip, *Arch. Biochem. Biophys.*, 1968, **127**, 741.
[269] K. B. M. Reid, P. T. Grant, and A. Youngson, *Biochem. J.*, 1968, **110**, 289.

found only in the insulins has now been discovered in the keratins.[270] None the less it remains an unusual disulphide bridge arrangement.

The availability of proinsulin has also enabled a previously embarrassing question to be answered. If the chains of insulin are completely reduced, the yield of native insulin that can be recovered by reoxidation of the disulphide bonds is remarkably small. This is in contrast to the many well-documented examples of proteins containing intrachain disulphide bridges, where the reoxidation is almost quantitative,[196] and, indeed, was a principal reason for the suggestion that insulin derived from a single-chain precursor. It has now been shown [271] that ox and rat proinsulins can be reversibly reduced and oxidised in high yield, which indicates that the complete polypeptide chain is required for this process. The interpretation of all such experiments has been that a polypeptide chain will assume the conformation of lowest free energy. In view of the low recovery when insulin is reduced and reoxidised, it is clear that the disulphide bonds (made originally in the precursor) impose an unfavourable conformation on the molecule. Similar considerations may well apply to other zymogen activation processes.

C. Calcitonin.—The amino-acid sequence of pig thyrocalcitonin (TCT) has been described independently by three groups of workers.[272] α-Thyrocalcitonin has been shown to be converted into the form originally designated β- by oxidation of the single methionine residue to the sulphoxide and the designations α- and β- have therefore been dropped.[272a] This result is reminiscent of the methionine oxidation in adrenocorticotropic hormone, which also led to the isolation of two forms of the peptide.[273] The amino-acid sequence of human calcitonin M has also been described and the sequences of the two hormones compared:[274]

Human M Cys-Gly-**Asn**-**Leu**-**Ser**-**Thr**-Cys-Met-**Leu**-Gly-Thr-**Tyr**-Thr-Gln-Asp-
-Phe-**Asn**-Lys-**Phe**-**His**-Thr-**Phe**-Pro-Gln-Thr-Ala-Ile-**Gly**-Val-Gly-
Ala-**Pro**·NH_2

Pig TCT Cys-Ser-**Asn**-**Leu**-**Ser**-**Thr**-Cys-Val-**Leu**-Ser-Ala-**Tyr**-Trp-Arg-Asn-
-Leu-**Asn**-Asn-**Phe**-**His**-Arg-**Phe**-Ser-Gly-Met-Gly-Phe-**Gly**-Pro-Glu-
Thr-**Pro**·NH_2

The interesting observation has been made that human calcitonin D is the antiparallel dimer of calcitonin M.[274] This seems to be the first recorded instance of natural antiparallel formation of disulphide bridges.

[270] H. Lindley and T. Haylett, *J. Mol. Biol.*, 1967, **30**, 63.
[271] D. F. Steiner and J. L. Clark, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 622.
[272a] R. Neher, B. Riniker, H. Zuber, W. Rittel, and F. W. Kahnt, *Helv. Chim. Acta*, 1968, **51**, 917. [b] J. T. Potts, H. D. Niall, H. T. Keutmann, H. B. Brewer, and L. J. Deftos, *Proc. Nat. Acad. Sci., U.S.A.*, 1968, **59**, 1321. [a] P. H. Bell, W. F. Barg, D. F. Colucci, M. C. Davies, C. Dziobkowski, M. E. Englert, E. Heyder, R. Paul, and E. H. Snedeker, *J. Amer. Chem. Soc.*, 1968, **90**, 2704.
[273] H. B. F. Dixon, 'The Hormones,' ed. G. Pincus, K. V. Thimann, and E. B. Astwood, Academic Press, London, 1964, p. 1.
[274] R. Neher, B. Riniker, W. Rittel, and H. Zuber, *Helv. Chim. Acta.*, 1968, **51**, 1900.

D. Gastrin and Caerulin.—The amino-acid sequences of gastrin from human, pig, ox, sheep, and dog are now known [275] and have been reviewed.[275b]

Human	Glp-Gly-Pro-Trp-*Leu*-Glu-Glu-*Glu*-Glu-*Glu*-Ala-Tyr-Gly-Trp-			
	1			**10**
	Met-Asp-Phe·NH_2			
	17			
Pig		*-Met-*		
Ox, Sheep		*-Val-*		*-Ala-*
Dog		*-Met-*	*-Ala-*	

Tyr-12 can occur naturally as the sulphate, as in the fibrinopeptide B of some species,[1] and both forms of gastrin are biologically active. (It might be as well to note a warning here that the *O*-sulphates of serine and threonine can form in protein hydrolysates during intensive drying.[276])

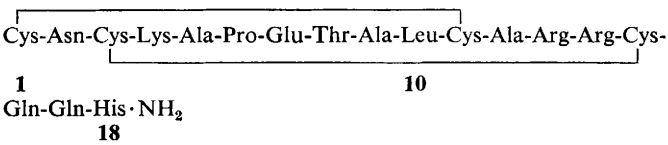
Given that the *C*-terminal tetrapeptide sequence of gastrin is all that is required to observe biological activity,[277] it comes as no surprise that the peptide caerulein from the skin of the Australian amphibian *Hyla caerulea* has a marked effect on gastric secretion, for caerulein has the amino-acid sequence:[278]

$$\begin{array}{c} \quad\quad\quad\quad SO_3H \\ \quad\quad\quad\quad\quad | \\ \text{Glp-Gln-Asp-Tyr-Thr-Gly-Trp-Met-Asp-Phe}\cdot NH_2 \end{array}$$

The tyrosine residue is fully sulphated, unlike that of gastrin. The biological significance of this remarkable homology remains obscure, although caerulein additionally has a potent and relatively long-lasting hypotensive action.

The presence of an amidated residue at the *C*-terminus of these peptides is worthy of note. This type of *C*-terminal group has been noted before in biologically active peptides, *e.g.* melanocyte stimulating hormone (α-MSH)[1] and, of course, inhibits hydrolysis by carboxypeptidase.

E. Venom Peptides.—Apamin is an 18-residue peptide from bee venom which exhibits excitatory properties on the central nervous system. Its complete amino-acid sequence is now established:[279]



Again a *C*-terminal amidated residue is found.

[275a] K. L. Agarwal, J. Beacham, P. H. Bentley, R. A. Gregory, G. W. Kenner, R. C. Sheppard, and H. J. Tracy, *Nature*, 1968, **219**, 614. [b] G. W. Kenner and R. C. Sheppard, *Proc. Roy. Soc.*, 1968, *B*, **170**, 89.
[276] K. Murray, *Biochem. J.*, 1968, **110**, 155.
[277] J. S. Morley, *Proc. Roy. Soc.*, 1968, *B*, **170**, 97.
[278] A. Anastasi, V. Erspamer, and R. Endean, *Arch. Biochem. Biophys.*, 1968, **125**, 57.
[279] G. L. Callewaert, R. Shipolini, and C. A. Vernon, *F.E.B.S. Letters*, 1968, **1**, 111.

The direct lytic factor (DLF) of cobra venom capable of haemolysing red blood cells has been characterised as a pure protein with a molecular weight of 7000.[280] DLF has 57 residues, *N*-terminal leucine, *C*-terminal serine, and is highly basic. The sequence remains to be determined.

5 Enzymes

A majority of work on the primary structure of proteins is devoted to enzymes, particularly if one includes the haem proteins in this category. Many studies are aimed at establishing the complete amino-acid sequence of the protein; other are limited to defining the chemistry of the active site and the quaternary structure of the enzyme molecule. The primary structure of several enzymes can now be given in some detail and many others are under intensive investigation.

A. Proteolytic Enzymes.—*Serine Proteinases.* See also chapter 2, part III, section 2B. The structure of the proteolytic enzymes was recently reviewed in detail.[281] The study of the serine proteinases, in particular, has moved on considerably since then. The disulphide bridge arrangement in trypsinogen, originally determined by the cystine diagonal technique,[282] has been confirmed by more conventional methods.[283] It has been known for some time [281, 284] that the sequences of bovine trypsin and chymotrypsin are highly homologous, leading one to suspect that the tertiary structures are also very similar. The correctness of this supposition has been elegantly demonstrated by showing [285] that it is possible to fit the two disulphide bridges of trypsin which are not homologous with chymotrypsin into the three-dimensional chymotrypsin model without distortion.

On the basis of comparable amino-acid sequence evidence from bovine chymotrypsinogen B and porcine pancreatic elastase, it is clear that similar conclusions apply here also. The complete amino-acid sequence of chymotrypsinogen B has now been established [286] and compared with that of chymotrypsinogen A and trypsinogen (Figure 5). Chymotrypsinogen A and B are more closely homologous one with another than is either with trypsinogen, suggesting that the evolutionary divergence of chymotrypsin A and B came after that of trypsin and chymotrypsin. Complementary studies of porcine elastase by the cystine diagonal technique [287] revealed that the four disulphide bridges are closely homologous with the four common

[280] S. Aloof-Hirsch, A. de Vries, and A. Berger, *Biochim. Biophys. Acta*, 1968, **154**, 53.
[281] H. Neurath, K. A. Walsh, and W. P. Winter, *Science*, 1967, **158**, 1638.
[282] D. L. Kauffman, *J. Mol. Biol.*, 1965, **12**, 929.
[283] V. Holeysovsky, B. Mesrob, V. Tomasek, O. Mikes, and F. Šorm, *Coll. Czech. Chem. Comm.*, 1968, **33**, 441.
[284] B. S. Hartley, J. R. Brown, D. L. Kauffman, and L. B. Smillie, *Nature*, 1965, **207**, 1157.
[285] P. B. Sigler, D. M. Blow, B. W. Matthews, and R. Henderson, *J. Mol. Biol.*, 1968, **35**, 143.
[286] L. B. Smillie, A. Furka, N. Nagabhushan, K. J. Stevenson, and C. O. Parkes, *Nature*, 1968, **218**, 343.
[287] J. R. Brown, D. L. Kauffman, and B. S. Hartley, *Biochem. J.*, 1967, **103**, 497.

```
A  Cys-Gly-Val-Pro-Ala-Ile-Gln-Pro-Val-Leu-Ser -Gly-Leu-Ser -Arg-Ile-Val-Asn-Gly-Glu-
B  Cys-Gly-Val-Pro-Ala-Ile-Gln-Pro-Val-Leu-Ser -Gly-Leu-Ala -Arg-Ile-Val-Asn-Gly-Glu-
T                                              Val -Asp-Asp-Asp-Asp-Lys -Ile-Val-Gly -Gly-Tyr-

   1                          10                                                  20

His-Phe-Cys-Gly-Gly-Ser-Leu-Ile-Asn-Glu-Asn-Trp-Val-Val-Thr-Ala-Ala-His-Cys-Gly-
His-Phe-Cys-Gly-Gly-Ser-Leu-Ile-Ser -Glu-Asp-Trp-Val-Val-Thr-Ala-Ala-His-Cys-Gly-
His-Phe-Cys-Gly-Gly-Ser-Leu-Ile-Asn-Ser -Gln-Trp-Val-Val-Ser -Ala-Ala-His-Cys-Tyr-
          III                                                                   IV
  40                                  50

Lys-----Ile -Gln-Lys-Leu-Lys-Ile -Ala-Lys-Val-Phe-Lys-Asn-Ser -Lys-Tyr-Asn-Ser-
Asp-----Thr-Gln-Val-Leu-Lys-Ile -Gly-Lys-Val-Phe-Lys-Asn-Pro-Lys-Phe-Ser -Ile-
Gly-Asn-Gln-Gln-Phe-Ile -Ser -Ala-Ser -Lys-Ser -Ile -Val-His -Pro-Ser -Tyr-Asn-Ser-
          80                                  90

Gln-Thr-Val-Ser -Ala-Val-Cys-Leu-Pro-Ser -Ala-Ser -Asp-Asp-Phe-Ala-Ala-Gly-Thr-
Glu-Thr-Val-Ser -Ala-Val-Cys-Leu-Pro-Ser -Ala-Asp-Glu-Asp-Phe-Pro-Ala-Gly-Met-
Ser -Arg-Val-Ala-Ser -Ile -Ser -Leu-Pro-Thr-----Ser -Cys-Ala -Ser -----Ala-Gly-Thr-
                          V                          VI
        120                                 130

Arg-Leu-Gln-Gln-Ala-Ser -Leu-Pro-Leu-Leu-Ser-Asn-Thr-Asn-Cys-Lys-Lys-Tyr-Trp-
Lys-Leu-Gln-Gln-Ala-Thr-Leu-Pro-Ile -Val-Ser-Asn-Thr-Asp-Cys-Arg-Lys-Tyr-Trp-
Val-Leu-Lys-Cys-Leu-Lys-Ala-Pro-Ile -Leu-Ser-Asn-Ser -Ser -Cys-Lys-Ser -Ala-Tyr-
        VIII                                                  IX
                  160                                 170

Ser-Cys-Met-Gly-Asp-Ser-Gly-Gly-Pro-Leu-Val-Cys-Lys-Lys-Asn-Gly-Ala-Trp-Thr-
Ser-Cys-Met-Gly-Asp-Ser-Gly-Gly-Pro-Leu-Val-Cys-Gln-Lys-Asn-Gly-Ala-Trp-Thr-
Ser-Cys-Gln-Gly-Asp-Ser-Gly-Gly-Pro-Val-Val-Cys-Ser -Gly-Lys----------------
   XI                          XII
                        200                                 210

Tyr-Ala-Arg-Val-Thr-Ala-Leu-Val -Asn-Trp-Val-Gln-Gln-Thr-Leu-Ala-Ala-Asn
Tyr-Ala-Arg-Val-Thr-Ala-Leu-Met-Pro-Trp-Val-Gln-Glu-Thr-Leu-Ala-Ala-Asn
Tyr-Thr-Lys-Val-Cys-Asn-Tyr-Val -Ser -Trp-Ile -Lys-Gln-Thr-Ile -Ala-Ser-Asn
              XIV
                        240                                 249
```

Figure 5 *Amino-acid sequences of bovine chymotrypsinogen* A, *chymotrypsinogen* B, (Reproduced by permission from *Nature*, 1968, **218**, 343)

Disulphide bridges in chymotrypsinogens A and B are at identical positions as follows: I–V XI–XIII (residues 194–223). Disulphide bridges in trypsinogen are II–VIII (residues 22–158);

```
Glu-Ala-Val-Pro-Gly-Ser-Trp-Pro-Trp-Gln-Val-Ser-Leu-Gln-Asp-Lys-Thr-Gly-Phe-   A
Asp-Ala-Val-Pro-Gly-Ser-Trp-Pro-Trp-Gln-Val-Ser-Leu-Gln-Asp-Ser-Thr-Gly-Phe-   B
Thr-Cys-Gly-Ala-Asn-Thr-Val-Pro-Tyr-Gln-Val-Ser-Leu-Asn--------Ser-Gly-Tyr-   T
   II
                                    30

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

60                                  70

Leu-Thr-Ile-Asn-Asn-Asp-Ile-Thr-Leu-Leu-Lys-Leu-Ser-Thr-Ala-Ala-Ser-Phe-Ser-
Leu-Thr-Val-Arg-Asn-Asp-Ile-Thr-Leu-Leu-Lys-Leu-Ala-Thr-Pro-Ala-Gln-Phe-Ser-
Asn-Thr-Leu-Asn-Asn-Asp-Ile-Met-Leu-Ile-Lys-Leu-Lys-Ser-Ala-Ala-Ser-Leu-Asn-

100                                 110

Thr-Cys-Val-Thr-Thr-Gly-Trp-Gly-Leu-Thr-Arg-Tyr-Thr-Asn-Ala-Asn-Thr-Pro-Asp-
Leu-Cys-Ala-Thr-Thr-Gly-Trp-Gly-Lys-Thr-Lys-Tyr-Asn-Ala-Leu-Lys-Thr-Pro-Asp-
Gln-Cys-Leu-Ile-Ser-Gly-Trp-Gly-Asn-Thr-Lys-Ser-Ser-Gly-Thr-Ser-Tyr-Pro-Asp-
   VII
                140                                 150

Gly-Thr-Lys-Ile-Lys-Asp-Ala-Met-Ile-Cys-Ala-Gly-Ala-----Ser-Gly-Val-----Ser-
Gly-Ser-Arg-Val-Thr-Asp-Val-Met-Ile-Cys-Ala-Gly-Ala-----Ser-Gly-Val-----Ser-
Pro-Gly-Gln-Ile-Thr-Ser-Asn-Met-Phe-Cys-Ala-Gly-Tyr-Leu-Glu-Gly-Gly-Lys-Asn-
                                    X
                180                                 190

Leu-Val-Gly-Ile-Val-Ser-Trp-Gly-Ser-Ser-Thr-Cys-Ser-Thr-Ser-Thr-----Pro-Gly-Val-
Leu-Ala-Gly-Ile-Val-Ser-Trp-Gly-Ser-Ser-Thr-Cys-Ser-Thr-Ser-Thr-----Pro-Ala-Val-
Leu-Gln-Gly-Ile-Val-Ser-Trp-Gly-Ser-----Gly-Cys-Ala-Gln-Lys-Asn-Lys-Pro-Gly-Val-
                                   XIII
                220                                 230
```

and trypsinogen (T) [286]

(residues 1–123); III–IV (residues 42–58); VII–XII (residues 137–204); IX–X (residues 169–183); III–IV; VI–XIV (residues 129–236); VII–XII; IX–X; XI–XIII.

to bovine chymotrypsin and trypsin, and further work on the primary structure has extended these observations. The overall homology with chymotrypsin A is 42%[288] and the tertiary structure is closely similar.[289] These correlations of primary and tertiary structure are neatly rationalised by recognising that the pattern of amino-acid replacement within the 'interior' of protein molecules must be very conservative. Internal positions are occupied almost without exception by non-polar residues[290] and replacements must be similarly non-polar, whereas 'surface' changes suffer considerably less restriction. The availability of the high-resolution three-dimensional structure of α-chymotrypsin[285] enables these considerations amply to be confirmed for the pancreatic serine proteinases.[291]

A good example of the fruitful collaboration between the crystallographer and the protein chemist has come from detailed examination of the three-dimensional structure of α-chymotrypsin. The *X*-ray data show that residue 103 (Figure 5) is close to the active site His-57, the essential histidine residue. Reinvestigation of the amino-acid sequence in this region[292] indicates that this residue should be aspartic acid, not asparagine. This in turn means that there is now a charged residue in an hydrophobic environment close to the active site His-57 and it may alleviate its position by hydrogen-bonding through His-57 to Ser-198, thereby explaining the peculiar nucleophilic reactivity of that serine residue.[292] Since the buried ion-pair Ile-16 : Asp-197 is also thought to play a major part in the formation of active enzyme,[285] the importance of occasional buried charged groups is clearly demonstrated. Bovine chymotrypsinogen A and B, trypsinogen, and porcine elastase are all highly homologous in these critical regions and doubtless have similar crucial interactions.

Study of the trypsinogens from other species reveals the expected close similarities. Sheep trypsinogen has been prepared in a pure state and turns out to be very similar to the bovine enzyme.[293] Of particular note is the presence of two *N*-terminal residues, Val and Phe, and the existence of two activation peptides, Val-Asp-Asp-Asp-Asp-Lys and Phe-Pro-Val-Asp-Asp-Asp-Asp-Lys, the first of which is identical to that of the bovine zymogen. The second peptide is closely related to the single activation peptide of the pig enzyme, Phe-Pro-Thr-Asp-Asp-Asp-Asp-Lys.[1] Presumably the sheep trypsinogens are the isoenzymic products of separate genes. Turkey trypsin, a non-mammalian trypsin, is also clearly highly homologous with the ox and pig enzymes, as revealed by studies of the histidine sequences:[294]

[288] D. M. Shotton and B. S. Hartley, quoted by B. S. Hartley, *Biochem. J.*, 1969, **110**, 1P.
[289] H. C. Watson and D. M. Shotton, quoted by L. N. Johnson, *F.E.B.S. Letters*, 1969, **2**, 201.
[290] M. F. Perutz, J. C. Kendrew, and H. C. Watson, *J. Mol. Biol.*, 1965, **13**, 669.
[291] B. S. Hartley, *Biochem. J.*, 1968, **110**, 1P.
[292] D. M. Blow, J. J. Birktoft, and B. S. Hartley, *Nature*, 1969, **221**, 337.
[293] R. Schyns, S. Bricteux-Grégoire, and M. Florkin, *Biochim. Biophys. Acta*, 1969, **175**, 97.
[294] T. Kishida and I. E. Liener, *Arch. Biochem. Biophys.*, 1968, **126**, 111.

Turkey
Asx-Ser-Gly-Tyr-**His-Phe-Cys-Gly-Glx-Ser-Leu** **Ala-Ala-His-Cys-Tyr-Lys**
Ox
Asn-**Ser-Gly-Tyr-His-Phe-Cys-Gly-Gly-Ser-Leu** **Ala-Ala-His-Cys-Tyr-Lys**
Pig
Asn-**Ser-Gly-Ser-His-Phe-Cys-Gly-Gly-Ser-Leu** **Ala-Ala-His-Cys-Tyr-Lys**

Turkey
Ala-Leu-Thr-**His-Pro**-Asx-Tyr
Ox
Ser-Leu-Val-**His-Pro**-Ser-Tyr
Pig
Ile-Ile-Thr-**His-Pro**-Asn-Phe

Obviously it is of great interest to compare the pancreatic proteinases with other serine proteinases, all of which can be inhibited with di-isopropyl phosphorofluoridate. Thrombin, which is synthesised as prothrombin in the liver, is a clear candidate for such a comparison, having a substrate specificity similar to, but narrower than, that of trypsin. The A-chain of thrombin, which is bound to the B-chain by a disulphide bridge, has the following amino-acid sequence:[295]

Thr-Ser-Glu-Asn-His-Phe-Glu-Pro-Phe-Phe-Asx-Glx-Lys-Thr-Phe-Gly-Ala-
1 **10**
Gly-Glu-Ala-Asp-Cys-Gly-Leu-Arg-Pro-Leu-Phe-Glu-Lys-Lys-Glu-Val-Glx-
20 **30**
Asx-Glx-Thr-Gln-Lys-Glu-Leu-Phe-Glu-Ser-Tyr-Ile-Glu-Gly-Arg
40 **49**

While this sequence shows no apparent homology with the pancreatic proteinases, the B-chain is reported [295] to show extensive homology with those parts of the chymotrypsin sequence that can best be described as 'internal' in the chymotrypsin model.[285] Given additionally that the *N*-terminal sequence of the B-chain of bovine thrombin is Ile-Val-Glu-Gly-, of human thrombin is Ile-Val-Gly-Gly-, and that the *N*-terminal isoleucine residue is critical for activity,[295] one begins to suspect that thrombin and the pancreatic proteinses have much in common, including perhaps elements of tertiary structure.

Although it is clear that the mammalian serine proteinases have derived from a common ancestral gene, it is highly probable that the subtilisins—bacterial serine proteinases closely related in mechanism to the mammalian enzymes—have a distinct genetic origin. The complete amino-acid sequences of subtilisin BPN′ (from *B. amyloliquefaciens*) [296] and subtilisin Carlsberg (from *B. subtilis*) [297] have been established (Figure 6). On the basis of amino-acid compositions, finger-prints, *etc.*, it seems likely that a third subtilisin, Novo, is identical to BPN′.[298] Comparison of the Carlsberg and

[295] S. Magnusson, *Biochem. J.*, 1968, **110**, 25P.
[296] F. S. Markland and E. L. Smith, *J. Biol. Chem.*, 1967, **242**, 5198.
[297] E. L. Smith, R. J. DeLange, W. H. Evans, M. Landon, and F. S. Markland, *J. Biol. Chem.*, 1968, **243**, 2184.
[298] S. A. Olaitan, R. J. DeLange, and E. L. Smith, *J. Biol. Chem.*, 1968, **243**, 5296.

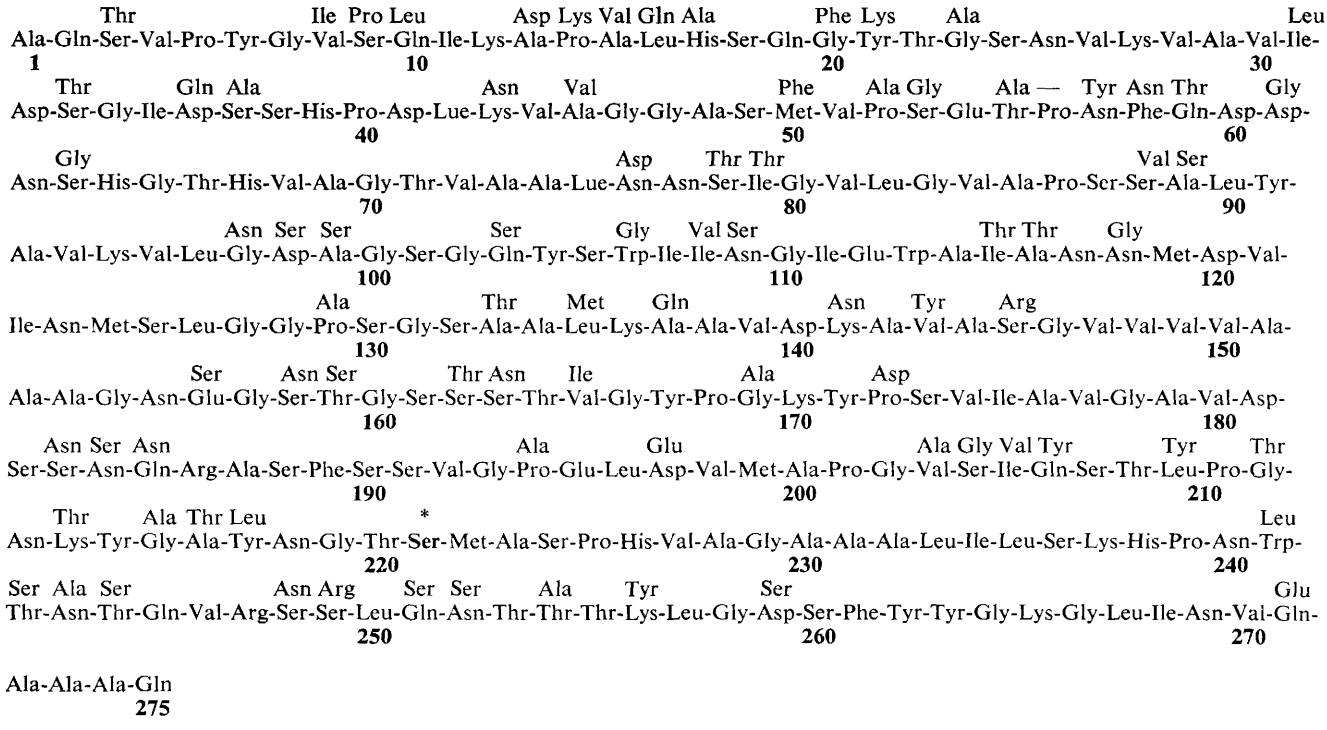


Figure 6 *Amino-acid sequences of subtilisin* BPN′ [296] *and subtilisin Carlsberg.*[297] *The continuous sequence is that of subtilisin* BPN′, *the differences in subtilisin Carlsberg being indicated above the line. The active site* Ser-221 *is shown in bold type*
(Reproduced by permission from *J. Biol. Chem.*, 1968, **243**, 2184)

BPN′ enzymes shows that there are 84 differences and one additional residue (position 56) in BPN′. Of these differences 61 are single-base changes and 23 are double-base changes in the *E. coli* code,[299] clear evidence of descent from a common precursor by a series of point mutations. A number of repetitions of similar sequence occur in different parts of the polypeptide chain, *e.g.*:

-His-Val-Ala-Gly-Thr-Val-Ala-Ala-Leu-
67 **75**
-His-Val-Ala-Gly- - - - -Ala-Ala-Ala-Leu-
226 **233**

It is possible that these arose by some form of partial gene duplication of the type discussed by Dixon.[300] However, this must have been a complex process since the repetitions do not occupy the same order in the *N*- and *C*-terminal halves of the molecule but are *inverted*, *i.e.* residues 39–42, 67–75, and 85–94 are repeated at positions 238–241, 226–233, and 126–136.

No sequence homology between the subtilisins and the mammalian serine proteinases can be detected [297] and this is in accord with there being no obviously shared three-dimensional structure.[301] Thus despite the fact that these enzymes work in a similar way, with an identical mechanism for activating the functional serine residue,[292, 301] it is highly probable that they represent true convergent evolution at the molecular level. The powerful forces operating on the selection of 'internal' residues in protein molecules are well exemplified by the fact that of the 84 differences between subtilisin BPN′ and Carlsberg, 83 are in 'exterior' positions.[301] Further, the deletion of Pro-56 in Carlsberg is in an outside loop; all of which may be taken as proof that the individual subtilisins have an almost identical three-dimensional structure.

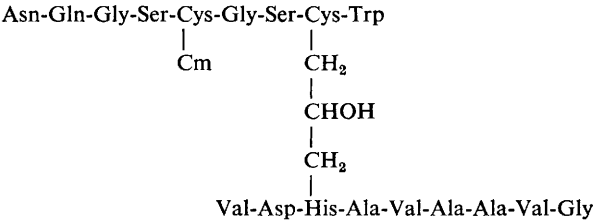
Other Proteinases. Work with other proteinases is not yet as advanced as that on the serine enzymes, although the rate of progress is rapid. Comparison of the amino-acid sequence around the functional cysteine residue of a group of plant proteolytic enzymes reveals the homology:

Stem bromelain [302]	-Cys-**Gly**-Ala-**Cys**-Trp-
Papain [303]	-Pro-Val-Lys-Asn-Gln-Gly-Ser-Cys-**Gly**-Ser-**Cys**-Trp-
Ficin [304]	-Pro-Ile-Arg-Gln-Gln-Gly-Gln-Cys-**Gly**-Ser-**Cys**-
Chymopapain [305]	-Lys-Arg-Val-Pro-Asp-Ser-**Gly**-Glu-**Cys**-Tyr-
Streptococcal protease [306]	-Ser-Phe-Val-Gly-Gln-Ala-Ala-Thr-**Gly**-His-**Cys**-Val-

The glycoprotein nature of stem bromalein has been reported.[307]

[299] F. H. C. Crick, *Cold Spring Harbour Symp. Quant. Biol.*, 1966, **31**, 1.
[300] G. H. Dixon, 'Essays in Biochemistry', ed. P. N. Campbell and G. D. Greville, Academic Press, London, 1966, **2**, 147.
[301] C. S. Wright, R. A. Alden, and J. Kraut, *Nature*, 1969, **221**, 235.
[302] L. P. Chao and I. E. Liener, *Biochem. Biophys. Res. Comm.*, 1967, **27**, 100.
[303] A. Light, R. Frater, J. Kimmel, and E. L. Smith, *Proc. Nat. Acad. Sci. U.S.A.*, 1964, **52**, 1276.
[304] R. C. Wong and I. E. Liener, *Biochem. Biophys. Res. Comm.*, 1964, **17**, 470.
[305] J. N. Tsunoda and K. T. Yasunobu, *J. Biol. Chem.*, 1966, **241**, 4610.
[306] T-Y. Liu, W. H. Stein, S. Moore, and S. D. Elliott, *J. Biol. Chem.*, 1965, **240**, 1143.
[307] T. Murachi, A. Suzuki, and N. Takahashi, *Biochemistry*, 1967, **6**, 3730.

The presence of histidine at the active site of papain has been nicely shown [308] by the isolation of a peptide cross-linking the functional cysteine and histidine residues. This can be achieved by allowing the protein to react with 1,3-dibromoacetone, reducing the keto-function with sodium borohydride, carboxymethylating, and isolating the cross-linked peptide from proteolytic digests:

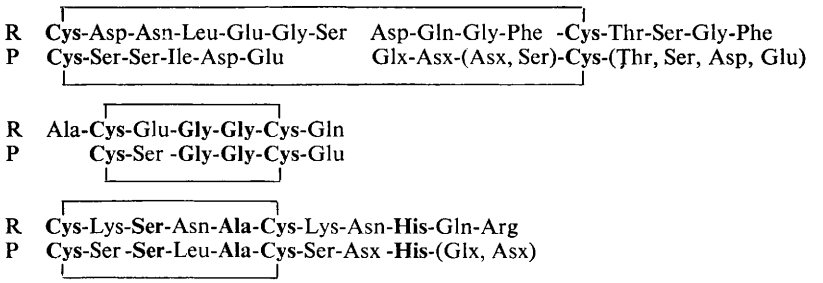


The histidine is linked through the 1-position of the imidazole ring. This association of cysteine and histidine at the active site of papain shows up well in the high-resolution *X*-ray structure of the enzyme.[309] Protein chemists, however, have been put on their mettle by the fact that the chemically derived amino-acid sequence is markedly at variance with the *X*-ray structure. Although all residues are not identifiable at 2·8 Å resolution, it seems clear that there are 211 residues in the single polypeptide chain and a complete unequivocal sequence is now awaited with much interest.

Studies on the primary structure of pig pepsin and calf rennin suggest that these proteins, too, form part of another family of proteinases. Comparison of part of the *C*-terminal sequence of pig pepsin established by the lysine diagonal technique [182] with that of the *C*-terminal sequence of calf rennin [310] indicates that they derive from a common ancestral gene:[182]

Pig pepsin -**Ala-Asn-Asn**-Lys-**Val-Gly-Leu-Ala**-Pro-Val-Ala
Calf rennin -**Ala-Asn-Asn**-Leu-**Val-Gly-Leu-Ala**-Lys-Ala-Ile

Similar conclusions are drawn from an examination of the amino-acid sequences around the three disulphide bridges of pepsin (P) and rennin (R):[311]



[308] S. S. Husain and G. Lowe, *Chem. Comm.*, 1968, 310.
[309] J. Drenth, J. N. Jansonius, K. Koekoek, H. M. Swen, and B. G. Wolthers, *Nature*, 1968, **218**, 929.
[310] B. Foltmann, *Compt. Rend. Trav. Lab. Carlsberg*, 1966, **35**, 143.
[311] B. Foltmann and B. S. Hartley, *Biochem. J.*, 1967, **104**, 1064.

Detailed studies on the complete amino-acid sequence of pepsinogen continue to progress.[312, 313] Based on the amino-acid sequence of tryptic peptides and activation peptides released during zymogen activation, an *N*-terminal sequence of 41 residues can be formulated for pepsinogen:[312]

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

Assuming that 41 residues are split off from pepsinogen during its activation to pepsin,[314] cleavage of a Glu-Ile bond at position 41 probably represents the decisive step in the activation process.[312] The *N*-terminal sequence of pepsinogen is most unusual in the clustering of thirteen basic amino-acids in this region. Three other basic amino-acids are also known to cluster towards the *C*-terminus of the molecule,[182, 315] leaving an intermediate region of some 300 residues devoid of basic amino-acids with the exception of a single histidine residue. This curious absence of basic amino-acids in the *N*-terminal 300 residues of pepsin may well contribute to the unusual acid stability and alkaline lability of the enzyme.[312, 315]

Although pancreatic phospholipase A is not to be confused with the proteolytic enzymes already discussed, it is of interest that the enzyme has been purified from pig pancreas and found to undergo analogous zymogen activation by trypsin, with the release of an activation peptide, Glp-Gly-Glu-Ile-Ser-Arg.[316] Partial sequences for many of the tryptic peptides of carboxypeptidase B have also been reported.[317] In view of the current interest in the high-resolution *X*-ray structure of carboxypeptidase A,[318] a comparison of the primary structures of these two exopeptidases will be of much value.

B. Other Enzymes.—See also chapter 2, part III, section 2B. During the past 18 months details of the complete amino-acid sequence of several enzymes other than proteinases have appeared. The evidence on which the complete sequence [1] of the A protein (α-subunit) of tryptophan synthetase is based has been given [319] and a correction has been made [320] to the

[312] P. V. Koehn and G. E. Perlmann, *J. Biol. Chem.*, 1968, **243**, 6099; E. B. Ong and G. E. Perlmann, *ibid.*, 1968, **243**, 6104.
[313] V. I. Ostoslavskaya, I. B. Pugacheva, E. A. Vakhitova, V. F. Krytsov, G. L. Muratova, E. D. Levin, and V. M. Stepanov, *Biokhimiya*, 1968, **33**, 331.
[314] T. G. Rajagopalan, S. Moore, and W. H. Stein, *J. Biol. Chem.*, 1966, **241**, 4940.
[315] T. A. A. Dopheide, S. Moore, and W. H. Stein, *J. Biol. Chem.*, 1967, **242**, 1833.
[316] G. H. de Haas, N. M. Postema, W. Nieuwenhuizen, and L. L. M. van Deenen, *Biochim. Biophys. Acta*, 1968, **159**, 118.
[317] M. Elzinga and C. H. W. Hirs, *Arch. Biochem. Biophys.*, 1968, **123**, 361.
[318] 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.
[319] J. R. Guest, G. R. Drapeau, B. C. Carlton, and C. Yanofsky, *J. Biol. Chem.*, 1967, **242**, 5442.
[320] H. Taniuchi, C. L. Cusumano, C. B. Anfinsen, and J. L. Cone, *J. Biol. Chem.*, 1968, **243**, 4775.

Ser-Asp-Lys-Ile-Ile-His-Leu-Thr-Asp-Asp-Ser-Phe-Asp-Thr-Asp-Leu-Val-Lys-Ala-Asp-Gly-Ala-Ile-Leu-Val-Asp-Phe-Trp-Ala-Glu-
1 10 20 30

Trp-Cys-Gly-Pro-Cys-Lys-Met-Ile-Ala-Pro-Ile-Leu-Asp-Glu-Ile-Ala-Asp-Glu-Tyr-Gln-Gly-Lys-Leu-Thr-Val-Ala-Lys-Leu-Asn-Ile-
40 50 60

Asp-Gln-Asn-Pro-Gly-Thr-Ala-Pro-Lys-Tyr-Ile-Gly-Arg-Gly-Ile-Pro-Thr-Leu-Leu-Leu-Phe-Lys-Asn-Gly-Glu-Val-Ala-Ala-Thr-Lys-
70 80 90

Val-Gly-Ala-Leu-Ser-Lys-Gly-Gln-Leu-Lys-Glu-Phe-Leu-Asp-Ala-Asn-Leu-Ala
100 108

Figure 7 *Amino-acid sequence of thioredoxin from* E. coli [323]

Ser-Thr-Ile-Glu-Glu-Arg-Val-Lys-Lys-Ile-Ile-Gly-Glu-Gln-Leu-Gly-Val-Lys-Gln-Glu-Glu-Val-Thr-Asp-Asn-Ala-Ser-Phe-Val-Glu-
1 10 20 30

Asp-Leu-Gly-Ala-Asp-Ser-Leu-Asp-Thr-Val-Glu-Leu-Val-Met-Ala-Leu-Glu-Glu-Glu-Phe-Asp-Thr-Glu-Ile-Pro-Asp-Glu-Glu-Ala-
* 40 50

Glu-Lys-Ile-Thr-Thr-Val-Gln-Ala-Ala-Ile-Asp-Tyr-Ile-Asn-Gly-His-Gln-Ala
60 70 77

Figure 8 *Amino-acid sequence of acyl carrier protein from* E. coli [324]

Met-Thr-His-Asn-Asp-Lys-Leu-Gln-Val-Ala-Glu-Ile-Lys-Arg-Gly-Thr-Val-Ile-Asn-His-Ile-Pro-Ala-Glu-Ile-Gly-Phe-Lys-Leu-Leu-
10 20 30

Ser-Leu-Phe-Lys-Leu-Thr-Glu-Thr-Gln-Asp-Arg-Ile-Thr-Ile-Gly-Leu-Asn-Leu-Pro-Ser-Gly-Glu-Met-Gly-Arg-Lys-Asp-Leu-Ile-Lys-
40 50 60

Ile-Glu-Asn-Thr-Phe-Leu-Ser-Glu-Asx-Glx-Val-Asx-Glx-Leu-Ala-Leu-Tyr-Ala-Pro-Gln-Ala-Thr-Val-Asn-Arg-Ile-Asn-Asp-Tyr-Glu-
70 80 90

Val-Val-Gly-Lys-Ser-Arg-Pro-Ser-Leu-Pro-Glu-Arg-Asn-Ile-Asp-Val-Leu-Val-Cys-Pro-Asp-Ser-Asn-Cys-Ile-Ser-His-Ala-Glu-Pro-
100 110 120

Val-Ser-Ser-Ser-Phe-Ala-Val-Arg-Arg-Ala-Asx-Asx-Ile-Ala-Leu-Lys-Cys-Lys-Tyr-Cys-Glu-Lys-Glu-Phe-Ser-His-Asn-Val-Val-Leu-
130 140 150

Ala-Asn

Figure 9 *Amino-acid sequence of regulatory subunit of aspartate transcarbamylase of* E. coli [328]

sequence [321] of the nuclease V8 from *Staph. aureus*. Residues 30–31 and 126–137 should now read Gln-Pro and Arg-Lys-Ser-Glu-Ala-Lys-Lys-Glu-Lys-Leu respectively. Partial analysis of nuclease Foggi reveals [322] that its sequence is almost identical to that of nuclease V8 with histidine replacing leucine at position 124. In a series of papers [323] the complete amino-acid sequence of thioredoxin from *E. coli* has been deduced (Figure 7). The protein functions as hydrogen carrier in the enzymic reduction of ribonucleotides through the reversible oxidation–reduction of its single disulphide bridge. Another *E. coli* protein for which the complete amino-acid sequence is now known [324] is the acyl carrier protein of fatty acid synthesis [325] (Figure 8). The pantotheine prosthetic group is attached in phosphodiester linkage to the hydroxyl group of Ser-36.

It is encouraging to learn that the long wrangle over the number of regulatory and catalytic subunits in aspartate transcarbamylase from *E. coli* may now be over. Recent reports of end-group and finger-printing analysis [326] had indicated that there were four of each type in a molecule of molecular weight 300,000. Even over the interpretation of this apparently straightforward fact there was some dispute.[327] However, determination of the amino-acid sequence of the regulatory subunit [328] (Figure 9) shows that its true molecular weight is 17,000 and, accepting a molecular weight of 33,000 for the catalytic subunit, there must clearly be six of each type of chain in the enzyme molecule;[328] a result entirely in agreement with preliminary crystallographic studies.[329] The *N*-terminal methionine residue of the regulatory subunit is not present in all chains,[328] in accord with the role of that amino-acid in chain initiation, and this may explain why previous investigations had indicated threonine as the principal *N*-terminal residue in too low a yield.

The complete amino-acid sequences of glyceraldehyde 3-phosphate dehydrogenase (GPDH) from lobster [184] and pig [185] muscle are now available (Figure 10), proving conclusively that the enzyme molecule contains four identical peptide chains. The homology is remarkable indeed, 72% of the residues occurring in identical positions. Further, the mammalian enzymes are thought to be almost identical in amino-acid sequence [330] and even the yeast enzyme is very closely related.[331] The amino-acid

[321] H. Taniuchi, C. B. Anfinsen, and A. Sodja, *J. Biol. Chem.*, 1967, **242**, 4752.
[322] C. L. Cusumano, H. Taniuchi, and C. B. Anfinsen, *J. Biol. Chem.*, 1968, **243**, 4769.
[323] A. Holmgren, R. N. Perham, and A. Baldesten, *European J. Biochem.*, 1968, **5**, 352; A. Holmgren, *ibid.*, 1968, **6**, 467, 475.
[324] T. C. Vanaman, S. J. Wakil, and R. L. Hill, *J. Biol. Chem.*, 1968, **243**, 6420.
[325] P. W. Majerus and P. R. Vagelos, *Adv. Lipid Res.*, 1967, **5**, 2.
[326] G. L. Hervé and G. R. Stark, *Biochemistry*, 1967, **6**, 3743; K. Weber, *J. Biol. Chem.*, 1968, **243**, 543.
[327] J. Monod, *Nature*, 1968, **218**, 106.
[328] K. Weber, *Nature*, 1968, **218**, 1116.
[329] D. C. Wiley and W. N. Lipscomb, *Nature*, 1968, **218**, 1119.
[330] R. N. Perham, *Biochem J.*, 1969, **111**, 17.
[331] R. N. Perham and J. I. Harris, *J. Mol. Biol.*, 1963, **7**, 316; G. M. T. Jones and J. I. Harris, Abstracts, 5th Meeting European Biochemical Society, Prague, 1968, p. 185.

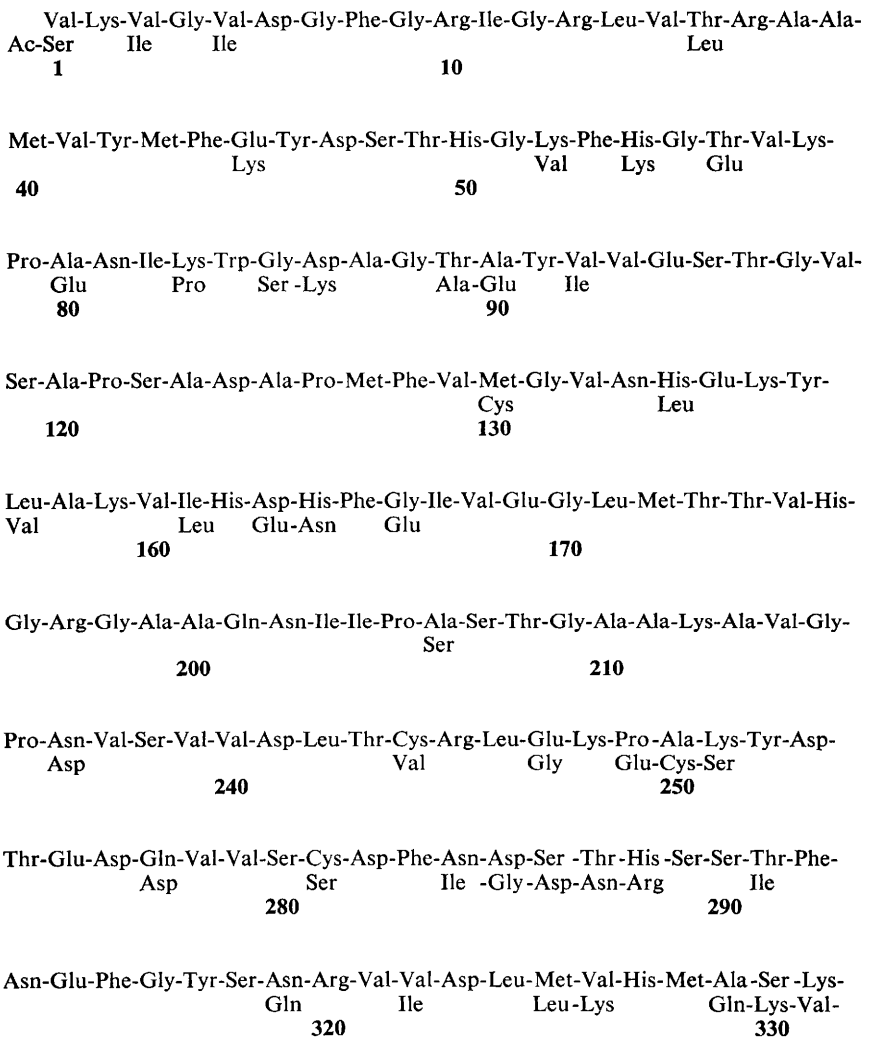


Figure 10 *Amino-acid sequence of the protein subunit of glyceraldehyde 3-phosphate is identical except where indicated below the continuous line of the pig sequence. sequence homology between the two protein chains. The active site Cys–149 is shown in*
(Reproduced by permission from *Nature*, 1968, **219**, 1025)

Phe-Asn-Ser -Gly-Lys-Val-Asp-Ile-Val-Ala-Ile-Asn-Asp-Pro-Phe-Ile-Asp-Leu-His-Tyr-
Leu-Ser -Cys Ala-Gln-Val Val Ala Glu
20 **30**

Ala-Glu-Asp-Gly-Lys-Leu-Val-Ile-Asp-Gly-Lys-Ala-Ile-Thr-Ile-Phe-Gln-Glu-Arg-Asp-
Met Ala Val Lys Val Asn Met Lys
60 **70**

Phe-Thr-Thr-Met-Glu-Lys-Ala-Gly-Ala-His-Leu-Lys-Gly-Gly-Ala-Lys-Arg-Val-Ile-Ile-
Ile Ser Phe Lys Val
100 **110**

Asp-Asn-Ser -Leu-Lys-Ile -Val-Ser-Asn-Ala-Ser-CYS-Thr-Thr-Asn-Cys-Leu-Ala-Pro-
Ser -Lys-Asp-Met-Thr-Val Thr
140 **150**

Ala-Ile-Thr-Ala-Thr-Gln-*Lys*-Thr-Val-Asp-Gly-Pro-Ser-Gly-Lys-Leu-Trp-Arg-Asp-
Val Ala Asp Gly
180 **190**

Lys-Val-Ile-Pro-Glu-Leu-Asp-Gly-Lys-Leu-Thr-Gly-Met-Ala-Phe-Arg-Val-Pro-Thr-

220 **230**

Asp-Ile-Lys-Lys-Val-Val-Lys-Gln-Ala-Ser-Glu-Gly-Pro-Leu-Lys-Gly-Ile-Leu-Gly-Tyr-
Ala-Ala-Met Thr Gln Phe
260 **270**

Asp-Ala-Gly-Ala-Gly-Ile-Ala-Leu-Asn-Asp-His-Phe-Val-Lys-Leu-Ile-Ser-Trp-Tyr-Asp-
Lys- Gln Ser -Lys-Thr Val-Val
300 **310**

Glu
Asp-Ser-Ala
334

dehydrogenase from pig skeletal muscle.[185] *The sequence of the lobster muscle enzyme* [184] *Residue* 24 *is represented as a deletion in the lobster sequence so as to maximise the bold type and capitals.*

sequences around the reactive cysteine residue of GPDH from a number of sources are listed in Table 1. This degree of homology is considerably greater than that found for the haemoglobins, cytochromes-*c*, or even the

Table 1 *Amino-acid sequence around the reactive cysteine residue of* GPDH *from a number of sources. The reactive cysteine residue is in italics*

Source	*Sequence*
Rabbit[a], Pig[b], Yeast[b], Ox[c], Chicken[d], Ostrich[d], Sturgeon[d], Badger[e], Monkey[e], Honey Bee[f]	Ile-Val-**Ser-Asn-Ala-Ser-*Cys*-Thr-Thr-Asn-Cys-Leu-Ala-Pro**-Leu-**Ala-Lys**
Human[d]	Ile-Ile-**Ser-Asn-Ala-Ser-*Cys*-Thr-Thr-Asn-Cys-Leu-Ala-Pro**-Leu-**Ala-Lys**
Halibut[d]	Val-Val-**Ser-Asn-Ala-Ser-*Cys*-Thr-Thr-Asn-Cys-Leu-Ala-Pro**-Leu-**Ala-Lys**
Blue Crab[f]	Val-Val-**Ser-Asn-Ala-Ser-*Cys*-Thr-Thr-Asn-Cys-Leu-Ala-Pro**-Val-**Ala-Lys**
Lobster[d]	Asp-Met-Thr-Val-Val-**Ser-Asn-Ala-Ser-*Cys*-Thr-Thr-Asn-Cys-Leu-Ala-Pro**-Val-**Ala-Lys**
E coli[f]	Tyr-(Glx, Asx, Gly)-Ile-Val-**Ser-Asn-Ala-Ser-*Cys*-Thr-Thr-Asn-Cys-Leu-Ala-Pro**-Leu-**Ala-Lys**

[a] I. Harris, B. P. Meriwether, and J. H. Park, *Nature*, 1963, **197**, 154. [b] R. N. Perham and J. I. Harris, *J. Mol. Biol.*, 1963, **7**, 316. [c] R. N. Perham, *Biochem. J.*, 1966, **99**, 14C. [d] W. S. Allison and J. I. Harris, *Abstracts 2nd Meeting European Biochemical Society, Vienna*, 1965, p. 140. [e] R. N. Perham, *Biochem. J.*, 1969, **111**, 17. [f] W. S. Allison, *Ann. New York Acad. Sci.*, 1968, **151**, 180.

subtilisins, which argues for a very close connection between sequence, three-dimensional structure, and mechanism of action for GPDH. Somewhat similar homologies are to be found amongst the aldolases,[332] suggesting that this may be a common feature of the glycolytic enzymes. Crystallographic analyses of these proteins are eagerly awaited to help to interpret these detailed chemical studies, for the tetrameric GPDH molecule is the largest protein yet attempted.

C. Haem Proteins.—*Haemoglobin and Myoglobin.* See also chapter 2, part III, section 2B. There is probably more information about the primary structures of the haemoglobins and myoglobin than about any other protein.[1, 333] The fascinating correlations between alterations in the primary structure, three-dimensional structure, and biological activity of the myriad human haemoglobins have most valuably been collected together and discussed.[334] It is clear that while many amino-acid replacements

[332] P. J. Anderson, I. Gibbons, and R. N. Perham, unpublished observations.
[333] C. Nolan and E. Margoliash, *Ann. Rev. Biochem.*, 1968, **37**, 727.
[334] M. F. Perutz and H. Lehmann, *Nature*, 1968, **219**, 902.

can be accommodated in the surface of the molecule, restrictions on the internal, non-polar, residues are severe.

Structural work on many more haemoglobins is still in progress. Improved techniques for the chromatography of various haemoglobin types on DEAE-Sephadex have been reported [335] and the amino-acid sequences of the soluble tryptic peptides from the β-chain of monkey (*Macaca mulatta*) haemoglobin have been described.[336] Compared with the corresponding peptides from human β-chains, six exchanges are noted. In a comparison of the α-chains of carp and human haemoglobins,[337] 67 replacements are observed, including two additions and one deletion, each of one amino-acid. Since one addition and one deletion occur in a helical region (E), it is thought that there might be slight differences in tertiary structure.[337]

The molecular weight of the haemoglobin from the insect larvae *Chironomus* has been reported as 16,000,[338] and the presence of heterogeneity in amino-acid sequence and dimer formation have been described.[339] A study of the amino-acid sequence of dolphin myoglobin is in progress,[340] the *N*-terminal 60 residues of humpback whale myoglobin have been determined in the sequenator,[79] and the amino-acid composition of the myoglobin from the Mediterranean mollusc, *Aplysia*, which is distinguished by the total absence of tyrosine and the presence of only one residue of histidine, has been reported.[341]

The haemoglobins of sheep and goats are of particular interest for a number of reasons. Both sheep and goats contain two types of haemoglobin, A and B, which are considered to be the products of allelic genes. In animals suffering from severe anaemia because of blood loss, a further variant, C, appears and can replace A entirely. The occurrence of haemoglobin C is restricted to animals carrying the A gene, no such variant appearing in animals homozygous for haemoglobin B. During recovery from anaemia, haemoglobin C diminishes and haemoglobin A returns. In sheep there is thought to be no variation in the αA- and αB-chains, but sequence studies show that βC differs from βA by at least 16 residues and from βB by at least 21 residues.[342] Chains βA and βB differ by at least seven scattered residues despite their being products of allelic genes. In goats the situation is even more interesting. The βA- and βB-chains are very similar, if not identical, but the βC-chain differs from the βA-chain

[335] A. M. Dozy, E. F. Kleihauer, and T. H. J. Huisman, *J. Chromatog.*, 1968, **32**, 723.
[336] G. Matsuda, T. Maita, M. Yamaguchi, H. Ota, M. Migita, and T. Miyauchi, *J. Biochem. (Japan)*, 1968, **63**, 136.
[337] K. Hilse and G. Braunitzer, *Z. physiol. Chem.*, 1968, **349**, 433.
[338] P. Thompson, W. Bleecker, and D. S. English, *J. Biol. Chem.*, 1968, **243**, 4463.
[339] V. Braun, R. R. Crichton, and G. Braunitzer, *Z. physiol. Chem.*, 1968, **349**, 197; G. Braunitzer, G. Buse, and S. Braig, *ibid.*, 1968, **349**, 263.
[340] P. Nedkov, N. Genov, M. Karadjova, and B. Keil, *Coll. Czech. Chem. Comm.*, 1968, **33**, 1974; N. Genov, M. Shopova, and M. Karadžova, *F.E.B.S. Letters*, 1968, **1**, 108.
[341] L. Tentori, G. Vivaldi, S. Carta, E. Antonini, and M. Brunori, *Nature*, 1968, **219**, 487.
[342] S. H. Boyer, P. Hathaway, F. Pascasio, J. Bordley, C. Orton, and M. A. Naughton, *J. Biol. Chem.*, 1967, **242**, 2211.

by at least 18 residues.[343] Moreover, the αA-chain is in fact two distinct polypeptide chains differing in at least four residues:[344]

$$\alpha A = \text{Gly-19, Ala-26, Leu-113, Asx-115,}$$
$$\alpha' A = \text{Ser-19, Thr-26, His-113, Ser-115}$$

and the αB- is identical to the αA-chain with the exception of the substitution of tyrosine at position 75.[344] Multiple differences of this sort have been reported before in the α-chains of individual rabbits and much consternation caused to protein chemists by the suggestion that such differences might arise by ambiguous translation of the genetic code.[345] However, in goats such ambiguity would be expected to give rise to an α′B- as well as an αB-chain, which is not found. It seems much more likely therefore that the αA- and α′A-chains in the goat are the products of two non-allelic, closely linked structural genes, with the αB-chain as the product of an allele of one of those two genes.[344] A multiplicity of haemoglobin molecules can then be formed by combination of the different α- and β-chains. Recent studies on the structure of the γ-chain of human foetal haemoglobin [346] and on the cell-free synthesis of rabbit haemoglobin [347] also point to the existence of multiple structural genes rather than ambiguous translation as the basis for polymorphism. Protein chemists will rest easier if the bogey of ambiguity is now finally laid.

Cytochromes. The amino-acid sequences of an enormous number of cytochromes are available [1] and have been extensively reviewed.[91, 333] Detailed evidence for the primary structure [1] of bovine cytochrome b_5 has now been published.[348] There is no striking homology with cytochrome *c* and further comparisons would better await the derivation of the three-dimensional structure. The amino-acid sequence of cytochrome c_2 of the photosynthetic bacterium *Rhodospirillum rubrum* has been determined and compared with other cytochromes.[349] With the postulation of various deletions, it is possible to obtain reasonable homology with the sequence of mammalian cytochrome *c* (Figure 11), but homology with cytochrome c_{551} from *Pseudomonas fluorescens* is rather less convincing. The postulation of gaps in sequences is, of course, something that must always be done with care, for even relatively dissimilar sequences can be made homologous by arranging judicious deletions.[350]

[343] T. H. J. Huisman, H. R. Adams, M. O. Dimmock, W. E. Edwards, and J. B. Wilson, *J. Biol. Chem.*, 1967, **242**, 2534.
[344] T. H. J. Huisman, G. Brandt, and J. B. Wilson, *J. Biol. Chem.*, 1968, **243**, 3675.
[345] G. von Ehrenstein, *Cold Spring Harbor Symp. Quant. Biol.*, 1966, **31**, 705.
[346] W. A. Schroeder, T. H. J. Huisman, J. R. Shelton, J. B. Shelton, E. F. Kleihauer, A. M. Dozy, and B. Robberson, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 537.
[347] C. Schapira, J. C. Dreyfus, and N. Maleknia, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 558.
[348] J. Ozols and P. Strittmatter, *J. Biol. Chem.*, 1968, **243**, 3376.
[349] K. Dus, K. Sletten, and M. D. Kamen, *J. Biol. Chem.*, 1968, **243**, 5507.
[350] C. R. Cantor, *Biochem. Biophys. Res. Comm.*, 1968, **31**, 410.

R. rubrum cytochrome c_2 Glu-**Gly**-**Asp**-Ala-Ala-Ala-**Gly**-Glu-**Lys**---------**Val**-Ser-**Lys**-Lys-**Cys**-Leu-Ala-**Cys**-**His**-**Thr**-Phe-Asp-
Horse heart cytochrome c **Gly**-**Asp**-Val-Glu-Lys-**Gly**-Lys-**Lys**-Ile-Phe-**Val**-Gln-**Lys**-----**Cys**-Ala-Gln-**Cys**-**His**-**Thr**-Val-Glu-
1 **10** **20**

Gln-**Gly**-**Gly**-Ala-Asn-**Lys**-Val-**Gly**-**Pro**-**Asn**-**Leu**-Phe-**Gly**-Val-**Phe**-Glu-Asn-Thr-Ala-Ala-His-Lys-Asp-Asn-Tyr-Ala-**Tyr**-Ser-
Lys-**Gly**-**Gly**-Lys-His-**Lys**-Thr-**Gly**-**Pro**-**Asn**-**Leu**-His-**Gly**-Leu-**Phe**-Gly-Arg-Lys-Thr-Gly-Gln-Ala-Pro-Gly-Phe-Thr-**Tyr**-Thr-
30 **40** **50**

Glu-Ser-Tyr-Thr-Glu-Met-**Lys**-Ala-**Lys**-**Gly**-Leu-**Thr**-**Trp**-Thr-**Glu**-Ala-Asn-**Leu**-Ala-Ala-**Tyr**-Val-Lys-**Asn**-**Pro**-Lys-Ala-Phe-
Asp-Ala-Asn----------**Lys**-Asn-**Lys**-**Gly**-Ile-**Thr**-**Trp**-Lys-**Glu**-Glu-Thr-**Leu**-Met-Glu-**Tyr**-Leu-Glu-**Asn**-**Pro**-Lys-Lys-Tyr-
60 **70**

Val-Leu-Glu-Lys-Ser-Gly-Asp-Pro-Lys-Ala-Lys-Ser-**Lys**-**Met**-Thr-**Phe**-----Lys-Leu-Thr-**Lys**-Asp-Asp-**Glu**-Ile-**Glu**-Asn-Val-
Ile-Pro-Gly----------Thr-**Lys**-**Met**-Ile-**Phe**-Ala-Gly-Ile-Lys-**Lys**-Lys-Thr-**Glu**-Arg-**Glu**-Asp-Leu-
80 **90** **100**

Ile-**Ala**-**Tyr**-**Leu**-**Lys**---------**Thr**-Leu-Lys
Ile-**Ala**-**Tyr**-**Leu**-**Lys**-Lys-Ala-**Thr**-Asn-Glu
110 **117**

Figure 11 *Comparative sequences of cytochrome* c *from horse* ***heart*** *and cytochrome* c_2 *from* R. rubrum.[349]

D. Studies on Quaternary Structure.—The methods of end-group analysis and finger-printing have been used to establish the quaternary structure of many enzymes. Some have already been discussed elsewhere in this Report; others are listed in Table 2.

Table 2 *Quaternary structure of some individual enzymes*

Enzyme	*Source*	*M.W.*	*No. of subunits*
Aldolase[a]	Mammals	160,000	4
Rhodanese[b]	Ox liver	37,000	2
Triose phosphate isomerase[c]	Rabbit muscle	60,000	2
Phosphofructokinase[d]	Rabbit muscle	92,000	4
Ceruloplasmin[e]		150,000	2 or more
Transferrin[f]	Human, rabbit, chicken	70,000–90,000	1
Glutathione reductase[g]	Yeast	124,000	2
Alcohol dehydrogenase[h]	Liver	80,000	2
	Yeast	150,000	4
Thiolase[i]	*E. coli*	170,000	4
Histidinol dehydrogenase[j]	*S. typhimurium*	80,000	2
Histidinol decarboxylase[k]	*Lactobacillus* 30a	190,000	10
Threonine deaminase[l]	*S. typhimurium*	194,000	4
ATP-guanidoacetate phosphotransferase[m]	*Nephthys coeca*	89,000	2
Met-*t*RNA synthetase[n]	*E. coli*	96,000	2
Thioredoxin reductase[o]	*E. coli*	66,000	2
Lac repressor[p]	*E. coli*	150,000	3–4

[a] K. Kawahara and C. Tanford, *Biochemistry*, 1966, **5**, 1578; E. Penhoet, M. Kochman, R. Valentine, and W. J. Rutter, *Biochemistry*, 1967, **6**, 2940; D. E. Morse and B. L. Horecker, *Adv. Enzymol.*, 1968, **31**, 125. [b] M. Volini, F. DeToma, and J. Westley, *J. Biol. Chem.*, 1967, **242**, 5220. [c] P. M. Burton and S. G. Waley, *Biochem. J.*, 1968, **107**, 737. [d] V. H. Paetkau, E. S. Younathan, and H. A. Lardy, *J. Mol. Biol.*, 1968, **33**, 721. [e] H. Mukasa, S. Kajiyama, K. Sugiyama, K. Funakubo, M. Itoh, Y. Nosoh, and T. Sato, *Biochim. Biophys. Acta*, 1968, **168**, 132. [f] F. C. Greene and R. E. Feeney, *Biochemistry*, 1968, **7**, 1366. [g] R. D. Mavis and E. Stellwagen, *J. Biol. Chem.*, 1968, **243**, 809. [h] I. Harris, *Nature*, 1964, **203**, 30; P. J. G. Butler, H. Jörnvall, and J. I. Harris, *F.E.B.S. Letters*, 1969, **2**, 239. [i] U. Gehring and J. I. Harris, *F.E.B.S. Letters*, 1968, **1**, 150. [j] J. Yourno, *J. Biol. Chem.*, 1968, **243**, 3273. [k] G. W. Chang and E. E. Snell, *Biochemistry*, 1968, **7**, 2012. [l] M. H. Zarlengo, G. W. Robinson, and R. O. Burns, *J. Biol. Chem.*, 1968, **243**, 186. [m] L.-A. Pradel, R. Kassab, C. Conlay, and N. V. Thoai, *Biochim. Biophys. Acta*, 1968, **154**, 305. [n] C. J. Bruton and B. S. Hartley, *Biochem. J.*, 1968, **108**, 281. [o] L. Thelander, *European J. Biochem.*, 1968, **4**, 407. [p] A. D. Riggs and S. Bourgeois, *J. Mol. Biol.*, 1968, **34**, 361.

6 Immunoglobulins

See also chapter 2, part III, section 2B, and chapter 5, section 2. The study of the amino-acid sequence of immunoglobulins is currently one of the most rapidly expanding areas of protein chemistry. Since more reviews [351] are devoted to this subject than perhaps to any other aspect of

[351a] R. R. Porter, 'Essays in Biochemistry,' ed. P. N. Campbell and G. D. Greville, Academic Press, London, 1967, p. 1, [b] S. Cohen and C. Milstein, *Adv. Immunol.*, 1967, **7**, 1, [c] M. Cohn and E. S. Lennox, *Ann. Rev. Biochem.*, 1967, **36**, 365. [d] E. Haber, *ibid.*, 1968, **37**, 497. [e] F. W. Putnam, *Science*, 1969, **163**, 633.

protein structure it would obviously be redundant to repeat them here. In any event, a discussion of the genetic implications is beyond the scope of this article. Until recently the study of immunoglobulins was hampered by the very great heterogeneity found within a family of basically similar molecules. This situation was revolutionised by the recognition of the connection between immunoglobulins and the homogeneous proteins secreted into the serum in the pathological condition of myelomatosis. This, together with the acceptance of the four-chain structure for IgG proposed by Porter, has led to the recent feverish activity in this field.

A. Light Chains.—Sequence information is now available on many light chains from a wide range of sources.[1, 351, 352] Complete amino-acid sequences for further κ-[353] and λ-chains[354] have recently been published and the pig π chains have been assigned to the κ-type.[355] Comparison of human Bence Jones proteins suggested[82, 356] that the sequence changes observed in the *N*-terminal halves of these molecules could be ascribed to variations of three basic sequences. These basic sequences have now been recognised in normal human light chains and shown to be non-allelic.[357] The *N*-terminal sequence of the light chain (as usual, κ-) from a cold agglutinin (a spontaneously arising monoclonal macroglobulin unassociated with myeloma) has been determined in the sequenator and also shown to fit one of the basic sequences.[358] With λ-chains, however, the number of basic sequences, if any, is in doubt and must await more evidence.[354] The κ- and λ-chains of both human and pig IgG can be separated by chromatography on SE-Sephadex[359] but are, of course, highly homologous.[351, 360] In view of the comparatively wide variation normally found between any two myeloma light chains, it is strange that the only two λ-chains so far available from mice appear to be very similar, if not identical.[361]

B. Heavy Chains. The almost complete amino-acid sequence for the F_c region of pooled rabbit IgG has now been described[362] showing it to be a unique sequence and proving, therefore, that heavy chains, like light chains, are essentially invariant for at least the *C*-terminal half of the molecule. Much effort is now being expended on establishing the amino-acid sequence of the heavy chains of myeloma proteins to determine exactly how much of the primary structure can vary.[351, 352, 363] The amino-acid

[352] Numerous papers in 'Antibodies,' *Cold Spring Harbor Symp. Quant. Biol.*, 1967, **32.**
[353] N. Hilschmann, *Z. physiol. Chem.*, 1967, **348**, 1718.
[354a] C. Milstein, J. B. Clegg, and J. M. Jarvis, *Biochem. J.*, 1968, **110**, 631. [b] B. Langer, M. Steinmetz-Kayne, and N. Hilschmann, *Z. physiol. Chem.*, 1968, **349**, 945.
[355] J. Novotny and F. Franek, *F.E.B.S. Letters*, 1968, **2**, 93.
[356] C. Milstein, *Nature*, 1967, **216**, 330.
[357] C. Milstein, C. P. Milstein, and A. Feinstein, *Nature*, 1969, **221**, 151.
[358] P. Edman and A. G. Cooper, *F.E.B.S. Letters*, 1968, **2**, 33.
[359] F. Franek, V. Brummelova, and F. Skvaril, *Biochim. Biophys. Acta*, 1968, **160**, 321.
[360] F. Franek, B. Keil, J. Novotny, and F. Šorm, *European J. Biochem.*, 1968, **3**, 422.
[361] E. Appella and R. N. Perham, *J. Mol. Biol.*, 1968, **33**, 963.
[362] R. DeLaney and R. L. Hill, *J. Biol. Chem.*, 1968, **243**, 4206.
[363] M. J. Waxdal, W. H. Konigsberg, and G. M. Edelman, *Biochemistry*, 1968, **7**, 1967.

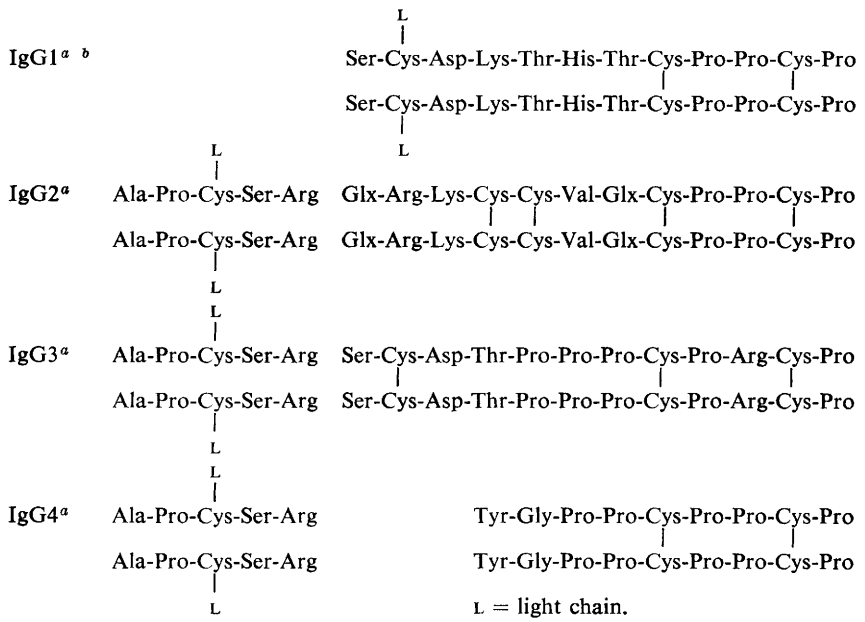
sequence of the *N*-terminal half (F_d region) of pooled rabbit IgG has been the subject of a very interesting study.[364] By limiting tryptic digestion to the arginine residues after trifluoroacetylation of the lysines, it was hoped that common structure would be emphasised and, indeed, an underlying sequence for most, if not all, of the F_d region is indicated. Similar results are obtained with purified antibodies, the simplest explanation being that comparatively few residues change each time with various antibody specificities.[364, 365]

C. Disulphide Bridges.—Intra-chain disulphide loops of *ca.* 60 residues are repeating patterns in both light and heavy chains of IgG,[366] in accord with their origin from a common genetic precursor through a series of gene duplications.[367] While it has been known for some time that the light chains are linked to the heavy chains by a disulphide bridge at their *C*-terminus, the number and location of the inter-heavy chain bridges have been difficult to determine.[351] These are now beginning to become clear for the four types of human IgG.[366, 368, 369] Although the number of inter-heavy chain bridges depends upon the type, striking homologies are apparent and some of these are shown in Table 3. Not all the bridges can be shown in this Table since some remain to be established.[366] It is particularly interesting that the cysteine residue which serves as the point of attachment of light chains in IgG1 forms an inter-heavy chain bridge in IgG3 and that the location of light chain attachment in IgG1 is far removed in the primary structure from that in the other three types. The biological significance of these differences remains obscure. The comparable region of rabbit IgG has also been reported,[369, 370] and the close resemblance to human IgG1 noted.[369] In all these proteins, the inter-heavy chain bridges are parallel and the unusual preponderance of proline in this region may explain the susceptibility of this portion of the primary structure to proteolytic attack in the native immunoglobulins.[351, 370]

Most of these experiments refer to immunoglobulin G. With other classes the situation may be somewhat different. Thus there has been a report[371] that in IgA the light chains are not bound covalently to the heavy chains by disulphide bridges and that monomer IgA molecules associate by disulphide bridge formation to yield aggregates of the basic structure.

[364] J. J. Cebra, D. Givol, and R. R. Porter, *Biochem. J.*, 1968, **107**, 69; J. J. Cebra, L. A. Steiner, and R. R. Porter, *ibid.*, 1968, **107**, 79.
[365] R. R. Porter, *Biochem. J.*, 1967, **105**, 417.
[366] B. Frangione, C. Milstein, and J. R. L. Pink, *Nature*, 1969, **221**, 145.
[367] R. L. Hill, H. E. Lebovitz, R. E. Fellows, and R. DeLaney, 'Gamma Globulins: Structure and Control of Biosynthesis,' ed. J. Killander, Nobel Symposium **3**, Interscience, New York, 1967, p. 109.
[368] B. Frangione and C. Milstein, *J. Mol. Biol.*, 1968, **33**, 893.
[369] L. A. Steiner and R. R. Porter, *Biochemistry*, 1967, **6**, 3957.
[370] D. S. Smyth and S. Utsumi, *Nature*, 1967, **216**, 332.
[371] C. A. Abel and H. M. Grey, *Biochemistry*, 1968, **7**, 2682.

Table 3 *Homologous sequences around the interchain bridges in human* IgG.



[a] B. Frangione, C. Milstein, and J. R. L. Pink, *Nature*, 1969, **221**, 145.
[b] L. A. Steiner and R. R. Porter, *Biochemistry*, 1967, **6**, 3957.

D. Haptoglobin and Haemagglutinin.—Since invertebrates are incapable of induced antibody formation,[351] it is interesting to compare agglutinins with immunoglobulins. A recent study[372] of haemagglutinin from the horse-shoe crab *Limulus* has revealed that it is probably composed of a ring structure of six subunits, each of which contains three polypeptide chains of molecular weight 22,500, giving a total molecular weight of *ca.* 400,000. No similarity of immunoglobulin light chains and these haemagglutinin chains is yet apparent.

In the sense that haptoglobin combines specifically with oxyhaemoglobin and is composed of two heavy and two light chains linked through disulphide bridges, it also shows superficial resemblance to antibodies. The complete amino-acid sequences of the smaller chains of human haptoglobin have been determined and reported to show slight but significant correlation with the sequence of human immunoglobulin light chains.[373] If there is a common evolutionary origin, appreciable divergence has obviously since occurred. These haptoglobin sequences are also very suggestive of gene

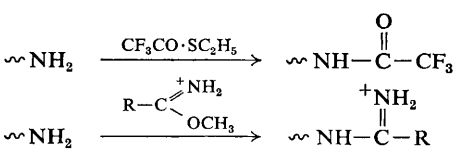
[372] J. J. Marchalonis and G. M. Edelman, *J. Mol. Biol.*, 1968, **32**, 453.
[373] J. A. Black and G. H. Dixon, *Nature*, 1968, **218**, 736.

duplication by unequal crossover at meiosis leading to the production of elongated peptide chains.[300, 373]

7 Chemical Modification of Proteins

The biological usefulness and, indeed, much of the chemical interest, of proteins lies in their unique capacity for variety within the confines of a fundamentally similar structure. It should cause no surprise, therefore, if chemical modifications of specific residues which succeed with one protein proceed differently with another. For example, it might have been supposed that the range of reactions with iodoacetic acid had been exhausted but recent investigations [374] add alkylation of carboxyl groups to the list. Many reagents introduced for the analysis of primary structure have found application in the study of protein modification and have been discussed in Section 2: others doubtless will suggest themselves. A recent review [375] gives a valuable summary of reagents and reactions and many experimental prescriptions have been collected.[4] Comment here will mostly be restricted to developments since that time. Enzymes are not considered here.

A. Reactions of Lysine and Arginine.—The reactivity of the nucleophilic amino-group has made it a favourite target for the protein modifier. Although lysine residues are almost entirely restricted to the 'surface' of protein molecules, the ε-amino-groups frequently display differing reactivities towards, for example, trinitrobenzenesulphonic acid.[376] Two highly specific reagents for amino-groups are *S*-ethyl trifluorothioacetate [152] and the imido esters.[161, 377] Since reaction with the former removes the positive charge from the lysine side-chain whereas reaction with the latter maintains it, these reagents can be used in complementary experiments to investigate the role of lysine residues in protein activity (Scheme 16). Thus



Scheme 16

it has been shown [153] that positive charge on one or both of the lysine residues in tobacco mosaic virus protein is essential for the smooth self-assembly into virus-like rods. These reagents have the added advantage

[374] K. Takahashi, W. H. Stein, and S. Moore, *J. Biol. Chem.*, 1967, **242**, 4682.
[375] L. A. Cohen, *Ann. Rev. Biochem.*, 1968, **37**, 695.
[376] R. B. Freedman and G. K. Radda, *Biochem. J.*, 1968, **108**, 5P; A. Loverde and P. Strittmatter, *J. Biol. Chem.*, 1968, **243**, 5779.
[377] W. F. Benisek and F. M. Richards, *J. Biol. Chem.*, 1968, **243**, 4267.

that the modification can be reversed and the regeneration of native protein studied. Although irreversible, reductive alkylation of amino-groups, which also maintains the positive charge on the side-chain, may find future application.[378]

Maleylation has been used to show that antibodies directed against neutral or positively charged haptens do not have amino-groups in the active site,[379] the reversibility of the modification again being turned to advantage. Similarly, modification of arginine residues by reaction with trimeric 2,3-butanedione [163] suggests that arginine is at the active site of antibodies directed against negatively charged haptens but not of those directed against neutral or positively charged haptens.[380] No doubt the recent development of arginine modifications that can be effected under mild conditions (p. 58) will lead to more studies of this kind.

Few reactions distinguish between α- and ε-amino groups, with the notable exception of guanidination [375] which seems to modify only the ε-amino-groups of lysine residues. However, in favourable cases, the α-amino-group of a protein can be removed by transamination with glyoxalate or pyridoxal as amino-acceptor and copper(II) or cobalt(I) ions as catalyst.[381] It seems likely that specificity for α-amino-groups is obtained because of the chelate formed at the *N*-terminus involving the *N*-terminal peptide carbonyl group. The terminal oxoacyl group can subsequently be removed by reaction with a suitable dinucleophile, shortening the protein chain by one residue, a reaction successfully carried out with several proteins.[382]

B. Reactions of Carboxyl Groups.— The unreactivity of carboxyl groups makes them difficult to modify. Esterification has not had much success and reduction to the corresponding alcohol is often too violent for most proteins. There have recently been reports [383] that reduction of protein carboxyl groups with diborane can be achieved under suitably mild conditions, although specific reduction of all available carboxyl groups appears difficult. Another approach to this problem is the use of water-soluble carbodi-imides to activate protein carboxyl groups for coupling with amines, *e.g.* glycine methyl ester.[384] While reaction with carboxyl groups predominates, it is probable that side-reaction with tyrosine hydroxyl groups can occur,[385] which, fortunately, can be reversed by treatment with a suitable nucleophile, *e.g.* hydroxylamine. The carbodi-imide modification has now been used successfully to demonstrate the functional significance

[378] G. E. Means and R. E. Feeney, *Biochemistry*, 1968, **7**, 2192.
[379] M. H. Freedman, A. L. Grossberg, and D. Pressman, *Biochemistry*, 1968, **7**, 1941.
[380] A. L. Grossberg and D. Pressman, *Biochemistry*, 1968, **7**, 272.
[381] H. B. F. Dixon, *Biochem. J.*, 1967, **103**, 38P.
[382] S. van Heyningen and H. B. F. Dixon, *Biochem. J.*, 1967, **104**, 63P.
[383] M. Z. Atassi and A. F. Rosenthal, *Biochem. J.*, 1969, **111**, 593.
[384] D. G. Hoare and D. E. Koshland jun., *J. Biol. Chem.*, 1967, **242**, 2447; H. Horinishi, K. Nakaya, A. Tani, and K. Shibata, *J. Biochem. (Japan)*, 1968, **63**, 41.
[385] K. L. Carraway and D. E. Koshland jun., *Biochim. Biophys. Acta*, 1968, **160**, 272.

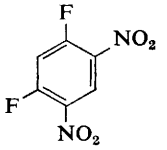
of the carboxyl side-chains of Asp-52 and Glu-35 in lysozyme [386] and the method will probably find wider application.

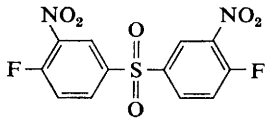
C. Reactions of Cysteine and Cystine.—Numerous reagents have been introduced for the selective alkylation of protein sulphydryl groups.[4, 375] New conditions for the reaction with α,β unsaturated compounds, *e.g.* acrylonitrile, have been described [387] under which the competing reaction with ε-amino-groups [388] is avoided. In this connection it is worth noting that, by analogy, unblocked reactive sulphydryl groups might add across the double bond of the reversible amino-group reagents, maleic and citraconic anhydrides. Certainly one should be aware of this possibility in using these reagents. Treatment of proteins containing sulphydryl groups with sodium sulphite and catalytic amounts of cysteine in 8M-urea is reported to lead to complete *S*-sulphonation,[389] presumably by the formation of disulphide bonds and subsequent sulphitolysis. Curiously, mercaptoethanol or dithiothreitol will not substitute for cysteine whereas 2-mercaptoethylamine will, pointing to the need for an amino-group in the thiol reagent. Since exposure to an excess of mercaptoethanol removes the *S*-sulpho-groups, this provides a method for the reversible blocking of cysteine residues. In tests with aldolase,[389] enzyme activity was completely recovered after removal of the sulpho-groups from the inactive derivative.

Reduction of the disulphide bridges of proteins can be made highly specific. It is well known that inter-chain disulphide bridges are more easily reduced than are intra-chain [390] and intra-chain bridges can also frequently be distinguished by their differences in reactivity. For example, one particular bridge can be cleaved in trypsinogen,[283] one in growth hormone,[391] and two in ribonuclease A.[392] Since these reduced molecules frequently retain biological activity [391-393] and therefore, presumably, three-dimensional structure, there must exist the exciting possibility that the thiol groups generated in the protein could be used to attach mercurials for *X*-ray crystallographic studies.

D. Cross-linking Reactions.—The use of bifunctional reagents for investigating the folding of polypeptide chains has been reviewed in detail by Wold.[394] 1,5-Difluoro-2,4-dinitrobenzene (F_2DNB) (10) and *p,p'*-difluoro-*m,m'*-dinitrophenylsulphone (F_2DPS) (11), the bridge spans of which differ

[386] T.-Y. Lin and D. E. Koshland jun., *J. Biol. Chem.*, 1969, **244**, 505.
[387] J. F. Cavins and M. Freedman, *J. Biol. Chem.*, 1968, **243**, 3357.
[388] J. F. Cavins and M. Freedman, *Biochemistry*, 1968, **6**, 3766.
[389] W. W.-C. Chan, *Biochemistry*, 1968, **7**, 4247.
[390] H. Zahn and H. G. Gattner, *Z. physiol. Chem.*, 1968, **349**, 373; A. Massaglia, F. Pennisi, U. Rosa, S. Ronca-Testoni, and C. A. Rossi, *Biochem. J.*, 1968, **108**, 247.
[391] J. B. Mills and A. E. Wilhelmi, *Ann. New York Acad. Sci.*, 1968, **148**, 343.
[392] H. Neumann, I. Z. Steinberg, J. R. Brown, R. F. Goldberger, and M. Sela, *European J. Biochem.*, 1967, **3**, 171.
[393] T. A. Bewley, J. S. Dixon, and C. H. Li, *Biochim. Biophys. Acta*, 1968, **154**, 420.
[394] F. Wold, *Methods in Enzymology*, 1967, **11**, 617.


(10)


(11)

considerably, have been among the most successfully used. Reaction with amino- and tyrosine phenolic-groups predominates but sulphydryl and imidazole groups are also affected. Improved chromatographic analysis of the reaction products with lysine and tyrosine residues has been described.[395] Hexamethylene di-isocyanate has been reported to react with ribonuclease to yield a fully active derivative with one intramolecular bridge between unidentified lysine residues.[396] However, great care must be exercised in interpreting cross-linking experiments to give information about polypeptide chain conformation. For example cross-links imputed to the amino-groups of lysozyme following reaction with phenol-2,4-disulphonyl chloride would appear to be possible only with rearrangement of the lysine side-chains in the accepted three-dimensional structure of lysozyme.[397] It is not unlikely that reaction with one end of a bulky bifunctional reagent can cause significant distortion of a 'native' protein structure, sufficient to allow previously impossible links to form. None the less, providing these limitations are recognised, cross-linking reactions clearly have considerable potential, particularly in proteins where *X*-ray crystallographic analysis is impossible.

A nice example of the use of these methods is provided by recent work on *Staphylococcal* nuclease.[398] Reaction of the nuclease with tetranitromethane in the presence and absence of the competitive inhibitor deoxythymidine 3′,5′-diphosphate results in the selective nitration of tyrosyl residues 15 and 85 respectively. The mononitrotyrosyl derivatives can then be reduced with sodium dithionite to monoaminotyrosyl derivatives. Since the p*K* of this amino-group is 4·7, well below that of other protein amino-groups, selective reaction of bifunctional reagents with the unprotonated form can be arranged at pH 5. After removal of excess of reagent by gel filtration, the pH is raised to 9·4 and the other end of the bound reagent allowed to react with susceptible groups in its vicinity. Depending on the bifunctional reagent used, various bridges can be formed and identified.

A short digression on nitration is pertinent here. Since its introduction,[399] nitration of tyrosine residues in proteins has become a popular modification

[395] H. W. Kunz, C. F. Bernard, and T. J. Gill, *J. Chromatog.*, 1968, **32**, 786.
[396] H. Ozawa, *J. Biochem. (Japan)*, 1967, **62**, 419.
[397] G. L. Moore and R. A. Day, *Science*, 1968, **159**, 210.
[398] P. Cuatrecasas, S. Fuchs, and C. B. Anfinsen, *J. Biol. Chem.*, 1969, **244**, 406.
[399] M. Sokolovsky, J. F. Riordan, and B. L. Vallee, *Biochemistry*, 1966, **5**, 3582.

reaction applied to many proteins, *e.g.* proteolytic enzymes,[400] pancreatic trypsin inhibitor,[405] and antibodies.[402] However, possible side-reactions with cysteine residues have been recognised [399] and, in the case of aldolase, substantial oxidation of thiol groups can occur with little concomitant tyrosine modification.[403] This oxidation can only partially be reversed with 2-mercaptoethanol, suggesting that disulphide bridge formation is not the entire answer. Other results with collagen [404] have indicated that some cross-linking involving tyrosine residues can occur during nitration and the single tryptophan of *Staphylococcal* nuclease will also react to form a yellow derivative.[405] All of which points to the need for circumspection in any chemical modification experiment, even when comparatively well-characterised reagents are being used.

In contrast, cross-linking of proteins with glutaraldehyde, the chemistry of which has been obscure to say the least, has served the enzymologist and crystallographer well even in the absence of precise knowledge of the cross-links induced.[406] While reaction is predominantly with amino-groups,[406-408] simple Schiff base formation is excluded since the reaction is irreversible.[406, 408] N.m.r. spectroscopy indicates that glutaraldehyde is largely polymeric, containing significant amounts of α- and β-unsaturated aldehydes.[409] Such oligomeric aldehydes, derived from aldol condensation products by loss of water, then react with amino-groups to give Michael-type adducts which will be stable to acid hydrolysis. Since the oligomers of glutaraldehyde can vary in size, the length of the cross-link formed can also vary.

PART II: Secondary Structures

by P. M. Hardy

1 Introduction

Three levels of organisation in proteins and peptides are recognised.[1] The primary structure concerns the linear sequence of the amino-acid residues, the secondary the types of folding of the backbone chain into three-dimensional structures often involving hydrogen bonding between the CO and NH groups of the peptide linkages, while the tertiary structure deals with the further folding of the secondary structure to interrelate regions of the chain not necessarily in contact with each other, and so

[400] J. F. Riordan, M. Sokolovsky, and B. L. Vallee, *Biochemistry*, 1967, **6**, 358; S. V. Shlyapnikov, B. Meloun, B. Keil, and F. Šorm, *Coll. Czech. Chem. Comm.*, 1968, **33**, 2292; R. A. Kenner, K. A. Walsh, and H. Neurath, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 353.

[401] B. Meloun, I. Frič, and F. Šorm, *European J. Biochem.*, 1968, **4**, 112.

[402] S. Fuchs and D. Givol, *F.E.B.S. Letters*, 1968, **2**, 45.

[403] J. F. Riordan and P. Christen, *Biochemistry*, 1968, **7**, 1525.

[404] R. J. Doyle, J. Bello, and O. A. Roholt, *Biochim. Biophys. Acta*, 1968, **160**, 274.

[405] P. Cuatrecasas, S. Fuchs, and C. B. Anfinsen, *J. Biol. Chem.*, 1968, **243**, 4787.

[406] F. A. Quiocho and F. M. Richards, *Biochemistry*, 1966, **5**, 4062.

[407] A. F. S. A. Habeeb and R. Hiramoto, *Arch. Biochem. Biophys.*, 1968, **126**, 16.

[408] W. H. Bishop and F. M. Richards, *J. Mol. Biol.*, 1968, **33**, 415.

[409] F. M. Richards and J. R. Knowles, *J. Mol. Biol.*, 1968, **37**, 231.

[1] K. Linderstrøm-Lang, 'Lane Medical Lectures,' Stanford University Press, 1952, p. 58.

determines the overall molecular shape. More recently, the quaternary structure has been defined as the association of the constituent polypeptide chains or subunits of the more complex proteins.[2] This section of the Report is primarily concerned with secondary structure, but in the discussion of oligopeptides and of the examination of proteins by conformational probes and spectroscopic methods all conformational aspects are considered.

The regular folding of the backbone chain of proteins to form secondary structures was suggested first in 1928 by *X*-ray studies on fibrous proteins.[3] Such work led to the theoretical prediction by Pauling and his coworkers[4] of the three basic secondary structures of polypeptides, the α-helix, the β-pleated sheet, and the random coil or non-periodic conformation. Such structures were first established for synthetic polyamino-acids and later for some fibrous proteins in the solid state by *X*-ray diffraction investigations. This early work on solid synthetic polyamino-acids showed many of these homopolymers to exist either as single secondary structures or as mixtures of the three basic types.[5, 6a] I.r. characterisation of β-structures was subsequently developed, and other helical forms were reported on the basis of *X*-ray results. A distorted α-helix was termed the ω-helix,[7] and poly-L-proline[6j] and polyglycine[8] were found to exist in helical forms without any intramolecular hydrogen bonds. The former has no NH group, while the latter hydrogen bonds intermolecularly. Conformational studies in solution were found more difficult than in the solid state, but were pursued in view of their relevance to the physiological functions of proteins. In solution there is no absolute method equivalent to *X*-ray diffraction for the solid state, so the assignment of conformation is generally based on the concordance of evidence from several sources, mainly spectroscopic. The existence of β-structure in solution was for a long time based on the assumption that the i.r. absorption of solid polypeptides underwent no change on solution, but o.r.d. and c.d. characterisation now provides a check on this.[9] The α-helical nature of some homopolymers in

[2] J. D. Bernal, *Discuss. Faraday Soc.*, 1958, **25**, 7.

[3] K. H. Meyer and H. Mark, *Chem. Ber.*, 1928, **61**, 1932.

[4] L. Pauling, R. B. Corey, and H. R. Branson, *Proc. Nat. Acad. Sci. U.S.A.*, 1951, **37**, 205; L. Pauling and R. B. Corey, *ibid.*, p. 729.

[5] Reviewed by [a] C. H. Bamford, A. Elliott, and W. E. Hanby in 'Synthetic Polypeptides,' Academic Press, New York, 1956; and recently by [b] S. N. Timasheff and M. J. Gorbunoff, *Ann. Rev. Biochem.*, 1967, **36**, 13.

[6] See the following review articles in 'Poly-α-Amino-acids,' ed. G. D. Fasman, Edward Arnold, London, 1967, and references therein: [a] *X*-Ray diffraction by synthetic polypeptides, A. Elliott, p. 1; [b] Infrared spectra, T. Miyazawa, p. 69; [c] Light scattering and hydrodynamics, H. Benoit, L. Freund, and G. Spach, p. 105; [d] Ultraviolet spectra, W. B. Gratzer, p. 177; [e] O.r.d. spectra, Y. T. Yang, p. 239; [f] C.d. spectra, S. D. Beychok, p. 293; [g] Hydrogen exchange, S. W. Englander, p. 339; [h] Theory of non-covalent structure, D. Poland and H. A. Scheraga, p. 391; [i] Factors responsible for conformational stability, G. D. Fasman, p. 499; [j] Poly-L-proline, L. Mandelkern, p. 675.

[7] E. M. Bradbury, L. Brown, A. R. Downie, A. Elliott, and W. E. Hanby, *Proc. Roy. Soc.*, 1960, *A*, **259**, 110.

[8] F. H. C. Crick and A. Rich, *Nature*, 1955, **176**, 780.

[9] P. K. Sarkar and P. Doty, *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **55**, 981; G. D. Fasman, 'Protides of the Biological Fluids,' vol. 14, Elsevier, Amsterdam, 1966, p. 453.

solution in only weakly interacting solvents was originally established by viscosity and light scattering measurements.[6c, 10] The use of spectroscopic measurements for the investigation of the secondary structure of proteins is based on their 'calibration' with polyamino-acids.[6] Since *X*-ray diffraction studies have now completely established the secondary structures of several globular proteins,[11] further spectroscopic correlations can be made.

2 Infrared

Infrared spectroscopy was one of the first physical techniques used to investigate polypeptide secondary structure.[5] The dichroism of oriented films can be measured as well as the characteristic absorption frequencies. Attempts, however, to extend dichroic studies to solutions by using flow techniques [12] or magnetic fields [6b, 13] to orient the molecules have met with little success. The frequencies of the amide I and amide II bands in the various polypeptide conformations were initially established by complementary i.r. and *X*-ray diffraction studies,[5a] and show good agreement with those predicted on theoretical grounds.[6b] The amide I band is generally sharper than the amide II and has therefore been of greater use in determining polypeptide conformations. A strong amide I band at 1630 cm.$^{-1}$ provides a clear means for identifying β-structures, and the weak but well-defined band at 1685–1700 cm.$^{-1}$ distinguishes between the parallel and antiparallel forms. The only material so far found to exhibit a pure parallel β-structure is β-keratin.[6b] In solution the nature of the dependence on concentration of the amide I band distinguishes between intermolecular pleated sheet structures and intramolecular cross-β conformations. The amide I band frequencies for the α-helix (1650 cm.$^{-1}$) and the random coil forms (1656 cm.$^{-1}$) are too close to allow clear distinction between these structures, although in oriented films the dichroism of the band enables this difficulty to be partly overcome.[14]

Although the amide I absorption for both *cis*- and *trans*-amide conformations occurs at the same wavelength, they can be differentiated in the amide II and amide III regions. The *cis*-form absorbs at 1450 and 1350 cm.$^{-1}$ and the *trans* at 1550 and 1250 cm.$^{-1}$. The amide V band is also sensitive to the type of secondary structure. Absorption is observed at 700, 650, and 620 cm.$^{-1}$ for the extended form, the disordered form, and the α-helical form respectively. The α-helical and random coil forms can therefore be distinguished even in unoriented films. In the far i.r. a band is found for the α-helix at 400 cm.$^{-1}$, well separated from the bands due to other

[10] P. Doty, J. H. Bradbury, and A. M. Holzer, *J. Amer. Chem. Soc.*, 1956, **78**, 947.
[11] L. Stryer, *Ann. Rev. Biochem.*, 1968, **37**, 25, and refs. therein.
[12] G. R. Bird and E. R. Blout, *J. Amer. Chem. Soc.*, 1959, **81**, 2499.
[13] G. Spach, *Compt. rend.*, 1959, **249**, 667.
[14] R. D. B. Fraser, B. S. Harrap, J. P. McRae, F. H. C. Stewart, and E. Suzuki, *J. Mol. Biol.*, 1965, **12**, 482; *ibid.*, 1965, **14**, 423.

secondary structures, and so of diagnostic value. The far i.r. spectra of poly-L-alanine in the helical and β-forms has recently been studied.[15]

3 Optical Rotatory Dispersion and Circular Dichroism

O.r.d. has been the most widely used spectroscopic technique for studying secondary structure.[6e, 9, 16] Despite its limitations, it provides at least a semi-quantitative estimate of the amounts of α-helical and disordered forms in solution. The optical rotatory power observed in polypeptides is due mainly to the optically active absorption bands of the peptide chromophore found at 145, 185, and 225 nm.[17] The contribution to the overall symmetry of the molecule by the individual asymmetric carbon atoms in each amino-acid residue is usually overshadowed by the much larger asymmetry due to the periodic folding of the peptide chain in the α-helical structure. The rotations measured in the visible spectral range are due to the trailing end of the Cotton effects found in the absorption band region. The more accessible long-wavelength region, 350–600 nm., was the first to be investigated.[6e, 16] The o.r.d. curves of polyamino-acids in the random coil form are found, in this spectral region distant from their absorption bands, to obey a one-term Drude equation (usually by use of a graphical treatment) and the dispersion is termed simple.[16c] Helical forms, however, generate dispersion curves which require the two-term Moffitt equation to describe them, and the dispersion is then called complex. When the wavelength term λ_0 in the Moffitt equation is 212 nm., most polyamino-acids in helix-promoting (helicogenic) solvents give values for the Moffitt equation parameter b_0 of *ca.* -630.[16c] The Moffitt equation was originally derived theoretically for the α-helix by considering it as a rigid array of identical chromophores which interact to form a co-operative unit acting as a single exciton system, but since certain important terms were omitted in the derivation [18] the equation is now used on an empirical basis. The Moffitt equation will also describe the dispersion of a β-structure; it cannot distinguish between α-helical and extended β-structures, but the constants are characteristically different. Values of b_0 in the range -240–$+190$ have been reported for β-structures, although more frequently they lie between 0 and 100. The presence of more than small amounts of β-structure in peptides and proteins can therefore interfere with the estimation of helix content.[6e, 6i]

A helical conformation made up of either L- or D-amino-acids can exist in two spiral forms; the screw sense can be either right-handed or

[15] K. Itoh, T. Nakahara, T. Shimanouchi, M. Oya, K. Uno, and Y. Iwakura, *Biopolymers*, 1968, **6**, 1759.

[16] Reviewed by [a] P. Urnes and P. Doty, *Adv. Protein. Chem.*, 1961, **16**, 401; [b] G. D. Fasman, 'Methods in Enzymology,' Academic Press, New York, 1963, vol. 6, p. 928; [c] Y. T. Yang, 'Conformations of Biopolymers,' Vol. I, Academic Press, New York, 1967, p. 173.

[17] W. Moffitt, *J. Chem. Phys.*, 1956, **25**, 467; *Proc. Nat. Acad. Sci. U.S.A.*, 1956, **42**, 736.

[18] D. D. Fitts and J. G. Kirkwood, *J. Amer. Chem. Soc.*, 1956, **78**, 2650.

left-handed. The suggestion that the screw sense of an α-helix is related to the sign of its b_0 value[17] has been confirmed by *X*-ray analysis, which has established that a negative b_0 value for a peptide composed of L-amino-acids corresponds to a right-handed helix.[19] Purely from energy considerations, the orientation and interaction of the side-chains with respect to the peptide backbone differ in the two spiral forms. In 1952 Huggins in fact predicted that a right-handed helix is a more stable conformation for most L-polypeptides than a left-handed one.[20] A few polyamino-acids, notably some esters of poly-L-aspartic acid, do form left-handed helices. Since the b_0 value for a random coil form is usually near zero, on the basis of a linear model the percentage of α-helix in a peptide $= -100b_0/630$. Estimates of the α-helical content of proteins in solution obtained in this way have given in general fair agreement with those determined for the 'wet' crystals by *X*-ray crystallography, although ferrichrome C and carbonic anhydrase C show large discrepancies.[21] In myoglobin, for example, a helical content of *ca.* 75% is found by both methods.[19] It is the fact that the λ_0 of helices is close to the parameter λ_c for the random coil in the Drude equation that enables one equation to characterise mixtures and so yield a parameter b_0 uniquely related to helical content.[16b] It should be noted that within normal experimental accuracy the rotatory dispersion of partly helical peptides (up to *ca.* 40–50%) is indistinguishable from a simple dispersion curve described by a one-term Drude equation. Attempts have been made to improve the Moffitt equation, *e.g.* by use of modified two-term Drude equations, but so far it remains the best general equation for the estimation of helix content.[22]

With the advent of recording spectropolarimeters, o.r.d. measurements down to 185 nm. can now be made. However, many helicogenic organic solvents absorb strongly at lower wavelengths, making 250 nm. the usual lower-wavelength limit of investigation except in aqueous or trifluoroethanol solutions. Extension of measurements into the far u.v. has enabled the Cotton effect peaks to be fully seen.[6b, 6e] The helical conformation is characterised by a trough at 233 nm. and a peak at 198 nm., with the crossover at 224 nm., and another trough at 184 nm., whereas the random-coil form has a negative Cotton effect with a trough at 205 and a peak at 190 nm., with a zero rotation near 198 and another much smaller Cotton effect near 230 nm.[6b, 6e] (see Figure 1). A fully helical peptide has a mean residue rotation of *ca.* −15,000 deg. cm.²/decimole at 233 nm. (whereas for a random coil the figure is −1800) and *ca.* +75,000 deg. cm.²/decimole at 198 nm. The magnitudes of these peaks can also be used to estimate helical content.[6e]

[19] J. C. Kendrew, H. C. Watson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore, *Nature*, 1960, **185**, 422.
[20] M. L. Huggins, *J. Amer. Chem. Soc.*, 1952, **74**, 3963.
[21] S. Beychok, *Ann. Rev. Biochem.*, 1968, **37**, 445, and refs. therein.
[22] E. Shechter, J. P. Carver, and E. R. Blout, *J. Chim. Phys.*, 1968, **65**, 118.

The presence of significant amounts of other chromophores such as aromatic residues or disulphide links [23] in the peptide which absorb in the same spectral region as the peptide bond can cause errors in the estimation of helical content.[6e] In the case of poly-L-tyrosine, two new Cotton effects with peaks at 288 and 254 nm. are observed, and a b_0 of +600 obtained in helicogenic solvents. This indicates a left-handed helix, although other

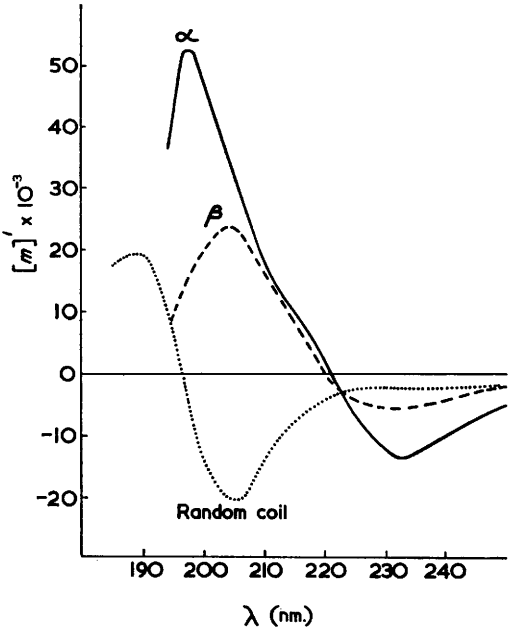


Figure 1 *The o.r.d. spectra of poly-L-lysine in three different conformations in water*
(Reproduced by permission from *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **55**, 981)

evidence shows that the helix is in fact right-handed.[24] A study of the optical activity of the disulphide bond in cystine and some of its derivatives indicates that disulphide Cotton effects would account for more than 10% of the observed rotation near 210 nm. for a peptide containing two S—S bonds per 100 residues when the helix content is between 25 and 40%. The 199 nm. Cotton effect of the disulphide bond may be directly related to its inherent asymmetry, whereas the 262 nm. Cotton effect may result primarily from environmental perturbations.[25]

Circular dichroism is an absorptive phenomenon, *i.e.* it is observed only at wavelengths where absorption occurs.[6b] The dispersive nature of

[23] D. L. Coleman and E. R. Blout, 'Conformations of Biopolymers,' vol. I, Academic Press, New York, 1967, p. 123.
[24] G. D. Fasman, E. Bodenheimer, and C. Lindblow, *Biochemistry*, 1964, **3**, 1665.
[25] D. L. Coleman and E. R. Blout, *J. Amer. Chem. Soc.*, 1968, **90**, 2405.

optical rotation, which makes it so useful in the near u.v., is more a handicap at lower wavelengths because it is difficult to separate the overlapping Cotton effects and so measure their rotatory strengths. For this reason c.d. is proving to be a more powerful technique in this region of the spectrum.[6b] Development of instruments for c.d. work lagged behind those for o.r.d., but the study of c.d. spectra has been catching up rapidly in the past two years. It is possible to calculate o.r.d. spectra from c.d. results

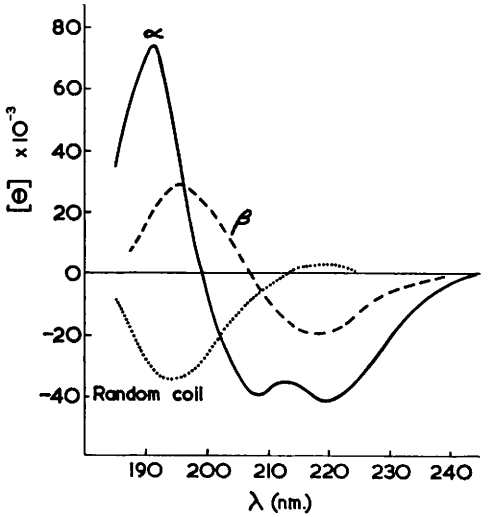


Figure 2 *The c.d. spectra of poly-α-L-lysine in water in three different conformations* (Reproduced by permission from 'Protides of the Biological Fluids', vol. 14, Elsevier, Amsterdam, 1966)

using the Kronig–Kramers transform, and the positions of the o.r.d. extrema as calculated from c.d. spectra agree well with those obtained experimentally.[26] The c.d. spectrum of the α-helix shows negative extrema at 222 and 208 and a positive extremum at 190 nm.[27a] (see Figure 2). Although earlier workers found quantitative differences in the rotatory strengths of the c.d. bands of both the helical and random coil forms,[6b] due partly perhaps to instrumental differences, estimates of helical content from c.d. band amplitudes have now been reported to give good agreement with o.r.d. results.[27b]

The Cotton effects of the β-structure initially proved difficult to measure, in part owing to the water insolubility and tendency to aggregation of

[26] J. P. Carver, E. Shechter, and E. R. Blout, *J. Amer. Chem. Soc.*, 1966, **88**, 2550.

[27] [a] G. M. Holzworth, W. B. Gratzer, and P. Doty, *J. Amer. Chem. Soc.*, 1962, **84**, 3194; G. M. Holzworth and P. Doty, *ibid.*, 1965, **87**, p. 218; E. Iizuka and J. T. Yang, *Biochemistry*, 1965, **4**, 1249; [b] H. Hashizume, M. Shiraki, and K. Imahari, *J. Biochem. (Japan)*, 1967, **62**, 543.

peptides in this conformation, but in the past two years the far-u.v.–o.r.d. curves of several polyamino-acid β-structures have been obtained and found to fall into two classes. The first class, which includes poly-L-lysine, has a trough at 230 and a peak at 205 nm. (see Figure 1), while those of the second class show a trough at 240 with a peak between 210 and 215 nm.[28] Since the extrema are of smaller magnitude than the absorption due to the helical conformation at the same wavelength, in mixtures of α- and β-conformations, the β-form can be overshadowed by the α-helix rotational strength.

O.r.d. curves have been computed for many different mixtures of α-helix, β-chain, and random coil, and comparison of observed protein spectra with these can allow decision among several alternative compositions in some cases.[29] No satisfactory assignment of the two forms of β-structure o.r.d. spectra to one or another of the three known types of β-structure on the basis of correlation with i.r. spectra can be made.[21] It has been predicted on theoretical grounds that the Cotton effects of the parallel and anti-parallel β-forms should have opposite signs, but the parallel β-structure has not yet been studied in the far-u.v.[30] O.r.d. results from β-structures in films and solutions agree well. Recently the c.d. spectra of films of poly-amino-acids in the β-form have also been found to fall into two classes which correspond with the o.r.d. results.[31] Those similar to the poly-L-lysine β-structure have a negative band near 219 crossing over at 207 to a positive band with a maximum near 195 nm. (see Figure 2); the others like poly-*S*-carboxymethyl-L-cysteine have a positive band at 198 and a negative one at 227 nm. While the band positions are essentially the same for various polyamino-acids, the relative intensities depend upon the particular side-chain.[31]

A new modulated method of linear dichroism has been reported. Linear dichroism is the small difference in absorption of parallel and perpendicularly polarised light, and reveals finer structure than the classical absorption spectrum. A new band observed by this method in the region of 160 nm. in an oriented poly-γ-ethyl-L-glutamate film is attributed to either an n–δ^* or an NV_2 transition.[32] The polarised u.v. spectrum of a poly-L-alanine film shows a band at 165 nm. when cast from a helicogenic solvent.[33] These are perhaps the first observations of a band recently calculated on theoretical grounds to occur near 160 nm. through constructive interference between bands at shorter and longer wavelengths.[34]

[28] G. D. Fasman and J. Potter, *Biochem. Biophys. Res. Comm.*, 1967, **27**, 209.
[29] B. Davidson and G. D. Fasman, *Biochemistry*, 1967, **6**, 1630; see also M. E. Magar, *Biochemistry*, 1968, **7**, 617.
[30] E. S. Pysh, *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **56**, 825.
[31] L. Stevens, R. Townend, S. N. Timasheff, G. D. Fasman, and J. Potter, *Biochemistry*, 1968, **7**, 3717.
[32] J. Brahms, J. Pilet, H. Damany, and V. Chandrasekharan, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 1130.
[33] P. K. Marnii and D. W. Urry, *Macromolecules*, 1968, **1**, 372.
[34] D. G. Barnes and W. Rhodes, *J. Chem. Phys.*, 1968, **48**, 817.

4 Nuclear Magnetic Resonance

In the past two years much attention has been devoted to a correlation of the n.m.r. spectra of polyamino-acids with their secondary structure. Application of n.m.r. to the study of peptide conformation stems from two observations that γ-esters of poly-L-glutamic acid in helix-promoting solvents resemble solids in showing only very broad bands in their n.m.r. spectra.[35, 36] This broadness has been attributed to the energy exchange between protons being so efficient in the helical form as to lead to very short spin–spin relaxation times.[37] A subsequent study of poly-γ-benzyl-L-glutamate in deuteriochloroform showed only one broad band due to the phenyl protons. In non-accumulated spectra the other bands were so broad as to be indistinguishable from the background noise.[38] The order of appearance of the bands on the gradual addition of trifluoroacetic acid showed the 'unfreezing' of the side-chains from the periphery of the helix in to the backbone chain. The appearance of the α-CH and peptide NH bands at a trifluoroacetic acid concentration between 20 and 30% coincided with a sudden change in the optical parameter b_0, indicating a helix-to-random-coil transition.[38]

Efforts to correlate the resonance peaks with helix content have been centred on the peptide NH and α-CH bands as these are the least developed in the helical form and are common to most polyamino-acids. The early suggestion that the helix–coil transition could be followed by measuring the peak area of the amide NH[36] was shown not to be generally useful since poly-L-alanine, for example, showed no change in the amide NH peak area over a solvent composition range where o.r.d. studies show a partial transition to the random coil.[39] The appearance of single peaks of unchanged area when optical techniques indicated a 50% helical content was ascribed at this time to the rapid interconversion of helical and random-coil segments of the same peptide chain, leading to a single time-averaged resonance.[40–42] Subsequent work, however, has established that this apparently single peak is in fact composite, and invalidates this explanation (see below). Because the NH peak broadens more rapidly than the α-CH and is subject to additional rate processes, the latter was suggested to be a more straightforward indicator of conformation.[41] The line-width of several polyamino-acids was observed to pass through a maximum during conformational transition. This was tentatively ascribed to the beginning

[35] F. A. Bovey, G. V. D. Tiers, and G. Filipovich, *J. Polymer Sci.*, 1959, **38**, 73.
[36] M. Goodman and Y. Masuda, *Biopolymers*, 1964, **2**, 107.
[37] D. I. Marlborough and H. N. Rydon, in 'Some Newer Methods in Structural Chemistry,' eds. R. Bonnett and J. C. Davis, United Trade Press, London, 1967, p. 211.
[38] D. I. Marlborough, K. G. Orrell, and H. N. Rydon, *Chem. Comm.*, 1965, 518.
[39] R. E. Glick, L. Mandelkern, and W. E. Stewart, *Biochem. Biophys. Acta*, 1965, **1**, 518; *Biochemistry*, 1967, **6**, 143.
[40] J. L. Markley, D. H. Meadows, and O. Jardetsky, *J. Mol. Biol.*, 1967, **27**, 25.
[41] E. M. Bradbury, C. Crane-Robinson, and H. W. E. Rattle, *Nature*, 1967, **216**, 862.
[42] E. M. Bradbury and H. W. E. Rattle, *Polymer*, 1968, 201.

of separation of two overlapping α-CH resonances differing only slightly in chemical shift.[41]

All the work so far described was carried out at 60 MHz using single run spectra. Subsequent studies at 100 MHz or higher frequency on poly-β-methyl-L-aspartate, poly-L-leucine,[43] and poly-L-methionine[44] in deuteriochloroform–trifluoroacetic acid solutions clearly showed two resonances for both the amide NH and α-CH protons, the peaks only partly overlapping. In the case of poly-L-methionine, at 60 MHz even with accumulated spectra only asymmetry of the amide NH peak was seen. This asymmetry, however, does indicate that the separation of two bands at 100 MHz is due to the increased resolution at the higher frequency and not to the frequency change spanning the helix–coil interconversion rate.[44] The observation of such twin peaks for the α-CH and NH protons corresponding to different chemical shifts for the helix and random coil forms seems to hold for polyamino-acids in general only at 100 MHz, although separations of this type have now been observed at 60 MHz for poly-D-α-amino-n-butyric acid and poly-D-norleucine.[41] The helical content of a peptide with separate peaks for the same proton in random-coil and helical environments can be estimated directly and unambiguously from the relative peak areas, and thereby an absolute method of determination of helix content is provided. Such an analysis has been carried out for poly-γ-benzyl-L-glutamate by a simple symmetry procedure,[45] but for poly-L-methionine a more quantitative procedure employing the curve-fitting of two Lorentzians was necessary.[44]

Although high-resolution n.m.r. studies on specimens of poly-γ-benzyl-L-glutamate of moderate molecular weight (D.P. $\leqslant 92$) show two separate bands in the α-CH region during the helix–coil transition, it has been reported that in a sample of higher molecular weight (D.P. = 640) under the same conditions only one band appears, although it passes through a bandwidth maximum during the conformational transition. This observation is attributed to a faster rate of helix–coil interconversion.[45] Poly-L-methionine of similar chain length (D.P. = 610), however, shows twin resonances, so this effect is not general to high molecular weight polyamino-acids. N.m.r. spectra of poly-γ-benzyl-L-glutamate of D.P. 13 in deuteriochloroform show that b_0 measurements of helix content are valid down to helical lengths involving only eight residues.[45] The helix–coil transition of poly-L-phenylalanine has recently been followed by n.m.r.; its poor solubility in helicogenic solvents has made its optical study difficult.[46] It is generally found in polyamino-acids that the random coil α-CH shifts upfield on helix formation, but the amide NH moves downfield. The latter

[43] J. A. Ferretti, *Chem. Comm.*, 1967, 1030.
[44] J. C. Haylock and H. N. Rydon in 'Peptides', ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 19.
[45] E. M. Bradbury, C. Crane-Robinson, H. Goldman, and H. W. E. Rattle, *Nature*, 1968, **217**, 812.
[46] F. Conti and A. M. Liquori, *J. Mol. Biol.*, 1968, **33**, 953.

movement is ascribed to the formation of strong hydrogen bonds, whereas the former shift is attributed to the magnetic anisotropy of the adjacent peptide bonds.[40, 47] Little work has so far been done on β-structures, but they are known to be disrupted by 1–2% trifluoroacetic acid. Poly-γ-benzyl-DL-glutamate exists in deuteriochloroform largely as the random coil, but the α-CH resonance occurs at τ6·05, the same position as for the helical poly-L isomer in this solvent. The occurrence in the latter of the α-CH at τ5·5 in deuteriochloroform–trifluoroacetic acid mixtures after disruption of the helix is therefore due only to solvation. This coincidence of the chemical shift of the α-CH for the helical and random coil forms does not hold for all pure solvents. In [$^2H_7$]dimethylformamide, for example, the helical α-CH is at τ5·90 and the random coil α-CH at τ5·75.[48] Studies on polypeptides such as poly-L-leucine which retain a partially helical conformation even in solutions of high trifluoroacetic acid content also show that the position of the helical α-CH peak is dependent on the trifluoroacetic acid concentration. N.m.r. studies of the water-soluble polymers poly-α-L-glutamic acid[40, 49] and poly-α-L-lysine[49] in D_2O also show only a very small chemical shift difference between the α-CH peaks for helix and random-coil forms. A prediction that the resonance peak of the α-CH protons of an infinitely long right-handed α-helix should occur at *ca.* 0·5 p.p.m. downfield of the position for a corresponding left-handed helix has proved difficult to test experimentally because of the factors discussed above, but it has been concluded that the chemical shift of the α-CH peak of a D-residue included in a right-handed helix is very similar to that for an L-residue in the same helix.[49]

N.m.r. studies, then, have in general correlated well qualitatively with o.r.d. studies of the helix–coil transition, but unless such studies are carried out at a resolution high enough to separate the helical and random-coil resonances for the α-CH and NH protons it is difficult to interpret peptide spectra in quantitative terms of percentage helix.

The 220 MHz n.m.r. spectrum of polysarcosine (poly-*N*-methylglycine) in [$^2H_6$]dimethyl sulphoxide shows eight peaks in the *N*-methyl region. This spectrum is interpreted as indicating nearly equal amounts of *cis*- and *trans*-peptide units. *N*-Methyl shielding is thought to be sensitive not only to the conformation of the peptide unit to which it is attached, but also to the conformation of the nearest neighbour units, and the eight peaks are therefore assigned to the eight possible different tripeptide conformational sequences. The preferred conformation varies widely with solvent, but the model compound *N*-acetylsarcosine methyl ester is always strongly *trans*.[50] The liquid crystalline nature of concentrated solutions of

[47] H. Sternlicht and D. Wilson, *Biochemistry*, 1967, **6**, 288.
[48] E. M. Bradbury, B. G. Carpenter, C. Crane-Robinson, and H. W. E. Rattle, *Nature*, 1968, **220**, 69.
[49] E. M. Bradbury, C. Crane-Robinson, H. Goldman, and H. W. E. Rattle, *Biopolymers*, 1968, **6**, 851.
[50] F. A. Bovey, J. J. Ryan, and F. P. Hood, *Macromolecules*, 1968, **1**, 305.

poly-γ-benzyl-L-glutamate results in nuclear dipole–dipole and quadrupole splittings of the solvent peaks. This occurs, for example, in the methyl absorption of dimethylformamide.[51] Determination of the alignment of CH_2Cl_2 or CD_2Cl_2 molecules when used as solvent for the same polyamino-acid has been made by studying the ^{35}Cl, ^{2}H, and ^{1}H n.m.r. signals, and it is concluded that all the solvent molecules contribute to the spectrum.[52] It is thought that study of the ^{13}C n.m.r. spectrum of proteins should aid conformational studies since the chemical shifts are an order of magnitude greater and the linewidths much sharper than in the ^{1}H spectra. Practical application, however, is contingent on an inexpensive ^{13}C source for enrichment beyond the present 1% natural abundance.[53]

5 Hydrogen Exchange

The hydrogen atoms attached to oxygen, nitrogen, and sulphur atoms in polypeptides exchange in general quite rapidly with the deuterium in D_2O, whereas the hydrogens attached to the carbon atoms exchange only under extreme conditions. The labile hydrogen atoms of simple peptides exchange with solvent D_2O often in only a few minutes, but native proteins are found to require up to 24 hr. for complete exchange of some of these protons.[6g, 54] Studies on polyamino-acids have correlated this slow exchange of the amide NH protons with the presence of secondary structures in which the intra-molecular hydrogen bonding slows down the rate of exchange.[55] Hydrogen exchange can be followed either by spectroscopic means or by physically separating the solute from solvent after allowing exchange and then measuring one or the other for hydrogen-isotope enrichment. The classical method of separation in the latter case is the freeze-drying technique developed by Linderstrøm-Lang. Some of the steps in this procedure may cause misleading results, and separations using Sephadex columns or rapid dialysis have been developed to ensure solution conditions throughout the experiment.[6g, 54] These separation methods, of course, do not give any information on the origin of the exchanged protons. Spectroscopic methods do indicate the site of exchange; the original observation that proteins possess hydrogen atoms with a range of exchange rates was made using the amide II absorption band in the i.r. spectrum of solutions.[56] However, the i.r. method tends to give less precise results. More recently n.m.r. has been used to follow exchange. It has so far proved more effective in the case of the smaller peptides such as gramicidin S.[57]

[51] M. Panar and W. D. Phillips, *J. Amer. Chem. Soc.*, 1968, **90**, 3880.
[52] D. Gill, M. P. Klein, and G. Katowycz, *J. Amer. Chem. Soc.*, 1968, **90**, 6870.
[53] W. J. Horsley and H. Sternlicht, *J. Amer. Chem. Soc.*, 1968, **90**, 3738.
[54] A. Hvidt and S. O. Nielson, *Adv. Protein Chem.*, 1966, **21**, 287.
[55] A. Elliott and W. A. Hanby, *Nature*, 1958, **182**, 654; E. R. Blout, C. de Lozé, and A. Asadourian, *J. Amer. Chem. Soc.*, 1961, **83**, 1895.
[56] H. Lenormant and E. R. Blout, *Nature*, 1953, **172**, 770.
[57] A. Stern, W. A. Gibbons, and L. C. Craig, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 734.

Many other physical techniques have been used to study secondary structures in peptides, including dilatometric and refractometric studies,[58] viscometry,[59] pressure effects,[60] ultrasonic absorption,[61] birefringence,[62] and temperature jump methods,[63] but space precludes any introduction to these topics.

6 Conformational Probes

The conformation of proteins, peptides, and polyamino-acids can also be investigated either by binding on groups or by making use of the properties of some groups naturally present which are sensitive to their local environment in the molecule through non-covalent interactions. Fluorescent probes have most frequently been used, and these have been briefly reviewed recently.[64] Use of the intrinsic fluorescence of tyrosine or tryptophan residues has the advantage that there is no possibility of modifying the native conformation through the addition of groups. Tryptophan fluorescence has been used to monitor the rate of denaturation of trypsin and some of its derivatives.[65] Tyrosyl residues with ionised hydroxyl groups are efficient quenchers of neutral tyrosyl fluorescence in oligo-L-tyrosines, one anion quenching *ca.* 50 un-ionised residues. Energy transfer is probably 100% efficient.[66] The influence of carboxyl, amino, and peptide groups on the quantum yield of tyrosyl fluorescence has been evaluated in a series of oligopeptides containing glycine and L-tyrosine. Quenching is caused by un-ionised carboxy-groups, charged amino-groups, and neighbouring peptide bonds.[67] Other studies on the phosphorescence of tyrosine oligopeptides[68] and on e.s.r. monitoring of the u.v.-induced triplet state in poly-L-tyrosine[69] have been reported. Measurements of fluorescence polarisation indicate that dioxopiperazines of tyrosine and tryptophan have folded conformations in 95% glycerol solutions.[70] It is suggested that the low yields of tyrosyl and tryptophyl emission in the native state of proteins may be partly accounted for by their interaction with the peptide group.[70] Bovine thyroglobulin contains 30 iodotyrosyl residues. From the change in tryptophan fluorescence upon

[58] *E.g.* H. Noguchi and J. T. Yang, *Biopolymers*, 1963, **1**, 359.
[59a] Reviewed by J. T. Yang, *Adv. Protein Chem.*, 1961, **16**, 323. [b] J. H. Bradbury and M. D. Fern, *J. Mol. Biol.*, 1968, **36**, 231.
[60] *E.g.* J. F. Rifkind and J. Applequist, *J. Amer. Chem. Soc.*, 1964, **86**, 4207.
[61] K. Saksena, B. Michels, and R. Zana, *J. Chim. Phys.*, 1968, **65**, 597.
[62] H. Matsumato, H. Watanabe, and K. Yoshiaka, *Biopolymers*, 1968, **6**, 929.
[63] Reviewed by M. Eigen and L. de Maeyer, in 'Techniques of Organic Chemistry,' ed. A. Weissberger, 2nd edn., vol. 8, Part 2, Wiley, New York, 1963.
[64] G. M. Edelman and W. O. McLure, *Accounts Chem. Res.*, 1968, **1**, 65.
[65] T. R. Hopkins and J. D. Spikes, *Biochem. Biophys. Res. Comm.*, 1968, **30**, 540.
[66] J. A. Knapp and J. W. Longworth, *Biochim. Biophys. Acta*, 1968, **154**, 436.
[67] H. Edelhoch, R. Perlman, and M. Wilchek, *Biochemistry*, 1968, **7**, 3893.
[68] R. F. Steiner, *Biochem. Biophys. Res. Comm.*, 1968, **30**, 502.
[69] J. J. ten Bosch, R. O. Rahn, J. W. Longworth, and R. G. Shulman, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 1003.
[70] H. Edelhoch, R. S. Bernstein, and M. Wilchek, *J. Biol. Chem.*, 1968, **243**, 5985.

ionisation of these iodoamino-acid residues average distances between emitter and recipient residues have been computed.[71] The fluorescence polarisation spectra of simple derivatives of tryptophan and peptides incorporating it have been found to be sensitive to conformation.[72] Similar studies on a copolymer of DL-glutamic acid and L-tryptophan (99 : 1), which is not amenable to o.r.d. measurements, indicate that it consists of runs of D- and runs of L-sequences.[72] The phosphorescence of the tryptophan in trypsin more closely resembles that from tryptophan in polyvinyl alcohol rather than in aqueous solution.[73] The intramolecular energy transfer of phycoerythrin, a strongly fluorescent photosynthetic accessory pigment of red algae which contains several chromophores, has been studied. O.r.d. results on this protein indicate that the phycoerythrobilin residues are probably buried in the molecule as their rotatory power is removed on denaturation.[74]

2-*p*-Toluidinonaphthalene-6-sulphonate and 1-anilinonaphthalene-4- and -8-sulphonates are practically non-fluorescent in water but fluoresce strongly when dissolved in organic solvents or bound to proteins.[64] It is thought that these residues are hydrophobic probes, that is, the fluorescence is responsive to the polarity of the solvent environment. These probes are highly sensitive to protein denaturation as the hydrophobic groups originally in the molecular interior are exposed to the solvent.[64] Although *X*-ray evidence indicates that the gross conformations of chymotrypsinogen and chymotrypsin are similar, the 2-*p*-toluidinonaphthalene-6-sulphonate derivative of the protein shows an increase in fluorescence which parallels the appearance of enzymic activity, indicating the refolding of the polypeptide chain which accompanies the formation of the active site.[75] Conformational changes in glutamate dehydrogenase induced by regulatory reagents have been studied using 1-anilinonaphthalene-8-sulphonate,[76] while use of 1-anilinonaphthalene-7-sulphonate has been suggested in order to avoid the intramolecular interaction possible in the 1,8-isomer.[77] Toluidinonaphthalenesulphonate has been found to be an excellent fluorescent probe for the β-structure of poly-L-lysine. Neither the α-helix nor the random coil causes enhanced fluorescence of the probe. The high polarisation spectrum of the β-form indicates that it is more rigid and has larger overall dimensions, causing slower rotational tumbling in solution, than other secondary structures.[78]

[71] R. L. Perlman, A. van Zyl, and H. Edelhoch, *J. Amer. Chem. Soc.*, 1968, **90**, 2168.
[72] J. Lynn and G. D. Fasman, *Biopolymers*, 1968, **6**, 159.
[73] E. Kuntz, *Nature*, 1968, **217**, 845.
[74] F. D. H. McDowell, T. Bednor, and A. Rosenberg, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 1356; see also J. Pecci and E. Fujimari, *Biochim. Biophys. Acta*, 1968, **154**, 332, and H. H. Krocs, *Biochem. Biophys. Res. Comm.*, 1968, **31**, 877.
[75] W. O. McLure and G. M. Edelman, *Biochemistry*, 1967, **6**, 567.
[76] W. Thompson and K. L. Yielding, *Arch. Biochem. Biophys.*, 1968, **126**, 399.
[77] D. C. Turner and L. Brand, *Biochemistry*, 1968, **7**, 3381.
[78] J. Lynn and G. D. Fasman, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 327.

The electronic excitation spectra of oligomers of L-proline of increasing chain length containing a 1-dimethylaminonaphthalene-5-sulphonyl donor group at one end and an α- or β-naphthylamide acceptor group at the other end have been studied, and a logarithmic plot of the transfer efficiency *versus* the donor–acceptor separation was found to agree well with the predicted theoretical values. This method is suggested as a tool for the estimation of distance between parts of macromolecules.[79] A conformation-sensitive intramolecular charge-transfer complex occurs in chymotrypsin alkylated at methionine-192 by α-bromo-4-nitroacetophenone. The donor is thought to be a tryptophan residue.[80] The phthalimido-group, which can be introduced into peptides at free amino-groups, has been used as a probe to signal the proximity of the phenyl group of phenylalanine. An n.m.r. study of the ring-current shielding occurring in diphthaloyl gramicidin S shows the proximity of the phenylalanine residues and supports the antiparallel pleated sheet conformation.[81]

Difference spectroscopy, which detects small perturbances in the absorbing species of chromophores, has been used to study not only changes in the conformation of proteins but also the number and nature of 'buried' chromophores. The change in environment accompanying disruption of the secondary structure and the central hydrophobic region of the molecule is reflected in small changes in the absorbing properties of a chromophore previously buried in the hydrophobic region. Addition of solvents like methanol and dimethyl sulphoxide to the native protein also causes spectral changes in the tyrosine and tryptophan residues on the outside of the protein, but those buried inside remain unaffected. Use of perturbants of various molecular sizes shows some chromophores to be accessible to small but not larger perturbants, furnishing data on the surface topography of the molecule. In bovine fibrinogen only 33% of both tyrosine and tryptophan residues are in contact with the solvent.[82] The additivity of solvent perturbation difference spectra has been checked and found valid for both tyrosyl and tryptophyl model compounds and proteins,[83] and application to pepsin, rabbit muscle aldolase, and bovine serum albumin has shown the number of buried tyrosine and tryptophan residues in these molecules.[84] The two tryptophyl residues of serum albumin are thought to be located at a polypeptide fold largely accessible to solvent penetration.[84] Solvent perturbation studies on 41-DNP ribonuclease A indicate that the 2,4-dinitrophenyl group is at least partially folded into the protein matrix.[85] The study of difference spectra of apoconalbumin has led to the proposal of an increase in molar absorbance of

[79] G. Gabor, *Biopolymers*, 1968, **6**, 809.
[80] D. S. Sigman and E. R. Blout, *J. Amer. Chem. Soc.*, 1967, **89**, 1747.
[81] R. Schwyzer and U. Ludescher, *Biochemistry*, 1968, **7**, 2514, 2519.
[82] E. Mihalyi, *Biochemistry*, 1968, **7**, 208.
[83] T. T. Herskovits and M. Sorenson, *Biochemistry*, 1968, **7**, 2523.
[84] T. T. Herskovits and M. Sorenson, *Biochemistry*, 1968, **7**, 2533.
[85] M. T. Ettinger and C. H. W. Hirs, *Biochemistry*, 1968, **7**, 3374.

ca. 10^3 at 275 nm. for the transfer of a buried tryptophyl residue from the interior of a protein to an aqueous environment.[86] It has been concluded from a comparison of the c.d. spectrum of poly-L-tyrosine with the spectra of proteins containing known ratios of exposed-to-buried tyrosine residues that the contribution at 280 nm. from buried tyrosine residues is *ca.* 12 times as great as that from exposed ones. This greater molar absorbance is ascribed to the restriction of the freedom of rotation of the side-chain in the buried residue causing the phenol chromophore to experience a large asymmetric field.[27b] Difference u.v. spectral studies indicate that the tyrosine residues of native soluble collagen are not buried, and only small conformational changes take place on denaturation. This accords with the accessibility of the tyrosines to chemical modification and the idea that the tyrosyl groups are in extrahelical appendages.[87]

Difference spectropolarimetry has been investigated as a probe for small conformational changes.[88] An extrinsic Cotton effect arises in complexes of Acridine Orange with poly-L-glutamic acid which corresponds to the absorption band of bound Acridine Orange when the polymer is in the α-helical conformation.[89] The corresponding complex of poly-*S*-carboxymethyl-L-cysteine, in the β-form, has now been studied and found also to exhibit a new Cotton effect, but only when the ordered secondary structure is present.[90]

The reactivity of amino-acid residues to specific chemical reagents can also be used to investigate whether they are normal (fully exposed) or abnormal (buried). Such studies are not covered in this section, but two examples are given. The use of cyanuric fluoride and *N*-acetylimidazole as reactive probes has shown the relative proportions of buried and exposed tyrosine residues in pepsinogen, soybean trypsin inhibitor, and ovomucoid. Different degrees of reactivity even of exposed groups were found.[91] Degradation of iodinated myoglobin indicates that the side-chain of tyrosine-146 is buried in the interior of the molecule since it remains free of iodine.[92]

7 Theoretical Approaches to Peptide Conformations

In the past few years computations have been made in an attempt to predict the conformation of a protein by starting simply with the amino-acid sequence.[93] If the computed conformation is found to agree with that

[86] A. T. Tan and R. C. Woodworth, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 739.
[87] R. J. Doyle and J. Bello, *Biochem. Biophys. Res. Comm.*, 1968, **31**, 869.
[88] B. J. Adkins and J. T. Yang, *Biochemistry*, 1968, **7**, 266.
[89] L. Stryer and E. R. Blout, *J. Amer. Chem. Soc.*, 1961, **83**, 1411.
[90] S. Makino, N. Murai, and S. Sugai, *J. Polymer Sci., Part B, Polymer Letters*, 1968, **7**, 477.
[91] M. J. Gorbunoff, *Biochemistry*, 1968, **7**, 2547.
[92] J. Hermans jun. and L. W. Lu, *Arch. Biochem. Biophys.*, 1968, **122**, 331.
[93] See the reviews: [a] H. A. Scheraga, *Adv. Phys. Org. Chem.*, 1968, **6**, 103; [b] G. N. Ramachandran and V. Sasisekharan, *Adv. Protein Chem.*, 1968, **23**, 283.

obtained experimentally in several cases, it is hoped that this concordance will lend credence to future predictions of conformation, especially of proteins not yet studied, and perhaps not amenable to study, by *X*-ray diffraction. A more limited objective is the refinement of 6 Å resolution *X*-ray data down to atomic resolution.[6h, 93a] Computations are being carried out on the assumption that the native protein has a narrow distribution of conformations about an equilibrium one, and that this distribution has the highest statistical weight. Calculations are usually based on the total system protein plus solvent water. The reversibility of denaturation and disulphide-bond regeneration are thought to support the idea of an equilibrium state in the native protein.

This approach, then, requires first obtaining an expression for the energy of the system as a function of the co-ordinates of the atoms of the system. Secondly, with the aid of a computer, the energy must be minimised, and finally the nature of the energy surface in the neighbourhood of the minimum must be explored to obtain the statistical weight.[93] Since the amide bond normally behaves as a rigid *trans*-unit, conformation is determined by the dihedral angles around the N—C^{α} (ϕ) and the C^{α}—CO (ψ) bonds. In some cases angular twisting from the plane of the amide bond is also considered (ω). As intramolecular contributions to the total energy, many factors have to be taken into consideration, including torsional barriers to rotation about single bonds, non-bonded interactions, electrostatic interactions, hydrogen bonding, bond stretching, bond-angle bending, and torsion about the peptide bond. Intermolecular solvent–polypeptide interactions such as hydrophobic bonding must also be taken into account. Expressions for these energies have been deduced from various types of physicochemical data on model systems of low molecular weight. In the last two years the expressions for these contributions have been refined by using them to compute known crystal structures of small molecules.[93] For conformational energy calculations on structures involving loops an empirical loop-closing potential has been developed.[94, 95]

Application to polyamino-acids gives results in general agreement with experimental observations. Data are normally presented in the form of maps of ϕ *vs.* ψ which display the allowed regions and outer limits for the polypeptide conformation [93] (see Figure 3). The dipole–dipole interaction of the ester group of poly-β-methyl-L-aspartate with the amide group of the backbone is the cause of its difference in screw sense to the majority of polyamino-acids; the extra methylene group of the corresponding glutamate polymer is sufficient to change the orientation and distance between the amide and ester dipoles.[96] It has been calculated that *para*-substitution of the nitro-, cyano-, chloro-, and methyl groups into the benzyl group of

[94] R. A. Scott, G. Vanderkooi, R. W. Tuttle, P. M. Shames, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 2204.
[95] K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 1317.
[96] T. Ooi, R. A. Scott, G. Vanderkooi, and H. A. Scheraga, *J. Chem. Phys.*, 1967, **46**, 4410.

poly-β-benzyl-L-aspartate, which forms a left-handed α-helix, should give rise to right-handed helices. This is in agreement with experimental results.[97]

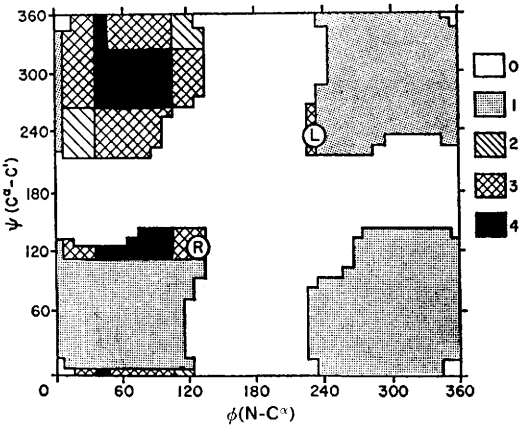


Figure 3 *The effect of variation of the side-chain on the allowed areas of the steric map for various dipeptides. In area* 0 *no conformations are allowed. Conformations in areas* 1 *to* 4 *are allowed for glycylglycine, in areas* 2 *to* 4 *for glycyl-*L*-alanine, in areas* 3 *to* 4 *for higher straight-chain homologues, while only area* 4 *is allowed for glycyl-*L*-valine and glycyl-*L*-isoleucine. The circles marked* R *and* L *indicate the location of the standard right- and left-handed* α*-helices*
(Reproduced by permission from *Adv. Phys. Org. Chem.*, 1968, **6**, 103)

Similar concordance with experiment is found for calculations on poly-L-serine [98] and poly-L-tyrosine.[94] Preferred conformations have also been calculated for poly-L-proline [99] and poly-*N*-methyl-L-alanine. The results on the latter indicate that the preferred conformation may be a right-handed threefold helix.[100] Theoretically derived conformations for gramicidin S [101, 102] and oxytocin [103] are considered in the section of this Report on cyclic peptides. Application to lysozyme [104] and ribonuclease S-peptide has been reported; in the latter case results suggest that when the restrictions of homogeneity and regularity, as found in polyamino-acids, are removed, the peptide can find irregular minimal energy conformations close to the regular one of a polyamino-acid.[105]

[97] J. Yan, G. Vanderkooi, and H. A. Scheraga, *J. Chem. Phys.*, 1968, **49**, 2713.
[98] T. M. Birshstein and O. B. Ptitsyn, *Biopolymers*, 1967, **5**, 785; K. P. Sarathy and G. N. Ramachandran, *ibid.*, 1968, **6**, 461.
[99] P. R. Schimmel and P. J. Flory, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 52.
[100] J. E. Mark and M. Goodman, *J. Amer. Chem. Soc.*, 1967, **89**, 1267; A. M. Liquori and P. De Santis, *Biopolymers*, 1967, **5**, 815.
[101] A. M. Liquori, P. De Santis, A. L. Kovacs, and L. Mozzarella, *Nature*, 1966, **211**, 1039.
[102] G. Vanderkooi, S. J. Leach, G. Némethy, and H. A. Scheraga, *Biochemistry*, 1966, **5**, 2991.
[103] K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 1317.
[104] D. A. Brant and P. R. Schimmel, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 428.
[105] K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 420.

The deviation from planarity of *trans*-amide bonds has been discussed.[106] The energy difference between a *cis*- and *trans*-peptide unit is calculated at >2 kcal./mole/residue, but for imino-residues the *cis*- and *trans*-forms differ only slightly in energy. The observed conformation of poly-proline I lies in a small fully allowed region on the conformational map.[107] A structure for the collagen triple helix having interchain hydrogen bonds *via* bound water molecules has been proposed on theoretical grounds, and explains more satisfactorily some of the properties of collagen.[108] A detailed study has been made of a single-chain triple helical structure for poly (Gly·Pro·Pro) following the conclusion that the subunit of *Ascaris* collagen is a single polypeptide chain which folds back upon itself to form a stable collagen-type triple helix.[109] In a system of three linked peptide units there are three types of conformation which contain NH···OC hydrogen bonding between the first and third units. Two are helical and both reverse the direction of chain progress. Such a conformation has been found in some peptides, and could occur as the turn in a cross-β structure.[110]

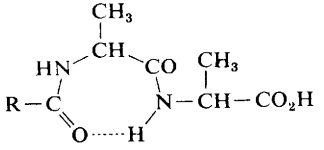
Calculations of the dependence of the conformational energy and the rotational strength of the amide n–π^* electronic transition of polyamino-acids on conformation have been reported.[111] The concept of partial molar rotatory powers has been used to calculate the rotational strength of the n–π^* transition of the helix of poly-L-alanine [112] and the antiparallel pleated sheet of poly-L-lysine.[113] The latter gives good agreement with reported data.

8 The Conformations of Small Peptides

X-Ray crystal structure studies of di- and tri-peptides have recently been reviewed.[114] Although peptides containing fewer than five or six residues cannot form α-helices, it is probable that they do not have random conformations in solution. I.r. spectra show that intramolecular hydrogen bonding of protected tetrapeptides occurs in chloroform solution and Boc·L-Ala·L-Val·L-Ala·L-Val·OMe (where Boc = t-butoxycarbonyl) shows evidence of β-structure. Di- and tri-peptides, on the other hand, show no evidence of any hydrogen bonding.[115] *N*-Acylalanylalanine methyl esters exist in dilute solutions in carbon tetrachloride as equilibrium

[106] G. N. Ramachandran, *Biopolymers*, 1968, **6**, 1494.
[107] G. N. Ramachandran and C. M. Venkatachalan, *Biopolymers*, 1968, **6**, 1255; see also J. Applequist, *ibid.*, 1968, **6**, 117.
[108] G. N. Ramachandran and R. Chandrasekharan, *Biopolymers*, 1968, **6**, 1649.
[109] G. N. Ramachandran, B. B. Doyle, and E. R. Blout, *Biopolymers*, 1968, **6**, 1771.
[110] C. M. Venkatachalan, *Biopolymers*, 1968, **6**, 1425.
[111] J. N. Vournakis, J. F. Yan, and H. A. Scheraga, *Biopolymers*, 1968, **6**, 1531.
[112] D. W. Urry, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 1114.
[113] D. W. Urry, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 394.
[114] R. E. Marsh and J. Donohue, *Adv. Protein Chem.*, 1967, **22**, 235.
[115] J. E. Shields, S. T. McDowell, J. Parker, and G. R. Gray, *J. Amer. Chem. Soc.*, 1968, **90**, 3549.

mixtures of the random coil and a conformation with an intramolecular hydrogen bond between the amide NH and the acyl carbonyl oxygen (1),



(1) R = alkyl group

the latter forming 60–80% of the total.[116] I.r. evidence indicates that the pentamer, hexamer, and nonamer of t-amyloxycarbonyl oligo-L-prolines are similar to poly-L-proline II in possessing a left-handed threefold screw-axis helical structure, but this of course is a special conformation containing no hydrogen bonds.[117] O.r.d. studies on all the diastereoisomers of tri- and tetra-leucines show that, with respect to the peptide chain and its immediate substituents, the conformations adopted in solution closely resemble those of the corresponding alanine stereoisomers.[118] Dielectric-constant measurements on oligopeptides in aqueous solution show that only a small proportion of the total possible conformations make major contributions to the observed average.[119] In the case of oligopeptides of both alanine and serine a preference for conformations intermediate between the extended chain and the pleated sheet β-structure is indicated; here the difference in side-chain seems unimportant.[120] Theoretical calculations of the mean square dipole moments of 14 diastereoisomeric oligomers of alanine [121] agree with the values previously obtained experimentally in aqueous solution.[120] The differences in charge separation of LL and LD dipeptide diastereoisomers also indicate a more compact conformation for the LD isomer.[119] The shift of the leucine side-chain resonance peaks to higher field in the n.m.r. spectrum of D-leucyl-L-tyrosine by the proximity of the aromatic ring of the tyrosyl residue shows also a more compact molecule than in the LL isomer.[122] The pK_a values of the DL isomer are lower than the LL isomer, again indicating a smaller charge separation in the DL form.[123] N.m.r. studies on glycine oligopeptides containing one L-leucine residue [124] have not revealed any correlation with optical effects

[116] S. L. Portnova, V. F. Bystrov, V. I. Tsetlin, V. T. Ivanov, and Yu. A. Ovchinnikov, *Zhur. obshchei Khim.*, 1968, **38**, 428.
[117] H. Okabayashi, T. Isemura, and S. Sakakibara, *Biopolymers*, 1968, **6**, 307, and 323.
[118] D. R. Dunstan and P. M. Scopes, *J. Chem. Soc.* (*C*), 1968, 1585.
[119] P. M. Hardy, G. W. Kenner, and R. C. Sheppard, *Tetrahedron*, 1963, **19**, 95.
[120] J. Beacham, V. T. Ivanov, G. W. Kenner, and R. C. Sheppard, *Chem. Comm.*, 1965, 386.
[121] P. J. Flory and P. R. Schimmel, *J. Amer. Chem. Soc.*, 1967, **89**, 6807.
[122] F. A. Bovey and G. V. D. Tiers, *J. Amer. Chem. Soc.*, 1959, **81**, 2870.
[123] E. Ellenbogen, *J. Amer. Chem. Soc.*, 1956, **78**, 369; N. C. Li, G. W. Miller, N. Solany, and B. T. Gillis, *ibid.*, 1960, **82**, 3737.
[124] A. F. Beecham and N. S. Ham, *Tetrahedron*, 1968, **24**, 2773.

which might result from conformational factors operating on the chromophore adjacent to the asymmetric carbon atom.[125] The observed chemical shift non-equivalence of the glycine methylene protons in many dipeptides containing glycine has been ascribed to the existence of preferred rotamers,[126] but temperature-dependence studies argue against this explanation.[127] A more likely cause is incomplete averaging of the electric-field gradients in the presence of free rotation. The observation of non-equivalence is then dependent upon an asymmetric centre.[128] It is suggested from n.m.r. studies on diastereoisomers of phenylalanylvaline that the conformations are determined primarily by non-bonded interactions and secondarily by some weighting of those conformations that bring the opposite charges of the dipolar form somewhat nearer.[128, 129]

9 Cyclic Peptides

Evidence for the existence of preferred conformations in oligopeptides comes also from differences in the ease of cyclisation of diastereoisomers.[130] DL Isomers of pentapeptides containing two leucyl residues give appreciably higher yields of cyclic products than the corresponding LL isomers.[119] Whilst dipeptide esters cyclise readily to dioxopiperazines, attempts to form cyclo-tripeptides have, with one exception, always yielded cyclohexapeptides. The only cyclo-tripeptide so far authenticated is cyclotriproline,[131] which constitutes a special case because of the lack of NH bonds. The amide bonds of cyclo-tripeptides are constrained to the *cis*-form, but models of cyclo-tetraglycine can be made with four *trans*-amide bonds.[132] Only a few cyclic tetrapeptides have been synthesised. Morozova and co-workers have obtained cyclic monomers from the action of ethoxyacetylene on H·Gly_3·Phe·OH,[133] H·Gly_3·Leu·OH,[134] and H·Gly·Leu·Gly·Leu·OH.[135] Cyclo-tetraglycine was the first cyclic tetrapeptide prepared.[132] Cyclo-tetrapeptides have also been prepared using a solid phase technique by attaching the peptide to the resin through an aromatic residue which acts as an active ester. This effectively prevents

[125] A. F. Beecham, *Tetrahedron*, 1967, **23**, 4481.
[126] A. Nakamura and O. Jardetsky, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 2212; J. W. Westley and B. Weinstein, *Chem. Comm.*, 1967, 1232.
[127] A. Nakamura and O. Jardetsky, *Biochemistry*, 1968, **7**, 1226.
[128] V. J. Morlino and R. B. Martin, *J. Amer. Chem. Soc.*, 1967, **89**, 3107.
[129] V. J. Morlino and R. B. Martin, *J. Phys. Chem.*, 1968, **72**, 2661.
[130] N. A. Poddubnaya and A. M. El-Haggar, *Zhur. obshchei Khim.*, 1968, **38**, 450, 732.
[131] M. Rothe and K.-D. Steffen, *Angew. Chem.*, 1965, **77**, 347.
[132] R. Schwyzer, B. Iselin, W. Rittel, and P. Sieber, *Helv. Chim. Acta*, 1956, **34**, 872.
[133] E. A. Morozova, L. V. Ionova, and N. A. Drobinskaya, *Zhur. obshchei Khim.*, 1964, **34**, 3888.
[134] E. A. Morozova, E. S. Oksenoit, and L. V. Ionova, *Vestnik Moskov. Univ., Ser. II, Khim.*, 1966, **21**, 102.
[135] E. A. Morozova and S. M. Zhenodarova, *Vestnik Moskov. Univ., Ser. II, Khim.*, 1965, **20**, 77.

reactions other than intramolecular cyclisation when the *N*-terminal amino-group is liberated.[136]

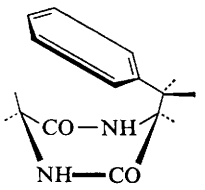
Cyclic pentapeptides are readily prepared by conventional high dilution cyclisation, but in some cases cyclodimerisation occurs. The tendency of linear tripeptides to cyclodimerise is not paralleled to the same degree in the pentapeptides. On the basis of the ready cyclodimerisation of tripeptides in general and the linear pentapeptide sequence of gramicidin S, Schwyzer proposed an antiparallel association of two molecules of peptide spatially suitable for a double condensation; this would naturally lead to peptides with an odd number of amino-acid residues being especially prone to doubling on cyclisation.[137] Some rather simple pentapeptides, however, containing several glycine residues have shown no tendency to doubling,[119, 138] while glycyl-L-prolyl-glycine [139, 140] gives the cyclic hexapeptide as easily as diglycyl-L-proline [139] although the prolyl unit carried no NH group with which to associate by hydrogen bondings. A recent study on the cyclisation of pentapeptide sequences to gramicidin S has shown that the ratio of cyclic dimer to monomer is very sensitive to the nature of the terminal amino-acids. The presence of an *N*-terminal glycine residue causes formation of only cyclic monomer. As the bulk of the *N*-terminal amino-acid side-chain is increased through alanine and leucine to valine, so the yield of cyclic monomer falls to 32%. *C*-Terminal glycine also seems to favour cyclic monomer.[141] That the configuration of the residues can affect dimerisation is shown by the fact that 'all-L' gramicidin S (that is, containing two L instead of D-phenylalanine residues) cannot be prepared by the doubling of the corresponding pentapeptide.[137] The frequent occurrence of D-amino-acids in natural cyclic peptides and depsipeptides and the higher synthetic yields in the cyclisation of oligopeptides with mixed configurations has been linked with a conformation for cyclic peptides derived theoretically.[142]

Hexapeptides, with one exception,[143] appear to always form cyclic monomers. The cyclodimerisation of tripeptides has been shown in several cases to be stereospecific;[144, 145] when H·Gly·DL-Ala·DL·Phe·OMe is treated with methanolic ammonia, for example, the pure *meso*-compound

[136] M. Fridkin, A. Patchornik, and E. Katchalski, *J. Amer. Chem. Soc.*, 1965, **87**, 4646.
[137] R. Schwyzer, 'Ciba Foundation Symposium on Amino-acids and Peptides with Antimetabolic Activity,' Churchill, London, 1958, p. 171; R. Schwyzer and P. Sieber, *Helv. Chim. Acta.*, 1958, **41**, 2186, 2190, 2199.
[138] G. W. Kenner and A. H. Laird, *Chem. Comm.*, 1965, 305.
[139] M. Rothe, K.-D. Steffen, and I. Rothe, *Angew. Chem. Internat. Edn.*, 1964, **6**, 64.
[140] R. Schwyzer, J. P. Carrión, B. Gorup, H. Nolting, and A. Tun-Kyi, *Helv. Chim. Acta*, 1964, **47**, 441.
[141] M. Waki and N. Izumiya, *J. Amer. Chem. Soc.*, 1967, **89**, 1278.
[142] J. Dale, *Angew. Chem. Internat. Edn.*, 1966, **5**, 1000.
[143] A. T. Moore and H. N. Rydon, *Acta Chim. Acad. Sci. Hung.*, 1965, **44**, 103.
[144] R. Schwyzer and A. Tun-Kyi, *Helv. Chim. Acta*, 1962, **45**, 859.
[145] Y. Chen-su, K. Bláha, and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1964, **29**, 2633; H. Brockmann and M. Springorum, *Tetrahedron Letters*, 1965, 837; H. Brockmann and K. Zelleroff, *ibid.*, p. 2291.

cyclo-(Gly·D·Ala·L-Phe·Gly·L-Ala·D-Phe) crystallises out.[144] Although in cyclo-hexapeptides the amide bonds are normally all assumed to be *trans*, i.r. evidence indicates that cyclo-(Gly·D-Phe·L-Leu·Gly·L-Phe·L-Leu) contains *cis*- as well as *trans*-amide bonds. The corresponding 'all-L' and DLDL diastereoisomers, however, seem to have all *trans*-bonds.[146]

An n.m.r. study of cyclo-(Gly·Phe) shows that in trifluoroacetic acid the phenyl group is positioned above the dioxopiperazine ring, which is itself buckled into a boat form so that the phenyl group occupies a 'flagpole' type orientation[147a] (2). Similar conformations have been deduced for cyclo-(Gly·Trp)[147a] and a series of dioxopiperazines containing one tyrosine



(2)

residue.[148] Experiments on *N*-methylated dioxopiperazines, which are more soluble in less polar solvents than are unsubstituted dioxopiperazines, show that this attractive interaction between an aromatic residue and the amide bonds persists in solvents of lower dielectric constant.[147a] The o.r.d. and c.d. spectra of some dioxopiperazines containing phenylalanine are also best interpreted on the basis of a non-planar boat-type conformation with the aromatic substituent above the ring. Cyclic hexapeptides containing two phenylalanine residues separated by two other amino-acids show a much stronger Cotton effect when both the aromatic residues are of the same configuration. Models show that only in these isomers can the aromatic rings approach one another closely enough for the aryl chromophores to interact.[147b] The molecular ellipticities of the phenol and indole chromophores of dioxopiperazines containing tyrosine and tryptophan residues are several times as large as those of the equivalent linear dipeptide forms. This enhancement is also ascribed to the rigid conformation of the ring systems.[148] O.r.d. studies on alanine dioxopiperazines have led to the conclusion that multiple Cotton effects and exciton behaviour is not unique to helical polypeptides as they arise in these simpler systems when the proper geometric disposition of identical chromophores is realised.[149] The similarity of the c.d. pattern of L-5-methylpyrrolid-2-one in cyclohexane

[146] K. Bláha, J. Smolikova, and A. Vitek, *Coll. Czech. Chem. Comm.*, 1966, **31**, 4296.

[147] 'Peptides', ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968: [a] G. Gawne, G. W. Kenner, N. H. Rogers, R. C. Sheppard, and K. Titlestad, p. 28; [b] K. Bláha and I. Frič, p. 40; [c] R. Walter, W. Gordon, I. L. Schwartz, F. Quadrifoglio, and D. W. Urry, p. 50.

[148] H. Edelhoch, R. E. Lippoldt, and M. Wilchek, *J. Biol. Chem.*, 1968, **243**, 4799.

[149] D. Balasubramanian and D. B. Wetlaufer, 'Conformations of Biopolymers,' Vol. I, Academic Press, New York, 1967, p. 147; *cf.* also J. A. Schellman and E. B. Nielson, *ibid.*, p. 109.

to that of helical polypeptides is also striking, the extrema differing only by a few nm.[150] In the crystal of cyclo-hexaglycine hemihydrate *X*-ray diffraction shows a pseudo-cell containing two different centrosymmetrical molecules. One has two internal transannular hydrogen bonds, while the other is internally hydrogen bonded to water.[151]

The use of cyclic peptides as models for enzyme active sites and as substrates for enzymes is of conformational interest. The size at which a cyclic peptide is unable to assume the shape of the transition complex with an enzyme seems to be about five or six amino-acid residues. The steric requirements of different enzymes may of course differ. Cyclo-(Gly$_5$-L-Lys)[152] is attacked by trypsin, but cyclo-(Gly·L·Lys·Gly·L-Lys·Gly)[138] is not. The naturally occurring heptapeptide evolidine, although containing a phenylalanine residue, is not cleaved by chymotrypsin.[153] Cyclic peptides are attractive as enzyme models because their greater rigidity compared to linear peptides allows attempts at tailor-making molecules with the side-chains in the right positions for co-operative interaction. This presupposes an accurate knowledge of the conformation of the framework. Cyclic pentapeptides containing tyrosine and histidine residues separated by two or three glycine residues undergo phenolic *O*-acylation in the presence of *p*-nitrophenyl acetate when the residues are both L, but not if the tyrosine has the D-configuration. This confirms the intramolecular nature of the reaction catalysing the ester hydrolysis.[154]

Doubling reactions have also been found to occur in the cyclisation of depsipeptides. The dipeptide analogue β-(*O*-t-butylseryloxy)propionic acid cyclodimerises, and in the case of the DL isomer both the *meso*-form and a racemic mixture of the DD and LL forms are produced.[155] Cyclisation of seven sterically heterogeneous isomers of *N*-methylvalyl-α-hydroxyisovaleryl-*N*-methylvalyl-α-hydroxyisovaleric acid gave cyclic monomers in 40–75% yield, but the homogeneous isomer gave only 8% accompanied by 13% of the doubled product.[156a] *N*-Methylated tri- and tetra-depsipeptides also cyclodimerise, the latter even tripling to some extent.[156b]

The conformation of the cyclic decapeptide gramicidin S continues to attract attention. In the crystal, *X*-ray diffraction has indicated an all-*trans*-amide antiparallel pleated-sheet structure with four transannular hydrogen bonds,[157] although the less favoured proposal of a mixed α and β structure[157] has received some support.[158] A conformation has been

[150] D. W. Urry, *J. Phys. Chem.*, 1968, **72**, 3035.
[151] I. J. Karle and J. Karle, *Acta Cryst.*, 1963, **16**, 969.
[152] M. Ohno and N. Izumiya, *Bull. Chem. Soc. Japan*, 1965, **38**, 1831.
[153] R. O. Studer and W. Lergier, *Helv. Chim. Acta.*, 1965, **48**, 460.
[154] K. D. Kopple, R. R. Jarabak, and P. L. Bhatia, *Biochemistry*, 1963, **2**, 958.
[155] C. H. Hassall and J. O. Thomas, *J. Chem. Soc.* (*C*), 1968, 1945.
[156] [a] Yu. A. Ovchinnikov, V. T. Ivanov, A. A. Kiryushin, and M. M. Shemyakin, *Doklady Akad. Nauk S.S.S.R.*, 1963, **153**, 122; [b] *Ibid.*, p. 1342.
[157] D. C. Hodgkin and B. C. Oughton, *Biochem. J.*, 1967, **65**, 752; also proposed by R. Schwyzer, ref. 136.
[158] D. Balasubramanian, *J. Amer. Chem. Soc.*, 1967, **89**, 5445.

proposed on theoretical grounds which involves two antiparallel right-handed 'α-helical' tripeptide segments connected by two 'non-helical' dipeptide segments,[101] and spectroscopic data have been interpreted as supporting this.[159] A second model proposed on theoretical grounds involves an antiparallel pleated β-sheet with only two hydrogen bonds.[102] More recent theoretical calculations indicate that the lowest energy conformation is not hydrogen-bonded; the low energy arises from favourable non-bonding and electrostatic energy contributions. Such a conformation, however, is probably not favoured in solution. The authors emphasise the tentative nature of these conclusions in view of the known possible sources of error in such calculations.[160] A recent very detailed n.m.r. study of gramicidin S has led to the proposal of a conformation similar to that put forward above on *X*-ray evidence.[57] The pattern of hydrogen–deuterium exchange of the amide NH protons in this work did not agree with previous findings.[159]

The importance of peptide hormones such as insulin and oxytocin which contain loops closed by disulphide bridges has led to much work on their conformations. O.r.d. and thin-film dialysis studies of angiotensin II indicate a remarkable increase in the size of the molecule coincident with the ionisation of the tyrosine phenolic hydroxyl group.[161] A c.d. study of oxytocin and four of its analogues shows a positive band at 250 nm. which is attributed to the disulphide group and corresponds to a sulphur dihedral angle close to 90°.[162] This suggests a right-handed helical screw sense for the disulphide bridge. On warming or adding dioxan a band appears in the region 225–240 nm., which is thought to be related to conformation and indicate distortion from the 90° sulphur dihedral angle.[147, 163c] Theoretical calculations on the conformations of heterodetic peptides such as oxytocin show little change in structure even if the hydration free energy is omitted. It is pointed out that in proteins, where one can identify an 'inside' and an 'outside', one would expect a greater influence of hydration on the conformation, leading to the observed preponderance in aqueous solution of non-polar groups inside and polar groups on the surface of the molecule.[103] Experiments on the oxidation of model peptides containing two cysteine residues to give cyclic disulphides show that when four or more glycine residues separate the cysteines the cyclic monomers are formed almost exclusively. The lower members of the series gave mixtures of cyclic monomer and antiparallel cyclic dimer.[164]

[159] A. M. Liquori and F. Conti, *Nature*, 1968, **217**, 635; *cf.* also F. Quadrifoglio and D. W. Urry, *Biochem. Biophys. Res. Comm.*, 1967, **29**, 785.

[160] R. A. Scott, G. Vanderkooi, R. W. Tuttle, P. M. Shames, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 2204.

[161] M. T. Franze de Fernández, A. E. Delius, and A. C. Paladini, *Biochim. Biophys. Acta*, 1968, **154**, 223.

[162] S. Beychok and E. Brewster, *J. Biol. Chem.*, 1968, **243**, 151.

[163] D. W. Urry, F. Quadrifoglio, R. Walter, and L. Schwartz, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 967.

[164] H. N. Rydon, in 'Ciba Foundation Symposium on Amino-acids and Peptides with Antimetabolic Activity,' 1958, p. 192.

10 Polyamino-acids: Studies in Non-aqueous Solution

Synthetic polyamino-acids offer a unique system for studying the contributions of side-chain interactions to the stabilisation of the α-helix. Studies in non-aqueous solvents eliminate competitive hydrogen bonding from water. Many polyamino-acids are also water-insoluble.[6i] The stability of peptide secondary structures was initially attributed solely to the inter-amide hydrogen bonding, but the formation of a helix by an acyclic analogue of poly-L-proline, poly-*N*-methyl-L-alanine,[165] without the assistance of hydrogen bonds has demonstrated the importance of side-chain interactions in determining conformation. As well as hydrogen bonding, hydrophobic bonds, van der Waals forces, and electrostatic effects are now considered important.[6i] Kauzmann introduced the term 'hydrophobic bond' to describe the attractive forces responsible for the tendency of non-polar groups to avoid contact with the aqueous phase and to adhere to one another in the form of an intramolecular micelle.[166] On this basis the most stable conformations of proteins are those in which the non-polar groups are in contact with each other away from the aqueous phase. Klotz has expressed this interaction in terms of the non-polar groups having crystalline hydrates.[167] The driving force for the transfer of the non-polar side-chains to the interior of the protein molecule then arises from the gain in entropy when these side-chains leave the aqueous phase and it becomes disordered. It has recently been claimed that the free energy change for dissociating an NH···OC bond to NH (aqueous solution) and CO (aqueous solution) is the same whether the initial amide hydrogen bond is in an aqueous or apolar solvent.[168]

Theoretical calculations of the α-helical structure of poly-L-alanine indicate that in simple polymers of this type van der Waals interaction may well control the conformation.[169] The lack of any α-helix formation by some polyamino-acids has led to the recognition of two criteria for preferential α-helix formation. These require no branching at the β-carbon atom of the amino-acid, as occurs in valine and isoleucine, and the absence of a heteroatom attached to the β-carbon atom, as occurs in cysteine and serine.[170] On this basis poly-*O*-acetyl-L-threonine and its *allo*-isomer, which have an oxygen atom attached to the β-carbon, would be expected to prefer the β-conformation, and recent experiments indicate that cross-β-structures are in fact formed.[171] The importance of the different side-chain interactions is best assessed from the relative stabilities of the α-helices of

[165] M. Goodman and M. Fried, *J. Amer. Chem. Soc.*, 1967, **89**, 1264; J. E. Mark and M. Goodman, *ibid.*, p. 1267.
[166] W. Kauzmann, *Adv. Protein. Chem.*, 1959, **15**, 1.
[167] I. M. Klotz, *Fed. Proc.*, 1965, **24**, 525.
[168] I. M. Klotz and S. B. Farnham, *Biochemistry*, 1968, **7**, 3879.
[169] P. De Santis, E. Criglio, A. M. Liquori, and A. Rapamonti, *Nature*, 1965, **206**, 456.
[170] E. R. Blout, in 'Polyamino-acids, Polypeptides, and Proteins,' ed. M. A. Stahmann, University of Wisconsin Press, Madison, 1962, p. 275.
[171] S. Kubata, S. Sugai, and T. Noguchi, *Biopolymers*, 1968, **6**, 1311.

polyamino-acids towards disruption by strong hydrogen-bonding solvents, usually dichloroacetic and trifluoroacetic acids.[6i] These transitions are followed by the spectroscopic methods outlined earlier in this Chapter. A wide spectrum of helix stabilities has been found; homopolymers which form strong helices, such as poly-L-alanine and poly-L-leucine, are still helical in pure dichloroacetic acid, and require high concentrations of trifluoroacetic acid to break them. Weak helices like poly-β-benzyl-L-aspartate are disrupted by 10% dichloroacetic acid. Ester side-chains do not necessarily lead to weak helices; γ-esters of poly-L-glutamic acid form quite strong helices.[6i] The high strength of the helices of those polyamino-acids which contain hydrocarbon side-chains is thought to be due to the hydrophobic bonds, which have a positive enthalpy but negative entropy of formation.

Polyamino-acid studies have shown the importance of the chain length in determining the secondary structure adopted. In poly-γ-benzyl-L-glutamate the degree of polymerisation for onset of helix formation is *ca.* seven residues, but extensive α-helix formation does not occur until high molecular weights; samples of medium molecular weight are mixtures of α- and β-structures. In this polyamino-acid the two conformations can be separated, only the β-form being soluble in formic acid.[172] It seems that 4–6 residues only are necessary for oligo-L-prolines to form a helical structure, but this of course is not an α-helix.[173] A prolyl residue has no NH group and hence cannot form an intra-chain hydrogen bond. The angle ϕ about the $N-C_\alpha$ bond is fixed, and for steric reasons it cannot form a fully extended chain, but is constrained to fold into a left-handed helix. A second form of poly-L-proline is known which spontaneously mutarotates in solution, and this is thought to be caused by a transition from *cis*- to *trans*-amide groups. The *trans*-form of this threefold screw-axis helix is not unique to poly-L-proline; it is also found in polyglycine.[6j] The non-α-helical nature of polyglycine has been ascribed to the unfavourable energy change on going to the α-helix due to the greater conformational freedom of the random coil form because of the absence of a side-chain.[174] The i.r. spectrum of polyglycine II at $-170°$ shows changes in frequency of the C—H stretching absorption which are attributed to the presence of CH···OC hydrogen bonds.[175] The effect of pressure on the *cis–trans* equilibrium of poly-L-proline has been investigated. The extremely slow approach to equilibrium enables the determination of the state at high pressure by measurements of optical rotation at atmospheric pressure after removal from the pressure vessel. The midpoint of transition from the *trans*- to the more compact *cis*-helix occurs at 2975 atmospheres.[176] The

[172] E. R. Blout and A. Asadourian, *J. Amer. Chem. Soc.*, 1956, **78**, 955; J. C. Mitchell, A. E. Woodward, and P. Doty, *ibid.*, 1957, **79**, 3955.
[173] A. Yaron and A. Berger, *Biochim. Biophys. Acta*, 1965, **107**, 307.
[174] G. Némethy, S. J. Leach, and H. A. Scheraga, *J. Phys. Chem.*, 1966, **70**, 998.
[175] S. Krimm and K. Kuroiwa, *Biopolymers*, 1968, **6**, 401.
[176] J. M. Rifkind and J. Applequist, *J. Amer. Chem. Soc.*, 1968, **90**, 3650.

c.d. spectrum of poly-L-proline in an unordered conformation has been reported.[177]

Transitions from the α-helix to the β-form in films of the water-insoluble polyamino-acids can be brought about under mechanical stress in steam. Reversion to the α-helix occurs on solution in helicogenic solvents.[178] Of a series of poly-γ-alkyl-L-glutamates only the n-propyl ester is non-helical in chloroform.[179] Although the different side-chains of poly-γ-alkyl-L-glutamates affect the helix strength, it has no apparent effect on the enthalpy of the helix–coil transition.[180] A considerable electric-field effect has been found on the helix–coil transition of poly-γ-benzyl-L-glutamate. This has been interpreted to indicate that conformational changes in biological systems could be potentially caused by the direct action of an electric field.[181] A form of poly-γ-benzyl-L-glutamate with helices packed into a two-dimensional oblique net, with the side-chains distorted, is indicated in an *X*-ray study of a film cast from dimethylformamide.[182] I.r. studies of poly-β-alkyl-L-aspartates show that those polymers which form left-handed helices show higher frequencies for the amide bonds than those which form helices of right-handed screw sense.[183] Spectroscopic studies have been made of poly-L-*p*-(*p'*-hydroxyphenylazo)phenylalanine and copolymers with *N*-(3-hydroxypropyl)-L-glutamine. In trimethyl phosphate and aqueous solutions between pH 10 and 12 right-handed helices are formed, and there is evidence of exciton coupling of spatially adjacent azoaromatic residues.[184a] At high trifluoroacetic acid concentrations the homopolymer assumes an extended polyelectrolyte conformation.[184b] All the absorption bands and Cotton effects of poly-L-*p*-(2'-hydroxy-5'-methylphenylazo)phenylalanine in hexafluoro-2-propanol are attributable to side-chain order and not to the peptide backbone.[184c] Detailed c.d. studies on poly-L-alanine and poly-L-serine have been reported recently,[185] and a further study of the far-u.v. Cotton effects of some aromatic homopolymers has appeared.[186]

11 Polyamino-acids: Studies in Aqueous Solution

Only polyamino-acids derived from those protein amino-acids having ionisable side-chains are readily water-soluble, and they undergo a helix-to-random-coil transition upon ionisation.[6i] Some polymers of this type also undergo this transition in the solid state as a result of changes in relative

[177] M. L. Tiffany and S. Krimm, *Biopolymers*, 1968, **6**, 1767.
[178] W. B. Gratzer, W. Rhodes, and G. D. Fasman, *Biopolymers*, 1963, **1**, 319.
[179] S. Sugai, K. Kamashima, and K. Nitta, *J. Polymer Sci.*, Part A-2, *Polymer Phys.*, 1968, **6**, 1065.
[180] G. Giacometti, A. Turola, and R. Boni, *Biopolymers*, 1968, **6**, 441; see also A. Kagemato and R. Fujishiro, *ibid.*, p. 1753.
[181] G. Schwartz and J. Seelig, *Biopolymers*, 1968, **6**, 1263.
[182] A. J. McKinnon and A. V. Tobolsky, *J. Phys. Chem.*, 1968, **72**, 1157.
[183] E. M. Bradbury, B. G. Carpenter, and R. M. Stephens, *Biopolymers*, 1968, **6**, 905.
[184] [a] M. Goodman and E. Benedetti, *Biochemistry*, 1968, **7**, 4226; [b] *Ibid.*, p. 4243; [c] *Ibid.*, p. 4242.
[185] F. Quadrifoglio and D. W. Urry, *J. Amer. Chem. Soc.*, 1968, **90**, 2755, 2760.
[186] M. Goodman, C. Toniolo, and E. Peggion, *Biopolymers*, 1968, **6**, 1691.

humidity.[178] A salt-induced helix–coil transition of poly-α-L-glutamic acid has also been observed.[187, 188] Stabilisation of the un-ionised helix of poly-α-L-glutamic acid has been ascribed to a super-helix of γ-carboxyl : γ-carboxyl hydrogen bonds. Addition of solvents such as dioxan, methanol, and chloroethanol to this polymer increases helicity.[6i] Poly-γ-D-glutamic acid occurs naturally in the cell wall of *Bacillus anthracis*. Since the peptide linkages involve the γ-carboxyl group it cannot form an α-helix, but seems to adopt a 3_{17} or 3_{19} left-handed helical conformation stabilised by α-carboxyl : α-carboxyl side-chain hydrogen bonding. In poly-α-L-lysine the helical conformation is assumed on the loss of charge from the side-chain amino-groups. Poly-α-L-aspartic acid, like poly-α-L-glutamic acid, forms a right-handed helix, but it is rather weaker; it is sterically impossible to form β-carboxyl : β-carboxyl hydrogen bonds between the side-chains. Helical poly-α-L-ornithine similarly seems less stable than poly-α-L-lysine.[6i]

The c.d. spectrum of poly-α-L-glutamic acid at high pH has recently been found to differ significantly from the characteristic random-coil spectrum. It has been suggested that the polymer may still be partly helical. Theoretical calculations favour a left-handed extended helix of *ca.* 2·5 residues per turn when the side-chains are ionised.[189] The helix-coil transition of poly-α-L-glutamic acid has been followed in D_2O by measuring the self-diffusion coefficient by pulsed-field gradient spin-echo n.m.r.[190] The temperature-induced aggregation and disaggregation of poly-α-L-glutamic acid gives rise to a reproducible hysteresis loop which can be followed by several physical methods.[191] The aggregated form has been separated chromatographically.[192] An aqueous solution of an Acridine Orange–poly-α-D-glutamic acid complex at pH 7·5 has been found to change from the random coil to the α-helix on compression to 4500 atmospheres.[193] Investigations on the action of papain on poly-α-L-glutamic acid and poly-α-L-lysine in aqueous ethanolic solution provides further evidence that the α-helix is not susceptible to enzymic hydrolysis.[194] O.r.d. and c.d. data suggest that in aqueous solution at pH 7 a mixture of poly-α-L-glutamic acid and poly-α-L-lysine forms a β-pleated sheet structure with a 1 : 1 stoicheiometry, although at this pH both components normally exist as random coils. Since β-structure in the water-soluble polyamino-acids is observed only with the homopolymers of those acidic and basic protein amino-acids with the longest hydrocarbon side-chains, hydrophobic interactions are apparently of importance in the stabilisation of the aggregate.[195] Spectroscopic studies on aqueous solutions of poly-α-L-

[187] G. D. Fasman, C. Lindblow, and E. Bodenheimer, *Biochemistry*, 1964, **3**, 155.
[188] A. Ciferri, D. Puett, L. Rajagh, and J. Hermans jun., *Biopolymers*, 1968, **6**, 1019.
[189] S. Krimm and J. E. Mark, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 1122.
[190] R. E. Moll, *J. Amer. Chem. Soc.*, 1968, **90**, 4739.
[191] B. R. Jennings, G. Spach, and T. M. Schuster, *Biopolymers*, 1968, **6**, 635.
[192] G. Spach and D. Constantin, *Biopolymers*, 1968, **6**, 653.
[193] K. Suzuki and Y. Taniguchi, *Biopolymers*, 1968, **6**, 215.
[194] W. G. Miller and J. Monroe, *Biochemistry*, 1968, **7**, 253.
[195] G. G. Hammes and S. E. Shullery, *Biochemistry*, 1968, **7**, 3882.

ornithine show only a 20% helical content above pH 11 when the side-chains are un-ionised. This contrasts with the complete helicity of poly-α-L-lysine above this pH. Intrinsic viscosity studies suggest that the polymer contracts on deprotonation.[196] Poly-α-L-lysine hydrochloride has been found to form α-helix when the methanol content of an aqueous solution rises above 87%. A maximum helical content is reached at 97% methanol, and the observation of a primary charge effect in sedimentation equilibrium runs indicates that the polyelectrolyte is charged under these conditions.[197] Attempts have been made to correlate ultrasonic absorption and the helix–coil transition of poly-α-L-lysine.[198] Poly-L-methionine-3-methylsulphonium salts undergo a transition from the random coil to a β-form on increasing the concentration of sodium iodide or sodium thiocyanate in the solution.[199] Collapsed monolayers of poly-β-benzyl-L-aspartate prepared after spreading chloroform–dichloroacetic acid solutions on water and allowing the organic solvent to evaporate seem to be present as right-handed α-helices; normally the left-handed helix is formed by this polymer.[200] Other studies on monolayers of poly-ε-carbobenzoxy-L-lysine [201] and esters of poly-γ-L-glutamic acid [202] have been described.

Some non-ionic polyamino-acids are water soluble. Poly-L-homoserine only has a small helical content in water,[203] and poly-N^5-(3-hydroxypropyl)-L-glutamine is less helical in water than in other organic solvents.[204] The relevance of work in non-aqueous media to the conformational stability of proteins can only be evaluated if the stabilisation forces act independently of solvent–solute interactions. The hydrophobic interior of proteins may more approximate to these conditions.[6a] Such problems can be partly overcome by studying copolymers of the water-insoluble polyamino-acids with flanking blocks of water-soluble DL-polyamino-acids. In the water-soluble copolymer $(\text{DL-Lys})_{15}(\text{L-Ala})_{10}(\text{DL-Lys})_{15}$ the helical content of the L-alanine section is *ca.* 50%, which is much higher than in the corresponding decamers of L-lysine and L-glutamic acid, indicating the greater helix-stabilising influence of the alanine side-chain. As the size of the alanine block is increased, so is the helical content, reaching 73% for a 50-residue central portion.[205] An L-valine block of 15 residues flanked by DL-lysine blocks is helical in 98% methanol, as predicted from a theoretical study, but forms intermolecular β-structure in water.[206] Block copolymers

[196] S. R. Chaudhuri and J. T. Yang, *Biochemistry*, 1968, **7**, 1379.
[197] R. F. Epand and H. A. Scheraga, *Biopolymers*, 1968, **6**, 1338.
[198] R. C. Parker, L. J. Slutsky, and K. R. Applegate, *J. Phys. Chem.*, 1968, **72**, 3177.
[199] S. Makino, K. Wakabayashi, and S. Sugai, *Biopolymers*, 1968, **6**, 551.
[200] B. R. Malcolm, *Nature*, 1968, **219**, 929.
[201] B. R. Malcolm, *Biochem. J.*, 1968, **110**, 733.
[202] G. I. Loeb, *J. Colloid Interface Sci.*, 1968, **26**, 236; *ibid.*, 27, p. 38.
[203] G. D. Fasman, Abstracts 142nd Meeting, American Chemical Society, September 1962, p. 54c.
[204] N. Lupu-Lotan, A. Yaron, A. Berger, and M. Sela, *Biopolymers*, 1965, **3**, 625.
[205] N. Lupu-Lotan, A. Berger, E. Katchalski, P. I. Ingwall, and H. A. Scheraga, *Biopolymers*, 1966, **4**, 239; *ibid.*, 1968, **6**, 331.
[206] R. F. Epand and H. A. Scheraga, *Biopolymers*, 1968, **6**, 1551.

of the type $(\text{DL-Ser})_{30}(\text{L-Ser})_{30}(\text{DL-Ser})_{30}$ have o.r.d. spectra in aqueous solution which resemble films of poly-L-serine in showing the characteristics of β-structure. Low molecular weight poly-L-serine, which is reasonably water-soluble, seems to be largely random chain in aqueous solution, but addition of methanol causes formation of β-structure. O.r.d. evidence suggests possible transient formation of α-helix during this transition.[207]

12 Copolymers

Studies on polyamino-acids, which are the simplest protein models, established some ideas about the behaviour of side-chains in modifying the secondary structural characteristics.[6i] The natural line of advance was then to study copolymers of two amino-acid residues to see the mutual influence of the side-chains in modifying the properties of the parent polyamino-acids. The *N*-carboxy-anhydride method lent itself readily to the preparation of copolymers of this type, but such studies have proved to be somewhat less useful than was originally hoped because of the non-random nature of some of the products.[6i] Many interesting copolymers have been explored, including mixtures of helix formers of the same screw sense of both water-soluble and insoluble parent polyamino-acids, and those of different screw sense. Incorporation of 31% of L-leucine into helical poly-γ-benzyl-L-glutamate actually increases the stability of the helix,[187] but glycine gives only block copolymers on copolymerisation with γ-benzyl-L-glutamate.[6i] Incorporation of L-leucine into the poly-α-L-glutamic acid helix contributes added stability, although the reason is not fully clear.[208] The random-chain homopolymer poly-γ-*N*-(2-morpholinoethyl)-α-L-glutamine is converted into a helix by incorporation of other residues such as L-alanine.[209] C.d. measurements on copolymers of L-tryptophan and poly-γ-ethyl-L-glutamate in ethylene glycol monomethyl ether indicate that poly-L-tryptophan, which is difficult to study directly because of the side-chain chromophores, is a right-handed helix in this solvent. *X*-Ray studies on films cast from dimethylformamide also show α-helix.[210] Poly-L-tryptophan block copolymerised with poly-γ-ethyl-DL-glutamate to render it soluble in solvents suitable for u.v. studies appears to be helical in trifluoroethanol, but the type of helix cannot be deduced.[211] The benzyl groups of poly-β-benzyl-L-aspartate can be replaced by alkyl groups using an acid-catalysed transesterification, leading to better than 99% replacement if required. Temperature-induced transitions from right-handed to left-handed helices have been observed for poly-β-n-propyl-L-aspartate and a copolymer of β-benzyl- and β-ethyl-L-aspartates prepared by this transesterification method.[212] Similar transitions occur in copolymers of β-benzyl- and

[207] N. M. Tooney and G. D. Fasman, *J. Mol. Biol.*, 1968, **36**, 355.
[208] R. E. Nylund and W. G. Miller, *J. Amer. Chem. Soc.*, 1965, **87**, 3537.
[209] P. Doty and R. E. Lundberg, *J. Amer. Chem. Soc.*, 1956, **78**, 4810.
[210] E. Peggion, A. Cosani, A. S. Verdini, A. De Pra, and M. Mammi, *Biopolymers*, 1968, **6**, 1477.
[211] A. Cosain, E. Peggion, A. S. Verdini, and M. Terbojevich, *Biopolymers*, 1968, **6**, 963.
[212] E. M. Bradbury, B. G. Carpenter, and H. Goldman, *Biopolymers*, 1968, **6**, 837.

β-nitrobenzyl-L-aspartates.[213] Although poly-β-methyl-L-aspartate forms a left-handed α-helix similar to poly-β-benzyl-L-aspartate, introduction of β-methyl-L-aspartate residues into poly-β-benzyl-L-aspartate prevents the formation of the ω-helix by the latter.[183] Detailed studies on a copolymer of L-glutamic acid and L-lysine have been reported.[214]

It has been established that the polymerisation of mixtures of *N*-carboxy-anhydrides is often not a random process. Even if the rates of reaction are similar, there can be stereochemical causes for non-randomness. For example, when a preformed poly-L-amino-acid α-helix is used to initiate polymerisation of a DL-*N*-carboxy-anhydride, the chain continues to grow in the same helical sense. There is a strong preference of the lengthening chain for its own isomer residue.[209] In copolymers of γ-benzyl-D- and γ-benzyl-L-glutamates, as the ratio of enantiomers approaches 1 : 1, both right- and left-handed helical sections seem to be present.[215] It is also impossible even in truly random copolymers to isolate specific interactions between side-chains as they will not occur in regular positions along the surface of any helical form.[6i]

13 Sequential Polypeptides

See also chapter 3, sections 2E and 4C.

Since polypeptides containing a known repeating sequence of amino-acids would be useful as protein models of rather more complexity than the polyamino-acids, and because of the difficulties with copolymers outlined above, attention has turned recently to the synthesis of such 'sequential' polypeptides. Most polypeptides of this type have been prepared by linking preformed oligopeptide units using an active ester polycondensation. In this method there is usually a risk of racemisation at the *C*-terminal amino-acid residue unless it is glycine or proline. The products are in general of lower average molecular weight and are more polydisperse than polyamino-acids prepared by the racemisation-free *N*-carboxy-anhydride method. A random copolymer of 10% L-valine and 90% L-methionine was found to be less helical than poly-L-methionine,[216] but a sequential polypeptide poly-(L-Met_4-L-Val) showed no decrease in helical content.[217] These observations illustrate the relative importance of side-chain interactions. Several series of sequential polypeptides containing varying proportions of two types of amino-acid residue merit attention as conformational models. Australian workers have studied the effect on conformation of separately introducing *S*-benzyl-L-cysteine [218] and glycine [219] residues

[213] C. Toniolo, M. L. Malxa, and M. Goodman, *Biopolymers*, 1968, **6**, 1579.
[214] K. Morita, E. R. Simons, and E. R. Blout, *Biopolymers*, 1968, **6**, 181.
[215] A. Wada, *J. Mol. Biol.* 1961, **3**, 507.
[216] S. M. Bloom, G. D. Fasman, C. de Lozé, and E. R. Blout, *J. Amer. Chem. Soc.*, 1962, **84**, 458.
[217] R. D. B. Fraser, B. S. Harrap, T. P. MacRae, F. H. C. Stewart, and E. Suzuki, *J. Mol. Biol.*, 1965, **12**, 482.
[218] R. D. B. Fraser, B. S. Harrap, T. P. MacRae, F. H. C. Stewart, and E. Suzuki, *J. Mol. Biol.*, 1965, **14**, 423.
[219] R. D. B. Fraser, B. S. Harrap, T. P. MacRae, F. H. C. Stewart, and E. Suzuki, *Biopolymers*, 1967, **5**, 251.

into poly-γ-ethyl-L-glutamate. The helix-breaking properties of these amino-acids were found to be largely independent of sequence. This is in accord with the idea of unfavourable contacts between the cysteine side-chain and the peptide backbone, while transfer of a glycyl residue from a random coil to a helix has an unfavourable entropy factor compared to other amino-acids. The introduction of L-valyl residues into poly-γ-ethyl-L-glutamate does, however, depend profoundly on sequence.[217] The polymers in this series fall into two groups, those with a similar helicity to poly-γ-ethyl-L-glutamate and those with a greatly decreased helical content. Analysis of the environments of the valine residues in such polymers on the surface of the α-helix has shown that only in the cases of decreased helicity do the side-chains of valyl residues in adjacent turns of the α-helix seriously interfere.[217] A series of sequential polymers examining the introduction of L-leucyl residues into poly-γ-benzyl-D-glutamate have shown similar helix stabilities when 2–4 γ-benzyl-D-glutamyl residues are present for each L-leucine residue. Such helices are much weaker than either of the parent polyamino-acids. In this series the two components are in opposition in determining the overall helical sense, but the higher fraction of γ-benzyl-D-glutamyl residues causes them to control the screw sense. Poly-(γ-benzyl-D-glutamyl-L-leucine) adopts an ordered conformation in solution which is neither an α-helix nor a β-structure. This secondary structure reverts to the random coil on addition of a few per cent of trifluoroacetic acid, is less rigid than an α-helix, and is intramolecularly hydrogen bonded. β-Structures containing alternate D- and L-residues have unfavourable side-chain contacts, and this new conformation may be generally adopted by poly-DL or LD dipeptides. Racemisation of the *C*-terminal residue in this series was found to be less than 2%.[220]

Sequential polypeptides have also been prepared for antigenic studies,[221] enzyme models,[222] silk fibroin models,[223] histone and protamine models,[224] anthrax polypeptide models,[225] and frequently as collagen models.[226]

[220] P. M. Hardy, H. N. Rydon, and R. C. Thompson, abstracts, Chemical Society Protein Group Meeting, Keele, September 1968.

[221] B. J. Johnson, *J. Chem. Soc.* (*C*), 1967, 2638; F. M. Richards, R. W. Sloane, and E. Haber, *Biochemistry*, 1967, **6**, 476.

[222] D. F. De Tar, F. F. Rogers jun., and H. Bach, *J. Amer. Chem. Soc.*, 1967, **89**, 3039, and refs. therein.

[223] F. H. C. Stewart, *Austral. J. Chem.*, 1966, **19**, 489.

[224] G. Spach, A. Brack, and F. Heitz, *Compt. rend.*, 1967, **264**, *C*, 19; V. D. Davydov and V. G. Debadov, *Zhur. obshchei Khim.*, 1967, **37**, 1000.

[225] J. Kovacs, G. N. Schmidt, and U. R. Ghatak, *Biopolymers*, 1968, **6**, 817.

[226] S. M. Bloom, S. K. Dasgupta, R. P. Patel, and E. R. Blout, *J. Amer. Chem. Soc.*, 1966, **88**, 2035; P. J. Oriel and E. R. Blout, *ibid.*, p. 2041; J. Engel, J. Kurtz, E. Katchalski, and A. Berger, *J. Mol. Biol.*, 1966, **16**, 404; *ibid.*, **17**, 255; M. Huggins, K. Ohtsuka, and S. Morimoto, *J. Polymer Sci., Part C*, 1968, 343; V. A. Shibnev, T. P. Chuvaeva, and K. T. Poroshin, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1968, 225; and IUPAC 5th International Symposium on the Chemistry of Natural Products, London, July 8–13, 1968, Abstract E-22, p. 270; V. A. Shibnev, K. T. Poroshin, T. P. Chuvaeva, and G. A. Martynova, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1968, 1144; A. M. Tamburro, A. Scatturin, and F. Marchiori, *Gazzetta*, 1968, **98**, 638.

Recently the collagen model poly-(L-Pro·L-Pro·Gly) has been synthesised by a solid-phase method by adding on successive tripeptide units.[227] In contrast to sequential polypeptides of this type prepared by the conventional active ester polymerisation,[226] material containing 20 units was practically monodisperse. The temperature dependence of its optical rotation was much closer to that of natural collagen than previously prepared polydisperse material.[227] Solid-phase synthesis has much to commend it for the preparation of sequential polypeptides, unlike the synthesis of natural peptides whose sequences are non-repeating; the non-occurrence of coupling of a small fraction of unblocked amino-groups at any one stage will only have the effect of slightly increasing the polydispersity of the product since the unit is coupled repetitively.

14 The Helix–Coil Transition

The transition of helical polyamino-acids to random coils can be brought about by several means. Polyamino-acids with ionisable side-chains undergo a transition on change of the pH of aqueous solutions coincident with ionisation of the side-chains, and the helix is thought to be broken in this case by electrostatic repulsion of the charged side-chains. Conformational disruption by lithium salts is attributed to the formation of ion-pair complexes with the polymer amide groups, or, in peptides with acidic side-chains, to possible interference with side-chain carboxyl association.[6i] Lithium salts will render many water-insoluble peptides soluble by interaction with the amide groups. In the case of random-coil formation in non-aqueous solutions caused by the addition of trifluoroacetic or dichloroacetic acids, those polyamino-acids with purely hydrocarbon side-chains might be expected to offer the greatest resistance to penetration by solvent. The presence of any polar group, such as an ester, should aid penetration of the side-chain shield and subsequently loosen the mutual interactions of the side-chains, eventually leading to breakdown of the helix. On this basis the only requirement for a compound to break the helix is the possession of a large enough permanent dipole moment.[6i] N.m.r. has shown the unfreezing of the poly-γ-benzyl-L-glutamate helix from the periphery inwards.[38] Temperature-induced transitions are often reported. These are thought to reflect the balance between the enthalpies and entropies of formation of the two forms. Often the helix becomes stronger with increasing temperature.[228] In aqueous solution the addition of salts other than lithium and also miscible organic solvents such as methanol can cause changes in helical strength. Such organic solvents raise the pK_a of the carboxyl groups, lower the hydrogen-bonding capacity of the solvent relative to water, and strengthen the intramolecular hydrogen bonds; all of these factors tend to increase helicity, but on the other hand they

[227] S. Sakakibara, Y. Kishida, Y. Kikuchi, R. Sakai, and K. Kakiuchi, *Bull. Chem. Soc. Japan*, 1968, **41**, 1273.

[228] A. Nakajima and T. Hayashi, *Chem. High Polymers (Japan)*, 1968, **25**, 281.

increase electrostatic repulsion interactions. The presence of salts also affects the degree of ionisation of the side-chains.[6i] Several theoretical studies have been made of the helix–coil transition, and predictions agree fairly well with experimental results.[229] Each polyamino-acid chain during transition can be a mixture of helical and random-coil segments. The helical segments ripple along the chain through thermal interactions.[229] N.m.r. results show that the rate of interconversion of the two types of segment can differ according to the nature of the side-chain and the helix content of the particular sample at any given solvent composition.

There is disagreement as to whether disruption of the helix in non-aqueous solvents by carboxylic acids is due to protonation of the amide group, leading to fission by electrostatic repulsion, or to hydrogen-bond competition as well as simple weakening of side-chain interactions. N.m.r. studies of poly-γ-benzyl-L-glutamate have been interpreted as indicating the absence of protonation before or during the helix–coil transition.[39] I.r. studies, however, have supported the theory of protonation.[230] An i.r. band due to carboxylate anion has been detected in the presence of helical polymer and trifluoroacetic acid.[231] Several workers have investigated possible protonation by studying the n–π^* transition band of the peptide amide group. Protonation of the amide group should not be possible without greatly modifying this band.[232-234] The n–π^* band of poly-γ-benzyl-L-glutamate is weak, and trifluoroacetic acid absorbs in the same region of the spectrum, making absorption measurements unreliable. However, the Cotton effect associated with this n–π^* transition is large, and can be followed conveniently. Only small changes in the o.r.d. spectrum were found to occur on addition of up to 10% of trifluoroacetic acid to an ethylene dichloride solution of the polyamino-acid, but as the percentage of acid was raised to 15, the Cotton effect virtually vanished. The b_0 shows the transition to the random coil occurring over just this range of solvent composition.[234] An independent c.d. study shows a similar change.[233] These results indicate that no protonation occurs before the onset of the helix–coil transition, but no decision is possible from this finding as to the nature of the interaction causing transition. Intermolecular hydrogen bonding may precede the transition if the method is insensitive to such changes.[233, 234] O.r.d. studies on D-oxolupanine (3), a model amide with a

[229] Reviewed by B. H. Zimm in 'Polyamino-acids, Peptides, and Proteins,' ed. M. A. Stahmann, University of Wisconsin Press, Madison, 1962, p. 229; see also No. Gō, M. Gō, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 1030; M. E. Craig and D. M. Crothers, *Biopolymers*, 1968, **6**, 385; T. R. Fink and D. M. Crothers, *ibid.*, p. 863.

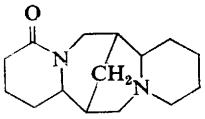
[230] S. Hanlon, *Biochemistry*, 1966, **5**, 2049, and refs. therein.

[231] B. Z. Volchek and A. V. Purkina, *Vysokomol. Soedineniya, Ser. A*, 1967, 1257.

[232] F. A. Bovey, *Pure Appl. Chem.*, 1968, **16**, 417, 'Proceedings of the International Symposium on Macromolecular Chemistry', Brussels–Louvain, Belgium, June 12–16, 1967.

[233] F. Quadrifoglio and D. W. Urry, *J. Phys. Chem.*, 1967, **71**, 2364.

[234] D. Balasubramian, *Biochem. Biophys. Res. Comm.*, 1967, **29**, 538.



(3)

strong n–π^* band but which is rigid and incapable of conformational change and therefore a simple model of a peptide link in an α-helix, shows no diminution in intensity of the n–π^* band on addition of large excesses of trifluoroacetic acid. This indicates the absence of protonation under transition conditions, and contrasts with the results described above. Addition of sulphuric acid to D-oxolupanine does diminish the intensity of the n–π^* band, indicating protonation by this stronger acid.[232]

N.m.r. studies on poly-L-methionine in solvent mixtures containing trifluoroacetic acid where both helix and coil are present show two partly overlapping peaks for both amide NH and α-CH protons, corresponding to the two conformations. The peak area for the amide NH in the helix is deficient by 25–30% compared to that for the αCH proton in the same conformation. This limited exchange is ascribed to partial protonation of the peptide linkages in the helical, but not in the random coil, form.[44] Skeletal models of helical structures remain rigid when no more than one in four of the hydrogen bonds is broken. It is suggested that the helical polymer is protonated by trifluoroacetic acid, but cannot tolerate breakage of more than about a quarter or a third of the intramolecular hydrogen bonds before collapsing to form the random coil hydrogen-bonded to trifluoroacetic acid.[44] A similar phenomenon has been shown for other polyamino-acids, but in the case of poly-β-benzyl-L-aspartate the deficiency of NH band area found when dichloroacetic acid was used as the helix-breaking solvent was not seen when [$^2H_6$]dimethyl sulphoxide was used. Since [$^2H_6$]dimethyl sulphoxide must break this relatively weak helix by some mechanism other than protonation, this observation is in accord with the explanation that the deficiency in NH band area before the disruption of the helix reflects the degree of protonation.[235]

Observation of polyelectrolyte behaviour in the viscometry of dichloroacetic acid solutions of poly-γ-benzyl-L-glutamate and several other polyamino-acids indicates that their random-coil forms are protonated in this solvent. These polyelectrolyte effects are suppressed by the addition of water (*ca.* 1·45%) through the formation of ions which shield the charges on the peptide groups and reduce their electrostatic repulsion. For acids stronger than dichloroacetic acid less water should be required to suppress the polyelectrolyte effect. In trifluoroacetic acid solution only poly-ε-benzyloxycarbonyl-L-lysine shows a polyelectrolyte effect. It is suggested that there is enough water present even in dried trifluoroacetic acid (*ca.* 0·18%) to cause suppression, and that the observation of the polyelectrolyte

[235] J. C. Haylock and H. N. Rydon, personal communication.

effect in poly-ε-benzyloxycarbonyl-L-lysine is due to protonation at the ε-NH group as well as the peptide NH group.[59b]

15 Some Conformational Aspects of Proteins

Early studies on fibrous proteins in the solid state by i.r. and *X*-ray diffraction established that they are made up of repeating units, often β-pleated sheet structures.[6i, 236a] The water-solubility of globular proteins is related to the arrangement of their amino-acid side-chains to give the molecule a hydrophobic interior and hydrophilic exterior.[237] Study of the globular proteins has proved more difficult than the fibrous proteins because they contain more than one type of secondary structure. They are, however, of great intrinsic interest because of their interaction with other types of molecules. Hydrodynamic studies initially showed that they behaved in solution as rigid, moderately asymmetric particles with dimensions sufficiently small to require that their polypeptide chains are folded in some compact manner. Spectroscopic studies provided evidence for their partial helicity in solution before *X*-ray studies on wet protein crystals reached their present sophistication.[6i] Myoglobin was the first protein whose conformation was completely described by *X*-ray analysis.[238] The α-helix was seen for the first time in detail at atomic resolution in myoglobin, but β-structure was characterised in this way for the first time only later in lysozyme.[239] In most cases the helical content of proteins as shown by spectroscopic measurements on solutions agrees quite well with the values obtained from *X*-ray diffraction studies,[6i] as in the recent case of papain, which contains 20% of α-helix.[240]

Experiments on polyamino-acids, copolymers, and sequential polypeptides have indicated how the constituent amino-acids may affect the conformation adopted. Proline is difficult to accommodate in an α-helical structure. It cannot easily be fitted into an α-helix if the helix is to continue in the same direction, and if present in the *cis*-form a sharp bend can be produced as well as a reversal of the helix sense. In myoglobin all four proline residues are situated at the end of helical segments.[238] In view of the high proline content of collagen it is not surprising that this molecule adopts a unique ordered structure. This structure consists of three intertwined peptide chains each in the poly-proline II conformation.[6i] The other protein amino-acids are less sterically demanding than proline. Although homopolymers of residues like glycine do not form α-helices and

[236] [a] Reviewed by S. Seifter and P. M. Gallop in 'The Proteins,' vol. 4, ed. H. Neurath, 2nd edn., 1966, Academic Press, New York, p. 155. [b] D. R. Helinsky and C. Yanofsky, *ibid.*, p. 1.

[237] C. B. Anfinsen, 'Harvey Lecture Series,' Academic Press, New York, 1966, p. 95.

[238] M. F. Perutz, J. C. Kendrew, and H. C. Watson, *J. Mol. Biol.*, 1965, **13**, 669.

[239] C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *Nature*, 1965, **206**, 757.

[240] J. Dreuth, J. N. Jansonius, R. Koekoek, H. N. Swen, and B. G. Walthers, *Nature*, 1968, **218**, 929.

those of valine and serine prefer to form β-structures, they can be accommodated to a certain extent in the α-helical segments of proteins without causing disruption. Several examples have been found in protein *X*-ray diffraction studies of secondary structures which are distorted from the normal bond angles.[6i] Attempts are being made to predict the conformations of proteins from their amino-acid sequences on the basis of comparisons with the sequences of proteins of known conformation.[241]

As far as is known, in protein biosynthesis there is genetic control of the primary structure only. The conformation enters into living processes only when a newly synthesised polypeptide chain coils, folds, and is stabilised into a functional protein molecule. No evidence has yet been found for enzymic catalysis of this folding. This has led to the idea that the secondary and tertiary structures of globular proteins are controlled solely by the amino-acid sequence.[236b, 237] Globular proteins can lose their native conformation under a variety of conditions, denaturing to the random-coil form. A spontaneous refolding or renaturing of the random coil derived from some proteins that do not contain disulphide bonds, *e.g.* tobacco mosaic virus protein, often occurs on elimination of the conditions which caused denaturation. The reversibility of denaturation in such proteins can be estimated by determining restored biological activity or physical properties related to the native conformation.[237]

Proteins which contain disulphide linkages constitute a rather more complex case. Ribonuclease, which contains four disulphide bridges, can, after reduction in 8M-urea solution, be oxidised back to the active protein in nearly theoretical yield.[237] Isomers containing incorrect disulphide links are formed if the oxidation is carried out without removing the urea. Such 'scrambled' isomers undergo disulphide interchange in the presence of thiols to regenerate the native form. Studies on derivatives of ribonuclease have indicated the relatively unimportant role of surface-oriented hydrophilic side-chains in directing the refolding and half-cystine pairing processes during the reoxidation of reduced chains. Ribonuclease, however, constitutes one of the more clear-cut cases of activity regeneration, but even γ-globulin (after modification to overcome the insolubility of one of the two component chains) seems to recombine specifically to give 50% of its initial activity even though 23 disulphide bridges are involved.[237]

An enzyme has been isolated which catalyses the rearrangement of disulphide-scrambled isomers of ribonuclease, lysozyme, and some other proteins containing single peptide chains to the native forms.[242] The rate of reactivation seems to depend on the number of bonds requiring rearranging. This disulphide isomerase rapidly inactivates insulin and chymotrypsin (proteins containing respectively two and three separate

[241] B. W. Low, F. M. Lovell, and A. D. Rudker, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 1519.

[242] D. Givol, F. De Lorenzo, R. F. Goldberger, and C. B. Anfinson, *Proc. Nat. Acad. Sci. U.S.A.*, 1965, **53**, 676.

peptide chains joined by disulphide bridges) by scrambling the disulphide links. In proteins containing more than one peptide chain linked in this way, the sequence of the chains seems able to exert only a limited control over disulphide bridging in favour of the native form on reoxidation of the reduced forms.[243] It has been known for some time that chymotrypsin is formed from a single-chain precursor or zymogen, chymotrypsinogen, which is unaffected by the disulphide isomerase.[237] This year the structure of the single-chain precursor of porine insulin, proinsulin, has been determined.[244] Proinsulin loses a segment of 33 amino-acid residues on conversion to insulin, and loss of part of the peptide chain in this way after formation of the disulphide bridges leads to a molecule with only a limited ability to re-form the native disulphide bridges on reoxidation of the reduced form.[243a] Proinsulin is, however, reversibly reduced and oxidised in high yield.[243b] It has been suggested on o.r.d. evidence that the insulin part of the proinsulin chain adopts a similar conformation in solution to insulin itself, and the lower overall α-helical content (14% compared to 25% in insulin) reflects the random-coil nature of the connecting chain. Of the 33 residues of this connecting chain, in porcine proinsulin seven are glycine and three are proline residues.[245] (See also chapter 2, part I, section 4B and chapter 5, section 3.)

Although the term 'random coil' is often used to describe peptides or proteins with no periodic secondary structure, it is probable that, as in oligopeptides, the chain is not completely random in the sense that there are some conformational preferences about the dihedral angles ϕ and ψ. The dimensions of protein random coils have been calculated for a variety of proteins of known amino-acid sequence. Glycine and proline contribute to reducing the dimensions. The mean square end-to-end distance assuming a random sequence was generally within 10% of the value obtained when a non-random sequence was considered. The results agree quite well with experimental values.[246]

16 Recent Applications of Physical Methods to the Study of Protein Conformations

Proteins are in general less homogeneous in conformation than polyamino-acids, but they differ widely amongst themselves with respect to their content of α-helix, β-structure, and random chain. There is increasing evidence for the common occurrence also of distorted secondary structures. Proteins possess varying amounts of side-chain chromophores and can be associated with non-peptidic prosthetic groups. As a result their o.r.d. and c.d. spectra are mostly very complex, and at the present time usually

[243] [a] P. G. Katsoyannis, in 'The Pharmacology of Hormonal Polypeptides and Proteins,' eds. N. Bach, L. Martini, and R. Paoletti, Plenum Press, New York, 1968, p. 278.
[b] D. F. Steiner and J. L. Clark, *Proc. Nat. Acad. Sci., U.S.A.*, 1968, **60**, 622.
[244] R. E. Chance, R. M. Ellis, and W. M. Bromer, *Science*, 1968, **161**, 165.
[245] B. H. Frank and A. J. Veros, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 155.
[246] W. G. Miller and C. V. Goebel, *Biochemistry*, 1968, **7**, 3925.

only lend themselves to qualitative interpretation. The large number of different protons and proton environments in proteins similarly hinders n.m.r. work, but the number of papers dealing with n.m.r. studies has increased sharply this year.

The appearance of proton resonances outside the range normally covered by diamagnetic compounds in the n.m.r. spectrum of sperm whale cyanmetmyoglobin has been discussed in terms of hyperfine interactions with the paramagnetic haem group.[247] Spin labels attached to the β-93 position in human haemoglobin have shown that changes in local protein conformation occur on oxygenation which depend on the state of ligation of the α- as well as the β-haem group.[248] The c.d. spectrum of deoxyhaemoglobin shows little change on oxygenation.[249] The c.d. spectrum of a respiratory haem-protein of *Chironomus thumin* indicates one-third less α-helix than is found in myoglobin.[250] O.r.d. studies show that conformational differences exist between both γA- and equine γT-globulins and their respective γG-globulins. While the rotatory dispersion spectra of the γA- and γT-globulins show overall similarities, comparison of their respective peptic fragments suggests some differences in conformation.[251] Significant differences between the conformations of γM- and γG-globulins are also indicated by o.r.d.[252] γ-Immunoglobin has been studied by c.d.,[253, 254] and no conformational change was detected on formation of the antibody–hapten complex.[253] Studies on bradykinin and some fragments derived from it indicate that antibodies recognise the peptide in a preferred conformation requiring the entire nonapeptide.[255] Examination of ten samples of Bence-Jones proteins showed in all cases a c.d. peak at *ca.* 217 nm., an extremum characteristic of β-structure.[256]

On dilution of a solution of reduced lysozyme in 8M-urea there is an immediate recovery of 70% of the native helical content. Either one or two of the six tryptophan residues is involved. During reoxidation, one or more tryptophans are buried in the internal fold and enzymic activity is recovered, presumably through disulphide interchange.[257] The c.d. spectra of bovine α-lactalbumin and egg white lysozyme, which share a high degree of sequence homology, seem to indicate similar conformational distributions. Differences in b_0 value are ascribed to side-chain optical effects.[258] On

[247] K. Wüthrich, R. G. Shulman, and J. Peisach, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 373.
[248] S. Ogawa, H. M. McConnell, and A. Horwitz, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 401.
[249] H. Takasada, K. Hamaguchi, and S. Migata, *J. Biochem.* (*Japan*), 1968, **63**, 654.
[250] H. Farmanek and J. Engel, *Biochim. Biophys. Acta*, 1968, **160**, 151.
[251] K. J. Dorrington and J. H. Rockey, *J. Biol. Chem.*, 1968, **243**, 6511.
[252] K. J. Dorrington and C. Tanford, *J. Biol. Chem.*, 1968, **243**, 4745.
[253] R. E. Cathon, A. Kulzycki jun., and E. Haber, *Biochemistry*, 1968, **7**, 3950.
[254] D. L. Ross and B. Jirgensons, *J. Biol. Chem.*, 1968, **243**, 2829.
[255] J. Spragg, R. C. Talamo, K. Suzuki, D. M. Appelbaum, K. F. Austen, and E. Haber, *Biochemistry*, 1968, **7**, 4086.
[256] G. K. Ikeda, K. Hamaguchi, and S. Migita, *J. Biochem.* (*Japan*), 1968, **63**, 654.
[257] K. Yutani, A. Yutani, A. Imanishi, and T. Isemura, *J. Biochem.* (*Japan*), 1968, **64**, 449.
[258] M. J. Kronman, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 535.

addition of 2-chloroethanol to β-lactoglobulin in aqueous solution the native protein changes in conformation to a form much richer in α-helix. Light-scattering results indicate that it remains monomeric.[259] An o.r.d. study indicates that fibrinogen has *ca.* 32% α-helix and 14% β-structure. No major conformational change seems to be associated with its thrombin proteolysis.[260] The o.r.d. spectrum of β-lipoprotein shows the presence of β-structure, and a shoulder at 1680 cm.$^{-1}$ in the i.r. spectrum indicates that it is the antiparallel form.[261] The conformation of human serum low-density lipoprotein, shown by c.d. to be a mixture of α-helix and random coil, seems to be rather resistant to disruption unless the lipid component is removed, indicating that the complexed lipid stabilises the protein conformation.[262] O.r.d. monitoring indicates that phosvitin undergoes a transition from an unordered conformation to a β-structure if the pH is lowered sufficiently to suppress ionisation of the side-chain carboxyls.[263]

Assignments have been made of four peaks in the n.m.r. spectrum of ribonuclease to the various C-2 protons of the four histidine residues. Further experiments have shown that the two imidazole C-2 proton peaks of ribonuclease A which are affected by inhibitor binding do indeed correspond to the two active-site histidine residues.[264] The binding of cytidine-3′-monophosphate to ribonuclease has been shown to affect three of the four histidine residues.[265] It has been suggested that the n–π^* Cotton effects of helices which are very short or badly distorted will be displaced towards lower wavelengths. Ribonuclease, which has only 2·5 turns of α-helical structure, shows such a shift.[266] The helicity of this short segment, as monitored by o.r.d., seems to be maintained even after splitting off the *N*-terminal eicosapeptide fragment (S-peptide), in which it occurs, from the rest of the protein.[267] Of the six tyrosine residues in ribonuclease only one exhibits altered chemical and spectroscopic properties when S-protein is formed. This may be the residue involved in an interaction with a carboxylic acid side-chain of the S-peptide.[268]

O.r.d. studies show that glucagon is largely helical in 2-chloroethanol, but mainly random coil in aqueous solutions at all values of pH. Antiparallel β-material precipitates out when acidified solutions of this hormone are set aside. Glucagon is known to be rich in serine and threonine,

[259] H. Inoue and S. N. Timasheff, *J. Amer. Chem. Soc.*, 1968, **90**, 1890.
[260] W. D. McCubbin and C. M. Kay, *Canad. J. Biochem.*, 1968, **46**, 617.
[261] A. M. Gotto, R. I. Levy, and D. S. Fredrickson, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 1436.
[262] A. Scanu and R. Hirz, *Nature*, 1968, **218**, 200.
[263] G. Taborsky, *J. Biol. Chem.*, 1968, **243**, 6014.
[264] D. H. Meadows, O. Jardetsky, R. M. Epand, H. H. Ruterjans, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 766.
[265] D. H. Meadows and O. Jardetsky, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 406.
[266] J. A. Schellman and M. J. Lowe, *J. Amer. Chem. Soc.*, 1968, **90**, 1070; see also A. M. Tamburro, A. Scatturin, and L. Moroder, *Biochim. Biophys. Acta*, 1968, **154**, 583.
[267] W. A. Klee, *Biochemistry*, 1968, **7**, 2731.
[268] E. R. Simons and E. R. Blout, *J. Biol. Chem.*, 1968, **243**, 218.

residues which are prone to form β-structures.[269] Proteins fully deuteriated except for the leucine side-chains have been obtained. The leucine residues in C-phycoerythrin and cytochrome *c* prepared in this way seem to occur in a variety of environments, a sizeable portion being in a very mobile aqueous environment.[270] The c.d. spectra of ferri- and ferro-cytochrome *c* are very similar below 240 nm.,[271, 272] but the aromatic bands and other evidence indicate conformational differences.[272]

Acetoacetate decarboxylase seems to contain very little α-helix, and exists partly in a β-structure which is destroyed on addition of sodium dodecyl sulphate.[273] O.r.d. work indicates that a reversible change of conformation occurs in uridine diphosphate-galactose-4-epimerase on oxidation or reduction of the pyridine nucleotide prosthetic group, which is bound by non-covalent forces to the enzyme. Additional β-structure appears on reduction.[274] Nitration of all three types of carbonic anhydrase at pH 8·5 or below does not change their enzymic activities. No change in o.r.d. or c.d. spectra occurs other than the appearance of new absorption bands due to the 3-nitrotyrosyl residue. Iodination of carbonic anhydrase B causes a decrease in esterase activity when two or more tyrosyl residues are attached, and changes in the c.d. spectrum of the product indicate some loss of native structure. Cotton effects in the 250–300 nm. region of the enzymes are irreversibly lost in 8M-urea even in cases where partial refolding after removal of the urea can be shown by difference spectra.[275] N.m.r. studies on lysozyme [276] and carboxypeptidase,[277] and c.d. studies on trypsin, pepsin,[278] and horseradish peroxidase [279] have also been reported.

Investigation by o.r.d. and c.d. shows that silk fibroin in aqueous solution lacks ordered structure, but viscosity studies suggest that the molecule is less extended than a polyion such as ionised poly-α-L-glutamic acid. Addition of more than 30% of methanol or dioxan induces a time-dependent coil to β-structure transition.[280] A detailed structure has been proposed to explain the *X*-ray diffraction pattern of the cross β-conformation found in *Chrysopa* silk.[281] The i.r. spectrum of human hair keratin obtained by means of the frustrated multiple internal reflectance phenomenon gives

[269] W. B. Gratzer, C. H. Beavan, H. W. E. Rattle, and E. M. Bradbury, *European J. Biochem.*, 1968, **3**, 276.
[270] J. L. Crespi, K. M. Rosenberg, and J. J. Katz, *Science*, 1968, **161**, 795.
[271] Z. Rand and S. Vinogradov, *Arch. Biochim. Biophys.*, 1968, **128**, 94; Y. P. Myer, *Biochim. Biophys. Acta*, 1968, **154**, 84.
[272] Y. P. Myer, *J. Biol. Chem.*, 1968, **243**, 2115.
[273] F. Lederer, *Biochemistry*, 1968, **7**, 2168.
[274] A. V. Bertland and H. M. Kalckar, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **61**, 629.
[275] J. A. Verpoorte and C. Lindblow, *J. Biol. Chem.*, 1968, **243**, 5993.
[276] J. S. Cohen and O. Jardetsky, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 92.
[277] G. Navon, R. G. Shulman, B. J. Wyluda, and T. Yamane, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 86.
[278] A. M. Tamburro, A. Scatturin, and R. Rocchi, *Gazzetta*, 1968, **98**, 1256.
[279] E. H. Strickland, *Biochim. Biophys. Acta*, 1968, **151**, 70.
[280] E. Eizuka and J. T. Yang, *Biochemistry*, 1968, **7**, 2218.
[281] A. J. Geddes, K. D. Parker, E. D. T. Atkins, and E. Beighton, *J. Mol. Biol.*, 1968, **32**, 343.

some information not usually obtained from conventional transmission spectroscopy without damaging the protein.[282] α-Elastin has recently been shown to contain *ca.* 10% of α-helix.[283] The c.d. spectrum of gliadin indicates that it has a helix content of *ca.* 30%. Changes in bands at 277 and 294 nm. on aggregation suggest that the environments of the aromatic side-chains change in a specific way.[284] Several recent reports concern the optical activity of membranes and membrane fragments.[285, 286] The red shift and decreased magnitude of c.d. curves of such systems are in part the result of their particulate nature.[286]

The spectroscopic study of protein conformations has been reviewed recently both in general[5b] and with special reference to i.r. and c.d.[287] Conformational studies of proteins have also been discussed with respect to aromatic side-chain effects.[288]

PART III: *X*-Ray Analysis

by C. C. F. Blake

1 Oligopeptides

The structure of the glycylglycine molecule in the α-crystal form has been determined.[1a] The configuration is rather different from that in the previously reported β-crystal form,[1b] in that there is an angle of 22·4° between the peptide and carboxyl groups which in the β-form were coplanar. The bond lengths and angles are otherwise very similar in the two crystal forms and the values for the peptide group are in good agreement with those listed by Marsh and Donohue[2] in their review of amino-acid and peptide structures.

2 Globular Proteins

A. Methods.—The methods used for the structure determination of crystalline proteins are still in essence those used in the first two determinations, *i.e.* of haemoglobin and myoglobin. The principal developments have been made in the technological and crystallographic side of the subject: in *X*-ray sources, automatic data collection and, of course, in the general area of data handling and processing in computers. The location of heavy atoms is normally carried out by means of difference Pattersons

[282] C. B. Baddiel, *J. Mol. Biol.*, 1968, **38**, 181.
[283] M. Mammi, L. Gotte, and G. Pezzin, *Nature*, 1968, **220**, 371.
[284] D. D. Kasarda, J. E. Bernardin, and W. Goffield, *Biochemistry*, 1968, **7**, 3950.
[285] J. M. Wrigglesworth and L. Packer, *Arch. Biochem. Biophys.*, 1968, **128**, 790.
[286] D. W. Urry and T. H. Ji, *Arch. Biochem. Biophys.*, 1968, **128**, 802.
[287] S. N. Timasheff, H. Susi, R. Townsend, L. Stevens, M. J. Gorbunoff, and T. F. Kumosinki, in 'Conformation of Biopolymers,' vol. I, Academic Press, New York, 1967, p. 173.
[288] M. Goodman and C. Toniolo, *Biopolymers*, 1968, **6**, 1673; see also M. Goodman, G. W. Davis, and E. Benedetti, *Accounts Chem. Res.*, 1968, **1**, 275.

[1] [a] A. B. Biswas, E. W. Hughes, B. D. Sharma, and J. N. Wilson, *Acta Cryst.*, 1968, **A, 24**, 40; E. W. Hughes, *ibid.*, 1968, **B, 24**, 1128. [b] E. W. Hughes and W. J. Moore, *J. Amer. Chem. Soc.*, 1949, **71**, 2618.
[2] R. E. Marsh and J. Donohue, *Adv. Protein Chem.*, 1967, **22**, 235.

of various kinds which have been critically reviewed by Phillips.[3] Recently, however, the usefulness of direct methods has been investigated.[4] Sayre's equation was used to solve three carboxypeptidase heavy-atom derivatives successfully, one of which contains four sites, and the conclusion was drawn that the method may be better than difference Patterson in the situation where a number of heavy-atom sites exist. However, the solution of the two outstanding chemical problems, namely the production of suitably large crystals and the development of general methods for the preparation of isomorphous derivatives, still eludes us.

The isomorphous replacement method has emerged as the only means, at least at present, of solving the phase problem. All protein structures reported so far have been solved exclusively by its use. The preparation of isomorphous derivatives has been recently reviewed.[5] Methods of preparation fall into two distinct categories. The first, the systematic method, requires some chemical knowledge of the protein in hand, since it exploits certain specific binding groups on the protein. The scope of the method has been well defined in some recent protein structure determinations. The preparation of a complete set of derivatives by the binding of mercurials to sulphydryl groups of differential reactivity has been used for all three haemoglobin low-resolution structures [6-8] and for the 2·8 Å electron-density map.[9] One of the five derivatives used in the work on papain also made use of the sulphydryl group.[10] In the solution of the phase problem for α-,[11] δ-,[12] and γ- [13] chymotrypsins, extensive use has been made of the ability of the active serine to react irreversibly with heavy-atom-labelled sulphonyl fluorides. This work is closely related to the use of enzyme competitive inhibitors carrying heavy atoms, which strictly speaking has only been reported for carbonic anhydrase [14] and ribonuclease S.[15] Another method, of less general use than either sulphydryl reagents or labelled competitive inhibitors, has been reported in work on carbonic anhydrase [14] and carboxy-

[3] D. C. Phillips, *Adv. Struct. Res. Diffraction Methods*, 1966, **2**, 72.
[4] T. A. Steitz, *Acta Cryst.*, 1968, **B**, **24**, 504.
[5] C. C. F. Blake, *Adv. Protein Chem.*, 1968, **23**, 59.
[6] A. F. Cullis, H. Muirhead, M. F. Perutz, M. G. Rossmann, and A. C. T. North, *Proc. Roy. Soc.*, 1961, **265**, *A*, 15.
[7] H. Muirhead and M. F. Perutz, *Nature*, 1963, **199**, 633.
[8] W. Bolton, J. M. Cox, and M. F. Perutz, *J. Mol. Biol.*, 1968, **33**, 283.
[9] M. F. Perutz, H. Muirhead, J. M. Cox, L. C. G. Goaman, B. W. Matthews, E. L. McGandy, and R. L. Webb, *Nature*, 1968, **219**, 29.
[10] J. Drenth, J. N. Jansonius, R. Koekoek, H. M. Swen, and B. G. Wolthers, *Nature*, 1968, **218**, 929.
[11] B. W. Matthews, P. B. Sigler, R. Henderson, and D. M. Blow, *Nature*, 1967, **214**, 652.
[12] J. Kraut, H. T. Wright, M. Kellerman, and S. T. Freer, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 304.
[13] B. W. Matthews, G. H. Cohen, E. W. Silverton, H. Braxton, and D. R. Davies, *J. Mol. Biol.*, 1968, **36**, 179.
[14] B. Tilander, B. Strandberg, and K. Fridborg, *J. Mol. Biol.*, 1965, **12**, 740.
[15] H. W. Wyckoff, K. D. Hardman, N. M. Allewell, T. Inagami, D. Tsernoglou, L. N. Johnson, and F. M. Richards, *J. Biol. Chem.*, 1967, **242**, 3749.

peptidase.[16] These enzymes contain an integral zinc atom, which has been exchanged in each case for a mercury atom to give useful derivatives.

None of these methods, however, is of universal application; suitable metalloproteins are not common and although this cannot be said of sulphydryl groups, the low occurrence of the appropriate number which can be differentially labelled without undesirable side-effects such as conformational changes or subunit rearrangements may limit the usefulness of the method. This latter disadvantage may also be a particularly serious limitation to the exploitation of labelled competitive inhibitors; the binding of such compounds (without the heavy-atom label) to carboxypeptidase [17] causes shifts in structure of 2–14 Å. A change of this magnitude, if it took place on forming a heavy-atom derivative, could result in serious distortion of the resulting electron-density map in the region of most immediate interest, the active site. In order to overcome these limitations some attempts to find more general methods have been made recently. One of the most interesting approaches has been the covalent attachment of metal-chelating groups to lysozyme.[18] Seven picolinamidinyl groups could be introduced into the native enzyme, probably reacting with lysines and the amino-terminus, but after the lysines had been guanidinated only a single group was introduced. These groups were found to chelate palladium, platinum, and gold compounds very strongly. Iodination of tyrosines has also been tried with myoglobin,[19] papain,[10] and α-chymotrypsin,[11] but only in γ-chymotrypsin [13] has this method contributed to the phase determination.

The second method of preparation of derivatives is a trial-and-error method in which crystals are soaked in solutions containing heavy atoms. In spite of its simplicity, the method has been notably successful, accounting for most of the protein structures so far solved. With the detailed solution of structures at high resolution has come knowledge of the way in which the heavy atoms prepared thus are bound to the proteins. It has become apparent that in general four types of protein side-chains will bind heavy metals, to some extent selectively.

The first, and least common, type of binding is represented by the major mercuri-iodide site in seal [20] and sperm-whale myoglobin.[21] The nature of the bound group has been determined as the planar HgI_3^- ion. It is located in the nonpolar interior of the molecules, close to and parallel with the haem group. This site involves large side-chain movements since there is no cavity present in the native protein large enough to accommodate the ion. The forces responsible for the binding of the ion appear to be mainly

[16] W. N. Lipscomb, J. C. Coppola, J. A. Hartsuck, M. L. Ludwig, H. Muirhead, J. Searl, and T. A. Steitz, *J. Mol. Biol.*, 1966, **19**, 423.
[17] 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.
[18] W. Benisek and F. M. Richards, *J. Biol. Chem.*, 1968, **243**, 4267.
[19] R. H. Kretsinger, *J. Mol. Biol.*, 1968, **31**, 315.
[20] H. Scouloudi and J. W. Prothero, *J. Mol. Biol.*, 1965, **12**, 17.
[21] R. H. Kretsinger, H. C. Watson, and J. C. Kendrew, *J. Mol. Biol.*, 1968, **31**, 305.

polarization forces, because other groups bound at the same site all include highly polarizable atoms—AuI_4^-, I_3^-, and Xe.

The most common type of binding appears to be nucleophilic chelation by nitrogen-containing side-chains, notably histidine but including, in some cases, lysine, arginine, the α-amino and amide groups. Two of the major sites of mercury binding to papain [10] each involve one of two histidine residues in the molecule. This behaviour is similar to that described for myoglobin and lysozyme in which, in addition to mercury, gold, platinum, and palladium compounds were found to be bound at this type of site (see ref. 5).

The remaining two kinds of binding each seem to be more specific for the type of heavy atom. Uranyl compounds have been observed to bind at a total of five sites in lysozyme. Each site involves a protein carboxyl group and in each case a uranium–oxygen distance of *ca.* 3 Å was found.[5] This suggests that, as in model compounds, a carboxyl oxygen is incorporated into the co-ordination sphere of the uranium. The binding of platinum complexes to ribonuclease S [22] and to α-chymotrypsin [23] also indicates a considerable tendency to interact with the sulphur atoms of methionines and disulphide bridges. One such compound K_2PtCl_4 makes a covalent bond to methionine 29 in ribonuclease S. However, platinum compounds may also be bound at nucleophilic sites (perhaps when methionine and cystine residues are unavailable because they are buried in the interior of the protein) whereas uranyl complexes do not appear to show this tendency. Nevertheless, it seems that by suitable choice of heavy-atom compounds a variety of potential binding sites may be examined, thus making the trial-and-error method more efficient than it has been to date.

B. Results.—*Chymotrypsin.* (See also p. 87.) In a series of papers by Blow and his colleagues the structure of chymotrypsin [11, 23] and a hypothesis of its activity [24] have been reported. A parallel series of papers by two American groups [12, 13, 25] have provided the structural basis of the interrelationships between the various chymotrypsins and of their mode of derivation from the inactive precursor, chymotrypsinogen.

Bovine α-chymotrypsin is a protease containing 241 amino-acids whose sequence has been determined. The enzyme is derived from the inactive zymogen, chymotrypsinogen A, by the stepwise enzymatic cleavage of four peptide bonds, with the elimination of two dipeptides to form a three-chain molecule. The three polypeptide chains (A chain, residues 1–13; B chain, residues 16–146; C chain, residues 149–245) are held together by two interchain disulphide bridges, residues 1–122 and 136–201,

[22] H. W. Wyckoff, K. D. Hardman, N. M. Allewell, T. Inagami, L. N. Johnson, and F. M. Richards, *J. Biol. Chem.*, 1967, **242**, 3984.
[23] P. B. Sigler, D. M. Blow, B. W. Matthews and R. Henderson, *J. Mol. Biol.*, 1968, **35**, 143.
[24] D. M. Blow, J. J. Birktoft, and B. S. Hartley, *Nature*, 1969, **221**, 337.
[25] H. T. Wright, J. Kraut, and P. E. Wilcox, *J. Mol. Biol.*, 1968, **37**, 363.

while there are three intrachain bridges, one (residues 42–58) in the B chain and two (residues 168–182 and 191–220) in the C chain.

The enzyme is known to contain an essential serine, residue 195, which may be specifically reacted with sulphonyl fluorides. This reaction has formed the basis for the preparation of the heavy-atom derivatives.[11] A pair of compounds with very high isomorphism was prepared by reacting the active serine with toluene-*p*-sulphonyl fluoride (the tosyl enzyme) and *p*-iodophenylsulphonyl fluoride (pipsyl fluoride) respectively. Additional derivatives were obtained by adding phenylmercuric acetate to the tosyl enzyme and potassium tetrachloroplatinate(II) to the native enzyme. The use of two parent compounds, the native and the tosyl enzyme, in the phase determination has caused certain deficiencies in the electron-density map.

The crystal form employed in the structure determination is monoclinic, containing two molecules in the asymmetric unit related by local twofold axes. These dimers found in the crystal may correspond to the dimers known in solutions of α-chymotrypsin at pH 4, the same pH value at which the crystals are grown.

The electron-density map showed the expected local twofold axes parallel to the crystallographic *a* axis and perpendicular to chains of molecules running in the C direction. The influence of the dyads does not, however, extend to the whole of each molecule, but deteriorates noticeably in parts of the molecules remote from the dyads. The electron-density map used for the interpretation of the molecular architecture was obtained by averaging the two independent molecules in the asymmetric unit. The course of the polypeptide chains could be traced and many side-chains identified. Among the more prominent of these were the five disulphide bridges and the tosylated serine residue. However, the electron density representing residues 30–40, 70–91, and 172–179 is not clearly defined, possibly because the local twofold symmetry does not hold in these regions.

The conformation of the polypeptide chains (Figure 1) is extremely complex. With the exception of eight residues at the carboxyl terminus of the C-chain which are α-helical, the chains tend to be fully extended. It is frequently found that segments of chain several residues in length run parallel to one another *ca.* 5 Å apart. This suggests that hydrogen bonds, possibly linking the chains in β-pleated sheet conformations, are largely involved in stabilizing the molecule.

There is no pronounced cleft in the molecule, marking the active site, unlike, for example, lysozyme [26] and ribonuclease.[22] There are, however, open regions in the surface of the molecule around the active serine where the structure is not closely packed. His-57 is found in the active site within hydrogen-bonding distance of the hydroxyl of the active serine. The role

[26] C. C. F. Blake, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *Proc. Roy. Soc.*, 1967, *B*, **167**, 365; C. C. F. Blake, L. N. Johnson, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *ibid.*, 1967, **167**, *B*, 378.

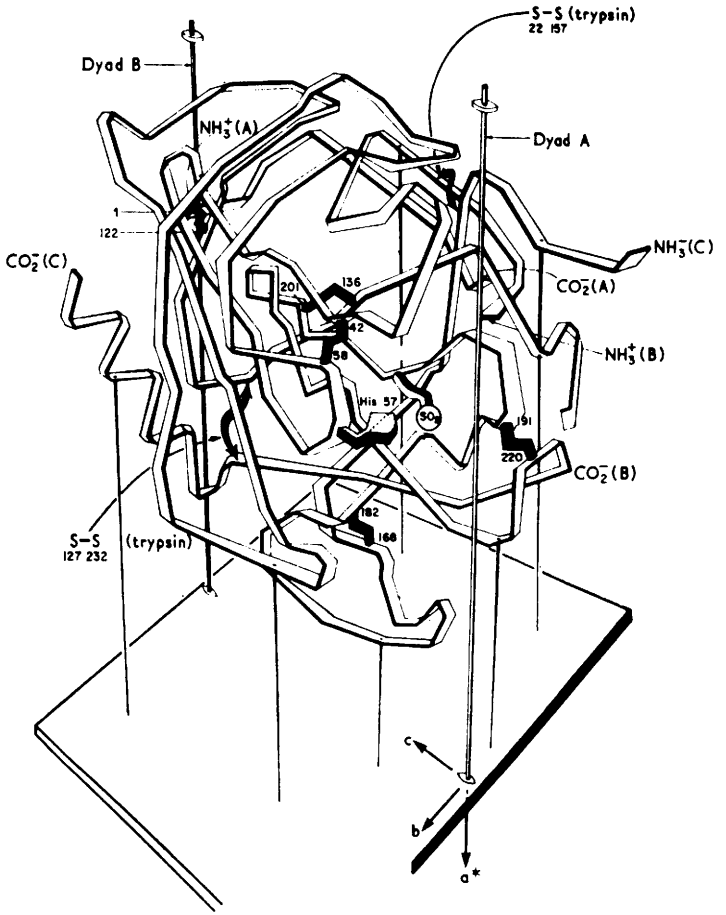


Figure 1 *The conformation of the polypeptide chains of α-chymotrypsin. The active site residues* His-57, *and* Ser-195 *with the sulphonate group of the inhibitor attached to it, are shown. Also shown in the diagram are the two additional disulphide bridges of the homologous enzyme trypsin*
(Reproduced by permission from *J. Mol. Biol.*, 1968, **35**, 143)

of Asp-194 is very specific. The carboxylate group of this residue is deeply buried in the interior of the molecule forming an ion pair with the positively charged amino-terminal residue of the B-chain, Ile-16. Both groups appear to be further stabilised by hydrogen bonds to a nearby peptide chain and a neighbouring serine residue. Because it is the production of this *N*-terminus which confers activity on the zymogen, it is presumably the strong ionic interaction which stabilizes the active conformation of the enzyme.

The local twofold axes relating the two molecules in the asymmetric unit are of two types, as shown in Figure 1, which alternate along the C axis. One type of axis (shown at the front of Figure 1) relates two chymotrypsin molecules so that their active sites are in close contact, only *ca.* 12 Å separating the sulphonyl groups of the related tosylated molecules. Among the interactions in this region, the α-carboxylate of tyrosine-146 hydrogen bonds to the peptide unit 57–58, while its phenolic group is close to the imidazole ring of His-57.[23] These interactions do not seem to interfere with the enzymatic activity as might be expected, since the crystalline enzyme is still active. The interaction about this local dyad may therefore represent the dimers of α-chymotrypsin found in solution.

The model of chymotrypsin has led to a proposal for the mechanism of the enzymatic activity,[24] largely through the identification of residue 102 as an aspartic acid. This residue, which in both sequence determinations was identified as an asparagine, is located in the molecule in a hydrophobic environment buried beneath His-57. Reinvestigation of the peptide containing residue 102 has shown that it is in fact an aspartic acid, as are the equivalent residues in the homologous proteins, chymotrypsinogen-B, trypsinogen, thrombin, and elastase.

The important side-chains in the active site are shown in Figure 2. Asp-102, His-57, and Ser-195 are linked by a system of hydrogen bonds, which allows the negative charge on Asp-102 to be relayed to Ser-195. The buried nature of Asp-102 does not allow water molecules to interfere with the relay system. The negatively charged oxygen of Ser-195 becomes a powerful nucleophile capable of attacking the carbon atom of the carbonyl group in a substrate. Figure 3*a* shows the acylation step, which disrupts the relay system because the proton linking His-57 and Ser-195 has been removed by the leaving group. In the deacylation step (Figure 3*b*) a water molecule replaces the leaving group re-establishing the relay system which can reverse the electron shifts that occurred in acylation, with the elimination of the carboxylic acid. This mechanism could function at pH 8, the pH of maximal activity of chymotrypsin, while a lowering of pH would result in the protonation of His-57, thus blocking the charge pathway between Ser-195 and His-57 and inactivating the enzyme.

The model of the α-chymotrypsin molecule in itself gives certain evidence on the mode of activation of the zymogen to the active enzyme. The four chain termini which arise during the activation process are in positions consistent with the cleavage of dipeptides from the surface of the zymogen without fundamentally altering its structure. The stereochemistry of the activation of the zymogen appears to be paralleled in the enzyme by a pH-dependent structural transition. In both transitions activity depends on the integrity of the ion pair between Ile-16 and Asp-194. The structural transition occurs at pH 8, above which the enzyme becomes inactive. The group with the pK_a 8–9 which controls the transition is the α-amino of Ile-16. When this amino-group loses its proton, the carboxylate of Asp-194

is released from its buried position to protrude into the active site, destroying its integrity and activity. This also appears to be the situation in the zymogen.

The structural relationships between the various chymotrypsins and the parent zymogen have been the concern of two papers from Kraut's group,[12, 25] which show elegantly the amount of information deducible from low-resolution *X*-ray maps when a comparable high-resolution

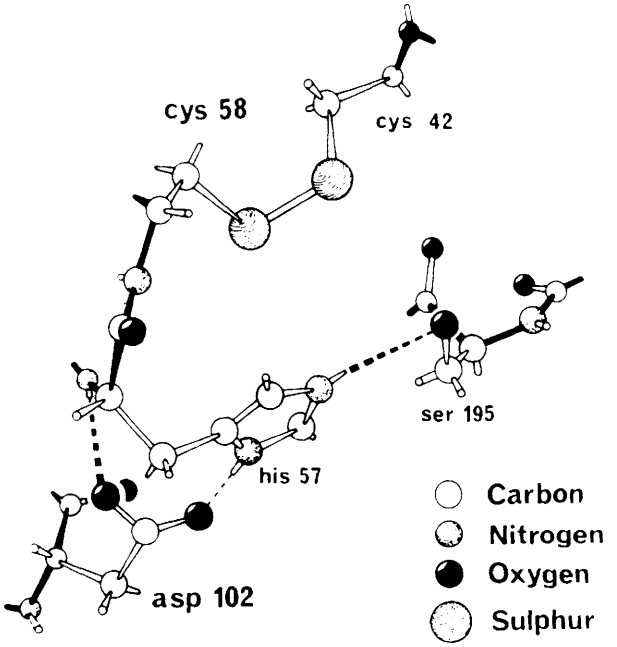


Figure 2 *The arrangement of the side-chains in the active site of α-chymotrypsin. The dotted lines are between atoms within hydrogen bonding distance*
(Reproduced by permission from *Nature*, 1969, **221**, 337)

structure is available. They have suggested a new scheme relating the various forms of chymotrypsin with the zymogen, which is shown in Figure 4.

The low-resolution structures used were a new 5 Å map of chymotrypsinogen calculated from previously collected intensity data after elimination of two of the most questionable heavy-atom derivatives and an entirely new 5 Å resolution map of δ-chymotrypsin inhibited with benzylsulphonyl fluoride (PMS). It was also found that crystals of PMS-π-chymotrypsin, PMS-γ-chymotrypsin, and γ-chymotrypsin itself were all isomorphous with PMS-δ-chymotrypsin although the intensities of each differed slightly, indicating small structural differences.

The maps of the zymogen and the inhibited δ-form of the enzyme indicated that the two molecules were essentially similar. There were, however, obvious differences in two particular regions. In the first a chain which is continuous in the zymogen is broken in the enzyme. The second difference is concerned with a loop of chain which moves to create a pocket in the enzyme structure. This pocket is the site where the specific serine

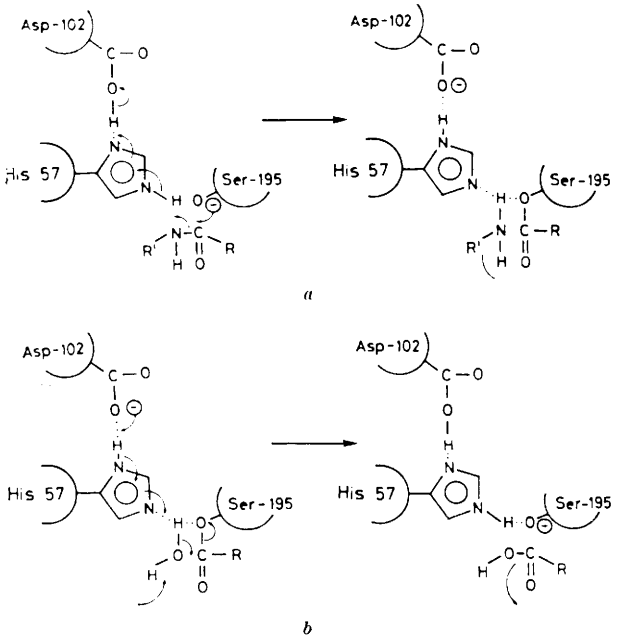


Figure 3 *The hypothesis of the mechanism of the activity of α-chymotrypsin suggested by X-ray studies.* (a) *The acylation step.* (b) *The deacylation step* (Reproduced by permission from *Nature*, 1969, **221**, 337)

reagents are bound. Comparison of the structure of α-chymotrypsin shows that it is identical in gross chain conformation with the δ-enzyme.

The isomorphism of the inhibited π-, δ-, and inhibited and native γ-chymotrypsins allows more detailed comparisons between them to be made by means of difference maps. These maps show the position of the dipeptide removed autocatalytically in going from the π- to the δ-enzyme. This position corresponds to the break in the chain found in the comparison of the zymogen structure with PMS-δ-chymotrypsin. The difference map between the δ- and γ-inhibited enzymes shows clearly the position of the dipeptide Thr-147—Asn-148. Thus it appears that the activation of the zymogen to form a series of active enzymes is not accompanied by any

extensive conformational change and that the essence of the activation process is the formation of the ion pair between Ile-16 and Asp-194.

These general conclusions have received additional support from the low-resolution study of γ-chymotrypsin.[13] A 5·5 Å electron-density map of this enzyme has been calculated using the pipsyl, tosyl, 3-chloromercuri-4-methoxy-benzenesulphonyl and iodinated-γ-chymotrypsins as derivatives. This map was compared with the high-resolution α-chymotrypsin

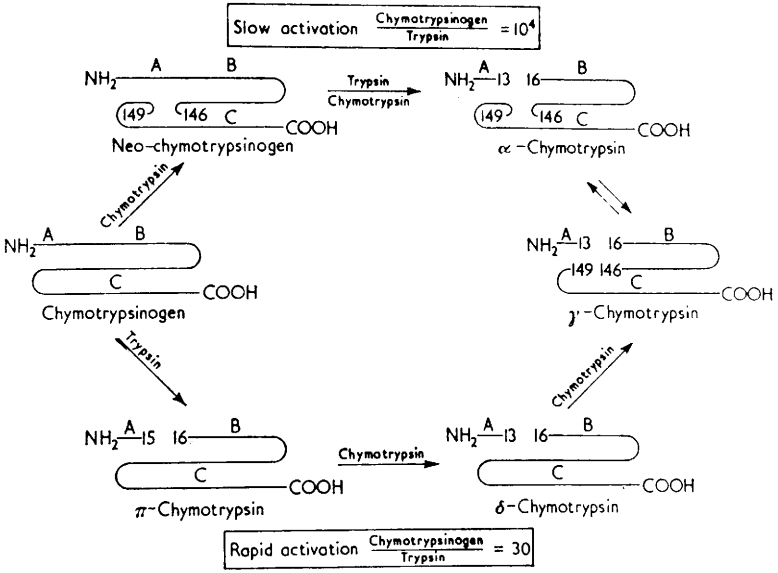


Figure 4 *The activation process of chymotrypsinogen suggested by* X-*ray studies* (Reproduced by permission from *J. Mol. Biol.*, 1968, **37**, 363)

map by calculating the transformation relating the tetragonal γ-chymotrypsin crystal form to the monoclinic crystal form of the α-enzyme. After transformation one of the sites of iodination in the γ-enzyme is only 3 Å from Tyr-171 of α-chymotrypsin, while the other two sites of iodination (consistent with di-iodination of a tyrosine) are not far from Tyr-146. The slight difference in position can be explained by a rotation and translation of Tyr-146 in γ-chymotrypsin from the position of this residue found in the α-form of the enzyme. The shift in position of this residue appears to be the main difference between the interconvertible α- and γ-chymotrypsins and also, in view of its involvement in the dimerisation of α-chymotrypsin, it may explain why the aggregation behaviour of these two very similar enzymes differs so markedly.

Haemoglobin. (See also p. 100.) The horse oxyhaemoglobin molecule is made up of four subunits, chemically identical in pairs; two α-chains and two β-chains. Low-resolution *X*-ray studies have shown that the four

subunits are arranged in a slightly distorted regular tetrahedron,[27] and that each type of subunit is essentially similar in gross conformation to the sperm-whale myoglobin molecule.[28]

Both the α- and β-chains carry a single sulphydryl group; the one on the α-chain, residue 104, is considerably less reactive than residue 93 in the β-chain. These sulphydryl groups are the loci of the heavy-atom derivatives [9] used for calculating the 2·8 Å resolution map.[29] One derivative was formed by the reaction of Cys-93β with *p*-chloromercuribenzoate (PCMB), the second by the reaction of 104α with mercuric acetate, after Cys-93 had been blocked with iodoacetamide. A third derivative was prepared by the reaction of Cys-104α and 93β with methylmercury ions and PCMB respectively.

The overall shape of the oxyhaemoglobin molecule is a spheroid of dimensions $64 \times 55 \times 50$ Å. Neighbouring molecules are in contact at only a few points and the salt solution fills the wide interstices between them and also the cavity in the centre of the molecule. With the aid of the sequence nearly every amino-acid of the 141 in the α-chains and the 146 in the β-chains were located. The haem groups appeared as well-defined discs surrounding a central peak representing the iron atom. The reactive Cys-93β is readily accessible from the interstitial liquid whereas the less reactive Cys-104α is enclosed by other residues on all sides and is accessible only from the central cavity.

The interpretation of the structure of the haemoglobin chains at high resolution indicates that, in detail, several features distinguish the α- and β-subunits from one another and from myoglobin. Figure 5 shows the gross arrangement of the helical and non-helical segments of the myoglobin chain. The differences between the α- and β-chains and myoglobin are expressed in terms of the myoglobin structure and use the same system of labelling the helical and non-helical sections of the chain.

The α-chains contain 141 residues, compared with 153 in myoglobin. Of the twelve missing residues, six lie in helix D which is absent in the α-chain and the remaining six are lost from the carboxyl terminus. The β-chains contain 146 residues; in comparison with myoglobin there is one additional residue in the NA segment, two fewer at the AB corner, and six fewer at the carboxyl terminus.

The non-helical amino-terminal section of the myoglobin molecule contains two residues as do the α-chains while the β-chains contain an additional residue. The 16-residue A helix of myoglobin is incorporated into the α- and β-chain without alteration. In myoglobin the AB region is a right-handed corner in which all residues are helical, while the α- and β-chains each turn this corner in different ways. The next helix (B) and the

[27] A. F. Cullis, H. Muirhead, M. F. Perutz, M. G. Rossmann, and A. C. T. North, *Proc. Roy. Soc.*, 1962, *A*, **265**, 161.
[28] J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore, *Nature*, 1960, **185**, 422.
[29] M. F. Perutz, H. Muirhead, J. M. Cox, and L. C. G. Goaman, *Nature*, 1968, **219**, 131.

following non-helical section (BC) are the same for all three chains. The short C helix is in the $3{\cdot}0_{10}$ conformation in both haemoglobin chains in distinction to the distorted α-helix conformation found in myoglobin. The next chain segments, the CD corner and the short D helix, are completely altered in the α-chain because of the absence of the entire D helix,

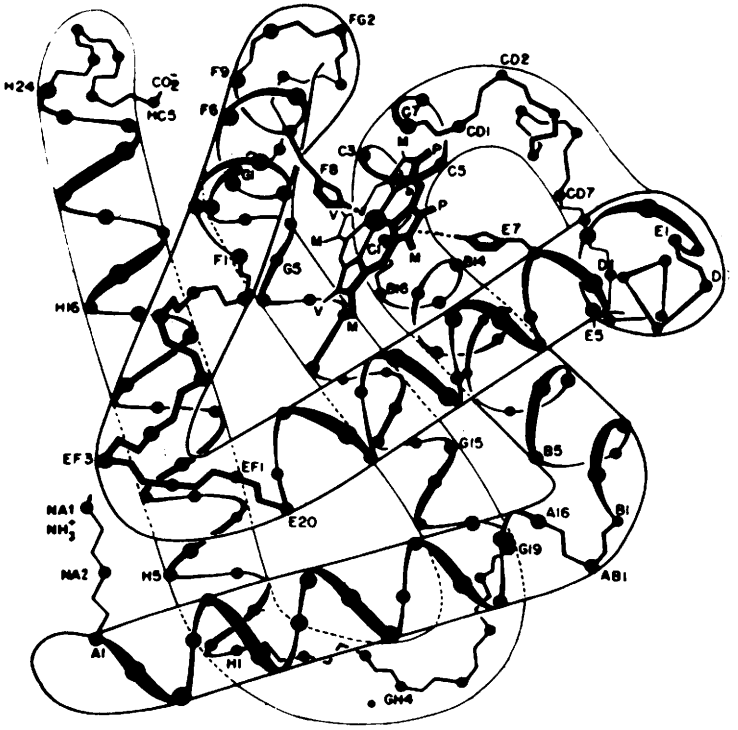


Figure 5 *A schematic diagram of the myoglobin molecule in which all side-chains, with the exception of the distal and proximal histidines, have been omitted for clarity*
(Reproduced by permission from 'The Proteins', 2nd edn., vol. 2, Academic Press, New York, 1964, p. 603)

while the β-chain conformation is not very different from myoglobin. The long E helix which carries the distal histidine contains a kink of 7° but is an otherwise regular α-helix in myoglobin. In the α-chain this helix is irregular throughout its length, but the irregularity is confined to the last three residues in the β-chain. In both haemoglobin chains the EF region is similar but differs somewhat from myoglobin. The segments F, FG, G, and GH are essentially similar in all three chains. The 26-residue H helix of myoglobin is shortened by 5 residues in both α- and β-chains, while the 4-residue HC tail of myoglobin consists of only 3 residues in haemoglobin.

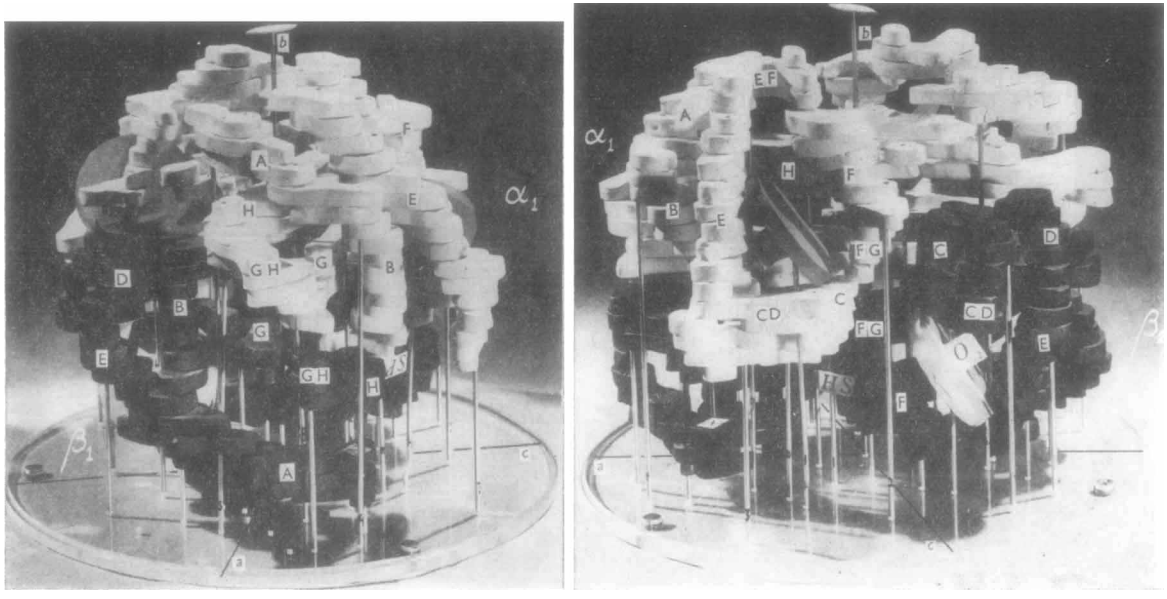
The environment of the haem groups has a similar nonpolar character to that found in myoglobin.[28] There are *ca.* 60 interactions between globin atoms and haem atoms, not counting the proximal histidine, of which all but one in the α-, and two in β-chains, are nonpolar. Comparison of the available sequences of mammalian haemoglobins shows almost total invariance of the residues surrounding the haem groups.

In a tetrameric molecule such as haemoglobin, each subunit may be in contact with the other three subunits. Thus an individual α-chain makes contacts of a different nature with the two β-chains and also makes contacts with the other α-chain. In addition, the β-chains can interact with one another. Thus four kinds of contacts, $\alpha_1\beta_1$, $\alpha_1\beta_2$, $\alpha_1\alpha_2$, and $\beta_1\beta_2$, need to be considered to fully define the stabilisation of the quaternary structure of haemoglobin.

Figure 6 shows the areas of contact between unlike chains; the residues involved are listed in Table 1. The most extensive is the $\alpha_1\beta_1$ contact: *ca.* 110 atoms of 34 residues (Table 1*a*) interact with one another. The great majority of these interactions are of a nonpolar nature. The contact $\alpha_1\beta_2$ is made up of 19 residues contributing *ca.* 80 atoms (Table 1*b*), nearly all the interactions of which are again nonpolar. Of the 34 residues making the $\alpha_1\beta_1$ contact, 21 are common to all known mammalian haemoglobin sequences, and all but 1 of the 19 residues forming the $\alpha_1\beta_2$ contact are also common; the one exception would not affect the stereochemistry of the contact. In contrast, no contacts between like chains are visible on the map, but they probably exist at low salt concentration in the form of salt bridges involving terminal residues.

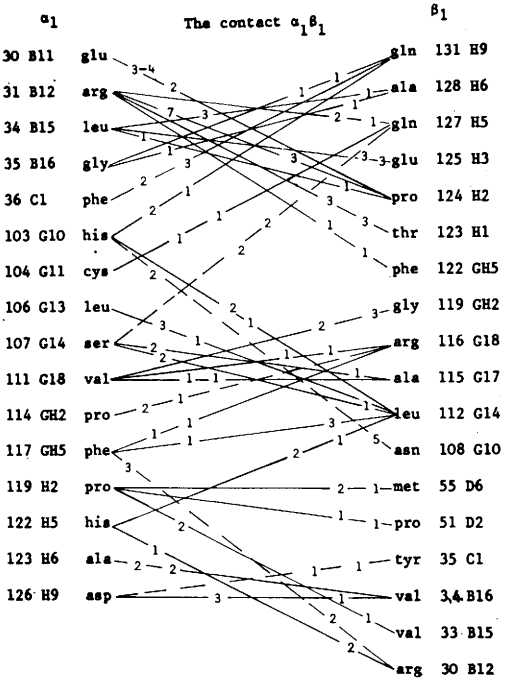
In the transition from oxy- to deoxy-haemoglobin some of the contacts are known to undergo alteration, with little or no change in the conformation of the individual subunits.[8] The extensive $\alpha_1\beta_1$ contact would change little during the transition. This contact appears to contain many interlocked side-chains which may be expected to translate a small change in secondary and tertiary structure into a larger change in quaternary structure. On the other hand, movement of the $\alpha_1\beta_2$ contact is large, and some atoms previously in contact may be displaced by up to 6 Å. The contact is both smaller and smoother and may allow subunits to slide past one another. It is closely connected to the haem groups, so that changes in the area of contact could be expected to change their environment. On deoxygenation the contact between the β-chains is broken, and that between the α-chains is altered.

In haemoglobin solutions the tetramer is in dynamic equilibrium with monomers and dimers. Dissociation into dimers occurs symmetrically so that $\alpha^2\beta^2$ splits into $2\alpha\beta$. Dissociation is favoured by high concentrations of neutral electrolyte which would be expected to weaken the polar contacts between the like subunits while strengthening the predominantly nonpolar contacts between unlike subunits. There have been indications that the $\alpha\beta$ dimers exhibit all the co-operative effects of which haemoglobin is

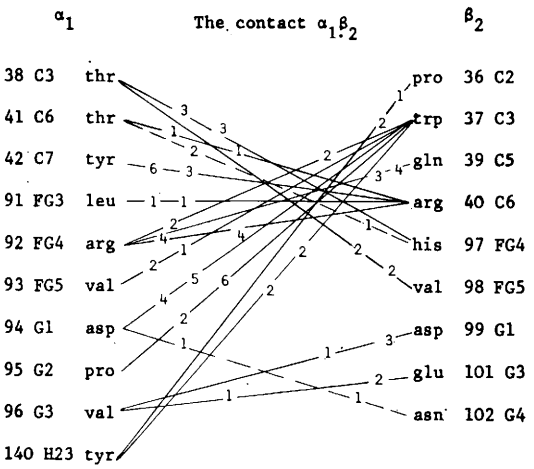


(a) (b)

Figure 6 *The areas of contact between the unlike subunits of haemoglobin, shown on the low-resolution model of the molecule.* (a) *The* $\alpha_1\beta_1$ *contact of haemoglobin.* (b) *The* $\alpha_1\beta_2$ *contact*
(Reproduced by permission from *Nature*, 1968, **219**, 131)



34 residues, including about 110 atoms in contact



19 residues, including about 80 atoms in contact

Table 1 *The residues involved in the unlike subunit contacts of haemoglobin. Residues in contact are joined by the full line when the interaction is non-polar and by the broken line when hydrogen bonds are involved. The number of atoms of each residue contributing to each contact is shown.* (Reproduced by permission from *Nature*, 1968, **219**, 131.)

capable and that interaction between dimers is of only secondary importance. It is probable that the dimers will have the structure $\alpha_1\beta_1$ because this contact is more extensive than the $\alpha_1\beta_2$. However, the pathway between the α_1 and β_1 haem groups is long and somewhat tenuous, while the $\alpha_1\beta_2$ contact offers a direct pathway for haem interactions. Thus the crystallographic evidence suggests that the tetramer, rather than the dimer, is the functional unit essential for the full expression of the co-operative effects, but that the structure of deoxyhaemoglobin at high resolution will be needed before the co-operation can be explained.

The structure determinations of myoglobins from sperm whale [28] and seal,[30] and of horse [29] and human [7] haemoglobin have provided an interesting example of structural and evolutionary relationships between molecules of similar, though not identical, function. The recent reports of low-resolution structures of two non-vertebrate haemoglobins, from the insect *Chironomus* [31] and the annelid worm *Glycera*,[32] should give us information on the evolution of molecules over a considerably longer period of time than that given by the mammalian species. In addition, since both these haemoglobins are monomeric molecules they may assist in defining both the functional and evolutionary relationships between themselves, myoglobin, and the tetrameric mammalian haemoglobins.

Both invertebrate haemoglobins show a marked similarity in tertiary structure to myoglobin and the subunits of mammalian haemoglobin. A considerable amount of analysis has been reported for the *Glycera* haemoglobin. It is suggested that the main differences between it and myoglobin are an increase in length of the F helix (see Figure 5) and a different chain orientation in the rather short C and D helices. In other respects the lengths and orientation of the helices are very similar. When the structures of these two proteins have been taken to higher resolution and complemented with sequence information, we may obtain a great deal of information about 'the myoglobin fold' and its role in causing the haem group to undergo reversible oxygenation. In addition, it may also be an important step in discovering precisely which factors are responsible for causing polypeptide chains of different sequence to coil into the characteristic 'myoglobin fold'.

Papain. (See also p. 93.) In a preliminary report [10] the structure of the sulphydryl protease papain and its relation to the previously reported sequence have been determined. The crystals, grown from methanol–water (2 : 1), are orthorhombic, space group $P2_12_12_1$, with cell dimensions $a = 45{\cdot}0$, $b = 104{\cdot}3$, $c = 50{\cdot}8$ Å containing one molecule in the asymmetric unit. The electron-density map was calculated at 2·8 Å resolution by using five heavy-atom derivatives to obtain the phase information.

[30] H. Scouloudi, *Proc. Roy. Soc.*, 1960, *A*, **258**, 81.
[31] R. Huber, H. Formanek, and O. Epp, *Naturwiss.*, 1968, **2**, 75.
[32] E. A. Padlan and W. E. Love, *Nature*, 1968, **220**, 376.

The map showed the main chain without interruption, the three disulphide bridges, and the active sulphydryl group which was located at the point of higher density in the map. In order to interpret the map in terms of the reported sequence, extensive rearrangement of the latter was needed together with the addition of 11 extra residues, bringing the total to 211. Residues 1–28 were found to be correctly placed. Thirteen residues were needed between Phe-28 and Arg-31 instead of only Ile-29 and Ile-30 and the peptide 138–176 had to be placed after the new 13-residue stretch and Arg-31. Other minor discrepancies also exist between the *X*-ray image and the sequence at a few other places.

The polypeptide chain (Figure 7) is folded in two distinct parts to the left and right of a cleft which contains the active site. The right-hand part, made up of residues 1–17 and 110–211 (numbered in the *X*-ray sequence), is much more hydrophobic than the left-hand part. It contains one short α-helix, residues 116–126, and a small segment of β-structure, residues 163–172, together with the disulphide bridge 152–199. The left-hand part contains three α-helices consisting of residues 26–41, 50–56, and 69–78 and the two other disulphide bridges, 22–63 and 56–95. This part of the molecule also contains some salt bridges of the arginine–carboxylate type, one of which is internal while four others are more or less on the surface. In the lysozyme [26] and carboxypeptidase [17] molecules the active site is flanked by an essentially hydrophobic area on the one side and a more extended hydrogen-bonded region on the other. In these two molecules, conformational changes associated with activity have been observed to take place only in the extended chain areas; it will be interesting to see if this is also the case in papain.

The active site was located by adding α-tosyl-lysyl chloromethyl ketone, an irreversible inhibitor of the enzyme, and calculating a difference map between the inhibited and native enzyme. The essential Cys-25 and the histidine, residue 158, which was known to be in the active site, were found *ca.* 4 Å apart on opposite sides of the cleft. Other residues in this general region include Trp-176, Gln-19, Asp-157, and Asp-64.

Aspartate Transcarbamylase. (See also p. 97.) An elegant demonstration of the ability of preliminary *X*-ray analysis to define quaternary structure has been reported [33] for aspartate transcarbamylase. This enzyme, which has a molecular weight of 310,000, catalyses the first unique reaction in the pathway of pyrimidine synthesis. It exhibits co-operative effects and is subject to feedback inhibition by cytidine triphosphate (CTP), an end-product of the pathway. The enzyme has been dissociated into two kinds of subunit, one of which catalyses the reaction, while the other binds the inhibitor CTP.

Two crystal forms have been observed. One is trigonal, space group *P*32, with two molecules in the cell lying on threefold axes. The second is

[33] D. C. Wiley and W. N. Lipscomb, *Nature*, 1969, **218**, 1119.

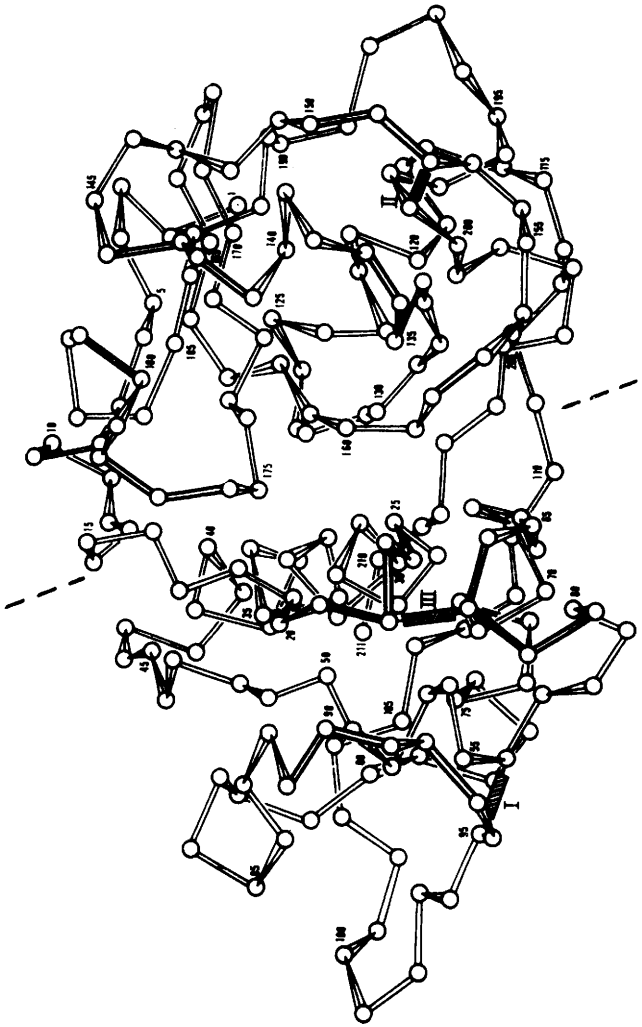


Figure 7 *The conformation of the polypeptide chain of papain. The α-carbons are represented by circles and the disulphide bridges labelled with Roman numerals. The run of the cleft which contains the active site is defined by the broken line*
(Reproduced by permission from *Nature*, 1968, **218**, 929)

tetragonal, either $P422$ or $P4_322$ with four molecules in the cell lying on twofold axes. In order to satisfy the molecular symmetry exhibited in these two crystal forms, each aspartate transcarbamylase molecule must contain at least six copies of each of the catalytic and regulatory subunits, a conclusion which is supported by the latest chemical data.[34]

Human Immunoglobulin G. (See also p. 104.) The report[35] of fairly large crystals of an immunoglobulin makes this, together with aspartate transcarbamylase, one of the potentially most interesting and exciting projects in protein crystallography. Immunoglobulin G from normal serum is heterogenous, but some patients with multiple myeloma produce homogeneous G myeloma protein, which can be crystallized. The crystals are monoclinic, space group $C2$ with cell dimensions $a = 194{\cdot}6$, $b = 92{\cdot}7$, $c = 51{\cdot}1$ Å and $\beta = 101° 7'$, and they contain two molecules lying on twofold axes, indicating that each immunoglobulin G molecule has a twofold axis of symmetry. The cell dimensions show that the molecule cannot be spherical because the C-axis is too short, but they do not exclude other proposals for the shape of the molecule.

In a similar preliminary investigation[36] of crystals of the C-terminal halves of the heavy chains of IgG myeloma protein (the so-called Fc fragment), the molecules were also found to lie on twofold axes. The crystal form of the native Fc fragment is $P2_12_12_1$, which contains no crystallographic twofold axes, but Patterson projections indicate *pseudo*-$A2_122$ symmetry. When crystals were treated with *o*-chloromercuriphenol, or with $KAuCl_3$ (*sic*) and partially dried, the symmetry became exactly $A2_122$, without much change in cell dimensions. In this latter space group the molecules lie on twofold axes, which in the native form are therefore local, non-crystallographic axes.

3 Fibrous Proteins

Muscle (see also p. 68)—The results of recent X-ray diffraction studies of muscle have been reviewed by Hanson.[37] The introduction of new X-ray cameras and the use of electron microscopy to complement X-ray results have given a considerable amount of data on this very complex system. Use of a toroidal mirror focusing system has indicated that the accepted two-strand coiled coil model of paramyosin is not correct.[38] A meridional reflection at 1·485 Å and a near-equatorial streak are not compatible with either two- or three-strand coiled coils in which the chains are parallel or antiparallel. It seems probable that side-chain interactions affect the diffraction pattern to a considerable extent and it is suggested that a model needs

[34] K. Weber, *Nature*, 1968, **218**, 1116.
[35] W. D. Terry, B. W. Matthews, and D. R. Davies, *Nature*, 1968, **220**, 239.
[36] D. J. Goldstein, R. L. Humphrey, and R. J. Poljak, *J. Mol. Biol.*, 1968, **35**, 247.
[37] J. Hanson, *Quart. Rev. Biophys.*, 1968, **1**, 177.
[38] A. Elliott, J. Lowy, D. A. D. Parry, and P. J. Vibert, *Nature*, 1968, **218**, 656.

to be developed which has different symmetries for the backbone and side-chains. In another paper concerned with side-chain symmetry,[39] it has been shown that heavy-metal stains and glutaraldehyde fixative reveal a 48 Å spacing in the thick filaments of molluscan catch muscle. These filaments have a well-known spacing of 145 Å, and it is supposed that lysines and other charged groups (which could combine with the heavy metals and the fixative) occur in register throughout the filaments separated axially by 48 Å. This fits in with the suggestion that charged groups are situated on the outside of the coiled coils of paramyosin.

The contractile apparatus of muscle is clearly of very great interest. The sliding-filament theory of contraction appears to have been confirmed, but the molecular mechanism has still not been settled. In insect flight muscle, the cross bridges have been implicated in the contractile process.[37] Recently, comparison of *X*-ray diffraction and electron micrographs of living striated muscles with those of muscles in rigor at the same sarcomere length have also suggested that the cross bridges are directly involved in the contractile process.[40] The cross bridges, which are believed to represent the heavy meromyosin part of the myosin molecules, are not attached to the actin filaments in relaxed muscle but in rigor they are extended farther out to combine with sites on the actin filaments. This model implies that contraction is brought about by structural changes which alter the orientation of the attachment of heavy meromyosin to actin.

That muscle is not simply a mechanical system, but behaves to some extent as a colloid or a liquid crystal, is suggested in a paper by Rome.[41] The low-angle equatorial *X*-ray reflections from the filament lattice of vertebrate striated muscle have been investigated as a function of sarcomere length and the nature of the bathing medium. It was found that the myosin–myosin centre-to-centre separation varied between 310 and 510 Å according to the bathing medium. The dependence of the filament separation on the electrical parameters of the inter-filament medium indicates the existence of long-range forces between the filaments. Thus muscle appears to behave as a colloid in which the filament separation is determined by the balance of electrostatic double-layer repulsive forces and van der Waals attractive forces. It remains to be seen if these observations have any bearing on the contractile process.

[39] A. Miller, *J. Mol. Biol.*, 1968, **32**, 687.
[40] H. E. Huxley, *J. Mol. Biol.*, 1968, **37**, 507.
[41] E. Rome, *J. Mol. Biol.*, 1968, **37**, 331.

3
Peptide Synthesis

BY J. H. JONES

1 Introduction

The first *ab initio* syntheses of material with enzymic properties (see p. 201) constitute a major landmark in the development of peptide synthesis. The role of peptide synthesis in protein studies has hitherto been largely peripheral, but now we stand on the threshold of an era in which more direct applications may be possible. We can anticipate investigations using synthetic analogues of natural proteins, and, when the factors involved in the control of structure and function become more intimately understood, the synthesis of enzymes specifically designed for the catalysis of particular reactions. It must be emphasised, however, that ribonuclease is one of the simplest possible enzymes for synthetic attack, not only because of its comparatively low molecular weight but also because the generation of higher orders of structure from the primary sequence is spontaneous. It is likely that many years (decades ?) of arduous effort in the development of synthetic methods will be required before the production of synthetic enzymes can become as commonplace as the preparation of synthetic oligopeptide hormones is today. Of particular importance is the need for procedures which are entirely devoid of ambiguity, and which can be used for the preparation of synthetic proteins in a pure condition; experiments with impure materials will be of uncertain value, because contaminants with erroneous amino-acid sequences might themselves be catalytically active, or might enter into non-productive association with substrates. This requirement is a very stringent one, and methods of approach which differ radically from those currently available may have to be employed. It may even be that advances in nucleotide chemistry will overtake improvements in peptide synthesis: the best way of obtaining synthetic enzyme analogues might then be by instruction of biological protein-synthesising systems with artificial messenger molecules.

Several other important synthetic achievements were reported during the year, and many valuable refinements in synthetic methods have been described. This chapter is confined to methods and results in the synthesis of peptides in which the amino-acids are linked together only by peptide and disulphide bonds: the synthesis of other types of peptide is reviewed in Chapter 4 of this volume.

2 Methods

A. Protective Groups.—As the aims of peptide synthesis become more ambitious and extend to larger and more complex peptides, there is an increasing need for more precisely designed protecting groups to meet ever more rigorous requirements of selective deprotection without side-reactions, convenient solubility of large protected peptide fragments, avoidance of racemisation, and compatibility with the various coupling methods. This need has stimulated an upsurge of activity in research on the chemistry of protective groups, and the year under review is no exception to this trend.

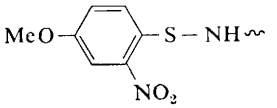
The use of anhydrous hydrogen fluoride for the cleavage of *S*-benzyl groups (the removal of which has until now generally been effected by the use of sodium in liquid ammonia, which is a troublesome reagent requiring most carefully controlled conditions to avoid further damage to the synthetic molecule) was first introduced by Sakakibara *et al.* in 1965.[1] It has since been shown that anhydrous hydrogen fluoride at 0° or 20° in the presence of anisole (a trap for electrophiles which might otherwise lead to undesirable side-reactions) is an excellent reagent for the removal of many of the protecting groups in current use. Thus alkoxycarbonyl groups, *O*- and *S*-benzyl groups, t-butyl groups, and N^{ω}-nitroarginine groups are rapidly and cleanly cleaved.[2, 3] Methyl esters, *S*-methyl, phthaloyl, and tosyl groups were among those claimed to be stable to the reagent, but the experiments on the tosyl group must have been confined to tosylamino derivatives, as it has recently been announced [4] that anhydrous hydrogen fluoride at 0° brings about rapid deprotection of N^{ω}-tosylarginine residues (which, like *S*-benzylcysteine residues, have until now usually been cleaved by the use of sodium in liquid ammonia). Special apparatus and techniques are required for handling hydrofluoric acid [2, 3, 5] and the removal of last traces of acid from the peptide has given some trouble,[5] but it has been emphasised [6] that the method is quite convenient once suitable apparatus has been constructed. Its application for deprotection in the final stages of peptide synthesis seems likely to become increasingly important. Several examples of the use of this procedure have now appeared, including both of the ribonuclease syntheses (see later).

Protection of Amino-groups. Nearly all of the amino-protective groups which are enjoying popularity at the present time can be removed by acidolysis. The most significant advances have been in improved methods

[1] S. Sakakibara and Y. Shimonishi, *Bull. Chem. Soc. Japan*, 1965, **38**, 1412.
[2] S. Sakakibara, Y. Shimonishi, M. Okada, and Y. Kishida, 'Peptides,' ed. H. C. Beyerman, A. van de Linde, and W. Maassen van den Brink, North-Holland Publishing Co., Amsterdam, 1967, p. 44.
[3] S. Sakakibara, Y. Shimonishi, Y. Kishida, M. Okada, and H. Sugihara, *Bull. Chem. Soc. Japan*, 1967, **40**, 2164.
[4] R. H. Mazur and G. Plume, *Experientia*, 1968, **24**, 661.
[5] S. Sakakibara, Y. Kishida, R. Nishizawa, and Y. Shimonishi, *Bull. Chem. Soc. Japan*, 1968, **41**, 438.
[6] J. Lenard and A. B. Robinson, *J. Amer. Chem. Soc.*, 1967, **89**, 181.

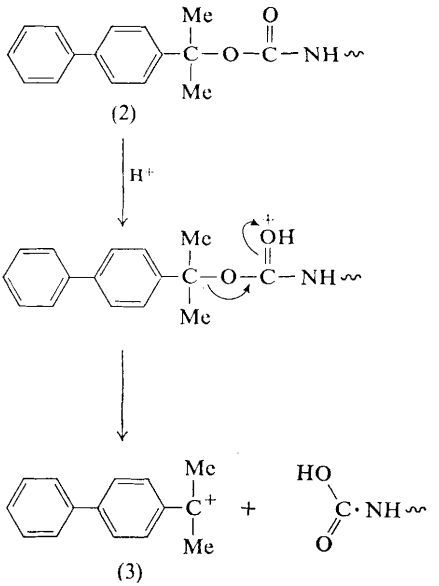
for the preparation of t-butoxycarbonylamino-acids and in the development of protecting groups which are so labile to acid that they can be removed selectively in the presence of protective groups derived from t-butanol. The increasing use of protecting groups with varying degrees of acid sensitivity means that selective deprotection steps and the final complete deprotection stage can be performed under milder conditions, with a concomitant reduction in the danger of undesirable side-reactions. A quantitative comparison of the rates of cleavage of some *N*-protecting groups by acid has been reported.[7]

The substituted sulphenyl protecting group (1) has been developed



(1)

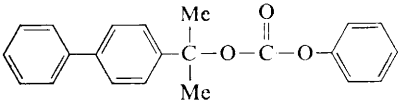
for use with penicillin derivatives. A number of amino-acids *N*-protected with this group have been characterised, and their application in some simple dipeptide preparations has been described. Cleavage was accomplished by means of acetic acid in dioxan.[8]



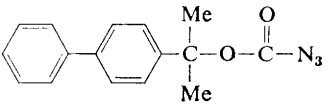
Scheme 1

[7] G. Losse, D. Zeidler, and T. Grieshaber, *Annalen*, 1968, **715**, 196.
[8] Y. Wolman, *Israel J. Chem.*, 1967, **5**, 231.

A study [9] of the acid lability of a series of aralkoxycarbonyl protecting groups shows that the fission (Scheme 1) is a first-order process, the rate depending on the stability of the carbonium ion which is produced. This investigation revealed the potential of the *N*-2-(*p*-diphenylyl)isopropoxycarbonyl group (2) [10] and this potential has been realised in a synthesis of porcine thyrocalcitonin.[11] The group (2) is very much more labile to acid than the t-butoxycarbonyl group because the carbonium ion (3) has the advantage of aromatic delocalisation which is not available to the t-butyl cation. Selective removal of (2) without disturbing protecting groups which contain t-butyl groups is therefore possible. The lability of (2) is comparable to that of the *N*-trityl group, but difficulties due to steric hindrance do not attend its use. The protecting group is easily introduced by use of the mixed carbonate (4), or the azide (5).



(4)



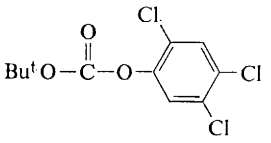
(5)

The t-amyloxycarbonyl group has received further attention,[12, 13] but it does not appear to have any clear advantage over the t-butoxycarbonyl group, except a marginal one in the slightly greater stability of the corresponding chloroformate. In fact use of t-amyl chloroformate is not very satisfactory [14] and so even this advantage does not seem very compelling.

There have been many papers in the past (summarised in ref. 15) on the use of mixed carbonates for the preparation of alkoxycarbonylamino-acids. This approach is especially valuable when the chloroformates are inconveniently unstable for direct use (*e.g.* for the preparation of t-butoxycarbonylamino-acids [16–18] and *p*-methoxybenzyloxycarbonylamino-acids [19, 20]) but is unlikely to find much application for the preparation of benzyloxycarbonylamino-acids [21] except in special circumstances. The most promising of the active ester acylating agents reported this year is t-butyl 2,4,5-trichlorophenyl carbonate (6): the acylamino-acid and phenolic

[9] P. Sieber and B. Iselin, *Helv. Chim. Acta.*, 1968, **51**, 614.
[10] P. Sieber and B. Iselin, *Helv. Chim. Acta.*, 1968, **51**, 622.
[11] W. Rittel, M. Brugger, B. Kamber, B. Riniker, and P. Sieber, *Helv. Chim. Acta*, 1968, **51**, 924.
[12] I. Honda, Y. Shimonishi, and S. Sakakibara, *Bull. Chem. Soc. Japan*, 1967, **40**, 2415.
[13] B. Rzeszotarska and S. Wiejak, *Angew. Chem. Internat. Edn.*, 1968, **7**, 379.
[14] S. Sakakibara, and M. Itoh, *Bull. Chem. Soc. Japan*, 1967, **40**, 646.
[15] B. Rzeszotarska and S. Wiejak, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 86.
[16] M. Fujino and C. Hatanaka, *Chem. and Pharm. Bull. (Japan)*, 1967, **15**, 2015.
[17] W. Broadbent, J. S. Morley, and B. E. Stone, *J. Chem. Soc.* (*C*), 1967, 2632.
[18] B. Rzeszotarska and S. Wiejak, *Annalen*, 1968, **716**, 216.
[19] J. H. Jones and G. T. Young, *Chem. and Ind.*, 1966, 1722.
[20] J. H. Jones and G. T. Young, *J. Chem. Soc.* (*C*), 1968, 53.
[21] B. Rzeszotarska and G. Palka, *Bull. Acad. polon. Sci. Ser. Sci. chim.*, 1968, **16**, 23.

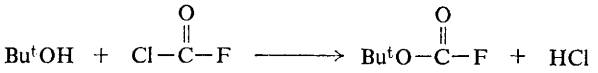
co-product which result from the reaction of (6) with amino-acids in the presence of a base can be extracted together into an organic solvent and



(6)

treated with dicyclohexylcarbodi-imide giving t-butoxycarbonylamino-acid 2,4,5-trichlorophenyl esters directly.[17]

The method developed by Ovchinnikov *et al.*[22] for the preparation of t-butyl chloroformate has been improved a little by some minor modifications, and it has been shown that t-butoxycarbonyl azide can be obtained directly from the unstable chloroformate in 35% overall yield by reaction with hydrazoic acid.[23] t-Butyl fluoroformate, prepared at $-25°$ in excellent yield as shown in Scheme 2, is very much more stable than the chloroformate and can be stored at 0–20° for several months. Yields of



Scheme 2

t-butoxycarbonylamino-acids obtained using this reagent (at *ca.* 0° and pH 8–10) are generally in the range 80–95%.[24–26] The alkoxycarbonyl fluorides corresponding to some other acid-labile protecting groups (*p*-methoxybenzyloxycarbonyl, 3,4,5-trimethoxybenzyloxycarbonyl, and furfuryloxycarbonyl) have also been prepared and shown to be similarly useful.[25, 26]

The removal of benzyloxycarbonyl groups from methionine-containing peptides is fraught with difficulties whatever method of cleavage is adopted: hydrogen bromide in acetic acid may lead to the formation of sulphonium derivatives, and hydrogenolysis usually fails because of catalyst poisoning. Although these problems can be avoided to some extent by using benzyl bromide scavengers or by conducting the hydrogenation in the presence of cyclohexylamine, it has generally been held that it is preferable to plan a synthesis so that the need for debenzyloxycarbonylation of a methionine-containing peptide does not arise. It now appears that quantitative

[22] Y. A. Ovchinnikov, A. A. Kiryushkin, and A. I. Miroshnikov, *Experientia*, 1965, **21**, 418.
[23] H. Yajima and H. Kawatani, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 182.
[24] E. Schnabel, H. Herzog, P. Hoffmann, E. Klauke, and I. Ugi, *Angew. Chem. Internat. Edn.*, 1968, **7**, 380.
[25] E. Schnabel, H. Herzog, P. Hoffmann, E. Klauke, and I. Ugi, 'Peptides 1968,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam 1968, p. 91.
[26] E. Schnabel, H. Herzog, P. Hoffmann, E. Klauke, and I. Ugi, *Annalen*, 1968, **716**, 175.

hydrogenolysis of benzyloxycarbonyl groups in the presence of methionine residues can be achieved in anhydrous methanol using a palladium catalyst if boron trifluoride etherate is added.[27] Conditions for the removal of benzyloxycarbonyl groups by hydrogenolysis without affecting a *p*-nitrophenyl ester have also been reported.[28]

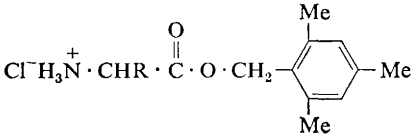
Phthaloyl groups may be removed under very mild conditions using hydroxylamine.[29] In coupling reactions involving excess of the amino-component with acetic acid catalysis, aminolysis can occur at the phthaloyl group, leading to phthalamoyl derivatives.[20]

Hydroxylamine or hydrazine [30] (also substituted hydrazines and aromatic amines [31]) can be used for the regeneration of amino-groups protected as their *N*-formyl derivatives. The resistance of the formyl group to many of the conditions used to remove other protecting groups and the ease with which it is removed by hydroxylaminolysis or hydrazinolysis combine to make it a useful protecting group for lysine side-chains. The value of the newly discovered means of cleavage has been demonstrated for removal of *N*-formyl groups in the final stages of syntheses of α-melanocyte stimulating hormone (α-MSH)[30] and monkey β-MSH.[32]

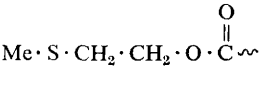
Protection of Carboxyl Groups. The rates of cleavage (by hydrogenolysis and by sodium thiophenoxide) of a series of *N*-protected amino-acid phenacyl esters have been determined.[33] Investigation of substituted benzyl esters also continues,[34-37] with interest focused on acid-labile groups. The most promising of these is the 2,4,6-trimethylbenzyl group,[35, 36] which is comparable with the t-butyl ester in its acid-sensitivity. Unfortunately, amino-acid 2,4,6-trimethylbenzyl ester hydrochlorides (7) can only be obtained by indirect routes (*e.g.* by selective deprotection of the corresponding *o*-nitrophenylsulphenyl derivatives), but this protecting group may be preferred to the t-butyl group in cases where the direct isobutene process is troublesome.

The β-methylthioethyl carboxyl protecting group (8) can be removed by treatment with mild alkali after molybdate-catalysed oxidation to the sulphone (9) using hydrogen peroxide in acetone.[38] Potential users of this

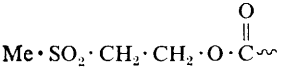
[27] M. Okamoto, S. Kimoto, T. Oshima, Y. Kinomura, K. Kawasaki, and H. Yajima, *Chem. and Pharm. Bull. (Japan)*, 1967, **15**, 1618.
[28] J. Kovacs and R. L. Rodin, *J. Org. Chem.*, 1968, **33**, 2418.
[29] O. Neunhoeffer, G. Lehmann, D. Haberer, and G. Steinle, *Annalen*, 1968, **712**, 208.
[30] H. Yajima, K. Kawasaki, Y. Okada, H. Minami, K. Kubo, and I. Yamashita, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 919.
[31] R. Geiger, W. König, G. Jäger, and W. Siedel, 'Peptides 1968' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 98.
[32] H. Yajima, Y. Okada, Y. Kinomura, and H. Minami, *J. Amer. Chem. Soc.*, 1968, **90**, 527.
[33] G. Losse and G. Berndsen, *Annalen*, 1968, **715**, 204.
[34] M.-H. Loucheux and J. Parrod, *Compt. rend.*, 1968, **267**, *C*, 614.
[35] R. Ledger and F. H. C. Stewart, *Austral. J. Chem.*, 1968, **21**, 1101.
[36] F. H. C. Stewart, *Austral. J. Chem.*, 1968, **21**, 2831.
[37] F. H. C. Stewart, *Austral. J. Chem.*, 1968, **21**, 2543.
[38] P. M. Hardy, H. N. Rydon, and R. C. Thompson, *Tetrahedron Letters*, 1968, 2525.



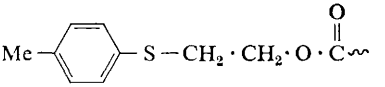
(7)



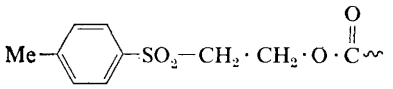
(8)



(9)



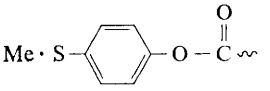
(10)



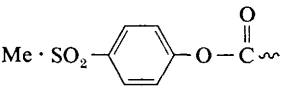
(11)

method should note, however, that hydrogen peroxide in acetone has caused several violent explosions.[39] The β-tolylthioethyl group (10) can be used as a protecting group and removed after oxidation to (11) in the same way as the β-methylthioethyl group, but it has been shown in a few simple examples that (11) itself provides sufficient protection for use in mixed carbonic anhydride, active ester, and dicyclohexylcarbodi-imide coupling reactions.[40]

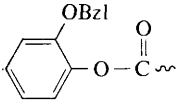
4-(Methylthio)phenyl esters (12) are sufficiently unreactive towards aminolysis to provide carboxyl protection, and oxidation with excess of hydrogen peroxide in acetic acid gives an activated ester (13) when required for further coupling.[41] However, activated esters such as (13) seem likely



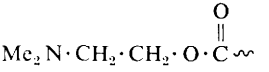
(12)



(13)



(14)



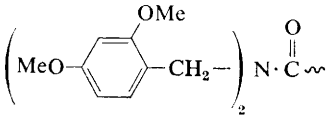
(15)

[39] C. J. M. Stirling, *Chem. in Britain*, 1969, **5**, 36.
[40] A. W. Miller and C. J. M. Stirling, *J. Chem. Soc.* (*C*), 1968, 2612.
[41] B. J. Johnson and P. M. Jacobs, *Chem. Comm.*, 1968, 73.

to parallel other conventional activated phenyl esters in their susceptibility to racemisation, so the use of activated peptides prepared in this way cannot be recommended with confidence. The *o*-benzyloxyphenyl ester group (14) might prove useful in an analogous fashion: the group (14) is unreactive as it stands but can be converted to a racemisation-free activated *o*-hydroxyphenyl ester by scission of the ether link.[42]

The use of the group (15) (which can be removed by treatment with dilute aqueous bicarbonate) in some very simple dipeptide preparations has been reported.[43]

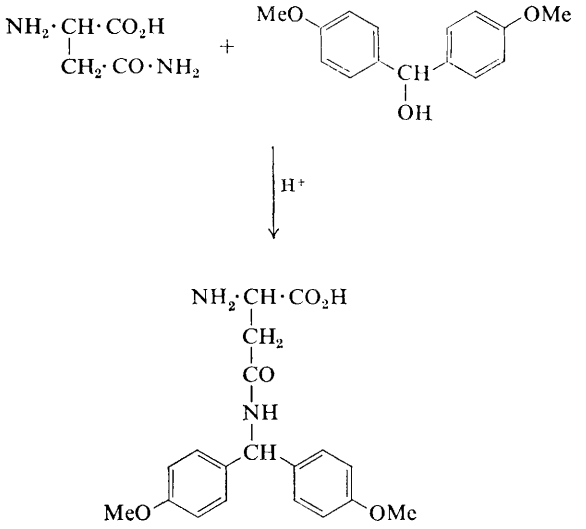
Protection of Side-chain Functional Groups. Amide side-chains have generally been left unprotected in the past through lack of suitable protecting groups. There are in fact several advantages to be gained from blocking amide side-chains: side-reactions such as nitrile formation would be eliminated and, most important now that synthetic work on very large peptides is in progress, protection of side-chain amide groups is expected to confer increased solubility in organic solvents. With these advantages in mind, a very detailed study of the susceptibility of variously substituted amides to cleavage has been undertaken.[44] Of the many groups examined, the bis(2,4-dimethoxybenzyl) method of protection (16) is the most



(16)

promising, and a number of useful glutamine and asparagine derivatives protected in this way have been characterised.[45] The protected amide function was introduced by coupling bis(2,4-dimethoxybenzyl)amine with a suitable protected aspartic or glutamic acid derivative having an exposed ω-carboxyl group. The 2,4-dimethoxybenzyl substituents are removed by treatment with trifluoroacetic acid, but resist hydrogenolysis, alkali, and treatment with methanolic hydrogen chloride. Use in conjunction with other well-tried protecting groups is therefore possible, and application of (16) in a number of examples was very successful.[45-47] The 4,4′-dimethoxydiphenylmethyl protecting group can be introduced directly on to side-chain amide groups as shown in Scheme 3.[31] Unlike the xanthyl protecting

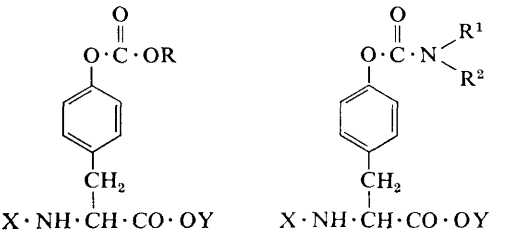
[42] J. H. Jones and G. T. Young, *J. Chem. Soc.* (*C*), 1968, 436.
[43] A. E. Greben', V. F. Martynov, and M. I. Titov, *Zhur. obshchei Khim.*, 1968, **38**, 664.
[44] F. Weygand, W. Steglich, J. Bjarnason, R. Akhtar, and N. Chytil, *Chem. Ber.*, 1968, **101**, 3623.
[45] F. Weygand, W. Steglich, and J. Bjarnason, *Chem. Ber.*, 1968, **101**, 3642; P. Pietta, F. Chillemi, and A. Corbellini, *Chem. Ber.*, 1968, **101**, 3649.
[46] P. G. Pietta, F. Chillemi, and A. Corbellini, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 104.
[47] F. Chillemi, P. G. Pietta, and A. Corbellini, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 107.



Scheme 3

group of Akabori,[48] to which it is related, the 4,4′-dimethoxydiphenylmethyl group is cleaved relatively easily with acid (trifluoroacetic acid, 20°, 2–3 hr.) and confers increased solubility on its derivatives.[31]

A wide range of *N*-acyltyrosines, *N*-acyltyrosine active esters, and tyrosine alkyl ester salts have been prepared with their phenolic oxygens blocked by alkoxycarbonyl[49] [17; R = CH_3, CH_3CH_2, $(CH_3)_2CH$, or $(CH_3)_2CHCH_2$] or carbamoyl groups[50] [18; $R^1 = R^2 = H$; $R^1 = H$, $R^2 = Ph$;

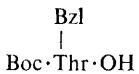


(17) (18)

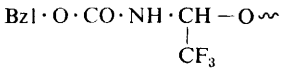
$R^1 = H$, $R^2 = (CH_3)_2CHCH_2$; $R^1 = H$, $R^2 = (p\text{-}NO_2)Ph$; or $R^1 = R^2 = Ph$]. These are useful intermediates in peptide synthesis. The protecting groups are stable to coupling conditions, are not affected by hydrogenation or strong acids, but are cleaved smoothly by nucleophiles such as hydrazine.[49, 50]

[48] S. Akabori, S. Sakakibara, and Y. Shimonishi, *Bull. Chem. Soc. Japan*, 1961, **34**, 739.
[49] R. Geiger, G. Jäger, A. Volk, and W. Siedel, *Chem. Ber.*, 1968, **101**, 2189.
[50] G. Jäger, R. Geiger, and W. Siedel, *Chem. Ber.*, 1968, **101**, 2762.

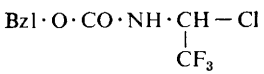
A convenient procedure for the preparation of *O*-benzyl-L-threonine by simultaneous acid-catalysed esterification and *O*-alkylation of the amino-acid followed by saponification has been reported. The *N*-t-butoxycarbonyl derivative (19) can be obtained from L-threonine without isolation of any intermediates.[51] The yields in the preparation of these intermediates were mediocre, but optical integrity was preserved: racemisation occurs if the benzylation is performed using sodium–ammonia. Alcoholic hydroxyl groups can be masked with the 2,2,2-trifluoro-1-benzyloxycarbonylamino-ethyl group (20), which is introduced by reaction of the chloride (21) with



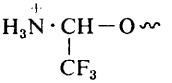
(19)



(20)



(21)

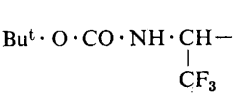


(22)

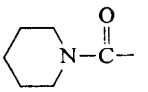
suitably protected derivatives. Protection with (20) is compatible with conventional methods of coupling and is cleaved by the usual conditions for removal of benzyloxycarbonyl groups. Removal of the benzyloxy-carbonyl part of the protecting group gives (22) which is very easily hydrolysed with regeneration of the original hydroxyl group.[52] A series of useful *N*-(*o*-nitrophenylsulphenyl)-*O*-t-butylhydroxyamino-acids and their activated esters have been prepared.[53]

The 2,2,2-trifluoro-1-benzyloxycarbonylaminoethyl group, and its t-butyloxycarbonyl analogue (23), can be used for the protection of the imino-nitrogen of histidine residues.[54, 55] These protecting groups are removed by the same conditions as are used for the cleavage of the corresponding *N*-alkoxycarbonyl groups. Adequate reduction in the basicity of the imidazole ring is provided by this type of protection, and the method proved satisfactory in the synthesis of a fragment (139–146) of human haemoglobin β-chain which contains two histidine residues.[47] The piperidinocarbonyl group (24) has been suggested as a means of protecting the imino-nitrogen of histidine: it is stable to acidolysis and hydrogenolysis but can be cleaved with hydrazine.[31, 56]

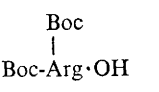
[51] T. Mizoguchi, G. Levin, D. W. Woolley, and J. M. Stewart, *J. Org. Chem.*, 1968, **33**, 903.
[52] F. Weygand, W. Steglich, F. Fraunberger, P. Pietta, and J. Schmid, *Chem. Ber.*, 1968, **101**, 923.
[53] E. Wünsch and A. Fontana, *Chem. Ber.*, 1968, **101**, 323.
[54] F. Weygand, W. Steglich, and P. Pietta, *Chem. Ber.*, 1967, **100**, 3841.
[55] F. Weygand, W. Steglich, A. Maierhofer, and A. Bauer, *Chem. Ber.*, 1968, **101**, 1894.
[56] G. Jäger, R. Geiger, and W. Siedel, *Chem. Ber.*, 1968, **101**, 3537.



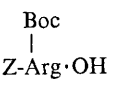
(23)



(24)



(25)



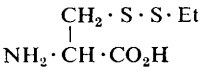
(26)

The finding that hydrogen fluoride can be used for cleavage of N^{ω}-tosylarginine residues has been mentioned earlier. The t-butoxycarbonyl derivatives (25) and (26) have been reported.[57]

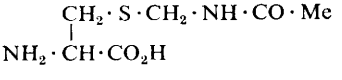
Protection of Thiol Groups and the Synthesis of Cystine Peptides. The *S*-benzyl group which was so successful in the early syntheses of natural oligopeptides, and which still enjoys considerable popularity, can only be removed under rather drastic conditions. Cleavage with sodium in ammonia is usually accompanied by side-reactions such as rupture of peptide bonds and desulphurisation, and hydrogen fluoride removes most other protecting groups. Furthermore, for the unambiguous synthesis of unsymmetrical cystine peptides with more than one disulphide bridge, there is a requirement for a range of thiol protecting groups to permit selective cross-linking between specified cysteine residues without affecting existing disulphide bonds. A large number of methods of *S*-protection have been proposed and used in recent years; literature relating to these is cited in the introductory sections of refs. 58–60.

The usefulness of the *S*-benzylthiomethyl group (which is cleaved by means of mercuric acetate in formic acid) has been confirmed in syntheses of glutathione and homoglutathione.[60] *S*-Ethylmercaptocysteine (27) has been used in syntheses of glutathione and oxytocin. The protecting group is removed by treatment with thiophenol or thioglycollic acid, but is not affected by acidolysis conditions. *S*-Acetamidomethylcysteine (28) has been prepared (by acid-catalysed *S*-alkylation of cysteine with the corresponding alcohol) and several useful derivatives of it have been characterised. This new *S*-masking group is stable under acidic (even hydrogen fluoride) and alkaline conditions but is cleanly cleaved by mercuric ions at pH 4.[61] It was used in one of the ribonuclease syntheses (see later). Details of a

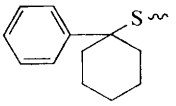
[57] H. Arold and S. Reissmann, *Z. Chem.*, 1968, **8**, 107.
[58] N. Inukai, K. Nakano, and M. Murakami, *Bull. Chem. Soc. Japan*, 1967, **40**, 2913.
[59] I. Photaki, I. Phocas, J. Taylor-Papadimitriou, and L. Zervas, 'Peptides 1968,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 201.
[60] R. Camble, R. Purkayastha and G. T. Young, *J. Chem. Soc.* (*C*), 1968, 1219.
[61] D. F. Veber, J. D. Milkowski, R. G. Denkewalter, and R. Hirschmann, *Tetrahedron Letters*, 1968, 3057.



(27)



(28)



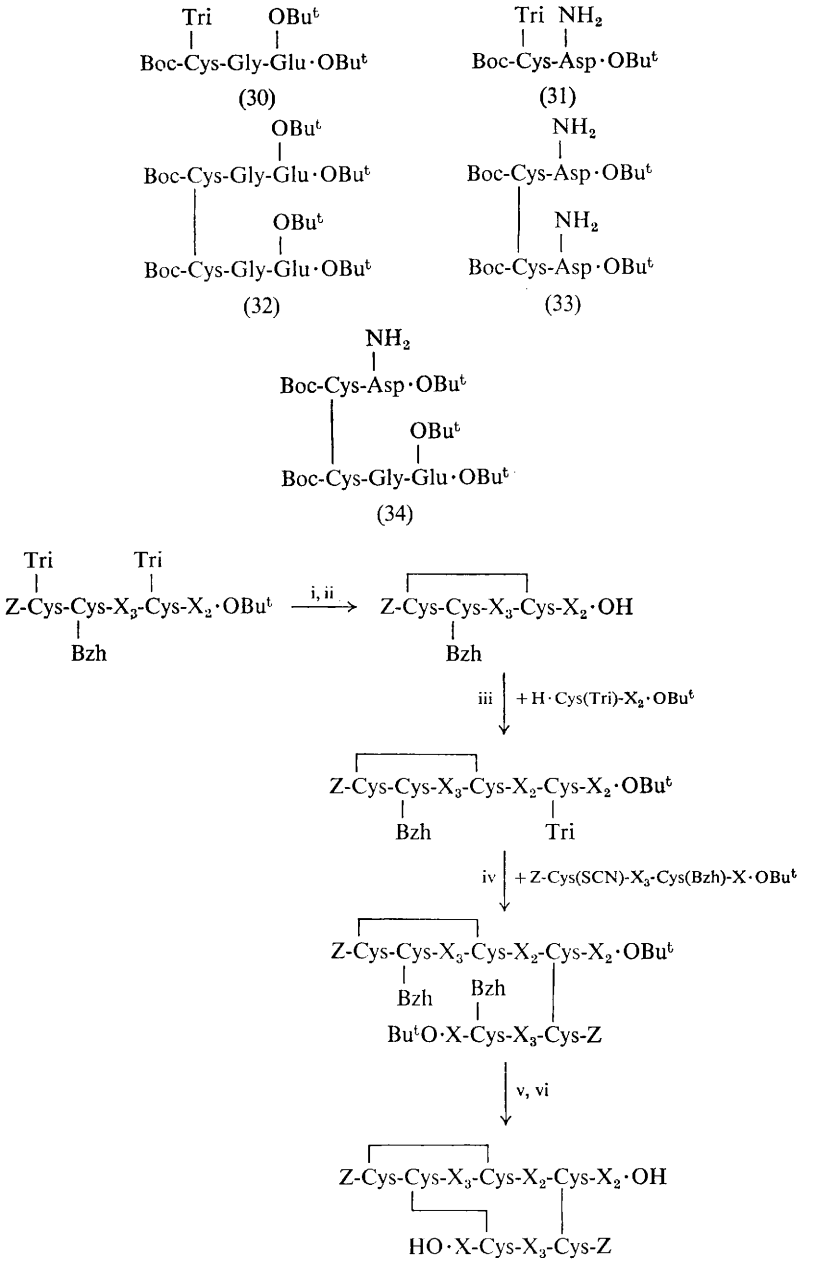
(29)

convenient preparation of *S*-benzhydryl-L-cysteine giving a crude yield of 90% have been published.[62] Despite claims [63] to the contrary, the *S*-trityl group is cleaved by hot trifluoroacetic acid:[59, 64] the confusion arose because removal of trifluoroacetic acid *in vacuo* after the reaction causes retritylation. The rates of fission by hot trifluoroacetic acid of a large number of *S*-protecting groups which would give stabilised carbonium ions on acidolysis have been determined.[63] Some of these [*e.g.* (29)] were cleaved very rapidly [5 min. in the case of (29)] and should therefore be selectively removable in the presence of groups such as benzhydryl, which is not affected by short exposure to hot trifluoroacetic acid. However, some of the conclusions of this investigation must be regarded with caution, as the extent of cleavage was determined after removal of the trifluoroacetic acid *in vacuo*, which may have permitted reintroduction of *S*-protection in some cases.

S-Tritylcysteine-containing protected peptides can be converted in one step to cystine peptides by treatment with iodine in boiling methanol. Thus treatment of a mixture of the two protected peptides (30) and (31) under these conditions gave the two symmetrical protected cystine peptides (32) and (33) and the protected insulin (A 20–21, B 19–21) fragment (34) in the statistically expected proportions.[65]

Further studies on the sulphenylisothiocyanate method for the construction of unsymmetrical disulphides have been reported,[66–68] and the feasibility of the procedure for the stepwise introduction of cystine bridges has been demonstrated by the synthesis of a *tris*-cystine derivative which is shown in Scheme 4.[66]

[62] R. G. Hiskey and J. B. Adams, *Biochem. Prep.*, 1968, **12**, 92.
[63] W. König, R. Geiger, and W. Siedel, *Chem. Ber.*, 1968, **101**, 681.
[64] L. Zervas, I. Photaki, and I. Phocas, *Chem. Ber.*, 1968, **101**, 3332.
[65] B. Kamber and W. Rittel, *Helv. Chim. Acta*, 1968, **51**, 2061.
[66] R. G. Hiskey, R. L. Smith, A. M. Thomas, J. T. Sparrow, and W. C. Jones, 'Peptides 1968,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 209.
[67] R. G. Hiskey and R. L. Smith, *J. Amer. Chem. Soc.*, 1968, **90**, 2677.
[68] R. G. Hiskey and M. A. Harpold, *J. Org. Chem.*, 1968, **33**, 559.

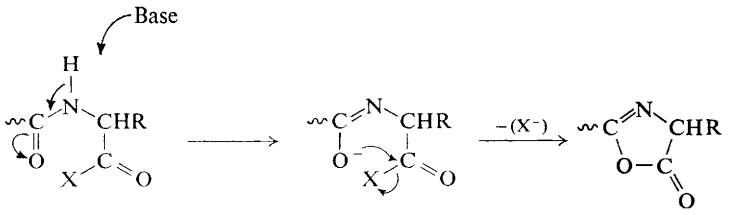


X = an amino-acid residue other than Cys

Reagents: i, $(SCN)_2$—AcOH; ii, BF_3—AcOH; iii, DCCI; iv, AcOH; v, $(SCN)_2$—TFA—AcOH; vi, BF_3—AcOH

Scheme 4

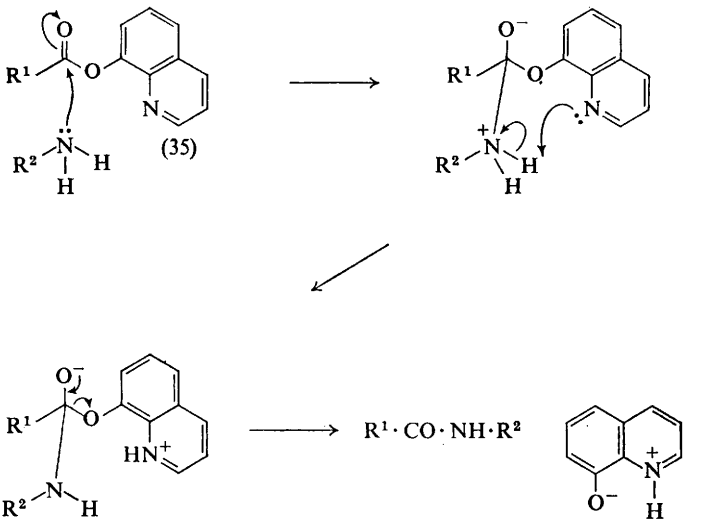
B. Formation of the Peptide Bonds.—*Activated Esters.* A very large number [69] of activated esters have been proposed for use in peptide synthesis, but few have been employed in actual syntheses. This wide range of intermediates has been investigated in the search for advantages such as improved crystallinity, rapidity of reaction, ease of preparation, and removal of co-products, *etc.* Most of the esters examined, however, depend for their activation on a strong electron-withdrawing effect in the leaving group. Such indiscriminate activation increases not only the rate of aminolysis but also the rate of oxazolone formation (and hence of racemisation), so that the use of most active esters is confined to stepwise procedures using activated acylamino-acids where the means of *N*-protection precludes oxazolone formation. Methods for selective activation of carboxyl groups towards aminolysis are therefore required. Now the nucleophile in oxazolone formation (Scheme 5) bears no hydrogen but the nucleophile in



Scheme 5

aminolysis does, so that one way of achieving the required selective activation is to exploit this difference by using intramolecular general base catalysis. The 8-hydroxyquinoline esters (35), which were first introduced [70] because the co-product is easily removed by extraction with acid, have been shown [42, 71] to owe their racemisation-free activation to intramolecular general base catalysis as shown in Scheme 6: a very similar mechanism had been proposed earlier [72] to explain the unexpectedly high rate of aminolysis of 1-piperidyl esters in the absence of acidic catalysis. The same principle is incorporated in *o*-hydroxyphenyl esters (36) [42] and the highly reactive esters of 2-mercaptopyridine (37):[73] it may even be that acyl azides are also selectively activated towards aminolysis by intramolecular general base catalysis (Scheme 7).[74] The criticism could be made that the transition state (38) involves an unfavourable ring shape, but in view of the consistency of Scheme 7 with the established mechanisms for the racemisation-free coupling of 8-hydroxyquinoline and *o*-hydroxyphenyl esters, this is perhaps

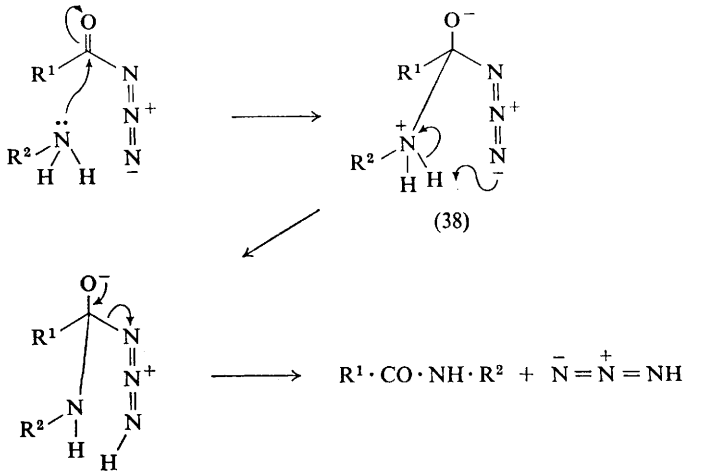
[69] H.-D. Jakubke, *Z. Chem.*, 1966, **6**, 52.
[70] H.-D. Jakubke, *Z. Naturforsch.*, 1965, **20b**, 273.
[71] H.-D. Jakubke, A. Voigt, and S. Burkhardt, *Chem. Ber.*, 1967, **100**, 2367.
[72] B. O. Handford, J. H. Jones, G. T. Young, and (in part) T. F. N. Johnson, *J. Chem. Soc.*, 1965, 6814.
[73] K. Lloyd and G. T. Young, *Chem. Comm.*, 1968, 1400.
[74] G. T. Young, 'Peptides,' ed. H. C. Beyerman, A. van de Linde, and W. Maassen van den Brink, North-Holland Publishing Co., Amsterdam 1967, p. 55.



(35)

Scheme 6

(36) (37)



(38)

Scheme 7

the most attractive of the solutions which have been proposed [75] for the long-standing problem of the freedom of azide couplings from the danger of racemisation.

Further studies of the use of 1-piperidyl esters in peptide synthesis have been reported:[20, 60, 76] these esters are useful in the synthesis of small peptides but the reaction is inconveniently slow when hindered amino-acids or larger peptides are involved.

The use of enol esters generated *in situ* by reaction with *N*-ethyl-5-phenylisoxazolium-3′-sulphonate (Woodward's Reagent K) is now well established, and it is therefore of interest to note that reaction of acylamino-acids with the isoxazolium salt (39) gives enol esters (40) which can be isolated as stable crystalline solids and used as active esters in the usual way.[77]

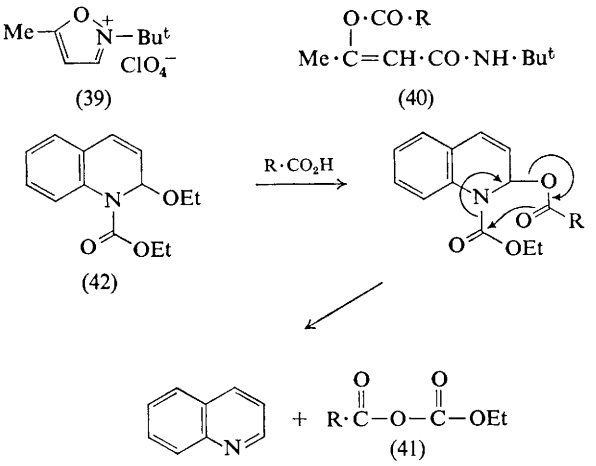
Removal of excess of acylating agent and the co-product is a problem which often attends the use of active esters. This difficulty is very neatly avoided if the active ester is an ester of a hydroxypolymer: filtration after the reaction then suffices to remove both excess of reagent and co-product. Esters of poly-(5-vinyl-8-hydroxyquinaldine),[78] copoly(ethylene-*N*-hydroxy-maleimide) [79] and cross-linked poly-(4-hydroxy-3-nitrostyrene) [80] have been suggested for use in this way, and the last has been used in a synthesis of bradykinin.[80]

Other papers have described studies on pentachlorophenyl esters,[81] 2,6-dichloro-4-nitrophenyl esters,[82] *p*(*N*-pyrimidyl)-sulphonamidophenyl esters,[83] *p*-propionylphenyl esters,[84] *p*-sulphonamidophenyl esters,[85] *O*-acyl derivatives of pivalohydroxamic acid,[86] and *O*-acyl derivatives of *N*-hydroxy-glutarimide.[87]

Other Methods. Mixed carbonic anhydrides (41) have been widely used over the years: such anhydrides are usually formed by reaction of acylamino-acid salts with ethyl chloroformate, and used without delay. One difficulty with the method is that racemisation (if permitted by the structure of the carboxyl component) and disproportionation into symmetrical anhydrides

[75] M. Bodanszky and M. A. Ondetti, 'Peptide Synthesis,' Interscience, New York, 1966, p. 149.
[76] J. H. Jones, B. Liberek, and G. T. Young, *J. Chem. Soc.* (*C*), 1967, 2371.
[77] R. B. Woodward and D. G. Woodman, *J. Amer. Chem. Soc.*, 1968, **90**, 1371.
[78] G. Manecke and E. Haake, *Naturwiss.*, 1968, **55**, 343.
[79] D. A. Laufer, T. M. Chapman, D. I. Marlborough, V. M. Vaidya, and E. R. Blout, *J. Amer. Chem. Soc.*, 1968, **90**, 2696.
[80] M. Fridkin, A. Patchornik, and E. Katchalski, *J. Amer. Chem. Soc.*, 1968, **90**, 2953.
[81] M. Fujino and C. Hatanaka, *Chem. and Pharm. Bull.* (*Japan*), 1968, **16**, 929; A. Kapoor, E. J. Davis, and M. J. Graetzer, *J. Pharm. Sci.*, 1968, **57**, 1514; J. Kovacs, M. Q. Ceprini, C. A. Dupraz, and G. N. Schmit, *J. Org. Chem.*, 1967, **32**, 3696; A. Kapoor, *J. Pharm. Sci.*, 1967, **56**, 1532.
[82] F. H. C. Stewart, *Austral. J. Chem.*, 1968, **21**, 477.
[83] G. Kupryszewski, F. Muzalewski, and J. Przybylski, *Roczniki Chem.*, 1968, **42**, 1009.
[84] M. Kaczmarek, G. Kupryszewski, and J. Wajcht, *Zeszyty. Nauk., Mat., Fiz., Chem.*, 1967, **7**, 143.
[85] G. Kupryszewski and F. Muzalewski, *Zeszyty Nauk., Mat., Fiz., Chem.*, 1967, **7**, 159.
[86] S. Rajappa, K. Nagarajan, and V. S. Iyer, *Tetrahedron*, 1967, **23**, 4805; T. R. Govindachari, S. Rajappa, A. S. Akerkar, and V. S. Iyer, *ibid.*, 1967, **23**, 4811.
[87] H. Jeschkeit, *Z. Chem.*, 1968, **8**, 20.

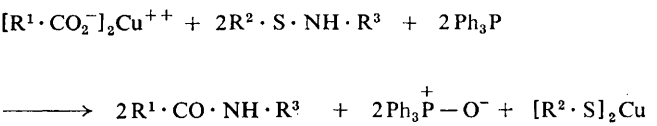
can occur in the interval between the formation and aminolysis of the mixed anhydride. An alternative to the classical procedure employs the novel reagent (42) which acts as a 'direct' coupling reagent (*cf.* the use of dicyclohexylcarbodi-imide). Addition of (42) to a mixture of the amino- and carboxyl-components causes generation of a mixed carbonic anhydride *in situ* as shown in Scheme 8: since the mixed anhydride is consumed



Scheme 8

immediately by aminolysis, side-reactions do not intervene and racemisation is reduced to such an extent that it is not detectable in the Young test.[88]

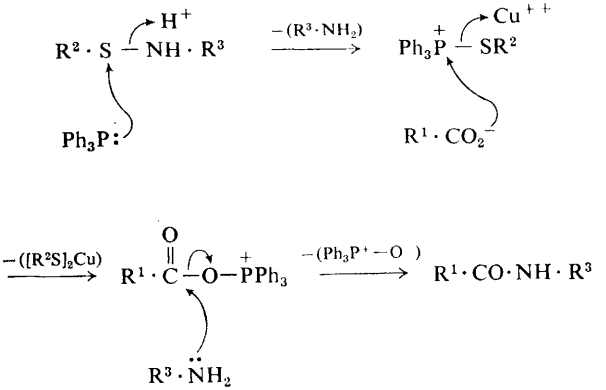
An interesting new method for the formation of peptide bonds uses the reaction of *N-o*-nitrophenylsulphenylamino-esters with acylamino-acid copper salts in the presence of triphenylphosphine (Scheme 9). Considering



Scheme 9

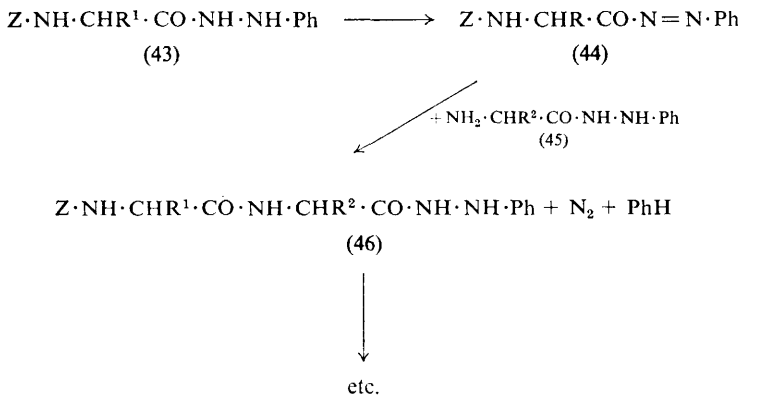
the complexity of the reaction mixture, surprisingly good yields were obtained in simple cases, but the danger of racemisation was not investigated.[89] No mechanism was put forward, but it seems likely that, as in other so-called *N*-activation methods, an activated carboxyl group is involved, perhaps as suggested in Scheme 10. If this is so, then racemisation of the carboxyl component could occur in certain circumstances.

[88] D. Belleau and G. Malek, *J. Amer. Chem. Soc.*, 1968, **90**, 1651.
[89] T. Mukaiyama, M. Ueki, H. Maruyama, and R. Matsueda, *J. Amer. Chem. Soc.*, 1968, **90**, 4490.



Scheme 10

Acylamino-acid phenylhydrazides (43) can be oxidised with lead tetra-acetate or *N*-bromosuccinimide to unstable phenyldi-imides (44), which can be treated with amino-acid phenylhydrazides (45) to give acyldipeptide phenylhydrazides (46): these can then be used in the same way for further coupling (Scheme 11). Further coupling reactions with (46), however, result in some racemisation (a few percent in the Anderson test).[90]



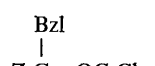
Scheme 11

Several other methods for the formation of peptide bonds have been the subjects of publications this year, including: the use of ynamines,[91] derivatives of arylsulphenic acids and phosphites,[92] 2-fluoropyridine *N*-oxide for

[90] H. B. Milne and C. F. Most jun., *J. Org. Chem.*, 1968, **33**, 169.
[91] F. Weygand, P. Huber, and K. Weiss, *Z. Naturforsch.*, 1967, **22b**, 1084; W. Steglich, G. Hofle, W. König, and F. Weygand, *Chem. Ber.*, 1968, **101**, 308.
[92] Yu. V. Mitin and G. P. Vlasov, *Doklady Akad. Nauk S.S.S.R.*, 1968, **179**, 353.

carboxyl activation,[93] silyl esters,[94] and the amino-acyl inclusion method.[95]

Racemisation. The mechanism of the most serious cause of racemisation during peptide synthesis (*via* oxazolones) is now well understood. The circumstances under which oxazolone formation can occur impose serious limitations on the strategy of peptide synthesis, and a knowledge of the mechanism of racemisation by this route has been vital in the development of new methods which reduce these restrictions (*e.g.* active esters which are reactive towards aminolysis by virtue of intramolecular general base catalysis—see p. 187). The oxazolone mechanism is, however, not the only means by which racemisation can occur, and some residues (in particular *S*-benzylcysteine residues) undergo slow base-catalysed racemisation even when *N*-protected with a group which prohibits oxazolone formation. It has been widely held in many quarters that the racemisation of *S*-benzylcysteine derivatives occurs by reversible β-elimination of benzyl mercaptan: this suggestion has now been disproved, as it has been shown that triethylamine-catalysed racemisation of (47) in the presence of benzyl[^{35}S]thiol



(47)

does not lead to incorporation of radioactivity.[96] The use of sterically hindered amines (*e.g.* ethyldi-isopropylamine) for the liberation of amino-groups from their salts has been suggested as a means of reducing the danger of this type of racemisation.[97] The exact means by which the thioether side-chain facilitates exchange at the α-carbon remains an open question.

Some important practical points concerning the application of the Young racemisation test have been discussed.[20] A simpler version of the gas chromatographic racemisation test has been devised by Weygand *et al.*:[98] this modification employs as its model reaction the coupling of trifluoroacetyl-L-prolyl-L-valine with L-proline methyl ester; racemisation in the valine residue is determined by gas chromatographic analysis of the diastereoisomeric tripeptides, which can be accomplished in 15 min. It was reported[99] in 1967 that the coupling of an acyl-L-alanine with L-phenyl-

[93] D. Sarantakis, J. K. Sutherland, C. Tortorella, and V. Tortorella, *J. Chem. Soc.* (*C*), 1968, 72.
[94] L. Birkofer and F. Muller, 'Peptides 1968,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 151.
[95] T. R. Telesnina, V. K. Antonov, and M. M. Shemyakin, *Zhur. obshchei Khim.*, 1968, **38**, 1691.
[96] J. Kovacs, G. L. Mayers, R. H. Johnson, and U. R. Ghatak, *Chem. Comm.*, 1968, 1066.
[97] M. Bodansky and R. J. Bath, *Chem. Comm.*, 1968, 766.
[98] F. Weygand, D. Hoffmann, and A. Prox, *Z. Naturforsch.*, 1968, **23b**, 279.
[99] B. Halpern, L. F. Chew, and B. Weinstein, *J. Amer. Chem. Soc.*, 1967, **89**, 5051.

alanine methyl ester (or of an acyl-L-phenylalanine with L-alanine methyl ester) could be employed as a model reaction for racemisation tests. The extent of racemisation in the carboxyl component can be estimated by examination of the methyl doublet resonance in the n.m.r. spectrum, as the chemical shift of the alanine side-chain is not the same in the L-L isomer as it is in the D-L isomer. The sensitivity of this test has recently been improved about tenfold by using the ^{13}C—H satellite peak of the predominant L-L isomer as an internal standard.[100, 101]

The influence of solvents on the degree of racemisation in the preparation of acylpeptide active esters using dicyclohexylcarbodi-imide has been examined,[102] and the effects of various active ester component additives (*N*-hydroxysuccinimide, *N*-hydroxypiperidine, *etc.*) on racemisation in dicyclohexylcarbodi-imide couplings have been compared.[103] Further studies on the hydrazinolysis of oxazolones have been described.[104, 105]

C. Repetitive Methods of Peptide Synthesis.—*Synthesis on a Polymeric Support.* At the time of writing, the publication of a laboratory manual [106] on this method is imminent, and the subject has recently been reviewed in the Japanese,[107] Russian,[108] Polish,[109] and German [110] languages. The year under review has seen the announcement of the synthesis of two very large peptides by the 'solid phase' method (ribonuclease, and apoferredoxin).[111, 112] These syntheses, particularly the former of course, are staggering achievements, but the purity of the final products leaves much to be desired. The outstanding problem is that unless every coupling reaction proceeds to completion, the final peptide is inevitably contaminated with peptides differing from the required sequence by one or more amino-acid deletions. The difficulties of attaining quantitative reaction and of separating mixtures of very similar peptides increase with the length of the peptide chain, and the prospects for the application of 'solid phase' methods for the unambiguous synthesis of proteins seem slim at the present time. In the case of small peptides, the separation of the required peptide from any

[100] R. S. Dewey, E. F. Schoenewaldt, H. Joshua, W. J. Paleveda jun., H. Schwam, H. Barkemeyer, B. H. Arison, D. F. Veber, R. G. Denkewalter, and R. Hirschmann, *J. Amer. Chem. Soc.*, 1968, **90**, 3254.
[101] R. Hirschmann, 'Peptides 1968,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 139.
[102] B. Bator-Sawicka, *Roczniki Chem.*, 1968, **42**, 47.
[103] J. E. Zimmerman and G. W. Anderson, *J. Amer. Chem. Soc.*, 1967, **89**, 7151.
[104] B. Liberek and A. Michalik, 'Peptides 1968,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 159.
[105] I. Z. Siemion, *Roczniki Chem.*, 1968, **42**, 237.
[106] J. M. Stewart and J. D. Young, 'Solid Phase Peptide Synthesis,' W. H. Freeman and Co., San Francisco, 1969.
[107] F. Sakiyama, *Kobunshi*, 1967, **16**, 790.
[108] V. S. Vesa, *Uspekhi Khim.*, 1968, **37**, 246.
[109] J. Izdebski and S. Drabarek, *Wiad. Chem.*, 1968, **22**, 35.
[110] T. Okuda, *Naturwiss*, 1968, **55**, 209.
[111] E. Bayer, H. Hagenmaier, G. Jung, and W. Konig, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 162.
[112] E. Bayer, G. Jung, and H. Hagenmaier, *Tetrahedron*, 1968, **24**, 4853.

'error' peptides does not pose serious problems: *e.g.* four recent reports [113, 114] describe syntheses of oxytocin in which the required protected nonapeptide precursor was obtained in a very pure condition by ammonolysis of the link between the protected peptide and the resin followed by simple reprecipitation or washing procedures.

The difficulty in obtaining complete coupling at every stage arises because the growing chains are not attached only to the surface of the polymer particles; if they were, the attainment of complete acylation would present no problems beyond those normally encountered in homogeneous solution reactions. In fact it has recently been shown (by autoradiography of resin beads containing tritium-labelled peptides) that the peptide chains are uniformly distributed throughout the polymeric matrix.[115] Presumably some of the peptide chains are attached to sites in cavities: for such chains there comes a time when the dimensions of the cavity prevent penetration of reagents, so that further growth does not occur. Acylation steps are usually performed using activated esters or dicyclohexylcarbodi-imide, but it has been suggested that the use of 1,2,4-triazole to increase the reactivity of active esters is advantageous, and this procedure has been applied in syntheses of oxytocin and oxytocinoic acid.[114] Acceleration of acylation reactions in this way is unlikely to ameliorate to any great degree the problems associated with incomplete reaction, however, as the accessibility of the growing peptide chain would seem to be a more important factor than the reactivity of the acylating agent.

It is now clear that it is desirable to monitor the progress of syntheses on polymeric supports; this is usually done by amino-acid analysis, but mass spectrometry after cleavage of a portion from the resin, partial degradation, chemical modification, and g.l.c., has been recommended as a convenient method of detecting 'error' peptides.[111] A simple system for the continuous determination of the optical density of the reacting solution at an appropriate wavelength enables the course of an individual step to be followed: the solution of reagent is circulated over the resin and through a spectrophotometer cell.[116]

The most appealing feature of the 'solid phase' method is its amenability to automation, and several technical advances have been made in this area recently.[117]

[113] E. Bayer and H. Hagenmaier, *Tetrahedron Letters*, 1968, 2037; M. Manning, *J. Amer. Chem. Soc.*, 1968, **90**, 1348; D. A. J. Ives, *Canad. J. Chem.*, 1968, **46**, 2318.

[114] H. C. Beyerman, C. A. M. Boers-Boonekamp, and H. M. van den Brink-Zimmermannova, *Rec. Trav. chim.*, 1968, **87**, 257.

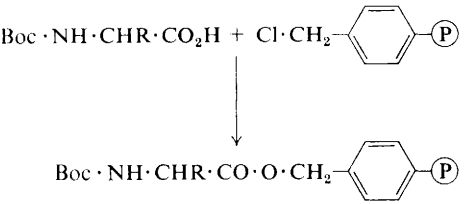
[115] R. B. Merrifield and V. Littau, 'Peptides,' ed. E. Bricsa, North-Holland Publishing Co., Amsterdam, 1968, p. 179.

[116] V. Gut and J. Rudinger, 'Peptides,' ed. E. Bricsa, North-Holland Publishing Co., Amsterdam, 1968, p. 185.

[117] G. W. H. A. Mansveld, H. Hindriks, and H. C. Beyerman, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 197; K. Brunfeldt, J. Halstrøm, and P. Roepstorff, *ibid.*, p. 194; A. Loffet and J. Close, *ibid.*, p. 189.

The fragment condensation method in 'solid phase' peptide synthesis has been examined.[118] The attractive feature of this approach is that if incomplete condensation occurs at any stage then the 'error' peptides which contaminate the final product are likely to differ sufficiently from the required sequence to permit relatively easy separation. Since the azide method was not satisfactory, however, the usual limitations on the strategy of fragment condensation must be taken into consideration if racemisation is to be completely avoided.

The original polymer introduced by Merrifield[119] consists of a chloromethylated co-polymer of styrene and divinylbenzene: the first amino-acid is introduced by treatment of the polymer with an acylamino-acid giving a benzyl ester link (Scheme 12). If the aromatic rings of the polymer are

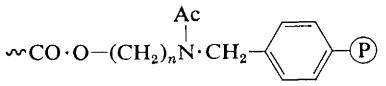


Scheme 12

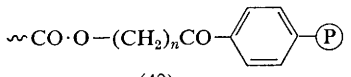
nitrated or chlorinated, the benzyl ester link survives the conditions necessary for acidolysis of benzyloxycarbonyl groups, but the use of more acid-labile *N*-protection in conjunction with unsubstituted Merrifield polymers has generally been preferred. One problem with this type of polymer–peptide bond is that rather drastic conditions are required for its cleavage by acid at the end of the synthesis. Alternative methods of cleavage have therefore been examined, including transesterification[120, 121] and hydrazinolysis.[122] The latter can, of course, be employed to prepare acylpeptide hydrazides for use in classical azide coupling reactions.

Because of the limitations on the methods of functional group protection imposed by the use of a benzyl ester polymer–peptide link and also because of the vigorous conditions required for scission of this link, modified resins with different types of polymer–peptide links have been examined. Thus polymers which can be bound to the carboxyl terminus by means of an alkyl ester bond[123, 124] [(48) and (49): bond formed by carbonyldi-imidazole activation; cleaved by alkaline saponification], a phenacyl ester bond[125]

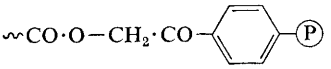
[118] G. S. Omenn and C. B. Anfinsen, *J. Amer. Chem. Soc.*, 1968, **90**, 6571.
[119] R. B. Merrifield, *J. Amer. Chem. Soc.*, 1963, **85**, 2149; G. R. Marshall and R. B. Merrifield, *Biochemistry*, 1965, **4**, 2394.
[120] B. Halpern, L. Chew, V. Close, and W. Patton, *Tetrahedron Letters*, 1968, 5163.
[121] H. C. Beyerman, H. Hindricks and E. W. B. de Leer, *Chem. Comm.*, 1968, 1668.
[122] M. Ohno and C. B. Anfinsen, *J. Amer. Chem. Soc.*, 1967, **89**, 5994.
[123] M. A. Tilak and C. S. Hollinden, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 173.
[124] M. A. Tilak and C. S. Hollinden, *Tetrahedron Letters*, 1968, 1297.
[125] F. Weygand, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 183.



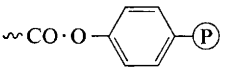
(48)



(49)



(50)

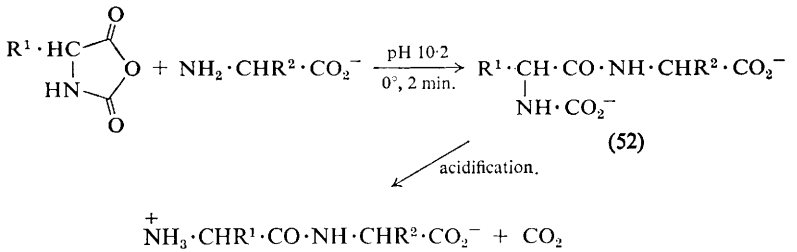


(51)

[(50): bond formed by reaction of an acylamino-acid with a bromoacyl-polystyrene; cleaved by sodium thiophenoxide], and a phenyl ester bond [126] [(51): bond formed by use of dicyclohexylcarbodi-imide; cleaved by mild alkali, methanolic ammonia, *etc.*] have been investigated and found to have some advantages.

. The use of polymeric activated esters has been mentioned in an earlier section (see p. 189).

The Use of N-*Carboxy-anhydrides.* The use of *N*-carboxy-anhydrides for the controlled synthesis of peptides in aqueous media without the isolation of intermediates was first reported in 1966.[127] High yields of peptides are obtained when an *N*-carboxy-anhydride is added to an aqueous solution of an amino-acid or peptide under very carefully controlled conditions (*e.g.* Scheme 13 for the preparation of a dipeptide). Detailed investigations of



(52)

Scheme 13

the mechanism of the reaction and of the optimum conditions for the preparation of dipeptides have now been described.[128] Control of the pH is vital because at pH values below *ca.* 10 the dipeptide carbamate (52) loses carbon dioxide giving the dipeptide at a rate sufficient to cause significant competition with the original amino-component for the *N*-carboxy-

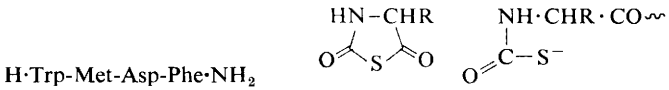
[126] N. Inukai, K. Nakano, and M. Murakami, *Bull. Chem. Soc. Japan*, 1968, **41**, 182.

[127] R. G. Denkewalter, H. Schwam, R. G. Strachan, T. E. Beesley, D. F. Veber, E. F. Schoenewaldt, H. Barkemeyer, W. J. Paleveda jun., T. A. Jacob, and R. Hirschmann, *J. Amer. Chem. Soc.*, 1966, **88**, 3163.

[128] R. Hirschmann, R. G. Strachan, H. Schwam, E. F. Schoenewaldt, H. Joshua, H. Barkemeyer, D. F. Veber, W. G. Paleveda jun., T. A. Jacob, T. E. Beesley, and R. G. Denkewalter, *J. Org. Chem.*, 1967, **32**, 3415.

anhydride (*i.e.* 'over-reaction' occurs). On the other hand, if the pH is greater than *ca.* 10·5, hydrolysis of the *N*-carboxy-anhydride and other side-reactions occur. Very rapid dissolution of the *N*-carboxy-anhydride is necessary to avoid 'over-reaction' as conventional methods of mixing are not adequate, and application of a Waring Blendor is recommended. The absence of racemisation in these reactions was proved by the fact that tritium was not incorporated into the product when the reaction was performed in tritiated water.[127]

The product from one reaction of this type can be coupled directly with a second *N*-carboxy-anhydride without further ado and the process can be repeated several times, so that oligopeptides can be prepared very rapidly without isolating any intermediates and using easily prepared starting materials (many of the side-chain functional groups which are usually protected in other methods of peptide synthesis can be left exposed without detrimental effects). As an example, the *C*-terminal tetrapeptide of gastrin (53) was prepared in *ca.* 1 hr., using phenylalanineamide as the amino-component in the first step.



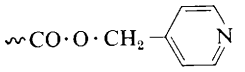
(53) (54) (55)

The thio-analogues (54) of *N*-carboxy-anhydrides (2,5-thiazolidinediones) can be used in the same way.[100, 101] The thiocarbamates (55) which are produced are more stable than ordinary carbamates, and over-reaction is therefore a less serious problem with these, and lower pH values can be used, with a consequent reduction in hydrolysis of the anhydride and other side-reactions induced by base. Because they are particularly prone to yield by-products as a consequence of conversion to isocyanates, the *N*-carboxy-anhydrides of histidine and glycine are not suitable, and in these cases the 2,5-thiazolidinediones are preferred. Unfortunately the thio-anhydrides (54) give rise to some racemisation.

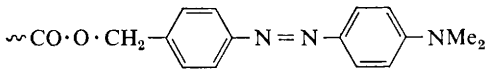
The *N*-carboxy-anhydride and *N*-thiocarboxy-anhydride methods were used in the preparation of the fragments required for the synthesis of ribonuclease S-protein; the only side-chains which were protected were those of lysine and cysteine (see later).

Other Methods. The advantage of peptide synthesis on a solid support is that removal of co-products and excess of reagents after each acylation stage is simply a matter of filtration and washing operations. This advantage is, however, to some extent offset by the fact that because the growing peptide is covalently linked to a polymeric matrix, the *N*-termini are not equally accessible for reaction with molecules in solution, with the consequent difficulties discussed earlier (see p. 193). Clearly what is required is a means by which a peptide can be reversibly attached to an insoluble

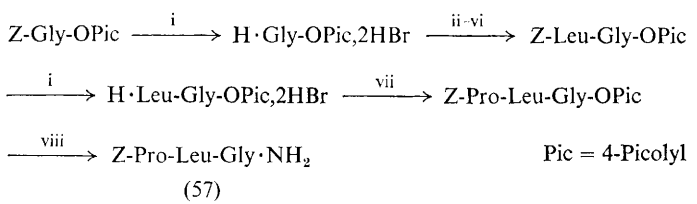
support to enable the reactions to be conducted in homogeneous solution whilst employing an insoluble peptide–polymer complex for separation of the peptide from the reaction mixture. One way of doing this is by giving the growing peptide a basic 'handle' at its carboxyl terminal by means of which it can be reversibly bound to an acidic ion-exchange resin. Preliminary studies of the use of 4-picolyl esters (56) in this way have been published.[129] This method is exemplified by the synthesis (Scheme 14) of the protected tripeptide amide (57), which was obtained in 52% overall yield in



(56)



(58)



(57)

Operations: i, HBr,AcOH; ii, Z-Leu-OPcp,NEt_3,THF; iii, filtered from NEt_3,HBr; iv, adsorbed on sulphoethyl-Sephadex; v, washed with THF; vi, eluted with DMF,H_2O,NEt_3; vii, Z-Pro-OPcp,NEt_3,THF, and then operations iii–vi repeated; viii, NH_3,MeOH; Pcp = pentachlorophenyl

Scheme 14

an analytically and chromatographically pure condition. All acylation steps were forced to completion by an excess of active ester. The 4-picolyl group resists acidolysis conditions but is cleaved by hydrogenation: it can also be converted to a *C*-terminal amide (as in Scheme 14) or to a hydrazide which could then be used for fragment condensation by the azide route. The group (58) has been examined for similar use as a basic handle; in this case the adsorption and elution from the acidic resin can be followed visually, as the derivatives are intensely coloured.[130]

The preliminary reports on this approach are very promising, although the method requires more time and effort than either the 'solid phase' or *N*-carboxy-anhydride methods, the danger of the formation of defective sequences (which can be produced in the *N*-carboxy-anhydride method by 'over-reaction' and in the 'solid phase' method by 'under-reaction') is completely eliminated.

D. Synthesis of Cyclic Peptides.—See also chapter 4, section 2. The self-condensation of peptide active esters in dilute solution is the most

[129] R. Camble, R. Garner, and G. T. Young, *Nature*, 1968, **217**, 247; R. Garner, D. J. Schafer, W. B. Watkins, and G. T. Young, 'Peptides 1968,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 145.

[130] T. Wieland and W. Racky, *Chimia* (*Switz.*), 1968, **22**, 375.

frequently employed method of cyclisation, but the reaction of 'direct' coupling reagents with free peptides is favoured by some workers. It has been shown that bis-*o*-phenylene pyrophosphite (in diethyl phosphite solution in the presence of imidazole) is a useful reagent for the cyclisation of penta- and hexa-peptides, but when tried for the cyclodimerisation of tripeptides it was found unsatisfactory.[131] Improved conditions for cyclisation reactions employing mixed carbonic anhydrides have been reported and used in syntheses of antamanide and some analogues of it.[132, 133] Two new methods for the preparation of cyclodipeptides (2,5-dioxopiperazines) have been described: one involves heating the free dipeptide in molten phenol,[134] and the other consists of heating the formate salt of the dipeptide methyl ester in a neutral solvent.[135]

E. Synthesis of Polymeric Models for Studies in Protein Chemistry.—See also chapter 2, part II, sections 10–13. The IUPAC–IUB Combined Commission on Biochemical Nomenclature has published tentative rules for the abbreviated nomenclature of synthetic polypeptides (polyamino-acids, sequential polypeptides, *etc.*).[136]

Polyamino-acids. The polymerisation of *N*-carboxy-anhydrides is still by far the best route to poly-α-amino-acids. The kinetics and mechanisms of such polymerisations (which are very complex) continue to attract attention.[137-139] Poly-L-serine has been prepared *via* poly-*O*-t-butyl-L-serine, which was itself obtained by polymerisation of *O*-t-butyl-L-serine *N*-carboxy-anhydride.[140] The use of alkyl esters of melphalan as initiating bases in *N*-carboxy-anhydride polymerisations has been reported; the resulting polypeptidyl derivatives of melphalan are tumor-inhibitors.[141] A detailed study of the production of modified poly-L-glutamic acids by aminolysis of poly-γ-methyl-L-glutamic acid has appeared.[142]

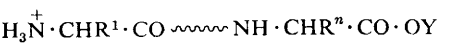
Further studies on the formation of polyamino-acids by heating *N*-dithiocarbonylalkoxycarbonylamino-acids (59) have been reported by Higashimura and his colleagues.[143] These authors conclude, on the basis of a very involved and circuitous argument, that the polymerisation of the

[131] A. W. Miller and P. W. G. Smith, *J. Chem. Soc.* (*C*), 1967, 2144.
[132] T. Wieland, J. Faesal, and H. Faulstich, *Annalen*, 1968, **713**, 201.
[133] T. Wieland, J. Faesal, and W. Konz, 'Peptides 1968,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 243.
[134] K. D. Kopple and H. G. Ghazarian, *J. Org. Chem.*, 1968, **33**, 862.
[135] D. E. Nitecki, B. Halpern, and J. W. Westley, *J. Org. Chem.*, 1968, **33**, 864.
[136] 'Tentative Rules for Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino-acids).' IUPAC–IUB Combined Commission on Biochemical Nomenclature, *Biochemistry*,1968, **7**, 483.
[137] A. Zilkha, H. Rosen, and D. Gertner, *European Polymer J.*, 1967, **3**, 13.
[138] Y. Iwakura, K. Uno, and M. Oya, *J. Polymer Sci., Part A*-1, *Polymer Chem.*, 1967, **5**, 2867; 1968, **6**, 2165.
[139] A. Zilkha, G. Friedman, and D. Gertner, *Canad. J. Chem.*, 1967, **45**, 2979.
[140] N. M. Tooney and G. D. Fasman, *Biopolymers*, 1968, **6**, 81.
[141] M. Szerkerke, R. Wade, and F. Bergel, *J. Chem. Soc.* (*C*), 1968, 1792.
[142] A. Kotai, *Acta Chem. Acad. Sci. Hung.*, 1967, **54**, 65.
[143] H. Kato, T. Higashimura, and S. Okamura, *Makromol. Chem.*, 1967, **109**, 9.

α-amino-acid derivatives (59) proceeds *via* isothiocyanates (60) rather than by formation and polymerisation of a thio-analogue of an *N*-carboxy-anhydride.

(59)

(60)



(61)

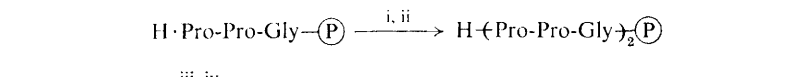
Sequential Polypeptides. The most popular method for the synthesis of sequential polypeptides is *via* the peptide active ester salts (61), which are treated with base, usually under conditions of high concentration in a polar organic solvent. If the *C*-terminal amino-acid of the monomer is other than glycine or proline there is inevitably some uncertainty about the optical integrity of the resulting polymer. Judicious choice of conditions can reduce the extent of racemisation to very low levels, but even very slight racemisation may lead to disproportionately serious conformational consequences. The severity of the problem is accentuated by the fact that reliable methods for the detection of small amounts of racemisation in sequential polypeptide synthesis are not available.

Several sequential polypeptide syntheses using active ester methods have been reported during the year: these are listed elsewhere in this chapter (see p. 209). It is unfortunate that many of the conclusions drawn from structural and biological studies of these products cannot be accepted with confidence simply because, in some cases, the polymers have been inadequately characterised. It is clearly unsatisfactory to attach any great significance to results obtained with materials of uncertain constitution. DeTar discussed this matter very forcefully in 1967 [144] and made a number of recommendations about the minimum characterisation which ought to be given in any paper on sequential polypeptides: it is urged that these recommendations should be adopted by all workers in the field.

Condensation-polymerisation of peptide active esters (61) always gives a polydisperse product. It is therefore of great interest that Sakakibara's group have recently reported [145] a stepwise 'solid phase' synthesis (Scheme 15) of practically monodisperse poly-L-prolyl-L-prolylglycine of molecular weight *ca.* 5000, The well-known difficulties associated with incomplete acylation at any stage in 'solid phase' syntheses are not serious in this

[144] D. F. DeTar, 'Peptides,' ed. H. C. Beyerman, A. van de Linde, and W. Maassen van den Brink, North-Holland Publishing Co., Amsterdam, 1967, p. 125.

[145] S. Sakakibara, Y. Kishida, Y. Kikuchi, R. Sakai, and K. Kakuichi, *Bull. Chem. Soc. Japan.*, 1968, **41**, 1273.



Operations: i, Aoc-Pro-Pro-Gly,OH·DCCI; ii, HCl,AcOH; iii, operations (i) and (ii) repeated 18 times; iv, HF, then dialysis, Aoc = t-amyloxycarbonyl

Scheme 15

special case: since the unit being added is the same at each step, incomplete reaction does not produce an erroneous sequence but only causes a slight departure from monodispersisity in the final sequential polypeptide.

3 Syntheses Achieved

Insufficient space is available for detailed consideration of all the syntheses reported, and the discussion in this section will be confined to a few selected natural polypeptides of particular interest which have been totally synthesised for the first time.

A. Bovine Pancreatic Ribonuclease.—Widespread coverage in the popular and daily press has ensured that all readers will be aware that two groups announced [146, 147] at the same time early in 1969 that they had achieved syntheses (by fundamentally different routes) of material with ribonuclease activity.

Gutte and Merrifield [146] used the automated 'solid phase' method, the 369 reactions and 11,931 operations being performed without any intermediate isolation stages. Each of the 124 amino-acid residues was introduced as an N^{α}-t-butoxycarbonyl derivative, usually by means of dicyclohexylcarbodi-imide: all functional groups except histidine side-chains were protected (with groups resistant to trifluoroacetic acid), and methionine was incorporated as its sulphoxide. Simultaneous removal of all protecting groups and cleavage of the resin–peptide link by means of hydrogen fluoride, followed by conversion to the *S*-sulphonate, purification, mercaptoethanol-reduction, and finally air-oxidation, gave a mixture. The major component of this mixture was eluted from the ion-exchange column at the same position as reduced and reoxidised natural ribonuclease A. After lyophilisation, 85 mg. of material were obtained, and a very careful comparison with the natural enzyme was therefore possible. The synthetic and natural enzymes were indistinguishable by electrophoresis, their

[146] B. Gutte and R. B. Merrifield, *J. Amer. Chem. Soc.*, 1969, **91**, 501.

[147] R. G. Denkewalter, D. F. Veber, F. W. Holly, 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.*, 1969, **91**, 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.*, 1969, **91**, 505; D. F. Veber, S. L. Varga, J. D. Milkowski, H. Joshua, J. B. Conn, R. Hirschmann, and R. G. Denkewalter, *ibid.*, 1969, **91**, 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.*, 1969, **91**, 507.

peptide maps after tryptic digestion of performic acid-oxidised samples were very similar, and amino-acid analyses of hydrolysates of the synthetic material were in reasonable agreement with those of natural ribonuclease A which had been subjected to the conditions of the final deprotection stage. Furthermore, not only was the synthetic enzyme biologically active, but it showed the expected substrate specificity, and its Michaelis constant was practically the same as that of the natural enzyme. The specific activity (which, compared to the native enzyme, was 13% when assayed with yeast RNA and 24% when 2′,3′-cyclic cytidine phosphate was substrate) was a little disappointing, but the authors are quick to point out that the presence of molecules closely related to but not identical with the natural enzyme cannot be excluded.

The synthesis by Denkewalter, Veber, Holly, Hirschmann, and their collaborators[147] follows a totally different strategy. These workers prepared the amino-acid sequence of ribonuclease S-protein by condensation of oligopeptide fragments which were themselves obtained by a combination of the *N*-carboxy-anhydride method (see p. 196) developed by the same group with the use of t-butoxycarbonylamino-acid succinimido esters. The oligopeptide fragments (19 in all) were linked together by azide coupling reactions under anhydrous conditions to give the partially protected tetrahectapeptide which was obtained in a crude yield of *ca.* 5 mg. The success of a synthesis involving so many azide condensations must owe much to the work of Honzl and Rudinger,[148] who developed techniques for performing azide coupling reactions under anhydrous conditions at low temperatures. Other important points about this synthesis which are of general significance in peptide synthesis include the use of a new *S*-protecting group (the *S*-acetamidomethyl group: see p. 184), the conditions developed for selective hydrazinolysis of an alkyl ester without affecting an arginine side-chain, and the success of the azide reactions in the presence of so many exposed functional groups (the only side-chains which were blocked were those of cysteine and lysine). Treatment of the tetrahectapeptide derivative with anhydrous hydrofluoric acid (using methionine and anisole, as scavengers) to remove *N*-benzyloxycarbonyl groups, followed by gel filtration gave *ca.* 78γ of *S*-acetamidomethylated S-protein, which was located in the eluate by calibration of the column with *S*-acetamidomethylated *S*-protein prepared from the natural enzyme. After removal of the cysteine blocking groups (mercuric acetate in aqueous acetic acid), treatment with mercaptoethanol, further gel filtration, and finally air-oxidation in the presence of S-peptide (previously synthesised by Hofmann and his colleagues[149]), a solution containing *ca.* 2γ of ribonuclease S′ activity was obtained. Treatment under the same conditions of a similar amount of *S*-acetamidomethylated S-protein derived from the natural

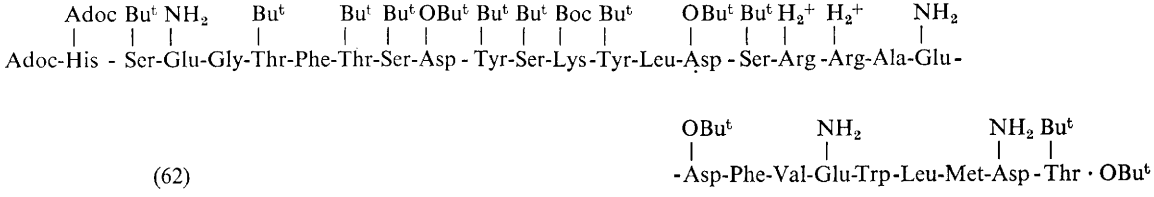
[148] J. Honzl and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1961, **26**, 2333.
[149] K. Hofmann, M. J. Smithers, and F. M. Finn, *J. Amer. Chem. Soc.*, 1966, **88**, 4107.

enzyme gave about three times as much activity, so the synthetic material had a specific activity of *ca.* 30%. Although the specific activity of this preparation was greater than that obtained by Gutte and Merrifield and the synthesis was unambiguous in the sense that intermediates were isolated, the minute amount of end-product obtained prevented full characterisation. However, as only preliminary results of the two syntheses have been published so far, we can anticipate that improvements, particularly in the final stages, will give products in better yield of higher specific activity; it would be invidious to make further comparison of these two great achievements at this early stage.

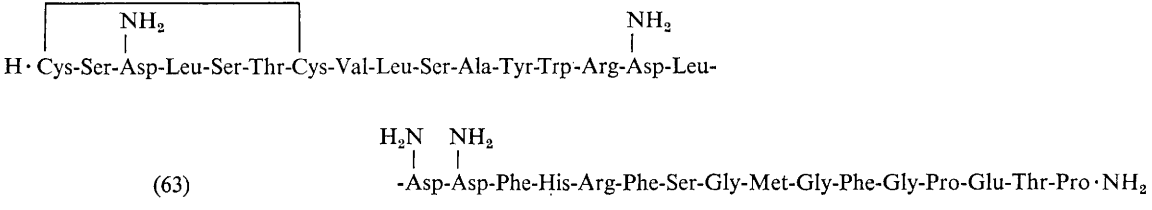
B. Apoferredoxin.—Bayer and his co-workers have reported[111, 112] the synthesis of the amino-acid sequence of ferredoxin, using Merrifield's method. This pentapentcontapeptide contains eight cysteine residues. These were blocked by benzyl protecting groups, which were removed in the last step by means of sodium in ammonia. The synthetic material was isolated and purified as the *S*-sulphonate: the spectroscopic, chromatographic, and sedimentation properties of the product were in good agreement with the behaviour of apoferredoxin *S*-sulphonate from natural sources. The results of amino-acid analysis were, however, far from convincing, and the presence of defective sequences seems probable. Part of the trouble may have been due to the use of sodium in ammonia, and it may be that a different *S*-protecting group or the use of hydrofluoric acid for simultaneous deprotection and cleavage from the resin will yield more satisfactory material.

C. Glucagon.—The pancreatic hyperglycaemic hormone glucagon has finally succumbed to total synthesis in the hands of Wünsch and his collaborators;[150–153] work towards this end has been in progress for some years (the final report[153] of success was number 19 in the series!). The main problems with this peptide are the preponderance of amino-acids with functional side-chains and the fact that the molecule contains only one glycine residue and no proline. Despite this difficulty, a fragment condensation approach was adopted, in which the dicyclohexylcarbodiimide-*N*-hydroxysuccinimide procedure was employed to minimise racemisation in the linking of the fragments. Protecting groups derived from t-butanol were used for all the side-chains except that of the histidine residue, which was introduced as its N^{α},N^{im}-bis-(adamantyloxycarbonyl) derivative. The deprotection of the blocked complete sequence (62) with trifluoroacetic acid required very carefully controlled conditions to reduce side-reactions, but the crude glucagon obtained after ion-exchange chromatography had a specific activity of *ca.* 50%. Crystalline synthetic glucagon

[150] E. Wünsch, *Z. Naturforsch.*, 1967, **22b**, 1269.
[151] E. Wünsch, G. Wendlberger, E. Jaeger, and R. Scharf, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 229.
[152] E. Wünsch and G. Wendlberger, *Chem. Ber.*, 1968, **101**, 3659.
[153] E. Wünsch, E. Jaeger, and R. Scharf, *Chem. Ber.*, 1968, **101**, 3664.


Adoc Bu^t NH$_2$ Bu^t Bu^t Bu^t OBu^t Bu^t Bu^t Boc Bu^t OBu^t Bu^t H$_2$$^+$ H$_2$$^+$ NH$_2$
Adoc-His - Ser-Glu-Gly-Thr-Phe-Thr-Ser-Asp - Tyr-Ser-Lys-Tyr-Leu-Asp - Ser-Arg -Arg-Ala-Glu-
OBu^t NH$_2$ NH$_2$ Bu^t
-Asp-Phe-Val-Glu-Trp-Leu-Met-Asp - Thr · OBu^t

(62)


NH$_2$ NH$_2$
H · Cys-Ser-Asp-Leu-Ser-Thr-Cys-Val-Leu-Ser-Ala-Tyr-Trp-Arg-Asp-Leu-
H$_2$N NH$_2$
-Asp-Asp-Phe-His-Arg-Phe-Ser-Gly-Met-Gly-Phe-Gly-Pro-Glu-Thr-Pro · NH$_2$

(63)

was obtained from the major fraction resulting from gel filtration. The synthetic material was identical with the natural hormone in every respect and had full biological activity.

D. Secretin.—The intestinal hormone secretin has an amino-acid sequence which is in parts closely related to that of glucagon, and its synthesis therefore presents similar problems. It has been synthesised by a fragment condensation route [154] (using azide reactions for the fragment coupling stages) and by a stepwise route [155] using active esters. Both syntheses yielded, after purification, the required heptacosapeptide amide with the biological activity and potency of the natural (porcine) hormone.

E. Gastrins.—Syntheses of human,[156, 157] ovine [158] and bovine [158] gastrins have been described. Of particular general interest is Morley's discussion [156] of the use, in a routine manner, of active esters for stepwise synthesis of oligopeptide fragments with minimal purification of intermediates.

F. Thyrocalcitonin.—Two syntheses of the (porcine) hypocalcaemic hormone thyrocalcitonin (63) have been reported. One of these (by Guttmann *et al.*[159]) used essentially conventional methods and strategy, but the other (by Sieber *et al.*[11]) employed a new extremely acid-labile protecting group to permit selective exposure of an *N*-terminal in the presence of t-butyl-containing protecting groups (see p. 177). Sieber and his colleagues have also reported a synthesis of the closely related human hormone calcitonin M using a strategy similar to that used by them for thyrocalcitonin, but with the additional interesting feature that the cystine bridge was formed by treatment of a bis-*S*-tritylcysteine derivative with iodine in boiling methanol (see p. 185).[160]

4 Appendix: A List of Syntheses Reported in 1968

A. Natural Acyclic Peptides, including Analogues and Partial Sequences.—Analogues and partial sequences are listed under the peptide to which they are related.

[154] M. A. Ondetti, V. L. Narayanan, M. von Saltza, J. T. Sheehan, E. F. Sabo, and M. Bodanszky, *J. Amer. Chem. Soc.*, 1968, **90**, 4711.
[155] M. Bodanszky, M. A. Ondetti, S. D. Levine, and N. J. Williams, *J. Amer. Chem. Soc.*, 1967, **89**, 6753.
[156] J. S. Morley, *J. Chem. Soc.* (*C*), 1967, 2410.
[157] J. Beacham, P. H. Bentley, G. W. Kenner, J. K. MacLeod, J. J. Mendive, and R. C. Sheppard, *J. Chem. Soc.* (*C*), 1967, 2520.
[158] K. L. Agarwal, J. Beacham, P. H. Bentley, R. A. Gregory, G. W. Kenner, R. C. Sheppard, and H. J. Tracy, *Nature*, 1968, **219**, 614.
[159] S. Guttman, J. Pless, E. Sandrin, P. A. Jacquenoud, H. Bossert, and H. Willems, *Helv. Chim. Acta*, 1968, **51**, 1155.
[160] P. Sieber, M. Brugger, B. Kamber, B. Riniker, and W. Rittel, *Helv. Chim. Acta.*, 1968, **51**, 2057.

Peptide	*Ref.*
Adrenocorticotropins	
Human ACTH	161
Analogues and partial sequences	161, 162
Angiotensins	
Human angiotensin I	163
Analogues of angiotensin II	164
Apoferredoxin	111, 112
Bradykinin	4, 6, 80, 165
Analogues	166
Caerulin	
Analogues	167
Calcitonin M	160
Cytochrome C	
Peptides related to the *C*-terminus of the horse heart protein	168
Eledoisin	
Analogues of the *C*-terminal hexapeptide sequence	169
Gastrins	
Human gastrin	156, 157
Ovine and bovine gastrins (which are identical)	158
Analogues and partial sequences	170

[161] S. Bajusz, Z. Paulay, Z. Láng, K. Medzihradsky, L. Kisfaludy, and M. Löw, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 237.

[162] K. Hofmann, H. Bohn, and R. Andreatta, *J. Amer. Chem. Soc.*, 1967, **89**, 7126; W. Oelofsen and C. H. Li, *J. Org. Chem.*, 1968, **33**, 1581; J. Blake and C. H. Li, *J. Amer. Chem. Soc.*, 1968, **90**, 5882.

[163] H. Aoyagi, K. Arakawa, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1968, **41**, 433.

[164] M. A. Cresswell, R. W. Hanson, and H. D. Law, *J. Chem. Soc.* (*C*), 1967, 2669; K. Hofmann, R. Andreatta, J. P. Buckley, W. E. Hageman, and A. P. Shapiro, *J. Amer. Chem. Soc.*, 1968, **90**, 1654; M. C. Khosla, N. C. Chaturvedi, R. R. Smeby, and F. M. Bumpus, *Biochemistry*, 1968, **7**, 3417; W. K. Park, R. R. Smeby, and F. M. Bumpus, *ibid.*, 1967, **6**, 3458.

[165] S. Sakakibara, N. Nakamizo, Y. Kishida, and S. Yoshimura, *Bull. Chem. Soc. Japan*, 1968, **41**, 1477.

[166] K. Suzuki and T. Abiko, *Chem. and Pharm. Bull.* (*Japan*), 1967, **15**, 1508.

[167] J. S. Morley, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, 251; A. Anastasi, L. Bernardi, G. Bosisio, R. De Castiglione, O. Goffredo, G. Bertaccini, V. Erspamer, and M. Impicciatore, *ibid.*, 247; A. Anastasi, L. Bernardi, G. Bosisio, R. De Castiglione, O. Goffredo, G. Bertacini, V. Erspamer, and M. Impicciatore, *Experientia*, 1968, **24**, 771.

[168] M. Frankel, Y. S. Klausner, and Y. Wolman, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam 1968, p. 278.

[169] F. Cuiban, V. Badea, and S. Cilianu, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 256.

[170] H. Gregory, A. H. Laird, J. S. Morley, and J. M. Smith, *J. Chem. Soc.* (*C*), 1968, 522; H. Gregory, D. S. Jones, and J. S. Morley, *ibid.*, 1968, 531; H. Gregory, J. S. Morley, J. M. Smith, and M. J. Smithers, *ibid.*, 1968, 715; J. S. Morley and J. M. Smith, *ibid.*, 1968, 726; H. Gregory and J. S. Morley, *ibid.*, 1968, 910; G. W. Kenner, J. J. Mendive, and R. C. Sheppard, *ibid.*, 1968, 761; K. L. Agarwal, G. W. Kenner, and R. C. Sheppard, *ibid.*, 1968, 1384; L. A. Shchukina, L. Yu. Sklyarov, and V. I. Gorbunov, *Khim. prirod. Soedinenii*, 1968, 120.

Peptide	*Ref.*
Glucagon	150–3
Partial sequences	171
Glutathione	58, 60
Homoglutathione	60
Haemoglobin	
Partial sequences of the β chain	47
Insulin	
A chain	172, 173
A chain analogues and partial sequences	174
B chain	173, 175
B chain partial sequences	123
Lysozyme	
Octapeptide partial sequence	176
Mellitin	
Partial sequences	177
Melanocyte stimulating hormone (MSH)	
α-MSH	30
Monkey β-MSH	32
Tetrapeptide fragment of β-MSH	178
Oxytocin	58, 111, 113, 114, 126, 179, 180
Analogues	97, 180, 181

[171] E. Wünsch, A. Zwick, and A. Fontana, *Chem. Ber.*, 1968, **101**, 326; E. Wünsch, A. Zwick, and E. Jaeger, *ibid.*, 1968, **101**, 336; E. Wünsch and G. Wendlberger, *ibid.*, 1968, **101**, 341; A. A. Costopanagiotis, B. O. Handford, and B. Weinstein, *J. Org. Chem.*, 1968, **33**, 1261; B. O. Handford, T. A. Hytton, K. T. Wang, and B. Weinstein, *ibid.*, 1968, **33**, 4251.

[172] S. Hornle, *Z. physiol. Chem.*, 1967, **348**, 1355; G. Weitzel, U. Weber, S. Hornle, and F. Schneider, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 171.

[173] H. Zahn, W. Danho, G. Schmidt, J. Dahlmans, A. Costopanogiotis, and E. Engels, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 217.

[174] K. Jošt and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1968, **33**, 109; H. Zahn and E. T. J. Folsche, *Annalen*, 1968, **716**, 164; U. Weber, S. Hornle, P. Kohler, G. Nagelschneider, K. Eisele, and G. Weitzel, *Z. physiol. Chem.*, 1968, **349**, 512.

[175] H. Zahn and G. Schmidt, *Tetrahedron Letters*, 1967, 5095.

[176] P. Jollès and J. Jollès, *Helv. Chim. Acta*, 1968, **51**, 980.

[177] E. Schröder, *Annalen*, 1968, **711**, 227.

[178] H. Yajima, Y. Kinomura, T. Oshima, and Y. Okada, *Chem. and Pharm. Bull. (Japan)*, 1967, **15**, 1922.

[179] O. V. Kesarev, V. F. Martynov, and M. I. Titov, *Zhur. obshchei Khim.*, 1968, **38**, 190.

[180] D. Yamashiro, D. B. Hope and V. du Vigneaud, *J. Amer. Chem. Soc.*, 1968, **90**, 3857.

[181] H. Takashima, V. du Vigneaud, and R. B. Merrifield, *J. Amer. Chem. Soc.*, 1968, **90**, 1323; D. Yamashiro, H. L. Aanning, L. A. Branda, W. D. Cash, V. V. S. Murti, and V. du Vigneaud, *ibid.*, 1968, **90**, 4141; D. Yamashiro and V. du Vigneaud, *ibid.*, 1968, **90**, 487; H. Schulz and V. du Vigneaud, *J. Medicin. Chem.*, 1967, **10**, 1037; E. Klieger, *Experientia*, 1968, **24**, 13; M. Manning, T. C. Wuu, J. W. M. Baxter, and W. H. Sawyer, *ibid.*, 1968, **24**, 659; A. Chimiak, K. Eisler, K. Jošt, and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1968, **33**, 2918; Zh. D. Bespalova, V. F. Martynov, and M. I. Titov, *Zhur. obshchei Khim.*, 1968, **38**, 1684; O. A. Kaurov, V. F. Martynov, and V. B. Morozov, *ibid.*, 1968, **38**, 720; O. A. Kaurov, I. M. Lushchitskaya and V. F. Martynov, *ibid.*, 1968, **38**, 724.

Peptide	*Ref.*
Physalaemin	
Analogues and fragments	126, 182, 183
Ribonuclease	
Ribonuclease A (bovine pancreatic)	146
Partial sequences of ribonuclease A	101, 147, 184
Analogues of the *N*-terminal eicosapeptide ('S-peptide')	185
Protected pentapeptide corresponding to the sequence 12–16 of ribonuclease T	186
Ribonuclease S′	147
Secretin	154, 155
Thyrocalcitonin	11, 159
Trypsinogen	
Peptides related to the *N*-terminal sequence	187
Vasopressins	
Lysine–vasopressin	188
Analogues of lysine–vasopressin	189

B. Cyclic Peptides.—

Albomycin–*cyclo*-hexapeptides related to the peptide part	190
Antamanide (*cyclo*-decapeptide)	133
Bacitracin–*cyclo*-dodecapeptide analogue	191
Cyclo-hexapeptides containing glycine, leucine and phenylalanine	192
Gramicidin S	193
Analogues	194

[182] E. Schnabel and A. Oberdorf, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 261.
[183] L. Bernardi, G. Bosisio, R. De Castiglione, and O. Goffredo, *Gazzetta*, 1967, **97**, 1837, 1848; R. De Castiglione and G. Bosisio, *ibid.*, 1967, **97**, 1858.
[184] S. Visser, J. Roeloffs, K. E. T. Kerling, and E. Havinga, *Rec. Trav. chim.*, 1968, **87**, 559; M. A. Ruttenberg, *J. Amer. Chem. Soc.*, 1968, **90**, 5598; V. G. Degtyar' and L. A. Shchukina, *Khim. prirod. Soedinenii*, 1968, 39 and 178.
[185] R. Rocchi, L. Moroder, F. Marchiori, E. Ferrarese, and E. Scoffone, *J. Amer. Chem. Soc.*, 1968, **90**, 5885; F. Marchiori, R. Rocchi, L. Moroder, A. Fontana, and E. Scoffone, *ibid.*, 1968, **90**, 5889.
[186] M. Waki, N. Mitsuyasu, T. Kato, S. Makisumi, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1968, **41**, 669.
[187] J. Savrda and E. Bricas, *Bull. Soc. chim. France*, 1968, 2423.
[188] J. Meienhofer and Y. Sano, *J. Amer. Chem. Soc.*, 1968, **90**, 2996.
[189] D. Gillessen and V. du Vigneaud, *J. Biol. Chem.*, 1967, **242**, 4806.
[190] A. M. El'Naggar and N. A. Poddubnaya, *Zhur. obshchei Khim.*, 1968, **38**, 44, 450; A. M. El'Naggar and N. A. Poddubnaya, *Vestnik Moskov. Univ.*, 1967, No. 5, 152.
[191] Y. Aryoshi, T. Shiba, and T. Kaneko, *Bull. Chem. Soc. Japan*, 1967, **40**, 2648.
[192] L. Mladenova-Orlinova, K. Blaha, and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1967, **32**, 4070.
[193] H. Klostermeyer, *Chem. Ber.*, 1968, **101**, 2823.
[194] M. Waki and N. Izumiya, *Bull. Chim. Soc. Japan*, 1968, **41**, 1909; M. Waki, O. Abe, R. Okawa, T. Kato, S. Makisumi, and N. Izumiya, *ibid.*, 1967, **40**, 2904; J. Halstrom and H. Klostermeyer, *Annalen*, 1968, **715**, 208; M. Rothe and F. Eisenbeiss, *Angew. Chem. Internat. Edn.*, 1968, **7**, 883.

Peptide	*Ref.*
C. Sequential Polypeptides.—	
poly-(Ala·Gly·Gly)	195
poly-(Glu·Ala·Gly)	196
poly-(Glu └—Ala·Gly)	196
poly-(Glu └—Gly), L and D	197
poly-{[Glu(OMe)$_x$]·Ser(Ac)·Glu(OMe)} (x = 0, 1, 2, or 3)	198
A series of polypeptides containing Glu(OBzl) residues	199
poly-(Gly·Hyp$_2$)	200
poly-{Gly[Lys(Tos)]$_2$}	201
poly-(Gly·Pro·Gly)	202
poly-[Gly·Pro·Lys(Tos)]	201
poly-(Lys·Ala·Gly)	203
poly-[Lys(Z)·Gly]	199
poly-(Met$_2$·Ala)	204
poly-(Phe·Pro·Gly)	205
poly-(Pro·Phe·Gly)	205
poly-(Pro·Pro·Gly)	145
D. Miscellaneous Peptides.—	
Peptides of β-(pyrazol-1-yl)- and β-(pyrazol-3-yl)-alanine	206
Peptides of *threo*-β-hydroxyaspartic acid	207
Selenium-containing peptides	208
Bacterial cell wall peptides	209

[195] A. Brack and G. Spach, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 45.
[196] B. J. Johnson, *J. Chem. Soc.* (*C*), 1967, 2638.
[197] J. Kovacs, G. N. Schmit, and U. R. Ghatak, *Biopolymers*, 1968, **6**, 817.
[198] F. H. C. Stewart, *Austral. J. Chem.*, 1968, **21**, 1935.
[199] R. Ledger and F. H. C. Stewart, *Austral. J. Chem.*, 1967, **20**, 2509.
[200] V. A. Shibnev, T. P. Chubaeva, and K. T. Poroshin, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1968, 225.
[201] V. A. Shibnev, V. S. Grechishko, and K. T. Poroshin, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1967, 2327.
[202] V. A. Shibnev, K. T. Poroshin, T. P. Chuvaeva, and G. A. Martynova, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1968, 1144.
[203] B. J. Johnson, *J. Chem. Soc.* (*C*), 1968, 3008.
[204] E. Dellacherie and J. Neel, *Comp. rend.*, 1968, **266**, *C*, 527.
[205] A. M. Tamburro, A. Scatturin, and F. Marchiori, *Gazzetta*, 1968, **98**, 638.
[206] K. Hofmann, R. Andreatta, and H. Bohn, *J. Amer. Chem. Soc.*, 1968, **90**, 6207.
[207] Y. Liwschitz, A. Singerman, and S. Sokoloff, *J. Chem. Soc.* (*C*), 1968, 1843.
[208] D. Theodoropoulos, I. L. Schwartz, and R. Walter, *Biochemistry*, **6**, 3927.
[209] P. Lefrancier and E. Bricas, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 293; P. Dezelee and E. Bricas, *ibid.*, p. 299; D. Jarvis and J. L. Strominger, *ibid.*, p. 305.

Peptide	*Ref.*
Peptides of amino-acids with purine or pyrimidine side-chains	210
Tobacco mosaic virus protein, tryptic peptide 2	211
Gly·Asp·Ser·Gly (Merrifield method, with useful experimental detail of general interest)	212
Thr·Ala·Cys·His·Asp	213

[210] C. I. Hong, C. Piantadosi, and J. L. Irvin, *J. Medicin. Chem.*, 1968, **11**, 588; Yu. P. Shvachin, G. A. Korshunova, and T. V. Ivanova, *Vestnik Moskov. Univ.*, 1967, No. 6, 94; R. G. Vdovina and A. V. Karpova, *Khim. prirod. Soedinenii*, 1968, 42.

[211] J. D. Young, C. Y. Leung, and W. A. Rombauts, *Biochemistry*, 1968, **7**, 2475.

[212] R. B. Merrifield and M. A. Corigliano, *Biochem. Prep.*, 1968, **12**, 98.

[213] I. Photaki, V. Bardakos, A. W. Lake, and G. Lowe, *J. Chem. Soc.* (*C*). 1968, 1860.

4
Peptides of Abnormal Structure

By J. S. DAVIES

1 Introduction

In this chapter of the Report the information published during 1968 on peptides containing 'abnormal bonds' and other features not commonly found in proteins has been summarised. An 'abnormal bond' for this purpose has been defined as a bond other than that between α-amino and α-carboxyl groups of amino-acids. A rigid classification of peptides into a group of this kind is very difficult. However, as a basis for this section the class has been taken to include: cyclic peptides (especially peptide antibiotics), depsipeptides, peptides containing thioether linkages (peptides containing disulphide links have not been included), peptides containing links between, for example, β or γ substituents of appropriate amino-acids, and peptides and amino-acids conjugated to lipids, carbohydrates, nucleotides, etc. Discussion on molecules containing mixed functions has been restricted to the peptide moiety and the nature of the linkage between the functions. Biochemical aspects have not been covered, the main emphasis being on structural and synthetic work.

Two excellent reviews,[1, 2] covering the literature up to mid-1966, contain relevant background references. Specific topics of immediate relevance to this section have also been reviewed recently.[3, 4] Aspects of the work published in 1967 have been summarised,[5] and the Proceedings of the Ninth European Peptide Symposium have been published.[6]

During the year, notable progress has been made in the elucidation of the structure of many compounds using mass spectrometry. 'Unusual features' often improve the volatility of peptides, which is a very desirable asset in mass spectrometry. The structure of a number of closely related peptide alkaloids has been elucidated, largely on the basis of mass spectral

[1] E. Schröder and K. Lübke, 'The Peptides', Academic Press, 1966, vol. 2.
[2] S. G. Waley, *Adv. Protein Chem.*, 1966, **21**, 2.
[3] M. M. Shemyakin and Y. A. Ovchinnikov, *Recent Developments in Chem. Natural Carbon Compounds*, 1967, **2**, 1; Y. A. Ovchinnikov, V. T. Ivanov, V. K. Antonov, A. M. Shkrob, I. I. Mikhaleva, A. V. Evstratov, G. G. Malenkov, E. I. Melnik, and M. M. Shemyakin, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 56.
[4] T. Wieland, *Fortschr. Chem. org. Naturstoffe*, 1967, **25**, 214; T. Wieland, *Science*, 1968, **159**, 946.
[5] H. D. Law, *Ann. Reports* (B), 1967, **64**, 451.
[6] 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968.

correlations. The determination of the structure of certain mycosides and of the antibiotic stendomycin demonstrates the importance of this technique for determining the basic structures of a mixture of homologues.

The structure of a number of peptide antibiotics has been elucidated using a combination of classical degradative techniques and modern physical methods. In many cases the antibiotics contain features closely related to antibiotics already characterised. Many are cyclic peptides and a number contain the peptide lactone nucleus. These antibiotics continue to be a source of many novel and unusual amino-acids. Information is also accumulating on the biological origins of these microbial peptides.

Studies on the structure–biological activity relationships of cyclic peptides have produced a number of new analogues of gramicidin S. Similar studies have been continued on cyclic depsipeptides. As a consequence of these studies, new procedures for the cyclisation of peptides and depsipeptides have been developed. A fundamental requirement for biological activity is that the molecule should have the correct topology. D-Amino-acids, and -hydroxy-acids in the case of depsipeptides, are very important residues in this context.

Complete syntheses of a number of actinomycins have been achieved and conclusive evidence has been obtained for the presence of a bis-pentapeptide lactone structure in these compounds.

Studies on the structure of bacterial cell walls have benefited from the availability of new enzyme preparations. The structure of many glycopeptides obtained from enzymic digests of cell walls has provided new information about the nature of the cell wall in the region around the peptide–carbohydrate linkages. The composition of the peptide moiety of a number of nucleotide murein precursors has also been elucidated.

The availability of synthetic compounds of well-defined configuration for structural comparison has been advantageous in the structural determination of a number of complex lipids.

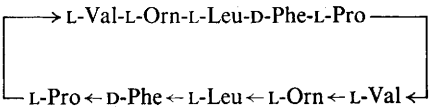
2 Cyclic Peptides

(See also Chapter 3, sections 2D and 4B.)

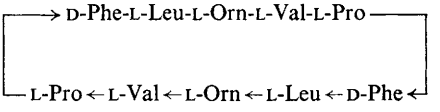
A. Gramicidins.—The Merrifield 'solid phase' technique has been used to synthesise the linear decapeptide analogues of gramicidin S (1),[7] and (Gly^{5,10})-gramicidin S.[8] *t*-Butoxycarbonyl protection of the α-amino groups was used throughout, and the δ-amino groups of the ornithine residues were protected by toluene-p-sulphonyl (tosyl) groups. Dicyclohexylcarbodi-imide (DCCI) and *N*-hydroxysuccinimide esters were employed in the coupling stages and cyclisation of the ditosylated linear decapeptides was achieved using *p*-nitrophenol–DCCI. Removal of the tosyl groups was carried out using sodium–liquid ammonia.

[7] H. Klostermeyer, *Chem. Ber.*, 1968, **101**, 2823.
[8] J. Halstrom and H. Klostermeyer, *Annalen*, 1968, **715**, 208.

Interest in the structure–biological activity relationship of the gramicidins continues, and the synthesis of many analogues have been reported. The cyclodiastereoisomer retrogramicidin S (2) has been synthesised via the linear decapeptide *p*-nitrophenyl ester.[9] Since the hydrophobic proline residues in retrogramicidin are directed to the same 'hydrophilic side' of the pleated plane (based on the *X*-ray crystallographic model) as the ornithine side-chains, this cyclodiastereoisomer, as expected, gave a much weaker antibacterial activity. Cyclodimerisation of the *p*-nitrophenyl ester of the pentapeptide D-Phe-L-Leu-L-Orn(Z)-L-Val-L-Pro, also gave the protected form of (2) in 25% yield, together with *cyclo*-semigramicidin S (26%). (Gly^{5,10})-Gramicidin S and its enantiomer retro(Gly^{5,10})-gramicidin S, on the other hand, showed antimicrobial activity very similar to that of gramicidin S. These analogues were synthesised[10] by cyclodimerisation of the appropriate *C*-terminal glycine pentapeptide nitrophenyl ester, the phthaloyl group being used to protect the δ-amino group of ornithine.



(1)



(2)

HCO·D-Phe-L-Pro-L-Val-L-Orn-L-Leu-D-Phe-L-Pro-L-Val-L-Orn-L-Leu·NH·CH_2·CH_2OH

(3)

Ornithyl residues in gramicidin S can be replaced by either lysine or α,γ-diaminobutyric acid (Dbu) residues without loss of biological activity.[11] In this study (Dbu^{2,7})- and (Lys^{2,7})-gramicidin S were synthesised by cyclodimerisation of the corresponding linear pentapeptide esters. A synthesis of 'all-L-' gramicidin S has also been achieved[12] utilising a stepwise synthesis, from the carboxyl end, of the pentapeptide Val-Orn(Boc)-Leu-Phe-Pro, using the *N*-hydroxysuccinimide ester method. Ring closure to give the cyclic dimer was achieved in 60% yield using *o*-phenylene chlorophosphite. The reaction also gave the cyclopentapeptide as a minor

[9] M. Waki and N. Izumiya, *Tetrahedron Letters*, 1968, 3083.
[10] M. M. Shemyakin, Y. A. Ovchinnikov, V. T. Ivanov, and I. D. Ryabova, *Experientia*, 1967, **23**, 326.
[11] M. Waki, O. Abe, R. Okawa, T. Kato, S. Makisumi, and N. Izumiya, *Bull. Chem. Soc. Japan*, 1967, **40**, 2904.
[12] M. Rothe and F. Eisenbeiss, *Angew. Chem. Internat. Edn.*, 1968, **7**, 883.

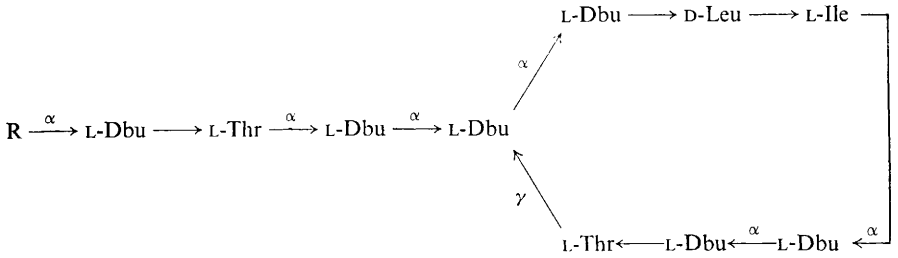
product. Synthesis of (Lys,[2] Gly[4])-gramicidin S and (Lys,[2] Gly,[4,9] Lys[7])-gramicidin S was also reported using the same method. The 'all-L-' isomer and the latter cyclopeptides show no antibiotic activity.

All the studies reported, so far, support the hypothesis[10] that topochemically similar cyclopeptides should show similar biological properties. The presence of the D-amino-acids appears to be an important factor in providing the correct topology for biological activity.

The decapeptide derivative (3), isolated from *Bacillus brevis*, has been suggested[13] as a possible biosynthetic precursor of gramicidin A.

The application of n.m.r. spectroscopy to the study of the conformation of gramicidin S has been explored. Peaks in the 100 MHz spectrum (dimethyl sulphoxide as solvent), at 8·05, 8·30, and 8·60 p.p.m. have been tentatively assigned[14] to groups of NH-protons in different environments. The solvent dependence of these protons was shown by the fact that in deuteriomethanol[15] the NH-protons of diphthaloylgramicidin S appeared at 6·78, 7·64, and 8·67 p.p.m. Using the phthaloyl group attached to the δ-amino group of the ornithine residues as a marker group, upfield shifts of the aromatic protons in gramicidin S have been interpreted[15] as being due to the close proximity of the phenylalanine and N^{δ}-phthaloylornithine side-chains, as required for the antiparallel pleated sheet conformation (based on the *X*-ray diffraction model). With the information available, however, it seems a little premature to draw very definite conclusions about the conformation of gramicidin S from n.m.r. data.

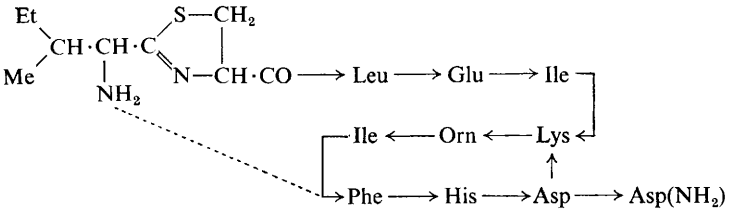
B. Circulin B.—A new modified structure (4) has been proposed[16] for this antibiotic. This structure differs from that of circulin A only in the nature of the fatty acid residue. The deacyl circulin B required for sequence determination was obtained by treating circulin B with polymyxin acylase, which released the *N*-terminal α,γ-diaminobutyric acid (Dbu) peptide. Examination of the partial hydrolysates, after treatment with subtilopeptidase A, led to the structure (4) which is structurally related to compounds of the polymyxin series.



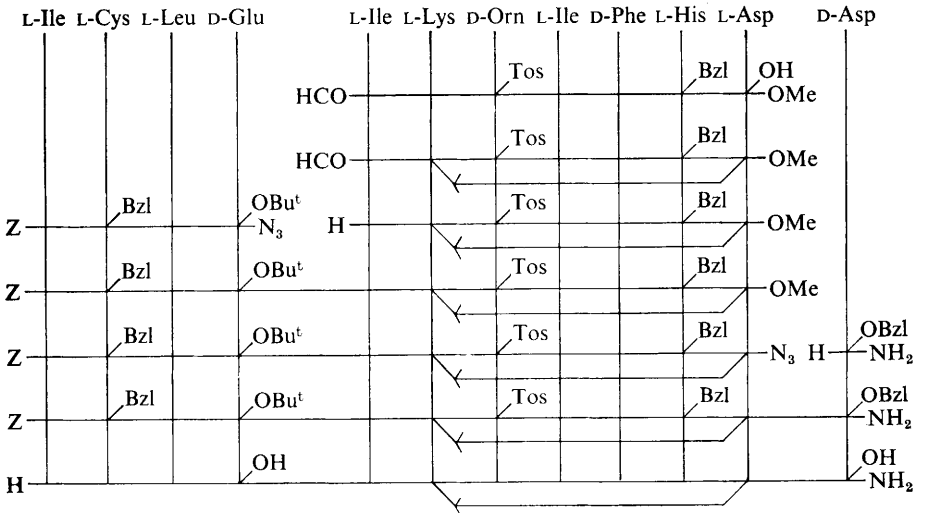
(4) R = iso-octanoyl

[13] L. W. Pollard, N. V. Bhagavan, and J. B. Hall, *Biochemistry*, 1968, **7**, 1153.
[14] A. M. Liquori and F. Conti, *Nature*, 1968, **217**, 635.
[15] R. Schwyzer and U. Ludescher, *Biochemistry*, 1968, **7**, 2519.
[16] B. K. Hayashi, Y. Suketa, and T. Suzuki, *Experientia*, 1968, **24**, 656.

C. Bacitracins.—Although a number of details concerning the postulated structure of bacitracin A remains ambiguous, the structure (5) has been taken as the basis for synthetic studies[17] on this cyclic peptide derived from *Bacillus licheniformis*. Clarification of the type of link between the



(5)



(6)

Scheme 1

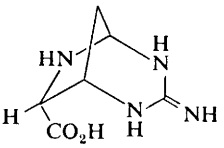
N-terminal leucine and the phenylalanine residue (a cyclol structure has been proposed) is still needed. The cyclic dodecapeptide analogue (6) was synthesised as outlined in Scheme 1, but showed no antibacterial activity. The azide coupling method was used extensively and cyclisation of the heptapeptide was carried out using dicyclohexylcarbodi-imide under high dilution conditions. Although cysteine peptides possessing a suitably protected *N*-terminal residue cyclise readily to peptides containing thiazolidine rings,[18] no evidence of thiazolidine ring formation was

[17] Y. Ariyoshi, T. Shiba, and T. Kaneko, *Bull. Chem. Soc. Japan*, 1967, **40**, 2648.
[18] Y. Hirotsu, T. Shiba, and T. Kaneko, *Bull. Chem. Soc. Japan*, 1967, **40**, 2945; 1967, **40**, 2950.

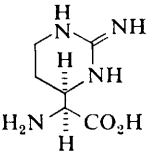
obtained when (6) was subjected to similar conditions. A possible explanation is that the *N*-terminal residue will need to be protected prior to the formation of the thiazolidine ring, and that cyclol formation takes place after the formation of this ring. A synthesis of a thiazole carboxylic acid found as a degradation product of bacitracin F has also been reported.[19]

The application of a micromethod for identifying asparagine and glutamine residues has confirmed[20] that the asparaginyl and aspartyl residues in bacitracin possess the L- and D-configurations respectively. The method involved dehydration of the peptide amides with ethylene chlorophosphite, to give the corresponding nitriles, which after reduction were hydrolysed to the readily recognisable 2,4-diaminobutyric acids.

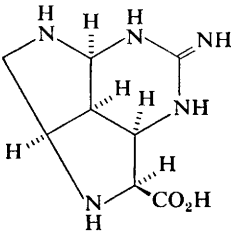
D. Viomycin.—Modifications to the tentative structure for this tuberculostatic antibiotic have been proposed in the light of new degradative evidence.[21] Viomycidine (7), obtained on acid hydrolysis[22] and whose structure was recently confirmed by *X*-ray crystallography,[23] is unlikely to be concerned with the chromophore since the loss of the chromophore on hydrolysis corresponded to an equivalent loss of urea. Viomycidine (7) was probably an artefact in the hydrolysate, since on hydrogenation of viomycin, the total hydrolysate afforded no viomycidine, but instead gave capreomycidine (8). Capreomycidine has also been isolated[24] from the capreomycin group of tuberculostatic antibiotics. The elucidation of the structure of viocidic acid (9), another degradation product, confirmed the presence of a capreomycidine-type residue in viomycin. The data now available have led to



(7)



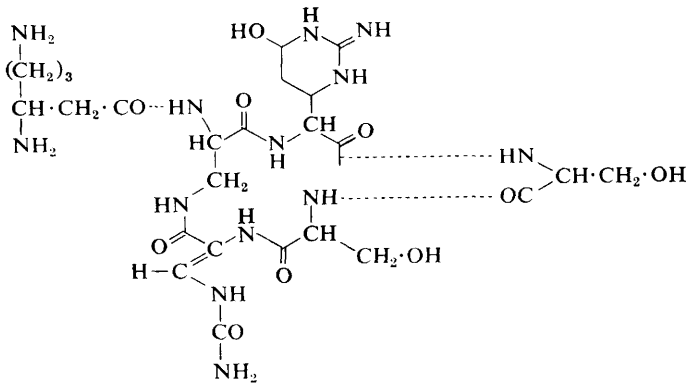
(8)



(9)

the tentative partial structure (10) for viomycin although the relative positions of the serine and β-lysine residues remain ambiguous. The chromophore is now associated with the dehydro-amino-acid residue.

[19] Y. Ariyoshi, T. Shiba, and T. Kaneko, *Bull. Chem. Soc. Japan*, 1967, **40**, 2654.
[20] E. Ratti, C. Lauinger, and C. Ressler, *J. Org. Chem.*, 1968, **33**, 1309.
[21] B. W. Bycroft, D. Cameron, L. R. Croft, A. Hassan-Ali Walji, and T. Webb, *Tetrahedron Letters*, 1968, 5901.
[22] B. W. Bycroft, D. Cameron, L. R. Croft, A. W. Johnson, T. Webb, and P. Coggon, *Tetrahedron Letters*, 1968, 2925; T. Kitigawa, Y. Sawada, T. Miura, T. Ozasa, and H. Taniyama, *ibid.*, 1968, 109.
[23] J. C. Floyd, J. A. Bertrand, and J. R. Dyer, *Chem. Comm.*, 1968, 998.
[24] B. W. Bycroft, D. Cameron, L. R. Croft, and A. W. Johnson, *Chem. Comm.*, 1968, 1301.



(10)

The presence of this residue, together with the fact that it is a cyclic peptide (absence of a free carboxyl group), reveal a possible relationship between viomycin and ostreogrycin, griseoviridin, and telomycin group of antibiotics.

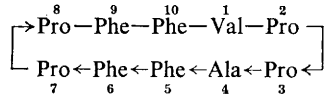
E. Peptides of *Amanita phalloides*.—Authoritative reviews on the poisonous principles of *Amanita* mushrooms have recently appeared.[4]

Antamanide (11), an antitoxic principle which counteracts the lethal action of phalloidine and α-amanitine, has been characterised.[25] A novel micromethod [26] was used to determine the configuration of the amino-acids. The method depends upon the identification, on thin-layer chromatography plates, of the diastereoisomers formed after treating amino-acids from the hydrolysates with the azide of a suitable benzyloxycarbonyl-L-amino-acid. Deprotection of the dipeptide formed was carried out directly on the thin-layer plate by applying drops of 0·1% palladium(II) chloride in methanol to the initial dipeptide, and then placing the plate in an atmosphere of hydrogen. Cellulose powder layers and a pyridine/water (80 : 20 v/v) solvent system gave a satisfactory separation of the diastereoisomers. The amino-acid sequence of (11) was determined by mass spectrometry, a high intensity peak at *m/e* 1146 being taken as indicative of a cyclopeptide. The molecule appeared to break into two pentapeptides in the mass spectrometer, which then gave a stepwise decomposition. Mass spectrometric analysis of fractions obtained by gas chromatography of the methyl esters (as their *N*-trifluoroacetyl derivatives) formed from partial methanolysis of antamanide confirmed the sequence of amino-acids. It is of interest to note that certain peptide methyl esters containing sequences

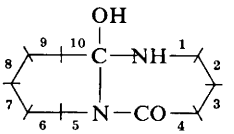
[25] T. Wieland, G. Lüben, H. Ottenheym, J. Faesel, J. X. De Vries, A. Prox, and J. Schmid, *Angew. Chem. Internat Edn.*, 1968, **7**, 204.

[26] T. Wieland and H. Ottenheym, 'Peptides,' eds. H. C. Beyerman, A. van den Linde, and W. Maassen van den Brink, North-Holland Publishing Co., Amsterdam, 1967, p. 195.

not readily explained by structure (11), were also obtained and have been tentatively interpreted as being due to the possible involvement of a cyclol intermediate, such as (12), which would give extraneous peptides on methanolysis. Final confirmation of the structure (11) came from the synthesis of antamanide *via* the cyclisation of a linear decapeptide using the anhydride method.



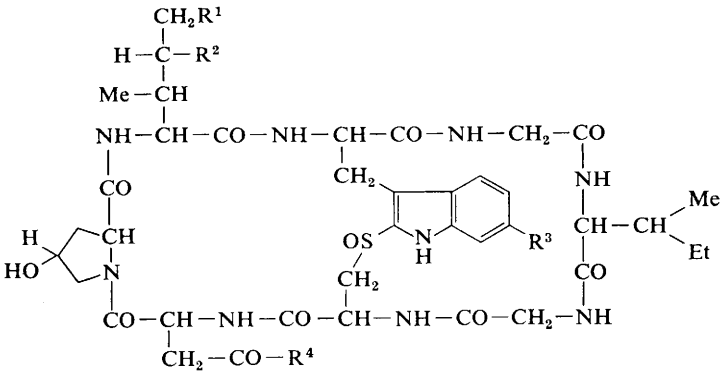
(11)



(12)

Phalloidine has been converted[27] into norphalloine by Raney nickel reduction of the dithiolane derivative of the ketophalloidine obtained by periodate oxidation of phalloidine. No damage to the thio-ether bridge was apparent.

The sulphoxide nature of α-, methyl α-, β-, and γ-amanitines (13–16) and amanine (17) has been proved[28] by comparison of the u.v. spectra with those of the 2-sulphoxides obtained from the peroxide oxidation of phalloidine and with 2-ethylsulphenylscatole. ε-Amanitine (18) has been shown to be the acid of γ-amanitine and structure (19) has been proposed for amanulline.[29]

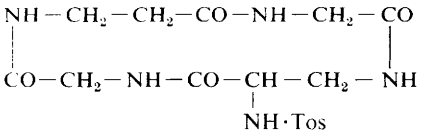


	R^1	R^2	R^3	R^4
(13)	OH	OH	OH	NH_2
(14)	OH	OH	OMe	NH_2
(15)	OH	OH	OH	OH
(16)	H	OH	OH	NH_2

	R^1	R^2	R^3	R^4
(17)	OH	OH	H	OH
(18)	H	OH	OH	OH
(19)	H	H	OH	NH_2

[27] T. Wieland and R. Jeck, *Annalen*, 1968, **713**, 196.
[28] H. Faulstich and T. Wieland, *Annalen*, 1968, **713**, 186.
[29] T. Wieland and A. Buku, *Annalen*, 1968, **717**, 215.

F. Cyclisation Reactions.—Penta- and hexa-peptides have been successfully cyclised using bis-*o*-phenylene pyrophosphite,[30] and good yields of cyclic products have been reported in the synthesis of gramicidin analogues [12] and the cyclic peptide (20) [31] using *o*-phenylene chlorophosphite. A one-step



(20)

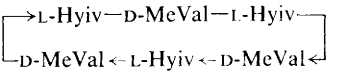
cyclisation of L-Trp-Gly-L-Leu-L-Ala-D-Thr and of seco-oxophalloidine occurred [32] in 30% and 15% yield respectively using ethoxycarbonyl chloride in the presence of pyridine. The former peptide was also cyclised in 13·5% yield using bis(2,4-dinitrophenyl) carbonate as a reagent.

The relative ease of cyclisation of a series of hexapeptides with the sequence $(L\text{-Orn})_3(Ser)_3$ has been shown [33] to be dependent on the sequence and the configuration of the serine residues. Cyclohexapeptides containing two D-serine residues formed in highest yield. The cyclopeptides containing the D-forms were also more labile and were completely hydrolysed by dilute lithium hydroxide.

3 Depsipeptides

A general review, and a summary of structure–biological activity relationships have appeared.[3]

A. Synthesis.—Using standard methods developed in recent years for the synthesis of depsipeptides, 'all L-', 'all D-', and the enantiomer (21) of enniatin B have been synthesised.[34] The synthesis of enantiomer (21) and other analogues has also been reported by another research group.[35] In studies [35] on the structure–biological activity relationship of the enniatins,



Hyiv = α-hydroxyisovaleric acid

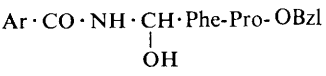
(21)

[30] A. W. Miller and P. W. G. Smith, *J. Chem. Soc.* (*C*), 1967, 2140.
[31] C. H. Hassall, D. G. Sanger, and J. O. Thomas, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 70.
[32] T. Wieland, J. Faesel, and H. Faulstich, *Annalen*, 1968, **713**, 201.
[33] A. M. El'Naggar and N. A. Poddubnaya, *Zhur. obshchei Khim.*, 1968, **38**, 444; A. M. El'Naggar, N. A. Poddubnaya, and M. Ahmed, *ibid.*, 1968, **38**, 450.
[34] G. Losse and H. Raue, *Chem. Ber.*, 1968, **101**, 1532.
[35] M. M. Shemyakin, Y. A. Ovchinnikov, V. T. Ivanov, and A. V. Evstratov, *Nature*, 1968, **213**, 412; I. I. Mikhaleva, I. D. Ryabova, T. A. Romanova, T. I. Tarasova, V. T. Ivanov, Y. A. Ovchinnikov, and M. M. Shemyakin, *Zhur. obshchei Khim.*, 1968, **38**, 1228.

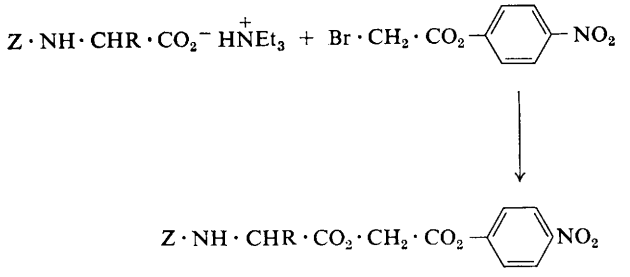
it appears that the amino-acid residues in enniatin B may be altered considerably with little change in biological activity, but modification of the hydroxy-acid residues reduced the activity drastically. Alteration of configuration in individual residues caused a loss of activity, but the optical antipode (21) of enniatin B displayed the same biological activities over the entire range of the antimicrobial spectrum. This appears to be the first case of a naturally occurring compound and its antipode displaying similar biological properties.

Ester bonds are generally more difficult to form than peptide amide bonds and therefore require a higher degree of activation of the carboxyl component. A *p*-nitrophenyl ester in the presence of imidazole activates the carboxyl group sufficiently to bring about depsipeptide ester bond formation at room temperature.[36] The reaction presumably involves the formation of an acyl-imidazole intermediate. Pentamethylbenzyl esters appear to be useful carboxyl-protecting groups for depsipeptide synthesis. The esters are crystalline and undergo rapid cleavage under mild acidic conditions.[37]

Depsipeptide links have been formed [38] in model peptides such as (22), by reacting the peptide with an acid chloride of a *N*-protected amino-acid in the presence of pyridine in tetrahydrofuran. *N*-Acylated depsipeptides



(22)



(23)

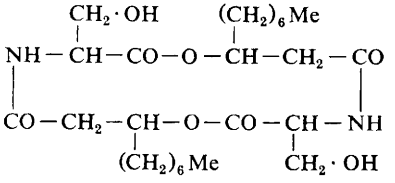
containing glycollic acid residues can be synthesised [39] using *p*-nitrophenyl bromoacetate, to give intermediates such as (23) which can then react further to form peptide bonds.

[36] F. H. C. Stewart, *Austral. J. Chem.*, 1968, **21**, 1639.
[37] F. H. C. Stewart, *Austral. J. Chem.*, 1968, **21**, 1327.
[38] N. A. Krit, G. A. Ravdel, and L. A. Shchukina, *Zhur. obshchei Khim.*, 1968, **38**, 1015.
[39] C. Wasielewski, *Roczniki Chem.*, 1967, **41**, 1915.

The solid-phase technique has been successfully applied [40] to the synthesis of tetradepsipeptide (24).* t-Butyl-protection of the hydroxyl groups was used throughout and t-butoxycarbonylvaline was used in the last stage. Activation of the carboxyl group for each coupling stage was achieved using the benzenesulphonic mixed anhydride. Cleavage of the peptide–resin link required hydrobromic acid in 100% trifluoroacetic acid. The synthesis of the depsipeptide analogues of antiogensin II on a polymer has also been reported.[41]

L-Val-Glyc-L-Hyiv-L-PhLac, HBr

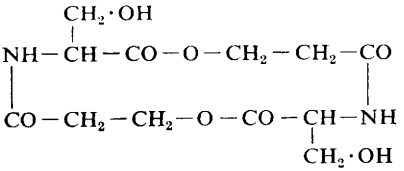
(24)



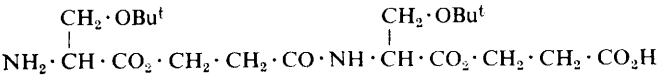
(25)

Experimental details for the synthesis of serratamolide (25) using the hydroxy-acyl insertion method have now been published.[42]

Both the D- and L-seryl enantiomers of the cyclodepsipeptide (26) have been synthesised by cyclodimerisation of the appropriate monomeric acid.[43] Cyclodimerisation of the racemic β-(*O*-t-butylseryloxy)propionic acid, however, gave a mixture containing the L-L and D-D cyclodimers (60%) and the *meso* D-L compound (40%). The latter was synthesised unambiguously by cyclisation of the acid chloride of (27) in the presence of triethylamine under conditions of high dilution.



(26)

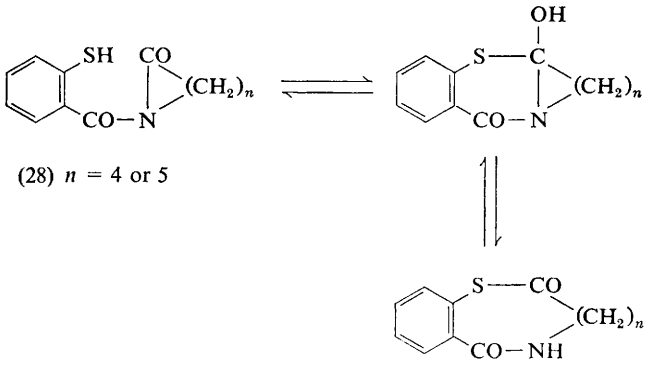


(27)

[40] L. A. Shchukina, E. P. Semkin, and A. P. Smirnova, *Khim. prirod. Soedinenii*, 1967, 358.
[41] E. P. Semkin, A. P. Smirnova, and L. A. Shchukina, *Zhur. obshchei Khim.*, 1968, **38**, 2358.
[42] A. A. Kiryushkin, V. I. Shchelokov, V. K. Antonov, Y. A. Ovchinnikov, and M. M. Shemyakin, *Khim. prirod. Soedinenii*, 1967, **3**, 267.
[43] C. H. Hassall and J. O. Thomas, *J. Chem. Soc.* (*C*), 1968, 1495.

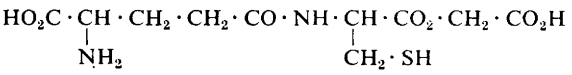
* Glyc = glycollic acid, PhLac = α-phenyllactic acid, Hyiv = α-hydroxyisovaleric acid.

cyclo-Thiodepsipeptides have been synthesised for the first time *via* thiacyclols.[44] Aromatic *o*-thioacyl lactams such as (28) appear to be unstable and react spontaneously to form thiacyclols which rearrange readily to the thiodepsipeptides (29).



(28) n = 4 or 5

(29)

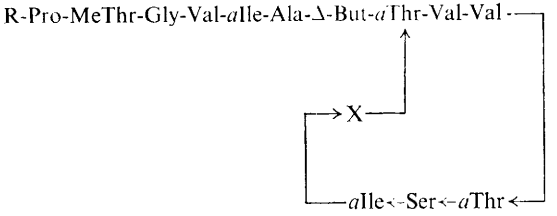


(30)

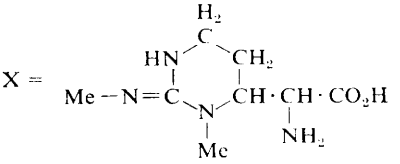
A synthesis of a depsipeptide analogue (30) of glutathione has been reported.[45] The ester link was introduced into the molecule during the first step of the synthesis, which involved condensing an appropriate bromoacetic acid ester with dibenzyloxycarbonylcystine. The molecule was then built up stepwise followed by cleavage of the disulphide linkage to give (30).

B. Stendomycin.—The tentative structure (31) (R = isomyristic or 11-methyllauric acid) has been proposed[46] for the main peptide antibiotics in the stendomycin mixture. The peptide lactone structure was based on the results of mild hydrolysis and ammonolysis, which gave the open chain stendomycin acid and amide respectively, and on the presence of an absorption at 1740 cm.$^{-1}$ in the i.r. spectrum. On chromic acid oxidation only one *allo*-threonine residue remained intact, confirming that the lactone link involves one *allo*-threonine hydroxyl group. The structure of (32) was based on n.m.r. data, which have not yet been published in detail. The *allo*-threonine residues, alanine, and all but one of the valine residues possess the D-configuration, all the remaining residues having the

[44] M. Rothe and R. Steinberger, *Angew. Chem. Internat. Edn.*, 1968, **7**, 884.
[45] L. A. Shchukina and A. L. Zhuze, *Zhur. obshchei Khim.*, 1967, **37**, 1980.
[46] M. Bodanszky, J. Izdebski, I. Muramatsu, and A. Bodansky, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 306; I. Muramatsu and M. Bodanszky, *J. Antibiot.* (*Tokyo*), 1968, **21**, 68.



(Δ-But = dehydrobutyrine) (31)



(32)

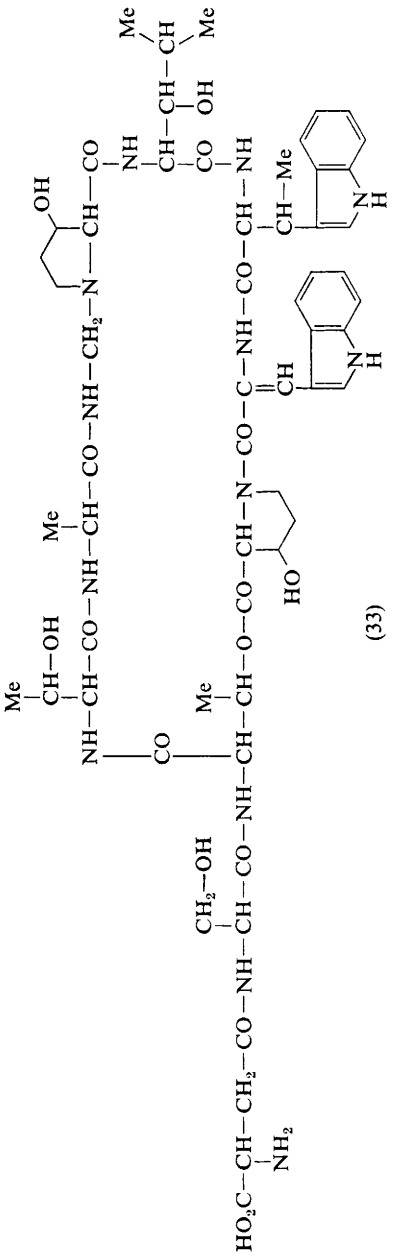
L-configuration. The u.v.-absorbing chromophore has been tentatively identified as the dehydrobutyrine residue.

Attempts to determine the amino-acid sequence from the mass spectrum of the permethylated derivative of stendomycin acid were complicated by the fact that permethylated threonine residues lose methanol readily in the spectrometer to give peaks at the same mass as proline.[47] It was, therefore, impossible to distinguish between *allo*-threonine, *N*-methylthreonine, dehydro-α-aminobutyric acid, and proline. The use of deuteriomethyl iodide for permethylation made it possible to divide the four residues into two groups. The mass of the proline residue and *N*-methyldehydro-α-aminobutyric acid (from *N*-methylthreonine) remained unchanged at 97 mass units, whereas the *N*-CD_3 group introduced into the *allo*-threonine and dehydro-α-aminobutyric acid residues gave peaks corresponding to loss of 100 mass units. Mass spectral cleavage occurred principally at the peptide—CO·N(Me)— bonds, with accompanying loss of methanol from each methylated threonine residue. Because the mass spectrum was determined on a mixture (containing fatty acid homologues), groups of peaks differing from each other by 14 mass units appeared in the spectrum. The mass spectral method confirmed the amino-acid sequence determined by classical methods, although sequence-determining peaks beyond the tenth amino-acid could not be detected above background.

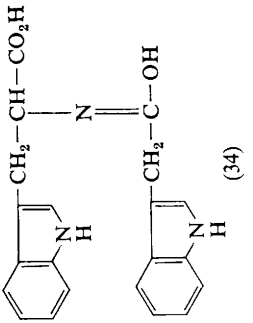
No definite interpretation[48] can yet be given for a trough appearing at 300 nm. in the o.r.d. spectrum of stendomycin, other than that it could be associated with the lactone ring. The trough disappeared on ring opening.

[47] D. W. Thomas, E. Lederer, M. Bodanszky, I. Izdebski, and I. Muramatsu, *Nature*, 1968, **220**, 580.

[48] M. Bodanszky and A. Bodanszky, *Nature*, 1968, **220**, 73.



(33)



(34)

C. Telomycin.—The complete structure of this peptide lactone (33) has been described.[49] Three new amino-acids, *erythro*-β-hydroxyleucine and *cis*- and *trans*-3-hydroxyproline, were identified in the hydrolysate mixtures, and these together with the other amino-acids are believed to have the L-configuration. The sequence was determined from the results of partial basic hydrolysis and a combination of the Edman, Sanger, and Akabori degradation methods. The lactone structure was based on the presence of an absorption at 1745 $cm.^{-1}$ in the i.r. spectrum, and the fact that telomycin underwent ready hydrolysis to telomycic acid. Comparison of the pK_a value and electrophoretic properties of telomycin with those for synthetic analogues confirmed the position of the free amino group. The presence of the β-methyltryptophan residue was proved by mass spectrometry. The source of the DL-tryptophan present in basic hydrolysates has been identified as dehydrotryptophan, which is believed to be the chromophore (λ_{max} 339 nm.) in telomycin. It has been suggested that the hydrolysis of dehydrotryptophan to give tryptophan proceeded via the indolylpyruvic acid, which then reacted with ammonia to give a ketone–ammonia adduct. Dehydration and decarboxylation to give (34), followed by hydrolysis, could then yield DL-tryptophan (50% max. yield) and indole-3-acetic acid. Both these products were identified in the hydrolysates.

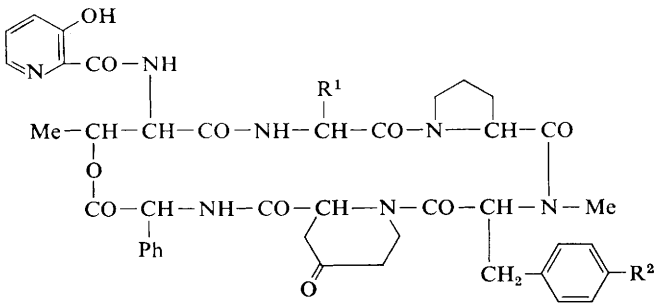
D. Patricin A.—Details of the complete synthesis of the antimicrobial agent, patricin A (35; R^1 = Et, R^2 = H, with proline instead of the 4-ketopipecolic acid residue) and its analogues, have been reported in a patent.[50]

E. Pristinamycins.—The two groups of antibiotics (classified in terms of their acidity), isolated from *Streptomyces pristinae spiralis* have been found to be identical with the synergistic antibiotics in the ostreogrycin series.[51] The detailed structures of the individual pristinamycins are shown in Table 1, together with a summary of the names used for the same compounds obtained from different sources.

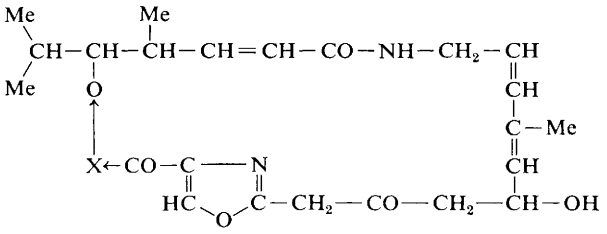
Table 1

	Pristinamycin	*Other names*
(35) R^1 = Et, R^2 = Me_2N	I_A	Ostreogrycin B, vernamycin Bα, mikamycin, PA 114 B_1
R^1 = Et, R^2 = MeNH	I_B	Ostreogrycin B_2, vernamycin Bβ
R^1 = Me, R^2 = Me_2N	I_C	Ostreogrycin B_1, vernamycin Bγ
(36) X = 2,3-dehydroproline	II_A	Ostreogrycin A, vernamycin A, virgimycin M_1, mikamycin A
X = proline	II_B	Ostreogrycin G, virgimycin

[49] J. C. Sheehan, D. Mania, S. Nakamura, J. A. Stock, and K. Maeda, *J. Amer. Chem. Soc.*, 1968, **90**, 462.
[50] M. Bodanszky, M. A. Ondetti, and J. T. Sheehan, U.S.P. 3,373,151/1968.
[51] J. Preud'homme, P. Tarridec, and A. Belloc, *Bull. Soc. chim. France*, 1968, 585.



(35)



(36)

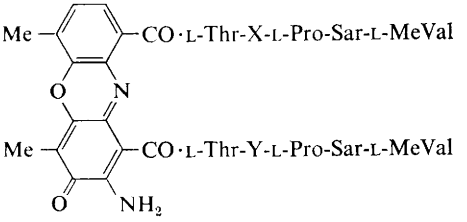
An analogue of pristinamycin I_A, in which the 4-ketopipecolic acid residue was replaced by L-4-hydroxyproline, has been synthesised.[52] A synthetic route involving ring closure at the depsipeptide link, as the final step, failed to give the desired product. However, prior incorporation of the ester linkage into a linear peptide, followed by cyclisation at the bond between threonine and α-aminobutyric acid using dicyclohexylcarbodiimide, gave (L-4-hydroxyproline [6])pristinamycin I_A. The presence of the ester linkage in the linear peptide necessitated the use of benzyl esters as carboxyl-protecting groups. *N*-Terminal protection of amino-acids was carried out using the *p*-methoxybenzyloxycarbonyl group which was selectively removed by trifluoroacetic acid.

F. Actinomycins.—Total syntheses of actinomycins C_1 (D) (38), C_2 (39), C_3 (40), and other actinomycin analogues possessing antibiotic activity have been reported.[53] Ring closure of acid (37) or its analogues, using an acetyl chloride–imidazole mixture, provided the last step in the synthetic sequence. The formyl group was used for *N*-protection in the preparation of the linear tetrapeptides, which were then coupled with *N*-(2-nitro-3-benzyloxy-4-methylbenzoyl)-L-threonine using the Woodward Reagent K.

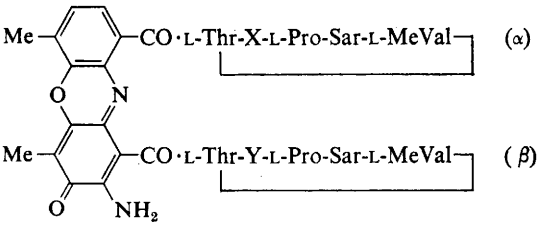
[52] G. Jolles and J. Bouchaudon, 'Peptides,' eds. H. C. Beyerman, A. van den Linde, and W. Maassen van den Brink, North-Holland Publishing Co., Amsterdam, 1967, 258.

[53] H. Brockmann and H. Lackner, *Chem. Ber.*, 1968, **101**, 1312; H. Brockmann and J. H. Manegold, *ibid.*, 1967, **100**, 3814.

Dimerisation of the linear protected peptides to produce the chromophore took place in the usual way by hydrogenation followed by oxidation with potassium ferricyanide. 4,6-Didemethyl-4,6-dimethoxyactinomycin was synthesised by the same route.[54]



(37)



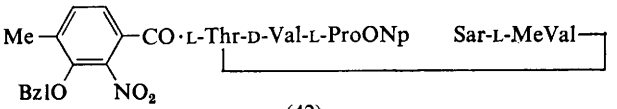
(38) X = Y = D-Val
(39) X = D-Val, Y = D-*a*Ile
(40) X = Y = D-*a*Ile
(41) X = D-*a*Ile, Y = D-Val

An alternative route for the synthesis of actinomycin C_1(D) (38) has been described.[55] Cyclisation to give the peptide lactone rings involved coupling between the proline and the sarcosine residues using the *p*-nitrophenyl ester (42) at high dilution in the presence of pyridine.

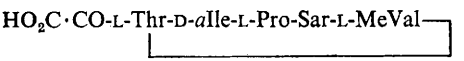
Further proof for the dipentapeptide lactone structure of the actinomycins has been obtained from the observation that the β-pentapeptide lactone ring from both actinomycins (39) and (40) was oxidised by hydrogen peroxide to (43).[56] The dipentapeptide lactone structure was further confirmed from the results of dimerisation experiments.[57] Oxidative dimerisation of a mixture of the lactones (44) and (45) gave the actinomycins (38–40), as well as the isomer (41).

In a study aimed at developing more active and less toxic actinomycins, synthesis of a number of actinocyl derivatives have been reported.[58] Esters

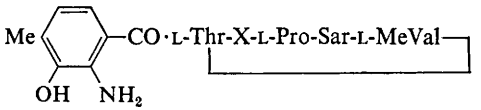
[54] H. Brockmann and F. Seela, *Tetrahedron Letters*, 1968, 161.
[55] J. Meienhofer, *Experientia*, 1968, **24**, 776.
[56] H. Brockmann and P. Boldt, *Chem. Ber.*, 1968, **101**, 1940.
[57] H. Brockmann and H. Lackner, *Chem. Ber.*, 1968, **101**, 2231.
[58] Y. Kameda, K. Matsui, H. Ooshiro, H. Suzuki, and H. Ozaki, *Chem. and Pharm. Bull. (Japan)*, 1968, **16**, 480.



(42)



(43)



(44) X = D-Val
(45) Y = D-*a*Ile

of various amino-acids (*e.g.* valine ethyl ester, serine methyl ester) were acylated with 3-benzyloxy-4-methyl-2-nitrobenzoyl chloride, and the product of hydrogenation was oxidised to the actinocyl derivatives.

4 Peptides and Amino-acids linked to Nucleosides and Nucleotides

A. Peptidyl-t-RNA.—Further work on the role of transfer RNA (t-RNA) in biosynthesis has necessitated the synthesis of peptidyl-t-RNA. General procedures have been reported [59] for the synthesis of peptidyl-t-RNA from amino-acyl-t-RNA. Conventional *N*-protecting groups could not be used for the synthesis because the ester link between the t-RNA and the peptide is labile to alkali and the t-RNA itself is sensitive to acid. Catalytic reduction would tend to reduce the pyrimidine bases. These restrictions, however, were largely overcome by using the mono-*p*-methoxy-triphenylmethyl group for *N*-protection. This group was removed under mild acidic conditions using 5% trichloroacetic acid at 4° for 5 min. ^{14}C-Valyl-, ^{14}C-phenylalanyl-, and ^{14}C-methionyl-t-RNA have been converted quantitatively into the corresponding dipeptidyl-t-RNA by reaction with *N*-monomethoxytrityl derivatives of alanine and glycine; *N*-hydroxysuccinimide was found to be a satisfactory coupling agent for this condensation. Gly-Gly-^{14}C-Phe-t-RNA and triglycyl-^{14}C-Phe-t-RNA have also been synthesised by the same procedure.[60]

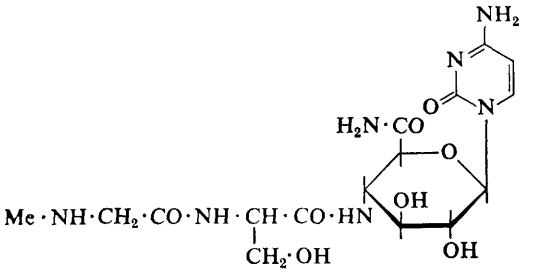
B. Adenylylated Peptide from Glutamine Synthetase.—A decapeptide having the amino-acid composition, $Asp_2Glu_2Pro_3Gly$,Leu,Tyr, and containing one

[59] Y. Lapidot, N. de Groot, M. Weiss, R. Peled, and Y. Wolman, *Biochim. Biophys. Acta*, 1967, **138**, 241; Y. Lapidot, N. de Groot, S. Rappoport, and A. D. Hamburger, *ibid.*, 1967, **149**, 532.

[60] Y. Lapidot, N. de Groot, S. Rappoport, and A. Panet, *Biochim. Biophys. Acta*, 1968, **157**, 433.

equivalent of adenylyl groups linked through a phosphodiester group, has been isolated[61] from adenylylated glutamine synthetase obtained from *Escherichia coli*. Tyrosine was confirmed as the only residue that could yield a phosphodiester derivative having the stability characteristics of the adenylylated peptide. Evidence that the adenylyl group was bound *via* a phosphodiester linkage to the phenolic hydroxyl group of tyrosine was obtained by comparing the u.v.-absorption data of adenylylated and un-adenylylated peptide at alkaline pH.

C. Gougerotin.—Results of spin-decoupling n.m.r. experiments have confirmed the structure (46) for the nucleoside antibiotic gougerotin.[62]



(46)

D. Nucleotide Murein Precursors.—Further studies have been reported on the structural elucidation of uridinediphosphate muramyl peptides from several sources, which are known to be precursors of peptidoglycans in bacterial cell walls. The nucleotide UDP-*N*-acetylmuramyl-L-alanyl-D-glutamyl-*meso*-2,6-diaminopimelic acid,* has been isolated from strains of *Bacillus cereus*.[63] Enzymic digestion of this nucleotide with sweet potato pyrophosphatase liberated the *N*-acetylamino sugar derivative of the peptide. On further treatment with *N*-acetylmuramyl-L-alanine amidase, the free peptide was obtained. Alanine was proved to be *N*-terminal from the results of dinitrophenylation experiments, and on hydrazinolysis of the sugar derivative of the peptide, alanine hydrazide, glutamyl-γ-monohydrazide, and *meso*-2,6-diaminopimelic acid were obtained. Thus, the *C*-terminal linkage must involve the γ-carboxyl group of glutamic acid.

An UDP-activated murein precursor has been isolated[64] from *Butyribacterium rettgeri* after inhibition by D-cycloserine for 1 hr. Partial hydrolysis and end-group analysis showed that the nucleotide contained muramic acid attached to the peptide L-Ser-D-Glu-L-Orn. Identification of γ-glutamylhydrazide as a hydrazinolysis product confirmed that ornithine

[61] B. M. Shapiro and E. R. Stadtman, *J. Biol. Chem.*, 1968, **243**, 3769.
[62] J. J. Fox, Y. Kuwada, and K. A. Watanabe, *Tetrahedron Letters*, 1968, 6029.
[63] T. Nakatni, Y. Araki, and E. Ito, *Biochim. Biophys. Acta.*, 1968, **156**, 210.
[64] I. Miller, R. Plapp, and O. Kandler, *Z. Naturforsch.*, 1968, **23b**, 217.

* UDP = uridinediphosphate, *N*-acetylmuramic acid = 2-acetamido-2-deoxy-3-*O*-(D-1-carboxyethyl)-D-glucose.

was linked, *via* its α-amino group, to the γ-carboxyl group of glutamic acid. The amino-acid sequence of this murein precursor agreed with the structure of the corresponding part of the complete murein.

Two murein precursors, accumulated by D-cycloserine inhibition of *Lactobacillus coryniformis* and *L. cellobiosus*, have been shown [65] to possess the sequences,* Mur-L-Ala-D-Gln-L-Lys-(D-Asn)-D-Ala, and Mur-L-Ala-D-Gln-L-Orn-(D-Asn)-D-Ala respectively. It seems likely that the aspartic acid residues in these peptides are involved in the cross-linking of the mureins by forming a peptide bond with the *C*-terminal D-alanine on an adjacent mucopeptide. A mucopeptide nucleotide precursor, accumulated in cultures of *Corynebacterium insidiosum* and *C. sepedonicum* after inhibition with vancomycin, has been shown [66] to contain the sequence muramyl-Gly-Glu-Dbu-Ala-Ala. The results of dinitrophenylation experiments carried out on the nucleotide indicated that both amino groups of diaminobutyric acid were blocked.

E. Polyoxins.—The structures of the main components, polyoxin A [47; X = (48)] and B (47; X = OH), of these agriculturally useful antifungal antibiotics have been elucidated.[67] The polyoxin antibiotic mixture produced by *Streptomyces cacaoi* var. *asoensis* contained nine components, some of which showed selective activity against phytopathogenic fungi.

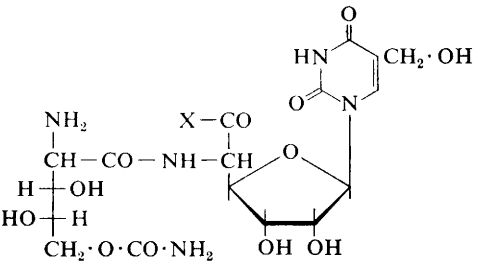
The constituent amino-acids found in the hydrolysates of polyoxin A were proved unambiguously to be 3-ethylidene-L-azetidine-2-carboxylic acid (48) and carbamyl polyoxamic acid (49) together with the amino-acid bearing the nucleoside moiety. Acid (49) consumed three equivalents of periodate and was readily hydrolysed to (50). The structure of (50) was confirmed by conversion to the *N*-acetyl-γ-lactone which underwent reduction with sodium amalgam yielding the known 2-acetamido-2-deoxy-β-L-xylose. The sequence of the three unusual L-α-amino acids was established by identification of the components of an alkali hydrolysate obtained after deamination of the parent molecule with nitrous acid.

5 Peptide–Carbohydrate Linkages

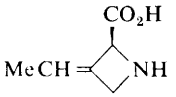
A review of the literature, with special emphasis on the carbohydrate aspects of the subject will be found in the Specialist Periodical Reports on Carbohydrate Chemistry.[68] Recent advances in the structure and biosynthesis of bacterial cell wall peptidoglycan have also been summarised.[69]

[65] R. Plapp and O. Kandler, *Z. Naturforsch.*, 1967, **22b**, 1062.
[66] H. R. Perkins, *Biochem. J.*, 1968, **110**, 47P.
[67] K. Isono and S. Suzuki, *Tetrahedron Letters*, 1968, 1133.
[68] 'Carbohydrate Chemistry,' R. D. Guthrie, The Chemical Society, London, 1968, vol. 1, p. 240.
[69] J. M. Ghuysen, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., 1968, p. 283.

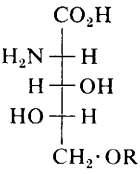
* The following abbreviations have been used for the carbohydrates: G = glucose, Gal = galactose, Mur = muramic acid, GA = glucuronic acid, GNAc = *N*-acetylglucosamine, Xyl = xylose, GalNAc = *N*-acetylgalactosamine.



(47)



(48)



(49) R = CO·NH$_2$
(50) R = H

A. Synthesis and General Aspects.*—Carbohydrates linked to the peptide moiety as *O*-glycosides bonded to the hydroxyl groups of serine or threonine are extremely widespread in glycoproteins. This type of linkage is characterised by its lability towards alkali, when it breaks down *via* a β-elimination process to give monosaccharides and the corresponding α-amino unsaturated acid. A detailed study of this alkaline cleavage has been reported.[70] This lability of sugar-peptide bonds to alkali has also been used[71] for the quantitative evaluation of the *O*-glycosidic-amino acid linkages in ovine submaxillary gland mucoprotein. Stoicheiometric conversion of serine to dehydro-alanine and of threonine to α-aminocrotonic acid was demonstrated in this case.

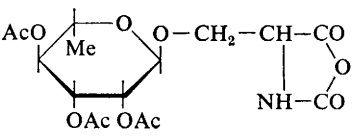
The importance of the *O*-glycosidic-serine and -threonine linkages has also stimulated research into the synthesis of a number of serine *O*-glycosides. A generally applicable method has been reported, which involves the Koenigs–Knorr glycosylation of *N*-benzyloxycarbonyl-L-serine methyl ester, followed by removal of the *N*-protecting group by hydrogenolysis.[72]

[70] V. A. Derevitskaya, M. G. Vafina, and N. K. Kochetkov, *Carbohydrate Res.*, 1967, **3**, 377.
[71] S. Harbon, G. Herman, and H. Clauser, *European J. Biochem.*, 1968, **4**, 265–272.
[72] V. A. Derevitskaya and M. G. Vafina, *Izvest. Akad. Nauk, S.S.S.R., Ser. khim.*, 1967, 1528.

* The following abbreviations have been used for the carbohydrates: G = glucose, Gal = galactose, Mur = muramic acid, GA = glucuronic acid, GNAc = *N*-acetylglucosamine, Xyl = xylose, GalNAc = *N*-acetylgalactosamine

L-Serine-*N*-methylamide-*O*-β-D-glucopyranoside, prepared from *N*-benzyloxycarbonyl-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranoside)-L-serine methyl ester using this method, was reacted further with benzyloxycarbonylglycine in the presence of dicyclohexylcarbodi-imide to give *N*-(*N*-benzyloxycarbonylglycyl)-L-serine-*N*-methylamide-*O*-β-D-glucopyranoside.

A method for linking sugars to synthetic polypeptides has been developed.[73] The preparation of the *N*-carboxy-anhydride (51) provided a suitable intermediate for attachment of further amino-acid residues. The anhydride (51) was synthesised via a Koenigs–Knorr condensation of the corresponding acetobromo sugars with *N*-benzyloxycarbonyl-L-serine benzyl ester, followed by hydrogenolysis to give the *O*-acetylated serine glycoside. Treatment of this glycoside with phosgene yielded (51).



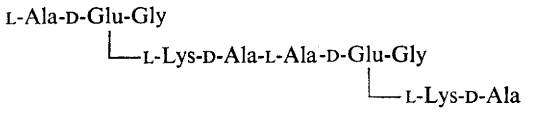
(51)

B. Glycopeptides from Cell Walls.—Enzymatic lysis continues to play a very important role in the structural elucidation of bacterial cell walls. Lysozyme, the first known bacteriolytic enzyme, is now only one member of a large family of enzymes which act on cell walls.[74] The L_{11} enzyme, another member of this family, has been shown[75] to hydrolyse the pentaglycine bridges in cell walls of *Staphylococcus aureus*. Lysis of the cell walls of *Corynebacterium diphtheriae*, using the L_3 enzyme preparation from *Streptomyces*, brought about cleavage of D-alanyl-*meso*-diaminopimelic acid linkages.[76] Use of the L_3 preparation, in conjunction with the peptidase from *Myxobacterium*, showed that the cell wall of *C. diphtheriae* contained both the tetrapeptide L-Ala-D-Glu-*meso*-Dpm-D-Ala and the tripeptide, L-Ala-D-Glu-*meso*-Dpm, subunits. The proportion of D- and L-alanine in the hydrolysate was determined enzymatically, the D-alanine being present as the *C*-terminal residue (from hydrazinolysis experiments) and the L-alanine as *N*-terminal (DNP method) (Dpm = α,α′-diaminopimelic acid).

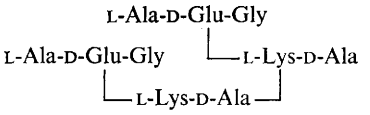
Six different types of degradation sequences carried out[77] on cell walls of *Micrococcus lysodeikticus* have given rise to a number of products including two new peptide dimers (52) and (53). In three of the degradation sequences used, N^{ϵ}-(D-alanyl)-L-lysine linkages were hydrolysed by *Streptomyces* ML endopeptidase, and the glycan obtained was degraded with either lysozyme

[73] E. Rüde and M. Meyer-Delius, *Carbohydrate Res.*, 1968, **8**, 219.
[74] J. L. Strominger and J. M. Ghuysen, *Science*, 1967, **156**, 213.
[75] K. Kato and J. L. Strominger, *Biochemistry*, 1968, **7**, 2754.
[76] K. Kato, J. L. Strominger, and S. Kotani, *Biochemistry*, 1968, **7**, 2762.
[77] J. M. Ghuysen, E. Bricas, M. Lache, and M. Leyhbouille, *Biochemistry*, 1968, **7**, 1450.

or F_1 endo-*N*-acetylmuramidase. The peptide units were isolated, either in the form of *N*-acetylmuramyl peptides after treatment with *Streptomyces* exo-*N*-acetylhexosaminidase, or in the form of the free peptides after treatment with *N*-acetylmuramyl-L-alanine amidase. Two peptide monomers isolated in this way were found to be identical with peptides previously isolated. The isolation of the dimers (52) and (53) cannot be explained by the structure of the cell wall previously suggested.[78]



(52)



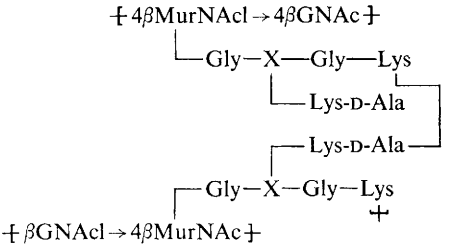
(53)

Examples of *O*-glycosidic amino-acid-sugar linkages and *N*-glycosidic amino-acid-sugar linkages occurring in the same glycoprotein have only been found very recently.[79] Good evidence to support the existence of both types of peptide-carbohydrate linkages has been obtained for a glycopeptide isolated from yeast cell wall.[80] The glycopeptide (*M* 76,000), containing approximately 4% amino-acid residues, after treatment with sodium hydroxide lost 64% of its threonine and 45% of its serine. This result would be expected for a β-elimination of a glycosidic serine or threonine bond. Difficulty in separating high molecular weight polysaccharide from the peptide material after alkaline hydrolysis suggested another possible link. Investigations on glycopeptide material after hot alkaline treatment of yeast cell walls showed that the amounts of glucosamine and amino-acids were equimolecular. Furthermore, aspartic acid was the only amino-acid present in equimolecular amounts to the glucosamine. This, therefore, suggests that the sugar-peptide linkage is *via* [*N*-(β-aspartyl)-β-D-(*N*-acetyl)glucosaminide].

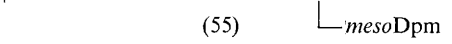
Cell walls of *Microbacterium lacticum*, on digestion with trypsin and extraction with trichloroacetic acid, gave a murein containing glutamic acid, lysine, glycine, and D-alanine in the ratio 1 : 2 : 2 : 1, in addition to the usual amino-sugars.[81] Dinitrophenylation experiments showed that 66%

[78] W. Katz and J. L. Strominger, *Biochemistry*, 1967, **6**, 930.
[79] G. Dawson, and J. R. Clamp, *Biochem. Biophys. Res. Comm.*, 1967, **26**, 349; R. G. Spiro, *J. Biol. Chem.*, 1967, **242**, 1926.
[80] R. Sentandreu and D. H. Northcote, *Biochem. J.*, 1968, **109**, 419.
[81] K. H. Schleifer, R. Plapp, and O. Kandler, *Biochim. Biophys. Acta*, 1968, **154**, 573.

of the total lysine yielded ε-DNP-lysine, while hydrazinolysis identified at least 40% lysine, 6% D-alanine, and 2·6% glycine as *C*-terminal. A tetrapeptide bound to muramic acid was shown to be Gly-Glu-Lys-D-Ala. In contrast to the other mureins the ε-amino group of lysine within the tetrapeptide appeared to be free and supported the structure (54) for the region around the peptide-sugar link.



(54) X = Glu or *threo*-3-hydroxyglutamic acid

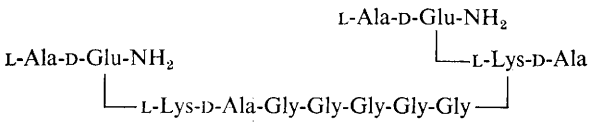


(55)

The disaccharide tripeptide (55) and its amide, containing *N*-glycoside amino-acid linkages, have been isolated[82] from the lysozyme digest of the cell walls of *Bacillus lichenformis*. The configurations of the amino-acids were assigned using enzymatic and chromatographic methods. End-group analysis was used to determine the amino-acid sequence and the isolation of γ-glutamylhydrazide as a product of hydrazinolysis confirmed the presence of an isoglutamine linkage in the molecule, which is of general occurrence in bacterial cell walls.

Partial hydrolysis of the cell walls of a strain of *Staphylococcus epidermidis* from raw milk has given rise to a tetrapeptide L-Ala-D-Gln-L-Lys-D-Ala linked to muramic acid.[83] A new mucopeptide containing diaminopimelic acid connected to glycine and D-alanine residues has been isolated from the cell wall of *Streptomyces roseochromogenes*.[84]

A synthesis of the tridecapeptide (56), the repeating peptide unit in the peptidoglycan from *Staphylococcus aureus*, has been outlined.[85] In an



(56)

[82] D. Mirelman and N. Sharon, *J. Biol. Chem.*, 1968, **243**, 2279.
[83] K. H. Schleifer, M. Reid, and O. Kandler, *Arch. Mikrobiol.*, 1968, **62**, 198.
[84] K. Arima, T. Nakamura, and G. Tamura, *Agric. and Biol. Chem. (Japan)*, 1968, **32**, 530.

attempt to define the exact structure of the tripeptide L-Ala-D-Glu-L-*meso*-diaminopimelic acid, obtained on enzymic degradation of the cell walls of certain bacteria, the synthesis of both the α- and the γ-isomers of the tripeptide has been reported.[86] The two isomers showed different migration properties when subjected to electrophoresis (3000 v for 24 hr.) at pH 4.

C. Glycopeptides from Miscellaneous Sources.—Low molecular weight glycopeptides have been identified[87] in the testicular hyaluronidase digests of pig skin dermatan sulphate. The glycopeptides accounted for 40% of the total polysaccharide-bound peptide. A general structure (57) was

$$(\text{GA-GalNAc})_n\text{-GA-Gal-Gal-Xyl-Ser (peptide)}$$

$$n = 1 \text{ or } 2 \qquad (57)$$

suggested for the carbohydrate–peptide link region. This was based on the fact that alkali treatment resulted in 90% destruction of serine residues due to β-elimination. Further confirmation for the *O*-glycosidic serine link came from the observation that serine was the only amino-acid present in equimolecular amounts to xylose.

The presence of an *N*-glycosidic-amino-acid link, in the pancreatic glycoprotein ribonuclease B, has been confirmed using a procedure designed to detect less than 1 nanomole quantities of 1-L-β-aspartamido-(2-acetamido)-1,2-dideoxy-β-D-glucose.[88] The technique involved the preparation of the dansyl derivative of the glycoprotein, followed by hydrolysis and the detection of the highly fluorescent dansyl-1-L-β-aspartamido-(2-acetamido)-1,2-dideoxy-β-glucose. A similar type of sugar–peptide bond has been strongly favoured for the glycopeptide isolated from the immunoglobulin secreted by a plasma cell tumour in mice.[89] The glycopeptide was obtained from the protein by treatment with mercuripapain, followed by ethanol fractionation and chromatography. It was shown to contain aspartic acid, serine, glutamic acid, glycine, and isoleucine, with a molar ratio of hexose to aspartic acid of 4·2 : 1.

Tryptic digestion of 'desialised' *S*-carboxymethyl human thyroglobulin has been reported to give a glycopeptide, which represented a major portion of the mannose-glucosamine units.[90] The structure (58) was proposed for the glycopeptide, based on the results of carboxypeptidase digestion and the Edman method of sequence determination. No phenylthiohydantoin derivative was detected, after the fourth step, in the Edman technique, thus

[85] P. Lefrancier and E. Bricas, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 293.
[86] P. Dezelee and E. Bricas, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 299.
[87] L.-Å. Fransson, *Biochim. Biophys. Acta*, 1968, **156**, 311.
[88] T. H. Plummer jun., A. Tarentino, and F. Maley, *J. Biol. Chem.*, 1968, **243**, 5158.
[89] T. J. Coleman, R. D. Marshall, and M. Potter, *Biochim. Biophys. Acta*, 1967, **147**, 396.
[90] A. B. Rawitch, T-H. Liao, and J. G. Pierce, *Biochim. Biophys. Acta*, 1968, **160**, 360.

Ala-Leu-Glu-Asp-Ala-Thr-Arg
|
(carbohydrate)

(58)

confirming the presence of a glycopeptide linkage at the aspartic acid residue. Aspartic acid was also found to be the only amino-acid present in stoicheiometric quantities in glycopeptides obtained from pronase digests of human and calf thyroglobulin. A preliminary report on the preparation and properties of glycopeptides released on enzymic digestion of α_2-human macroglobulin has appeared.[91]

The glycopeptide isolated from pronase and trypsin proteolysates of a phytohaemagglutinin from *Robinia pseudoaccacia* possessed essentially all the activity of the parent phytohaemagglutinin.[92] The structure of the peptide has not yet been elucidated but preliminary results showed a predominance of aspartic and glutamic acids and serine. Similarly, a glycopeptide from a phytohaemogglutinin of *Phaseolus vulgaris* has been shown to be composed of twelve amino-acid residues: aspartic acid, threonine, glutamic acid, and tyrosine in the ratio 6 : 3 : 2 : 1, together with nineteen sugar residues.[93] The precise nature of the linkages has not yet been determined.

Two further reports of the isolation of glycopeptides from fibrinogen have appeared. Of the three principal glycopeptides isolated [94] from partial hydrolysates, one contained a tripeptide fragment Asp-Lys-Thr, while the other two were shown to have the same pentapeptide sequence, Val-Gly-Glu-Asp-Arg. Four other glycopeptides were also proved to be present but have not been characterised. Four major glycopeptide fractions have been obtained from the pronase digests of both bovine fibrinogen and fibrin.[95] Two of these fractions have been found to contain lysine, while the other two contained arginine, glutamic acid, glycine, and valine.

Soluble and insoluble collagen from guinea pig skin gave similar amounts of *O*-Hyl(-Gal-G) and *O*-Hyl(-Gal) on alkaline hydrolysis.[96] Both types of collagen also gave rise to the glycohexapeptide Gly-Met-Hyl(-Gal-G)-Gly-His-Arg. However, a glycopeptide fraction of a higher molecular weight, but containing no hydroxylysine, (Hyl), was found only in insoluble collagen.

Studies on the biosynthesis of plasma glycoprotein have led to the isolation of two glycopeptides from the deoxycholate-soluble fraction of rat

[91] R. Bourrillon, E. Razafimahaleo, and M. A. Parnaudeau, *Biochim. Biophys. Acta*, 1968, **154**, 405.
[92] R. Bourrillon and J. Font, *Biochim. Biophys. Acta*, 1968, **154**, 28.
[93] T. Takahashi and I. E. Liener, *Biochim. Biophys. Acta*, 1968, **154**, 560.
[94] L. Mester, E. Moczár, and L. Szabados, *Compt. rend.*, 1967, **265**, *C*, 877.
[95] B. A. Bray and K. Laki, *Biochemistry*, 1968, **7**, 3119.
[96] L. W. Cunningham and J. D. Ford, *J. Biol. Chem.*, 1968, **243**, 2390.

liver microsomes.[97] One of the peptides contained only *N*-acetylglucosamine, mannose, and aspartic acid, while the other glycopeptide contained sialic acid, galactose, *N*-acetylglucosamine, mannose, aspartic acid, serine, glutamic acid, proline, and glycine in the ratio 1 : 1 : 4 : 5 : 3 : 1 : 2 : 1 : 1. In the latter peptide, aspartic acid was shown to be the only *N*-terminal amino-acid. The peptide–carbohydrate link does not appear to involve serine, since no change was observed in the serine content on treatment with alkali.

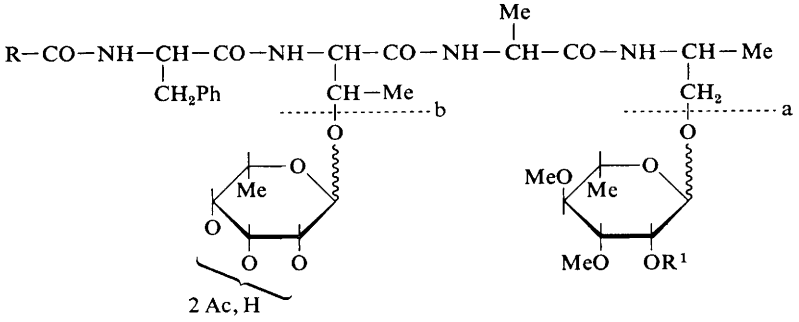
Enzymatic digestion of stem bromelain with pronase P gave,[98] after chromatographic and electrophoretic separation, a glycopeptide containing aspartic acid, serine, glutamic acid, glucosamine, mannose, xylose, and fucose, in the mole ratios 3 : 3 : 2 : 4 : 3 : 1 : 1. The results of dinitrophenylation experiments showed that the *N*-terminal residue was aspartic acid.

The presence of *O*-glycoside-threonine links in the glycopeptides obtained from χ-casein of cow has been reported.[99] The result was based on the determination of the α,β-unsaturated amino-acid content of alkali hydrolysates, and the destruction of threonine on treatment with alkaline lithium borohydride. Similarly, *O*-glycoside-serine and -threonine linkages have been found in gluco-amylase of *Aspergillus niger*.[100]

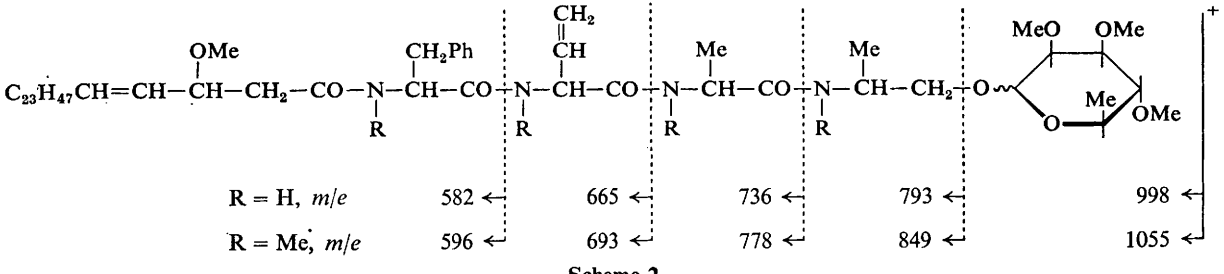
D. Mycosides.—Mass spectrometry has proved to be an extremely useful tool for determining the structure of a number of closely related mycosides (peptidoglycolipids). The structure of mycoside C (59) was elucidated from mass spectral data, and chemical data accumulated from partial acid and base hydrolysis, and dinitrophenylation experiments.[101] This mycoside isolated from *Mycobacterium scrofulaceum* was a mixture of components which differed from each other only in the nature of their fatty acid residues. The structure of mycoside C_{b1} (60) isolated from *Mycobacterium butyricum* has also been determined using mass spectrometry.[102] The interpretation of the mass spectral data was facilitated by comparison of the breakdown pattern of the parent mycoside with that of its *N*-methylated derivative. *N*-Methylation was carried out by the Hakamori method using a solution of methyl sulphinyl carbanion (from sodium hydride in dimethyl sulphoxide) and excess methyl iodide in dimethyl sulphoxide. Characteristic peaks in the mass spectrum appeared at *m/e* 1041 (1111 for the methylated compound) corresponding to cleavage at position a, and at *m/e* 998 (1055 when methylated) due to cleavage at b. Other ions obtained after loss of the 6-deoxytalose residue and hydrogen transfer are given in Scheme 2.

The mycoside C_{1217} (61) has been shown to be the main component in a mixture of glyco-peptidolipids isolated[103] from a strain of *Mycobacteria*.

[97] Y. T. Li, S. C. Li, and M. R. Shetlar, *J. Biol. Chem.*, 1968, **243**, 656.
[98] T. Murachi, A. Suzuki, and N. Takahashi, *Biochemistry*, 1967, **6**, 3730.
[99] A.-M. Fiat, C. Alais, and P. Jollès, *Chimia* (*Switz.*), 1968, **22**, 137.
[100] D. R. Lineback, *Carbohydrate Res.*, 1968, **7**, 106.
[101] E. Vilkas, C. Gros, and J. C. Massot, *Compt. rend.*, 1968, **266**, *C*, 837.
[102] E. Vilkas and E. Lederer, *Tetrahedron Letters*, 1968, 3089.
[103] G. Laneele and J. Asselineau, *European J. Biochem.*, 1968, **5**, 487.



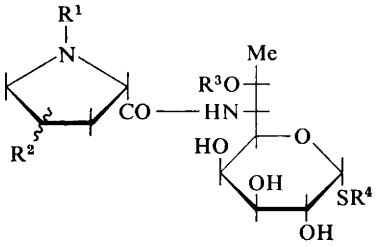
(59) R = $Me(C_nH_{2n-2})CH(OMe)—$, with n = 28–31, R^1 = H
(60) R = $Me(C_{24}H_{46})CH(OMe)CH_2—$, R^1 = Me
(61) R = $C_{25}H_{51}CH(OH)CH_2—$, R^1 = Me



Scheme 2

The amino-acids and the amino-alcohol residue were identified by gas chromatography of their *N*-trifluoroacetyl derivatives. Alkali degradation confirmed that the deoxytalose residue was attached to the hydroxyl group of the *allo*-threonine. The amino-acid sequence was determined using the dansyl and Edman techniques.

E. Amino-acid–Carbohydrate Antibiotics.—The preparation of several analogues of lincomycin (62) has been reported. 4′-Depropyl-4′-ethoxylincomycin (*cis* and *trans*) (63) have been synthesised by coupling the appropriate amino-acid to the sugar moiety using the mixed anhydride method.[104] The products possessed only 2% of the activity of lincomycin. A synthesis of *N*-ethyl-lincomycin D (64) has been patented.[105]



(62) R^1 = Me, R^2 = Pr, R^3 = H, R^4 = Me
(63) R^1 = Me, R^2 = OEt, R^3 = H, R^4 = Me
(64) R^1 = Et, R^2 = Pr, R^3 = H, R^4 = Me
(65) R^1 = Me, R^2 = H, R^3 = Me, R^4 = $CH_2 \cdot O \cdot CO$

(66) R^1 = Me, R^2 = H, R^3 = Me, R^4 = $CH_2 \cdot OH$
(67) R^1 = Me, R^2 = Pr, R^3 = Me, R^4 = $CH_2 \cdot O \cdot CO$

The antibiotic celesticetin (65) and an analogue desalicetin (66) have been found to be structurally related to lincomycin.[106] Hydrazine hydrate cleaved the ester and the amide bond in celesticetin, and the amino sugar then released was converted to a product that was also obtained from lincomycin under similar conditions. The stereochemistry of the molecule was determined using n.m.r. The analogue (67) showed antibacterial activity midway between the lincomycins and celesticetin.

6 Peptide Alkaloids

A number of peptide alkaloids isolated from various sources have now been shown to contain a *p*-alkoxystyrylamino residue as part of a 14-membered ring. Good progress has recently been made in the structural

[104] B. J. Magerlein, *J. Medicin. Chem.*, 1967, **10**, 1161.
[105] A. D. Arquodelis and F. Kagan, U.S.P. 3,361,739/1968.
[106] H. Hoeksema, *J. Amer. Chem. Soc.*, 1968, **90**, 755.

elucidation of these peptide alkaloids, due to the application of mass spectrometry and other physical methods. A review of the structural elucidation of peptide alkaloids using mass spectrometry has appeared.[107] Two research groups [108] have suggested a revised structure for ceanothine B (see Table 2). The new structure contains the *p*-alkoxystyrylamino-residue instead of the *ortho*-fused ring system proposed previously. The structures of a number of peptide-alkaloids recently isolated are summarised in Table 2.

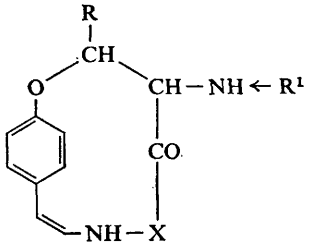


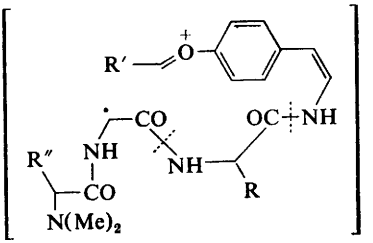
Table 2

Name	*Source*	*Structure*
Ceanothine B[108]	*Ceanothus americanus*	R = —CHMe$_2$, R^1 = MePro, X = Phe
Americine[a]	*Ceanothus americanus*	R = —CHMe$_2$, R^1 = MeVal, X = Trp
Scutianine[b]	*Scutia buxifolia* Reiss	R = —CHMe$_2$, R^1 = Me$_2$Phe-Pro, X = Phe
Integerressine[c]	*Ceanothus integerrimus*	R = Ph, R^1 = Me$_2$Val, X = Phe
Integerrenine[c]	*Ceanothus integerrimus*	R = Ph, R^1 = Me$_2$Ile, X = Leu
Integerrine[d]	*Ceanothus integerrimus*	R = Ph, R^1 = Me$_2$Val, X = Trp
Franganine[e]	*Rhamnus frangula* and *Melochia corchorifolia*[f]	R = —CHMe$_2$, R^1 = Me$_2$Leu, X = Leu
Frangufoline[e]	*Rhamnus frangula* and *Melochia corchorifolia*[f]	R = —CHMe$_2$, R^1 = Me$_2$Phe, X = Leu
Frangulanine[g]	*Rhamnus frangula*	R —CHMe$_2$, R^1 = Me$_2$Ile, X = Leu
Adouetine Y′[f]	*Melochia corchorifolia*	R = —CHMe$_2$, R^1 = Me$_2$Phe, X = Ile
Adouetine X[h]	*Waltheria americana*	R —CHMe$_2$, R^1 = Me$_2$Leu, X = Ile
Adouetine Y[h]	*Waltheria americana*	R = Ph, R^1 = Me$_2$Phe, X = Ile
Adouetine Z[h]	*Waltheria americana*	R = Phe, R^1 = Me$_2$Phe-Pro, X = Phe

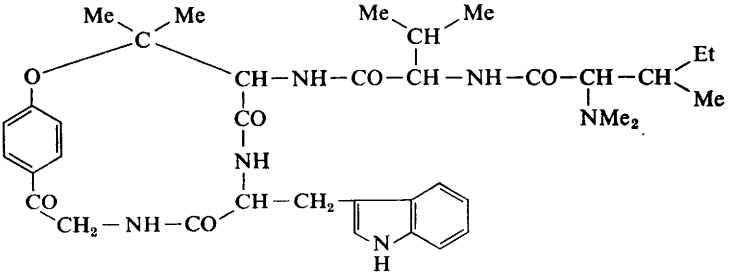
[a] F. Klein and H. Rapoport, *J. Amer. Chem. Soc.*, 1968, **90**, 2398.
[b] R. Tschesche, R. Welters, and H.-W. Fehlhaber, *Chem. Ber.*, 1967, **100**, 323.
[c] R. Tschesche, J. Rheingans, H.-W. Fehlhaber, and G. Legler, *Chem. Ber.*, 1967, **100**, 3924.
[d] R. Tschesche, E. Frohberg, and H.-W. Fehlhaber, *Tetrahedron Letters*, 1968, 1311.
[e] R. Tschesche and H. Last, *Tetrahedron Letters*, 1968, 2993.
[f] R. Tschesche and I. Reutel, *Tetrahedron Letters*, 1968, 3817.
[g] R. Tschesche, H. Last, and H.-W. Fehlhaber, *Chem. Ber.*, 1967, **100**, 3937.
[h] M. Pais, J. Marchand, F.-X. Jarreau, and R. Goutarel, *Bull. Soc. chim. France*, 1968, 1145.

[107] H.-W. Fehlhaber, *Z. analyt. chem.*, 1968, **235**, 91.
[108] R. E. Servis and A. I. Kosak, *J. Amer. Chem. Soc.*, 1968, **90**, 4179; F. K. Klein and H. Rapoport, *ibid.*, 1968, **90**, 3576.

The fragmentation pattern in the mass spectra of the alkaloids listed in Table 2 was substantially independent of the nature of the amino-acid residues.[109] Significant fragmentation peaks appeared due to cleavage of the carbon–carbon bonds adjacent to the *N*-dimethyl group and due to the breakdown of the 14-membered rings to give (68) (after H shift). Fragment (68) appeared to undergo further cleavage at the amide bonds. Other important fragmentations were due to typical cleavages undergone by phenol ethers, and due to stepwise degradation of the side-chain.



(68)

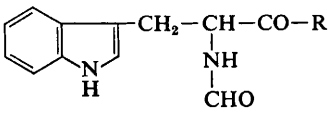


(69)

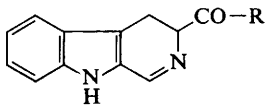
A peptide-alkaloid isolated from *Hymencardia acida* has been shown [109a] to possess the structure (69), from the results of acid and base hydrolysis and mass spectral data. In this structure a $-CH_2 \cdot CO-$ group replaces the styryl double bond found in all the examples listed in Table 2.

Peptides containing the β-carboline residue, for example (71), have been synthesised by treating N^{α}-formyltryptophyl-peptide esters (70) with trifluoroacetic acid,[110] or with a polyphosphate.[111] In contrast compounds such as (72) have been reported to undergo cyclodehydration to yield (73) under similar conditions.

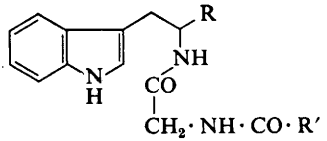
[109] H.-W. Fehlhaber, *Angew. Chem. Internat. Edn.*, 1967, **6**, 968.
[109a] M. Pais, J. Marchand, G. Rattle, and F.-X. Jarreau, *Bull, Soc. chim. France*, 1968, 2979.
[110] A. Previero, M-A. Coletti-Previero, and L. G. Barry, *Canad. J. Chem.*, 1968, **46**, 3404.
[111] Y. Kanaoka, E. Sato, and O. Yonemitsu, *Tetrahedron*, 1968, **24**, 2591.



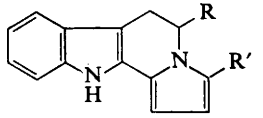
(70)



(71)



(72)



(73)

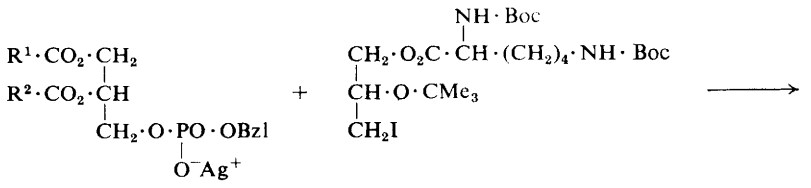
It has been reported that small peptides can be readily linked to the free amine group of emetine using conventional methods for peptide synthesis.[112] Thus, 2′-(L-prolyl-L-prolyl)-emetine was prepared by condensing Boc-L-Pro-L-Pro with emetine, using the water-soluble 1-ethyl-3-(3′dimethyl-aminopropyl) carbodi-imide hydrochloride as condensing agent.

7 Peptides and Amino-acids Conjugated to Lipids

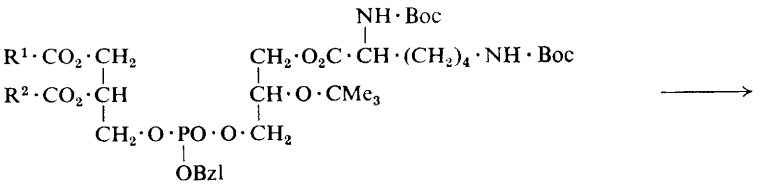
Studies on the structural elucidation of naturally occurring lipo-amino-acids and peptides continue to benefit from the development of methods for the preparation of synthetic analogues with well-defined configurations. The structure and configuration of a phosphatidyl-glycerol from *Staphylococcus aureus* has been confirmed[113] by the synthesis of 1-[(1′-oleoyl-2′-palmitoylglyceryl)phosphoryl]-3-L-lysylglycerol (75) using the synthetic route summarised in Scheme 3. The glycerol derivative (74) was synthesised from the di-*N*-protected lysyl residue and an iodoglycerol in the presence of triethylamine. The synthetic product (75) was shown to be identical with that of the natural product. The synthesis of a number of isomeric *O*-lysyl esters of phosphatidyl glycerol has also been reported[114] using a similar approach to that outlined in Scheme 3.

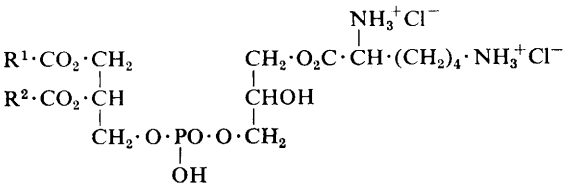
Both the *R* and *S* forms of 1-L-[(1′,2′-distearoylglyceryl)phosphoryl]-3-L-alanylglycerol have now been synthesised.[115] The lipo-amino-acid isolated from *Clostridium welchii* showed very similar properties to the *S*-epimer. The synthesis followed a similar approach to that outlined in Scheme 3.

[112] G. R. Pettit and S. K. Gupta, *Canad. J. Chem.*, 1967, **45**, 1600.
[113] P. P. M. Bonsen, G. H. DeHass, and L. L. M. Van Deenen, *Biochemistry*, 1967, **6**, 1114.
[114] Y. G. Molotkovskii and L. D. Bergel'son, *Chem. Phys. Lipids*, 1968, **2**, 1.
[115] Y. G. Molotkovskii and L. D. Bergel'son, *Izvest. Akad. Nauk S.S.S.R., Ser. khim.*, 1967, **11**, 2498.



(74)





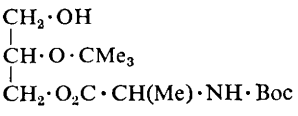
(75) $R^1 = C_{17}H_{33}$, $R^2 = C_{15}H_{31}$

Scheme 3

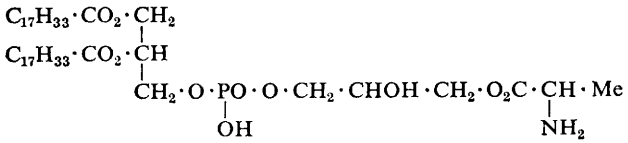
The *R* and *S* forms of 1-(benzyloxycarbonyl-L-alanyl)-2-*O*-benzyl-3-iodoglycerol were obtained from the appropriate stereoisomer of 2-*O*-benzyl-3-iodoglycerol on reaction with benzyloxycarbonyl-L-alanine in the presence of dicyclohexylcarbodi-imide. Debenzylation, followed by benzyloxycarbonylation at the same position, gave products which, on treatment with the silver salt of benzyl distearoylglyceryl phosphate, yielded the protected precursors of the required phosphatidylglycerols.

A synthesis of *O*-amino-acid esters of phosphatidylglycerols has been reported using phosphorus oxychloride as a condensing reagent.[116] 1,2-Dioleoylglycerol, on treatment with phosphorus oxychloride and the protected glycerol (76), readily formed (77) after removal of protecting groups. Complex lipids containing hydroxy-amino-acids directly linked to the phosphate group have also been synthesised by direct condensation of the amino- and carboxyl-protected amino-acid with the appropriate

[116] V. I. Shvets, S. M. Presnyakova, L. V. Malysheva, and N. A. Preobrazhenskii, *Zhur. Vsesoyuz. Khim. obshch. im. D. I. Mendeleeva*, 1968, **13**, 112.



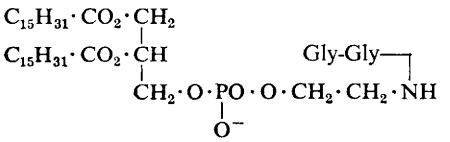
(76)



(77)

glycerol, in the presence of phosphorus oxychloride. Syntheses of *O*-(L-2,3-dioleoyloxy-1-glycerylphosphoryl)-L-tyrosine,[117] 1-(1,2-dilinoleoyl)glyceryl-phosphoryl serine,[118] and a phosphatidyl-hydroxy-L-proline [119] have been reported using this method.

Further work on the synthesis of model phosphatidopeptides for use as prototypes in the investigation of biological activities has been reported.[120] In the synthesis of (78) the diglycyl residue was attached to ethanolamine prior to condensation with the phosphatidyl residue. The condensation of *N*-phthaloylglycylglycine with ethanolamine took place using ethyl chloroformate in the presence of triethylamine at $-10°$. The substituted ethanolamine after treatment with sodium iodide was then reacted with the silver salt of benzyl 1,2-dipalmitoylglyceryl phosphate to give the protected form of (78). A synthesis of *O*-[1-(1,2-distearoyl)glycerylphosphoryl-(1-myoninositol)-*N'*-(glycylalanyl)] ethanolamine has also been described.[121]



(78)

The partial structure, $H_2N\text{-}(CH_2)_4CH(CO_2R)NH\cdot CO\cdot C_{15}H_{31}$ has been proposed [122] for siolipin A, a new lipo-amino-acid ester from *Streptomyces*

[117] V. I. Shvets, L. F. Pogrebnaya, A. A. Kraevskii, and N. A. Preobrazhenskii, *Zhur. org. Khim.*, 1968, **4**, 971.
[118] V. I. Shvets, M. K. Petrova, G. Kazenova, and N. A. Preobrazhenskii, *Zhur. obshchei Khim.*, 1967, **37**, 1454.
[119] D. L. Turner, M. J. Silver, R. R. Holburn, and E. Baczynski, *Lipids*, 1968, **3**, 228.
[120] S. Y. Mel'nik, M. A. Miropol'skaya, and G. I. Samokhvalov, *Zhur. obshchei Khim.*, 1967, **37**, 2452.
[121] A. I. Lyutik, A. V. Luk'yanov, E. S. Zhdanovich, and N. A. Preobrazhenskii, *Zhur. obshchei Khim*, 1968, **38**, 2251.
[122] J. Kawanami, A. Kimura, and H. Otsuka, *Biochim. Biophys. Acta*, 1968, **152**, 808.

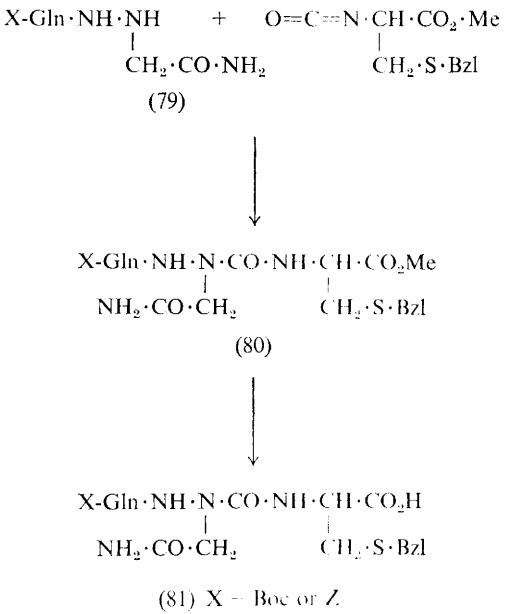
sioyaensis. A peptidyl phosphatidylglycerol has also been isolated from lamellas of *Phaseolus* kidney bean chloroplasts.[123]

8 Amino-acids and Peptides Conjugated to Steroids

Further work on the synthesis of amino-acid and peptide derivatives of steroids has been described.[124] The amino-steroid funtamine (3-amino-pregnan-20-one) condensed readily with an *N*-protected amino-acid or peptide using either dicyclohexylcarbodi-imide or Woodward's Reagent K as coupling agents. The active ester method using 2,4,5-trichlorophenyl esters also gave satisfactory results.

9 Peptide Analogues

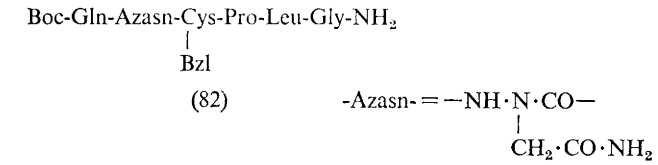
Interest continues in the synthesis of compounds containing modified peptide links. The synthesis of novel peptides containing an aza-asparagine residue has been outlined.[125] The peptide (81) was synthesised from the hydrazinopeptide (79) as shown in Scheme 4. The peptide analogue (79)



Scheme 4

[123] V. S. Chigirev, V. I. Shvets, and E. N. Bezinger, *Doklady Akad. Nauk, S.S.S.R.*, 1968, **181**, 747.
[124] Nguytn-Dang Tâm and E. Philogene, *Bull. Soc. chim. France*, 1967, 3805.
[125] H. Niedrich, 'Peptides,' ed. E. Bricas, North-Holland Publishing Co., Amsterdam, 1968, p. 267.

was prepared from 2-hydrazino-acetamide and *N*-protected glutamine nitrophenyl ester. Because of the lability of (80) towards alkali, hydrolysis of the methyl ester group was achieved using chymotrypsin. Aza-peptide (81) was reacted further with Pro-Leu-Gly-NH_2 using conventional coupling methods to yield (82) an analogue of the *C*-terminal hexapeptide of oxytoxin.



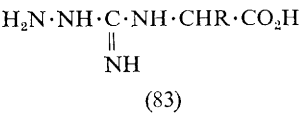
Boc-Gln-Azasn-Cys-Pro-Leu-Gly-NH_2

Bzl

(82) -Azasn- = —NH·N·CO—

$CH_2 \cdot CO \cdot NH_2$

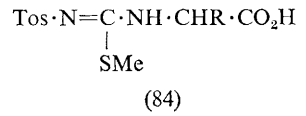
The synthesis of *N*-(aminoguanyl)amino-acid derivatives has been reported.[126] Treatment of the sodium salt of an amino-acid with *S*-methyl isothiosemicarbazide hydroiodide ($H_2NNHC(=NH)SCH_3$,HI) gave the dipeptide analogue (83). Salts of amino-acids on reaction with *S*,*S*-dimethyl- *N*-tosyldithiocarbonimidate [Tos—N=C(SMe)$_2$] gave the derivatives (84). These derivatives were found to be reactive towards nucleophiles and could be converted into the 'peptide-like' analogues (85), (86), (87), and (88) on treatment with hydrazine, amines, alcohols, and water respectively.



$H_2N \cdot NH \cdot C \cdot NH \cdot CHR \cdot CO_2H$

NH

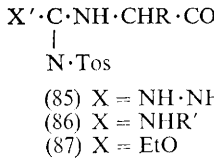
(83)



$Tos \cdot N{=}C \cdot NH \cdot CHR \cdot CO_2H$

SMe

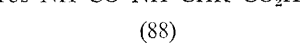
(84)



$X' \cdot C \cdot NH \cdot CHR \cdot CO_2H$

$N \cdot Tos$

(85) X = $NH \cdot NH_2$
(86) X = NHR′
(87) X = EtO

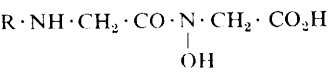


$Tos \cdot NH \cdot CO \cdot NH \cdot CHR \cdot CO_2H$

(88)

Hydroxyaminolysis of *N*-substituted phthalimides under mild conditions has provided a method for the preparation of *N*-hydroxypeptides.[127] The *N*-hydroxy derivative (89) of glycylglycine was prepared by hydroxyaminolysis of the *N*-phthaloyl derivative (90). The compound was also prepared from *N*-phthaloylglycyl chloride and *N*-hydroxyglycine.

[126] J. Gante, *Chem. Ber.*, 1968, **101**, 1195.
[127] O. Neunhoeffer, G. Lehmann, D. Haberer, and G. Steinle, *Annalen*, 1968, **712**, 208.

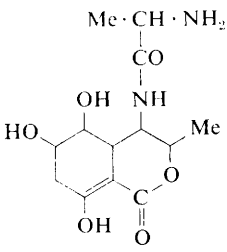


(89) R = H
(90) R = Pht

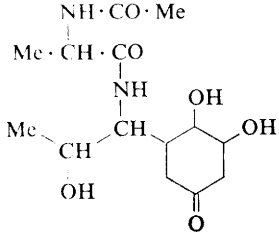
10 Miscellaneous

The biosynthesis of peptide antibiotics has been reviewed [128] and an empirical rule has been introduced to explain the biosynthetic origin of D-amino acids.[129]

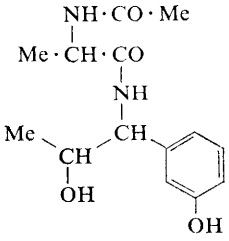
A. Actinobolin.—The structure of the antibiotic actinobolin (91) isolated from *Streptomyces griseoviridus* var. *atrofaciens* has been elucidated.[130] The compound (92) was obtained on treatment of the *N*-acetyl derivative of (91) with ammonia. The structure (92) was consistent with the formation of



(91)



(92)

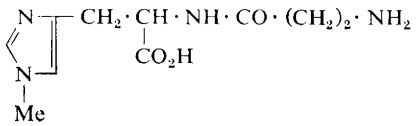


(93)

L-alanine on vigorous hydrolysis and L-alanyl-L-threonine on mild permanganate oxidation. The n.m.r. spectrum of the derived aromatic compound (93) also confirmed the point of attachment of the side-chain to the ring system.

B. Ophidine.—A modified structure has been proposed [131] for ophidine (94) obtained from snake muscle or whale tissues. The new structure was based

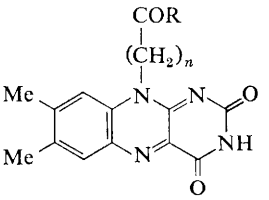
[128] A. B. Mauger, *Experientia*, 1968, **24**, 1068.
[129] M. Bodanszky and D. Perlman, *Nature*, 1968, **218**, 291.
[130] M. E. Munk, D. B. Nelson, F. J. Antocz, D. L. Herald jun., and T. H. Haskell, *J. Amer. Chem. Soc.*, 1968, **90**, 1087.
[131] J. Wolff, K. Horisaka, and H. M. Fales, *Biochemistry*, 1968, **7**, 2455.



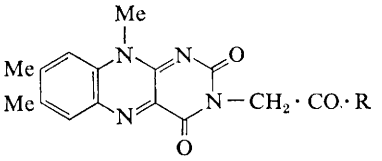
(94)

on the results of *N*-methyl analysis using the gold chloride-hydriodic acid method, and on the n.m.r. spectrum. Two peaks appeared in the n.m.r. spectrum in the region 2–3·1τ, corresponding to the two olefinic protons in the 3-methylhistidine residue.

C. Flavin-peptides.—Attempts to characterise the nature of the interaction between the flavin nucleus [7,8-dimethyl(isoalloxazine)] and aromatic amino-acids have been described.[132] The study involved the synthesis of flavin-peptides with varying side-chain lengths. Reaction of 10-ω-carboxyalkylflavins (95) or 3-ω-carboxyalkylflavins (97) with *N*,*N*′-



(95) R = OH
(96) R = Trp, Tyr, or Phe(methyl esters) } n = 3, 4, 5



(97) R = OH
(98) R = Trp, Tyr, or Phe(methyl esters)

carbonyldi-imidazole gave the corresponding acylimidazoles, which were reacted, without isolation, with an aromatic amino-acid methyl ester to give (96) and (98) respectively. Interaction of the aromatic amino-acid residue with the flavin resulted in a broadening of the visible flavin absorption towards green, the fluorescence increasing with increasing number of methylene groups in the side-chain.

[132] W. Föry, R. E. MacKenzie, and D. M. McCormick, *J. Heterocyclic Chem.*, 1968, **5**, 625.

5

The Relationship between the Structure and Biological Activity of Some Peptides and Proteins (excluding Enzymes)

BY D. G. SMYTH

1 Introduction

Since the last General Report on Proteins and Peptides,[1] spectacular advances have been made in the determination of structure and in synthesis; understanding of biological function at the molecular level will follow. Outstanding achievements are: a total synthesis of the first enzyme, pancreatic ribonuclease,[2] by the solid-phase method;[3] preliminary synthesis of the same protein by classical techniques;[4] synthesis of the fragments spanning the sequence of ribonuclease-T_1;[5] synthesis of the peptide hormone secretin.[6] The complete synthesis of crystalline glucagon [7] stands as a model demonstrating the precision and effort of the synthetic method in solution. The refined techniques now available for the elucidation of primary structure are reflected in the successful determination of the total amino-acid sequence of a human myeloma IgG protein,[8] subunit *M* 75,000. Detailed *X*-ray crystallographic analysis of ribonuclease [9, 10]

[1] D. G. Smyth, *Ann. Reports Chem. Soc.*, 1965, **62**, 488.
[2] B. Gutte and R. B. Merrifield, *J. Amer. Chem. Soc.*, 1969, **91**, 501.
[3] R. B. Merrifield, *Biochemistry*, 1964, **3**, 1385.
[4] R. G. Denkewalter, D. F. Veber, F. W. Holly, and R. Hirschmann, *J. Amer. Chem. Soc.*, 1969, **91**, 502; R. G. Strachan, W. J. Paleveda, R. F. Nutt, R. A. Vitali, D. F. Veber, M. J. Dickinson, V. Garsky, J. E. Deak, E. Walton, S. R. Jenkins, F. W. Holly, and R. Hirschmann, *ibid.*, 1968, **91**, 503; S. R. Jenkins, R. F. Nutt, R. S. Dewey, D. F. Veber, F. W. Holly, W. J. Paleveda, T. Lanza, R. G. Strachan, E. F. Schoenewaldt, H. Barkemeyer, M. J. Dickinson, J. Sondey, R. Hirschmann, and E. Walton, *ibid.*, 1968, **91**, 505; D. F. Veber, S. L. Varga, J. D. Milowski, H. Joshua, J. B. Conn, R. Hirschmann, and R. G. Denkewalter, *ibid.*, 1968, **91**, 506; R. Hirschmann, R. F. Nutt, D. F. Verber, R. A. Vitali, S. L. Varga, T. A. Jacob, F. W. Holly, and R. G. Denkewalter, *ibid.*, 1968, **91**, 507.
[5] N. Yanaihara, C. Yanaihara, G. Dupuis, J. Beacham, R. Camber, and K. Hofmann, *J. Amer. Chem. Soc.*, 1969, **91**, 2184.
[6] M. A. Ondetti, V. L. Narayanan, M. Von Saltza, J. T. Sheehan, E. F. Sabo, and M. Bodanszky, *J. Amer. Chem. Soc.*, 1968, **90**, 471.
[7] E. Wünsch, *Z. Naturforsch*, 1967, 22b, 1269; E. Wünsch, G. Wendelberger, E. Jaeger, and R. Scharf, in 'Peptides,' 1968, ed. E. Bricas, North-Holland Publishing Co., Amsterdam. p. 229.
[8] 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, **62**, in press.
[9] G. Kartha, J. Bello, and D. Harker, *Nature*, 1967, **213**, 862.
[10] H. W. Wycoff, K. D. Hardman, N. M. Allwell, T. Inagami, L. N. Johnson, and F. N. Richards, *J. Biol. Chem.*, 1967, **242**, 3984.

and of chymotrypsinogen [11] lays open their three-dimensional structures; *X*-ray studies proceed on a crystalline IgG myeloma protein.[12] The isolation, determination of sequence, and total synthesis of porcine and human calcitonins, 32 amino-acid residues, have been accomplished.[13]

Progress in understanding the molecular basis of hormone action is waiting upon the isolation and chemical characterisation of receptors and on the elucidation of the nature of the processes involved in the stages connecting receptor activation and biological effect. In this Report, advances have been selected which indicate the general direction of experimental design and illustrate current concepts.

2 Immunoglobulin-G (IgG)

(See also pp. 104 and 172.)

A homogeneous IgG with definite antigen-binding ability is still sought. An antibody isolated from hyperimmune rabbits immunised against Group-A variant streptococcal vaccines exhibited a restricted range of electrophoretic properties,[14] which contrasts with the wide range exhibited by the normal complement of circulating IgG or the average heterogeneous antibody.[15] The relative purity of the antibody was attributed to the simple chemical nature of the immunising hapten, which consisted almost entirely of a single sugar constituent, L-rhamnose, present as a repeating disaccharide unit. Reports are accumulating of homogeneous myeloma proteins that exhibit antibody activity.[16] The binding affinities, however, are mostly lower than those of a normal antibody and in the one case where a high affinity was present, the myeloma was an IgA protein with a single binding site per four-chain molecule.[16b] Caution should be exercised in assuming that the binding properties of these myeloma proteins represent the physiological binding of an antibody for its specific antigen.

An enterprising plan to produce a homogeneous antibody was developed by Brenneman and Singer.[17] A chemically and structurally homogeneous

[11] B. W. Matthews, P. B. Sigler, R. Henderson, and D. M. Blow, *Nature*, 1968, **214**, 62; J. Kraut, T. H. Wright, M. Kellerman, and S. T. Freer, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **58**, 304.

[12] H. P. Avey, R. J. Poljak, J. Rossi, and A. Nisonoff, *Nature*, 1968, **220**, 1248.

[13] F. W. Kahnt, B. Riniker, I. MacIntyre, R. Neher, *Helv. Chim. Acta*, 1968, **51**, 214; R. Neher, B. Riniker, H. Zuber, W. Rittel, and F. W. Kahnt, *ibid.*, 1968, **51**, 917; W. Rittel, M. Brugger, B. Kamber, B. Riniker, and P. Sieber, *ibid.*, 1968, **51**, 924; J. T. Potts, H. D. Niall, H. T. Keutmann, H. B. Brewer, and L. J. Deftos, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 1321; P. H. Bell, W. F. Barg, D. F. Colucci, M. C. Davies, C. Dziobkowski, M. E. Englert, E. Heyder, R. Paul, and E. H. Snedeker, *J. Amer. Chem. Soc.*, 1968, **90**, 2705; S. Guttman, J. Pless, E. Sandrin, P. A. Jaquenoud, H. Bossert, and H. Williams, *Helv. Chim. Acta*, 1968, **51**, 1155.

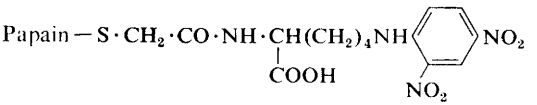
[14] J. B. Fleischmann, D. G. Braun, and R. M. Krause, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 134; D. G. Braun, K. Eichmann, and R. M. Krause, *J. Exp. Med.*, 1969, **129**, 809.

[15] Z. L. Awdeh, A. R. Williamson, and B. A. Askonas, *Nature*, 1968, **219**, 66.

[16] (*a*) H. Metzger, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **57**, 1490. (*b*) H. N. Eisen, E. S. Simms, and M. Potter, *Biochemistry*, 1968, 7, 4126. (*c*) D. Schubert, A. Jobe, and M. Cohn, *Nature*, 1968, **220**, 882.

[17] L. Brenneman and S. J. Singer, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 258.

hapten–protein conjugate was synthesised by the attachment of a single haptenic group at the unique cysteine in the active site of papain. Modification was effected by carboxymethylation at the SH-group and the product was coupled to ε-dinitrophenyl-lysine, as in (1). The hapten was



(1)

thus orientated in a precise manner with respect to its carrier protein and immunisation was found to lead to the production of antibody with electrophoretically homogeneous L-chains. The effect, however, was especially prominent only during early stages of the immunisation, when the antibody titres were low. The problem remains of accumulating homogeneous, functional IgG in sufficient quantity for chemical and structural analysis.

Analysis of the primary structure of the H-chain of rabbit IgG is well advanced [18] and of H- and L-chain in a human myeloma IgG is complete (Figure 1).[19] The absence of antigen-binding ability unfortunately precludes interpretation in terms of function. The sequence data obtained for the myeloma protein will be of value in comparative studies with other myeloma proteins in order to establish which regions of the IgG molecule are invariable. Comparison can already be made with the structure of a large fragment surrounding the interchain disulphide bonds of another human myeloma IgG.[20] To these results must be added the complete sequence of the κ-chain of a Bence-Jones protein [21] and the λ-chain of a second Bence-Jones protein.[22] It is anticipated that further studies will allow the piecing together of a fixed framework of structure on which the variability of antibody is superimposed.

From the comparison of the sequences of the variable region in a number of Bence-Jones proteins, Kabat has called attention to some unique features

[18] (*a*) J. M. Wilkinson, *Biochem. J.*, 1969, **112**, 183. (*b*) J. J. Cebra, L. A. Steiner, and R. R. Porter, *Biochem. J.*, 1968, **107**, 79; 1968, **107**, 79. (*c*) H. E. Lebovitz, R. DeLaney, R. E. Fellows, and R. L. Hill, *J. Biol. Chem.*, 1968, **243**, 4197; R. E. Delaney and R. L. Hill, *ibid.*, 1968, **243**, 4206; R. L. Hill, H. E. Lebovitz, R. E. Fellows, and R. DeLaney, 'Gamma Globulins: Structure and Control of Biosynthesis,' ed. J. Killander, Nobel Symposium, **3**, Interscience, New York, 1967, p. 109.

[19] (*a*) G. M. Edelman, W. E. Gall, M. J. Waxdal, and W. H. Konigsberg, *Biochemistry*, 1968, **7**, 1950; (*b*) M. J. Waxdal, W. H. Konigsberg, W. L. Henley, and G. M. Edelman, *ibid.*, 1968, **7**, 1959; (*c*) M. J. Waxdal, W. H. Konigsberg, and G. M. Edelman, *ibid.*, 1968, **7**, 1967; (*d*) W. E. Gall, B. A. Cunningham, M. J. Waxdal, W. H. Konigsberg, and G. M. Edelman, *ibid.*, 1968, **7**, 1973; (*e*) B. A. Cunningham, P. D. Gottlieb, W. H. Konigsberg, and G. M. Edelman, *ibid.*, 1968, **7**, 1983.

[20] L. A. Steiner and R. R. Porter, *Biochemistry*, 1967, **6**, 3957.

[21] N. Hilschmann, *Z. physiol. Chem.*, 1967, **348**, 1718.

[22] C. Milstein, J. B. Clegg, and J. M. Jarvis, *Biochem. J.*, 1968, **110**, 631.

L-*chain*

Asp-Ile-Gln-Met-Thr-Gln-Ser-Pro-Ser-Thr-Leu-Ser-Ala-Ser-Val-Gly-Asp-Arg-Val-Thr-Ile-Thr-Cys-Arg-Ala-Ser-Gln-Ser-Ile-Asn-
1 **10** **20** **30**

Thr-Trp-Leu-Ala-Trp-Tyr-Gln-Gln-Lys-Pro-Gly-Lys-Ala-Pro-Lys-Leu-Leu-Met-Tyr-Lys-Ala-Ser-Ser-Leu-Glu-Ser-Gly-Val-Pro-Ser-
40 **50** **60**

Arg-Phe-Ile-Gly-Ser-Gly-Ser-Gly-Thr-Glu-Phe-Thr-Leu-Thr-Ile-Ser-Ser-Leu-Gln-Pro-Asp-Asp-Phe-Ala-Thr-Tyr-Tyr-Cys-Gln-Gln-
70 **80** **90**

Tyr-Asn-Ser-Asp-Ser-Lys-Met-Phe-Gly-Gln-Gly-Thr-Lys-Val-Glu-Val-Lys-Gly-Thr-Val-Ala-Ala-Pro-Ser-Val-Phe-Ile-Phe-Pro-Pro-
100 **110** **120**

Ser-Asp-Glu-Gln-Leu-Lys-Ser-Gly-Thr-Ala-Ser-Val-Val-Cys-Leu-Leu-Asn-Asn-Phe-Tyr-Pro-Arg-Glu-Ala-Lys-Val-Gln-Trp-Lys-Val-
130 **140** **150**

Asp-Asn-Ala-Leu-Gln-Ser-Gly-Asn-Ser-Gln-Glu-Ser-Val-Thr-Glu-Gln-Asp-Ser-Lys-Asp-Ser-Thr-Tyr-Ser-Leu-Ser-Ser-Thr-Leu-Thr-
160 **170** **180**

Leu-Ser-Lys-Ala-Asp-Tyr-Glu-Lys-His-Lys-Val-Tyr-Ala-Cys-Glu-Val-Thr-His-Gln-Gly-Leu-Ser-Ser-Pro-Val-Thr-Lys-Ser-Phe-Asn-
190 **200** **210**

Arg-Gly-Glu-Cys
214

H-*chain*

Pca-Val-Gln-Leu-Val-Gln-Ser-Gly-Ala-Glu-Val-Lys-Lys-Pro-Gly-Ser-Ser-Val-Lys-Val-Ser-Cys-Lys-Ala-Ser-Gly-Gly-Thr-Phe-Ser-
1 **10** **20** **30**

Arg-Ser-Ala-Ile-Ile-Trp-Val-Arg-Gln-Ala-Pro-Gly-Gln-Gly-Leu-Glu-Trp-Met-Gly-Gly-Ile-Val-Pro-Met-Phe-Gly-Pro-Pro-Asn-Tyr-
40 **50** **60**

Ala-Gln-Lys-Phe-Gln-Gly-Arg-Val-Thr-Ile-Thr-Ala-Asp-Glu-Ser-Thr-Asn-Thr-Ala-Tyr-Met-Glu-Leu-Ser-Ser-Leu-Arg-Ser-Glu-Asp-
70 **80** **90**

Thr-Ala-Phe-Tyr-Phe-Cys-Ala-Gly-Gly-Tyr-Gly-Ile-Tyr-Ser-Pro-Glu-Glu-Tyr-Asn-Gly-Gly-Leu-Val-Thr-Val-Ser-Ser-Ala-Ser-Thr-
100 **110** **120**

Lys-Gly-Pro-Ser-Val-Phe-Pro-Leu-Ala-Pro-Ser-Ser-Lys-Ser-Thr-Ser-Gly-Gly-Thr-Ala-Ala-Leu-Gly-Cys-Leu-Val-Lys-Asp-Tyr-Phe-
130 **140** **150**

Pro-Glu-Pro-Val-Thr-Val-Ser-Trp-Asn-Ser-Gly-Ala-Leu-Thr-Ser-Gly-Val-His-Thr-Phe-Pro-Ala-Val-Leu-Gln-Ser-Ser-Gly-Leu-Tyr-
160 **170** **180**

Ser-Leu-Ser-Ser-Val-Val-Thr-Val-Pro-Ser-Ser-Ser-Leu-Gly-Thr-Gln-Thr-Tyr-Ile-Cys-Asn-Val-Asn-His-Lys-Pro-Ser-Asn-Thr-Lys-
190 **200** **210**

Val-Asp-Lys-Arg-Val-Glu-Pro-Lys-Ser-Cys-Asp-Lys-Thr-His-Thr-Cys-Pro-Pro-Cys-Pro-Ala-Pro-Glu-Leu-Leu-Gly-Gly-Pro-Ser-Val-
220 **230** **240**

Phe-Leu-Phe-Pro-Pro-Lys-Pro-Lys-Asp-Thr-Leu-Met-Ile-Ser-Arg-Thr-Pro-Glu-Val-Thr-Cys-Val-Val-Val-Asp-Val-Ser-His-Glu-Asp-
250 **260** **270**

Pro-Gln-Val-Lys-Phe-Asn-Trp-Tyr-Val-Asp-Gly-Val-Gln-Val-His-Asn-Ala-Lys-Thr-Lys-Pro-Arg-Glu-Gln-Gln-Tyr-Asx-Ser-Thr-Tyr-
280 **290** **300**

Arg-Val-Val-Ser-Val-Leu-Thr-Val-Leu-His-Gln-Asn-Trp-Leu-Asp-Gly-Lys-Glu-Tyr-Lys-Cys-Lys-Val-Ser-Asn-Lys-Ala-Leu-Pro-Ala-
310 **320** **330**

Pro-Ile-Glu-Lys-Thr-Ile-Ser-Lys-Ala-Lys-Gly-Gln-Pro-Arg-Glu-Pro-Gln-Val-Tyr-Thr-Leu-Pro-Pro-Ser-Arg-Glu-Glu-Met-Thr-Lys-
340 **350** **360**

Asn-Gln-Val-Ser-Leu-Thr-Cys-Leu-Val-Lys-Gly-Phe-Tyr-Pro-Ser-Asp-Ile-Ala-Val-Glu-Trp-Glu-Ser-Asn-Asp-Gly-Glu-Pro-Glu-Asn-
370 **380** **390**

Tyr-Lys-Thr-Thr-Pro-Pro-Val-Leu-Asp-Ser-Asp-Gly-Ser-Phe-Phe-Leu-Tyr-Ser-Lys-Leu-Thr-Val-Asp-Lys-Ser-Arg-Trp-Gln-Glu-Gly-
400 **410** **420**

Asn-Val-Phe-Ser-Cys-Ser-Val-Met-His-Glu-Ala-Leu-His-Asn-His-Tyr-Thr-Gln-Lys-Ser-Leu-Ser-Leu-Ser-Pro-Gly
430 **440** **646**

Figure 1 *The amino-acid sequence of a human* IgG *myeloma protein* (Eu)
(Reproduced from G. M. Edelman, B. A. Cunningham, W. E. Gall, P. D. Gottlieb, V. Rutishauser, and M. J. Waxdal, *Proc. Nat. Acad. Sci. U.S.A.*, in press; from the results of Edelman and his colleagues and Konigsberg and his colleagues [19a–e])

and their possible relation to complementarity.[23] Certain stretches were attributed to play a role in binding of L-chain to H-chain, which is considered to provide a cradle for the antigen-binding site. Other stretches of the variable region of the L-chain were assigned to contain the actual contact amino-acids, whose variation is believed to determine the specificity of antigen binding.

Electric charge relationships between antigen and antibody have been demonstrated. Positively charged groups predominated in an antibody made against a negatively charged antigen, a polymer of glutamic acid and tyrosine, whereas the antibody against a basic antigen was negatively charged.[24] Similar findings have been reported concerning an antibody made against *p*-azobenzenearsenate, which was found to possess a surfeit of NH_4^+ and guanidinium groups as alternatives in the combining site; chemical modification of carboxyl groups did not affect the binding properties.[25]

The carbohydrate structure of rabbit IgG [26] has been elucidated (Figure 2). The C_3-oligosaccharide chain, which includes five glucosamine residues, is present on both H-chains in all molecules. The C_2-oligosaccharide, which includes one galactosamine residue, is present on 40% of the H-chains; it appears to be attached to one H-chain and not the other in the four-chain structure. The C_1-oligosaccharide, which includes two glucosamine residues, is present on 15% of the H-chains. L-Chain in rabbit IgG is devoid of carbohydrate. It may be noted, however, that a human L-type Bence-Jones protein has been reported to contain an oligosaccharide unit.[29]

The polypeptide chains of IgG are biosynthesised in the microsome fraction of the lymph node cells [30] whereas carbohydrate attachment takes place in the particle-free cell sap.[31] Evidence was presented that the initial carbohydrate is added close to the time of synthesis of the polypeptide chains while the last residues are added immediately before secretion of the finished molecule into the extracellular medium. A specific enzymic mechanism was postulated for the synthesis of the carbohydrate component of IgG in the lymph node cells. IgG is, in any case, practically the sole protein secreted by these cells. A suitable system for further study of carbohydrate attachment is available in a cell-free preparation obtained from Plasmacytoma tumours, which have been shown to biosynthesise a specific IgG.[32]

[23] E. A. Kabat, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 613; S. J. Singer and N. O. Thorpe, *ibid.*, 1968, **60**, 1371.
[24] E. Rude. E. Mozes, and M. Sela, *Biochemistry*, 1968, **7**, 2971.
[25] M. H. Freedman, A. L. Grossberg, and D. Pressman, *J. Biol. Chem.*, 1968, **243**, 6186.
[26] M. W. Fanger and D. G. Smyth, *Biochem. J.*, in the press.
[27] D. G. Smyth and S. Utsumi, *Nature*, 1967, **216**, 332.
[28] D. Givol and D. DeLorenzo, *J. Biol. Chem.*, 1968, **243**, 1886.
[29] A. B. Edmundson, F. A. Sheber, K. R. Ely, N. B. Simonds, N. K. Hutson, and J. L. Rossiter, *Arch. Biochem. Biophys.*, 1968, **127**, 725.
[30] R. M. Swenson and M. Kern, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **57**, 417.
[31] R. M. Swenson and M. Kern, *J. Biol. Chem.*, 1967, **242**, 3242.
[32] B. Mach, H. Koblet, and D. Gros, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 445.

Figure 2 *Chain structure of rabbit immunoglobulin. Oligosaccharide chains are attached at sites indicated by* C_1, C_2, *and* C_3.[26] *The sequence in the hinge region is taken from* refs 27 *and* 28

That some individual rabbits possess IgG devoid of the C_2-oligosaccharide [33] (see Figure 2) precludes an important role for this carbohydrate in secretion of IgG from the cell. In this context, it may be noted that removal of terminal sialic acid from ceruloplasmin results in a rapid disappearance of the asialo-ceruloplasmin from the serum and concomitantly it appears in the parenchymal cells of the liver.[34] The carbohydrate apparently mediates against transfer across the cell membrane. Similarly, rabbit IgG lacking the C_2-oligosaccharide is able to cross the placental membrane more rapidly than the IgG molecule that contains this oligosaccharide.[33] The precise composition of each oligosaccharide chain in rabbit IgG and the complete specificity involved in the sites of attachment of the carbohydrate to the IgG molecule favours the concept that the carbohydrate fulfils an essential role in a biological function, yet to be elucidated.

3 Insulin

(See also pp. 82 and 150)

An insulin precursor has been isolated from human beta-cell tumour and from rat islet tissue.[35] The single chain 'proinsulin' was shown to consist of the B-chain coupled at its carboxyl-terminus through a connecting peptide to the amino-terminus of the A-chain. A proinsulin has now been isolated as a minor component from crystalline porcine insulin and its total sequence [36] has been determined (see Figure 4, chap. 2, part I). The molecule was observed to possess a weak intrinsic insulin-like activity, which was not due to contamination by insulin. Digestion of proinsulin with trypsin gave rise to a fully active two-chain insulin molecule lacking alanine from the carboxyl-terminus of the B-chain; the 31-residue connecting peptide was released at the same time.

There are strong indications that the connecting peptide between B- and A-chains forms a random coil. Of the 33 amino-acids in this region of porcine proinsulin, three are proline residues which will not allow α-helix formation and seven are glycine residues which cannot contribute to stabilisation of an α-helix by side-chain interaction. Of the 30 amino-acids in this region in bovine proinsulin, four are proline and eight are glycine.[37] The insulin moiety of the proinsulin molecule appears to exist in the same conformation as insulin itself, on the basis of c.d. and o.r.d. measurements.[38]

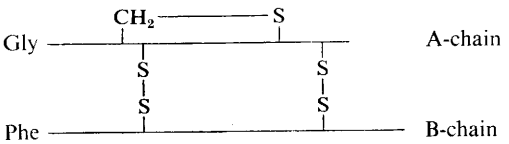
[33] W. Hinrich and D. G. Smyth, unpublished data.
[34] A. G. Morell, R. A. Irvine, I. Sternlieb, and I. H. Scheinberg, *J. Biol. Chem.*, 1968, **243**, 155.
[35] D. F. Steiner and P. E. Oyer, *Proc. Nat. Acad. Sci. U.S.A.*, 1967, **57**, 473; D. F. Steiner, D. Cunningham, L. Spigelman, and B. Aten, *Science*, 1967, **157**, 697.
[36] R. E. Chance, R. M. Ellis, and W. W. Bromer, *Science*, 1968, **161**, 165.
[37] R. E. Chance, quoted in ref. 38; D. D. Schmidt and A. Arens, *Z. Physiol. Chem.*, 1968, **349**, 1157.
[38] B. H. Frank and A. J. Veros, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 155.

The physiological significance of proinsulin and the mechanism responsible for the transformation of proinsulin to insulin are unknown. It seems likely that trypsin or a trypsin-like enzyme may be responsible for the cleavage of the B_{63}—A_1 bond whereas another enzyme may hydrolyse the basic residues around the terminal B_{30}—B_{31} connection to give insulin. The discovery of the proinsulin molecule will stimulate a new field of investigation centering on the determination of proinsulin levels in pathologies involving carbohydrate metabolism.

The previous failure to obtain better than 10–20% recovery of biological activity, as assessed by the mouse convulsion assay, when insulin is reduced and then reoxidised,[39] is now explained on the basis that the disulphide bridges may be formed *in vivo* immediately after the single-chain precursor is biosynthesised and before it undergoes cleavage to release the two-chain insulin molecule. Confirming this hypothesis, it has been demonstrated that both beef and rat proinsulins can be reduced and reoxidised with correct pairing of the disulphide bridges in high yield.[40]

The impairment of protein synthesis in diabetes and the role of insulin in the regulation of protein biosynthesis are currently attracting attention. Insulin insufficiency certainly impairs the formation of amino-acyl-t-RNA molecules in liver [41] but the effect may be due to insufficiency in the amino-acyl-t-RNA synthetases. The hormone has been shown to stimulate amino-acid transport by increasing the affinity of the carrier mechanisms for amino-acids and further it has been suggested that insulin initiates the synthesis of a specific protein which enhances transport across rat diaphragm.[42] That a unique biological action for insulin has not yet been agreed prevents, at this time, detailed analysis of the relation between its structure and function.

The careful synthesis of an insulin analogue (2) in which one of the sulphur atoms of the intra-A-chain disulphide bridge is replaced by a methylene



(2)

group [43] deserves special mention. The considerable biological activity of the product precludes a functional role for the intra-chain disulphide bridge of insulin, which thus cannot participate in a disulphide exchange reaction with a receptor protein.

[39] H. Zahn, personal communication.
[40] D. F. Steiner and J. J. Clark, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 622.
[41] Y. L. Germanyuk and V. I. Mironenko, *Nature*, 1969, **222**, 486.
[42] L. J. Elsas, I. Albrecht, and L. E. Rosenberg, *J. Biol. Chem.*, 1968, **243**, 1846.
[43] K. Jöst, J. Rudinger, H. Klostermeyer, and H. Zahn, *Z. Naturforsch.*, 1968, **23**, 1059.

4 Oxytocin and the Vasopressins

Two comprehensive reviews listing 120 structural analogues and their biological activities are available.[44, 45] In derivatives where activity remains high, unequivocal conclusions have been drawn: the amino-acid or functional group replaced is assumed not to perform an essential role in the biological activity of the hormone. Where an analogue exhibits weak or no activity, on the other hand, less information is obtained. The loss could be related to a deficiency in either of two components: ability of the molecule to approach and bind to the receptor or ability of the molecule to initiate contraction after binding has occurred.

The finding of specific inhibitors of oxytocin without intrinsic activity (Figure 3) [46-48] has opened the way to investigation of the range of structure

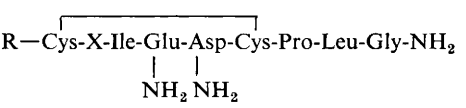


Figure 3 *Specific inhibitors of oxytocin without intrinsic activity.*[46, 47] R = $NH_2 \cdot CO$; X = Phe (*p*Me), Phe (*p*Et), Tyr (Me), or Tyr (Et)

that is compatible with binding as distinct from structure necessary for overall biological activity. Such inhibitors appear to attach at the receptor site, blocking the action of the natural hormone, but do not themselves possess the ability to trigger the process of smooth muscle contraction.

The three-dimensional conformation of oxytocin and vasopressin is little known; the molecular chemistry of their interaction with the receptor sites is totally unknown. The replacement of a single amino-acid could result in a minor alteration in a discrete portion of the molecule or it could lead to disruption of the entire three-dimensional structure. Attempts to probe the conformation of hormone analogues have been made with the aid of circular dichroism measurements.[49] The indications were that the disulphide bonds had dihedral angles close to 90°.

Deamino analogues exhibited decreased absorption bands, which was attributed to divergence in the disulphide region. It is clear that more sophisticated methods will be required to pursue the matter further. It is known that within a protein both 'near-neighbour' and 'distant' interactions determine the configuration of stretches of the peptide chain.[50] By analogy, the

[44] I. L. Schwartz and L. M. Livingston, *Vitamins and Hormones*, 1964, **22**, 261.
[45] R. Walter, J. Rudinger, and I. L. Schwartz, *Amer. J. Med.*, 1967, **42**, 653.
[46] D. G. Smyth, 'Proceedings of the 7th European Peptide Symposium,' ed. V. Bruckner and K. Medzihradszky, Budapest, 1964 *Hung. Chim. Acta*, 1965, **44**, 197; *J. Biol. Chem.*, 1967, **242**, 1579, 1592.
[47] A. Chimiak, K. Eisler, K. Jöst, and J. Rudinger, *Coll. Czech. Chem. Comm.*, 1968, **33**, 2918.
[48] H. Schulz and V. duVigneaud, *J. Medicin. Chem.*, 1966, **9**, 647.
[49] S. Beychak and E. Breslow, *J. Biol. Chem.*, 1968, **243**, 151.
[50] D. C. Phillips, *Proc. Nat. Acad., Sci. U.S.A.*, 1967, **57**, 484; C. Levinthal, *J. Chim. phys.*, 1968, **65**, 44; W. A. Klee, *Biochemistry*, 1968, **7**, 2731.

physiologically active conformation of a peptide hormone may be expected to be determined not only by its inherent tendency to assume a certain shape but also by the intermolecular forces induced between the hormone and its complementary receptor.

5 The Biological Role of Adenosine-3′,5′-monophosphate (Cyclic AMP)

Cyclic AMP has been established as a second 'messenger' mediating the effects of a variety of hormones, including glucagon, adrenocorticotropin

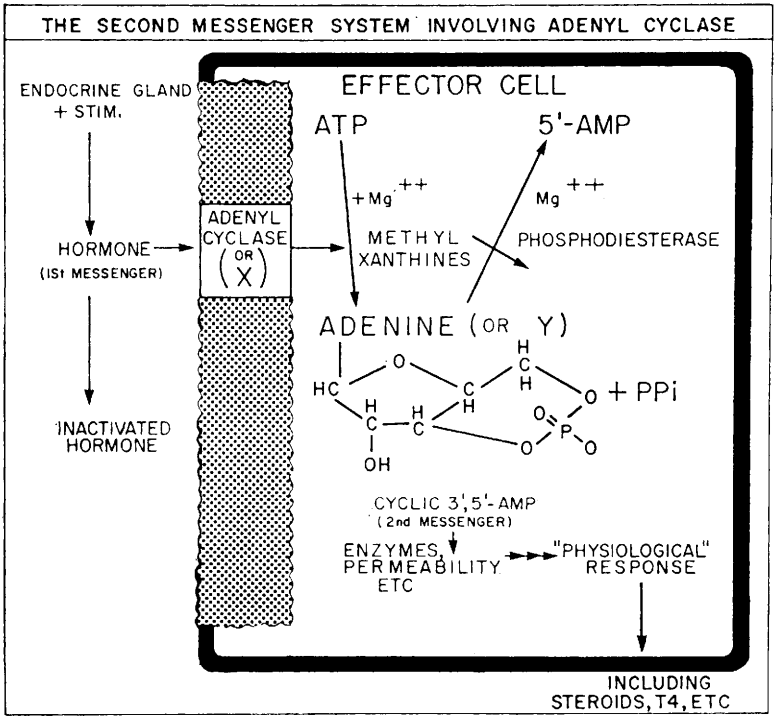


Figure 4 *Diagram illustrating the hypothesis of the 'second messenger' in hormone action*
(Reproduced by permission from *Circulation*, 1968, **37**, 279)

(ACTH), vasopressin, and adrenaline.[51, 52] The first messenger, the hormone itself, travels from its cells of origin to the cells of the target tissue and causes there an alteration in the intracellular level of the second messenger (Figure 4).[53] The formation of cyclic AMP from ATP is catalysed

[51] E. W. Sutherland and G. A. Robson, *Pharmacol. Rev.*, 1966, **18**, 145.
[52] M. Rodbell, A. B. Jones, G. E. Chiappe de Cingolani, and L Birnbaumer, *Rec. Progr. Hormone Res.*, 1968, **24**, 215.

by the hormone-stimulated adenyl cyclase [54] and also by a specific cyclic 3′,5′-nucleotide phosphodiesterase [55] (cyclic phosphodiesterase), both of which occur in the cell membrane and participate in the regulation of metabolic systems.

Adenyl cyclase is affected by different hormones in different tissues.[56] By this mechanism the initial extracellular hormone is able to trigger an intracellular 'signal' which can be amplified, the ultimate response depending on the enzymatic profile of the cell involved. It should be noted, however, that non-hormonal factors such as potassium concentration may also be important, in some tissues, in regulating the level of the cyclase enzymes. The only second messenger identified so far is cyclic AMP, but it seems possible that other cyclic nucleotides may function in a similar capacity.

6 Hormone Receptors

No receptor to a peptide hormone has yet been isolated. Paton and Rang [57] studied the absolute concentration of receptors to acetylcholine and demonstrated a very low density of receptor sites. In experiments on the binding of steroid hormones to uterus, a macromolecular fraction has been isolated which retains a high binding affinity for oestradiol.[58] It must be borne in mind, however, that the marked specificity of oestradiol binding *in vivo* is not preserved during *in vitro* experiments.[59] The receptor appears to be a lipoprotein of moderate size, having a sedimentation coefficient of 4*S*, but it exhibits a strong tendency to aggregate.[60] The mechanism of binding has been considered to involve two steps, 'uptake' and 'retention'.[61]

Structure–affinity relationships have been calculated by Rasmussen [62] for a series of oxytocin, vasopressin analogues in their action on the isolated toad bladder; the affinity constants deduced did not take into account possible differences in the rate of flow of each hormone into the cell. When a permeability factor (α) was considered in conjunction with the affinity constant (K), the following order of physiological effectiveness was found: arginine vasopressin > oxytocin > deamino-oxytocin.[63] In the same bioassay, the pH dependence of the antidiuretic action of deamino-oxytocin was found to be almost identical with that of oxytocin, with a

[53] E. W. Sutherland, A. Robison, and R. W. Butcher, *Circulation*, 1968, **37**, 279.
[54] E. W. Sutherland, T. W. Rall, and T. Menon, *J. Biol. Chem.*, 1962, **237**, 1220.
[55] R. W. Butcher and E. W. Sutherland, *J. Biol. Chem.*, 1962, **237**, 1244.
[56] J. Orloff and J. S. Handler, *Amer. J. Med.*, 1964, **36**, 686.
[57] W. D. M. Paton, and H. P. Rang, *Proc. Roy. Soc.*, 1965, *B*, **163**, 1.
[58] G. P. Talwar, S. J. Segal, A. Evans, and D. W. Davidson, *Proc. Nat. Acad. Sci. U.S.A.*, 1964, **52**, 1059.
[59] R. J. B. King, J. Gordon, D. M. Cowan, and D. R. Inman, *J. Endocrinol.*, 1966, **36**, 139.
[60] T. Erdos, *Biochem. Biophys, Res. Comm.*, 1968, **32**, 338.
[61] E. V. Jensen, T. Suzuki, T. Kawashima, W. E. Stumpf, P. W. Jungblut, and E. R. DeSombre, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 632.
[62] H. Rasmussen, I. L. Schwartz, R. Young, and J. Marc-Aurele, *J. Gen. Physiol.*, 1963, **46**, 1171.
[63] P. F. Gulyassy and I. S. Edelman, *Biochim. Byophys. Acta*, 1965, **102**, 185.

maximum at pH 8·0. This suggests that the dissociable group involved is present on the receptor and not on the hormone.[63] As the antidiuretic response to AMP passes through a maximum at pH 7·0, and not pH 8·0, it seems likely that the dissociable group in question is a component of the first reaction step involving the hormone rather than a component of the membrane site where the changes in permeability occur.

Theoretical approaches to the description of hormone–receptor interactions [64] are generally based on complex mathematics.[65] The complexity is increased further when such phenomena as allosteric effects are invoked to provide *ad hoc* explanations for experimental observations. There is pressing need for the chemical isolation of hormone receptors in a form that will allow expression of their physiological properties.

[64] D. R. Waud, *Pharmacol. Rev.*, 1968, **20**, 49.
[65] D. Mackay, *J. Pharm. Pharmacol.*, 1966, **18**, 201.

6

Metal Derivatives of Amino-acids, Peptides, and Proteins

BY R. D. GILLARD AND S. H. LAURIE

1 Introduction

This article is primarily concerned with those facets of the chemistry of metal derivatives of amino-acids, peptides, and proteins which reflect the alteration of the properties of the organic ligand consequent upon bonding to the metal. We have paid little attention to the great amount of work concerned with elucidating the electronic nature of the metal in such environments. The literature published during the year 1968 has been covered and relevant work prior to this is also included.

There are a considerable number of recent reviews concerning the present topic. A selection of the more relevant titles are: 'Stereoselectivity and Reactivity in Complexes of Amino Acids and Peptides;'[1] 'Role of Transition-metal Ions in Biological Processes;'[2] 'Enzyme Actions: Views Derived from Metalloenzyme Studies;'[3] 'Transition-metal Ions as Reagents in Metallo-enzymes;'[4] 'Crystal Structures of Metal–Peptide Complexes;'[5] 'Chemical and Biological Applications of Mössbauer Spectroscopy;'[6] 'Catalytic Functions of Metal Ions and Their Complexes;'[7] 'The Biochemistry of Copper;'[8,9] 'Ferredoxin and other Non-haem Iron Compounds;'[10] 'Biochemical Aspects of Iron-Sulphur Linkages in Non-haem Iron Proteins, with Special Reference to Adrenodoxin.'[11]

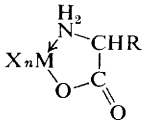
2 Amino-acids

A. Binding Sites.—Depending upon the nature of the particular metal ion, an amino-acid may bind through either, or both, of the $-NH_2$ and

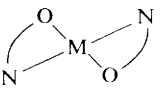
[1] R. D. Gillard, *Inorg. Chim. Acta, Rev.*, 1967, **1**, 69.
[2] R. J. P. Williams, *Roy. Inst. Chem. Rev.*, 1968, **1**, 13.
[3] B. L. Vallee and R. J.P. Willams, *Chem. in Britain*, 1968, **4**, 397.
[4] A. E. Dennard and R. J. P. Williams, *Transition Metal Chem.*, 1967, **2**, 115.
[5] H. C. Freeman, *Adv. Protein Chem.*, 1967, **22**, 257.
[6] N. N. Greenwood, *Endeavour*, 1968, **27**, 33.
[7] E. Ochiai, *Co-ordination Chem. Rev.*, 1968, **3**, 49.
[8] J. Peisach, P. Aisen, and W. E. Blumberg, 'Biochemistry of Copper,' Academic Press, New York, 1966.
[9] E. Frieden, *Scientific American*, 1968, **218**, 102.
[10] [a] J. B. Neilands, *Structure and Bonding*, 1966, **1**, 59. [b] B. B. Buchanan, *ibid.*, p.109.
[11] T. Kimura, *Structure and Bonding*, 1968, **3**, 1.

$-COO^-$ residues. Thus it has been suggested [12] that at pH 4·5–5, hexa-aquochromium(III) combines with glycine to give $[Cr(Gly)_3(H_2O)_3]$, in which the glycines are unidentate only, binding through their amino-residues. Similar unidentate attachment of glycine has been achieved to other kinetically inert metal ions, such as iridium.[13]

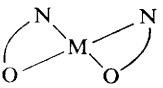
In general, however, chelation lends stability to the metal complex, so that glycinate, α-alaninate, leucinate, *etc.* form such complexes as (1). Considerable attention has focused recently on two structural points for such metal complexes: the possibility of structural isomers for such complexes as $[Cu(Gly)_2]$, and the possibility of the polar substituents on amino-acids like glutamic acid, or lysine, or histidine, acting as a metal-binding site in addition to the chelate ring shown in (1).



(1)



trans (2) *cis*



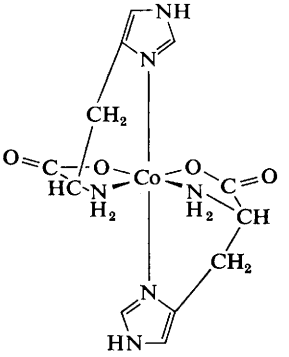
cis- and *trans*-Isomers of the type (2) have been known for many years for the non-labile M = Pt^{II} or Pd^{II}, where kinetic control ensures accessibility of both isomers. Recently it has been found, however, that even for kinetically labile metal ions such as copper(II), where the isomers of (2) are in rapid equilibrium, it is possible to isolate both geometric isomers as solids. This has been proved for $[Cu(L\text{-}Ala)_2]$ (crystal-structure determination of the *trans*-form by Dijkstra [14] and that of the *cis*-form by Gillard *et al.*[15]), and it seems likely that similar geometric isomers exist for the racemic complexes of copper(II) with glycine, DL-phenylalanine, and DL-tyrosine.[16]

It has commonly been assumed that the configuration of the kinetically labile complexes in solution is *trans*- (2); this has yet to be proved.

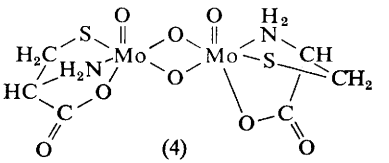
Attention has been paid to the possibility of three-point binding of amino-acids with polar substituents, both in the solid-state (by crystallographic methods) and in solution (largely by use of optical activity methods). The structure determined by Harding and Long [17] for the complex bis-L-histidinatocobalt(II) monohydrate is shown in (3); this compound is of particular interest because of its ready oxidation. A number of other terdentate attachments of amino-acids have been established; for example, the first compound of molybdenum and cysteine to be isolated [18] was formulated

[12] D. Banerjea and S. D. Chaudhuri, *J. Inorg. Nuclear Chem.*, 1968, **30**, 871.
[13] K. A. Gladyshevskaya, I. V. Prokof'eva, and O. E. Zvyaginstev, *Zhur. neorg. Khim.*, 1968, **13**, 1370.
[14] A. Dijkstra, *Acta Cryst.*, 1966, **20**, 588.
[15] R. D. Gillard, R. Mason, N. C. Payne, and G. B. Robertson, *Chem. Comm.*, 1966, 155.
[16] S. H. Laurie, *Austral. J. Chem.*, 1967, **20**, 2609; *Chem. Comm.*, 1967, 155.
[17] M. M. Harding and H. A. Long, *J. Chem. Soc.* (*A*), 1968, 2554.
[18] A. Kay and P. C. H. Mitchell, *Nature*, 1968, **219**, 267.

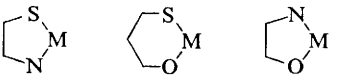
as $Na_2[Mo_2^{V}O_4(S\cdot CH_2\cdot CHNH_2\cdot CO_2)_2(H_2O)_2],3H_2O$, with the cysteinates binding through their —S and $-NH_2$ groups only. Knox and Prout subsequently showed[19] the molecule to contain terdentate cysteinate molecules (4).



(3)



(4)

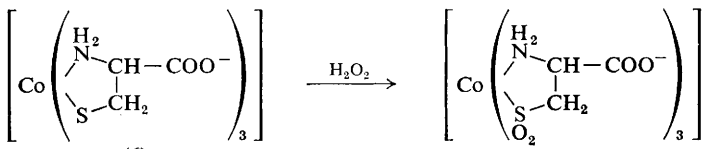


(a) (b) (c)

(5)

Cysteine may of course also be bidentate, and the three possible bidentate attachments (5a–c) are all known.

In connection with the general interest in the non-haem iron–sulphur proteins such as ferredoxin, rubredoxin, and adrenodoxin, model systems have been studied. Thus, two possible orange diastereoisomers of $[Fe(L\text{-}Cys)_2(CO)_2]$ were isolated[20] with N—S type bonding (5a); these isomers on treatment with silver ions (very strong binding preference for thiol groups) readily convert to the crimson N—O bonding form (5c), *i.e.* $[Fe(L\text{-}CysS\text{-}Ag)_2(CO)_2]$. Similarly, nickel(II) forms a red complex (5a) which converts to a green form (5c) on treatment with silver(I). By use of low working temperatures with iron(III)–cysteinate solutions, the 1 : 1, 1 : 2, and 1 : 3 complexes of the S—O bonded type (5b) and the 1 : 3 complex of the S,N form (5a) were established[21] from spectroscopic studies. The stable cobalt(III) complexes with three cysteinate ligands (6) have been studied[22] and the conversion shown in Scheme 1 established. This *in situ*



(6)

Scheme 1

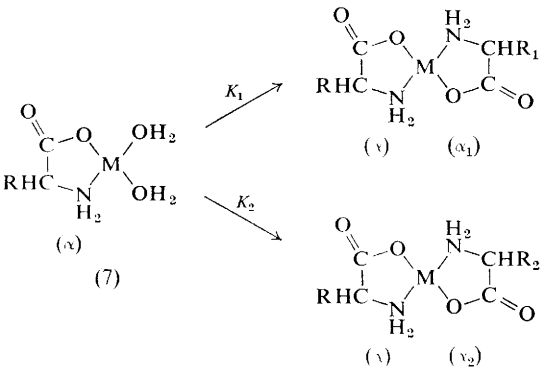
[19] J. R. Knox and C. K. Prout, *Chem. Comm.*, 1968, 1227.
[20] A. Tomita, H. Hirai, and S. Makishima, *Inorg. Nuclear Chem. Letters*, 1968, **4**, 715.
[21] A. Tomita, H. Hirai, and S. Makishima, *Inorg. Chem.*, 1967, **7**, 760.
[22] R. D. Gillard and R. Maskill, *Chem. Comm.*, 1968, 160.

conversion of a thiolate bonded to a metal ion to the corresponding sulphinic acid may prove useful elsewhere.

Hare and his group have proposed [23] a modified octant rule which enables a distinction to be made between bidentate and terdentate attachment in solution. Thus, for complexes of the type [Cu(Gly)(α)] an increasingly positive Cotton effect in the visible region with increase in alkaline pH is observed when α is an amino-acid with a polar substituent but not when α has no polar substituent; the change in c.d. spectrum was therefore taken as indicative of α functioning as a terdentate ligand.

Very recently [24] circular dichroism has been applied as a means of linking the available stereochemical information from the solid-state (usually crystallographic) with the solution structure. By comparing the c.d. spectrum of an asymmetric solid of known structure dispersed in potassium halide discs with that of the corresponding solution, it is possible to say whether a conformational change has occurred in solution. In this way it appears that, for example, Cu(L-Tyr)$_2$ has the *trans*-structure in solution.

Studies of the thermodynamics of complex formation of α-amino-acids appear to have gone through three stages: the earlier studies being concerned mainly with the determination of stability constants, these were followed by detailed calorimetric measurements to evaluate thermodynamic quotients and to correlate with stability constants from standard techniques. More recently the emphasis has been on stereospecificity and mixed ligand complex formation. If a metal ion is in a particular defined environment [*e.g.*, it already has one α-amino-acid attached to it as in (7)] then it may in principle discriminate between possible substrates (see Scheme 2).



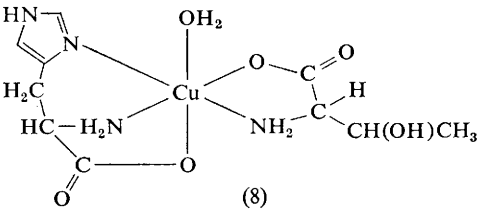
Scheme 2

Specificity entails $K_1 \neq K_2$. If α_1 and α_2 are chemically different, then the process is due to the differing stabilities of 'mixed complexes', [M(α)(α$_1$)]

[23] K. M. Wellman, T. G. Mecca, W. Mungall, and C. R. Hare, *J. Amer. Chem. Soc.*, 1968, **90**, 805.

[24] R. D. Gillard and S. H. Laurie, *Chem. Comm.*, 1969, 489.

as against [M(α)(α$_2$)]. If chemically α = α$_1$ = α$_2$ but α$_1$ and α$_2$ are enantiomers, then the competing processes are the formation of the diastereoisomers [M(L-α)(L-α)] and [M(L-α)(D-α)]. The study of the formation of mixed complexes in solution has received considerable impetus from the computer-based calculations of Perrin and co-workers.[25-27] A mixed complex [Cu(L-His)(L-Thr)] has been detected [28] in human serum and a crystalline



(8)

form of this complex has been shown [29] to have the structure (8). Stereospecificity for terdentate ligands, *e.g.* histidinate and aspartate, is known, but in the case of the bidentate ligands, such as alaninate, the complete absence of stereospecificity has been demonstrated.[30, 31] Some of the results are summarised in Table 1.

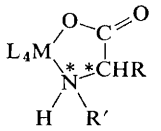
Table 1 *Stability constants of copper(II) complexes of alaninate, phenylalaninate, valinate, and prolinate from potentiometric titrations*

Ligand	*Enantiomeric form*	*log* K_1	*log* K_1K_2	*Reference*
Alaninate	L	8·16 ± 0·05	14·98 ± 0·05	*a*
	D	8·14	15·00	*a*
	DL	8·20	15·04	*a*
	L,D, or DL	8·22	15·07	*b*
Phenylalaninate	L	7·93 ± 0·05	15·10 ± 0·05	*a*
	D	8·00	15·08	*a*
	LD	8·03	15·14	*a*
Valinate	L,D, or DL	8·19	15·18	*b*
Prolinate	L,D, or DL	8·92	16·58	*b*

[a] From ref. 31; conditions: 20·0°, I = 0·37 ($NaNO_3$). [b] From ref. 30; conditions: 20·0°, I = 0·10 (NaCl).

[25] D. D. Perrin and V. S. Sharma, *J. Chem. Soc.* (*A*), 1967, 82.
[26] R. W. Hay, P. J. Morris, and D. D. Perrin, *Austral. J. Chem.*, 1968, **21**, 1073.
[27] D. D. Perrin and V. S. Sharma, *J. Chem. Soc.* (*A*), 1968, 446.
[28] B. Sarkar, M. Bersohn, Y. Wigfield, and T. C. Chiang, *Canad, J. Biochem.*, 1968, **46**, 595.
[29] H. C. Freeman, J. M. Guss, M. J. Healey, R.-P. Martin, and C. E. Nockolds, *Chem. Comm.*, 1969, 225.
[30] R. D. Gillard, H. M. Irving, R. M. Parkins, N. C. Payne, and L. D. Pettit, *J. Chem. Soc.* (*A*), 1966, 201.
[31] Vl. Simeon and O. A. Weber, *Croat. Chem. Acta*, 1966, **38**, 161.

B. Stereochemistry.—A number of examples of a phenomenon first discussed by Meisenheimer have recently been observed in complexes of *N*-substituted amino-acids, such as sarcosine. On attachment of the nitrogen to a metal ion, it becomes asymmetric (9). The attachment to the metal may be stereospecific in two cases. First, if there is a preferred conformation for the whole chelate ring then there is (even in labile systems) a favoured configuration for the nitrogen atom. Secondly, if in kinetically inert systems, by virtue of interaction with other ligands [L_4 in (9)] the



(9)

N → M link forms stereospecifically, then, since dissociation of the N group is slow and since exchange of the H (which may lead to inversion) is slow, the configuration at the N atom is largely retained.

Thus, in $[Co(NH_3)_4(Me\alpha)](NO_3)_2$, where α is L-Ala or L-Leu, Saturi and Yoshikawa [32] suggest on the basis of o.r.d., c.d., and n.m.r. evidence that the co-ordination of the *N*-methyl group is stereospecific. Similarly, the interactions between the N—Me and ethylenediamine chelate rings allow only one (stereospecific) mode of bonding of the sarcosine in the complex [33] (−)-$[Co\ en_2Sar]^{2+}$. Finally, under conditions close to neutrality, $[Co(NH_3)_4(Sar)]^{2+}$ may be optically resolved,[34] since the nitrogen atom is co-ordinated stereospecifically and there is no mechanism for its inversion.

Proline has special interest in the sense that it is atypical in structure. Its metal complexes often show similar features. Thus, the c.d. spectra of the series $[Cu(L\text{-}\alpha)_2]$ show [for the copper(II) *d–d* transitions] a net negative Cotton effect for all bidentate L-α, except for L-proline and L-hydroxyproline where the c.d. band is positive. Yasui explained [35] this apparent anomaly on the ground of the stereospecific attachment of L-proline to the metal, giving a new asymmetric centre at the N of the proline ligand.

The notion that the conformation of the ligand will affect the rotational strength of attached metal chromophores stems from Pfeiffer's use in 1937 of copper(II)-amino-acid complexes to compare the configuration of the acids. Several series of complexes have been studied from this point of view. These include $[Co^{III}(NH_3)_4\alpha]^{n+}$, studied by o.r.d. and c.d. techniques [36], where the cobalt(III) band at 490 nm. develops a pronounced

[32] M. Saturi and S. Yoshikawa, *Inorg. Chem.*, 1968, **7**, 1890.
[33] J. F. Blount, H. C. Freeman, A. M. Sargeson, and K. R. Turnbull, *Chem. Comm.*, 1967, 324.
[34] B. Halpern, A. M. Sargeson, and K. R. Turnbull, *J. Amer. Chem. Soc.*, 1966, **88**, 4630.
[35] T. Yasui, *Bull. Chem. Soc. Japan*, 1965, **38**, 1746.
[36] C. T. Lui and B. E. Douglas, *Inorg. Chem.*, 1964, **3**, 1356.

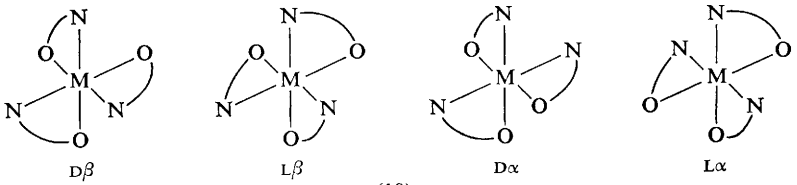
negative Cotton effect if the chelating amino-acid has the L-configuration; $[Co^{III}(NH_3)_5\alpha]^{n+}$, in which the amino-acid is attached to the cobalt ion through its carboxylate end only,[37] where L-amino-acids develop a negative Cotton effect in the band due to cobalt(III) at *ca.* 520 nm. This correlation has been shown[38] to be reliable only in the absence of ion-pairing. The latter series has also been studied kinetically, and the hydrolysis:

$$[Co(NH_3)_5(H\alpha)]^{3+} + H_2O \longrightarrow [Co(NH_3)_5(H_2O)] + H\alpha$$

follows the rate-law:

$$\text{Rate} = [\text{complex}]\,(k_{H_2O} + k_{H^+}[H^+])$$

A second kind of stereospecificity has already been discussed for bis-chelate complexes of bidentate amino-acids. There has been a good deal of work on what might be called 'operational stereospecificity' in complexes containing three bidentate amino-acids disposed about an octahedral metal ion such as cobalt(III). The general finding is that of the four possible geometric optical isomers (10) it is usually possible to isolate all forms.



(10)

This has been achieved for[39] Gly, L-Ala, L-Leu, L-Val, and L-Asp. There appear to be no major differences in relative stabilities in water, regardless of the particular amino-acids, unless some special stereochemical factor intervenes. Thus, with L-proline, Piper and Denning[40] were unable to isolate the Dα-isomer; the preparative non-accessibility of this isomer is predictable on steric grounds.

C. Reactivity.—In an attempt to evaluate the possibility that deprotonation of the amide group is significant for peptide and protein complexes of copper(II) within the physiological pH range, Sigel[41] studied deprotonation in a ternary complex of copper(II) resembling an enzyme–metal ion–substrate complex. He chose 2,2′-bipyridyl and glycineamide (HL). On titration of a 1 : 1 molar ratio of Cu^{2+} : HL,HCl, three protons were liberated, from the protonated amide group, from the neutral amide group, and from the co-ordinated water molecule, $[Cu{-}OH_2 \rightarrow Cu{-}OH + H^+]$.

[37] J. H. Dunlop and R. D. Gillard, *Tetrahedron*, 1967, **23**, 349.
[38] C. J. Hawkins and P. J. Lawson, *Chem. Comm.*, 1968, 177.
[39] R. D. Gillard, *Inorg. Chim. Acta Rev.*, 1967, **1**, 69, and references therein.
[40] R. G. Denning and T. S. Piper, *Inorg. Chem.*, 1966, **5**, 1056.
[41] H. Sigel, *Angew. Chem. Internat. Edn.*, 1968, **7**, 137.

In the presence of an equimolar amount of bipyridyl, only two protons were liberated (from $-NH_3^+$ and then from the amide). For the second proton release:

$$\begin{array}{rcll} Cu(HL) & \rightleftharpoons & CuL + H^+ & pK = 7{\cdot}01 \\ Cu\ bipy\ (HL) & \rightleftharpoons & Cu\ bipy\ L + H^+ & pK = 7{\cdot}71 \end{array}$$

The acidity of the amide thus decreases on going from the binary to the ternary complex, but the group still deprotonates in the physiological pH range. Under more extreme conditions (potassium amide in liquid ammonia), solid derivatives of glycinate and similar complexes of cobalt(III) may be obtained [42] in which the amino-group has been deprotonated. Attempts to methylate the deprotonated complexes were unsuccessful.

Kinetic measurements on the combination of amino-acids with nickel(II) [43] and copper(II) [44] showed that only free anion actually combines with the metal; the zwitterionic species is unreactive. In the case of copper(II) and glycinate, the rate-determining step appears to be the release of the water bound to the copper ion.

The promotion of the rate of hydrolysis of amino-acid esters by metal ions is well known, and studies continue on the effect of varying other ligands on the metal. For example, the equilibrium constants for the reaction

$$[M(NTA)]^- + Gly \rightleftharpoons [M(NTA)(Gly)]^-$$

where NTA is nitrilotriacetate, follow the [45a] Irving-Williams order $Mn < Co < Ni < Cu > Zn$. In a second paper, the rates for the metal-complex-promoted hydrolyses for a series of ethyl esters of α-amino-acids were found [45b] to be *ca.* 200 times faster than the rates in the absence of metal complex. It was proposed that the promoted hydrolysis occurs by activation (through chelation) of the carbonyl carbon atom to intermolecular attack by hydroxide ion. Similar conclusions were reached [46] for the mechanism of the metal-promoted hydrolysis of ethyl valinate-*N*,*N*-diacetic acid and related compounds.

Hay and Morris [47] studied the base hydrolysis of methyl 2,3-diaminopropionate and its metal complexes and were able to account for their results on the grounds that the reactivity of a species towards hydroxide ion depends largely on the total positive charge carried by the species. Hix and Jones [48] examined the possibility of attaining kinetic stereoselectivity in ester hydrolysis. They found that *R*(−)-histidine methyl ester is hydrolysed

[42] G. Watt and J. F. Knifton, *Inorg. Chem.*, 1968, **7**, 1159.
[43] J. C. Cassatt, and R. G. Wilkins, *J. Amer. Chem. Soc.*, 1968, **90**, 6045.
[44] A. F. Pearlmutter and J. Stuehr, *J. Amer. Chem. Soc.*, 1968, **90**, 858.
[45] [a] D. Hopgood and R. J. Angelici, *J. Amer. Chem. Soc.*, 1968, **90**, 2508. [b] R. J. Angelici and D. Hopgood, *J. Amer. Chem. Soc.*, 1968, **90**, 2514.
[46] [a] R. J. Angelici and B. E. Leach, *J. Amer. Chem. Soc.*, 1968, **90**, 2499. [b] B. E. Leach and R. J. Angelici, *J. Amer. Chem. Soc.*, 1968, **90**, 2504.
[47] R. W. Hay and P. J. Morris, *Chem. Comm.*, 1968, 732.
[48] J. E. Hix and M. M. Jones, *J. Amer. Chem. Soc.*, 1968, **90**, 1723.

40% faster in the presence of the 1 : 1 complex [Ni{*S*(+)-histidinate}]$^+$ than in the presence of its enantiomer [Ni{*R*(−)-histidinate}]$^+$. The results were interpreted as arising from differences in the specific rates of hydrolysis for the co-ordinated ligands, rather than from stereoselectivity in complex formation. However, in the case of these potentially terdentate amino-acids, further complications are possible in that geometrical environments of the metal may differ in the diastereoisomers [M(*R*)(*S*)] and [M(*R*)$_2$].

The hydrolysis of the complexed glycine ester in the compound [Co en$_2$(Gly-OC$_3$H$_7$)](ClO$_4$)$_3$ has been studied in detail [49] using oxygen-18, and it was shown that the hydrolysis occurs without opening the glycine ester chelate ring.

The autocatalytic oxygenation of several amino-acids in the presence of metal ions is known, and Gorton and Jameson [50] have studied the copper(II) and iron(III) complexes of L-β-(3,4-dihydroxyphenyl)alanine (DOPA) using equilibrium methods in solution. Small amounts of copper(II) inhibit oxygenation of DOPA because the initial complex formed involved interaction with the amino-acid function. Iron(III), on the other hand, is an active catalyst, possibly because it preferentially chelates with the catechol function of DOPA. The kinetics of the rapid uptake of molecular oxygen by the cobalt(II)–L-histidine system have been interpreted [51] as arising from the stepwise process

$$\text{Co(L-His)}_2 + \text{O}_2 \rightleftharpoons \text{Co(L-His)}_2\text{O}_2 \rightleftharpoons [\text{Co(L-His)}_2]\text{O}_2$$

The catalytic autoxidation of ferrocytochrome *c* has been studied [52] in the presence of copper(II) with two histidines (whether L or DL was not stated). The kinetic results are consistent with the simple mechanism:

$$\begin{array}{ccc} \text{ferrocytochrome } c + [\text{Cu(His)}_2] & \rightleftharpoons & [\text{cytochrome } c\ \text{Cu(His)}_2] \\ & & \Big\downarrow \text{H}^+, \text{O}_2 \\ & & \text{ferricytochrome } c + [\text{Cu(His)}_2] \end{array}$$

The radiation chemistry (as solids) of several metal complexes of glycine and alanine was found to be similar to that of the free amino-acids though a slight increase in stability to radiation resulted from chelation.[53]

Stereoselective interactions of complexed amino-acids have received attention. Harada [54] achieved a partial resolution of [Cu(D-Asp)(L-Asp)]

[49] D. A. Buckingham, D. M. Foster, and A. M. Sargeson, *J. Amer. Chem. Soc.*, 1968, **90**, 6032.
[50] J. E. Gorton and R. F. Jameson, *J. Chem. Soc.* (*A*), 1968, 2615.
[51] J. Simplicio and R. G. Wilkins, *J. Amer. Chem. Soc.*, 1967, **89**, 6092.
[52] A. J. Davison, *J. Biol. Chem.*, 1968, **243**, 6064.
[53] W. C. Gotschall and B. M. Tolbert, *J. Phys. Chem.*, 1968, **72**, 922.
[54] K. Harada, *Nature*, 1968, **218**, 199.

by seeding its supersaturated solution with wool or cotton. A partial resolution of DL-alanine was obtained [55] from the reaction of triscarbonatocobaltate(III) with L-aspartic acid, to give (−)-$[Co(L\text{-}Asp)_2CO_3]^{3-}$, which on treatment with DL-alanine gave insoluble *cis*-(−)-$[Co(Ala)_3]$. This contained D-alanine of 38% optical purity. Another stereoselective interaction of a metal complex with amino-acids involves the resolution of the complex anion [Co edta]$^-$ using the L-histidinium cation. The resolved (−)-[Co edta]$^-$ then can be used to resolve DL-lysine.[56]

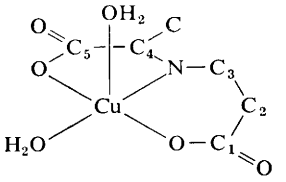
D. Complexes of Schiff Bases.—Pfeiffer discovered thirty years ago that some reactions (notably ester exchange, racemisation, and oxygenation) of α-amino-acid esters are effected very readily when their Schiff bases with salicylaldehyde are complexed to a transition-metal ion (notably copper). The analogy has often been drawn between the pyridoxal phosphate-promoted enzymic reactions of α-amino-acid metabolism and reactions in the model systems involving metal ion, α-amino-acid, and an aromatic hydroxy-aldehyde (such as 4-nitrosalicylaldehyde or pyridoxal). A good deal of work is now beginning to appear on these related activations of amino-acid derivatives.

Work on the species present in solutions containing salicylaldehyde, glycinate, and divalent metal ions (Mn, Ni, Cu, and Zn) showed [57] that species of the type $[M(sal)(Gly)_2]$, $[M(sal)(Gly)_2]^-$, and $[M(sal)_2(Gly)_2]^{2-}$ were present. Copper(II) forms only [Cu(sal)(Gly)]; the enhanced stability of the 1 : 1 : 1 complexes is attributed to their existence as Schiff bases, and u.v. spectra ($\pi \rightarrow \pi^*$ transitions) were thought to confirm this. Comparison of the results for salicylaldehyde Schiff bases with those [58] for pyridoxal Schiff bases shows that the latter form less stable complexes. Spectrophotometric studies of chelate formation in mixtures of pyridoxal with amino-acids (or peptides) have also been reported [59] for aluminium, gallium, and copper(II) ions.

The crystal structures of several compounds isolated from solutions of these Schiff base complexes have appeared. (±)-Phenylalanine-(pyridoxylidene-5-phosphate)copper(II) contains [60] penta-co-ordinated copper(II). Three in-plane donors are the N(imine), O(phenol), and the O(carboxyl) of the pyridoxylidene group. The fourth ligand is a water molecule. The square pyramid is completed by a phosphate oxygen of a neighbouring molecule (Cu—O = 2·31 Å). The α-amino-acid and pyridoxal co-factor are almost coplanar and the metal ion may act by maintaining a conjugated system suitable for facilitated reactions. In this structure, it is likely that the pyridoxal is present in a protonated form (at the ring nitrogen).

[55] M. Shibata, H. Nishikawa, and K. Hosaka, *Bull. Chem. Soc. Japan*, 1968, **41**, 130.
[56] R. D. Gillard, P. R. Mitchell, and H. C. Roberts, *Nature*, 1968, **217**, 949.
[57] D. L. Leussing and K. S. Bai, *Analyt. Chem.*, 1968, **40**, 575.
[58] D. L. Leussing and N. Huq, *Analyt. Chem.*, 1966, **38**, 1388.
[59] C. Cennamo, *Ital. J. Biochem.*, 1967, **16**, 179.
[60] G. A. Boutley, J. M. Waters, and T. N. Waters, *Chem. Comm.*, 1968, 988.

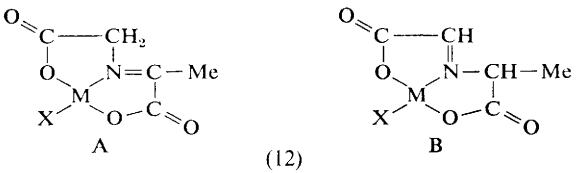
Following an earlier study [61a] of the structure of *N*-salicylideneglycinato-aquocopper(II) hemihydrate, the same group have presented their results on pyruvidene-β-alaninatoaquocopper(II) dihydrate.[61b] This was studied because, whereas solutions of copper(II), α-alanine, and pyruvate give transamination, similar solutions where β-alanine replaces the α-isomer do not transaminate. The general features of the structure are shown in (11).



(11)

The bond C^4—N is abnormally short (1·24$_7$ Å); the sum of double-bonded radii C=N is 1·29 Å. The bond N—C^3 is normal at 1·48 Å. A transamination is expected if the 'single' bond N—C^3shows some double-bond character; this shortening was, as expected, found in the *N*-salicylideneglycinatocopper(II) complex, where transamination is well established.

The possibility of proton transfers occurring in these transaminations has been studied.[62] The copper(II) complexes made from pyruvic acid and glycine, and from glyoxylic acid and α-alanine, are not the tautomers (12)A and (12)B, but are both (12)A. In the reported work, palladium(II) com-

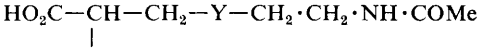


A B

(12)

plexes were made, and again both methods led to the structure (12)A. In heavy water, only a single methyl resonance was found, so that the methylene protons had completely exchanged.

In related work, the effect of γ-substitution in the amino-acid upon the rate of their copper(II)-catalysed transaminations with glyoxylic acid is reported.[63] For the amino-acids of formula (13) the rates of transamination



(13)

[61] [a] T. Ueki, T. Ashida, Y. Sasada, and M. Kakudo, *Acta Cryst.*, 1967, **B22**, 870. [b] *Ibid.*, 1968, **B24**, 1361.

[62] H. Yoneda, Y. Morimoto, Y. Nakao, and A. Nakahara, *Bull Chem. Soc. Japan*, 1968, **41**, 255.

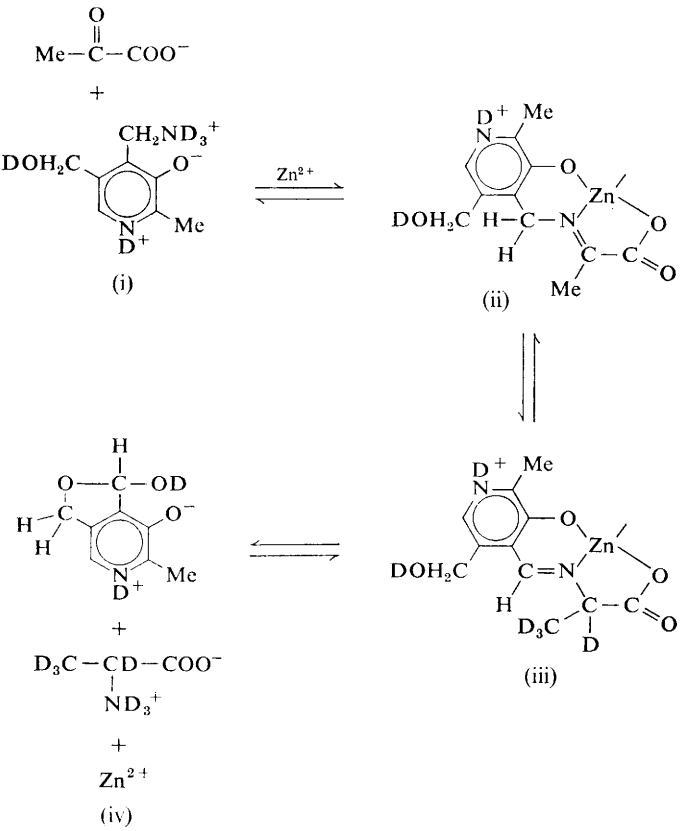
[63] P. Hermann and I. Willhardt, *Z. physiol. Chem.*, 1968, **349**, 395.

increased in the order $Y = CH_2 < S < (+)\text{-}SO < (-)\text{-}SO < SO_2$. This increase of rate with electronegative substitution was attributed to an increased lability of the α-proton, favouring a prototropic rearrangement within the copper(II) complex of the Schiff base.

Holm has continued his careful study (using n.m.r. methods) of these Schiff base reactions. With the aim of detecting and identifying the species involved in the reaction:

$$\text{pyridoxamine} + \text{pyruvate} \rightleftharpoons \text{pyridoxal} + \text{alanine}$$

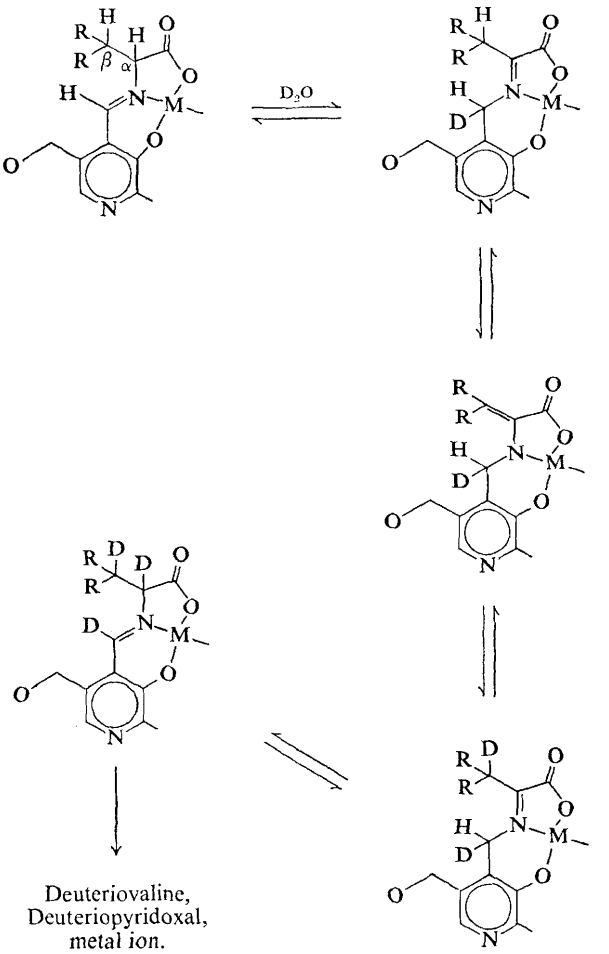
the following systems were examined;[64] (*i*) 0·1M-pyridoxal, 0·1M-alanine, 0·05M-Zn^{II}; (*ii*) as for (*i*) but without Zn^{II}; (*iii*) 0·1M-pyridoxamine, 0·1M-pyruvate, 0·05M-Zn^{II}; (*iv*) as for (*iii*) but without Zn^{II}. All n.m.r. signals were assigned, and the mechanism proposed on the basis of the observed exchanges with solvent is shown in Scheme 3. It is noteworthy that



Scheme 3

[64] O. A. Gansow and R. H. Holm, *J. Amer. Chem. Soc.*, 1968, **90**, 5629.

exchange of the methyl protons of the pyruvate with solvent is rapid, and a closely related observation is that of Abbott and Martell.[65] They found that in a solution containing aluminium ions, pyridoxal, and valine, at pD 5·9, kept on the steam-bath for 4 days, the β-protons of valine had exchanged with solvent. A similar but faster exchange of β-protons was found with α-aminobutyric acid. Here, copper(II) and zinc(II) ions were also promoters of exchange. The proposed mechanism for the α and the novel β exchange is shown in Scheme 4. Since many vitamin B_6-catalysed reactions occur at



Scheme 4

[65] E. H. Abbott and A. E. Martell, *Chem. Comm.*, 1968, 1501.

the β-position, the imine–enamine tautomerism should represent an important step for extending the catalytic effect of vitamin B_6 into the amino-acid backbone. One apparent contradiction between the mechanisms in Schemes 3 and 4, which are in fact very similar, is the noted lack of proton exchange at the azomethine carbon in the former mechanism while apparently this readily occurs (as might be expected) in the latter mechanism.

Synthetic applications of these Schiff base complexes have been suggested. Following the copper(II)-promoted condensation of acetaldehyde with glycinate to give threonine isomers, the Schiff base complex of glycine has been employed.[66] *N*-Salicylideneglycinatoaquocopper(II), preferably in dimethylformamide, in the presence of potassium hydroxide is treated with various alkyl halides, to give new amino-acids. The results are given in Table 2.

Table 2 (ref. 66)

Halide	*Amino-acid obtained*	*Ratio of derived amino-acid to unreacted glycine*
MeI	α-Alanine	1·13
$Me_2CH \cdot Br$	Valine	0·11
$Me_2CH \cdot CH_2I$	Leucine	0·09
$Me(Me \cdot CH_2)CH \cdot Br$	Isoleucine	Trace
$Ph \cdot CH_2Br$	Phenylalanine	0·75
$Br \cdot CH_2 \cdot CO_2H$	Aspartic acid	0·41

Houghton has extended his earlier findings [67] on the synthetic utility of magnesium derivatives of these Schiff bases. The reaction of benzylamine and magnesium methoxide with the Schiff bases formed from glycine ethyl ester and salicylaldehyde, 2-hydroxy-1-naphthaldehyde, or acetylacetone results [68] in *N*-benzyl-amide formation.

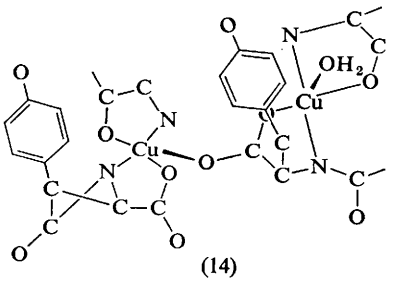
3 Peptides

A. Binding Sites.—A very large number of suggestions are still current in the literature on possible binding sites of metals to peptides and on non-specific metal–protein binding. Work continues on simple peptides, and the current position on copper–peptide structures resulting from the work of his school has been most admirably summarised by Freeman.[69] His general conclusions are: (*i*) there is little change in dimensions of the peptide after co-ordination to a metal ion; the amide group also remains essentially planar; (*ii*) protonated N(peptide) atoms are both energetically and geometrically unfavourable for binding to a metal ion, and hence are

[66] A. Nakahara, S. Nishikawa, and J. Mitani, *Bull. Chem. Soc. Japan*, 1967, **40**, 2212.
[67] R. P. Houghton, *J. Chem. Soc.* (*C*), 1967, 2030.
[68] K. Blažević, R. P. Houghton, and C. S. Williams, *J. Chem. Soc.* (*C*), 1968, 1704.
[69] H. C. Freeman, *Adv. Protein Chem.*, 1967, **22**, 258.

not found as metal binding sites; (*iii*) when a metal ion is bound to three donor groups, including the N(peptide), the three donor atoms lie in the same plane; and (*iv*) metal binding by peptide molecules is pH-dependent for Cu, Co, and Ni, but not for Zn. In acidic solutions the donor atoms are the N(amino) and O(peptide) atoms closest to it, the terminal COO group then binds to a different metal ion. As the pH is raised the N(peptide) displaces the O(peptide) as the donor atom for Cu, Co and Ni, with a corresponding dissociation of a peptide proton. With di- and tri-peptides, at the higher pH values, the COO may also participate in chelation. With larger peptides N(peptide) atoms displace the COO groups.

Other work on copper–peptide interactions has been directed to rather specific questions. For example, the crystal structure of a blue compound of formula $Cu_2(C_{17}H_{25}N_3O_5)_2,8H_2O,(C_2H_5)_2O$, made by treating copper sulphate with barium hydroxide and Gly-L-Leu-L-Tyr, shows [70] the unusual feature that the aromatic ring of the tyrosine moiety lies below the copper(II) ion as shown (14) with the plane of the benzene ring parallel to the chromophoric plane containing the Cu^{II} ion. The distance of closest approach is



(14)

3·17 Å so that an interaction between copper and tyrosine cannot be ruled out. A similar type of environment is envisaged [16] for the Cu^{II} ion in [Cu(D-Tyr)(L-Tyr)]. The possible relevance of this type of environment to the mode of action of the enzyme tyrosinase is noteworthy.

Following an earlier suggestion that copper(II) ions may form a bridge between the terminal NH_2 of a protein polypeptide chain and an imidazole group in a side-chain, and the finding that for the equilibrium

$$[Cu(Gly\text{-}Gly)] + Im \rightleftharpoons [Cu(Gly\text{-}Gly)Im]$$

(where Im is imidazole) log $K = 3{\cdot}85$; the recent work on solid, mixed copper(II) complexes of imidazole and glycylglycinate is of interest.[71] Denoting $[Gly\text{-}Gly]^2$ as A, the following compounds were isolated: $CuA,3H_2O$, $Cu(HA)(Im)_2,ClO_4,H_2O$, $Cu_2A_2Im,8H_2O$, and $CuAIm,H_2O$.

[70] D. van der Helm and W. A. Franks, *J. Amer. Chem. Soc.*, 1968, **90**, 5627.
[71] R. Driver and W. R. Walker, *Austral. J. Chem.*, 1968, **21**, 671.

The first of these was used to establish log $K = 3{\cdot}8 \pm 0{\cdot}3$ for the above equilibrium. The crystal structure of the perchlorate salt is described [72] and shows that the peptide is attached to the copper(II) through the N(amino) and the O(peptide) atoms; the square planar structure is completed by the two imidazole residues.

Interest in the interaction in solution between copper ions and peptides continues. At this point, it is perhaps well to comment on the possible distinctions between the structures in the solid-state and those in solution. While the assumption may be true that for the large protein molecules there is unlikely to be any very major change of shape on passing from solution to solid, this is not the case for molecules as small as these peptide complexes. The crystallographic studies are of the utmost value in delineating modes of attachment and suggesting conformations for the complexed peptides, but unless the metal complex is kinetically inert, there is no reason to suppose that structures in the solid-state necessarily also refer to solutions.

In some painstaking work using computer refinement, e.m.f. measurements [73] (at 25°, in 3·0M-$NaClO_4$) on copper complexes of glycylhistidinylglycine(L) (pK_a 3·57, 7·54, and 8·62), suggest the predominant species to be CuL_2, Cu_3L_4, Cu_5L_6, and $Cu_{15}L_{16}$. The authors point out that it is tempting to regard the species as fragments of an infinite chain . . . -pep-Cu-pep-Cu-pep-Cu- . . . . These chains are known to occur in the solid-state for complexes with glycylglycylglycine and with glycylhistidine.[69] Similar measurements [74] on the analogous glycylglycylglycine system suggested that at low concentrations, monomers [$CuHqA^{(1+q)+}$, $q = 1$, 0, -1, and -2] predominated. Above 10mM, evidence for possible dimers was obtained.

The free energies (from potentiometric stability constants) and enthalpies (by direct calorimetry) have been measured [75] for the formation of copper(II) complexes with Gly-Gly and with Gly-Gly-Gly. The results for the reactions (involving the singly charged anions of the amino-acid or peptide) are collected in Table 3. The relatively unfavourable free energy for the

Table 3 *Thermodynamic data for copper*(II) *complex formation with glycinate* (G^-), *glycylglycinate* (GG^-), *and glycylglycylglycinate* (GGG^-), 25°, $I = 0{\cdot}10$M. (Ref. 75)

Equilibrium	ΔG (*kcal. mole*$^{-1}$)	ΔH (*kcal. mole*$^{-1}$)	ΔS (*cal. mole*$^{-1}$ *deg.*$^{-1}$)
$Cu^{2+} + G^- \rightleftharpoons CuG^+$	−11·71	−6·76	+16·6
$CuG^+ + G^- \rightleftharpoons CuG_2$	−9·47	−6·89	+8·7
$Cu^{2+} + GG^- \rightleftharpoons CuGG^+$	−7·58	−6·1	+4·97
$Cu^{2+} + GGG^- \rightleftharpoons CuGGG^+$	−6·87	−6·3	+1·9

[72] H. C. Freeman and J. D. Bell, personal communication, quoted in ref. 69.
[73] R. Österberg and Bo Sjöberg, *Acta Chem. Scand.*, 1968, **22**, 689.
[74] R. Österberg and Bo Sjöberg, *J. Biol. Chem.*, 1968, **243**, 3058.
[75] A. P. Brunetti, M. C. Lim, and G. H. Nancollas, *J. Amer. Chem. Soc.*, 1968, **90**, 5120.

peptides is attributed to the entropy term. Binding is thought to be from the N(amino) and the O(peptide) of the first peptide linkage in solutions of pH < 7.

Kinetic studies (using the temperature-jump method) have been made of the combination of carnosine with copper(II) ions [76] and with cobalt(II) ions.[77] The rates of proton transfer are considerably slower than the

$$\begin{array}{ccc} CuHL^{2+} & \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} & CuL^{+}+H^{+} \\ & & \uparrow\downarrow \\ & & k_{-2}\ \ k_2 \\ & & \downarrow\uparrow \\ & & Cu(L-H)+H^{+} \end{array}$$

$k_1 = 230$ sec.$^{-1}$, $k_{-1} = 1{\cdot}0 \times 10^7$M^{-1} sec.$^{-1}$, $k_2 = 360$ sec.$^{-1}$, $k_{-2} = 5{\cdot}0 \times 10^7$M^{-1} sec.$^{-1}$.

diffusion-controlled limit, which is attributed to intramolecular changes concurrent with proton attack. Similarly, Pagenkopf and Margerum [78] have studied the rates of proton transfers of a peptide hydrogen in the copper(II)–glycylglycylglycine system. These rates are much slower than for 'normal' acid–base reactions, and this is attributed to the need for co-ordinate bonds to the copper(II) ion to break and rearrange.

Emery [79] has observed a similar slow exchange of 2–4 protons (depending upon pH), using tritium–hydrogen exchange, in ferrichrome and ferrichrome A, both iron complexes of cyclic hexapeptides. In this case, by use of the known (solid-state) conformation of ferrichrome A, two of the slowly exchanging hydrogens were assigned to the buried ornithine groups and the other two (whose presence depends on pH) to ring H-bonds [N(amides) of Orn_3 and Ser_2]. Similar observations were made with ferrioxamine B. Interestingly, the author does not believe that a conformational change need be invoked to explain these slow exchanges.

Sporadic reports concern peptide complexes with metal ions other than copper(II). Crystal structures of four complexes of nickel(II) with glycine and glycine peptides have been reported.[80] In $[Ni(Gly)_2(H_2O)_2]$, the nickel ion is surrounded by two (*trans*) glycinate chelates in one plane with the two water molecules completing the octahedron. At high pH, the yellow complexes $Na_2Ni(Gly\text{-}Gly)_2{,}8H_2O$ and $Na_2Ni(Gly\text{-}Gly)_2{,}9H_2O$ may be obtained. These complexes have the same structure which is similar to that found [81] in $NH_4[Co^{III}(Gly\text{-}Gly)_2]{,}2H_2O$, *i.e.* co-ordination of the peptides

[76] R. F. Pasternak and K. Kustin, *J. Amer. Chem. Soc.*, 1968, **90**, 2295.
[77] K. Kustin and R. F. Pasternak, *J. Amer. Chem. Soc.*, 1968, **90**, 2805.
[78] G. K. Pagenkopf and D. W. Margerum, *J. Amer. Chem. Soc.*, 1968, **90**, 6963.
[79] T. F. Emery, *Biochemistry*, 1967, **6**, 3858.
[80] H. C. Freeman, J. M. Guss, and R. L. Sinclair, *Chem. Comm.*, 1968, 485.
[81] R. D. Gillard, E. C. McKenzie, R. Mason, and G. B. Robertson, *Nature*, 1966, **209**, 1347.

occurs *via* the N(amino), N(peptide), and O(carboxyl) atoms. Finally, the crystals of $Na_2Ni(Gly\text{-}Gly\text{-}Gly),8H_2O$ obtained from a solution at high pH are isomorphous with a known Gly-Gly-Gly complex of copper(II). The nickel ion is then in a planar environment with the peptide attached through its N(amino) and both N(peptide) atoms. The bonds from metal to peptide in this four-co-ordinated nickel complex are significantly shorter than those in similar octahedral (six-co-ordinate) compounds, illustrating one possible effect of varying metal co-ordination number in metallo-enzymes. The nickel(II)–glycylglycinate system has also been studied [82] in solution and equilibrium constants obtained at 10°, 28°, and 30°.

B. Stereochemistry.—It has been known for some time that diastereoisomeric dipeptides form complexes with metals of differing stability. Some further stability constants have appeared [83] during the year and are summarised in Table 4. Under the conditions employed, *i.e.* pH 4·5, copper(II) forms only a 1 : 1 complex, while zinc(II) forms both 1 : 1 and 1 : 2 complexes, in keeping with their pH dependence mentioned earlier.

Table 4 *Stability constants from potentiometric measurements at* 20°, pH 4·5, $I = 0{\cdot}1$M (Ref. 83)

Ligand	Cu^{II}	Zn^{II}		pK_{a1}	pK_{2a}
	β_1	β_1	β_2		
L-Leu-L-Ala	4×10^3	$2{\cdot}8 \times 10^3$	$8{\cdot}4 \times 10^3$	3·20	8·05
L-Leu-D-Ala	$1{\cdot}15 \times 10^3$	$8{\cdot}5 \times 10^2$	$3{\cdot}5 \times 10^3$	3·10	8·20

Related to this is the finding, in work [84] on the structure of albomycin, that the thermodynamic stabilities of iron complexes of certain tri- and hexa-peptides depended on both the sequence and configuration of the polyfunctional amino-acids. Of particular interest are the observations that the peptides and their iron complexes showed specific activity against certain micro-organisms.

In many cases, when an asymmetric peptide is co-ordinated to a metal ion, the metal serves as a new centre of asymmetry. Thus, the complexes of cobalt(III) with dipeptides of the general structures (15) may be described as D-$[Co(\alpha_1\alpha_2)_2]^-$ and its enantiomer (diastereoisomer if one of the amino-acids is asymmetric) L-$[Co(\alpha_1\alpha_2)_2]^-$. One remarkable recent finding is that, if the dipeptide contains two L-amino-acid residues or one L-amino-acid plus glycine, then only one diastereoisomer of the cobalt(III) complex is produced.[85] This is most probably the L-isomer, since it has a very large

[82] S. Morazzani-Pelletier, *Compt. rend.*, 1968, 267, *C*, 925.
[83] F. Karcyzynski and G. Kupryszewski, *Roczniki. Chem.*, 1967, **41**, 1665.
[84] N. A. Poddubnaya, A. M. el Naggar, I. N. Skvortsova, and G. N. Balandina, *Zhur. obshchei Khim.*, 1968, **38**, 732.
[85] R. D. Gillard and A. Spencer, *Disc. Faraday Soc.*, 1968, **46**, 213.

(15)

negative circular dichroism for the cobalt(III) visible absorption band. It seems likely that this stereoselectivity is thermodynamically controlled, since any one of the six distinct preparative routes gives the same result.

C. Reactivity.—*Hydrolysis and Synthesis.* Two remarkable hydrolyses of peptides have been noticed as a result of preparative work for *X*-ray crystallographic studies. First, in an attempt to obtain crystals of the hexaglycine complex of copper(II), in molar base solution, the pentaglycine complex of copper(II) formed together with free glycine.[86] Nickel(II) acted in the same manner. Over a longer period, further, less specific, hydrolysis occurs. Solution studies suggest that the peptide is co-ordinated to the copper ion *via* nitrogen groups, so that the hydrolytic cleavage probably occurred at the carboxylate end.

Second, the compound made by treating aqueous diketopiperazine and potassium hydroxide in the molar ratio 1 : 2 with excess of 'copper hydroxide', and originally formulated (naturally enough) as a complex of diketopiperazine, turned out on elemental and *X*-ray analyses to contain two glycylglycinate molecules per copper ion, *i.e.* $K_1[Cu(H_2NCH_2 \cdot CO \cdot NCH_2 \cdot COO)_2(H_2O)_2],4H_2O$, the glycylglycine having been formed by hydrolysis.[87] Co-ordination of the dipeptide is, as expected, through the N(amino) and N(peptide) atoms.

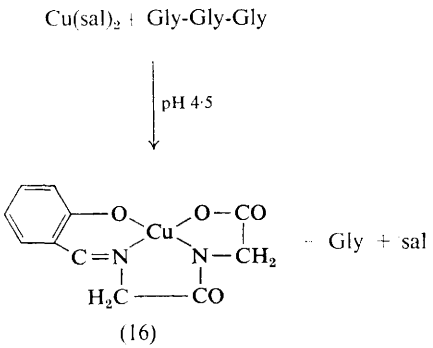
A further metal-promoted hydrolysis was noted [88a] during a study of the formation of Schiff-base complexes of copper(II) with Gly, Gly-Gly, and Gly-Gly-Gly. Gly-Gly reacted with bis-salicylaldehydatocopper(II) and bisbenzoylacetaldehydocopper(II) to form the glycylglycinato-Schiff-base copper(II) complexes. With Gly-Gly-Gly, however, at pH 4·5, the product with the salicylaldehyde complex was the copper complex of the Schiff base of glycylglycine and free glycine [(16), Scheme 5]. Later work [88b] indicates that the hydrolytic cleavage occurred at the carboxyl end, although not with 100% specificity. Quite clearly this type of reaction may well prove useful, particularly if the specificity can be improved.

In connection with hydrolytic metal-promoted cleavage of peptides, the best known reagent is undoubtedly that discovered by Collman and

[86] R. H. Andreatta, H. C. Freeman, A. V. Robertson, and R. L. Sinclair, *Chem. Comm.*, 1967, 203.

[87] A. Sugihara, T. Ashida, Y. Sasada, and M. Kakudo, *Acta Cryst.*,1968, **B24**, 203.

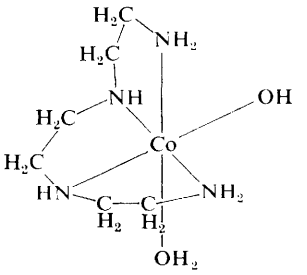
[88] [a] A. Nakahara, K. Hamada, I. Miyachi, and K. Sakurai, *Bull. Chem. Soc. Japan*, 1967, **40**, 2826. [b] A. Nakahara, K. Hamada, Y. Nakao, and T. Higashiyama, *Co-ordination Chem. Rev.*, 1968, **3**, 207.



(16)

Scheme 5

Buckingham,[89] the hydroxoaquotriethylenetetraminecobalt(III) ion. An extensive programme of most careful and elegant work has clarified and extended the original observations. Of the various possible isomers of [Co trien $(H_2O)(OH)]^{2+}$, trien = triethylenetetramine, it is the β-isomer (17) which is the more reactive towards peptide hydrolysis. This combines



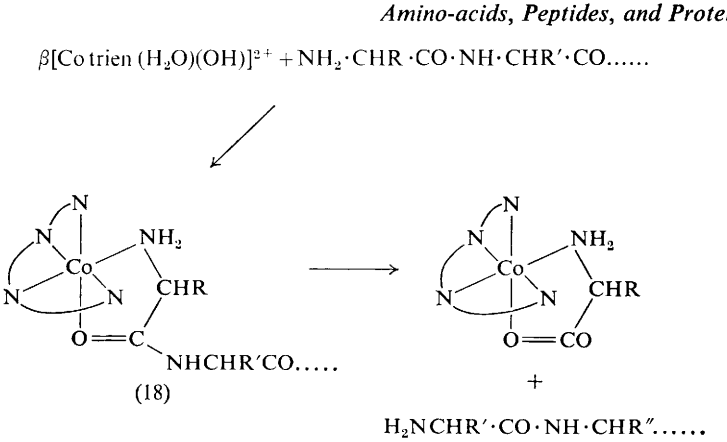
(17)

rapidly with the *N*-terminal amino-acid residue of the peptide or peptide ester, which is then slowly hydrolysed (pH *ca.* 7, 60–65°, 30–180 min.), possibly as shown in Scheme 6, leading to a specific *N*-terminal hydrolysis.

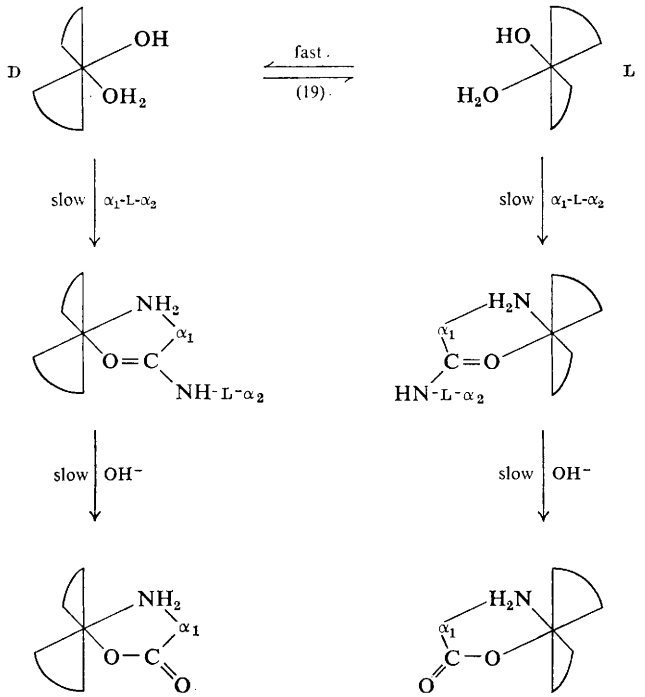
More recently, the glycylglycine-*O*-ester intermediate of Scheme 6 has been isolated and the structure shown (18) confirmed by *X*-ray analysis.[90] It seems likely that the hydrolysis of the peptide then proceeds without rupture of the chelate ring.

Some related observations [91] are rather unexpected. The system studied involved the *cis*-hydroxoaquobisethylenediaminecobalt(III) ion (19) which

[89] J. P. Collman and D. A. Buckingham, *J. Amer. Chem. Soc.*, 1963, **85**, 3039.
[90] D. A. Buckingham, P. A. Marzilli, I. E. Maxwell, A. M. Sargeson, H. Fehlmann, and H. C. Freeman, *Chem. Comm.*, 1968, 488.
[91] D. E. Allen and R. D. Gillard, *Chem. Comm.*, 1967, 1091.



Scheme 6



Scheme 7

racemises very rapidly at room temperature in the sequence of reactions shown in Scheme 7; it was found that, irrespective of the nature of α_1 (whether it was D-, L-, or Gly) provided that the *C*-terminal α_2 had the L-configuration, then the product always contained more D-$[Co\ en_2\alpha_1]^{n+}$

than L-[Co $en_2\alpha_1]^{n+}$. The stereochemical control in this case then is by that part of the peptide which is never directly bonded to the metal ion. To account for this it was suggested that the reactive species is in fact an ion-pair (which forms stereospecifically) perhaps through hydrogen-bonding with an NH_2 group of the ethylenediamine ligand.

Perhaps the most startling revelation was the accidental finding that the [Co trien]$^{3+}$ moiety is also effective in promoting the formation of peptide bonds; this was observed by two independent schools at approximately the same time.[92, 93] The reaction shown in Scheme 8 takes place in non-aqueous solvents.

$$cis\text{-[Co trien } L_2]^{n+} + 2\text{Gly-OEt} \longrightarrow cis\text{-[Co trien Gly-GlyOEt]}^{n+} + EtOH + 2L$$

Scheme 8

When L is tributyl phosphate the reaction is complete[92] within a few minutes at 25°; for $L = Cl^-$, the reaction is considerably slower.[93] The expected product [Co trien (GlyOEt)L]$^{n+}$ was not obtained, although this is probably an intermediate in the reaction. Interestingly, the peptide complexes obtained were shown to be intermediates in the previously described[89] hydrolysis reaction. Although the reaction appears to have exciting possibilities as a new method of peptide synthesis, it must be remembered that the co-ordinated peptide would be extremely difficult to remove without at the same time hydrolysing the peptide bond, and, furthermore, co-ordination labilises the α-hydrogens in the chelate ring [see section 3 C] which would favour racemisation of any asymmetric amino-acid residue. These problems, however, do not detract from the elegant nature of the observations.

Other Reactions. It has been known for some time that when complexed to metals, the methylene groups in glycinate chelate rings (or substituted glycinate) may become activated. Thus, the condensation of aldehydes and pyruvic acid to form new C—C bonds has been described.[66] Similarly, in the cobalt(III) complex [Co en_2(Gly)]$^{2+}$ the methylene protons of the glycinate chelate ring exchange rather readily with slightly alkaline D_2O, as shown by n.m.r. studies.[94] It was recently shown[95] that the methylene protons of the *C*-terminal glycinate in [Co(Gly-Gly)$_2$]$^-$ exchange with deuterons in alkaline D_2O whereas the protons of the *N*-terminal glycine residue do not. In the case of the stereospecifically formed cobalt(III) complex of glycyl-L-alanine, the protons of the asymmetric carbon of

[92] D. A. Buckingham, L. G. Marzilli, and A. M. Sargeson, *J. Amer. Chem. Soc.*, 1967, **89**, 2772.
[93] J. P. Collman and E. Kimura, *J. Amer. Chem. Soc.*, 1967, **89**, 6096.
[94] D. H. Williams and D. H. Bush, *J. Amer. Chem. Soc.*, 1965, **87**, 4644.
[95] R. D. Gillard, P. R. Mitchell, and N. C. Payne, *Chem. Comm.*, 1968, 1150.

L-alanine exchange at a much faster rate than that of racemisation. Hence, the exchange is stereoselective.

In addition to such activation of methylene groups in peptides by chelation, some attention has been given to the reactivity of co-ordinated peptides towards oxidation. Thus, oxidation of the copper(II) complexes of tetra-L-alanine at pH 6–9 with hexachloroiridate(IV) yielded equimolar quantities of di-L-alanineamide and pyruvoyl-L-alanine. Similarly, di-L-alanineamide and pyruvoyl-di-L-alanine were obtained from the chelated penta-L-alanine.[96] Copper(II) complexes of other peptides showed similar fragmentation, in which the third amino-acid residue is destroyed. The authors expressed the opinion that owing to its specificity the reaction could be of use in locating copper-binding sites in natural copper proteins. The uptake of molecular oxygen by cobalt(II) complexes of amino-acids and peptides has been known for some time; recently it has been reported[85] that, when the peptide is glycyl-L-tryptophan, the uptake is catalytic, leading to the oxygenation of the tryptophan and the formation of glycyl-kynurenine.

4 Proteins

A. Binding Sites.—Much work continues to be done on the type, affinity, and specificity of metal-binding sites in proteins. Gurd and his school have carried out some recent elegant studies on the copper(II) binding of the peptide comprising the first 24 residues of bovine serum albumin (isolated after a controlled peptic digestion). In the first report,[97] the peptide was shown to bind two copper(II) ions at neutral pH. One site had a higher copper-binding capacity than the other and at a copper : peptide ratio of 1 : 1 is exclusively occupied. This particular binding-site was said to involve the *N*-terminal atom of aspartyl (residue 1), the deprotonated N(peptide) atoms of threonyl (residue 2) and histidyl (residue 3), and the N-1 of the imidazole side-chain of histidyl (residue 3). The second report[98] confirmed this suggestion; by experiments with bromoacetate, it was shown that, in the presence of one equivalent of copper(II) ion, the terminal amino-group and histidyl (residue 3) were quantitatively protected against alkylation. Similarly, the binding of two equivalents of copper(II) also protects these same components and in addition the histidyl residues 9 and 18 are largely and about equally protected. The binding of the second equivalent of copper was therefore said to involve these two histidine residues (*i.e.* 9 and 18). Finally, the tetrapeptide L-aspartyl-L-threonyl-L-histidyl-L-lysine (*i.e.* the first four residues of the large peptide discussed above) was shown to form an equimolar complex of copper(II) which has substantially the same properties (titration, electronic spectrum, and circular dichroism) as the 1 : 1 complex of copper(II) with the 24-peptide

[96] A. Levitzki, M. Anbar, and A. Berger, *Biochemistry*, 1967, **6**, 3757.
[97] W. T. Shearer, R. A. Bradshaw, F. R. N. Gurd, and T. Peters, *J. Biol. Chem.*, 1967, **242**, 5451.
[98] R. A. Bradshaw, W. T. Shearer, and F. R. N. Gurd, *J. Biol. Chem.*, 1968, **243**, 3817.

residue. Equimolar mixtures of copper(II) and bovine serum albumin also had very similar properties.

In polarographic work on the binding of copper(II) to ovalbumin, it was concluded [99] that whereas at pH 5·5, copper(II) binding was by carboxyl groups, above pH 5·5 binding to histidyl residue occurs. This change of binding site from an oxygen donor at lower pH to a nitrogen donor at higher pH is apparently rather typical of copper(II) ions. The nature of the copper-binding site in serum ceruloplasmin (which contains *ca.* 90% of all serum copper) has been studied.[100] Photo-oxidised ceruloplasmin lost enzymic activity concomitantly with copper loss, and lack of enzymic activity was also found after reaction of the histidyl residues with diazo-1-H-tetrazole. Hence, the conclusion reached was that the histidyl residues were the copper(II)-binding sites and that tyrosyl residues were not involved as had been previously thought.

In work by Sarkar and Wigfield,[101] continuing attention is given to the mode of occurrence of 'free' copper in normal human serum. The binding of the first copper(II) ion to serum albumin is stronger than that of subsequent sites. When copper was added in the form of its amino-acid complexes, significant amounts of L-His and of L-Thr were attached to serum albumin. In the absence of copper no such binding occurred at pH 7·4. It was concluded that ternary complexes between albumin, copper, and these amino-acids are formed at both the first and subsequent sites. A particularly interesting finding [102] is that the binding *in vitro* of an encephalitogenic protein (from the central-nervous system) to myelin was enhanced by copper ions but not by other cations; this is of particular relevance to multiple sclerosis.

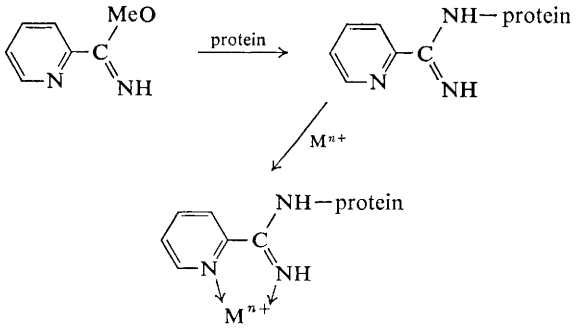
A study [103a] of the binding constants and pH dependence of the two most avid copper(II)-binding sites of ox pancreatic ribonuclease, and of the effect of copper(II)-binding on enzymic activity, implicates the histidyl residues 12 and 119 as the binding sites. Later work,[103b] by the same authors, on the protected histidine derivative, 1-carboxymethylhistidine[119]-ribonuclease, establishes the histidyl residue 119 as both being involved in the active enzyme site and as the most avid binding site for copper(II).

In other work on metal binding, the attachment of copper(II) to mohair keratin has been studied [104] and related to commercial treatments of wool and to the effect of a copper-deficient diet on the properties of the wool fibre. Similarly, the binding of chromium(III) to insoluble keratins has been discussed [105] in terms of the mechanism of wool dyeing.

[99] W. U. Malik and M. R. Jindal, *Electroanalyt. Chem. and Interfacial Electrochem.*, 1968, **19**, 436.
[100] E. J. Wood, C. M. Salisbury, and W. H. Bannister, *Experientia*, 1968, **24**, 1193.
[101] B. Sarkar and Y. Wigfield, *Canad. J. Biochem.*, 1968, **46**, 601.
[102] M. A. Moscarello, D. Murdoch, and D. D. Wood, *Canad. J. Biochem.*, 1968, **46**, 235.
[103] [a] R. H. Saundry and W. D. Stein, *Biochem. J.*, 1967, **105**, 107; [b] *Ibid.*, 1968, **108**, 583.
[104] R. E. Guthrie and S. H. Laurie, *Austral. J. Chem.*, 1968, **21**, 2437.
[105] F. R. Hartley, *Austral. J. Chem.*, 1968, **21**, 2277, 2723.

The binding of metal ions to insulin has been studied in solution by Graae,[106] who concludes that in dilute solutions, zinc and cadmium ions are bound to N(amino) groups, and in the solid-state by Brill and Venable.[107] *X*-Ray data suggest that copper(II)–insulin crystals are isomorphous with the much studied zinc–insulin crystals (zinc is found in association with insulin in pancreatic extracts). On the basis of e.s.r. experiments, copper(II) ions are said to be attached to three nitrogen atoms (from the imidazole groups of histidyl residues). However, this work employs the titration data of Tanford and Epstein which are criticised in Graae's paper.

In connection with the use of heavy-atom derivatives in protein structural studies, a new method of locating the heavy atoms is described.[108] The method was successfully employed with three heavy-atom derivatives of carboxypeptidase A. An interesting approach to the problem of making heavy-atom derivatives suitable for *X*-ray study has been described.[109] This depends on a covalent attachment to the protein of a metal-chelating group. For example, lysozyme reacts with methyl picolinimidate to introduce *ca.* seven picolinamidinyl groups into the protein as shown in Scheme 9.



Scheme 9

Model picolinamidines strongly chelate transition-metal ions. A different and well-known approach, labelling the protein with an active site-specific reagent which contains a suitable heavy atom, has been used in determining the structures of tosyl-chymotrypsin [110] and ribonuclease-S.[111] t-RNA has been crystallised in the presence of metal ions (such as copper, cadmium, and magnesium). These ions were thought [112] to be present in the crystals; this is comparable to the case of insulin where the formation of good crystals is known to require the presence of metal ions.

[106] J. Graae, *Biochem. J.*, 1968, **106**, 777.
[107] A. S. Brill and J. H. Venable, *J. Mol. Biol.*, 1968, **36**, 343.
[108] T. A. Steitz, *Acta Cryst.*, 1968, **B24**, 504.
[109] W. F. Benisck and F. M. Richards, *J. Biol. Chem.*, 1968, **243**, 4267.
[110] B. W. Mathews, P. B. Sigler, R. Henderson, and D. M. Blow, *Nature*, 1967, **214**, 652.
[111] H. W. Wyckoff, K. D. Hardman, N. M. Allewell, T. Inagrami, D. Tsernoglou, L. N. Johnson, and F. M. Richards, *J. Biol. Chem.*, 1967, **242**, 3749.
[112] H. H. Paradies, *F.E.B.S. Letters*, 1968, **2**, 112.

B. Conformations and Stereochemistry.—The divalent metal ions magnesium, calcium, and manganese have a marked effect on mixtures of DNA with homogeneous globular proteins,[113] in line with the profound influence of such metal ions on the structures of many nucleoproteins, including ribosomes. For example, with horse heart cytochrome and DNA, insoluble fibres were obtained only in the presence of the divalent ions. In the solid form the haem groups were shown by polarisation studies to be oriented at 60° to the fibre axis. The suggested mechanism is that the metal ions bind to the phosphate groups of the nucleic acids and hence reduce interchain electrostatic repulsion. The formation of fibres may be due to cross-linking by metal ions between DNA molecules; the haem protein molecules are then co-operatively bound into the nucleoprotein structure.

The nature of the conformational transitions observed in sperm whale ferrimyoglobin in the presence of several equivalents of copper(II) ion has been investigated.[114a] There appears to be a linear dependence between the amount of copper(II) ion added and the reduction in the characteristic e.s.r. signal of high-spin iron(III). Some degree of dipolar coupling of neighbouring paramagnetic ions is indicated, although direct Fe—Cu coupling can be ruled out. Concomitant with the apparent electronic change in the haem moiety is a decrease in helix content and a loss of solubility.[114b] These changes do not occur when the ratio Cu : protein is only 1 : 1.

C. Reactivity.—Some interesting results have appeared on the modification of collagen amino-acid composition attendant on the presence of iron-binding agents.[115] Biosynthetically, hydroxyproline is derived from proline, and similarly hydroxylysine is derived from lysine (*i.e.* hydroxylysine as such is not incorporated into collagen). In the presence of chelating agents, no collagen hydroxyproline was formed, whereas the amount of proline doubled. The formation of collagen hydroxylysine is completely prevented by 2,2′-bipyridyl, 1,10-phenanthroline, or 8-hydroxy-quinoline, all effective iron-binding agents. Ethylinediamine tetra-acetate has no effect (probably because it fails to penetrate the cell walls). The authors conclude that the mechanism involved in hydroxylation of proline and of lysine are the same, especially in respect of metal-ion participation.

It has been shown [116] that the effect of penicillamine on soft-tissue collagen from rats is unrelated to its copper-binding properties. Penicillamine, of course, is effective in treating Wilson's disease and its therapeutic properties have been attributed to its copper(II)-chelating properties. However, it has recently been found [117] that penicillamine (as already known

[113] D. Kabat, *Biochemistry*, 1967, **6**, 3443.
[114] [a] F. R. N. Gurd, K.-E. Falk, Bo G. Malmström, and T. Vanngård, *J. Biol. Chem.*, 1967, **242**, 5731; [b] C. R. Hartzell, K. D. Hardman, J. M. Gillespie, and F. R. N. Gurd, *J. Biol. Chem.*, 1967, **242**, 47.
[115] J. Hurych and A. Nording, *Biochem. Biophys. Acta*, 1967, **140**, 168.
[116] I. A. Jaffe, P. Merriman, and D. Jacobus, *Science*, 1968, **161**, 1016.
[117] D. D. Perrin and I. G. Sayce, *J. Chem. Soc.* (*A*), 1968, 53.

in the case of cysteine) readily reduces copper(II) to copper(I). Somewhat related to this is the work by Adam and Kühn [118a] which had as its origin the fact that gold thiosulphate has been used in the therapy of rheumatoid arthritis for 40 years, without a knowledge of the mechanism. After rats had been dosed with gold(I) complex, $Na_3[Au(S_2O_3)_2]$, a hardening of collagen was observed attributed to displacement of the thiosulphate ligands by polypeptide side-chain groups giving cross-linkages. This notion of cross-linkages is supported [118b] by the decrease in solubility of skin collagen after the treatment *in vivo* of lathyritic rats with gold thiosulphate.

5 Metalloproteins

A. Non-haem Metalloproteins.—It has been true for several years that the primary area of effort in purified metal–proteins has been in the nature of the metal-binding, and in the mechanistic function of the metal. The past year's reported work reflects this situation. A most interesting general paper by Vallee and Williams [119] makes some useful suggestions, which should perhaps serve as a mild warning against attempts to reproduce the specific catalytic properties of metalloenzymes using simple models. These authors point out that the spectroscopic properties and redox potentials of many Cu, Fe, and Zn(Co) metalloenzymes point to their having unusual geometric states, which are not normally found in simple co-ordination molecules. They consider that these 'unusual' geometric states correspond to highly energetic transition states, and that the metal and its ligands (through the secondary and tertiary protein structures) should be considered to generate the 'entatic' state jointly.

Copper Proteins. A molecular orbital model has been proposed to account for the intense blue colour and low nuclear hyperfine coupling constants for two varieties of *Pseudomonas* blue proteins.[120] This model represents a partial revival of the earlier views of Pauling. An experimental model for copper(II) in oxidoreductases has also been described,[121] in which at least some of the features of the e.s.r. and electronic absorption spectrum of the proteins are reproduced. For example, by γ-irradiation at 77°K of the copper(I) complex $[Cu(MeCN)_4]ClO_4$, a complex of copper(II) in a nearly tetrahedral environment is produced. The suggestion was made that the copper site in oxidoreductases is in an environment intermediate between the planar one [common for copper(II)] and the tetrahedral one [normal for copper(I)]. As expected from the Franck–Condon principle, this renders electron transfer particularly easy.

In contrast to these models based on distorted geometric states, a suggestion has also been made that, in the case of a number of enzymes

[118] [a] M. Adam and K. Kühn, *European J. Biochem.*, 1968, **3**, 407; [b] M. Adam, P. Fietzek, and K. Kühn, *ibid.*, p.411.
[119] B. L. Vallee and R. J. P. Williams, *Proc. Nat. Acad. Sci, U.S.A.*, 1968, **59**, 498.
[120] A. S. Brill and G. F. Bryce, *J. Chem. Phys.*, 1968, **48**, 4398.
[121] D. S. Gould and A. Ehrenberg, *European J. Biochem.*, 1968, **5**, 451.

containing more than one copper ion per molecule, a direct interaction between copper(I) and copper(II) ions is involved.[122] In support of this the authors prepared a mixed copper(I)–copper(II) acetate complex, the e.s.r. spectrum of which showed the copper ions to be indistinguishable, but which was not typical of that of copper enzymes.

Studies of the conformation in metal proteins by means of the Cotton effect (o.r.d. and c.d.) continue. It is perhaps not unfair to say that such studies are currently more useful in showing that a conformational change has occurred than in commenting on the detailed nature of such changes. The c.d. spectra of oxy-, deoxy-, and apo-haemocyanin have been measured. Oxyhaemocyanin shows Cotton effect at 700 (+), 570 (−), 450 (+), and 350 nm. (−), and positive bands in the u.v. region. In both deoxy- and apo-haemocyanin, the visible bands disappear and small changes occur in the u.v. region.[123a] Similar results were obtained on a different variety of haemocyanin by Van Holde.[123b] In a similar study[124] using *Pseudomonas* blue protein and laccase B from *Polyporas versicolor*, both proteins show six Cotton effects at wavelengths above 300 nm. which can be associated with the copper ion. The aromatic bands also show (in both proteins) distinct Cotton effects, perhaps suggesting that the environments of the aromatic residues are highly asymmetric.

Zinc Proteins. An elegant study of the mode of action of carbonic anhydrase by Riepe and Wang[125] employs i.r. spectroscopy in the region 2000–2800 cm.$^{-1}$, on aqueous solutions of the enzyme. For carbon dioxide loosely bound to a hydrophobic surface at the active site of the enzyme, $\nu_{asym}(CO_2)$ occurs at 2341 cm.$^{-1}$. The binding constant of CO_2 to the protein was estimated from difference i.r. spectra at varying pressures of gas. Nitrous oxide was shown to compete on about equal terms with carbon dioxide. Azide ion, which was also found to bind to the zinc ion, apparently offers steric hindrance to the attachment of carbon dioxide in the hydrophobic cavity near the zinc ion. Similar i.r. data on competition between bicarbonate or hydroxide with azide ion for zinc sites lead to the suggestion that carbonic anhydrase operates by attack on the bound carbon dioxide by hydroxide ion co-ordinated to zinc. Further comments on carbonic anhydrase are made as a result of a kinetic study[126] on the reaction of zinc(II) with apo-enzyme. The reaction was followed by the corresponding liberation of protons, u.v. spectral changes and regeneration of enzymic activity. In common with the findings[119, 120] on copper enzymes, it was found that the reaction could not be understood in terms of a reaction of the zinc ion with simple ligands of known structure, presumably

[122] C. Sigwart, P. Hemmerich, and J. T. Spence, *Inorg. Chem.*, 1968, **7**, 2545.
[123] [a] H. Takesada and K. Hamaguchi, *J. Biochem.*, 1968, **63**, 725; [b] K. E. Van Holde, *Biochemistry*, 1967, **6**, 93.
[124] S.-P. W. Tang, J. E. Coleman, and Y. P. Miller, *J. Biol. Chem.*, 1968, **243**, 4286.
[125] M. E. Riepe and J. H. Wang, *J. Biol. Chem.*, 1968, **243**, 2779.
[126] R. W. Henkens and J. M. Sturtevant, *J. Amer. Chem. Soc.*, 1968, **90**, 2669.

because of the different geometric requirements of the co-ordination sites.

Alkaline phosphatase from *Escherichia coli* has been shown [127] to contain four zinc ions per mole of protein (whose molecular weight is 89,000). Two of the zinc ions may be readily removed (by 8-hydroxyquinoline-5-sulphonic acid), but enzymic activity is then lost; activity is regained on readdition of the zinc ions. Tritium–hydrogen exchange has been studied for insulin, both containing and lacking zinc, in crystalline and solution phases.[128] The authors conclude that a conformational change occurs at some site independent of the regions that contain the bound zinc.

Kinetic studies of substrate hydrolysis, using carboxypeptidase A, containing manganese, cobalt, zinc, nickel, or cadmium, show certain anomalies which are not explicable on the basis of the particular metal ion present.[129] An elegant study of the properties of the tightly bound manganese in pyruvate carboxylase, which catalyses the reaction shown in Scheme 10,

$$\text{pyruvate} + \text{ATP} + \text{HCO}_3^- \underset{\text{Mg}_2^+}{\overset{\text{acetyl-CoA}}{\rightleftharpoons}} \text{oxalacetate} + \text{ADP} + \text{Pi}$$

Scheme 10

has been made,[130] using the now well-established n.m.r. technique involving relaxation rates of the bound water. The findings are consistent with inhibition (*e.g.* by oxalate, phenylpyruvate, and malonate) arising through bidentate chelation of the bound manganese, and three *cis*-ligand positions on the manganese are shown to be involved in the enzyme mechanism. Cysteamine oxygenase, isolated from horse kidney, has been shown [131] to contain iron, copper, and zinc at approximately one atom each per enzyme molecule. Not all of the metal ions are involved in enzymic activity.

Iron Proteins. Non-haem iron proteins (in particular those related to ferredoxins) have attracted much attention, in particular concerning the electronic configuration and the stereochemistries of the centres containing iron, though without general agreement. Mössbauer spectroscopy has been widely applied.[132] The reactions of ferredoxins from *Clostridium acidiurici* with chelating agents are described by Malkin and Rabinowitz.[133]

[127] R. T. Simpson and B. L. Vallee, *Biochemistry*, 1968, **7**, 4343.
[128] M. Praissman and J. A. Rupley, *Biochemistry*, 1968, **7**, 2431.
[129] R. C. Davies, J. F. Riordan, D. S. Auld, and B. L. Vallee, *Biochemistry*, 1968, **7**, 1090.
[130] M. C. Scrutton and A. S. Mildvan, *Biochemistry*, 1968, **7**, 1490.
[131] D. Cavallini, S. Duprè, R. Scandurra, M. T. Graziani, and R. Cotta-Ramusino, *European J. Biochem.*, 1968, **4**, 209.
[132] R. Cooke, J. C. M. Tsibris, P. G. Debrunner, R. Tsai, I. C. Gunsalus, and H. Fruenfelder, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 1045; C. E. Johnson, E. Elstner, J. F. Gibson, G. Benfield, M. C. W. Evans, and D. O. Hall, *Nature*, 1968, **220**, 1291; T. H. Moss, A. J. Bearden, R. G. Bartsch, M. A. Cusanovich, and A. San Pietro, *Biochemistry*, 1968, **7**, 1591.
[133] R. Malkin and J. C. Rabinowitz, *Biochemistry*, 1967, **6**, 3880.

The site of oxygen sensitivity in ferredoxin (*C. acidi-urici*) is shown to be the inorganic sulphide moiety. The 'labile sulphur' has been the subject of an ingenious experiment [134] in which two iron–sulphur proteins (one from *Pseudomonas*, the other from adrenals) were allowed to exchange their labile sulphur for selenium (^{77}Se, with $I = \frac{1}{2}$). The hyperfine structure of the e.s.r. spectra then showed that after reduction, the unpaired electron interacts with two ^{77}Se nuclei. In the original sulphur-containing proteins therefore the unpaired electron interacts with both iron and both labile sulphur atoms present.

Newman and Postgate [135] have isolated a new rubredoxin from a nitrogen-fixing variety of *Desulphovibrio desulphuricans*. It is somewhat similar to other rubredoxins in molecular weight (6500 ± 60), electronic and optical activity spectra, and chemical properties, including the relative inaccessibility of the chromophoric iron. The sequence and iron-binding site (four cysteine sulphurs, the indole nitrogen of tryptophan residue 35, and possibly a phenolic oxygen from tyrosine residue 11 or 13) have been established [136] for rubredoxin from *Micrococcus aerogenes*.

The complex polymer $[Fe_4O_3(OH)_4(NO_3)_2]_n$, with radius *ca.* 70 Å, has properties (Mössbauer spectrum and magnetic susceptibility) closely similar to those of the ion in the storage protein ferritin.[137] The exchange of ^{59}Fe among transferrin molecules is found to be rapid in the presence of citrate, whereas in the absence of complexing agent there is no exchange. The suggested mechanism [138] involves formation of a ternary complex transferrin–iron–citrate through displacement of bicarbonate by citrate. Complementary results (showing that certain anions such as oxalate compete with bicarbonate for the iron) have been reported by Young and Perkins.[139]

The thermodynamics and kinetics of oxygen binding by the respiratory pigment haemerythrin (from *Sipunculus nudus*) have been studied.[140] There is only a slight site–site interaction, and no Bohr effect. The complex kinetic behaviour may arise from conformational changes.

B. Haem Proteins.—A description of the three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2·8 Å resolution has appeared.[141]

[134] W. H. Orme-Johnson, R. E. Hansen, H. Beinert, J. C. M. Tsibris, R. C. Bartholomans, and I. C. Gunsalus, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **60**, 368.
[135] D. J. Newman and J. R. Postgate, *European J. Biochem.*, 1968, **7**, 45.
[136] H. Bachmayer, K. T. Yasunobu, and H. R. Whiteley, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 1273.
[137] G. W. Brady, C. R. Kurkjian, E. F. X. Lyden, M. B. Robin, P. Saltman, T. Spiro, and A. Terzis, *Biochemistry*, 1968, **7**, 2185.
[138] P. Aisen and A. Leibman, *Biochem. Biophys. Res. Comm.*, 1968, **32**, 220.
[139] J. W. Young and D. J. Perkins, *European J. Biochem.*, 1968, **4**, 385.
[140] G. Bates, M. Brunori, G. Amiconi, E. Antonini, and J. Wyman, *Biochemistry*, 1968, **7**, 3016.
[141] M. F. Perutz, H. Muirhead, J. M. Cox, L. C. G. Goaman, F. S. Mathews, E. L. McGandy, and L. E. Webb, *Nature*, 1968, **219**, 29; M. F. Perutz, H. Muirhead, J. M. Cox, and L. C. G. Goaman, *ibid.*, p.131.

As in myoglobin, the haem groups lie in non-polar pockets of the globin chain. A preliminary account of features of the *X*-ray structural analysis of horse-heart cytochrome *c* has been given.[142] Again the haem disc is located in a crevice in the protein structure; one co-ordination position is occupied by a histidine residue and the sixth probably by the methionyl residue 80. In support of this is the study by Stellwagen [143] on the carboxymethylation of ferricytochrome *c* and cyanoferricytochrome *c* who also concludes that methionyl residue 80 is a ligand for the haem iron in ferricytochrome *c*.

The Mössbauer spectra of peroxidase and some derivatives,[144] a number of cytochromes,[145] and a series of iron(III) compounds in strong-field tetragonal environments [146] have been reported.

Conformational changes in haem proteins have been further studied. Myer [147] showed that the c.d. of horse-heart ferricytochrome *c* below 240 nm. is unaffected by reduction, and concluded that the protein tertiary structure remains unaltered. The conformational transitions of ferricytochrome *c* have also been studied [148] as a function of pH and extrinsic chloride ion. In another conformational study [149] it was shown that the binding of one equivalent of haemin per globin dimer causes restoration of the helix content not only of the chain which has haemin bound but also almost all the helical content of the other chain in the dimer as well. Related to this finding is the discovery [150] using e.s.r. that, according to the spin-state of the β haem iron, there are small protein conformational changes near cysteine β93, but these do not cause haem–haem interactions. During the oxygenation of deoxyhaemoglobin, however, there are protein conformational changes near cysteine β93 and these are almost certainly involved in the haem–haem interaction.

The functional properties of a haemoglobin carrying haem only on its α chains have been evaluated.[151] Interestingly, the oxygen affinity of these chains is *ca.* 3–10 times higher than that of normal haemoglobin. Further, the reactivity of the haem moiety in a given chain is evidently modified by interaction with the partner chain, even if the latter is devoid of prosthetic groups.

In concluding this first report, it is worth pointing out that the enormous interest in the catalytic effects of metals in protein environments has led to

[142] R. E. Dickerson, H. L. Kopka, J. Weinzierl, J. Varnum, D. Eisenberg, and E. Margoliash, *J. Biol. Chem.*, 1967, **242**, 3015.
[143] E. Stellwagen, *Biochemistry*, 1968, **7**, 2496.
[144] Y. Maeda, *J. Phys. Soc. Japan*, 1968, **24**, 151.
[145] G. Lang, D. Herbert, and T. Yonetani, *J. Chem. Phys.*, 1968, **49**, 944; T. H. Moss, A. J. Bearden, R. G. Bartsch, and M. A. Cusanovich, *Biochemistry*, 1968, **7**, 1583.
[146] B. W. Dale, R. J. P. Williams, P. R. Edwards, and C. E. Johnson, *Trans. Faraday Soc.*, 1968, **64**, 620.
[147] Y. P. Myer, *J. Biol. Chem.*, 1968, **243**, 2115.
[148] D. Fung and S. Vinogradov, *Biochem. Biophys. Res. Comm.*, 1968, **31**, 596.
[149] K. Javacherian and S. Beychok, *J. Mol. Biol.*, 1968, **37**, 1.
[150] H. M. McConnell, S. Ogawa, and A. Horowitz, *Nature*, 1968, **220**, 787.
[151] K. H. Winterhalter, G. Amiconi, and E. Antonini, *Biochemistry*, 1968, **7**, 2228.

complementary work in areas varying from pure inorganic chemistry to pure biochemistry. As knowledge and techniques in these diverse disciplines improve, such collaborative work should become an essential feature in this field and rapid progress may therefore be expected.

Author Index