



MBL/WHOI



0 0301 0020723 9





SYMPOSIUM ON PROTEIN STRUCTURE

The Symposium was sponsored by the Protein Commission of the Section of Biological Chemistry of I.U.P.A.C. and was held at the College de France, Paris, 25-29 July 1957.

MEMBERS OF THE PROTEIN COMMISSION

A. Neuberger (*President*)
Stanford Moore (*Secretary*)
J. Roche (*Vice-President*)

Kenneth Bailey
Per Edman
C. Fromageot
K. Linderstrøm-Lang
Hans Neurath
J. L. Oncley
K. O. Pedersen

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY
(Paris Meeting 1957)

Symposium on Protein Structure

EDITED BY
ALBERT NEUBERGER

Professor of Chemical Pathology
St Mary's Hospital Medical School
London

LONDON · METHUEN & CO LTD
NEW YORK · JOHN WILEY & SONS INC

First published 1958
Copyright in all countries signatory
to the Berne Convention
© 1958 by Methuen & Co Ltd

Catalogue No. 6078/U
Printed in Great Britain by
The Camelot Press Ltd.,
London and Southampton

Preface

In 1951 the Section of Biological Chemistry of the International Union of Pure and Applied Chemistry (IUPAC) established a Protein Commission largely for the purpose of investigating the possibility of making available on an international basis a few protein preparations meeting reasonable criteria of homogeneity. This task was found to be more difficult than was foreseen, but early in 1957 an insulin preparation, made by the British Insulin Manufacturers, and examined by countercurrent and chromatographic methods, was made available. During discussion on a possible extension of this programme, it became apparent that the work of the Commission would be greatly helped and the value of any future plans be better assessed, if the question of homogeneity or 'purity' of proteins could be considered by a special symposium at which many of the leading protein chemists would be present. Financial support by IUPAC was promised in the most generous manner and it was decided to hold such a conference in Paris in the summer of 1957. As the planning of this symposium progressed it became clear that it was not advisable to restrict the subjects to be discussed too closely to questions of immediate relevance to the problem of homogeneity and in the end the programme of the conference was arranged to give an all-round survey of protein chemistry at the present time.

The IUPAC Symposium on Protein Structure took place in Paris from 25 to 29 July 1957 under the general chairmanship of Professor Jean Roche. The Symposium was attended by over 150 protein chemists, but it was felt that a much larger number of people might like to see in print the main papers given in Paris. These are now collected in the present volume, amplified by the contributions of many speakers in the Discussion. Unfortunately it has not been possible to include all the comments which were made and only those which gave new experimental facts directly relevant to the defined topics of the Conference are presented. In editing this book containing papers written in three different languages, a considerable degree of latitude has been adopted. The system of quoting literature references has been allowed to vary greatly, and both English and American spelling has been permitted.

We would like to thank all those who have helped to make the Paris symposium a success and the publication of this volume possible. Professor Murray Luck, the President of the Section of Biological Chemistry, and other Honorary Officers of IUPAC have given helpful support at all stages.

Professor Roche, Vice-President of the Commission, was responsible for the local arrangements and organized the Symposium in a most competent manner with unfailing courtesy. Dr Stanford Moore deserves the main credit for the efficient way in which the scientific side of the meeting was planned. We are also indebted to Dr Green for his help with the proof reading and for providing the index. Finally, I wish to express our appreciation to the publishers for their helpful co-operation in bringing out this volume so quickly and so efficiently.

ALBERT NEUBERGER

LONDON

May 1958

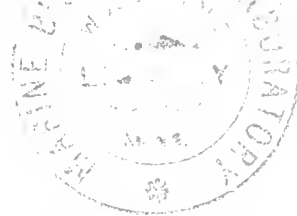
Claude Fromageot

1899-1958

It is with profound regret that we have to record the untimely death on January 10th, 1958 of Professor Claude Fromageot, whose contribution to the Protein Symposium appears in this volume. Fromageot was Secretary of the Biological Section of I.U.P.A.C., and with Professor Jean Roche contributed a great deal to the success of the Symposium. By his death, France has lost one of her leading biochemists, who was known internationally for his distinguished work in many fields, and whose eminence was achieved by the most selfless and disciplined devotion to his researches. These are notable not only for their quality but also for their unusual variety, ranging from structural aspects of protein chemistry—exemplified by papers on vasopressin, ovomucoid and lysozyme—to pioneer work on sulphur metabolism and the stability of enzymes. In addition to the responsibilities incurred by his duties as Professor of Biological Chemistry in the Faculty of Science at the Sorbonne and the guidance of several teams working under him, he found time to contribute most valuable review articles, of which perhaps that with Bricas on 'Naturally occurring peptides' is outstanding.

The loss to biochemistry is great, but even greater to those who knew him is the loss of a colleague whose integrity was a shining example to all.

Contents



Preface	page 7
Note on Claude Fromageot 1899-1958	9

GENERAL PROBLEMS AND METHODS

1 Factors affecting the structure of hemoglobins and other proteins <i>Linus Pauling</i>	17
2 Deuterium exchange and protein structure <i>K. Linderström-Lang</i>	23
3 The configuration of globular proteins in aqueous solution and its dependence on pH <i>Charles Tanford</i>	35
4 Über die Artspezifität der Proteinstruktur <i>H. Tuppy</i>	66
5 Comparative studies in the field of protein microstructure <i>F. Šorm</i>	77
Discussion on the paper of Professor Šorm <i>B. Keil</i>	90
6 Zone electrophoresis and chromatography as methods for the purification and characterization of proteins <i>A. Tiselius</i>	93
Mutual displacement in protein chromatography <i>Hans G. Boman</i>	100
7 The characterization of lower molecular weight proteins by dialysis <i>L. C. Craig, Wm. Konigsberg, A. Stracher and T. P. King</i>	104
8 Quantitative Papierchromatographie der Aminosäuren durch Isotopenverdünnung <i>F. Turba</i>	116

SPECIFIC PROTEINS

Haemoglobin and Myoglobin

9 Crystallographic studies of myoglobin <i>J. C. Kendrew</i>	125
---	-----

	X-ray analysis of haemoglobin <i>M. F. Perutz</i>	page 136
	Further study on myoglobin: I. Heterogeneity of human myoglobin. Some properties of Mb I and Mb II <i>A. Rossi-Fanelli and E. Antonini</i>	140
	II. Chemical and Biochemical properties of a new type of myoglobin in molluscs <i>A. Rossi-Fanelli, E. Antonini and D. Povoledo</i>	144
	The chemical effects of gene mutations in some abnormal human haemoglobins <i>J. A. Hunt and V. M. Ingram</i>	148
<i>Proteolytic Enzymes</i>		
10	Etude de quelques protéolyses limitées du chymotrypsinogène de bœuf <i>P. Desnuelle et M. Rovero</i>	155
	Studies on the autocatalytic activation of trypsinogen <i>J. F. Pechère and Hans Neurath</i>	169
	Comments on the structure of α -chymotrypsin <i>B. S. Hartley</i>	175
	Reaction of trypsin with organophosphorus esters <i>T. Viswanatha</i>	176
	Modification of pepsin by autodigestion <i>Gertrude E. Perlmann and Mary J. Mycek</i>	179
11	Some studies on the structure and activity of papain <i>Emil L. Smith, Robert L. Hill and J. R. Kimmel</i>	182
<i>Ribonuclease</i>		
12	Studies on the structure of ribonuclease <i>C. H. W. Hirs, William H. Stein and Stanford Moore</i>	211
13	The structure of ribonuclease in relation to its enzymatic activity and physical properties <i>C. B. Anfinsen</i>	223
	Immunological approaches to the study of ribonuclease <i>B. Cinader and J. H. Pearce</i>	240
<i>Tobacco Mosaic Virus</i>		
14	Degradation and structure of tobacco mosaic virus <i>H. Fraenkel-Conrat and K. Narita</i>	249

15	Strukturuntersuchungen am Protein des Tabakmosaikvirus <i>G. Schramm, G. Braunitzer, F. A. Anderer, J. W. Schneider und H. Uhlig</i>	page 262
	X-ray diffraction studies of the structure of the protein of tobacco mosaic virus <i>Rosalind E. Franklin</i>	271
<i>Other Proteins and Peptides</i>		
16	Nouvelles données concernant la structure du lysozyme d'œuf de poule <i>P. Jollès, J. Jollès-Thaureaux et C. Fromageot</i>	277
17	The isolation of an immunologically active fragment of bovine serum albumin <i>R. R. Porter</i>	290
	Serum protein changes during differentiation <i>Earl Frieden</i>	296
	La variabilité allotypique de certaines protéines du serum <i>Jacques Oudin</i>	298
18	Species variation and structural aspects in some pituitary hor- mones <i>Choh Hao Li</i>	302
	Intervention sur l'exposé du Professeur Li a propos des groupes C-terminaux de l'hormone de croissance et de l'hormone lactogène <i>M. Jutisz</i>	330
	The structure and activity of melanocyte-stimulating and adrenocorticotropic peptides <i>J. Ieuan Harris</i>	333
	Some observations on the structural basis of enolase activity <i>Bo G. Malmström</i>	338
	Insulin crystals: zinc atoms in the unit-cell, nucleation, growth and shape <i>Jørgen Schlichtkrull</i>	341
	Subject Index	344

GENERAL PROBLEMS AND METHODS

Factors affecting the structure of hemoglobins and other proteins

LINUS PAULING

Gates and Crellin Laboratories of Chemistry
California Institute of Technology, Pasadena, Calif.

(Communication No. 2215)

An understanding of the structural basis of the difference in properties of normal adult human hemoglobin and the abnormal hemoglobins, such as sickle-cell-anemia hemoglobin, will be obtained only when complete structure determinations of the molecules of these substances have been made. The properties of hemoglobin depend not only on the sequence of amino-acid residues in the polypeptide chains but also on the way in which the polypeptide chains are folded.

It is probable that the folding of the polypeptide chains of a protein is in considerable part, perhaps in some cases entirely, determined by the amino-acid sequence in the chains. For example, it has been shown by Doty, Moffitt, and their co-workers^{1,2,3} that the chains of poly- α -L-glutamic acid and of poly- γ -benzyl-L-glutamate have the configuration, in solution in certain solvents, of the right-handed α -helix, and that the molecules resume this right-handed helical configuration after they have been unfolded by change in the nature of the solvent, and then have been returned to the original solvent in which the α -helix is stable. It is evident that the interactions of the side chains of the L-amino-acid residues stabilize the right-handed α -helix relative to the left-handed α -helix for these polypeptides. Yang and Doty⁴ have also reported that the molecules of certain proteins—especially silk fibroin and the B chain of insulin—can, in certain solvents, be brought into the configuration of the right-handed α -helix. These assignments of sense to the α -helix, made by application of the theory of optical activity developed by Moffitt,⁵ are probably more reliable than the earlier assignments made by Riley and Arndt.⁶

There is a possibility that the difference in properties of the abnormal hemoglobins and normal adult human hemoglobin can be ascribed entirely to the replacement of a few amino acid residues by residues of other

amino acids. Such a change in even two or three residues at the surface of the molecule might introduce or destroy the self-complementariness in surface configuration to which is attributed the low solubility of sickle-cell-anemia hemoglobin, and the sickling of the cells consequent upon the formation of tactoids. There is, however, some evidence that a change in configuration of the polypeptide chains is also involved. The principal evidence indicating this change in configuration is the observed heterogeneity of globin made from normal adult human hemoglobin and a similar heterogeneity of the globin from sickle-cell-anemia hemoglobin, as illustrated by the work of Havinga and Itano.⁷ Havinga and Itano reported that carefully prepared globins of the two kinds show a difference in electrophoretic mobility similar to that of the hemoglobins, but that slight denaturation causes the two globin preparations to become heterogeneous, the heterogeneity being of such a nature as to suggest that there are two ways of folding the polypeptide chains, one characteristic of normal hemoglobin and the other of sickle-cell-anemia hemoglobin. It is possible, of course, that even a change in a couple of amino acid residues could effect a change in the way of folding the chains in one part of the molecule.

The difficulties of experimental determination of the complete molecular structure of globular proteins are so great as to make it essential that every possible aid be utilized. In the attacks on the problem of protein structure one of the most valuable aids has been found to be the formulation of structural principles for polypeptide chains. There is now no doubt that stable configurations of polypeptide chains involve the planar configuration of the amide group, with distances $\alpha\text{C}-\text{C}'=1.53 \text{ \AA}$, $\text{C}'-\text{O}=1.24 \text{ \AA}$, $\text{C}'-\text{N}=1.32 \text{ \AA}$, $\text{N}-\alpha\text{C}=1.47 \text{ \AA}$ (all $\pm 0.01 \text{ \AA}$), and angles $\text{N}-\alpha\text{C}-\text{C}'=110^\circ$, $\alpha\text{C}-\text{C}'-\text{N}=114^\circ$, $\text{O}-\text{C}'-\text{N}=125^\circ$, and $\text{C}'-\text{N}-\alpha\text{C}=123^\circ$. The planarity of the amide group and the shortening of the $\text{C}'-\text{N}$ bond to 1.32 \AA are

attributed to resonance between structures

$$\begin{array}{c} \text{H} \\ | \\ \text{C}-\ddot{\text{N}} \\ // \quad \backslash \\ \text{O} \quad \quad \quad \end{array} \quad \text{and} \quad \begin{array}{c} \text{H} \\ | \\ \text{C}=\ddot{\text{N}} \\ // \quad \backslash \\ \text{O} \quad \quad \quad \end{array} ,$$

with about 60% and 40% contributions respectively, corresponding to 40% double-bond character for the $\text{C}'-\text{N}$ bond. In addition, the stable configurations of polypeptide chains involve the formation of $\text{N}-\text{H}\cdots\text{O}$ bonds, with the oxygen atom close to the $\text{N}-\text{H}$ axis and the nitrogen-oxygen distance equal to $1.79 \pm 0.10 \text{ \AA}$.⁸

The trans configuration of the amide group seems to be significantly more stable than the cis configuration. At the present time the only known peptide structure in which the amide groups have the cis configuration is that of diketopiperazine,⁹ in which the nature of the covalent bonds (a six-membered ring, for the cyclic dipeptide) is such as to require this configuration. Spectroscopic evidence showing that the trans configuration in simple

amides is significantly more stable than the *cis* configuration has been summarized by Mizushima.¹⁰ From infrared absorption studies of amides and *N*-substituted amides Badger and Rubalcava¹¹ concluded that the *trans* configuration may well be stabilized by more than 2 kcal/mole.

The principal degrees of freedom in the polypeptide chain are the azimuthal orientations around the N— α C single bond and the α C—C' single bond. It was suggested a few years ago by Pauling and Corey¹² that six values of the azimuthal angle for each of these single bonds are more stable than other values. Some doubt has been thrown on this conclusion by the observation that the amplitude of the Fourier term with six minima for rotation around the C—N single bond in nitromethane is extremely small, only about 0.003 kcal/mole. At the present time it seems to be wise to allow all azimuthal orientations about the single bonds in the attempted formulation of configurations of polypeptide chains.

There exists now evidence indicating that an additional structural feature for polypeptide chains can be accepted. In their recent review of the configuration of polypeptide chains in proteins Pauling and Corey⁸ stated their opinion at that time of the answer to the question of the orientation of the N—H \cdots O hydrogen-bond axis relative to the O—C' axis in the following words: 'Although it might be anticipated from consideration of the partial covalent character of the H \cdots O interaction that the favored orientation of the N—H \cdots O hydrogen bond about an oxygen atom of an amide group would be in the plane of the amide group and at an angle of about 125° with the C'—O bond, there is little evidence that this orientation is at all favored with respect to others, and at the present time it may be assumed that the oxygen atom may form a hydrogen bond in any direction.'

There exists, however, significant indication that the stable angle between the hydrogen-bond axis and the O—C' axis in polypeptides is not about 125°, but is close to a straight angle. The evidence supporting this structural feature, an angle of 180° at the oxygen atom, is summarized in the following paragraphs.

First, all of the known (reasonably well substantiated) structures for polypeptide chains in nature, including synthetic polypeptides, have the angle approximately 180° for the hydrogen bond at the oxygen atom. These structures include the α -helix,¹³ the antiparallel-chain pleated sheet,¹⁴ the parallel-chain pleated sheet,¹⁴ the structure of polyglycine II,¹⁵ and the several three-chain structures of collagen similar to those suggested by Crick and Rich.¹⁶

In addition, there are several simple peptides and closely related substances in which this angle is nearly 180°. Among these structures are β -glycylglycine,¹⁷ *N,N'*-diacetylhexamethylenediamine,¹⁸ glycyl-L-tyrosine hydrochloride monohydrate,¹⁹ and creatinine;²⁰ in all of these structures the deviation from a straight angle is less than 30°. There are also several simple peptides in which the angle at the oxygen atom is as small as 120° to 140°

(*N*-acetylglycine,²¹ *N,N'*-diglycyl-L-cysteine dihydrate,²² glycyl-L-tryptophan hydrate²³), and even some in which the N—H···O hydrogen bond involving the NH and O of the peptide groups is not formed, as for one of the peptide oxygens in glycyl-L-asparagine.²⁴ It should not be surprising that the molecules of a simple substance, in seeking the way of arranging themselves in a crystal that minimizes the free energy of the crystal, should occasionally have to sacrifice the most favorable circumstances for hydrogen-bond formation in order to satisfy some other structural features leading to stability.

In many of these substances the N—H···O hydrogen bonds that are formed by oxygen atoms other than those in peptide groups involve angles of about 120°, with the atoms adjacent to the oxygen atom coplanar. Further examples of this sort are the carboxyl oxygens in the amino acids glycine,²⁵ alanine,²⁶ L_S-threonine,²⁷ and DL-serine.²⁸

A simple theoretical consideration can be made that supports the proposed principle that under ordinary circumstances an oxygen atom attached to a carbon atom by a double bond tends to form hydrogen bonds that lie in the plane formed by the carbonyl group and adjacent atoms, with the angle at the oxygen atom equal approximately to 120°, whereas in hydrogen bonds formed by a peptide group the angle tends to be a straight angle. Let us consider the electrons of the oxygen atom in a carbonyl group. There are two electrons involved in the double bond (shared with two electrons of the carbon atom), and two electron pairs occupy the remaining two orbitals of the neon shell of the oxygen atom. We may discuss the deviation of the electronic structure from spherical symmetry by considering only the distribution of one electron in each of the two orbitals for unshared pairs inasmuch as there is, as a first approximation, spherical symmetry when one electron is placed in each of the four orbitals of the neon shell. Also, as a first approximation, we may treat the orbitals as tetrahedral *sp*³ orbitals. The treatment of the double bond as involving two bent tetrahedral bonds is mathematically equivalent to its treatment as a sigma bond and a pi bond.

When the orbitals for the shared pairs are taken as $\frac{1}{2}s + \frac{1}{\sqrt{2}}p_z + \frac{1}{2}p_x$ and $\frac{1}{2}s - \frac{1}{\sqrt{2}}p_z + \frac{1}{2}p_x$ (the bond direction being the *x* axis), it is found that the electron distribution for one electron in each of these orbitals, with arbitrary normalization, is given by the expression

$$\rho(\pi_z) = 1 + 6 \sin^2 \theta \cos^2 \phi + 3 \cos^2 \theta + 2\sqrt{3} \cos \theta \quad (1)$$

When this is differentiated with respect to the angle θ and the angle ϕ , it is found that there are two maxima in the density, both lying in the plane determined by the atoms around the carbon atom and at the angle 125° 16' with the C—O axis.

In discussing the distribution of the electrons around the oxygen atom

of the amide group, as in a peptide, we must take into consideration the contribution of structures in which there is an interaction between the ionic character of the sigma bond and the unshared pair of the oxygen atom lying in the plane of the group. This causes the nature of the carbon-oxygen bond to assume to some extent the character of the pi bonds in carbon dioxide. In carbon dioxide we may, as a first approximation, describe each carbon-oxygen bond as a double bond to which the p_z and the p_y electrons contribute equally (there is also, of course, some contribution of the triply bonded and singly bonded structures, as well). The contribution for the bond in which the double bond has 50% π_z character and 50% π_y character, normalized to the same basis as for Equation 1, is equal to

$$\rho\left(\frac{1}{2}\pi_z + \frac{1}{2}\pi_y\right) = 4 + 2\sqrt{3} \cos \theta \quad (2)$$

It is found by differentiating this expression that the maximum is at $\theta=0^\circ$; that is, along the carbon-oxygen axis.

In the amide group the ionic structure for the sigma bond between the carbonyl carbon and the nitrogen atom would effectively liberate a p orbital for carbon lying in the plane of the group, and permit a double bond to be formed with use of the p_y orbital of the oxygen atom and the pair of electrons written as an unshared pair in the usual structural formulas. The electron distribution corresponding to a contribution by $n\%$ of this structure, with a contribution of $(100-n)\%$ of the structure involving the double bond orbitals perpendicular to the plane of the group, is found on differentiating the corresponding expression for ρ to have the maximum in electron density along the carbon-oxygen axis when the double bond has 21% or more π_y character and 79% or less π_z character. It seems not unreasonable to expect that there is approximately this amount of π_y character in the double bond in the amide group, and accordingly that the electron distribution for the oxygen atom in the amide group is such as to favor a linear arrangement of the hydrogen bond at this atom, whereas in other compounds containing carbon-oxygen double bonds the electron distribution has its maximum in the plane of the ring, with the angle approximately equal to the tetrahedral angle.

From these considerations we reach the conclusion that an additional structural principle can be formulated for folded polypeptide chains; namely, that the stable configurations will tend to be those in which the $\text{N}-\text{H}\cdots\text{O}-\text{C}$ hydrogen bond is not only linear with respect to the three atoms $\text{N}-\text{H}\cdots\text{O}$, but also with respect to the carbon atom adjacent to the oxygen atom, with the bond angle at the oxygen atom close to 180° . This structural principle may be of value in the search for additional polypeptide-chain structures.

REFERENCES

1. P. DOTY, J. H. BRADBURY, and A. M. HOLTZER, *J. Am. Chem. Soc.*, **78**, 947 (1956).
2. P. DOTY and R. D. LUNDBERG, *Proc. Nat. Acad. Sci. (USA)* **43**, 213 (1957).
3. W. MOFFITT and J. T. YANG, *Proc. Nat. Acad. Sci. (USA)*, **42**, 596 (1957).
4. J. T. YANG and P. DOTY, *J. Am. Chem. Soc.*, **79**, 761 (1957).
5. W. MOFFITT, *J. Chem. Phys.*, **25**, 467 (1956).
6. D. P. RILEY and U. W. ARNDT, *Proc. Roy. Soc. (London)*, **B 141**, 93 (1953).
7. E. HAVINGA and H. A. ITANO, *Proc. Nat. Acad. Sci. (USA)*, **39**, 65 (1953).
8. L. PAULING and R. B. COREY, *Fortschritte der Chemie organischer Naturstoffe* **XI**, 180 (1954).
9. R. B. COREY, *J. Am. Chem. Soc.*, **60**, 1598 (1938).
10. S. MIZUSHIMA, T. SIMANOUTI, S. NAGAKURA, K. KURATANI, M. TSUBOI, H. BABA, and O. FUJIOKA, *J. Am. Chem. Soc.*, **72**, 3490 (1950).
11. R. M. BADGER and H. RUBALCAVA, *Proc. Nat. Acad. Sci. (USA)*, **40**, 12 (1954).
12. L. PAULING and R. B. COREY, *Proc. Nat. Acad. Sci. (USA)*, **37**, 729 (1951).
13. L. PAULING, R. B. COREY, and H. R. BRANSON, *Proc. Nat. Acad. Sci. (USA)*, **37**, 205 (1951).
14. L. PAULING and R. B. COREY, *Proc. Nat. Acad. Sci. (USA)*, **39**, 255 (1953).
15. A. RICH and F. H. C. CRICK, *Nature*, **176**, 780 (1955).
16. F. H. C. CRICK and A. RICH, *Nature*, **176**, 915 (1955).
17. E. W. HUGHES and W. J. MOORE, *J. Am. Chem. Soc.*, **71**, 2618 (1949).
18. M. BAILEY, *Acta Cryst.*, **8**, 575 (1955).
19. D. W. SMITS and E. H. WIEBENGA, *Acta Cryst.*, **6**, 531 (1953).
20. S. DUPRÉ and H. MENDEL, *Acta Cryst.*, **8**, 311 (1955).
21. G. B. CARPENTER and J. DONOHUE, *J. Am. Chem. Soc.*, **72**, 2315 (1950).
22. H. L. YAKEL, JR., and E. W. HUGHES, *Acta Cryst.*, **7**, 291 (1954).
23. R. A. PASTERNAK, *Acta Cryst.*, **9**, 341 (1956).
24. R. A. PASTERNAK, L. KATZ, and R. B. COREY, *Acta Cryst.*, **7**, 225 (1954).
25. G. ALBRECHT and R. B. COREY, *J. Am. Chem. Soc.*, **61**, 1087 (1939).
26. H. A. LEVY and R. B. COREY, *J. Am. Chem. Soc.*, **63**, 2095 (1941).
27. D. P. SHOEMAKER, J. DONOHUE, V. SCHOMAKER, and R. B. COREY, *J. Am. Soc.*, **72**, 2328 (1950).
28. D. P. SHOEMAKER, R. E. BARRIEAU, J. DONOHUE, and C. S. LU, *Acta Cryst.*, **6**, 241 (1953).

Deuterium exchange and protein structure

K. LINDERSTRØM-LANG

The Carlsberg Laboratory, Copenhagen, Denmark

1. INTRODUCTION

In a series of recent papers by Hvidt, Johansen, Vaslow, Berger, and myself^{1,2,3,4,5,6,7,8,9} a study was made of the rate of deuterium exchange between ordinary water and proteins or peptides in which all the easily exchangeable hydrogen atoms, i.e. those bound to oxygen, nitrogen, and sulphur atoms, had been replaced by deuterium atoms. The conditions under which this rate of exchange was studied were varied within rather wide limits, the temperature e.g. was varied from 0° to 40°C, pH from 2 to 8. In some cases buffer, in others denaturing agents like urea or guanidinium chloride were added.

The results may be summarized as follows:

1. Neglecting the possibility that the exchange of one deuterium atom may influence the rate of exchange of another, the rate curve for any peptide or protein may be represented by a sum of first order terms, viz.:

$$n_{\infty} - n = \sum n_i e^{-\beta_i t}; \quad \sum n_i = n_{\infty} \quad (1)$$

where n_{∞} is the total number of exchangeable deuterium atoms per peptide or protein molecule and n_i the number of atoms within any group characterized by the same rate constant β_i . The quantity n is the average number of exchanged deuterium atoms at the time t . The *half-time* of the rate group i is given by

$$\left(t_{\frac{1}{2}}\right)_i = \frac{\ln 2}{\beta_i} = \frac{0.6932}{\beta_i} \quad (2)$$

As everybody knows, most kinetic data (and ours too) can be represented by such exponential series and the number of constants β_i required is the smaller, the less accurate the experiments.

2. The deuterium atoms bound to oxygen and nitrogen in the end groups or side chains of peptides and proteins seem to exchange fast in most cases, viz. with a half-time of less than 0.5 min. at pH 3 and 0°, i.e. with a velocity which is too high to be picked up by our methods, a point to be discussed

later. In a few cases, where the side chains are engaged in hydrogen bonding, lower reaction rates might be observed, but we have been unable to distinguish such cases from those discussed in the following paragraph.

3. The deuterium atoms bound to nitrogen in the backbone of the peptide chain show a highly varying rate of exchange. Half-times from less than 0.5 min. to more than 24 hours have been observed at pH 3 and 0°. Since the fast reactions are primarily found in unfolded, denatured, proteins and peptides, whilst the slow reactions are predominant in native molecules, it is natural to ascribe the sluggishness of the reaction in the latter case to internal hydrogen bonding in the secondary structure of these native molecules, viz. to the presence of groups of the type



where the deuterium atom is less available for exchange. It is to be expected that the stronger the hydrogen bond in question, the slower the observed exchange, i.e. the smaller β_i .

4. The rate of deuterium exchange is strongly dependent upon temperature and generally the more so, the lower the rate.

5. The rate of deuterium exchange varies with pH, the most general feature being that it increases with increasing pH from a pH about 3. In certain cases this may be due to a change in the stability of the hydrogen bonds caused by ionization or desionization of the protein molecule but in the best investigated case, which I shall discuss in a moment, the determining factor seems to be the rate of the exchange process proper which is strongly catalysed by hydroxyl ions.

6. Addition of denaturing agents generally increases the rate of exchange in agreement with the fact that deuterium atoms released from internal hydrogen bonds of the secondary structure may react faster with the solvent.

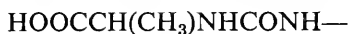
The rapidly growing literature on the secondary structure of proteins and polypeptides has made it likely that the α -helix of Pauling and Corey is an essential element of this structure^{10,11,12,13,14,15}, and the bonds of type (i) may therefore in many cases be regarded as situated in α -helices. One of these cases is represented by poly-DL-alanine (PDLA) which in its soluble form, according to Elliott,^{10,11} shows an infrared band at 1662 cm^{-1} , considered to be characteristic of the internal CO-HN-bond. The question whether this polypeptide in aqueous solution maintains its helical configuration is of course of considerable interest, and since PDLA contains no exchangeable hydrogen atoms in its side chains and furthermore is unassociated in aqueous solution, it appears to be an ideal object for the study of the stability of helical hydrogen bonds by means of the deuterium exchange method.

In the following a discussion of some of the results obtained with PDLA will therefore be given.

2. THE PDLA PREPARATION

The preparation of PDLA used was synthesized in Katchalski's laboratory and was reported to contain an average of 30 residues per molecule. The material, which contained pyridine, was purified by dissolution in dilute NaOH, lyophilization over P_2O_5 , redissolution in an equivalent amount of HCl, and, finally, lyophilization after removal of a small quantity of undissolved gel. The purified PDLA dissolved in water to a clear solution. The pH of the solution was 2.8, and the peptide was therefore present mainly as a positive charged ion.

A simple PDLA molecule containing 30 residues has a molecular weight of 2151 and a nitrogen content of 19.53%. The ratio of α -amino groups to total nitrogen atoms should be 1:30. The present preparation when dissolved in 0.1N-KCl at 22°C gave a titration curve which showed the presence of one amino group (pK 8.1) per 36.7 nitrogen atom, a figure which is compatible with the assumed average chain length, if end groups of the type



(see Sela and Berger¹⁶) replace α -amino groups in about 20% of the molecules. PDLA molecules with such end groups have a molecular weight of 2266 and a nitrogen content of 19.16%, and the average nitrogen content of our preparation should therefore be 19.46%. The experimental value 19.0% (corrected for NaCl and acid content) is in fair agreement with the theoretical one. The molar concentration, c_s , of PDLA in a given solution was calculated from the nitrogen content, and the number of exchangeable hydrogen atoms per molecule was assumed to be 29 (backbone)+4 (end groups) at pH 2.8, i.e. a total of 33.

Since the hydrogen exchange was measured in 0.1M citrate buffers which were 0.1M with respect to glycerol (cf. 7), the sedimentation constant of PDLA was determined at pH 3 in this medium. The sedimentation constant corrected to water at 20°C was found to be represented by the equation

$$S_{20} \cdot 10^{13} = 0.4 - 0.07 \times (\% \text{ PDLA})$$

in the concentration range 0.5 to 2%, which is compatible with a molecular weight of 2000 ($D \sim 20 \cdot 10^{-7}$, $V \sim 0.74$). Hence there is no indication of significant association. The question of the distribution of molecular sizes in our preparation is unsettled but, since the boundary spreading in the centrifuge appeared to be normal, it may be inferred that the size distribution is within rather narrow limits.

3. EXCHANGE METHOD

The method used in all cases was Method 2 described in (4) and (7); i.e.

PDLA was first deuterated completely and the rate of back-exchange with ordinary water was followed.

A volume of 150 $\mu\text{l.}$ of a 2% PDLA solution, pH 2.8, was lyophilized and dried in a vacuum over P_2O_5 . Next 150 $\mu\text{l.}$ of 99.7% D_2O was added and allowed to exchange with the peptide at 38.6°C for 4-15 hr. The deuterated peptide was freed from D_2O by cryosublimation and drying for 4 hr. at 68°C,^{4,7} after which time 150 $\mu\text{l.}$ of 0.1 M citrate buffer of the desired pH and temperature were added, and samples of 15 $\mu\text{l.}$ were taken out at intervals for measurement of the deuterium exchange (cryosublimation and determination of the density of the water in the gradient tube.^{1,4,7})

It should be noted that back-exchange is measured here in buffers in order to fix pH in the otherwise poorly buffered PDLA solutions. We also call attention to the fact that the solid, the dissolution of which is the initial step in the kinetic experiment, is the same in all cases (obtained by lyophilization of 2% PDLA solutions at pH 2.8), and that consequently pH is determined by the composition of the buffers added. The value obtained for the total exchange should therefore be the same in all experiments ($n_\infty = 33$).

4. RESULTS OF THE EXCHANGE EXPERIMENTS

In each kinetic experiment the total exchange, n_∞ , at the given conditions was determined by placing from one to two 15 $\mu\text{l.}$ samples at 38.6°C for 4-15 hr. and determining the deuterium content of the water. As the mean of 24 determinations under widely different conditions, a value of 33.2 was found with a standard deviation of 0.4, a result which is in good agreement with the above assumptions. Nevertheless the n -values of a given kinetic run were corrected by multiplying them by the ratio $33/(n_\infty)_x$, where $(n_\infty)_x$ is the value for n_∞ determined in this particular run. A rough correction for losses during lyophilization and incomplete removal of D_2O is obtained in this way.

Figs. 1 and 2 show the kinetic results. The curves drawn are theoretical ones and will be explained in the following section.

5. THEORY

It is obvious that the curves in Figs. 1-2 may be satisfactorily described by Eq. 1 (see (4)) if the number and values of β_i can be adjusted arbitrarily to fit the curves. The variations of β_i with a given parameter X (pH) may be most adequately determined by the variation of t at constant extension of the reaction (constant n). From Eq. (1) we get

$$\left(\frac{\delta \ln t}{\delta X}\right)_n = -\frac{\sum \frac{\delta \beta_i}{\delta X} e^{-\beta_i t}}{\sum \beta_i e^{-\beta_i t}} \equiv -\frac{\delta \beta_i / \delta X}{\beta_i} \quad (3)$$

If now the β_i may be regarded as products of a general factor, B , which

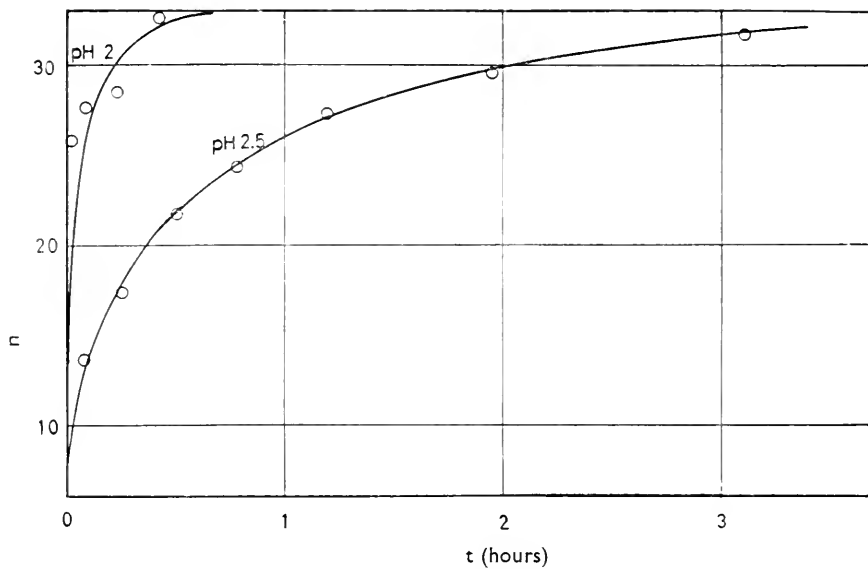


Fig. 1. Deuterium exchange of PDLA at 0°C.

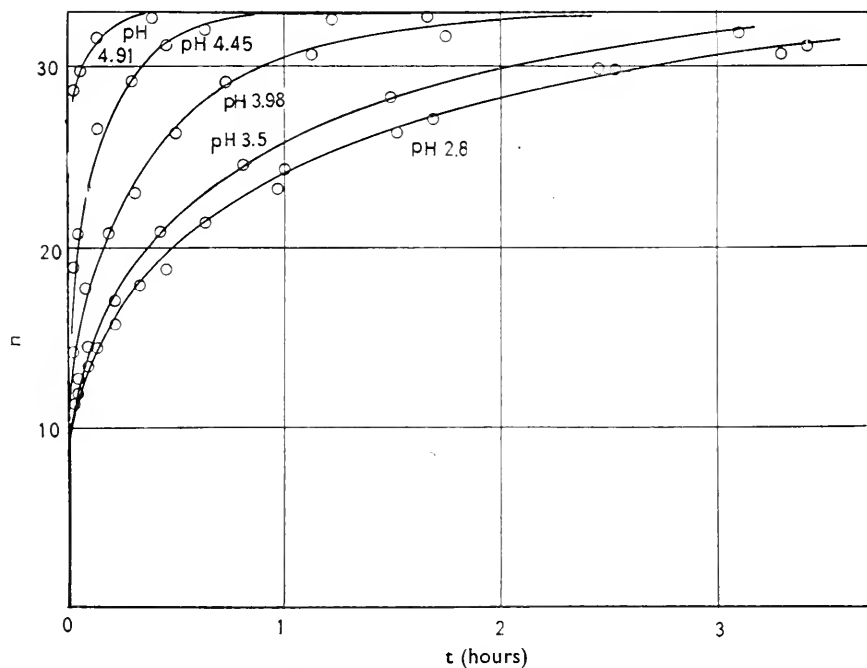


Fig. 2. Deuterium exchange of PDLA at 0°C.

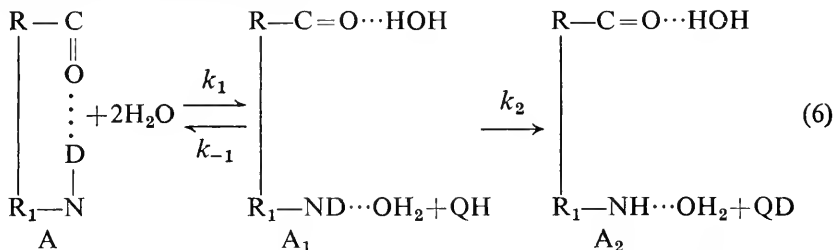
is a function of X , and an individual one, α_i , which is not, we get from Eq. (3)

$$\left(\frac{\delta \ln t}{\delta X}\right)_n = -\frac{\delta \ln B}{\delta X} \quad (4)$$

for

$$\beta_i = B\alpha_i \quad \text{and} \quad \frac{\delta \beta_i}{\delta X} = \alpha_i \frac{\delta B}{\delta X} \quad (5)$$

In our attempt to explain the rate curves at 0°C and different pH we have considered the simple reaction



where QH is a hydrogen donor or acceptor (H_3O^+ , H_2O , or OH^-) present in constant concentration at constant pH. We have here taken the primitive attitude to assume that the D-atom in the internal hydrogen bond exchanges only by way of the 'hydrolysed' form (A_1), the instability of the bond, and therefore the 'availability' of the deuterium atom, being measured by the ratio of the concentrations of forms A_1 and A at equilibrium.

Using the stationarity principle, we obtain the rate constant

$$\beta = k_2[\text{QH}] / \left\{ 1 + \frac{k_{-1}}{k_1} + \frac{k_2[\text{QH}]}{k_1} \right\} \quad (7)$$

and if the exchange reaction proper ($\text{A}_1 \longrightarrow \text{A}_2$) is considered the bottleneck of the over-all reaction, Eq. (7) is transformed into

$$\beta = k_e \times k_2[\text{QH}] \quad (8)$$

where

$$k_e = \frac{1}{1 + \frac{k_{-1}}{k_1}} = \frac{1}{1 + e^{\Delta F_r / RT}} \quad (9)$$

denoting by ΔF_r the free energy of 'hydrolysis' of the hydrogen bond in A.

Assuming now that the α -helix is a predominant secondary structure of PDLA in aqueous solution, we may utilize Eq. (9) in the following way: The complete helix of our PDLA contains a maximum of 26 internal hydrogen bonds with four end-group and three terminal imide hydrogen atoms exposed to water. Adopting the treatment given by Schellman^{17,18,19} and assuming that the opening of any hydrogen bond along the helix takes place exclusively through an unwinding of the helix from the ends, we obtain

$$\beta_i = \frac{1 - y^{14-i}}{1 - y^{14}} \cdot y^i \cdot k_2[\text{QH}] \quad (10)$$

for the rate constant assigned to the two bonds situated i places from the ends. The quantity y is $e^{-\Delta F_r/RT}$ and i may assume values from 1 to 13, the two halves of the helix being considered symmetrical ($n_i=2$). The calculations leading to Eq. (10) are rather trivial and will be given in greater detail elsewhere. The terms involving β_i for the seven deuterium atoms which are not hydrogen-bonded were assumed to be negligible.

It will appear from Eq. (10) that the more positive ΔF_r , the more the β_i will decrease toward the middle of the helix ($i=13$). It was therefore tempting to apply Eq. (10) to our results and see whether reasonable values for ΔF_r and $k_2 \times [\text{QH}]$ could be obtained on the basis of these very simple assumptions. It became immediately clear that it was impossible to fit this theory to the experimental data in an acceptable way. In order to satisfy the condition that deuterium atoms fully exposed to water (7 in our case) should exchange with a half-time too small to be measured, ΔF_r has to be given a high positive value so that hydrogen bonds only a few places away from the ends are suitably inhibited in their exchange. This choice of ΔF_r caused, however, the β_i of bonds nearer to the middle of the helix to become too small so that the calculated rate curves fell off too rapidly.

This negative result led us to consider the possibility of breaks in the interior of the helix as a means to increase the value of the internal β_i , and we have adopted the treatment of Schellman,¹⁸ p. 236, Eq. (9), putting

$$(\beta_i)_{break} = k_2[\text{QH}] \cdot 3 \cdot e^{-\Delta F_{break}/RT} \quad (11)$$

and taking into account that a particular hydrogen atom becomes exposed in three different breaks and that breaks occur not only in completely folded helices, but also in partly unfolded ones. Formula (11) is correct only if ΔF_r is large so that the concentration of intact helices is high, and even then it is a crude approximation. A more exact expression could be derived, but it is hardly worth the effort to introduce these highly complex corrections.

The possibility that deuterium ions are pulled out one by one from the middle of the helix and replaced by protium ions cannot be excluded entirely. Since, however, the mechanism of this process is not clear to us—except for the belief that it may lead to a term of the same type as Eq. (11), i.e. to β_i values which are approximately independent of i —we shall postpone the discussion of this point to a later publication.

The curves in Figs. 1 and 2 are calculated on the basis of Eqs. (11) and (10) combined, viz.,

$$B_i = \{e^{-i\Delta F_r/RT} + 3 \cdot e^{-\Delta F_{break}/RT}\} k_2[\text{QH}] \quad (12)$$

or

$$\beta_i = \{(0.376)^i + 0.00572\} k_2[\text{QH}] \quad (13)$$

neglecting higher powers of y in Eq. (10). The numerical value introduced for ΔF_r was +530 cal. and for $\Delta F_{break} = +3400$ cal., a somewhat lower value than that adopted by Schellman (4500), who neglected the entropy increase in the process of breaking. The calculation is in any case uncertain

considering especially the additional stabilization of the helix by non-polar bonds discussed below. In fact it mainly serves to show the necessity of introducing a third constant besides y and k_2 in the description of the kinetic curves. It will be observed that Eq. (12) is of the form $B \times a_i$ (Eq. (5)), so that the whole set of curves in Figs. 1 and 2 is determined by Eq. (13) introducing the following expression for $k_2[\text{QH}]$:

$$k_2[\text{QH}] = 6 \cdot 2 \cdot 10^{10}[\text{OH}^-] + 4 \cdot 6 \cdot 10^2[\text{OH}_3^+] + 0 \cdot 72 \quad (14)$$

It appears therefore from Figs. 1 and 2 that the exchange is 'catalysed' both by hydroxyl ions, hydrogen ions, and water ($[\text{H}_2\text{O}]$ is put equal to 1).

In the present stage of our investigation we shall refrain from going more fully into the mechanism of the exchange reaction proper, but a few words may be said about Eq. (14) which ought to apply generally to the exchange of deuterium atoms in ND groups which are exposed to water ($k = 1$). At the rate minimum, $\text{pH} \sim 2 \cdot 8$, $k_2[\text{QH}]$ assumes the value 1.7 corresponding to a half-time of 0.4 min., and the rate is therefore so high that a quantitative measurement by our method is impossible.

Table 1

DEUTERIUM EXCHANGE OF GLYCYLGLYCYLGLYCINE AT 0°C
($n_\infty = 5$)

at pH 3		at pH 5.6	
<i>t</i> (min.)	<i>n</i>	<i>t</i> (min.)	<i>n</i>
0.8	4.61	2.0	4.93
2.3	4.72	4.4	5.02
4.0	4.85	6.3	5.01
7.0	4.89	48.3	4.96
27.3	5.03	57.3	5.01
—	—	65.2	4.96

In general the process of dissolution of the lyophilized, deuterated peptide or protein takes about 15-30 seconds, and the first 15 μl . sample for deuterium oxide estimation is taken about 1-2 minutes after the start of the kinetic run. Our experiments with the short peptide DL-leucyltriglycine⁸ could therefore only demonstrate that all six of the exchangeable hydrogen atoms in this substance exchanged within one minute at 0°. The pH was, however, in the neighbourhood of 6 in this case, and in view of the strong dependence upon pH found for the exchange of PDLA, a reinvestigation of the exchange of short peptides at pH 2.5-3.5 and 0°C was indicated. Such a reinvestigation has been carried out by Dr. Wierzchowski in our laboratory using glycyglycylglycine (GGG) as test substance. Table 1 shows the results of two typical kinetic runs carried out in essentially the same way as the PDLA experiments. It appears from this table that the exchange

reaction is faster at pH 5.6 than at pH 3, but that even in the latter case about 85% of the 2 ND-groups have reacted with water in the course of 2 minutes. This finding is compatible with a half-time of 0.5 min. for the reaction of these groups at pH 3, but as an argument in favour of our value for $k_2[\text{QH}]$ it is not very strong from an experimental point of view, and is further considerably weakened by the fact that about 15% of ND-groups apparently exchange at a lower rate. This phenomenon requires further study but may be due to association of GGG in solution.

If, however, the value 530 cal. is adopted for ΔF_r , it is possible to calculate the degree of unwinding of PDLA in aqueous solution. Some of the calculations and results are summarized in Table 2, where the double indices refer to the two ends of the peptide (c_{xy} is the concentration of a molecule with x bonds opened at one end, and y bonds at the other).

Table 2
COMPOSITION OF PDLA-SOLUTION
For explanation see text

Type of molecules		Relative concentration, %
Number of hydrogen bonds opened	Concentration of different types of molecules	
0	c_0 (intact helix)	38.9
1	$c_{01} + c_{10} = 2yc_0$	29.3
2	$c_{11} + c_{02} + c_{20} = 3y^2c_0$	16.5
3	$c_{12} + c_{21} + c_{03} + c_{30} = 4y^3c_0$	8.3
4	$c_{22} + c_{13} + c_{31} + c_{04} + c_{40} = 5y^4c_0$	3.9
5	$c_{23} + c_{32} + c_{14} + c_{41} + c_{05} + c_{50} = 6y^5c_0$	1.8
total		98.7

Although these figures should be accepted with proper reservation considering their highly approximate character, they suggest an interesting qualitative picture of the state of PDLA in aqueous solution. 99% of the molecules have more than 21 intact hydrogen bonds out of 26, and PDLA may therefore be regarded as being essentially in the helical form. It is also inherent in the considerations leading to the value suggested for ΔF_r that the rate of opening and closing of the hydrogen bonds by unwinding the helix is very high, corresponding to a low free energy of activation. A molecular structure of this kind may be said to possess a high motility (a term suggested to the author at the Gordon Conference 1957) and any qualitative distinction between a native and a denatured state in aqueous solution is meaningless.

In a previous paper⁴ on the deuterium exchange of insulin, a different

attitude was taken to the question of the relative velocities of the reactions $A \longrightarrow A_1$, and $A_1 \longrightarrow A_2$ (see scheme (6)). It was assumed that $A_1 \longrightarrow A_2$ was the fast reaction so that Eq. (7) assumed the form

$$\beta = k_1 \quad (15)$$

and it was to k_1 , the rate constant for the unfolding, that high values of the energies of activation ΔH^\ddagger and ΔF^\ddagger were assigned. The truth probably lies somewhere between the assumption leading to Eq. (8) and to Eq. (15). In the case of PDLA the maximum free energy of unfolding the helix is that of breaking in the middle, *viz.*, 3000-5000 cal., and $\Delta F^\ddagger_{break}$, the free energy of activation of the breaking, must be at least that high and probably considerably higher. From the first term on the right side of Eq. (14) the standard free energy of activation of the hydroxyl ion catalysed exchange may be calculated from the theory of absolute reaction rates. The value obtained, *viz.* $\Delta F^\ddagger_{exch} = 4700$, indicates that somewhere in the alkaline region there is a definite chance that $k_2 \times [QH] \gg k_1$. In the case of insulin, where the breaking of the helices has to involve much greater energies, if our treatment used for PDLA applies here, the point where $k_2[QH] \gg k_1$ may be reached at lower pH. In agreement with this general picture the overall activation energy for the exchange of PDLA is found to be considerably lower than that found for insulin.^{4,9} The portion of the insulin molecule⁴ in which the ND-groups have half-times of exchange of 24 hours or more at 0° may then be said to have low motility. The easy way of exchange by unwinding the helix is blocked by S-S bonds and the hard one by breaking the helix is made harder by stabilizing interaction between sidechains.^{4,19}

Native β -lactoglobulin²⁰ in which about 100 hydrogen atoms out of 550 do not exchange at all during several days at 38°C at pH 5.4, represents a class of proteins with very low motility, *i.e.* low k_1 , but not necessarily low k_1/k_{-1} . Significant, irreversible denaturation occurs, if at higher temperatures (50-60°C) the firmly bound hydrogen atoms are forced to exchange. Some parts of the β -lactoglobulin molecule may therefore be said to be kinetically stable but thermodynamically unstable at lower temperature, and a relatively clear distinction between native and denatured forms may be made here. Unfortunately the interesting denaturing effect of high concentrations of urea at 0°²¹ could not be studied by back-exchange kinetics, because β -lactoglobulin could not be fully deuterated without damage to the protein. Otherwise experiments of this kind would have served to throw considerable light on the process of denaturation.

It remains now to find an explanation of the fact that PDLA exchanges so relatively slowly with water while the A-chain of insulin and even oxidized ribonuclease react almost instantaneously. As pointed out by Schellman¹⁸ and by Harrington and Schellman,¹⁹ the tertiary structure plays a very important role in the stabilization of helices, which when 'naked' seem to be very loose with a ΔF_7 close to zero. Since PDLA is unassociated in

aqueous solution, the stabilizing factor must be an intrinsic one. It is therefore natural to think of the methyl side groups which, if there is a random distribution of D and L forms along the chain, will protrude in two different directions from the core of the helix, and so that a quarter of them may form close pairs over the gap between two turns of the helix. The distance between such methyl groups, one reaching down from a given α -carbon of the helix, the other reaching up from an α -carbon in the next turn (separated from the first α -carbon by 3 peptide groups and 2 α -carbons), is 3.6 Å. reckoned from center to center of the β -carbon atoms. Since the average distance between the molecules in liquid methane at -164°C is of the order of 4 Å., there is reason to believe that a hydrophobic or nonpolar bond may be formed.

If we assume that ΔF_r for the naked helix is zero, we shall have to adopt the value of $4 \times (+530) \sim +2000$ cal. for the contribution of each nonpolar bond to the set of ΔF_r of PDLA (since only one-fourth of the group form pairs). With a random distribution of the pairs, it may be permissible as a first approximation to split the contribution in four parts of $+530$ assigned equally to four hydrogen bonds. An exact treatment would involve a complete statistical analysis and is therefore extremely complex.

It is difficult to say whether the value $+2000$ cal. is a reasonable one or not. Due to the small additional entropy change involved in the formation of the nonpolar bond in the helix, we may here have to do with a ΔH value. Figures of the order 1 to 2 kcal. have been suggested for nonpolar bond strengths of CH_2 groups,^{22,23,24} but it seems that very little has been done to determine experimentally the thermodynamic functions for the formation of nonpolar bonds in water. We shall therefore leave the question open.

The assumption concerning the internal stabilization of PDLA by nonpolar bonds is supported by the fact that poly-DL-serine of similar chain lengths (25 residues) exchanges 'instantaneously'. It is apparently contradicted by the recent findings^{14,15,25} that polypeptides of L-amino acids have a tendency to form righthanded helices, the stability of which decreases with increasing admixture of D-amino acids. Similarly D-amino acids have a preference for lefthanded helices. However, these important results were obtained with poly amino acids having more bulky sidechains than PDLA, so the question is not quite settled. It is clear, however, that our preparation must contain equal quantities of righthanded and lefthanded helices; the question is only whether D-alanine predominates in the lefthanded helices and L-alanine in the righthanded ones, in which case the frequency of methyl pairs along the helices may be greatly reduced. This question may be solved by digesting our sample with leucylaminopeptidase or carboxypeptidase, the action of which will stop whenever a D-alanyl group is encountered. In the extreme case, in which the preparation consists of pure poly-L-alanine (righthanded helix) and pure poly-D-alanine (lefthanded helix), these enzymes should split half of the peptide bonds (14-15 per mole). If, however,

the D- and L-forms are randomly distributed, only 1/2 to 1 peptide bond per mole will be opened. Preliminary experiments have given results which are in favor of a random distribution.

The author is indebted to J. Schellman, W. Kauzmann, and M. Ottesen for valuable discussions.

REFERENCES

1. AASE HVIDT, G. JOHANSEN, K. LINDERSTRÖM-LANG, and F. VASLOW, *Compt. rend. trav. lab. Carlsberg. Sér. chim.*, **29**, 129 (1954).
2. AASE HVIDT and K. LINDERSTRÖM-LANG, *Biochim. et Biophys. Acta*, **14**, 574 (1954).
3. AASE HVIDT and K. LINDERSTRÖM-LANG, *Biochim. et Biophys. Acta*, **16**, 168 (1955).
4. K. LINDERSTRÖM-LANG, *Chem. Soc. (London), Spec. Publ.*, **2**, 1 (1955).
5. AASE HVIDT, *Biochim. et Biophys. Acta*, **18**, 307 (1955).
6. AASE HVIDT and K. LINDERSTRÖM-LANG, *Biochim. et Biophys. Acta*, **18**, 308 (1955).
7. I. M. KRAUSE and K. LINDERSTRÖM-LANG, *Compt. rend. trav. lab. Carlsberg. Sér. chim.*, **29**, 367 (1955).
8. AASE HVIDT and K. LINDERSTRÖM-LANG, *Compt. rend. trav. lab. Carlsberg. Sér. chim.*, **29**, 385 (1955).
9. A. BERGER and K. LINDERSTRÖM-LANG, Sci. Program and abstracts of papers, Intern. Symposium on Macromolecular Chem. Israel, 1956, p. 52. *Arch. of Biochem. and Biophys.*, **69**, 106 (1957).
10. A. ELLIOTT, *Nature*, **170**, 1066 (1952).
11. C. H. BAMFORD, A. ELLIOTT and W. E. HANBY, *Synthetic polypeptides*, Academic Press, New York (1956).
12. P. DOTY, A. WADA, J. T. YANG, and E. R. BLOUT, *J. Polymer Science*, **23**, 851 (1957).
13. P. DOTY and R. D. LUNDBERG, *J. Amer. Chem. Soc.*, **78**, 4810 (1956).
14. P. DOTY and R. D. LUNDBERG, *Proc. Nat. Acad. Sci.*, **43**, 213 (1957).
15. J. T. YANG and P. DOTY, *J. Amer. Chem. Soc.*, **79**, 761 (1957).
16. M. SELA and A. BERGER, *J. Amer. Chem. Soc.*, **77**, 1893 (1955).
17. J. A. SCHELLMAN, *Compt. rend. trav. lab. Carlsberg. Sér. chim.*, **29**, 223 (1955).
18. J. A. SCHELLMAN, *Compt. rend. trav. lab. Carlsberg. Sér. chim.*, **29**, 230 (1955).
19. W. F. HARRINGTON and J. A. SCHELLMAN, *Compt. rend. trav. lab. Carlsberg. Sér. chim.*, **30**, 21 (1956).
20. K. LINDERSTRÖM-LANG, Soc. Biol. Chemists, India. Silver Jubilee Souvenir, 1955, p. 191.
21. C. F. JACOBSEN and L. KORSGAARD CHRISTENSEN, *Nature*, **161**, 30 (1948).
22. P. DEBYE, *J. Phys. and Colloid Chem.*, **53**, 1 (1949).
23. P. DEBYE, *Ann. N.Y. Acad. Sci.*, **51**, 573 (1949).
24. I. REICH, *J. Phys. Chem.*, **60**, 257 (1956).
25. A. ELLIOTT and B. R. MALCOLM, *Nature*, **178**, 1170 (1956).

The configuration of globular proteins in aqueous solution and its dependence on pH

CHARLES TANFORD

Department of Chemistry
State University of Iowa, Iowa City, Iowa

MACROMOLECULAR CONFIGURATIONS

Any long chain molecule can acquire a very large number of possible spatial configurations as a result of rotation about single bonds. Schematic diagrams representing some of the possible configurations of a linear polymer without branching points or cross-links are shown in Fig. 1. We are interested in this paper in asking which of these or other configurations will actually represent a given macromolecule in solution. The answer will clearly depend on the affinity of the segments of the macromolecule for one another, the affinity of solvent molecules for one another, and the affinity of chain segments for solvent molecules. It is profitable to define three extreme situations which may arise.

(1) *Random coiling*.²⁴ Molecules are said to be randomly coiled if all possible configurations have the same energy. In that event, all spatial configurations, such as those of Fig. 1, occur with equal probability. Randomly

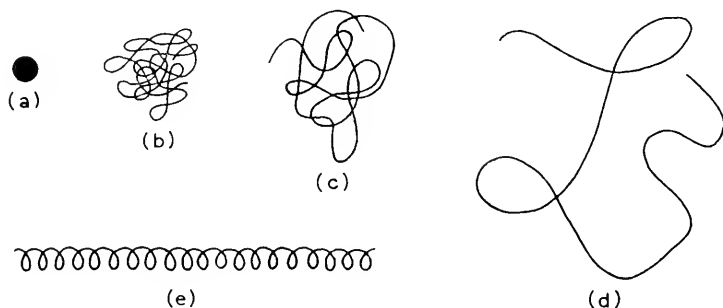


Fig. 1. Schematic representation of possible configurations of an unbranched polymer chain. The fully extended length is the same for each configuration.

coiled macromolecules are ordinarily described in terms of an *average* configuration. Fig. 1c may be taken as a schematic representation of this average configuration. Exactly random coiling can occur only if the reaction, 2 segment-solvent neighboring pairs \rightarrow 1 segment - segment neighboring pair + 1 solvent - solvent neighboring pair, has exactly zero free energy change, a situation which can arise only fortuitously ('ideal' solvent*). The class of randomly coiled polymers in solution is therefore ordinarily defined so as to include polymer-solvent systems in which there are slightly more ('good' solvent) or slightly fewer ('poor' solvent) segment-solvent contacts than would be predicted on the basis of statistics alone. The corresponding average configurations can be described schematically as moving from Fig. 1c towards Fig. 1d or Fig. 1b, respectively. Most of the common non-polar polymers, dissolved in non-polar solvents, are randomly coiled if this more flexible definition of the term is used.

(2) *Helical rods*. If the predominant attractive force in a polymer-solvent system is a specific attraction between a particular atom of *each* segment of a chain molecule for another particular atom of another segment, then the probable result is a helical configuration.^{37,58} The resulting overall shape is that of a long, thin rod, as illustrated by Fig. 1e. The most familiar helical configuration is the Pauling α -helix.⁵⁸ This configuration describes certain polypeptides in some solvents, e.g. poly- γ -benzyl glutamate in dimethylformamide.¹⁹

(3) *Compact, rigid, impenetrable spheres*. If the only important force in a polymer-solvent system is either solvent-solvent attraction or non-specific segment-segment attraction (or both) then segment-solvent contacts will be minimized, a result best achieved by insolubility of the polymer in the solvent. If, however, the polymer contains more than one kind of segment, and if some of the segments have strong affinity for solvent, segment-segment or solvent-solvent attraction being otherwise the predominant force, then the polymer will be soluble and its optimum configuration will be one which minimizes segment-solvent contacts except for those segments which have strong affinity for solvent. An idealization of the resulting configuration is a compact, rigid sphere, impermeable to solvent (Fig. 1a). The segments attracted to solvent are on the outside. The remainder fill the interior space, which is devoid of solvent.

The three extremes just discussed were chosen because they may be unequivocally distinguished by almost any physical measurement. For instance, compact rigid spheres have much smaller radii of gyration and much larger

* The definition here given in terms of the stability of individual segment-solvent pairs is, of course, an oversimplification. Flory²⁴ should be consulted for a full discussion of this subject.

sedimentation and diffusion coefficients than molecules of the same molecular weight which are random coils or rigid rods. To distinguish between random coils and rigid rods is equally easy if the macromolecule under consideration can be prepared with a range of molecular weights. The radius of gyration, for example, would then be proportional to the molecular weight for rods and roughly to the square root of the molecular weight for random coils. For naturally-occurring substances of fixed molecular weight the distinction between rods and random coils generally requires more subtle analysis. The present paper is concerned only with macromolecules bearing electric charges. It will appear subsequently that such charges have a considerable effect on the configuration of randomly-coiled molecules, enabling us to distinguish them readily from rigid rods.

The simplest physical measurement which will distinguish between the extreme configurations here under discussion is the intrinsic viscosity, $[\eta]$. For rigid spheres of unit density $[\eta]=2.5$ cc./gram, independent of molecular weight. For rigid rods or random coils, $[\eta]$ is much larger: it varies roughly as M^2 for the former and as $M^{0.5-1.0}$ for the latter (where M =mol. wt.).

It would be naive, of course, to expect all macromolecules to fall into one of the extreme configuration classes just discussed. Intermediate configurations which undoubtedly exist are helical rods with a small number of breaks, random coils with some cross-links, compact spherical regions joined by short lengths of randomly-coiled chain, etc. To distinguish these from one another in an unequivocal way is at present not possible.

ISOELECTRIC GLOBULAR PROTEINS IN AQUEOUS SOLUTION

It is clear from the preceding discussion that the configuration of macromolecules in solution is determined by the properties of the solvent as much as by those of the macromolecules. Thus, the configuration of a protein molecule is likely to differ from one solvent to another, and in any solvent may differ from the configuration in the crystalline state. This paper will concern itself with the configuration of common globular proteins in dilute aqueous solution. The conclusions reached will not apply to crystalline proteins or to proteins in other solvents. Indeed, it is well-established that configurations of proteins in solvents such as concentrated aqueous urea differ markedly from what they are in water.

When many of the common globular proteins, in aqueous solution, are examined at or near their isoelectric points, they can usually at once be assigned to a configuration much closer to that of a compact, rigid, impenetrable sphere than to the other extremes mentioned earlier. This is illustrated, for example, by the intrinsic viscosities listed in Table 1. A similar conclusion is reached from all other pertinent measurements.

It should be noted that three of the proteins listed in Table 1, fibrinogen,

collagen and myosin, obviously do not resemble compact spheres in aqueous solution. The configuration of these proteins thus presents a problem different from that of the other proteins listed. We shall not discuss these proteins in this paper. It is worth noting, however, that their different configuration is not primarily a result of their high molecular weight. Other proteins of high molecular weight, such as fumarase²⁹ and catalase,^{13,83} appear from sedimentation and diffusion studies to have a compact configuration near their isoelectric points.

Table 1

INTRINSIC VISCOSITIES OF ISOELECTRIC GLOBULAR PROTEINS
IN AQUEOUS SOLUTION

	Mol. Wt.	$[\eta]$, cc./g.
Rigid impenetrable sphere ^a	any	1.9
Random coil ^b	50,000	47
Random coil ^b	560,000	91
Helical rod ^c	66,000	45
Helical rod ^c	340,000	262
Ribonuclease ¹⁰	13,683	3.3
Myoglobin ¹¹⁰	17,000	3.1
β -Lactoglobulin ⁹	35,000	3.4
Pepsin ²¹	35,000	3.4
Ovalbumin ⁶³	44,000	4.3
Serum Albumin ⁹²	65,000	3.7
Hemoglobin ^{17, 89}	67,000	3.6
Conalbumin ⁶²	77,000	3.8
Fibrinogen ^{35, 75}	330,000	27
Collagen ⁵	345,000	1150
Myosin ³⁶	550,000	230

(a) Unsolvated sphere of density 1.33, by Einstein's equation.²²

(b) Polyisobutylene in cyclohexane.²⁵

(c) Poly- γ -benzyl glutamate in dimethyl formamide.¹⁹

Two questions arise in connection with those proteins which do possess a compact, close to spherical shape. The first (and the less interesting) concerns the meaning of the relatively small deviations of the observed viscosity (or of other pertinent properties) from the theoretical values for perfect rigid, solvent-free spheres. Part of the deviation undoubtedly results from incorporation of some solvent in the dissolved particle.*.¹⁰⁴ Deviations from spherical shape¹ and non-rigidity presumably account for the remainder. Attempts to obtain quantitatively the contribution of each of these factors

* It is important to note, however, that the water so incorporated must be largely bound or trapped water. That the protein particle is not permeable to the solvent as a whole is shown unequivocally by the effect of ionic strength on titration curves.⁸⁷ For compact proteins this effect is much smaller than it would be if salt ions could enter the space between charges, where they would most effectively shield interaction between charges.

have had poor success.⁵⁶ (See, for instance, the criticisms^{32,88,91,104} of a recent attempt by Scheraga and Mandelkern.⁷³)

More interesting than the small deviations from the ideal spherical model is the question why the behavior of these proteins is as close as it is to the behavior of rigid spheres. In other words, what are the forces which maintain so compact a structure? As was pointed out earlier, these can be attractive forces between solvent molecules or attractive forces between segments of the polypeptide chain, or both. In addition, the fact that aqueous solutions of these proteins are stable at all requires the existence, at the surface of the molecular particle, of a few segments of the polypeptide chain which have strong affinity for solvent. The following pages will list in turn each of the forces which may make a contribution, and will try, on the basis of general chemical principles, to assess the relative importance which each might have. Most of the points which will be brought out have been made before, notably in a recent review by Kauzmann.⁴⁴ They bear repeating, however, as a deterrent against the dangerous (though popular) practice of ascribing the stability of compact proteins to one force alone.

Attraction between solvent molecules. One of the relevant forces which without doubt contributes to the stability of compact configurations is the hydrogen bonding between water molecules. This force repels non-polar groupings since these would create holes in the hydrogen bonded structure of the solvent. In aqueous detergent solutions this force acts *intermolecularly*, causing the non-polar portions of detergent molecules to aggregate, with the formation of micelles. In proteins there are many non-polar side chains on a single molecule. The same force is therefore expected to act *intramolecularly* and to lead to a compact structure with the non-polar side chains in the center. Kauzmann⁴⁴ has used the term 'hydrophobic' bonding to describe the intramolecular attraction which results from this force, and we shall employ the same term.

Electrostatic forces between charged segments of the protein chain. Isoelectric proteins have a net charge of zero. The molecules possess, however, a large number of charges, both positive and negative. Coulombic forces between these charges must exist, and they provide a second factor which inevitably affects the stability of a compact configuration. The contribution of the interaction between these charges to the free energy of a compact protein molecule can be calculated by the method of Kirkwood,⁴⁷ which takes into account the fact that the dielectric constant within the molecule is different from that of the solvent. Kirkwood's method has recently been adapted by the author^{97,90} to estimation of the electrostatic interaction energy of spherical protein molecules with specified locations of charges. In this method, the molecule is treated as being impermeable to solvent and the charges are taken to be at the surface, in contact with solvent, such that the distance

of closest approach between a charge and a water molecule is the same as it would be in a low molecular weight organic ion in aqueous solution.

Fig. 2 shows three model structures for which calculations have been made for this paper. All calculations were averaged over all configurations resulting from permutation of charges over identical sites. They are approximate in that the Bragg-Williams method¹⁰¹ was used. (This is equivalent to retention of only one term in equation 33 of Tanford and Kirkwood.⁹⁷) For each model, two isoelectric forms are likely to occur, one with protons removed from every carboxyl group and from no other groups, the second with protons removed from all but one of the carboxyl groups and from one of the imidazolium groups. The corresponding electrostatic interaction

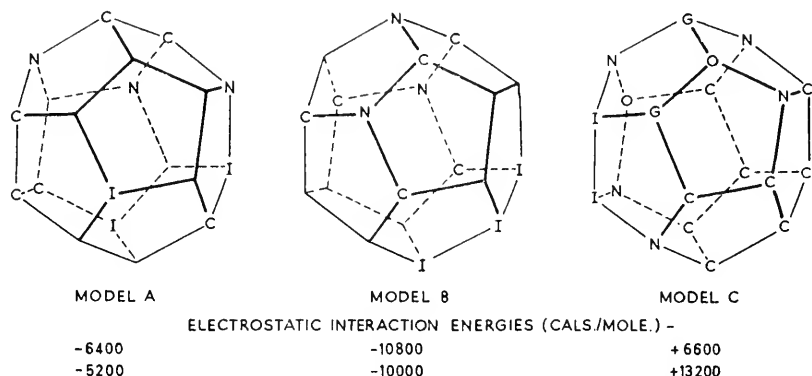


Fig. 2. Models used for calculation of the electrostatic contribution to the free energy of compact configurations. The letters C, G, I, N and O represent, respectively, carboxyl, guanidine, imidazole, amino and phenolic groups. The dodecahedra are inscribed in spheres of radius 10 Å, with the vertices 1 Å below the surface, this being the distance of closest approach of solvent to the charged sites on any organic acid or base.⁹⁰ The distance between nearest neighboring vertices is 6.5 Å.

energies are: for model A, -6400 and -5200 calories/mole; for model B, -10,800 and -10,000 calories/mole; for model C, +13,200 and +6600 calories/mole. These calculations are for isolated protein molecules at zero ionic strength. The addition of inorganic salts reduces these values, but for impermeable molecules, the effect is relatively small, as Table 2 shows, because the salt ions cannot enter the space between charges, where they would be most effective.

These calculations clearly show that the contribution of electrostatic interaction to the free energy of isoelectric protein molecules can be very large, and that the magnitude of the effect depends markedly on the positions of charged sites relative to one another. The interaction energy may be negative or positive. However, models leading to positive interaction energies (such as model C) require unlikely arrangement of the charges with most of the positive charges well separated from most of the negative charges.

The fact that most protein molecules in solution have very low permanent dipole moments,⁴³ strongly suggests that positive and negative charges are in fact more or less evenly distributed. Such an arrangement always leads to a net negative electrostatic interaction energy for the isoelectric state, with the result that electrostatic forces contribute materially to the stability of compact configurations.*

Further calculations for these same models will be reported in Table 2. It should be noted that the isoelectric point is not necessarily the point at which the electrostatic interaction energy is a minimum. For model C, for instance, the minimum occurs when the net charge is -4 .

Intramolecular hydrogen bonds. A third force is provided by intramolecular hydrogen bonds.^{54,18,52,60} It is convenient to distinguish between two kinds of intramolecular hydrogen bonds. The first is the familiar peptide hydrogen bond ($-C=O \dots HN=$).^{37,58} The second kind is the hydrogen bond between side-chain groups.⁵² Examples of the latter are the bond between undissociated carboxyl groups, well known in the dimeric form of carboxylic acids in some nonaqueous solvents, and the frequently postulated phenol-carboxylate bond. Such hydrogen bonds are probably the principal factor determining the configuration of proteins in the solid state.⁵⁹ In aqueous solutions, however, there is a competition between formation of such intramolecular bonds and formation of hydrogen bonds between the same groups and water molecules, so that their effect on configuration is likely to be reduced. When appropriate small molecules are studied, to which the same kind of competition applies, equilibrium is found to lie on the side of hydrogen bonds to water. Thus Schellman⁷⁰ has calculated, from the thermodynamic behavior of urea in water solution, that in such solutions the association of two urea molecules by hydrogen bonds of the peptide type is accompanied by a *positive* free energy change ($\Delta F^\circ = +1990$ calories/mole). Similarly, for the association of tyrosine with acetate¹⁰⁷ by formation of phenol-carboxylate hydrogen bonds, $\Delta F^\circ > 1400$ calories/mole.

Of course, these model reactions are between small molecules which are free in solution, and are therefore accompanied by a loss of translational entropy. The formation of the corresponding bonds between segments of a polypeptide chain involves in place of this factor a loss of rotational entropy. The difference between these quantities could well be sufficiently large to change the sign of the free energy of formation of such hydrogen bonds in aqueous protein solutions. In fact, Schellman⁷¹ made a calculation to show that this is just what would happen in the case of peptide

* The contribution of electrostatic forces to the stability of a compact configuration is, of course, the difference between the figures calculated and the corresponding interaction energies in a randomly coiled configuration. The latter, however, are comparatively small because the space between charges in solvent-permeated configuration is filled by material of high dielectric constant.

hydrogen bonds, but in a later paper,³³ this prediction was withdrawn. For formation of phenol-carboxylate hydrogen bonds in proteins, Laskowski and Scheraga⁵² have calculated a large loss of rotational entropy which is of the same order of magnitude as the probable loss of translational entropy in the tyrosine-acetate reaction mentioned above.

Perhaps even more important is the fact that hydrogen bonds in proteins can occur in conjunction with other intramolecular forces, such as hydrophobic forces. For instance, when several amino acid residues, adjacent in sequence, are all non-polar, then they will all tend to be located in the interior of the molecule and will carry with them the part of the peptide chain to which they are attached. This part of the polypeptide chain, being removed from contact with water, will presumably tend to form peptide hydrogen bonds, and the ability to form such bonds will contribute to the stability of the hydrophobic bonding.*

That hydrogen bonds are in fact present in the compact forms of globular proteins is strongly suggested by several kinds of experiments. The determination of optical rotatory power, for example by Yang and Doty,¹¹¹ indicates that globular proteins in aqueous solutions are partially in the form of helices. This almost certainly reflects the presence of peptide hydrogen bonds. The deuterium-hydrogen exchange experiments of Linderström-Lang and Hvidt^{38,53} show that some exchangeable hydrogen atoms are more slowly exchanged than others. This result could arise if peptide or amide groups (for example) were in the hydrophobic interior of the molecule without hydrogen bonding, but it is usually assumed that if such groups are so located they will also tend to be hydrogen-bonded. Finally, it will be seen below that in at least two proteins the stability of the compact form towards pH suggests the presence of hydrogen bonds or intramolecular ion pairs between titratable side-chain groups. It is of interest in this connection that the titration curves of some simple dicarboxylic acids¹⁰⁶ with bulky non-polar side chains indicate the presence of weak intramolecular bonds between $-\text{COOH}$ and $-\text{COO}^-$.

Intramolecular ion pairs.† The attraction between positive and negative ions leads to close association between them in the crystalline state and to the formation of ion pairs in non-aqueous solvents. In water, however, the free energy of hydration is sufficiently negative to prevent ion pair formation, at least between univalent ions. Even guanidinium and acetate ions, the association of which would be favored by hydrogen bonds as well as electro-

* Hydrogen bonds can be important even if they do not contribute to thermodynamic stability by blocking kinetic pathways to more extended configurations. For example, the activated state for unfolding could be one in which intramolecular hydrogen bonds are ruptured, but in which new hydrogen bonds to water are not yet formed.

† This new term is suggested as more descriptive than the term 'salt bridge' which has sometimes been used in the past to describe this kind of bond.

static attraction, do not associate appreciably.⁸⁵ However, as in the case of hydrogen bonds, there will be a difference in entropy between the formation of intermolecular ion pairs between small ions and the formation of corresponding bonds between protein side chains of opposite charge. There is no experimental evidence at present which would enable one to arrive at a decision concerning the presence of such intramolecular ion pairs in compact protein molecules or on their contribution to the maintenance of a compact structure. (In calculating the energy of electrostatic interaction earlier in this paper, the absence of ion pair formation was assumed: the closest distance of approach of two charges in the models used is 6.5 Å.)

Hydrophilic groups at the surface. We have discussed the attractive forces between solvent molecules and the attractive forces between segments of the polypeptide chain, which, acting alone or together, could stabilize a compact configuration for protein molecules in aqueous solution. As was pointed out earlier, an additional requirement for the stability of such a configuration is the presence of a small number of surface sites with strong affinity for the solvent. For proteins in aqueous solution, the side chain groups which bear charges would seem most likely to fill this requirement. Indeed, there is strong evidence from titration curves that most of these groups are in fact in contact with solvent in every globular protein. Ribonuclease, for example, maintains its compact configuration through most of the accessible pH range (see below). The titration curve⁹⁵ shows that every one of the carboxyl, imidazole and amino groups (but not all phenolic groups: see below) which this protein is known to possess is titrated reversibly in this pH range. Moreover, there is no conflict between the observed titration curve and a calculated curve based on (a) intrinsic dissociation constants identical with these found in appropriate substances of low molecular weight, (b) electrostatic interaction calculated with the assumption that charges are as close to the surface as in low molecular weight organic acids. Since any variation in the closeness of charged sites to the solvent results in large changes in electrostatic self-energy,⁹⁷ the presence of as few as one or two of these titratable groups in the interior of the molecule would at once be detected, either by failure to titrate these groups at all or at least by pronounced shifts in their dissociation constants.

Ovalbumin is another protein which maintains a compact configuration over most of the accessible pH range. The titration curve¹¹ again requires that essentially all charged sites be at the surface. The same conclusion applies with somewhat less force to most other proteins. Many of these do not maintain their compact configuration over a sufficient range of pH to allow observation of the actual titration of the majority of charged sites in the absence of configurational change. One can often, however, deduce the presence of titratable groups with normal properties even if they are not titrated directly. It is unlikely that any of the small globular proteins has

as many as ten per cent of its carboxyl, amino, imidazole or guanidine groups anywhere other than in direct contact with solvent.

Hemoglobin has been cited as a possible exception to this generalization.^{79,81} This protein undergoes a reaction below pH 4 which is accompanied by the sudden binding of 18 protons per molecule (molecular weight 33,500). The protein acts as if 18 carboxyl groups in their anionic form were *not* available to titration above pH 4. This conclusion is erroneous, however. The reaction of hemoglobin below pH 4 is accompanied by breakdown of the compact structure (see below). As a result, there is a decrease in electrostatic interactions and this necessarily results in the sudden binding of about 15 protons⁸⁹ to sites which were available all the time and were not occupied by protons solely because of the repulsive force of other positively charged sites. Thus, if any of the carboxylate side chains of hemoglobin are not available to titration, the number must be quite small. It is probable that the explanation here given accounts for the apparent unavailability of prototropic sites in some other proteins.⁸¹

Of the side chain groups subject to titration with acid or base, there is one kind which has been found 'unavailable', i.e. presumably in the interior, in some proteins. These groups are the phenolic groups. In ovalbumin¹⁸ all of these groups are unavailable as long as the compact configuration is maintained; in ribonuclease⁹⁶ three out of six phenolic groups are unavailable; in conalbumin,¹⁰⁸ too, about half are unavailable. Phenolic groups, however, are not charged at the pH at which these proteins exist in nature. Uncharged phenolic groups have no particular affinity for water and to find some of them in the 'hydrophobic' interior of protein molecules is therefore no contradiction of the generalization made in this section.

Disulfide bonds. The discussion so far has made only incidental reference to the effect of cross-linkages on the configuration of polymeric chain molecules. Most proteins possess such links in the form of disulfide bonds. Their principal effect is to limit the range of possible *extended* configurations: configurations such as illustrated by Fig. 1*d* may become unattainable as long as disulfide bonds are intact. This factor will indirectly increase the stability of a compact configuration, as compared to that of a randomly-coiled configuration. The chief factor favoring the latter is its high configurational entropy. The presence of cross-links, however, decreases the number of possible configurations available to the randomly coiled state, and hence reduces the entropy gain which would result from a transition to this state.

There is another possible effect of the presence of disulfide cross-links which works in the opposite direction. To gain maximum benefit from 'hydrophobic' bonds, electrostatic interactions, etc., requires freedom to bring appropriate portions of a polypeptide chain into each other's vicinity. Disulfide bonds interfere with this freedom so that their presence may make a compact structure less stable than it otherwise would have been.

There is no way of predicting which of these effects will predominate. If ribonuclease is oxidized with performic acid so that all disulfide bonds are broken, it loses its compact configuration even at the isoelectric point.³³ On the other hand, ovalbumin, which contains just a single disulfide bond in its natural state is one of the proteins most resistant to loss of its compact form, whether by acid, base or urea. Finally, serum albumin, with about 16 disulfide bonds, loses its compact configuration more readily than most proteins. Another aspect of the presence of disulfide bonds is that they are likely to facilitate the reversibility of configurational changes, as Kauzmann⁴⁴ has pointed out.

It is evident from the preceding discussion that no basic principles of chemistry and no conclusive experimental tests allow us to state precisely what forces maintain a compact configuration for so many isoelectric proteins in aqueous solution. Much experimental work is currently going on in many laboratories towards the elucidation of this problem. The most fruitful approach has been to study the conditions under which compact configurations break down. Such breakdown occurs quite generally when the solvent is changed, when an aqueous solution is heated, when acid or base is added, etc. It is usually called denaturation, although some definitions of this term would not include rapidly reversible changes in configuration. A recent general review by Putnam⁶⁴ and the paper by Kauzmann⁴⁴ earlier referred to provide a summary of results which have been obtained. Unfortunately, these results have not as yet progressed to where they can provide an answer to the questions which we have posed. In general, kinetic studies of denaturation predominate. There have been few attempts, however, to characterize the configuration of the denatured form or to determine the thermodynamics of the denaturation process.* The studies have gone far enough, however, to indicate that every protein behaves in an individual way towards various agents which cause configurational change. As a result, it may well be that different proteins possess a compact configuration for quite different reasons.

The remainder of this paper will summarize recent information obtained by the present author and co-workers, and by others, on configurational changes which accompany the titration of proteins in aqueous solution. The results will serve to illustrate the conclusion that we are still far from a solution of the problems which we set out to solve.

THE EFFECT OF pH ON CONFIGURATION

When acid or base is added to an isoelectric protein solution, three things

* The lack of thermodynamic data results in large part from the fact that denaturation has usually been studied under irreversible conditions. An unfortunate consequence is that we do not know for most proteins whether the observed range of stability of a compact configuration reflects conditions under which this configuration is truly the form of lowest free energy, or whether maintenance of compactness is merely a kinetic phenomenon.

happen: (1) the pH of the solution is changed, (2) protons are removed from or added to appropriate acidic or basic groups, (3) the net charge of the protein molecule is increased. Of these changes, the last-named is the one most likely to affect the configuration, as was pointed out as early as 1931 by Wu.¹⁰⁹ For the increase in net charge necessarily results in repulsion between charges and a net positive contribution of electrostatic interaction to the free energy of a compact configuration. Some calculations for the models of Fig. 2 are shown in Table 2. They were made using the same equations and assumptions discussed earlier. Table 2 shows changes in electrostatic interaction energy between 10,000 and 20,000 calories for the models used, which contain 12 or 20 charged sites. The corresponding changes for real proteins can easily amount to 100,000 calories/mole.

Table 2

THE CONTRIBUTION OF ELECTROSTATIC INTERACTION TO
THE FREE ENERGY OF COMPACT SPHERICAL PROTEINS

Net Charge	Interaction Energy, cal./mole*				
	$\mu=0$	Model A $\mu=0.015$	$\mu=0.060$	Model B $\mu=0$	Model C $\mu=0$
+9					+25340
+6	+7530	+5880	+5110	+ 8510	+14100
+5					+11950
+4	+ 380	- 30	- 170	+ 250	+10570
+3	-2260	-2220	-2130	- 3190	+10030
+2	-4260	-3880	-3610	- 6170	+10270
+1	-5640	-5030	-4640	- 8700	+11320
0	-6390	-5650	-5200	-10770	+ 5210
	-5240	-4630	-4260	-10030	+ 6610
-1	-5440	-4820	-4440	-11630	+13150
					+ 8800
-2	-4060	-3670	-3400	-11110	+ 3660
					+ 6200
-3	-2270	-2220	-2120	- 9210	+ 5970
	-1210	-1170	-1100	- 6570	+ 3920
-4	+1100	+ 700	+ 550	- 4350	+ 8510
					+ 1960
-5					+ 7060
					+ 5910
-6	+9410	+7750	+6930	+ 6860	+10610
-9					+15240
					+25930

* The model structures to which these calculations apply are shown in Fig. 2. Where two values are shown for the same net charge they refer to alternate forms which are likely to be present with that net charge. For model C, for example, with a charge of -4, the first figure is for the form with protons removed from two phenolic groups, the second for the form with protons removed from one phenolic and one amino group. μ stands for ionic strength.

The effect of ionic strength on interaction energies, for impenetrable models such as these, is quite small. This is shown by the calculation of this effect for model A. It should be noted, incidentally, that the choice of impenetrable models is not an arbitrary one. As already mentioned, it is forced on us by the small experimentally observed effect of ionic strength on titration curves of compact proteins.⁸⁷

Table 2 indicates that, whatever the forces stabilizing a compact configuration near the isoelectric point, electrostatic interaction is the principal force which makes such a configuration unstable as one moves away from the isoelectric region. For the strength of 'hydrophobic' bonding, of peptide hydrogen bonds and of disulfide bonds is not affected at all by pH changes. Hydrogen bonds between side chain groups and intramolecular ion pairs are affected, since titration of the side chain groups involved in such bonds would lead to their weakening or rupture. However, all the energy contributed in this way to a shift in the free energy balance between a compact and an extended configuration must be reflected by an exactly equal anomalous free energy of ionization of the groups concerned. Whether such anomalies in fact occur, is debatable. If one assumes a uniform density of charged sites on the surface of the protein molecule, then anomalies certainly exist.^{52,86} If, on the other hand, one allows for reasonable variation from one protein to another in the relative locations of charged sites, then such apparent anomalies can be explained on the basis of electrostatic interaction of these sites with the acidic or basic groups under consideration.⁹⁰ In any event, any instability contributed by hydrogen bond rupture would be an addition to and not a substitute for the instability due to electrostatic forces.

It should be noted again, however, that, though Coulombic forces probably provide the principal basis for the *thermodynamic* instability of a compact configuration, hydrogen bonds may block the kinetic pathway to a configurational change. It is not impossible to visualize mechanisms by which a change in configuration is acid- or base-catalysed so that a change in pH per se may affect the rate at which a change in configuration can occur.

Before considering the effect of pH on globular proteins, it is of interest to examine its effect on a simple synthetic polyelectrolyte, such as polymethacrylic acid. This polymer is isoelectric when all of its carboxyl groups are un-ionized. It then has no charges at all, and is randomly coiled. When it becomes charged, by ionization of the carboxyl groups, Coulombic repulsion sets in. Relatively compact configurations, such as that represented by Fig. 1*b*, acquire a much higher electrostatic interaction energy than relatively extended configurations, such as that represented by Fig. 1*d*. The average configuration therefore becomes much more extended, and this is reflected by corresponding changes in viscosity, sedimentation coefficient, etc. (A theoretical treatment of this effect has recently been presented by Harris and Rice.^{34,67} The same authors have also considered polyampholytes with equal numbers of acidic and basic groups.⁶⁸) Experimental verification of

this effect is provided by the viscosity data of Arnold and Overbeek³ for polymethacrylic acid, which are shown in Fig. 3.

An increase in ionic strength has a much greater effect on electrostatic interactions in configurations in which the molecular domain is permeated by solvent than it does for the impenetrable configuration to which the calculation of Table 2 applies. As a result, a moderate ionic strength suffices to eliminate electrostatic forces almost entirely with a resultant return to essentially the randomly coiled configuration of the uncharged molecule. This effect is also illustrated by Fig. 3.

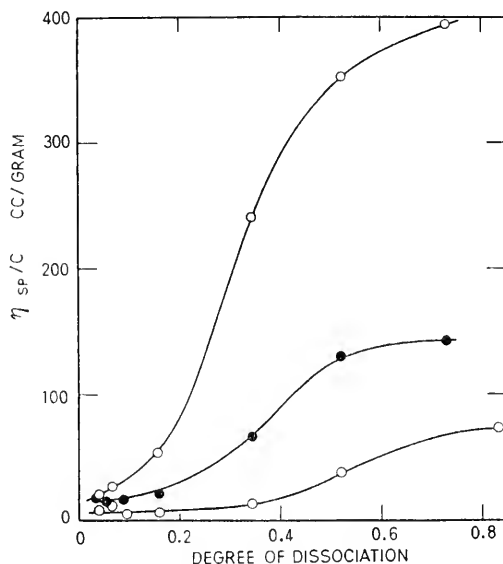


Fig. 3. Reduced viscosities in dilute solutions of polymethacrylic acid (Arnold and Overbeek³). Ionic strength is 0.010 to 0.017 for the top curve, 0.10 for the middle curve, and 1.0 for the lowest one. The concentration c was 0.00085 g./cc. in all experiments.

Exactly parallel behavior on the part of globular proteins is, of course, not to be expected. Isoelectric polymethacrylic acid is randomly coiled, i.e. all possible configurations have approximately the same energy. Thus, even small energy changes in the more compact configurations are immediately (and reversibly) reflected in an increase in the proportion of molecules with more extended configurations. In the globular proteins here under discussion, on the other hand, the various possible configurations in the isoelectric state do not have identical energy. Instead, the most compact configuration is stabilized by what may be a large energy inherent in the various forms of intramolecular attraction described earlier in this paper. The effect of titration is to make the electrostatic contribution to the free energy a positive one (and, perhaps, to change the strength of specific bonds between side chain groups). Only when the resulting change in free energy becomes

sufficiently large to overcome the energy initially stabilizing a compact form will transition to an extended randomly coiled configuration become possible. When a transition does occur, the configurational change will be much more drastic than the shift in average configuration which takes place in simple polyelectrolytes, with the result that the transition may be a slow process and may not occur under reversible conditions. Another difference between the anticipated behavior of globular proteins and the behavior of simple polyelectrolytes is that the former may be prevented by the presence of disulfide bonds from acquiring configurations as extended as those of simple polyelectrolytes.

It has been noted already that the stabilization of a compact configuration for isoelectric proteins depends markedly on the arrangement of charged sites, and that disulfide bonds may have a stabilizing effect in some proteins and not in others. 'Hydrophobic' bonding is clearly subject to similar variation, and so is any stabilization energy which may result from hydrogen bonding or intramolecular ion pairs. It is clear that some proteins may be expected to withstand a much higher net charge without configurational change than others, and this proves indeed to be the case.

Ribonuclease¹⁰ is an example of a protein with unusual stability in this regard. Its intrinsic viscosity (Fig. 7) remains essentially equal to that observed at the isoelectric point from pH 2 to pH 11.5, and in this range it is essentially independent of ionic strength. The small changes (<10%) in $[\eta]$ which do occur are of the order of magnitude to be expected for the effect of charge ('electroviscous effect') on the viscosity of rigid impenetrable spheres.⁶

We shall now list some globular proteins which are known to undergo a major change in configuration as a result of pH changes. Many of the examples given are well-known instances of acid- or base-induced 'denaturation'. The most thoroughly studied example (serum albumin), however, represents a very rapid and reversible reaction which would not ordinarily be classified as a 'denaturation'. We shall not, on the basis of the data available, be able to shed much light on the forces maintaining the initially compact configurations. We shall be able, however, to demonstrate that the product of the pH-induced configurational change, in every instance which has been subjected to experimental test, is, as anticipated, a flexible molecule exhibiting behavior characteristic of simple linear polyelectrolytes.

Two criteria of such behavior will be used. The first is the expansion of the molecular domain with increasing net charge and/or the contraction, at constant charge, with increasing ionic strength. This behavior is illustrated by the viscosity data for polymethacrylic acid shown in Fig. 3. Sedimentation or diffusion coefficients provide the same information.

The second criterion is based on the titration curve. In particular, we shall examine the dependence on net charge of the quantity $\text{pH} - \log \alpha / (1 - \alpha)$, in which α represents the experimentally observed degree of dissociation of

a particular kind of group on the molecule. This quantity is an effective pK for the hydrogen ion dissociation of the kind of group being considered, and, at constant ionic strength and temperature, it would be a constant quantity independent of net charge if there were no interaction between the many acidic and basic sites present on a protein or simple polyelectrolyte molecule. Such interaction does in fact exist, however, as a result of ordinary Coulombic forces between electrostatic charges.

Consider first a globular protein in its compact impenetrable configuration, or any other large molecule in which the distances between sites and the penetrability by solvent do not change during the course of titration. Then each proton added to the molecule will provide a repulsive force for all subsequent protons, and each proton removed will have the opposite effect. Thus $pH - \log \alpha/(1-\alpha)$ will decrease with increasing net positive charge and increase with increasing net negative charge.⁸⁶ The effect will be greatest at low ionic strengths, but for impenetrable configurations will persist to high ionic strengths. The resulting behavior is illustrated by Fig. 5, which is obtained from the titration of carboxyl groups of ribonuclease,⁹⁵ a protein known from viscosity measurements to remain compact

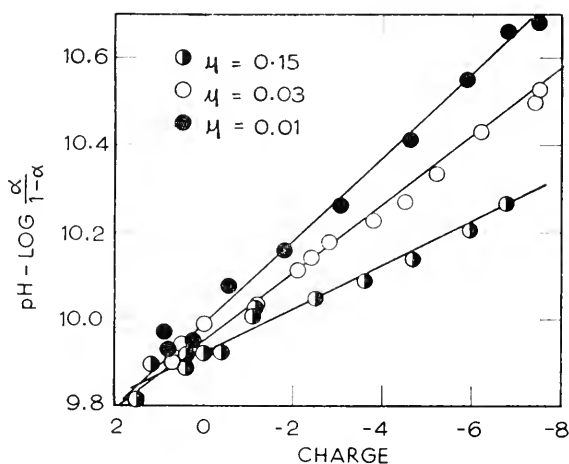


Fig. 4. Logarithmic plot for titration of the free phenolic groups of ribonuclease (Tanford, Hauenstein and Rands⁹⁶).

and impenetrable throughout the pH range under consideration, and by Fig. 4, obtained from the titration of the three surface phenolic groups of the same protein.⁹⁶

For simple flexible polyelectrolytes the effect of Coulombic interaction is quite different. The first few protons added or removed exert the expected effect. But each increment of net charge results, as we have seen, in a greater extension of the molecular domain and this in turn leads to an increase in

the distance between charges and to the presence of more shielding electrolyte between them. The typical result, illustrated by Fig. 6, is that $\text{pH} - \log \frac{\alpha}{1-\alpha}$ becomes virtually independent of net charge.³

Two additional criteria will be used as circumstantial evidence for the absence of configurational change. These are optical rotatory power and the rotational relaxation time of bound fluorescent dye molecules. These properties are quite sensitive to configuration, although they do not lend

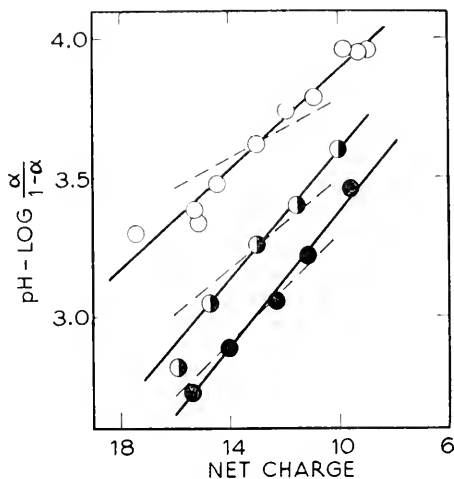


Fig. 5. Logarithmic plot for titration of the carboxyl groups of ribonuclease (Tanford and Hauenstein⁹⁵). Ionic strength is 0.15 for the top curve, 0.03 for the middle curve, and 0.01 for the lowest one. The dashed lines show the slopes of Fig. 4 for corresponding ionic strengths. The differences (20 to 40%) presumably reflect the relative positions of various kinds of charged sites.⁹⁰

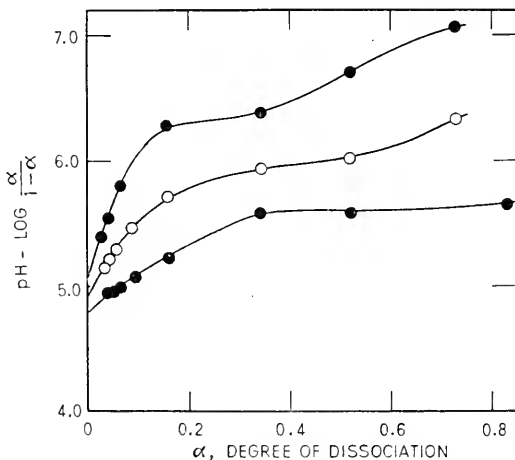


Fig. 6. Logarithmic plot for titration of the carboxyl groups of polymethacrylic acid (Arnold and Overbeek³). Ionic strength is 0.010 to 0.017 for the top curve, 0.10 for the middle curve, and 1.0 for the lowest one.

themselves easily to interpretation of the kind of configurational change which is taking place. No change in these properties, however, indicates that no major configurational change at all is occurring.

Serum Albumin. Serum albumin undergoes a reversible change in configuration near pH 4, most clearly demonstrated by the electrophoretic behavior near this pH.² Below this pH the protein shows the behavior typical of simple polyelectrolytes, as demonstrated by the effect of net charge and ionic strength on viscosity,^{93,112} shown in Fig. 7. The effect of charge on

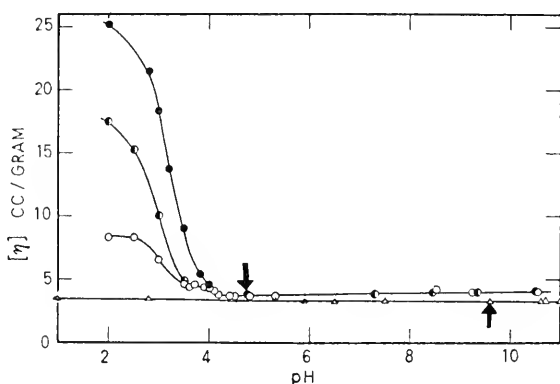


Fig. 7. Intrinsic viscosity of ribonuclease (Buzzell and Tanford¹⁰) and of bovine serum albumin (Yang and Foster¹¹², Tanford, Buzzell *et al*^{92,93}). Ribonuclease at ionic strengths ranging from 0.02 to 0.25 (Δ), serum albumin at ionic strengths 0.01 (\bullet), 0.03 (\bullet) and 0.15 (\circ). The arrows indicate the isoelectric points of the two proteins.

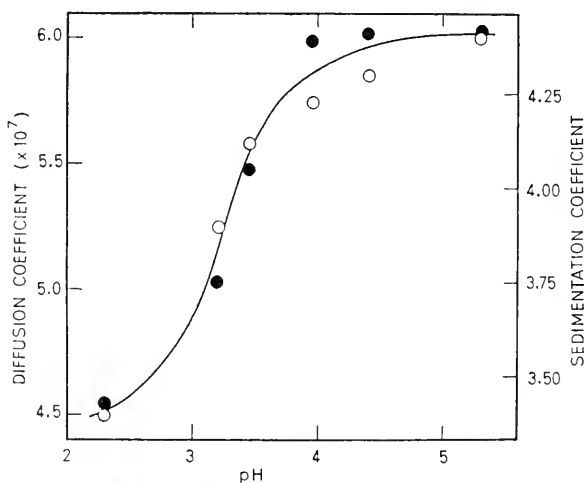


Fig. 8. Sedimentation (\bullet) and diffusion (\circ) coefficients for serum albumin. Sedimentation coefficients (Harrington *et al*³²) at ionic strength 0.1, diffusion coefficients (Champagne and Sadron,¹⁴ Kekwick⁴⁶) at ionic strength 0.15, except that at pH 2.3, which is at ionic strength 0.33.

sedimentation³² and diffusion coefficients^{14,46} lead to the same conclusion,* as shown by Fig. 8. The effect of ionic strength on the sedimentation coefficient has been measured at one pH,⁶⁶ and it also supports this conclusion. Essentially similar data have been obtained by several other workers.^{4,7,30,84} The titration of carboxyl groups below pH 4 is also in accord with polyelectrolyte-like behavior.⁹⁹

Serum albumin also undergoes a reversible change in configuration on the alkaline side of the isoelectric point, above pH 11. The polyelectrolyte-behavior of the product is indicated strikingly by the titration data of Fig. 9.⁹⁹ There is a very pronounced increase in viscosity near the same

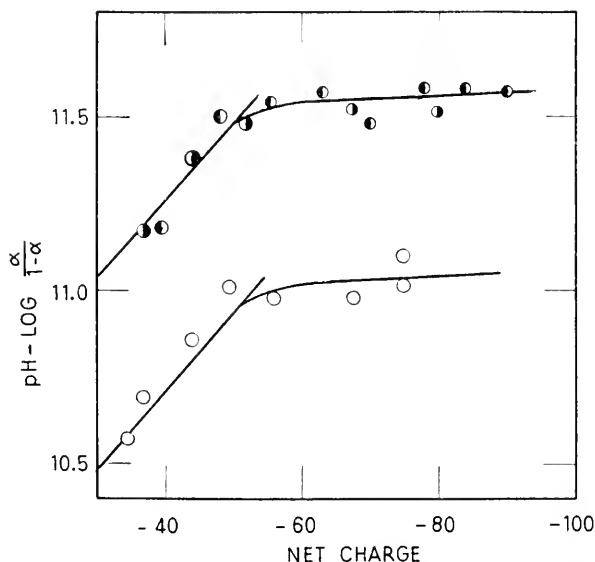


Fig. 9. Logarithmic plot for titration of the phenolic groups (upper curve) and amino groups (lower curve) of bovine serum albumin (Tanford, Swanson and Shore⁹⁹).

pH.¹⁴ Both the acid and the alkaline reaction are accompanied by marked changes in fluorescence polarization (Fig. 10)¹⁰⁵ and in optical rotatory power.^{112,39}

An interesting aspect of these configurational changes is that they are not symmetrical about the isoelectric point. The acid change occurs when the net charge is about +10, the alkaline change occurs at a net charge of about -50. It seems improbable that the electrostatic interaction energy of a unique compact configuration could be sufficiently asymmetric about the isoelectric point to account for this observation. These data therefore provide one of the strongest arguments in favor of the idea that intramolecular

* Needless to say, a change in sedimentation coefficient alone or in diffusion coefficient alone need not signify a change in configuration, but could reflect instead, a change in molecular weight. Combination of the sedimentation and diffusion data of Fig. 8 leads to the conclusion that no change in molecular weight occurs.

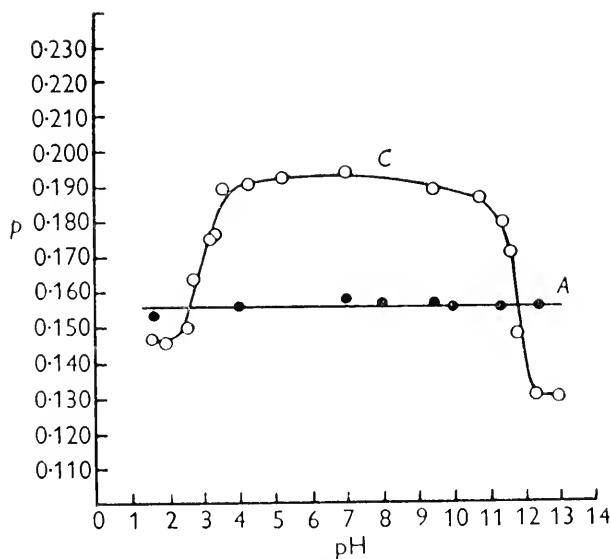


Fig. 10. Fluorescence polarization of ovalbumin (curve A) and of serum albumin (curve C) (Weber¹⁰⁵).

hydrogen bonds (or ion pairs) between side chain groups contribute to the stability of a compact configuration. (Since both the acid and alkaline reaction are reversible, we are dealing in this case with true thermodynamic stability.) However, an alternative explanation is possible. It is not necessarily true that serum albumin, or any other protein, has a *unique* compact configuration. More than one such configuration may exist. That at the isoelectric point and acid to it may have a relatively high electrostatic interaction energy. On the alkaline side, where some of the charge-bearing side chains are different from sites which are charged on the acid side, it may be possible for the molecule to rearrange itself to an alternate compact configuration of lower electrostatic free energy. (A transition from model A of Fig. 2 to model B would be an example of the kind of transition here suggested.) As we shall mention later, it has in fact often been suggested, for reasons quite unconnected with the present discussion, that serum albumin has the ability to undergo configurational changes of this kind.

Hemoglobin. Hemoglobin undergoes a major configurational change, characterized by a change in absorption spectrum and in the titration curve, near pH 3.5.^{79,81} The reaction is relatively slow, becoming much faster with decreasing pH, and is slowly reversible. The reaction is accompanied by an increase in intrinsic viscosity (at pH 3.5 and ionic strength 0.04) from 3.5 to 13.5 cc./gram.⁸⁹ This is presumptive evidence for a randomly-coiled configuration, though the effect of charge and ionic strength must be determined to establish this conclusion. The titration curve¹⁶ on the alkaline

side (above pH 10) is sufficiently steep to suggest that a similar change in configuration occurs there.

A detailed study of the kinetics of the acid reaction, in both directions, has been made by Steinhardt and Zaiser.^{80,114,115,116} Combination of the data establishes the pH-dependence of the equilibrium between the compact and expanded form. This information is more complete than similar information for any other protein, although it is unfortunate, in view of the probable importance of electrostatic forces as a driving force of the reaction, that all experimental studies have been performed at a single ionic strength. Both the equilibrium and the rate constant for formation of the expanded configuration depend strongly on pH, the former being roughly proportional to $(H^+)^5$, the latter to $(H^+)^{2.5}$. This marked dependence on pH has been used as basis for the hypothesis that a number of critical hydrogen bonds must be ruptured to initiate the reaction. This hypothesis, however, is entirely speculative. As Kauzmann⁴⁴ has shown, kinetic behavior similar to that observed for hemoglobin will result if the reaction depends solely on an increase in electrostatic interaction energy consequent upon a decrease in pH. Kauzmann's argument applies with equal force to the equilibrium of the reaction.

Conalbumin. Conalbumin undergoes a major change in configuration below pH 4,¹⁰⁸ which results in loss of solubility under conditions where the original form of the molecule is soluble. That the product of the reaction is of the simple polyelectrolyte type is shown by the viscosity and sedimentation data of Phelps and Cann.⁶² Whereas the compact form has an intrinsic viscosity of 3.8 cc./gram., the intrinsic viscosity, after the change in configuration, is much larger and strongly dependent on ionic strength. At pH 3.0, $[\eta] = 16.0, 11.0$ and 8.4 cc./gram, respectively, at ionic strengths 0.02, 0.07 and 0.10. The sedimentation coefficient decreases with increasing charge, as is shown by Fig. 11, and the decrease is partly suppressed by ionic strength.

On the alkaline side of the isoelectric point, near pH 10.5, there is a break in the logarithmic plot of titration data for phenolic and amino groups,¹⁰⁸ which is much like the similar break shown for serum albumin in Fig. 9. The net charge at this point is about the same as the net charge at which the acid reaction begins.

Wishnia¹⁰⁸ has made a detailed study of the kinetics of the acid reaction, both in the forward and reverse direction. An interesting feature is that the reverse reaction shows a pH maximum.

Pepsin. Pepsin undergoes irreversible loss of its enzymatic activity between pH 6 and 7.⁷⁸ Edelhoeh²¹ has studied the molecular change which accompanies this reaction. It appears that a small part of the molecule is split off. The remaining portion clearly shows polyelectrolyte-like behavior. At pH 6, the value of η_{sp}/c for a dilute solution (i.e. essentially the intrinsic viscosity)

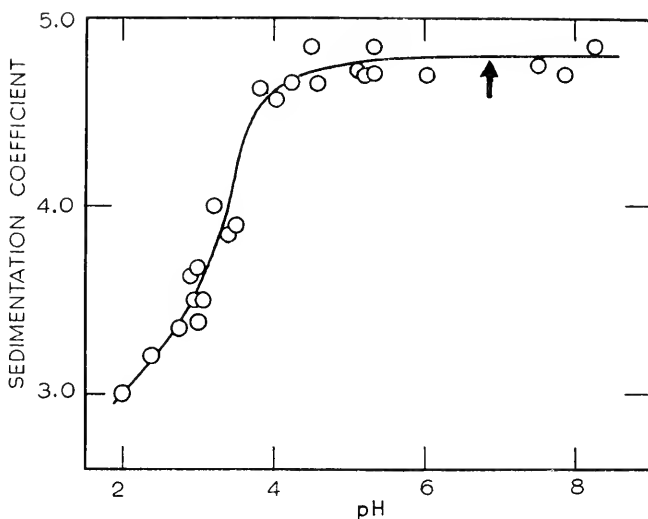


Fig. 11. Sedimentation coefficient of conalbumin at ionic strength 0.1 (Phelps and Cann⁶²). Osmotic pressure measurements over the same pH range indicate that no change in molecular weight takes place. The arrow shows the location of the isoelectric point.

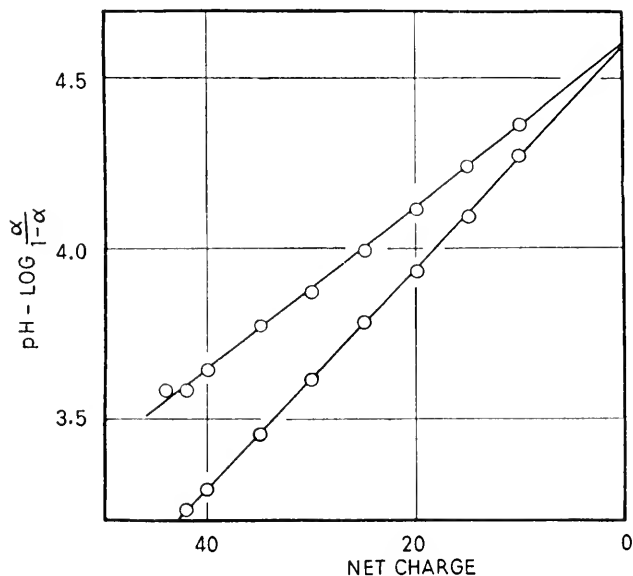


Fig. 12. Logarithmic plot for titration of the carboxyl groups of β -lactoglobulin (Cannan, Palmer and Kibrick¹²). Ionic strength is 0.27 for the upper curve and 0.069 for the lower curve.

is about 3.5 cc./gram, independent of ionic strength. At pH 7, on the other hand, the same parameter ranges from 21.2 to 6.7 cc./gram as the ionic strength is increased from 0.01 to 0.55. Corresponding changes were observed in sedimentation and diffusion coefficients.

A kinetic study of the loss of enzymatic activity of pepsin was performed by Steinhart⁷⁸ as early as 1937. The pH dependence was used as evidence that rupture of specific bonds triggers the reaction. This conclusion is subject to the same comment as was made above for hemoglobin.

β -Lactoglobulin. β -Lactoglobulin undergoes an irreversible reaction above pH 9, the product of which is insoluble at the isoelectric point. The kinetics of the reaction have been investigated by Groves, Hipp and McMeekin.³¹ That the product is of the simple polyelectrolyte type is indicated by the titration data for the phenolic groups shown in Fig. 13.⁹⁸

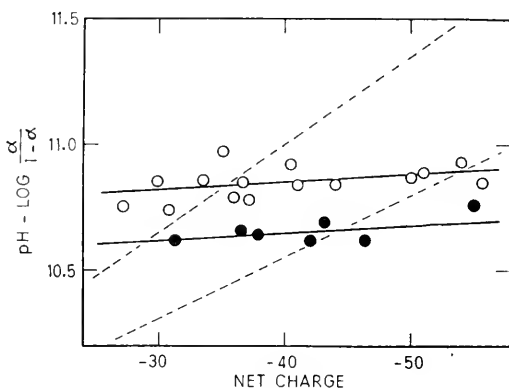


Fig. 13. Logarithmic plot for titration of the phenolic groups of β -lactoglobulin (Tanford and Swanson⁹⁸) at ionic strengths 0.08 (○) and 0.27 (●). The dashed lines show the slopes of corresponding plots for titration of carboxyl groups (Fig. 12). The difference in slopes is roughly a factor of ten.

β -Lactoglobulin is another protein with marked asymmetry about the isoelectric point. The titration data for carboxyl groups,¹² shown in Fig. 12, are characteristic of a compact protein up to the highest positive charges reached. Viscosity measurement as acid as pH 1.7 also give no indication of any change in configuration.¹¹³ In this protein it is thus certain that electrostatic forces alone cannot account for the effect of pH on configuration.

γ -Globulin. Jirgensons⁴⁰ has shown that the viscosity and optical rotatory power of γ -globulin depend on pH much like the corresponding properties of serum albumin. His data are shown in Fig. 14. It should be noted that the isoelectric point of γ -globulin is considerably higher than that of serum

albumin, so that the changes indicated by Fig. 14 represent behavior which is much more nearly symmetrical about the isoelectric point.

It should be noted that the expanded configurations of all of the proteins discussed above, though they behave *qualitatively* like simple polyelectrolytes, do not have intrinsic viscosities as large as those of simple polyelectrolytes of similar molecular weight and charge. This result is to be expected, as has been mentioned, because disulfide cross-links limit the expansion which

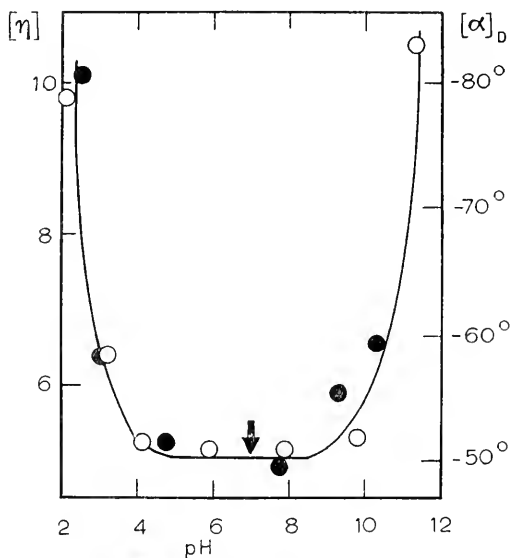


Fig. 14. Intrinsic viscosity (●) and specific rotation (○) of γ -globulin at ionic strength 0.1 (Jirgensons⁴⁰). This protein is a mixture of components with different isoelectric points, the average of which is near pH 7.

protein molecules can undergo. It is important to point out, in addition, that it is virtually impossible to distinguish configurations consisting of a mixture of compact and randomly coiled parts from a configuration which is completely random except for cross-links. In other words, though the measurements cited clearly show polyelectrolyte-like properties, this does not yet provide a complete picture of the expanded configurations.

We shall list next, some globular proteins which are much more resistant to configurational change than those discussed above.

Ribonuclease. In contrast to the proteins just discussed, ribonuclease is highly resistant to configurational change. As Fig. 7 shows, its intrinsic viscosity is virtually independent of charge and ionic strength between pH 1 and pH 11.5.¹⁰ At pH 11.5, there is a configurational change which brings three phenolic groups from the interior of the molecule to the surface.⁹⁶ Whether this reaction also leads to expansion is not known. At pH 1.0 and ionic

strength 0.15, the intrinsic viscosity is 4.5 cc./gram, in contrast to the value of 3.5 cc./gram observed at the same pH at ionic strength 0.25.¹⁰ At pH 2 (ionic strength unspecified) there is also a change in absorption spectrum,⁷² the steepness of which with respect to pH suggests that a configurational change is occurring. It is thus possible that ribonuclease tends to acquire a polyelectrolyte-like configuration at the extreme acid end of its titration curve if the ionic strength is sufficiently low.

The release of phenolic groups by ribonuclease, which occurs at or above pH 11.5, does not take place under reversible conditions. As was pointed out earlier, this prevents us from deciding whether the stability of the original compact configuration of this protein represents true thermodynamic stability, or whether it merely reflects a requirement for a larger free energy of activation. One of the most fruitful approaches to an understanding of ribonuclease and other proteins which change configuration under irreversible conditions would be a search for conditions under which a reversal to a compact configuration occurs. From a study of such a reversal, one could obtain a measure of the true thermodynamic stability of the compact configuration.

Ovalbumin. Ovalbumin, like ribonuclease, undergoes no major configurational change between pH 2 and pH 11.5. This is shown by the titration curve of Cannan and coworkers,¹¹ which fits the compact sphere model throughout this range of pH, and by the absence of change in the polarization of fluorescence, shown in Fig. 10.¹⁰⁵ The viscosity data of Bull⁸ are not entirely unequivocal, because η_{sp}/c was determined at each pH at a single concentration and not extrapolated to give the intrinsic viscosity. At the very low ionic strengths used in some of his experiments, this creates considerable uncertainty. Within this uncertainty the results confirm the conclusion reached from the other data cited.⁶

Ovalbumin behaves like ribonuclease in another respect. Its phenolic groups, initially inaccessible to titration, are slowly released near pH 12 or higher.¹⁸ Whether this process is accompanied by expansion is not known.

Insulin. In many ways, insulin is the best known of all globular proteins. However, only scant information is available about the effect of pH on its configuration. There is certainly no configurational change in acid solutions, as has been shown by viscosity measurements of Yang and Foster.¹¹³ The titration curve in acid solution is anomalous,⁹⁴ but the anomaly is fully explained on the basis of molecular weight changes which this protein is known to undergo (see below).

Lysozyme. Lysozyme is another protein for which no expansion has yet been observed. Steiner⁷⁷ has shown that the polarization of fluorescence of dye molecules bound to lysozyme remains unchanged between pH 3.5 and

pH 10. Yang and Foster¹¹³ have demonstrated that the intrinsic viscosity is constant between pH 2.3 and pH 7.5. The titration curve of lysozyme^{100,90} shows no feature suggestive of expansion at any pH.

Apart from the major configurational changes discussed so far, there is likelihood of the occurrence of minor changes in configuration. There may, for example, be several different arrangements of the polypeptide chain (or chains) which have 'hydrophobic' regions of comparable stability. A change in pH and the resultant change in the number and positions of charged sites may stabilize one such form at the expense of another. A similar situation can arise if hydrogen bonds contribute to the stability of the original compact form. The rupture of some of these bonds by titration could lead, if the net charge is still low, to a new compact configuration with a new set of hydrogen bonds.

There are small changes in the sedimentation coefficient of β -lactoglobulin near pH 5 and pH 8,⁶¹ which might reflect such minor changes in configuration. A similar small change in sedimentation coefficient occurs for ovalbumin, near pH 4.¹⁵ It would be desirable to obtain confirmatory evidence of these changes by other methods.

Minor configurational changes need not necessarily be accompanied by changes in hydrodynamic properties. Klotz and coworkers, for example, have observed changes in the ability of serum albumin to bind calcium⁴³ and certain anions⁵⁰ and uncharged molecules,⁴⁹ which they have ascribed to a configurational change. The change is observed near pH 7, where no change in hydrodynamic properties can be detected.⁹¹ Karush⁴² has suggested that the unusual ability of serum albumin to bind ions of all kinds is evidence for quite general configurational adaptability. That these observations based on binding properties reflect configurational changes is not an unassailable conclusion, for alternative interpretations are possible.

POLYMERIZATION OF PROTEINS

This paper has dealt with the various forces which, in water solution, attract one portion of a protein molecule to another portion of the same molecule. It is obvious that the same forces must be able to act *intermolecularly* and thus bring about polymerization. Whether polymerization will actually take place, depends on the extent to which these forces have been able to be utilized intramolecularly.

Association between protein molecules in their compact configurations occurs in some proteins and not in others. It does not occur in ribonuclease,⁶⁵ the molecular weight of which is independent of pH or ionic strength between pH 2 and pH 11. It does not occur in serum albumin.³² The very slow association^{7,51,69} which this protein undergoes is observed only on standing over a period of days, and then only subsequent to loss of the compact configuration.⁸² Insulin, on the other hand, is polymerized at all pH values

in aqueous solution.^{20,26,57} The extent of polymerization is lowest when the molecule bears a high net charge, positive or negative. It increases as the isoelectric point is approached, and, in the isoelectric region itself, leads to insolubility of the protein. This is clearly the expected result if electrostatic forces predominate in the pH-dependence of intermolecular attraction. Hemoglobin²³ is another globular protein which fits this picture. It exists as a dimer near its isoelectric point and dissociates reversibly as the pH is lowered. Dissociation is probably complete before the expansion earlier referred to occurs.

The most interesting example of the polymerization of a compact protein molecule is that which occurs for β -lactoglobulin.^{55,102,103} Experimental molecular weights for this protein, as observed by Townend and Timasheff, are shown in Fig. 15. The most interesting feature is the polymerization of

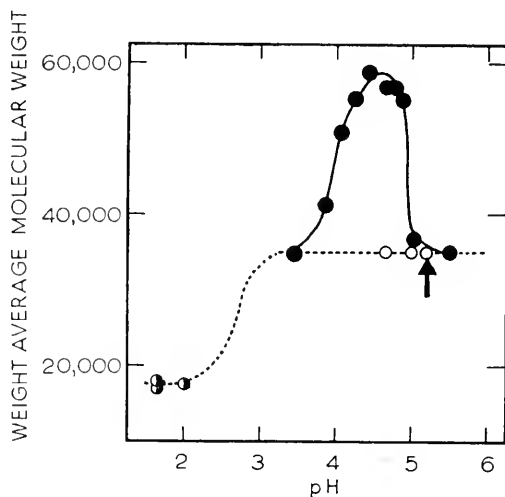


Fig. 15. The molecular weight of β -lactoglobulin near 25° (dashed line) and near 2° (solid line) (Townend and Timasheff^{102,103}): ● 25° at the lowest feasible concentration, as determined by the Archibald method; ○ 25°, miscellaneous observations; ● 2°, at concentrations of 2 to 3%, calculated from the relative areas of *separate* sedimentation peaks representing species of molecular weight 35,000 and 70,000. The arrow locates the isoelectric point.

the dimer, at 0°C, which occurs entirely on one side of the isoelectric point. These observations, coupled with the earlier observation that β -lactoglobulin becomes randomly coiled on the alkaline side of the isoelectric point, and not on the acid side, make this one of the most interesting of all globular proteins.

It might at first sight seem probable that globular proteins in a randomly coiled configuration should be more liable to polymerization than the same proteins in compact configurations. However, the forces, such as electrostatic repulsion, which lead to the preference for a randomly coiled configuration,

will also operate against polymerization.* An interesting situation arises, however, when an expanded protein is suddenly returned to near its isoelectric point. The randomly coiled configuration is now no longer the stable one, but return to a compact configuration may require specific coiling of the polypeptide chains, and, as a result, may be a slow reaction. Under these conditions polymerization may compete with return to a compact shape, even if the latter is thermodynamically stable. Insolubility at the isoelectric point, which is so often a consequence of denaturation, may be explained in this way. The configurational changes described earlier for serum albumin and hemoglobin are reversible, so that return to a compact shape is the favoured reaction for these proteins. In view of Kauzmann's suggestion⁴⁴ that the number of disulfide bonds may determine the rate of return to a compact configuration, it is interesting to note that, of these two proteins, serum albumin has a large number of disulfide bonds (about 16), but hemoglobin has none.

CONCLUSION

The most important conclusion which can be drawn from this paper is that our ignorance of the configuration of proteins in aqueous solution greatly exceeds the few facts which have been established. It appears possible that the final answer to the problems which have been posed may be different for each of the proteins which have been discussed. Moreover, a complete theory will have to account not only for the properties of these globular proteins, but it must also explain why proteins such as collagen appear to have no stable compact configuration at all in aqueous solution.

ACKNOWLEDGMENTS

The work of the author and coworkers described in this paper has been supported by research grants from the National Science Foundation and from the National Institutes of Health, Public Health Service. This paper was written while the author was a John Simon Guggenheim Memorial Fellow at Yale University.

REFERENCES

1. J. W. ANDEREGG, W. W. BEEMAN, S. SHULMAN and P. KAESBERG, *J. Am. Chem. Soc.*, **77**, 2927 (1955).
2. K. AOKI and J. F. FOSTER, *J. Am. Chem. Soc.*, **78**, 3538 (1956).
3. R. ARNOLD and J. TH. G. OVERBEEK, *Rec. trav. chim. Pays-Bas*, **69**, 192 (1950).
4. S. BJORNHOLM, E. BARBU and M. MACHEBOEUF, *Bull. soc. chim. biol.*, **34**, 1083 (1952).

* The slow polymerization of serum albumin in acid solutions^{7,51,69} appears to be due to a sulfhydryl-disulfide exchange reaction.⁸² The association of serum albumin²⁸ and ovalbumin²⁷ after expansion by urea has been explained on a similar basis. These reactions therefore involve changes in primary bonding.

5. H. BOEDTKER and P. DOTY, *J. Am. Chem. Soc.*, **78**, 4267 (1956).
6. F. BOOTH, *Proc. Roy. Soc. (London)*, **A 203**, 533 (1950).
7. P. BRO, S. J. SINGER, and J. M. STURTEVANT, *J. Am. Chem. Soc.*, **77**, 4924 (1955).
8. H. B. BULL, *Trans. Faraday Soc.*, **36**, 80 (1940).
9. L. BUNVILLE and C. TANFORD, unpublished data.
10. J. G. BUZZELL and C. TANFORD, *J. Phys. Chem.*, **60**, 1204 (1956).
11. R. K. CANNAN, A. C. KIBRICK and A. H. PALMER, *Am. N.Y. Acad. Sci.*, **41**, 243 (1941).
12. R. K. CANNAN, A. H. PALMER and A. C. KIBRICK, *J. Biol. Chem.*, **142**, 803 (1942).
13. R. CECIL and A. G. OGSTON, *Biochem. J.*, **43**, 205 (1948).
14. M. CHAMPAGNE and C. SADRON, *Conf. High Polymers*, Strasbourg, 1954.
15. P. A. CHARLWOOD and A. ENS, *Can. J. Chem.*, **35**, 99 (1957).
16. E. J. COHN, A. A. GREEN and M. H. BLANCHARD, *J. Am. Chem. Soc.*, **59**, 509 (1937).
17. E. J. COHN and A. M. PRETISS, *J. Gen. Physiol.*, **8**, 619 (1927).
18. J. L. CRAMMER and A. NEUBERGER, *Biochem. J.*, **37**, 302 (1943).
19. P. DOTY, J. H. BRADBURY and A. M. HOLTZER, *J. Am. Chem. Soc.*, **78**, 947 (1956).
20. P. DOTY and G. E. MYERS, *Disc. Faraday Soc.*, No. 13, 51 (1953).
21. H. EDELHOCH, *J. Am. Chem. Soc.* **79**, 6100 (1957).
22. A. EINSTEIN, *Ann. Physik*, **19**, 289 (1906); **34**, 591 (1911).
23. E. O. FIELD and J. R. P. O'BRIEN, *Biochem. J.*, **60**, 656 (1955).
24. P. J. FLORY, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, N.Y., 1953.
25. T. G. FOX, JR., and P. J. FLORY, *J. Phys. and Colloid Chem.*, **53**, 197 (1949).
26. E. FREDERICQ, *Nature*, **171**, 570 (1953).
27. H. K. FRENSDORFF, M. T. WATSON and W. KAUZMANN, *J. Am. Chem. Soc.*, **75**, 5157 (1953).
28. H. K. FRENSDORFF, M. T. WATSON and W. KAUZMANN, *J. Am. Chem. Soc.*, **75**, 5167 (1953).
29. C. FRIEDEN, R. M. BOCK and R. A. ALBERTY, *J. Am. Chem. Soc.*, **76**, 2482 (1954).
30. E. GAVRILESCO, E. BARBU and M. MACHEBOEUF, *Bull. soc. chim. biol.*, **32**, 924 (1950).
31. M. L. GROVES, N. J. HIPPI and T. L. MCMEEKIN, *J. Am. Chem. Soc.*, **73**, 2790 (1951).
32. W. F. HARRINGTON, P. JOHNSON and R. H. OTTEWILL, *Biochem. J.*, **62**, 569 (1956).
33. W. F. HARRINGTON and J. A. SCHELLMAN, *Compt. rend. trav. lab. Carlsberg*, **30**, 21 (1956).
34. F. E. HARRIS and S. A. RICE, *J. Phys. Chem.*, **58**, 725 (1954).
35. C. S. HOCKING, M. LASKOWSKI, JR., and H. A. SCHERAGA, *J. Am. Chem. Soc.*, **74**, 775 (1952).
36. A. HOLTZER and S. LOWEY, *J. Am. Chem. Soc.*, **78**, 5954 (1956).
37. M. L. HUGGINS, *Chem. Revs.*, **32**, 195 (1943).
38. A. HVIDT and K. LINDERSTRÖM-LANG, *Biochim. et Biophys. Acta*, **16**, 168 (1955).
39. B. JIRGENSONS, *Arch. Biochem. Biophys.*, **39**, 261 (1952).
40. B. JIRGENSONS, *Arch. Biochem. Biophys.*, **48**, 154 (1954).
41. B. JIRGENSONS, *Makromol. Chem.*, **18**, 48 (1956).
42. F. KARUSH, *J. Am. Chem. Soc.*, **72**, 2705 (1950).
43. S. KATZ and I. M. KLOTZ, *Arch. Biochem. Biophys.*, **44**, 351 (1953).
44. W. KAUZMANN, *The Mechanism of Enzyme Action* (W. D. McElroy and B. Glass, ed.), Johns Hopkins Press, Baltimore, Md., 1954.
45. W. KAUZMANN and R. B. SIMPSON, *J. Am. Chem. Soc.*, **75**, 5154 (1953).
46. R. A. KEKWICK, quoted by W. F. Harrington *et al.*, ref. 32.
47. J. G. KIRKWOOD, *J. Chem. Phys.*, **2**, 351 (1934).
48. J. G. KIRKWOOD and J. B. SHUMAKER, *Proc. Natl. Acad. Sci., U.S.*, **38**, 855 (1952).

49. I. M. KLOTZ and J. AYERS, *J. Am. Chem. Soc.*, **74**, 6178 (1952).
50. I. M. KLOTZ, R. K. BURKHARD and J. M. URQUHART, *J. Am. Chem. Soc.*, **74**, 202 (1952).
51. M. J. KRONMAN, M. D. STERN and S. N. TIMASHEFF, *J. Phys. Chem.*, **60**, 829 (1956).
52. M. LASKOWSKI, JR., and H. A. SCHERAGA, *J. Am. Chem. Soc.*, **76**, 6305 (1954).
53. K. LINDERSTRØM-LANG, *Chem. Soc. (London) Spec. Publ. No. 2*, 1 (1955).
54. A. E. MIRSKY and L. PAULING, *Proc. Natl. Acad. Sci., U.S.*, **22**, 439 (1936).
55. A. G. OGSTON and J. M. A. TILLEY, *Biochem. J.*, **59**, 644 (1955).
56. J. L. ONCLEY, in E. J. COHN and J. T. EDSELL, *Proteins, Amino Acids and Peptides*, Reinhold Publishing Corp., New York, N.Y., 1943.
57. J. L. ONCLEY, E. ELLENBOGEN, D. GITLIN and F. R. N. GURD, *J. Phys. Chem.*, **56**, 85 (1952).
58. L. PAULING, R. B. COREY and H. R. BRANSON, *Proc. Natl. Acad. Sci., U.S.*, **37**, 205 (1951).
59. L. PAULING and R. B. COREY, *Proc. Natl. Acad. Sci., U.S.*, **37**, 251, 256, 261, 272 (1951).
60. L. PAULING and R. B. COREY, *Proc. Natl. Acad. Sci., U.S.*, **37**, 282 (1951).
61. K. O. PEDERSEN, *Biochem. J.*, **30**, 961 (1936).
62. R. A. PHELPS and J. R. CANN, *Arch. Biochem. Biophys.*, **61**, 51 (1956).
63. A. POLSON, *Kolloid-Z.*, **88**, 51 (1939).
64. F. W. PUTNAM, *The Proteins* (H. Neurath and K. Bailey, ed.), Academic Press, New York, N.Y., 1953.
65. D. G. RANDS, Ph.D. Thesis, State University of Iowa, 1957.
66. M. E. REICHMANN and P. A. CHARLWOOD, *Can. J. Chem.*, **32**, 1092 (1954); cf. footnote 25 of ref. 93.
67. S. A. RICE and F. E. HARRIS, *J. Phys. Chem.*, **58**, 733 (1954).
68. S. A. RICE and F. E. HARRIS, *J. Chem. Phys.*, **24**, 326, 336 (1956).
69. H. A. SAROFF, G. I. LOEB and H. A. SCHERAGA, *J. Am. Chem. Soc.*, **77**, 2908 (1955).
70. J. A. SCHELLMAN, *Compt. rend. trav. lab. Carlsberg*, **29**, 223 (1955).
71. J. A. SCHELLMAN, *Compt. rend. trav. lab. Carlsberg*, **29**, 230 (1955).
72. H. A. SCHERAGA, *Biochim. et Biophys. Acta*, **23**, 196 (1957).
73. H. A. SCHERAGA and L. MANDELKERN, *J. Am. Chem. Soc.*, **75**, 179 (1953).
74. D. SHUGAR, *Biochem. J.*, **52**, 142 (1952).
75. S. SHULMAN, *J. Am. Chem. Soc.*, **75**, 5846 (1953).
76. R. B. SIMPSON and W. KAUZMANN, *J. Am. Chem. Soc.*, **75**, 5139 (1953).
77. R. F. STEINER, *Arch. Biochem. Biophys.*, **46**, 291 (1953).
78. J. STEINHARDT, *Kgl. Danske Videnskab. Selskab, Mat. fys. Medd.*, **14**, No. 11 (1937).
79. J. STEINHARDT and E. M. ZAISER, *J. Biol. Chem.*, **190**, 197 (1951).
80. J. STEINHARDT and E. M. ZAISER, *J. Am. Chem. Soc.*, **75**, 1599 (1953).
81. J. STEINHARDT and E. M. ZAISER, *Advances in Protein Chem.*, **10**, 180 (1955).
82. J. M. STURTEVANT, S. J. SINGER and P. BRO, Personal Communication.
83. J. B. SUMNER, A. L. DOUNCE and V. L. FRAMPTON, *J. Biol. Chem.*, **136**, 343 (1940).
84. T. SVEDBERG and B. SJOGREN, *J. Am. Soc.*, **52**, 2855 (1930).
85. C. TANFORD, *J. Am. Chem. Soc.*, **76**, 945 (1954).
86. C. TANFORD, *Electrochemistry in Biology and Medicine* (T. Shedlovsky, ed.), John Wiley and Sons, New York, N.Y., 1955.
87. C. TANFORD, *J. Phys. Chem.*, **59**, 788 (1955).
88. C. TANFORD, *J. Phys. Chem.*, **61**, 1023 (1957).
89. C. TANFORD, *J. Am. Chem. Soc.*, **79**, 3931 (1957).
90. C. TANFORD, *J. Am. Chem. Soc.*, **79**, 5340 (1957).
91. C. TANFORD and J. G. BUZZELL, *J. Am. Chem. Soc.*, **76**, 3356 (1954).
92. C. TANFORD and J. G. BUZZELL, *J. Phys. Chem.*, **60**, 225 (1956).
93. C. TANFORD, J. G. BUZZELL, D. G. RANDS and S. A. SWANSON, *J. Am. Chem. Soc.*, **77**, 6421 (1955).
94. C. TANFORD and J. EPSTEIN, *J. Am. Chem. Soc.*, **76**, 2170 (1954).

95. C. TANFORD and J. D. HAUENSTEIN, *J. Am. Chem. Soc.*, **78**, 5287 (1956).
96. C. TANFORD, J. D. HAUENSTEIN and D. G. RANDS, *J. Am. Chem. Soc.*, **77**, 6409 (1955).
97. C. TANFORD and J. G. KIRKWOOD, *J. Am. Chem. Soc.*, **79**, 5333 (1957).
98. C. TANFORD and S. A. SWANSON, *J. Am. Chem. Soc.*, **79**, 3297 (1957).
99. C. TANFORD, S. A. SWANSON and W. S. SHORE, *J. Am. Chem. Soc.*, **77**, 6414 (1955).
100. C. TANFORD and M. L. WAGNER, *J. Am. Chem. Soc.*, **76**, 3331 (1954).
101. D. TER HAAR, *Elementary Statistical Mechanics*, Rinehart and Co., New York, N.Y., 1954.
102. R. TOWNEND and S. N. TIMASHEFF, *Arch. Biochem. Biophys.*, **63**, 482 (1956).
103. R. TOWNEND and S. N. TIMASHEFF, *J. Am. Chem. Soc.*, **79**, 3613 (1957).
104. J. H. WANG, *J. Am. Chem. Soc.*, **76**, 4755 (1954).
105. G. WEBER, *Biochem. J.*, **51**, 155 (1952).
106. F. H. WESTHEIMER and O. T. BENFY, *J. Am. Chem. Soc.*, **78**, 5309 (1956).
107. D. WETLAUFER, *Compt. rend. trav. lab. Carlsberg*, **30**, 135 (1956).
108. A. WISHNIA, Ph.D. Thesis, New York University, 1957.
109. H. WU, *Chinese J. Physiol.*, **5**, 321 (1931).
110. J. WYMAN and E. N. INGALLS, *J. Biol. Chem.*, **147**, 297 (1943).
111. J. T. YANG and P. DOTY, *J. Am. Chem. Soc.*, **79**, 761 (1957).
112. J. T. YANG and J. F. FOSTER, *J. Am. Chem. Soc.*, **76**, 1588 (1954).
113. J. T. YANG and J. F. FOSTER, *J. Am. Chem. Soc.*, **77**, 2374 (1955).
114. E. M. ZAISER and J. STEINHARDT, *J. Am. Chem. Soc.*, **73**, 5568 (1951).
115. E. M. ZAISER and J. STEINHARDT, *J. Am. Chem. Soc.*, **76**, 1788 (1954).
116. E. M. ZAISER and J. STEINHARDT, *J. Am. Chem. Soc.*, **76**, 2866 (1954).

Über die Artspezifität der Proteinstruktur

H. TUPPY

II. Chemisches Institut der Universität Wien

In Organismen, welche verschiedenen Arten, Gattungen, Ordnungen usw. angehören, werden übereinstimmende physiologische und biochemische Leistungen in der Regel von Proteinen nicht gleicher, aber ähnlicher Struktur vollbracht. Mit anderen, der biologischen Terminologie entlehnten Worten: *Analoge* Proteine, d.h. solche, die in verschiedenartigen Organismen vorkommend einander in funktioneller Hinsicht gleichen, stehen zueinander häufig auch im Verhältnis der *Homologie*, d.h. einer weitreichenden strukturellen Übereinstimmung, die auf eine genetische Verwandtschaft hinweist. Die Tatsache, dass funktionsgleiche Proteine sich im Rahmen der Homologie voneinander durch spezifische strukturelle Merkmale unterscheiden, wird gewöhnlich mit dem Terminus *Artspezifität* gekennzeichnet (womit jedoch nicht präjudiziert werden soll, dass sich die Variationen der Proteinstruktur an die Einteilung der Organismen in Arten hält).

Während die Klassifikation der Organismen ausschliesslich nach morphologischen Merkmalen, nach der Homologie des anatomischen Baues, erfolgt, steht bei der Beschreibung und Einteilung der physiologisch wirksamen Proteine der funktionelle Aspekt im Vordergrund. Dennoch ist auch bei Proteinen ein 'vergleichend-anatomisches' Studium von Interesse. Denn einerseits kann man füglich einen Zusammenhang suchen zwischen mehr oder weniger weitreichenden Homologien der Proteinstruktur und engeren oder fernerer genetischen Beziehungen der Organismen, denen die untersuchten homologen Proteine entstammen. Andererseits lässt sich von der Kenntnis der Artspezifität der Proteine ein besseres Verständnis der strukturellen Voraussetzungen ihrer biologischen Wirkungen erhoffen: dürften doch nur die unveränderlichen Strukturmerkmale mit der gleichbleibenden biologischen Wirkung in Zusammenhang stehen, nicht jedoch die je nach Herkunft variablen.

Nur in Ausnahmefällen ist die Artspezifität der Proteinstruktur leicht erkennbar, wie z. B. bei Vertebraten-Hämoglobinen durch auffallende Unterschiede der Kristallform und der Löslichkeiten. Im allgemeinen bedarf die Feststellung der Artunterschiede einer eingehenderen Prüfung. Eine seit Jahr-

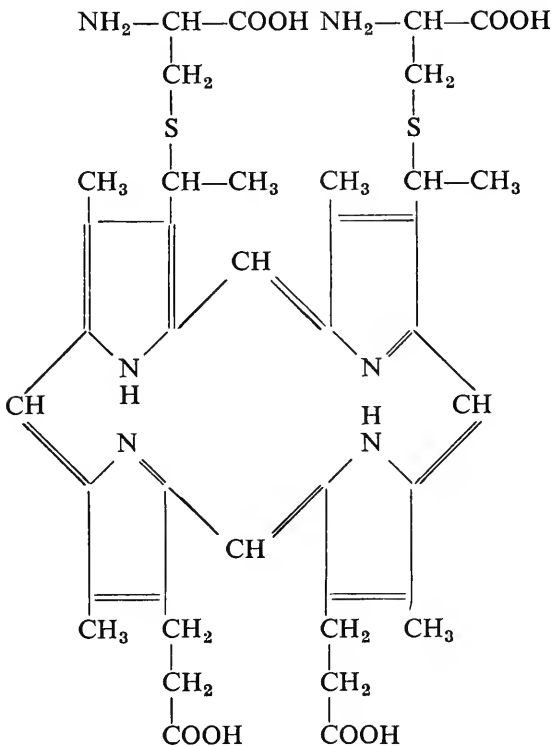
zehnten erprobte Methode ist die immunologische: Homologe Proteine sind (mit gewissen Ausnahmen) spezifische Antigene und ihre Antigen-Spezifität ist umso ähnlicher, je näher sie einander in der biologischen Klassifizierung stehen. Bisweilen lassen sich Proteine verschiedener Herkunft auch durch eine unterschiedliche elektrophoretische Beweglichkeit unterscheiden. Seitdem in den letzten anderthalb Jahrzehnten die Aminosäure-Analytik einen so hohen Grad der Genauigkeit erreicht hat, dass sie auch die geringfügigen Unterschiede in der Aminosäure-Zusammensetzung sehr ähnlicher Proteine eindeutig feststellen kann, verdanken wir der Baustein-Analyse homologer Eiweisskörper ebenfalls wichtige Hinweise. Hier sei beispielsweise die von Harfenist und Craig (1952 b)¹⁰ am Insulin dreier Säugetierarten vorgenommene Aminosäurebestimmung hervorgehoben; diese zeigte, dass bei ansonsten gleicher Zusammensetzung der homologen Insulinmoleküle an drei ihrer 51 Aminosäurebausteine auf Konto der Artspezifität gehende Verschiedenheiten nachweisbar sind. Besonders hervorzuheben ist auch die analytische Untersuchung, die Knight (1947) an einer Reihe von Pflanzenviren, die dem Tabakmosaikvirus nahestehen, vorgenommen hat; er fand, dass einander in pathologischer Hinsicht nahe verwandte Stämme eine identische oder sehr ähnliche, biologisch entfernter verwandte Stämme hingegen eine deutlich verschiedene Aminosäure-Zusammensetzung besitzen. Seine Befunde gaben der Vermutung Nahrung, dass sich in der Proteinstruktur engere oder entferntere genetische Beziehungen widerspiegeln. Noch einen Schritt weiter führten Feststellungen der Natur der in den Polypeptidketten homologer Proteine endständigen Aminosäurereste; bei der ersten vergleichenden Endgruppen-Untersuchung, die von Porter und Sanger (1948) vorgenommen worden ist, wurde ebenfalls der ermutigende Befund erhoben, dass chemische Gleichheiten und Verschiedenheiten bis zu einem gewissen Grade mit serologischen parallel gingen und der biologischen Klassifizierung der Organismen, von denen die Hämoglobine stammten, entsprachen.

Für einen genaueren Einblick in die Natur der Artspezifität der Proteinstruktur ist jedoch die Bestimmung der Aminosäure-Reihenfolge in Polypeptidketten homologer Proteine erforderlich. Solche vergleichende Sequenz-Untersuchungen sind in den letzten Jahren an einigen Stellen durchgeführt worden. Hier ist vor allem die detaillierte Aufklärung der Artdifferenzen im Insulin fünf verschiedener Säugetier-Spezies hervorzuheben, die wir Sanger und seinen Mitarbeitern verdanken (Brown, Sanger und Kitai, 1955; Harris, Sanger und Naughton, 1956). Diese und andere Homologie-Studien werden zusammenfassend besprochen, nachdem über eigene vergleichende Untersuchungen an Cytochrom-c verschiedener Herkunft berichtet worden ist.

Cytochrom-c ist für eine chemische Bearbeitung der Artspezifität der Proteinstruktur besonders geeignet. Denn es zeichnet sich (wie keineswegs die meisten anderen Eiweisstoffe) durch ein nahezu ubiquitäres Vorkommen in der Welt der Organismen aus: Von den niedersten bis zu den höchsten

Lebewesen, in Mikroorganismen, in Pflanzen und in Tieren ist Cytochrom-c nachgewiesen worden. Bei seiner Auffindung und Isolierung leisten seine charakteristischen Absorptionsbanden unschätzbare Hilfe. Es besitzt ein niedriges Molekulargewicht (in der Grössenordnung von 12000) und überlebt dank seiner verhältnismässig robusten Konstitution die Unbilden einer Gewinnung und Reinigung leichter als andere, empfindlichere Proteine.

Eine besondere und vorteilhafte Eigentümlichkeit der Cytochrom-c-Struktur besteht darin, dass im Hämoprotein-Molekül die prosthetische Gruppe mit dem Eiweissanteil durch zwei ziemlich stabile Thioätherbrücken verknüpft ist. Diese überdauern sogar einen hydrolytischen Abbau des Proteins. Theorell (1939) sowie Zeile und Meyer (1939) konnten aus sauren Hydrolysaten des Cytochroms-c eine als Porphyrin-c bezeichnete Verbindung isolieren, welche man sich durch Addition der Thiolgruppen von zwei Cysteinresten an die Vinylseitenketten in Stellung 2 und 4 eines Protoporphyrins aufgebaut denken kann.



Porphyrin c

Im intakten Cytochrom-c enthält die prosthetische Gruppe natürlich noch Eisen als Zentralatom und die beiden porphyrin-gebundenen Cysteinreste sind mit ihren Amino- und Carboxylgruppen peptidisch in die Eiweisskomponente des Hämoproteins eingefügt. Man kann sich diese Verhältnisse

zunutze machen und beim Abbau des Cytochroms-c die gefärbte prosthetische Gruppe als Markierung des ihr benachbarten Teiles des Apoproteins verwenden. Bei einem Studium der Artunterschiede in Cytochrom-c verschiedener Herkunft bietet die Untersuchung gerade des porphyrinnahen Teiles ihrer Aporoteine noch einen weiteren Vorteil; sie gibt absolute Sicherheit, dass in den homologen Molekülen analoge Regionen verglichen werden, die in der gleichen sterischen und funktionellen Beziehung zur prosthetischen Gruppe stehen.

Wenn Cytochrom-c mit verdünnter Schwefelsäure schonend hydrolysiert wird, so geht der Abbau nicht bis zum Porphyrin-c, sondern es finden sich im Hydrolysat auch Peptide des Porphyrins-c, Verbindungen also, in denen mit den porphyrin-gebundenen Cysteinresten noch andere Aminosäuren verknüpft sind. Diese gefärbten Produkte lassen sich von den im Hydrolysat vorkommenden freien Aminosäuren und porphyrinfreien Peptiden durch Adsorption an Talk abtrennen. Unterwirft man Porphyrin-c und Porphyrin-c-Peptide der Silbersulfat-Behandlung, welche Paul (1950) mit Erfolg zur Spaltung des intakten Cytochroms-c in Hämatoporphyrin und Apoprotein verwendet hat, so werden unter Sprengung der Thioätherbrücken Cystein und Cysteinpeptide in Freiheit gesetzt. Oxydation mit Perameisensäure verwandelt diese in Cysteinsäure und Cysteinsäurepeptide, welche durch Ionophorese und Papierchromatographie voneinander getrennt werden können. (Tuppy und Bodo, 1954 a). Die bei der Untersuchung von *Rinder*-Cytochrom-c erhaltenen und identifizierten Cysteinsäurepeptide zeigten (siehe Tabelle 1, Teil A), dass einer der beiden Cysteinreste des Porphyrins-c von Lysin und Alanin, der andere von Glutaminsäure und Histidin flankiert ist (Tuppy und Bodo, 1954 c).

Ein weiterer Einblick in die der prosthetischen Gruppe benachbarte Region des Apoproteins wurde durch einen enzymatischen Abbau des Cytochroms-c gewonnen (Tuppy und Paléus, 1955). Bei der Verdauung des Hämoproteins mit Pepsin entsteht, wie Tsou schon 1951 gezeigt hat, ein niedrigermolekulares Abbauprodukt, welches rot gefärbt ist und die intakte prosthetische Gruppe des Ferments enthält. Tsou's 'pepsinmodifiziertes Cytochrom-c' war jedoch nicht einheitlich; es bedurfte noch einer Reihe von Reinigungsoperationen (Fällungen mit Ammonsulfat und Trichloressigsäure, Ausflockung beim isoelektrischen Punkt und Verteilungschromatographie auf Hyflo-Supercel-Säulen), bis ein wirklich reines Produkt vorlag. Dieses 'Hämopeptid' enthielt, wie eine Aminosäureanalyse nach Stein und Moore ergab, abgesehen von den zwei porphyrin-gebundenen Cysteinresten, je einen Threonin-, Alanin-, Histidin- und Lysin-Rest, zwei Valin- und drei Glutaminsäure-Reste, insgesamt also 11 Aminosäurebausteine. Diese sind in einer einzigen Polypeptidkette angeordnet, an deren Aminoende, wie mit der DNP-Technik (Sanger, 1945) festgestellt wurde, ein Valinrest sitzt, gefolgt von Glutaminsäure. Vom Aminoende stammt auch ein Tripeptid, Val-Glu(NH₂)-Lys, welches vom Hämopeptid bei Inkubation mit Trypsin

Tabelle 1

ERMITTLUNG EINER AMINOSÄURESEQUENZ IM CYTOCHROM-C DES RINDES

A	Ermittlung der Nachbarschaft der porphyrin-gebundenen Cysteinreste durch partielle saure Hydrolyse des Cytochroms-c	<p>DNP-Val DNP-Val-Glu</p> <p style="text-align: center;">NH₂ Val-Glu-Lys</p> <p>Val-Glu Val-[Glu, Lys] Lys-CySO₃H [CySO₃H, Ala, Glu] Glu-[CySO₃H, His]</p> <p style="text-align: center;">Lys-CySO₃H Lys-[CySO₃H, Ala] CySO₃H-His</p> <p style="text-align: center;">Glu-[CySO₃H, His] CySO₃H-His</p>
B	Untersuchung eines aus Cytochrom-c durch Abbau mit <i>Pepsin</i> erhaltenen Hämopeptids	<p>Val-Glu Val-[Glu, Lys] Lys-CySO₃H [CySO₃H, Ala, Glu] Glu-[CySO₃H, His]</p> <p style="text-align: center;">NH₂ Val-Glu Ala-Glu [Lys, CySO₃H]</p> <p style="text-align: center;">NH₂ Val-Glu-Lys-CyS Ala-Glu-CyS His-Thr-Val-Glu</p> <p style="text-align: center;">NH₂ Val-Glu Ala-Glu [CySO₃H, His]</p> <p style="text-align: center;">NH₂ Val-Glu-Lys Ala-Glu-CyS His-Thr-Val-Glu</p> <p style="text-align: center;">NH₂ Val-Glu Ala-Glu [CySO₃H, His]</p> <p style="text-align: center;">NH₂ Val-Glu-Lys Ala-Glu-CyS His-Thr-Val-Glu</p>
C	Untersuchung eines aus Cytochrom-c durch Abbau mit <i>Trypsin</i> erhaltenen Hämopeptids	<p style="text-align: center;">NH₂ Val-Glu-Lys Ala-Glu-CyS His-Thr-Val-Glu</p> <p style="text-align: center;">NH₂ Val-Glu Ala-Glu [CySO₃H, His]</p> <p style="text-align: center;">NH₂ Val-Glu-Lys Ala-Glu-CyS His-Thr-Val-Glu</p> <p style="text-align: center;">NH₂ Val-Glu Ala-Glu [CySO₃H, His]</p> <p style="text-align: center;">NH₂ Val-Glu-Lys Ala-Glu-CyS His-Thr-Val-Glu</p> <p style="text-align: center;">NH₂ Val-Glu Ala-Glu [CySO₃H, His]</p> <p style="text-align: center;">NH₂ Val-Glu-Lys Ala-Glu-CyS His-Thr-Val-Glu</p> <p style="text-align: center;">NH₂ Val-Glu Ala-Glu [CySO₃H, His]</p> <p style="text-align: center;">NH₂ Val-Glu-Lys Ala-Glu-CyS His-Thr-Val-Glu</p>

abgespalten wird. Zur Aufklärung der vollständigen Aminosäuresequenz im Hämopeptid wurde dieses zuerst mittels Paul's Silbersalzmethode in seine Hämatin- und seine Polypeptid-Komponente zerlegt; in letzterer wurden mit Perameisensäure die zwei Cysteinreste oxydiert, mit dem Ergebnis, dass wir ein zwei Cysteinsäure-Reste enthaltendes, farbloses, porphyrin-freies Undekapeptid erhielten. Dieses wurde sowohl mit Salzsäure partiell hydrolysiert als auch mit Subtilisin proteolytisch gespalten; in beiden Fällen entstanden eine Reihe von Abbauprodukten, aus deren Struktur die Anordnung der 11 Aminosäurereste in der ungespaltenen Polypeptidkette ohne weiteres erschlossen werden konnte (Tabelle 1, Teil B).

Wird Cytochrom-c statt mit Pepsin mit Trypsin abgebaut, so erhält man (nach entsprechender Reinigung) ein dem oben beschriebenen Hämopeptid sehr ähnliches Spaltprodukt (Tuppy und Bodo, 1954 b), dessen Struktur ebenfalls durch partielle Säurehydrolyse ermittelt wurde (Tabelle 1, Teil C). Der Polypeptidanteil besteht hier aus nur 9 Aminosäureresten; die Nonapeptid-Sequenz ist auf der N-terminalen Seite um 3 Aminosäurereste kürzer und auf der C-terminalen um einen Rest länger als die im peptischen Abbauprodukt vorgefundene Undekapeptid-Sequenz.

Da, wie schon erläutert, die durch die prosthetische Gruppe markierte Region des Cytochrom-c-Apoproteins ein dankbares Objekt für Homologie-Studien ist, wurde sie ausser in Rinder-Cytochrom-c auch in zwei anderen Säugetier-Cytochromen, dem des Pferdes und des Schweines, untersucht. Tint und Reiss (1950) waren auf Grund elektrophoretischer Versuche zur Auffassung gelangt, dass die Säugetier-Cytochrome artspezifisch seien. Die von uns untersuchte, auf eine Sequenz von 10 Aminosäuren beschränkte Region war jedoch in den 3 Cytochromen vollständig gleich (Tuppy und Bodo, 1954 c) (Tabelle 2). Die nächste Prüfung betraf die peptischen Abbauprodukte eines Fisch- und eines Vogel-Cytochroms-c (Tuppy und Paléus, 1955). In dem des Lachses erwies sich die Aminosäuresequenz ebenfalls

Tabelle 2

VERGLEICH EINER HOMOLOGEN REGION IN CYTOCHROM-C
VERSCHIEDENER HERKUNFT

Rind	...-Val-Glu(NH ₂)-Lys-CyS-Ala-Glu(NH ₂)-CyS-His-Thr-Val-Glu-Lys-...
Pferd-Lys-CyS-Ala-Glu(NH ₂)-CyS-His-Thr-Val-Glu-Lys-...
Schwein-Lys-CyS-Ala-Glu(NH ₂)-CyS-His-Thr-Val-Glu-Lys-...
Lachs	...-Val-Glu(NH ₂)-Lys-CyS-Ala-Glu(NH ₂)-CyS-His-Thr-Val-Glu-.....
Huhn	...-Val-Glu(NH ₂)-Lys-CyS-Ser-Glu(NH ₂)-CyS-His-Thr-Val-Glu-.....
Seiden- spinner	...-Val-Glu(NH ₂)-Arg-CyS-Ala-Glu(NH ₂)-Cys-His-Thr-Val-Glu-.....
Hefe	...-Phe-Lys-Thr-.....-Arg-CyS-Glu--Leu-.....-CyS-His-Thr-Val-Glu-.....

als mit der in Säugetier-Cytochromen ermittelten identisch, in Hühner-Cytochrom-c hingegen war an die Stelle eines Alanin-Restes ein Serin-Rest getreten. Ein Invertebraten-Cytochrom-c, das des Seidenspinners Bombyx

mori (Tuppy, 1957), enthielt zwischen den beiden porphyrin-gebundenen Cystein-Resten wieder einen Alanin-Rest, so wie die Säugetier-Cytochrome; dafür ersetzte ein Arginin- den bisher angetroffenen Lysin-Rest. Eine bedeutendere Abweichung von diesen sehr ähnlichen Aminosäuresequenzen wurde erst beobachtet, als nicht mehr tierisches, sondern Hefe-Cytochrom-c zur Untersuchung kam (Tuppy und Dus, 1957). Aus Bäckerhefe konnten zwei Hämoproteine mit Cytochrom-c-Aktivität extrahiert und voneinander durch Ionenaustauscher-Chromatographie getrennt werden. Eines von ihnen, welches im Hefe-Extrakt in reichlicherer Menge vorkommt, wurde mit Pepsin abgebaut; die unterste Zeile der Tabelle 2 zeigt die Aminosäuresequenz im peptischen Spaltprodukt.

Die untersuchten homologen Cytochrome, ob sie nun vom Säugetier, Fisch, Vogel, Insekt oder sogar von Hefe stammen, sind einander spektroskopisch äusserst ähnlich; ihre Absorptionsbanden im sicht-baren Teil des Spektrums sind ununterscheidbar. Auch entsprechen sie einander in ihrer katalytischen Wirksamkeit; sie sind alle imstande, die spezifische Cytochrom-c-Funktion im Succinoxidase-System des Säugetiers zu erfüllen. Offenbar ist nicht nur die prosthetische Gruppe im engeren Sinne stets die gleiche, sondern es bleiben auch die Art ihrer Verknüpfung mit dem Proteinanteil und die Natur jener Gruppen des Proteins, die dem 'aktiven Zentrum' zuzurechnen sind, von Artunterschieden unberührt.

Zu den charakteristischen Invariablen der aufgeklärten homologen Aminosäuresequenzen gehören der stets gleiche Abstand der beiden porphyrin-gebundenen Cysteinreste und das benachbarte Vorkommen eines Histidin-Restes. Wie ein von Ehrenberg und Theorell (1955) gebautes Modell zeigte, ist die Verknüpfung der zwei Cysteinreste der Polypeptidkette mit den Vinylgruppen der prosthetischen Gruppe in zwangloser Weise möglich, wenn die Cysteinreste voneinander durch zwei andere Aminosäurereste getrennt sind und die Polypeptidkette die Form einer α -Helix hat; durch diese sekundäre Struktur der Polypeptidkette kommt der Histidin-Rest in eine räumliche Lage, die eine Bindung eines seiner Imidazol-Stickstoffatome an das Eisen der prosthetischen Gruppe geradezu nahelegt. Dieser Histidinrest stellt aller Wahrscheinlichkeit nach einen Teil des 'aktiven Zentrums' des Cytochrom-c-Moleküls dar; er ist als einer von jenen zwei Aminosäureresten der Proteinkomponente zu betrachten, die mit 'häm-gebundenen' basischen Gruppen die Hämochromogen-Natur des Cytochroms-c bewirken (Theorell und Åkeson, 1941).—Ob auch noch andere in allen untersuchten Cytochromen als gleich befundene Strukturelemente (so z. B. die Sequenz Thr-Val-Glu) darum invariabel sind, weil sie eine besondere funktionelle Bedeutung besitzen, ist derzeit nicht zu beantworten.

Wenn wir von Hefe-Cytochrom-c absehen und vorerst nur die untersuchten Tier-Cytochrome vergleichend betrachten, stellen wir fest, dass sich die Artspezifität der Struktur in der aufgeklärten Polypeptidregion auf den Austausch einzelner Aminosäurereste gegen andere, ähnliche beschränkt

(Lysin gegen Arginin, Alanin gegen Serin). Gleichartige Befunde sind von anderen Autoren an anderen homologen Polypeptiden und Proteinen erhoben worden. Es ist ebenfalls der Erstaz eines Lysin-Restes durch einen Arginin-Rest, der das Vasopressin des Schweines von dem des Rindes unterscheidet (Du Vigneaud, Lawler und Popenoe, 1953), und der Austausch eines Alanins gegen Serin ist einer der geringen Unterschiede zwischen Schweine- und Schaf-Corticotropin (Bell, 1954; Li, Geschwind, Cole, Raacke, Harris und Dixon, 1955). Drei Paare ähnlicher Aminosäurereste, Alanin und Threonin, Glycin und Serin, Valin und Isoleucin, vertreten einander in den von Sanger und Mitarbeitern miteinander verglichenen fünf homologen Säugetier-Insulinen (Tabelle 3) (Harris, Sanger und Naughton, 1956). Das alternative

Tabelle 3

ARTUNTERSCHIEDE IN INSULIN
(NACH HARRIS, SANGER UND
NAUGHTON, 1956)

Rind	...-CyS-Ala-Ser-Val-CyS...
Schwein	...-CyS-Thr-Ser-Ileu-CyS...
Schaf	...-CyS-Ala-Gly-Val-CyS...
Pferd	...-CyS-Thr-Gly-Ileu-CyS...
Wal	...-CyS-Thr-Ser-Ileu-CyS...

Vorkommen von Alanin und Threonin ist ausserdem auch in homologen Serumalbuminen (Thompson, 1954), von Valin and Isoleucin im Hypertensin des Rindes und Pferdes (Elliott und Peart, 1956; Skeggs, Lentz, Kahn, Shumway und Woods, 1956), von Glycin und Serin in Hämoglobin verschiedener Herkunft (Ozawa und Satake, 1955) beobachtet worden; in Hämoglobinen wurde auch ein Ersatz von Valin durch Methionin und von Glutaminsäure durch Asparaginsäure gefunden (Porter und Sanger, 1948; Ozawa und Satake, 1955). Eine der Artspezifität entsprechende 'Stammesspezifität' stellten Niu und Fraenkel-Conrat (1955) in der Proteinkomponente des Tabakmosaikvirus fest; in vier Stämmen lautet die C-terminale Sequenz Thr-Ser-Gly-Pro-Ala-Thr, in einem fünften kommen anstelle von Serin und Glycin die diesen ähnlichen Aminosäuren Alanin und Threonin vor.

Wenn wir nun die aufgeklärte Aminosäuresequenz im Hefe-Cytochrom-c mit derjenigen im Cytochrom-c verschiedener Tiere vergleichen, so sehen wir an mehreren einander entsprechenden Stellen der Polypeptidketten unähnliche Aminosäurereste stehen. Diese Feststellung trifft z. B. auf jene zwei Aminosäuren zu, welche die beiden porphyrin-gebundenen Cystein-Reste voneinander trennen; hier finden wir Alanin bzw. Serin durch Glutaminsäure ersetzt und Glutamin durch Leucin. Für eine solche gegenseitige Vertretung unähnlicher Aminosäuren sind bisher nur vereinzelte andere Beispiele beschrieben worden. In zwei homologen Melanophoren-stimulierenden

Hormonen ersetzen einander Serin und Glutaminsäure (Geschwind, Li und Barnafi, 1957; Harris und Roos, 1956), in zwei homologen corticotropen Hormonen Alanin und Leucin (Bell, 1954; Li, Geschwind, Cole, Raacke, Harris und Dixon, 1955).

Ein Ziel solcher vergleichender Untersuchungen wäre die Kenntnis der Regeln, welche die Abwandlung der Struktur in homologen Proteinen in Abhängigkeit von ihrer biologischen Stellung bestimmen; leider lassen sich derartige Regeln, wenn es solche überhaupt gibt, aus den beschränkten bisher erhobenen Befunden nicht ableiten.

Eine wichtige Frage ist die, ob kleine Unterschiede der Aminosäuresequenz in Proteinen, etwa Auswechslungen einzelner Aminosäurereste gegen ähnliche andere, wie sie im Rahmen von Artunterschieden festgestellt worden sind, auch innerhalb derselben Organismenart auftreten können. Bei der Strukturauflklärung des Insulins ist kein Hinweis gefunden worden, dass es Alternativen der Aminosäuresequenzen in den Polypeptidketten gäbe, wodurch eine Mikroheterogenität der primären Proteinstruktur zustande käme (Sanger und Tuppy, 1951; Sanger und Thompson, 1953). Zwei voneinander durch Gegenstromverteilung getrennte Formen des Insulins (Harfenist und Craig, 1952 a) unterscheiden sich nicht in der Aminosäureanordnung, sondern nur durch eine Säureamidgruppe (Harfenist, 1953). Neuerdings halten jedoch Grassmann, Strobel, Hannig und Deffner-Plöckl (1956) eine Mikroheterogenität der Fraktion A des Rinder-Insulins (ein alternatives Vorkommen von Isoleucin und Valin) dennoch für möglich. Die Untersuchung der porphyrin-nahen Region des Cytochrom-c-Apoproteins förderte, wenigstens bisher, keinen intraspezifischen Sequenzunterschied zutage. Auch bei der Strukturauflklärung der Polypeptid-Hormone Oxytocin, Vasopressin, Corticotropin, Intermedin und Glukagon wurden diese als homogen befunden.

Andererseits sind in letzter Zeit mehrere Fälle bekannt geworden, in denen Proteine, aus Individuen der gleichen Art gewonnen, in zwei oder zahlreicheren ähnlichen Formen auftreten. Das bekannteste Beispiel ist das menschliche Hämoglobin, das nicht nur in einer normalen adulten Form und einer fötalen Form existiert, sondern auch in einer Reihe abnormaler Formen (Zusammenfassung bei Itano, 1956), unter denen das Sichelzellanämie-Hämoglobin das erstgefundene (Pauling, Itano, Singer und Wells, 1949) und bestuntersuchte ist. Die Synthese einiger von diesen verschiedenen, elektrophoretisch unterscheidbaren Hämoglobinen steht unter der Kontrolle aller Gene. Noch ist eine genaue strukturelle Differenzierung der einzelnen Hämoglobine nicht erfolgt, durch die sicher erwiesen wäre, ob die Verschiedenheit in der Aminosäuresequenz oder aber in der sekundären Struktur der Proteinmoleküle liegt. Die vergleichende Untersuchung von Ingram (1956) an normalem und an Sichelzell-Hämoglobin spricht für das erstere; die Unterschiede scheinen nur sehr gering zu sein und könnten in der Grössenordnung der bei Artspezifitätsstudien gefundenen liegen.

Vom biologischen Standpunkt aus erscheint es durchaus befriedigend, dass

ein und dasselbe Protein innerhalb derselben Spezies in zwei oder mehreren homologen Formen vorkommen kann, die durch Einzelheiten der Aminosäuresequenz unterschieden sind. Wenn die Proteinsynthese genetisch gesteuert ist und wenn eine Änderung der Aminosäuresequenz, wie sie im Rahmen von Artunterschieden auftritt, Ausdruck und Folge einer interspezifischen Verschiedenheit von Erbfaktoren ist, welche die Proteinsynthese kontrollieren, so lässt sich auch bei intraspezifischem Vorkommen verschiedener, für die Proteinsynthese verantwortlicher Erbfaktoren (bei Heterozygotie eines Individuums bzw. Heterogenie einer Population artgleicher Individuen) das Auftreten eines Proteins mit Verschiedenheiten in der Aminosäuresequenz erwarten. In einem heterozygoten Organismus sollte ein Protein in zwei verschiedenen Formen, in einer heterogenen Population in ebensovielen Formen vorkommen können, als Allele des betreffenden Erbfaktors existieren.

Ob dem wirklich so ist, dass sich in Änderungen der Aminosäureanordnung in Proteinen—direkt oder eher indirekt—mutative Veränderungen von Genen widerspiegeln, wird die Zukunft zeigen. Jedenfalls aber sichert der genetische und entwicklungsgeschichtliche Aspekt den experimentellen Untersuchungen sowohl der intra- als auch der interspezifischen Unterschiede der Proteinstruktur ein besonderes Interesse.

REFERENCES

- BELL, P. H. (1954), *J. Amer. Chem. Soc.*, **76**, 5565.
 BROWN, H., SANGER, F. und KITAI, R. (1955), *Biochem. J.*, **60**, 556.
 DU VIGNEAUD, V., LAWLER, H. C. und POPENOE, E. A. (1953), *J. Amer. Chem. Soc.*, **75**, 4880.
 ELLIOTT, D. F. und PEART, W. S. (1956), *Nature*, **177**, 527.
 EHRENBERG, A. und THEORELL, H. (1955), *Acta Chem. Scand.*, **9**, 1193.
 GESCHWIND, I. I., LI, C. H. und BARNAFI, L. (1957), *J. Amer. Chem. Soc.*, **79**, 620, 1003.
 GRASSMANN, W., STROBEL, R., HANNIG, K. und DEFFNER-PLÖCKL, M. (1956), *Z. physiol. Chem.*, **305**, 21.
 HARFENIST, E. J. (1953), *J. Amer. Chem. Soc.*, **75**, 5528.
 HARFENIST, E. J. und CRAIG, L. C. (1952 a), *J. Amer. Chem. Soc.*, **74**, 3083.
 HARFENIST, E. J. und CRAIG, L. C. (1952 b), *J. Amer. Chem. Soc.*, **74**, 4216.
 HARRIS, J. I. und ROOS, P. (1956), *Nature*, **178**, 90.
 HARRIS, J. I., SANGER, F. und NAUGHTON, M. A. (1956), *Arch. Biochem. Biophys.*, **65**, 427.
 INGRAM, V. M. (1956), *Nature*, **178**, 792.
 ITANO, H. A. (1956), *Ann. Rev. Biochem.*, **25**, 331.
 KNIGHT, C. A. (1947), *J. Biol. Chem.*, **171**, 297.
 LI, C. H., GESCHWIND, I. I., COLE, R. D., RAACKE, I. D., HARRIS, J. I. und DIXON, J. S. (1955), *Nature*, **176**, 687.
 NIU, C. I. und FRAENKEL-CONRAT, H. (1955), *Arch. Biochem. Biophys.*, **59**, 538.
 OZAWA, H. und SATAKE, K. (1955), *J. Biochem. (Japan)*, **42**, 641.
 PAUL, K. G. (1950), *Acta Chem. Scand.*, **4**, 239.
 PAULING, L., ITANO, H. A., SINGER, S. J. und WELLS, I. C. (1949), *Science*, **110**, 543.
 PORTER, R. R. und SANGER, F. (1948), *Biochem. J.*, **42**, 287.

- SANGER, F. und THOMPSON, E. O. P. (1953), *Biochem. J.*, **53**, 353, 366.
- SANGER, F. und TUPPY, H. (1951), *Biochem. J.*, **49**, 463, 481.
- SKEGGS, L. T., LENTZ, K. E., KAHN, J. R., SHUMWAY, N. P. und WOODS, K. R. (1956), *J. Exp. Med.*, **104**, 193.
- THEORELL, H. (1939), *Enzymologia*, **6**, 88.
- THEORELL, H. und ÅKESON, Å. (1941), *J. Amer. Chem. Soc.*, **63**, 1804.
- THOMPSON, E. O. P. (1954), *J. Biol. Chem.*, **208**, 565.
- TINT, H. und REISS, W. (1950), **182**, 397.
- TSOU, C. L. (1951), *Biochem. J.*, **49**, 362.
- TUPPY, H. (1957), *Z. Naturf.*, **126**, 784.
- TUPPY, H. und BODO, G. (1954 a), *Monatsh.*, **85**, 807.
- TUPPY, H. und BODO, G. (1954 b), *Monatsh.*, **85**, 1024.
- TUPPY, H. und BODO, G. (1954 c), *Monatsh.*, **85**, 1182.
- TUPPY, H. und DUS, K. (1958), im Druck.
- TUPPY, H. und PALÉUS, S. (1955), *Acta Chem. Scand.*, **9**, 353.
- ZEILE, K. und MEYER, H. (1939), *Naturwiss.*, **27**, 596.

Comparative studies in the field of protein microstructure

F. ŠORM

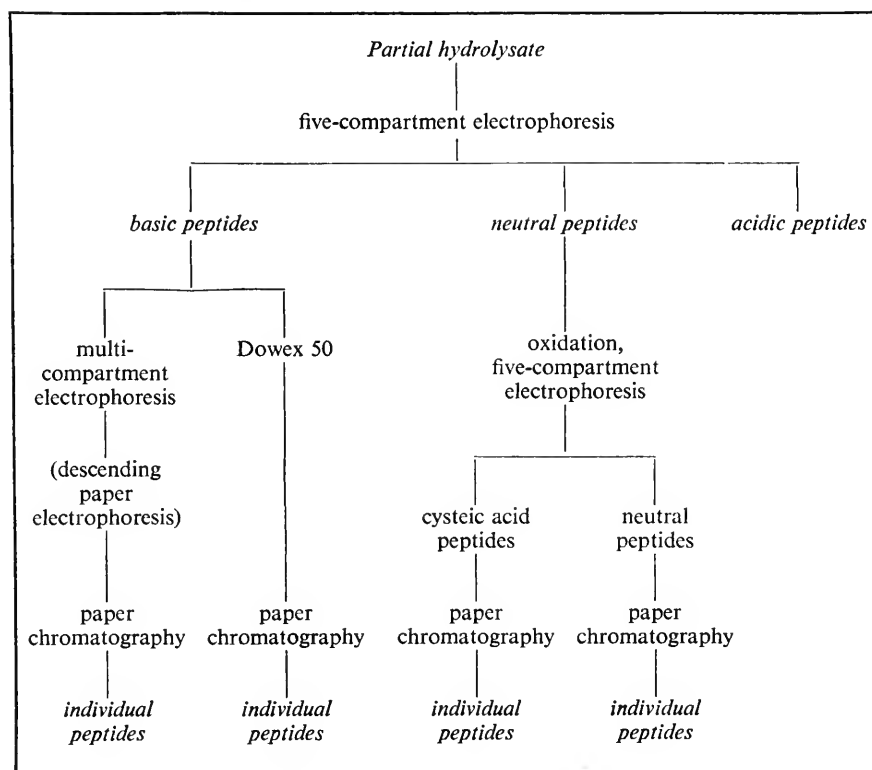
Institute of Chemistry,
Czechoslovak Academy of Science, Prague

Work in the field of protein microstructure is at present being directed along two fundamental lines: the first of these is characterized by efforts to determine the complete structure of the simplest proteins; work along the second line seeks to discover general regularities in protein structure on the basis of comparative studies of lower fission products of the proteins. Practical difficulties at present limit the scope of the first line of research to the higher natural peptides and proteins of very low molecular weight. In spite of these limitations, a number of notable successes has, as is generally known, been achieved in this field, culminating at present in the determination of the complete structure of ribonuclease.¹ The exact determination of the amino acid sequence in the peptide chains of the simpler proteins and of natural peptides undoubtedly also contributes valuable information to our knowledge of the general laws governing protein structure. The results of the work carried out so far must, however, for the present be evaluated with some caution in this respect, since the compounds concerned are mostly possessed of highly specific biological activities which very probably reflect a corresponding specificity of chemical structure. For these reasons, the second fundamental line of protein research, which is not limited by the molecular weight of the materials studied, is gaining increasing importance. The aim of such research may be either the elucidation of regularities in the arrangement of amino acids within the protein chains—a line hitherto pursued mainly in work on the fibrous proteins—or a search for resemblances in the chemical structure of proteins related in their biological function or origin. The search for such resemblances in protein structure has up to the present more particularly involved those parts of the peptide chains whose nature makes them amenable to direct comparison at the present level of experimental technique. Such is the case, for instance, with the amino acid groupings terminating the peptide chains, which may be exactly determined by identification of the terminal amino acids or peptide sequences. A second

approach to this problem involves the separation from partial hydrolysates and the further study of such peptides as differ markedly in some manner from the other peptides in the hydrolysate, and which may by virtue of these differences be isolated as a separate group. A very interesting contribution of this kind is the work of Tuppy who, by a comparison of the coloured peptides obtained by the partial hydrolysis of cytochrome-*c*, established the identity of the amino acid sequence in the neighbourhood of the point of attachment of the pigment in a number of animal species.²

In our laboratory we have for some time past been carrying out comparative studies on the microstructure of certain proteins,³ in particular of chymotrypsinogen and trypsinogen, and of the haemoglobins and myoglobins of various animal species. In this work we concentrate on the isolation and comparison of lower peptides containing amino acids which constitute only a relatively small fraction of the total residue number of the proteins studied and which, moreover, are easily separated from the other peptides. These conditions are met particularly in the case of peptides containing cysteic acid (derived from cystine and cysteine by oxidation), arginine, lysine, or

Table 1
FRACTIONATION SCHEME FOR PEPTIDE MIXTURES



histidine. For practical reasons we initially confined ourselves to the analysis of partial acid hydrolysates in which about 50 per cent. of the peptide bonds had been preserved intact and which consisted of relatively low molecular weight peptides. The isolation of the cysteic acid peptides was at first carried out by means of ion exchange on Amberlite IRA-4B, that of the arginine peptides on Amberlite IRA-400, particularly in those cases where only a statistical analysis of the overall composition of the appropriate peptide fractions was carried out. Later, when we undertook the isolation and identification of the individual peptides contained in each fraction, we developed a fractionation procedure whose essentials are set out schematically in Table 1.

As an example of the statistical procedure in comparing the microstructure of proteins, some of our results on the serum albumins of a number of species—man, ox, horse, sheep, and duck—may be given.⁴ As is shown in Table 2, the amino acid composition of all five proteins is very similar. The

Table 2

AMINO ACID COMPOSITION OF SOME SERUM ALBUMINS

Amino acid	Species				
	ox	man	horse	duck	sheep
Cysteic acid	6.5 ± 0.3	6.3 ± 1.1	7.6 ± 0.6	6.9 ± 0.2	6.1 ± 0.28
Lysine+					
histidine	16.8 ± 0.05	15.3 ± 0.19	17.05 ± 0.18	13.3 ± 0.3	17.1 ± 0.28
Arginine	5.2 ± 0.82	6.25 ± 0.30	5.2 ± 0.21	5.4 ± 0.16	5.6 ± 0.14
Aspartic acid	10.9 ± 0.3	11.9 ± 0.0	10.2 ± 0.1	13.1 ± 0.2	11.6 ± 0.25
Serine	4.2 ± 0.12	3.53 ± 0.4	4.18 ± 0.22	6.1 ± 0.12	4.3 ± 0.12
Glycine	1.8 ± 0.05	1.62 ± 0.05	2.34 ± 0.14	2.5 ± 0.14	2.5 ± 0.16
Glutamic acid	16.5 ± 0.3	17.2 ± 0.0	16.65 ± 0.48	16.1 ± 0.22	15.33 ± 0.36
Threonine	5.8 ± 0.5	3.93 ± 0.36	3.6 ± 0.26	4.1 ± 0.12	5.2 ± 0.12
Alanine	6.25 ± 0.08	8.2 ± 0.2	7.85 ± 0.45	6.6 ± 0.15	6.7 ± 0.07
Proline	4.7 ± 0.0	4.6 ± 0.0	4.3 ± 0.0		
Tyrosine	5.0 ± 0.1	3.75 ± 0.8	3.88 ± 0.02	10.8 ± 0.15	9.3 ± 0.3
Methionine	0.8 ± 0.03	1.77 ± 0.12	—	4.5 ± 0.2	1.0 ± 0.09
Valine	5.9 ± 0.35	7.76 ± 0.3	4.0 ± 0.23	4.2 ± 0.26	6.8 ± 0.23
Phenylalanine	6.6 ± 1.02	8.1 ± 0.87	8.25 ± 0.77	6.6 ± 0.36	9.0 ± 0.57
Leucine+					
isoleucine	14.8 ± 0.42	15.18 ± 0.5	15.10 ± 0.77	11.3 ± 0.39	14.6 ± 0.26
Methionine	0.76	1.14	0.46	1.14	0.81
Tyrosine	5.62	5.20	4.43	5.66	5.67
Tryptophan	1.43	0.88	1.08	0.62	1.11

The figures denote grams of amino acid per 100 g. of protein; they are average values from 8 independent determinations (6 in the case of glutamic acid, threonine, methionine, valine, and phenylalanine). The probable error is indicated after each value. The values for methionine, tyrosine and tryptophan recorded in the last three lines were obtained by specific colorimetric methods.

quantitative determination of the individual amino acids in the total hydrolysate was here, as in all other analyses recorded in this paper, carried out by photometric measurement on paper chromatograms by the method of B. Keil.⁵ Table 3 shows the overall amino acid composition of the arginine peptide fraction isolated from the partial acid hydrolysates of the five serum albumins by means of Amberlite IRA-400. The table reveals marked differences in the amounts of several of the amino acids, evidently connected with differences in the microstructure of the individual serum albumins. In particular, the conspicuously high threonine content of the peptide fractions obtained from ox and sheep albumin may be noted, as well as the very low tyrosine content of the peptides from duck serum albumin. This last feature is particularly characteristic since all the native proteins contain about the same amount of tyrosine.

Table 3

AMINO ACID COMPOSITION OF TOTAL HYDROLYSATES OF THE ARGININE PEPTIDES FROM SOME SERUM ALBUMINS

Amino acid	Species				
	ox	man	horse	duck	sheep
Cysteic acid	4.5	5.3	4.8	7.4	3.8
Lysine+histidine	9.2	13.3	10.5	12.4	9.7
Arginine	28.5	29.9	29.6	30.1	24.4
Serine+glycine	6.6	4.2	5.1	6.0	5.4
Glutamic acid+					
+threonine	9.8	2.8	4.3	4.1	9.3
Alanine	5.2	5.0	5.9	6.3	6.0
Proline	11.5	12.3	13.2	11.3	11.0
Tyrosine	3.7	4.5	3.6	0.4	1.6
Methionine+valine	3.7	5.3	2.9	3.3	6.0
Phenylalanine	0.9	1.6	1.6	0.9	3.2
Leucine+isoleucine	16.2	18.7	18.3	16.7	19.8

Our main concern in the work carried out so far has been with the comparison of chymotrypsinogen and trypsinogen. These two proteins—both of approximately the same molecular weight—are formed by a single type of pancreatic cell, resemble each other in their enzymic function, and can be obtained crystalline in a state of high purity. In choosing these proteins we further considered that the results of our work might contribute to the determination of their full structures—a problem which will doubtlessly receive considerable attention in the near future.

The amino acid composition of the two enzymogens as determined in our laboratory is set out in Table 4, which reveals considerable differences between the two proteins in this respect. Chymotrypsinogen contains much less serine than trypsinogen and more threonine and alanine. The Table

Table 4

AMINO ACID COMPOSITION (RESIDUE NUMBERS)
OF BOVINE CHYMOTRYPSINOGEN AND
TRYPSINOGEN

Amino acid	Chymotrypsinogen	Trypsinogen
Alanine	19	13
Glycine	20	21
Valine	20	15
Leucine	16	12
Isoleucine	9	12
Proline	8	7
Phenylalanine	6	4
Tyrosine	4	9
Tryptophan	6	—
Serine	26	38
Threonine	20	(9-11)
Cysteine (half-cystine)	9-10	11-13
Methionine	(1)	1
Arginine	4	2
Histidine	2	3
Lysine	12	14
Aspartic acid	19	24
Glutamic acid	10	10
Amide N	23	(23±3)

further shows that the residue numbers of the basic amino acids, particularly of arginine and histidine, in chymotrypsinogen and trypsinogen (4 and 2, and 2 and 3 respectively) are well suited for our comparative studies, as well as the relatively low number of cysteine (half-cystine) residues (9-10 and 11-13, respectively). At the outset we again carried out a statistical comparison of the amino acid composition of total hydrolysates of the cysteic acid peptide fractions isolated from the partial hydrolysates of the performic acid—oxidized proteins by means of Amberlite IRA-4B.⁶ Table 5, in which the amino-acid compositions of the total hydrolysates of several proteins and those of the corresponding cysteic acid peptides are set out, indicates that the degree of resemblance between the amino acid composition of hydrolysates obtained from the cysteic acid peptides of chymotrypsinogen and trypsinogen, respectively, is about the same as that between the composition of the two whole proteins. At the same time, however, the table reveals significant differences between the composition of hydrolysates of the cysteic acid peptides of the two enzymogens on the one hand, and of the remaining proteins on the other.

The statistical evaluation of the composition of the cysteic acid peptide fractions naturally gives only very approximate information about resemblances in the structure of the proteins under examination; the results have

STATISTICAL COMPARISON OF THE AMINO ACID COMPOSITION OF SOME

	Proteins														$\bar{\delta}$
	Leu+Ileu		Val		Ala		Thr		Glu		Ser, Gly, Met		Asp		
	%	Δ	%	Δ	%	Δ	%	Δ	%	Δ	%	Δ	%	Δ	
Chymotrypsinogen	<i>16.1</i> 13.4 100		<i>10.1</i> 8.4 100		<i>7.6</i> 6.3 100		<i>11.4</i> 9.5 100		<i>9.0</i> 7.5 100		<i>18.2</i> 15.1 100		<i>11.3</i> 9.4 100		
Trypsinogen	<i>18.2</i> 14.9 111	11	<i>9.0</i> 7.3 87	13	<i>5.8</i> 4.7 75	25	<i>5.4</i> 4.4 46	54	<i>10.7</i> 8.7 116	16	<i>20.5</i> 16.7 111	11	<i>13.6</i> 11.1 118	18	± 10.4
Serum albumin bovine	<i>14.8</i> 13.1 98	2	<i>5.9</i> 5.2 62	38	<i>6.2</i> 5.5 87	13	<i>5.8</i> 5.1 54	46	<i>16.5</i> 14.6 195	95	<i>6.8</i> 6.0 40	60	<i>10.9</i> 9.6 102	2	± 19.7
Serum albumin human	<i>12.7</i> 11.9 89	11	<i>7.7</i> 7.2 86	14	<i>6.2</i> 5.8 92	8	<i>5.0</i> 4.7 49	51	<i>17.4</i> 16.3 218	118	<i>6.6</i> 6.2 41	59	<i>10.4</i> 9.7 103	3	± 22.0
Ovalbumin	<i>16.2</i> 14.8 110	10	<i>7.1</i> 6.5 77	23	<i>6.7</i> 6.1 97	3	<i>4.0</i> 3.6 38	62	<i>16.5</i> 15.0 200	100	<i>16.4</i> 15.0 99	1	<i>9.3</i> 8.5 90	10	± 18.6
Insulin	<i>16.0</i> 14.1 105	5	<i>7.8</i> 6.9 82	18	<i>4.5</i> 4.0 63	37	<i>2.1</i> 1.9 20	80	<i>18.6</i> 16.4 218	118	<i>9.5</i> 8.4 56	44	<i>6.8</i> 6.0 64	36	± 24.5
Globin bovine	<i>15.4</i> 14.4 107	7	<i>9.1</i> 8.5 101	1	<i>7.4</i> 6.9 110	10	<i>4.4</i> 4.1 43	57	<i>8.5</i> 7.9 105	5	<i>12.4</i> 11.6 77	23	<i>10.6</i> 9.9 105	5	± 9.7
Fibrinogen bovine	<i>11.9</i> 10.8 81	19	<i>4.1</i> 3.7 44	56	<i>3.7</i> 3.4 54	46	<i>6.1</i> 5.5 58	42	<i>14.5</i> 13.2 176	76	<i>15.2</i> 13.8 91	9	<i>13.1</i> 11.9 126	26	± 18.2

In the columns headed '%' the italic numbers denote the number of grams of amino acid obtained from 100 g. of protein, the numbers in roman type denote the same results recalculated to add up to 100%, and the figures in bold-face type denote the amount of the amino acid expressed as a percentage of the amount of the same amino acid present in chymotrypsinogen. In the columns headed 'cm²' the italic figures denote the area

PROTEINS AND THE CORRESPONDING ACIDIC PEPTIDE FRACTIONS

Acidic Peptides														
Leu+Ileu		Val		Ala		Thr		Glu		Ser, Gly, Met		Asp		$\bar{\delta}$
cm ²	Δ	cm ²	Δ	cm ²	Δ	cm ²	Δ	cm ²	Δ	cm ²	Δ	cm ²	Δ	
24.6 89.0 100		13.6 49.2 100		22.5 81.4 100		15.3 55.3 100		54.9 198.4 100		55.5 200.5 100		21.7 78.5 100		
29.5 117.7 132	32	11.0 43.9 89.2	11	20.2 80.6 99.1	1	10.5 42.0 76.0	24	48.0 163.0 82.2	18	52.8 210.5 105.0	5	22.2 88.5 113.0	13	± 7.5
22.4 94.3 106	6	7.5 31.6 64.1	36	18.4 77.5 95.2	5	11.6 49.0 88.5	11	73.2 308.5 155.5	56	16.2 68.4 34.1	66	18.4 77.5 98.8	1	± 14.8
20.7 90.1 101	1	11.5 50.0 101.5	2	25.2 109.8 135	35	7.0 30.5 54.2	46	76.4 332.2 167.2	67	17.8 77.5 38.6	61	24.7 107.5 137.0	37	± 17.5
19.1 94.4 106	6	20.4 100.9 205.0	105	19.8 97.8 120.1	20	5.5 22.5 40.7	59	56.2 278.0 140.0	40	26.6 131.2 65.5	34	18.0 89.0 113.2	13	± 12.8
1.4 18.2 155	55	23.2 77.5 157.6	58	19.1 63.8 78.5	21	4.0 13.4 24.2	76	104.6 349.0 176.0	76	46.7 156.0 77.8	22	24.3 81.2 103.5	4	± 21.1
27.5 5.0 73.0	27	53.6 126.8 257.5	158	22.5 53.2 65.4	35	15.4 36.4 65.8	34	105.3 249.0 125.5	26	54.0 119.0 59.4	41	39.5 93.4 119	19	± 27.0
4.5 6.0 85.4	15	33.1 16.7 208.0	108	19.4 102.7 74.0	26	12.0 61.0 67.2	33	86.4 37.2 135	35	52.8 268.0 81.5	18	33.3 163.7 135.2	35	± 19.8

under the peaks of the photometric curves, the figures in roman type the same results referred to 1000 as the sum of all areas, and the numbers in bold-face type the results expressed as percentages of the appropriate value for chymotrypsinogen. The columns headed Δ list the differences from the chymotrypsinogen values, and $\bar{\delta}$ is the mean deviation from the chymotrypsinogen values.

been given here more particularly for the reason that this method might have some justification in the comparison of highly complex proteins where there is no possibility of isolating and identifying a reasonable proportion of individual peptides. In our further work on the cysteic acid peptides of chymotrypsinogen and trypsinogen* we undertook the isolation and identification of individual peptides. For this purpose we used a cysteic peptide fraction obtained by oxidizing the neutral peptide fraction, isolated by electrophoresis in a five-compartment apparatus, with performic acid; the

Table 6

CYSTEIC ACID PEPTIDES ISOLATED FROM PARTIAL ACID
HYDROLYSATES OF BOVINE CHYMOTRYPSINOGEN AND
TRYPSINOGEN

Chymotrypsinogen		Trypsinogen
1	his.cys	his.cys
2	leu.(pro, val).cys gly.(leu, pro, val).cys	val.cys gly.(pro, val).cys
3	val.cys ala.val.cys (val, cys, glu)	ser.cys cys.ala cys.ala.gly
4	thr.cys cys.val thr.cys.val ser.thr.cys	thr.cys
5	phe.cys cys.gly (phe, cys).gly	phe.cys cys.gly
6	gly.cys.ser	ser.gly.cys
7	ileu.cys ileu.cys.ala cys.ala	cys.leu
8	ser.cys cys.met ser.cys.met ser.(cys, met, ileu) (cys, met, lieu).ala	
9	thr.asp.cys	(ser, glu).cys

* Actually, trypsin was used in these experiments since it differs from the enzymogen by a peptide sequence irrelevant to this particular study.

strongly acidic portion was roughly fractionated on Dowex 2 and the individual peptides were isolated by repeated paper chromatography. The individual cysteic acid peptides obtained from both proteins are listed in Table 6.⁷ It is evident at first sight that our partial hydrolysates of trypsin yielded a far smaller number of peptides than the partial hydrolysates of chymotrypsinogen; this may be due to differences in the extent of hydrolysis in the two cases.

It is further evident from Table 6 that, taking in conjunction those peptides which are certainly or probably connected, it is possible to make out nine different sequences for the immediate environment of the cysteic acid residues in chymotrypsinogen; this corresponds exactly to the number of cysteine (half-cystine) residues thought to be present in the chymotrypsinogen molecule. It may also be seen that, although the number of cysteic acid peptides isolated from the partial hydrolysate of trypsin is much smaller, their structures conform very closely to four of the peptide sequences found in the case of chymotrypsinogen.

For the isolation of the peptides of arginine, histidine and lysine from the partial acid hydrolysates of chymotrypsinogen and trypsin we used the basic peptide fractions obtained in a five-compartment electrophoretic apparatus. These were further subjected to electrophoresis in a multi-compartment apparatus under conditions permitting the direct separation of a weakly basic fraction composed predominantly of histidine-containing peptides, and at the same time a rough fractionation of the strongly basic peptides of arginine and lysine. The final isolation of individual peptides was achieved by repeated paper chromatography in several solvent systems.

The compositions or structures of the individual peptides isolated in this manner are summarized in Table 7;^{8,13} the results show that we have succeeded in establishing the amino acid sequence in the immediate vicinity of two of the four arginine residues present in chymotrypsinogen, and confirm a third sequence (leu.ser.arg.ileu.)* which had previously been established from activation studies. The amino-acid sequence about the fourth residue is probably val.arg.asp or val.arg.glu, since we found bound arginine to occur in the neutral peptide fractions of the hydrolysate. Of the dipeptides isolated from partial hydrolysates of trypsin, two correspond in structure with sequences identified in chymotrypsinogen.

Table 7 also shows the structures of the histidine peptides isolated in this work.⁹ Here again we were able, to a fair degree of certainty, to establish the immediate environment of both the histidine residues present in chymotrypsinogen. The peptides isolated from trypsin show agreement with chymotrypsinogen only for one of the three histidine residues; it further appears that the sequence phe.his in chymotrypsinogen may be replaced by tyr.his in trypsinogen.

* This sequence has been confirmed by enzymic hydrolysis just before this paper was given.

Table 7

PEPTIDES OF ARGININE AND HISTIDINE ISOLATED
FROM PARTIAL ACID HYDROLYSATES OF BOVINE
CHYMOTRYPSINOGEN AND TRYPSINOGEN

Chymotrypsinogen		Trypsinogen	
1	ser.arg arg.ileu leu.ser.arg.ileu*	1	ser.arg arg.leu
2	ala.arg. arg.val ala.arg.val ala.(arg, val, lys)		
3	thr.arg thr.arg.tyr		
4	val.arg thr.(asp, ala).arg (enzymic)	2	val.arg ser.(val, arg).ala
1	ala.his his.cys ala.his.cys	1	ala.his his.cys
2	phe.his his.phe	2	tyr.his
		3	pro.his

* This sequence has been confirmed very recently by enzymic hydrolysis.

Table 8 shows that there is also a marked resemblance in the composition and structure of the lysine-containing peptides of the two proteins, although in this case, owing to the large number of lysine residues present and the small number of peptides isolated, no definite conclusions as to their structural similarities can be drawn.

A general review of our preliminary results on the comparison of the microstructure of chymotrypsinogen and trypsinogen first of all shows that our findings are in full agreement with the structural fragments so far detected in these proteins by other workers. The extensive structural similarity established for a number of peptides at defined sites in the molecules of the two proteins permits the conclusion that there is considerable resemblance in their microstructures. However, the molecules must necessarily also contain larger peptide fragments which differ in the two cases; this follows

Table 8

PEPTIDES OF LYSINE ISOLATED FROM PARTIAL
ACID HYDROLYSATES OF BOVINE CHYMO-
TRYPSINOGEN AND TRYPSINOGEN

Chymotrypsinogen	Trypsinogen
lys.lys gly, lys	gly.lys (i)leu.lys ser.lys thr.lys
ser.lys thr.lys val.lys	tyr.lys lys.ala
lys.(ala, ileu) lys.ileu lys.leu	lys.leu lys.pro
lys.val lys.(val, phe) lys, cys lys, glu lys, glu, ileu	lys, cys lys, glu lys, glu, (i)leu lys, leu, ileu
ser.(asp, cys, thr).lys (enzymic)	

particularly from the small number of different peptide sequences involving cysteine⁶ detected in partial hydrolysates of trypsinogen, which contrasts with the total number of cysteine (half-cystine) residues (11-13) present in this protein. Finally, the results of our work on the arginine peptides may prove of value in the concluding phases of the full structure determination of the two proteolytic enzymes since they establish—unambiguously in the case of chymotrypsinogen—the amino acid sequence at those very sites at which fission of the peptide chains takes place in controlled enzymic hydrolysis.

I should further like to give an account of some of our preliminary findings in a comparison of the microstructure of two haemoglobins (horse and hog)—proteins where, for practical reasons, we may not for the time being expect an attack on the total structure. In this work we have so far concentrated on the comparison of the peptides of arginine¹⁰ and histidine¹¹ (of which the haemoglobin molecule contains 14 and 36-38 residues, respectively), and partly those of lysine. The isolation of the basic peptides from partial hydrolysates of the globins prepared from pure, crystalline samples of the haemoglobins was carried out by a combination of electrophoretic

methods, ion exchange on Dowex 50, and paper chromatography. Table 9 shows the composition or structure of the individual basic peptides isolated. The close resemblance between the structures found in the two proteins is quite evident. Although the isolation of both the arginine and the histidine peptides was carried out with particular care in this instance, the number of individual peptides eventually isolated is small when compared with the total number of residues of the two basic amino acids known to be present in the haemoglobins. Moreover, no evidence was found for the presence of histidine or arginine in the neutral or acid peptide fractions. Our results

Table 9

PEPTIDES OF ARGININE, HISTIDINE, AND
LYSINE ISOLATED FROM PARTIAL HYDRO-
LYSATES OF HORSE AND HOG HAEMO-
GLOBIN

horse	hog
ala.arg. leu.arg. phe.arg. ser, arg. tyr.arg. arg.leu arg.phe arg, lys lys, his, arg	ala.arg. phe.arg. ser.arg. tyr.arg. arg.leu arg.phe arg, lys
ser.his ala.his leu.his his.ala his.lys ser.(his, lys) his, val	ser.his ala.his leu.his his.ala his.lys ser.(his, lys) his.(lys, ala) his.(lys, ala, val) his, lys, gly
gly.lys ser.lys thr, lys his.lys lys.ala lys, val lys, leu ser.(his, lys) lys, ala, val lys, gly, thr	gly.lys thr.lys his.lys lys.ala lys.val lys, leu ser.(his, lys) lys.(ala, val) leu, lys, val lys, his, gly lys, ser, ala

thus indicate not only an extensive similarity in the structure of the two hemoglobins examined, but also either the repetition of certain peptide sequences involving the basic amino acids, or, more probably, the existence of several identical units in the molecules of these proteins. This last assumption is borne out by the work of Reichmann who found that under certain conditions the haemoglobin molecule may dissociate into four parts.¹²

In this review, I have attempted to demonstrate the principles on which we base our approach to the problem of protein microstructure and some of the first results of this approach.

REFERENCES

1. The results so far obtained in the field of protein microstructure are reviewed in a paper by F. ŠORM, B. KEIL *et al.*, *Chem. listy*, **51**, 1171 (1957); *Coll. Czechoslov. Chem. Commun.*, **22**, 1310 (1957).
2. H. TUPPY and G. BODO, *Monatsch.*, **85**, 807, 1024, 1182 (1954); H. TUPPY and S. PALÉUS, *Acta Chem. Scand.*, **9**, 353 (1955).
3. F. ŠORM, *Coll. Czechoslov. Chem. Commun.*, **19**, 1003 (1954).
4. V. KNESSLOVÁ, V. KOSTKA, B. KEIL and F. ŠORM, *Coll. Czechoslov. Chem. Commun.*, **20**, 1311 (1955).
5. B. KEIL, *Coll. Czechoslov. Chem. Commun.*, **19**, 1006 (1954).
6. B. KEIL and F. ŠORM, *Coll. Czechoslov. Chem. Commun.*, **19**, 1018 (1954).
7. B. KEIL and F. ŠORM, *Coll. Czechoslov. Chem. Commun.*; in the press.
8. J. VANĚČEK, B. KEIL and F. ŠORM, *Coll. Czechoslov. Chem. Commun.*, **20**, 363 (1955).
9. J. VANĚČEK, B. MELOUN and F. ŠORM, *Coll. Czechoslov. Chem. Commun.*; in the press.
10. P. MĀSIAR, B. KEIL and F. ŠORM, *Chem. listy*, **51**, 352 (1957); *Coll. Czechoslov. Chem Commun.*, **22**, 1203 (1957).
11. P. MĀSIAR, B. KEIL and F. ŠORM, *Chem. listy*; in the press; *Coll. Czechoslov. Chem. Commun.*; in the press.
12. M. E. REICHMANN and J. R. COLVIN, *Can. J. Chem.*, **34**, 411 (1956.)
13. B. MELOUN, J. VANĚČEK and F. ŠORM, *Coll. Czechoslov. Chem. Commun.*; in the press.

Discussion on the paper of Professor Šorm

B. KEIL

Allow me to supplement Professor Šorm's lecture by a few remarks concerning methods and techniques. In the work discussed by Professor Šorm we have been using a number of procedures worked out in our laboratories, in addition to the well-known and well-tried methods of separation and identification. For following the course of fractionation on Dowex columns or by multi-compartment electrophoresis we use paper electrophoretic analysis of samples in place of the usual ninhydrin colorimetry. This method gives rapid approximate information not only of the amount, but also of the nature of the material in the samples. The arrangement used for this purpose is very simple.³ A sheet of Whatman paper is loosely stretched vertically between two troughs containing pyridine acetate buffer of a low and defined conductivity. Voltages as high as 1500 V, corresponding to a potential gradient of about 40 V/cm., can be applied without cooling (Fig. 1).

The analytical separation is very uniform; the time of each run is 1.0 to 1.5 hours. Up to 200 mg. of a peptide mixture can be resolved on a single sheet of Whatman No. 3 paper. Sanger has used a four-compartment electrophoretic apparatus for the primary fractionation of acid hydrolysates of insulin. We have found a five-compartment apparatus most convenient, with formic acid, acetic acid and ammonia as the electrolytes.⁵ The omission of the sulphuric acid and the introduction of the fifth compartment removes the danger of loss or contamination of the material at the electrodes. The separation of partial hydrolysates into basic, acidic and neutral peptide fractions is clear-cut, as may be seen from the paper electrophoretic check of the individual fractions.

An extension of the membrane system of electrophoresis led to the development in our laboratories of a multi-compartment apparatus for the fractionation of peptides and proteins.⁴ In a set-up of one hundred compartments, with water-cooling and pyridine acetate buffer, and working at 5000 V, a progressive fractionation of peptide mixtures can be achieved practically without losses. The fractions are withdrawn from the individual compartments by pipetting, and directly lyophilized. A similar arrangement of twenty compartments with paper membranes has been used for the preparative fractionation of proteins and of mixtures of nucleic acids, polysaccharides and proteins (Fig. 2).

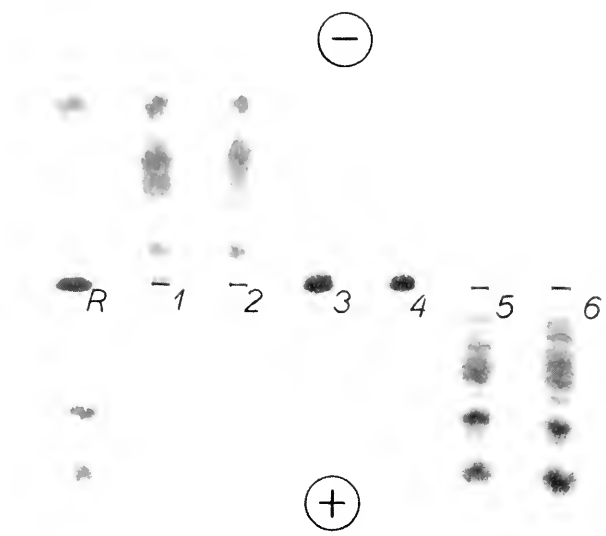


Fig. 1

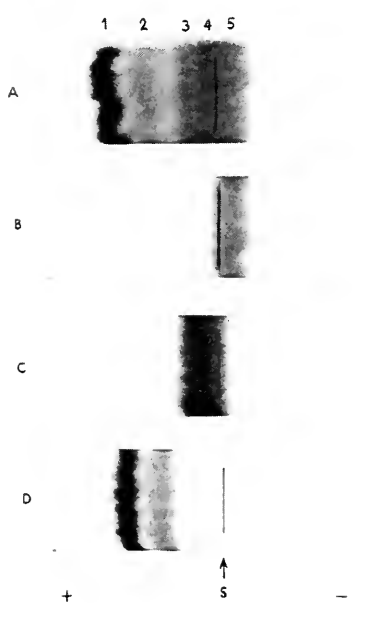
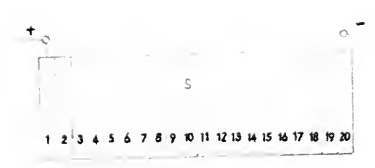


Fig. 2

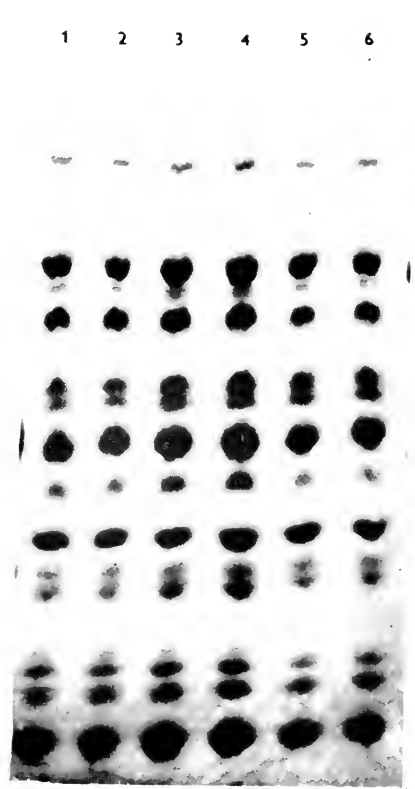


Fig. 3

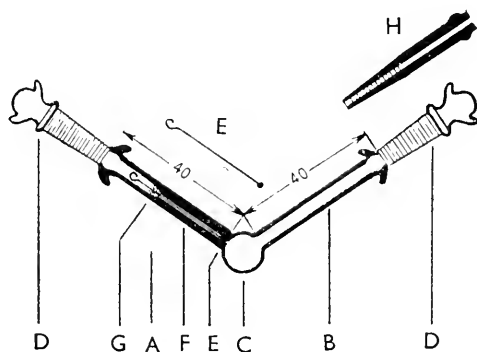
With compartments of 10 ml. capacity each, up to 1 g. of material may be fractionated.

There are two further techniques which I would like to mention: one for the quantitative determination of amino acids, and the second for carrying out dinitrophenylation analysis.

The most precise method for the quantitative analysis of amino acid mixtures today is, of course, the technique of Moore and Stein; however, for the routine quantitative determination of the amino acid composition of peptide fractions we have worked out four years ago a simpler procedure.¹ This required, first, an adaptation of the paper-chromatographic technique which would permit the separation of as many of the amino acids as possible in a single direction, so that a number of samples might be analysed in parallel (Fig. 3).

This problem was essentially solved by repeated development of the chromatogram, with drying in between each solvent run. After detection, the chromatogram is photographed on cinefilm and the blackening of the film is measured using a precision spectrographic microphotometer. A blank reading for any part of the paper is readily obtained from the photographic record. Since the photographs are taken by reflected light the inhomogeneity of the paper fibres is of no consequence. The same technique can be used for example for the direct quantitative determination of glutamic, aspartic, or cysteic acid, histidine, etc., after paper-electrophoretic separation, or, in general, for the estimation of any compound giving a coloured spot. In another laboratory of our Institute this technique is for instance being used for the determination of elementary sulphur. The areas under the curves obtained—which resemble those obtained by the Moore and Stein method—are converted graphically into equivalents of each particular compound.

Finally, a few words about the dinitrophenylation technique: With lower peptides on a micro scale, i.e. with amounts of about 5-30 μg , we use various modifications of the following procedure:²



The peptide is dinitrophenylated in the microflask and then selectively retained on a microcolumn of Zerolite H, consisting of about 20 μg . of the

ion exchanger. All other components of the mixture, including unreacted peptide, are eluted. The DNP-peptide is then desorbed back into the flask, evaporated to dryness and, after removal of the ion-exchanger, hydrolysed in situ. The DNP amino acid is adsorbed on a fresh batch of the ion exchanger, the free amino acids being eluted. The adsorbed material is eluted back into the flask, the dinitrophenol is removed by vacuum sublimation and the residual DNP amino acid is identified by paper chromatography. This procedure eliminates the extraction with organic solvents and the separation into water-soluble and ether-soluble fractions. The technique, which readily lends itself to work with whole series of samples, was used for most of the peptides mentioned in Professor Šorm's paper.

REFERENCES

1. B. KEIL, *Chem. Listy*, **48**, 725 (1954); *Coll. Czechoslov. Chem. Commun.* **19**, 1006 (1954).
2. B. KEIL, *Chem. Listy*, **51**, 1927 (1957); *Coll. Czechoslov. Chem. Commun.* **22**, in press.
3. O. MIKEŠ, *Chem. Listy*, **51**, 138 (1957); *Coll. Czechoslov. Chem. Commun.* **22**, 831 (1957).
4. O. MIKEŠ, J. VANĚČEK, B. MELOUN, V. KOSTKA and J. KÁRA, *Chem. Listy*, **51**, 1562 (1957); *Coll. Czechoslov. Chem. Commun.* **22**, in press.
5. J. VANĚČEK, B. MELOUN and F. ŠORM, *Chem. Listy*, **51**, 1367 (1957); *Coll. Czechoslov. Chem. Commun.* **22**, in press.

Zone electrophoresis and chromatography as methods for the purification and characterization of proteins

A. TISELIUS

Biochemical Institute, Uppsala University, Sweden

The Commission on Proteins was created at the IUPAC Congress in New York and Washington 1951 and had as its first task to investigate the possibility of finding a standard protein preparation which would meet reasonable requirements of purity and homogeneity. This is not the time to discuss the very valuable work of this commission. I think, however, that I am right in saying that the task proved much more difficult than could be foreseen. Still, I believe that if nothing else had come out of the work than a realization of the difficulty of defining a pure protein, this would have been well worth the effort spent.

Some may perhaps ask how far we should go in our demands as to the purity of proteins, and if there is not a danger that we might become protein 'purists', interested in purity for its own sake. Such an argument neglects, however, several facts of importance in protein research to-day. One is that the very character of the modern sequence analysis methods, developed by Sanger, assumes a high degree of homogeneity. It would be an almost hopeless task to fit the puzzle of peptide fragments together, if these residues come from more than one homogeneous protein. Secondly, the high resolution of modern methods of protein separation has in an increasing number of cases revealed a complexity or micro-heterogeneity which appears to be extremely interesting. I am thinking now not only of the various pathological varieties of hemoglobin which were first demonstrated by Pauling, Itano, Singer and Wells (1949), but also of a number of enzyme proteins which appear to consist of mixtures of closely similar but distinctly different components. We cannot yet judge the significance of these results—for example, how these differences are reflected in specificity or activity—but this is certainly an important field for further investigations. In some cases this variation appears to be genetically controlled. Besides the pathological hemoglobins this complexity has now also been established for normal

proteins (e.g. the sheep hemoglobins studied by Harris and Warren (1956; see also Evans *et al.*, 1956) and the haptoglobins in human serum studied by Smithies (1955), Smithies and Walker (1955) and by Smithies and Poulik (1956).

I shall confine myself here to a discussion of some recent developments of two methods, namely *electrophoresis* and *chromatography*, in their application to the separation of proteins and protein-like substances. Of these methods electrophoresis has long been known as a particularly efficient method of studying the homogeneity of proteins. The application of chromatography to such problems is more recent, but the last few years have given us several examples demonstrating that the specificity of this method, well known from its application to low-molecular weight substances, is also quite striking with large molecular weight material.

In *electrophoresis* the most important recent advance appears to me to be the development of zone methods and their application to separation problems on a macro and micro scale. Fig. 1 illustrates schematically the

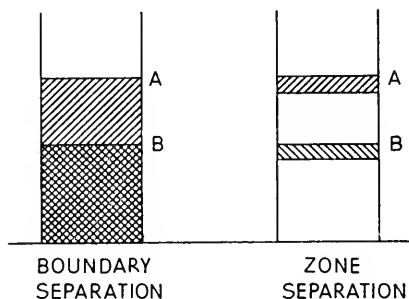
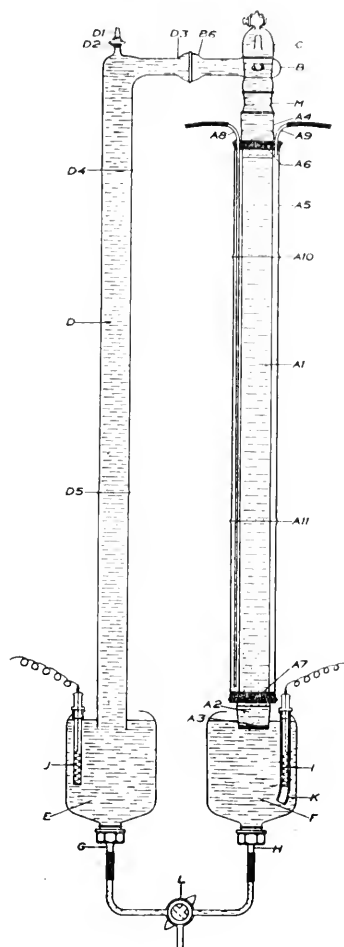


Fig. 1. Schematic representation of the principal difference between boundary and zone methods of separation.

difference in principle between boundary methods (as applied in the standard moving boundary procedure) and zone methods in the separation of two substances, A and B. It is seen that the zone method requires a supporting medium to stabilize the zones against gravity convections. Thus the experiments are performed in columns or troughs packed with cellulose, starch, glass powder or similar materials, or in filter paper strips (paper electrophoresis). In certain cases zone experiments have been performed in gels (especially when serological reagents are to be applied—'immuno-electrophoresis') or in density gradients. The apparatus required is rather simple. As an example, Fig. 2 shows a zone column electrophoresis apparatus constructed by Porath (1956, 1957 *a, b*) in Uppsala.

The column, which is surrounded by a cooling jacket, is packed with cellulose powder. Different types of columns may be used, depending upon the desired degree of resolution and the capacity required. Quantities of up to 50 g of protein have been separated in columns of this type. Fig. 3

Fig. 2. Zone column electrophoresis apparatus according to Porath (1956). The packed column is seen to the right, surrounded by a cooling mantle. Below are the electrode vessels with a levelling arrangement.



shows the zone electrophoresis laboratory in the Institute of Biochemistry, Uppsala, with several columns in operation. In this procedure the zones, after having been separated by electrophoretic migration, are eluted from the column and collected in a standard automatic fraction collector of the type commonly used in chromatographic work. This has the advantage that the columns may be used many times without repacking. In other types of apparatus (see especially Kunkel, 1954) flat horizontal troughs are used. This makes the surface of the supporting medium more easily accessible, so that staining reactions can be applied for the localization of the zones, which may then be isolated by cutting the trough contents accordingly. This, however, makes it necessary to pack the trough anew for each experiment. Experiments in gels often give very sharp separations, but elution is difficult. Paper strip electrophoresis has become extremely popular on account

of its simplicity and good resolution, and has found wide application as a micro-method. To a certain extent its relation to other types of zone electrophoresis is similar to that of paper chromatography to column chromatography, but the heat generated by the current may give rise to evaporation disturbances which make the method less suitable for quantitative work.

Recent experiments in Uppsala with very small columns show that quantities as small as those commonly used in paper strip electrophoresis may also be studied in columns under more rigidly controlled conditions than in the paper strip method.

In the discussion which is to follow Dr Porath will give some examples of the application of the zone column technique. I shall limit myself to a few comments on the advantages and limitations of the zone methods in general.

The advantages of the zone methods over the classical boundary procedure are obvious, especially that in the former it is possible to achieve complete separation of the migrating components and to isolate them for further investigation. In the boundary method the zones overlap partially and conclusions are drawn only from optical observation of their migration. A further study of the properties and difference in the characteristics of components is difficult or (with closely similar substances) impossible. In addition the zone methods are much easier to apply to low-molecular weight compounds, like amino acids and peptides. Boundary electrophoresis of such substances has proved very difficult because of their considerable contribution to the conductivity of the system, which results in boundary anomalies. In zone electrophoresis these anomalies are apparently much less pronounced, probably because of the easy interchange of ions by diffusion between the front and the rear of a zone and by a gradual automatic adjustment of the concentration to lower values. A further advantage, which has been demonstrated by Porath (1957 *b*), is that the reduction in diffusion in powdered supporting medium is much greater than the reduction in current; the heat generated by the current being carried away by the stabilizer almost as efficiently as in free solution. This contributes very much to the separation efficiency, particularly with low-molecular weight substances.

One has to pay a price, however, for all these advantages. Most materials suitable for use as supporting media show some interaction with proteins and other substances. There is of course a mechanical hindrance but, above all, specific interaction by adsorption is very common, leading to a spreading of the zones by tailing and influencing their mobilities in a way which is difficult to predict. This is the key-problem of the zone method and much work has already been done to find more inert supporting materials which would show as small interaction as possible with the migrating substances. Flodin and Kupke (1956) have prepared a partially esterified cellulose by ethanolysis, which has proved much superior to ordinary cellulose for zone electrophoresis columns and is now generally used in the Uppsala laboratory for protein and peptide separations. No doubt other possibilities will

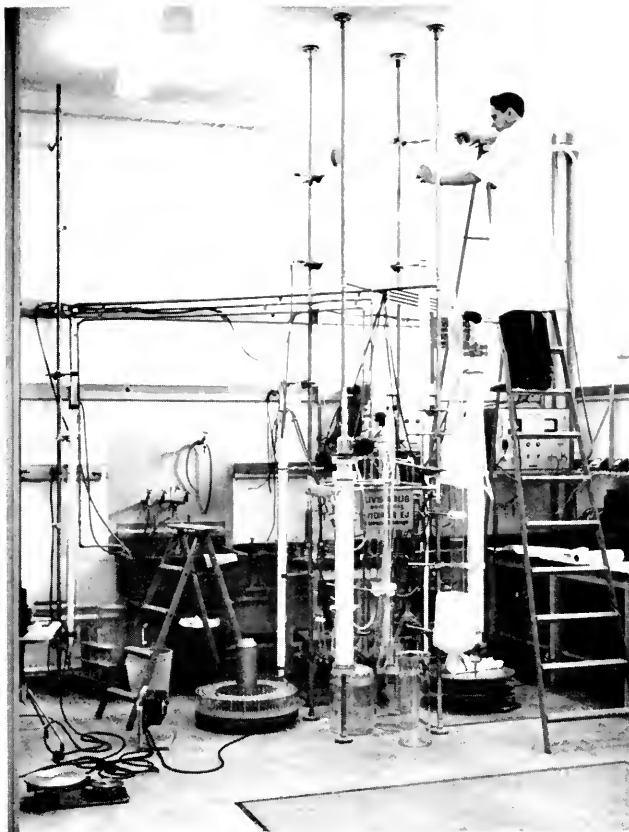


Fig. 3. Zone electrophoresis laboratory at the Institute of Biochemistry, Uppsala.

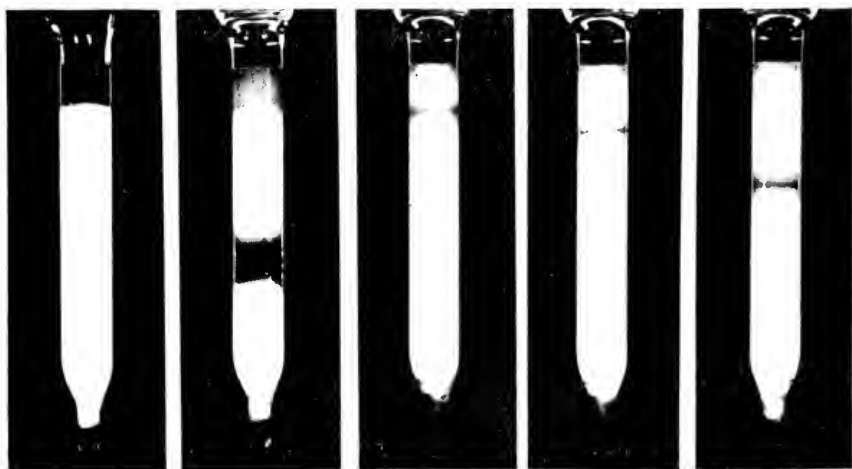


Fig. 4. Chromatographic separation of R-phycoerythrin and R-phycoecyanin in an extract of *Ceramium rubrum*. Calcium phosphate (Brushite) column, saturated with 0.005M phosphate buffer, pH=6.8. Elution of phycoerythrin (same buffer, 0.02M, exposures 1-2) and of phycoecyanin (same buffer, 0.04M, exposures 3-5). (From A. Tiselius, S. Hjertén and Ö. Levin, *ib.* (1956).)

be explored. It is of course desirable that the water content of the column is as high as possible. In this respect cellulose has advantages: a column packed by sedimentation contains approximately 85% water, as compared with 40% for glass powder and 50% in starch, of which only 33% is free (Porath, 1957 *b*).

The effects mentioned represent a difficulty not only by interfering with the separation but also as regards the evaluation of the results in terms of mobility. It is true, of course, that most zone electrophoresis experiments are made with the separation as chief purpose, and that for accurate mobility determinations the free moving boundary electrophoresis method is preferred. The possibility of quantitative control of the separation has, however, been an attractive feature of the electrophoretic methods, and it would be a pity if this would have to be sacrificed in order to gain the advantages offered by the zone methods. If there is no interaction between the migrating substances and the supporting material except the purely mechanical hindrance (or, strictly speaking, if such interaction and mechanical hindrance influence the mobilities of all substances present, including the buffer ions, to the same extent) it ought to be possible to determine mobilities by measuring the volume distances between the peaks in the elution diagrams obtained from zone electrophoresis columns, and comparing them with the distance of migration of a substance of known mobility. Attempts in this direction have been made, but I believe it is very important to pursue such investigations with the more inert supporting materials now available.

When turning now to *protein chromatography* I wish to say first of all that this field is much more recent and considerably more difficult to discuss. It appears, however, as if the application of chromatography to proteins of large molecular weight has now left the 'lag phase' (to use an expression from bacteriology). Whether it has definitely entered a 'log phase' remains to be seen, but there are already a number of beautiful results which indicate that the method offers great possibilities. It is equally clear, however, that the underlying principles are not yet well understood and that the methods involve certain risks of making grave mistakes.

There is an excellent recent review of this whole field by Moore and Stein (1956) and for this reason I shall limit myself to some brief remarks. The particular difficulties offered by proteins in chromatography depend on their large molecular size, which may prevent them from entering into the pores of the column material or make their reactions with this material too slow. Also, the instability of proteins and particularly the risk of denaturation at surfaces is an obstacle. Most of the common types of chromatography as applied to aqueous solutions have been used for protein separations, i.e. ionic exchange, partition and adsorption. In each case, however, one has found that only certain systems have the properties required. Both cationic and anionic exchange resins have been used, as well as ionic exchangers prepared from cellulose. Partition systems have been worked out, consisting

of glycols, water and salts, in which insulin and recently also γ -globulins have been fractionated. Among adsorbents calcium phosphate seems to offer special advantages, and I will show some pictures illustrating this method (Fig. 4; for details see figure text).

This experiment illustrates a behaviour which is often observed in protein chromatography, and which is markedly different from what is usually experienced in elution chromatography of low-molecular weight substances. It is very difficult or impossible to find a buffer solution which will elute both the phycoerythrin and the phycocyanin as reasonably well-defined zones, migrating with different R_f -values. It appears that the R_f of many proteins is highly dependent on the concentration of the eluting buffer, so that its value will shift from almost zero to almost 1.0 in a comparatively narrow interval of buffer concentration. This is to be expected, according to the law of mass action, if the attachment of one protein molecule to the column will lead to a liberation of several small ions. The consequence is, however, that elution in the ordinary way is not suitable, but rather a step-wise or gradient elution must be applied. The step-wise elution is often very efficient, but involves the risk that a slowly eluted component may appear in more than one step. In the interpretation of such experiments it is essential to rechromatograph the fractions obtained to ascertain if their different chromatographic properties can be reproduced in separate experiments. I want to stress this particularly, as the results so far often indicate a complexity of proteins which were earlier believed to be reasonably homogeneous. The pronounced dependence of the partition coefficient of proteins on the composition of the two-phase system is, of course, an analogous phenomenon.

There is no doubt that chromatography of proteins already has shown its usefulness both in preparative and analytical work, and that the method appears to possess a high degree of specificity. I shall quote only a few examples:

Synthetic polycarboxylic acid resins: cytochrome-c ribonuclease, lysozyme, chymotrypsinogen, chymotrypsin, hemoglobins, histones.

Cellulose ionic exchangers: serum proteins.

Synthetic anion exchange resins: serum proteins, phosphatases.

Partition chromatography: ribonuclease, insulin, γ -globulins.

Adsorption chromatography: various proteins and enzymes.

References are found in the above-mentioned review by Moore and Stein (1956).

With the 'all or none' elution behaviour of many proteins discussed above one may of course ask, whether column operation has any particular advantage or if one could not just as well apply a batch operation. This is no doubt practical in many cases. However, column operation has certain practical advantages, particularly with systems of unknown behaviour. It should be noted also that even if the length of the column does not contribute

essentially to the separation and if each component can be eluted at a buffer concentration giving $R_f=1$, mutual displacement effects may contribute very much to the separation of closely similar substances, and in this case a comparatively long column is of value. Such displacement effects are very common and often very pronounced among proteins. Zones may often be found which during migration through the column show marked separation without splitting up into two distinct peaks, the slower material being carried along with the faster by the reduction in affinity to the column caused by the displacement effect. It appears likely that a closer study of such displacement effects may lead to a better understanding of the phenomena underlying protein chromatography and perhaps also to more efficient and rational modes of operation.

In the discussion to follow, Dr H. Boman will show some results which illustrate the importance of displacement effects in the chromatography of some proteins.

REFERENCES

- EVANS, J. V., KING, J. W. B., COHEN, B. L., HARRIS, H. and WARREN, F. L. (1956), *Nature*, **178**, 849.
- FLODIN, P. and KUPKE, D. W. (1956), *Biochim. et Biophys. Acta*, **21**, 368.
- HARRIS, H. and WARREN, F. L. (1956) *Biochem. J.*, **60**, XXIX.
- KUNKEL, H. (1954), in Glick, *Methods of Biochemical Analysis*, **1**, 163 (Interscience Publ., New York).
- MOORE, S. and STEIN, W. H. (1956) in *Advances in Protein Chemistry*, **11**, 191 (Academic Press, New York).
- PAULING, L., ITANO, H. A., SINGER, S. J. and WELLS, I. C. (1949), *Science*, **110**, 543.
- PORATH, J. (1956), *Biochim. et Biophys. Acta*, **22**, 151.
- PORATH, J. (1957 a), *Ark. Kemi*, **11**, No. 18.
- PORATH, J. (1957 b), *Dissertation*, Uppsala, 1957.
- SMITHIES, O. (1955), *Biochem. J.*, **61**, 629.
- SMITHIES, O. and WALKER, N. F. (1955), *Nature*, **176**, 1265.
- SMITHIES, O. and POULIK, M. D. (1956), *Nature*, **117**, 1033.
- TISELIUS, A., HJERTÉN, S. and LEVIN, Ö. (1956), *Biochim. Biophys. Acta*, **65**, 132.

Mutual displacement in protein chromatography

HANS G. BOMAN

Institute of Biochemistry, University of Uppsala

Similar substances have sometimes a marked tendency to compete for the same sites on an adsorbent. This mechanism was in 1943 used for chromatographic separations by Tiselius¹ who showed that when a mixture of sucrose and maltose were eluted with 0.5% phenol, the sugars came out from the column as narrow zones, bordering each other. This type of chromatography was named displacement development.

Ten years later, Polis and Shmukler² were the first to describe and utilize protein-protein displacement in their purification of coloured milk proteins with calcium phosphate columns. Unfortunately, they did not give any figures illustrating their displacement experiments and nobody else seems to have made use of their interesting findings.

Less than two years ago, when working out the technique for protein chromatography on the anion exchange resin Dowex 2,³ we found that when a mixture of horse-radish enzymes was eluted with a concentrated buffer, peroxidase and acid phosphatase were separated into two narrow bands as shown in Fig. 1, upper part. The same separation has later been obtained also on a cellulose anion exchanger, Fig. 1, lower part.⁴

In ideal displacement, the length of the zone should be proportional to the amount of material, and the concentration should depend on the affinity constant of the substance and the concentration of the displacing agent.¹ When the loading of the columns were varied, we obtained variations in peak width indicating that protein-protein displacement was a major factor contributing to the separation of peroxidase and the first phosphatase. This was different with the second phosphatase which obviously was eluted by the chloride concentration.

However, formally, these experiments (Fig. 1) could be characterized as elution with a very steep gradient of chloride and, in fact, the referee of our second paper⁴ did never believe that any protein-protein displacement was demonstrated by our experiments.

It was therefore thought to be of interest to show that a moderately adsorbed enzyme, prostatic acid phosphatase, could be desorbed by a more strongly adsorbed protein. Dowex 2 was chosen as adsorbent only because of the fact that we had elution data for a variety of proteins just on this resin.³ The dimensions of the column were 19×1.2 cm. The buffer con-

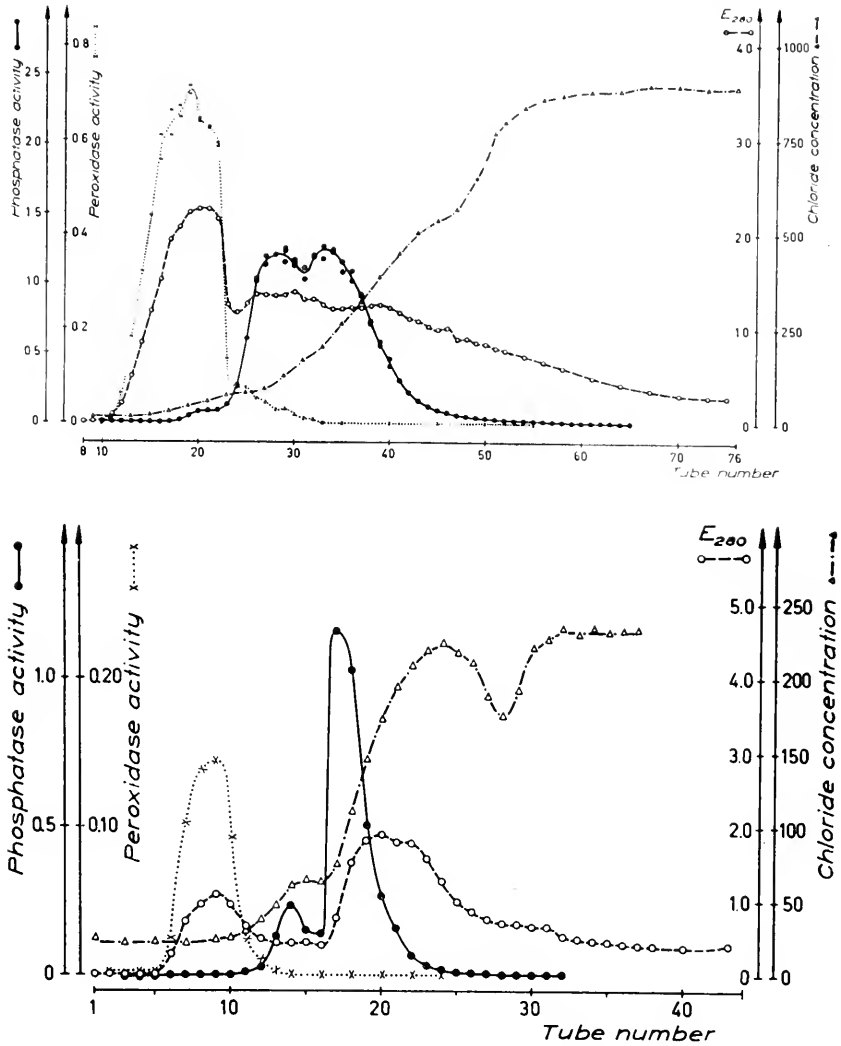


Fig. 1. (a) Upper part: One-step elution of 100 mg. horse-radish peroxidase and acid phosphatase on a 180 ml. column of Dowex 2. Fraction volumes about 0.65 ml. For details see ref. 3. (b) Lower part: One-step elution of 35 mg. horse-radish peroxidase and acid phosphatase on a 17 ml. column of TEAE cellulose. Fraction volumes 0.5–0.6 ml. For details see ref. 4.

centration has been kept constant through each experiment and the developer used was a α_2 -globulin fraction from human serum (kindly supplied by AB KABI, Stockholm). The results from three experiments carried out at different buffer concentrations are given in Fig. 2. At the start, the column was in equilibrium with a THAM (2-amino-2-hydroxymethylpropane-1 : 3 diol)-HCl buffer of pH 7.3. In each experiment, the same amount of freeze-

dried prostatic extract (1 mg.) was dissolved in the corresponding buffer and applied to the column. Elution was then directly started with a 0.1% solution of α_2 -globulin, also dissolved in the corresponding buffer. Phosphatase activity was determined and expressed as described earlier.³

The upper part of Fig. 2 shows that with 0.02M buffer, only small amounts

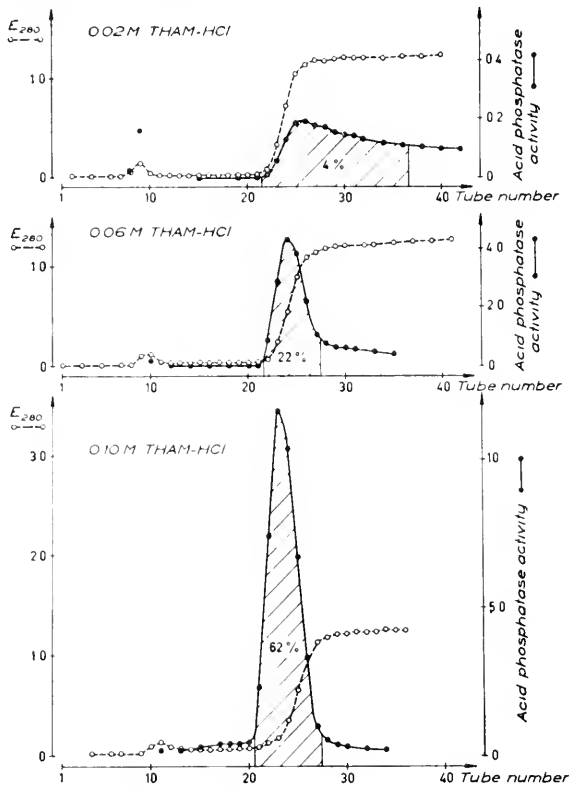


Fig. 2. Displacement analysis of prostatic acid phosphatase at three different concentrations of THAM-HCl buffers of pH 7.3. In each experiment, 1 mg. of prostatic extract was applied to a column with 20 ml. of Dowex 2 and developed with a 0.1% solution of α_2 -globulin from human serum. Fraction volumes around 0.8 ml.

of phosphatase (recovery 4%) are eluted as a very broad zone and that the front of the activity coincides with that of the developer. As a whole, this experiment looks like an unsuccessful frontal analysis. However, when the adsorption of the phosphatase is decreased by increasing the buffer concentration to 0.06M, 22% of the activity appears as a localized zone, just at the beginning of the front of the developing α_2 -globulin (Fig. 2, middle part). With a further decrease in the adsorption of the enzyme, when using a 0.10M THAM-HCl buffer, 62% of the activity is recovered in a sharp zone which has moved forward with regard to the front of the developer (Fig. 2, lower part). The small peak around tube 10 is non-adsorbed material.

These experiments demonstrate that protein-protein displacement may be a practically useful phenomenon and that the optimal buffer concentration is an important factor. Sober *et al.*⁵ describe how serum, during adsorption on cellulose anion exchanger, separates into a number of 'distinct coloured bands', which, however, are spread out during the gradient elution. Similar observations have been made in Uppsala, especially on calcium phosphate (Hjertén, to be published). The most likely explanation is that a good separation is obtained by mutual protein displacement during the adsorption step, which then is spoiled by the salt gradient which passes over the optimal condition. It may therefore be possible, that in the case of marked displacement effects, a better resolution of a chromatogram is obtained with a protein solution as developer [cf. (2)], than that resulting from changes in buffer composition.

It has been suggested that it is easier to find the optimal conditions for the separation of a protein mixture with zone electrophoresis than with chromatography. Personally I do not want to make any general statement on this question, but I feel that the one-step elution on an anion exchanger is a rapid and convenient way to obtain valuable information about a complicated enzyme mixture.^{3,4,6}

REFERENCES

1. A. TISELIUS, *Arkiv Kemi*, **16A**, No. 18, 1 (1943).
2. B. D. POLIS and H. W. SHMUKLER, *J. Biol. Chem.*, **201**, 475 (1953).
3. H. G. BOMAN and L. E. WESTLUND, *Arch. Biochem. Biophys.*, **64**, 217 (1956).
4. H. G. BOMAN and L. E. WESTLUND, *Ibid.*, **70**, 572 (1957).
5. H. A. SOBER, F. J. GUTTER, M. M. WYCKOFF and E. A. PETERSON, *J. Am. Chem. Soc.*, **78**, 756 (1956).
6. H. G. BOMAN and U. KALETTA, *Biochim. Biophys. Acta*, **24**, 619 (1957).

The characterization of lower molecular weight proteins by dialysis

L. C. CRAIG, WM. KONIGSBERG,*
A. STRACHER† AND T. P. KING

The Laboratories of the Rockefeller Institute for Medical Research, New York, N.Y.

During the course of the fractionation studies required to isolate a new chemical individual from a naturally occurring or synthetic mixture much is usually learned about the nature of the substance. Thus distillation or the ultracentrifuge will set certain limits concerning its size, electrophoresis gives information about its net charge at specific pH levels, countercurrent distribution can tell something of its polarity, etc. In considering the various fractionation techniques now available for isolating relatively large molecules it occurred to us that a technique which would separate substances on the basis of their size or shape would not only be a very useful technique supplementing those now used, but could also give very informative data about size and shape. With this idea in mind we have turned to the study of membrane diffusion.

The viewpoint taken in this work differs from that of the numerous studies on membranes already in the literature in a number of ways. Perhaps most important—much less emphasis has been placed on pore size. Membranes are compared on the basis of the rate they permit standard solutes of known size to pass by diffusion as well as on the basis of total exclusion. An attempt is made to minimize charge effects. The separate units are designed with fractionation in mind and with the intention of making them part of a countercurrent process.

Dialysis is one of the earlier separation methods to have been used, and is familiar to everyone for separating dialysable from nondialysable solutes. However, few studies attempting to refine the technique beyond this point are to be found in the literature. Probably the most promising were those of Signer¹ and co-workers, who experimented with a continuous countercurrent dialysis train of unique design.

* National Science Foundation Fellow.

† Fellow of the National Foundation for Infantile Paralysis.

As regards the measurement of size, the approach we have followed is similar to that in the well-known method of molecular weight determination of Northrop and Anson,² who measured the rate of diffusion of a solute through sintered glass. However, as will be seen later, there are marked differences.

It was realized in the beginning that the rate of diffusion through a membrane would be a function of a number of interdependent variables, when the sole driving force is a concentration gradient across the membrane. These are the nature of the membrane, charge effects on the membrane, effect of solvent, pH, ideality of the solute, temperature, etc.

In preliminary studies cellophane appeared to be the most promising membrane for the investigation. A simple cell was constructed from a cellophane sac in which both the contents of the sac and the solvent around it were stirred. By measuring the rates of escape of solutes of known size through the sac, it was soon found³ that the escape rates did not bear the relationship to size that would be expected from free diffusion alone and as would be expected from the Northrop and Anson cell. A greater selectivity was noted which suggested a membrane effect in addition to the differential provided by diffusion.

Fig. 1. Schematic drawing of the microdialysis cell. A 7-mm. glass tube fitted at the top with a 2-ml. syringe passes through two larger sections of glass tubing A and B. The dialysis membrane represented by the dashed line is made by tying a knot in a wet piece of cellophane tubing with care to avoid any stretching of the membrane. A is of such an outside diameter (ca. 15 mm.) that it is a little less than the diameter of the extended sac. B is slightly larger in diameter than A and serves to hold the membrane in position. C is a test tube which is large enough to provide about 2 mm. clearance from the membrane. It holds the outside solvent amounting to about 10-20 times the volume of the solution inside the sac. The upper rim of C is collapsed somewhat in order to keep it centered. The lower end of C rests on the table and can be easily centered. During operation the height of the solutions both inside the sac and outside are adjusted to reach only the top of A. A small hole is punched through the membrane just below B in order to avoid pressure being built up during the run.



In order to enhance this effect it appeared desirable to try less porous membranes so that the solutes would be even more restricted in their passage, or alternatively, use larger test solutes with the same membrane. Either course implied a much slower process, although the time involved had already proved a serious handicap. Therefore, an attempt was made to design a cell which would give a faster rate of net transfer across the membrane.

It is known⁴ that the net rate of transfer under a concentration gradient

106 L. C. CRAIG, W. KONIGSBERG, A. STRACHER, T. P. KING
of solute through a membrane is directly proportional to the cross sectional area of the membrane. It is also proportional to the concentration gradient and, therefore, for a given weight of solute, inversely proportional to the volume inside the sac. The most rapid net rate of transfer of solute, therefore, would be given when the solution to be dialysed is spread evenly as a thin film over the entire inside surface of the dialysis sac. The concentration outside should be kept as low as possible. This can be accomplished very easily by the simple apparatus⁵ shown schematically in Fig. 1.

In this cell all the solution inside is held in a thin film of the order of 0.1-0.2 mm. in thickness. No stirring is required. The rate of escape of a solute is measured by periodic removal of the solution outside for analysis and replacement with fresh solvent. At these times the film inside can be redistributed by moving the syringe plunger at the top of the cell. Actual experiment showed that with this arrangement the rate of dialysis for a given membrane was of the order of 10-100 fold that afforded by the usual dialysis procedures.

If the experimental arrangement is satisfactory and the solute behaves ideally its rate of escape should be proportional only to the concentration gradient. The rate should thus follow first order reaction kinetics, and when the logarithm of the decrease in concentration is plotted against time, a straight line should be obtained.

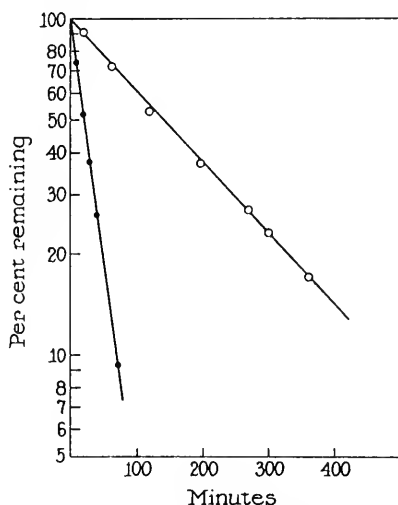


Fig. 2. Escape rate curves of bacitracin and subtilin.

This has now been accomplished with many different solutes. Fig. 2 is an example with two different peptides—bacitracin (M.W. 1,422) and subtilin (M.W. 3,300) using 18/32 Visking cellophane. If a mixture of equal parts of the two should be taken the dialysis could be interrupted after 50 minutes when 80% of the bacitracin would be out, but only 20% of

the subtilin. A series of such cells might be employed in countercurrent manner as in countercurrent distribution. At each stage the dialysate could be evaporated and put inside the sac of the next member of the series. Thus from each cell 80% of the bacitracin would be moved forward at each stage, but only 20% of the subtilin. The familiar mathematics of countercurrent distribution (the binomial theorem) is directly applicable and will tell us that the bacitracin-subtilin mixture is an easy separation requiring only a few stages. Useful separations with peptides and derivatives have been accomplished this way.^{3,6}

The straightness of the line with a supposedly pure substance affords a measure of adherence to ideality in the solvent used and of the homogeneity with respect to size and shape. In the countercurrent process a straight line represents ideality and is desirable just as a linear partition isotherm represents ideality in countercurrent distribution. An unfavorable solvent can produce molecules of different sizes from an otherwise pure solute by promoting association or theoretically by causing a fraction of the molecules to assume a shape different from the others but perhaps not instantaneously interconvertible.

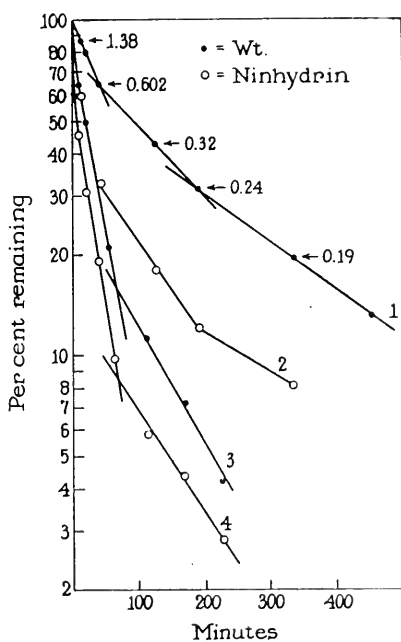


Fig. 3. Escape rate curves obtained with a sample of protamine.

Where a mixture of sizes is present the deviation from a straight line may take the form of distinct breaks in the curve as shown in Fig. 3 which was obtained with a 20 mg. sample of a crude protamine. Curve 1 was obtained by weight analysis of the dialysate fractions, but Curve 2 was obtained by

108 L. C. CRAIG, W. KONIGSBERG, A. STRACHER, T. P. KING
 the Moore-Stein ninhydrin procedure.⁷ The figures at each point on the weight curve give the ninhydrin values per mg. and definitely indicate a range of sizes.

The first 35% of the solute to pass was recovered and put in a second cell. Now weight analysis gave Curve 3 and ninhydrin gave Curve 4. Here a mixture of sizes is definitely being fractionated.

A rough survey of the solvent question and other variables intimately involved with it such as pH, salt content, temperature, etc., seemed to show that with the polypeptides and smaller proteins 0.01N acetic acid at 25° seemed to offer an environment most favorable to ideality. A survey was then made of a number of solutes of increasing molecular weight in a cell with Visking 20/32 cellophane. These data⁵ are given in Table 1, where the escape rate is expressed as a 50% escape time. For comparison a few data are given in 0.1N acetic acid and in 0.01N acetic acid in 18/32 Visking, a less porous casing. The molecular weights listed are those from the literature.

Table 1 does not list all the solutes studied. A number were not included because they showed too great a deviation from ideality. There is not time

Table 1

50% ESCAPE TIMES OF VARIOUS POLYPEPTIDES AND PROTEINS

Solute	M. W.	50% Escape Time		
		0.1 N HAc 20/32	0.01 N HAc 20/32	0.01 N HAc 18/32
Tryptophan	204	—	4 Min.	6 Min.
Bacitracin	1,422	—	15 Min.	21 Min.
Salmiridin*	2,000-8,000	—	40 Min.	150 Min.
Subtilin	3,300	—	54 Min.	138 Min.
B Chain from Insulin	3,600	—	42 Min.	Will not pass
Glucagon	4,000	—	50 Min.	↓
Insulin	5,733	1.5 Hrs.	—	
Cytochrome C	12,000	—	60 Min.	
Ribonuclease	13,600	3.5 Hrs.	2 Hrs.	
Lysozyme	14,000	3.5 Hrs.	2.3 Hrs.	
Trypsin	20,000	—	4 Hrs.	
Trypsinogen	20,000	—	6.5 Hrs.	
Chymotrypsin	24,500	—	5 Hrs.	
Chymotrypsinogen	25,000	12 Hrs.	9 Hrs.	
Pit. Lact. Hormone	26,000	—	13 Hrs.	
Gliadin	27,000	—	29 Hrs.	
Ovomucoid	28,000	—	35 Hrs.	
Pepsin	35,500	—	80 Hrs.	
Ovalbumin	45,000	—	More slowly than Pepsin	

* A protamine from the rainbow trout obtained through the courtesy of the Nordisk Insulinlaboratorium, Denmark.

here to discuss the individual behavior of each solute, but a few generalizations seem to be appropriate. Firstly, the increase in molecular weight (M.W.) consistently follows the same sequence as the decrease in escape rate, but not in direct proportion. A relative increase in M.W. has a greater effect as the maximum range of size capable of passing is approached. Thus the rate ratio for bacitracin/subtilin in 20/32 casing is a little more than three-fold, but in 18/32 casing it is more than six-fold. The selectivity, therefore, becomes greater as the limiting size is approached. Since the list of proteins in Table 1 includes acidic, neutral, and strongly basic ones with no obvious effect on the relationship of escape rate to M.W., it would appear that charge effects are relatively unimportant in the determination of escape rate.

The results with two pairs of proteins are particularly striking. These are trypsin-trypsinogen and chymotrypsin-chymotrypsinogen. Although each enzyme is so closely related in size to its zymogen that the accepted methods of molecular weight do not differentiate them, a striking difference in escape rate is noted here. It would seem that this must reflect a difference in shape. Indeed, such a theory is in line with the published results of Neurath and co-workers.^{8,9}

If the escape rate of proteins in this type of cell is really governed almost entirely by its overall outside dimensions, then the shape of the molecule should also have a strong influence. The proteins in Table 1 are all known to be quite globular in shape except under conditions which denature them. We have accordingly investigated the effect of denaturing conditions on dialysis rate. These include the effect of temperature, pH, urea, trichloroacetic acid, oxidation with performic acid, etc. In all cases where these conditions are known to cause denaturation, reversible or otherwise, a strikingly slower dialysis rate has been observed.

Table 2

EFFECT OF TEMPERATURE (IN DISTILLED WATER)

Substance	50% Escape Time at			
	4°	25°	41°	65°
Bacitracin A	50 Min.	15 Min.	6 Min.	6 Min.
Ribonuclease	7.0 Hrs.	2.0 Hrs.	1.2 Hrs.	2.5 Hrs.
Lysozyme	—	2.2 Hrs.	1.1 Hrs.	48 Min.
Chymotrypsin	—	8.0 Hrs.	—	>60 Hrs.

The effect of temperature is shown in Table 2. Ribonuclease is known¹⁰ to undergo reversible unfolding at a temperature of 65°. This is reflected here by a slower escape rate at 65° than at 41°. Lysozyme, on the other hand, is known to be very stable and shows a faster escape rate at 65° than at 41°. One might infer from these data that a mixture of the two proteins could

110 L. C. CRAIG, W. KONIGSBERG, A. STRACHER, T. P. KING
 be separated by dialysis at 65° even though their molecular weights are very similar.

Table 3
 EFFECT OF pH ON 50% ESCAPE TIME

Substance	HCl 0.03 N pH 1.5	HCl 0.003 N pH 2.6	Acetic 0.1 N pH 2.9	Acetic 0.01 N pH 3.4	Distilled water pH 5.6	α -Picoline 0.01 N pH 7.2	Ammonia 0.01 N pH 9.2
Bacitracin A	48 Min.	35 Min.	15 Min.	15 Min.	15 Min.	77 Min.	—
Insulin	—	2.4 Hrs.	1.5 Hrs.	—	—	—	—
Ribonuclease	—	3.5 Hrs.	3.2 Hrs.	2.2 Hrs.	2.0 Hrs.	2.8 Hrs.	Very slow
Oxidized Ribonuclease	—	—	—	5.2 Hrs.	3.8 Hrs.	Very slow	—
Lysozyme	7.2 Hrs.	2.8 Hrs.	3.2 Hrs.	2.3 Hrs.	2.3 Hrs.	1.8 Hrs.	4.7 Hrs.
Trypsin	—	—	—	4.0 Hrs.	3.6 Hrs.	3.6 Hrs.	—
Chymotryp- sin	—	—	—	5.0 Hrs.	8.0 Hrs.	3.8 Hrs.	—
Chymotryp- sinogen	—	—	—	9.0 Hrs.	4.9 Hrs.	16.3 Hrs.	—

Table 3 shows the effect of pH at 25°. The results with bacitracin were obtained with 18/32 Visking, but 20/32 Visking was used for all the others. Bacitracin is known to be unstable below a pH of 2 and above a pH of 7. Ribonuclease shows the behavior to be expected and is quite different from oxidized ribonuclease. Lysozyme shows a markedly slowed rate of dialysis only at the pH extremes. Autodigestion probably plays a role with trypsin and chymotrypsin at the higher pH.

Table 4
 DIALYSIS IN UREA SOLUTION

Substance	50% Escape Time in				
	H ₂ O	1 M	3 M	6 M	9 M
Bacitracin A	15 Min.	19 Min.	19 Min.	20 Min.	33 Min.
Insulin*	1.5 Hrs.	2.5 Hrs.	2.5 Hrs.	5.0 Hrs.	5.0 Hrs.
Ribonuclease	2.0 Hrs.	2.0 Hrs.	2.7 Hrs.	3.7 Hrs.	8.0 Hrs.
Ox. Ribonuclease	3.8 Hrs.	3.8 Hrs.	6.8 Hrs.	—	8.6 Hrs.
Lysozyme	2.3 Hrs.	1.9 Hrs.	2.1 Hrs.	2.8 Hrs.	3.4 Hrs.
Cyt. C.	—	1.5 Hrs.	1.3 Hrs.	1.9 Hrs.	4.2 Hrs.
Trypsin	3.6 Hrs.	3.5 Hrs.	3.5 Hrs.	7.6 Hrs.	20.0 Hrs.
Chymotrypsin	8.0 Hrs.	6.2 Hrs.	6.2 Hrs.	6.2 Hrs.	45.0 Hrs.
Chymotrypsinogen	4.9 Hrs.	5.5 Hrs.	6.4 Hrs.	10.5 Hrs.	68.0 Hrs.

* Done in 0.1N acetic acid.

The effect of urea at 25° is shown in Table 4. Again there is good correlation with the known effect of various concentrations of urea on these proteins. The tendency to unfold could be followed by viscosity measurements or by measuring specific optical rotations. Thus far, only the latter has been studied. A solution giving an increase in optical activity has also shown a decrease in escape rate, both in the urea studies and in the pH studies.

Proteins are known to associate in a favorable environment to give dimers, trimers, etc. Indeed, with certain representatives this tendency has been so strong that the smallest dissociable unit has been recognized only after prolonged study. Insulin has been an example of this type of substance,¹¹ but was finally recognized as having a dissociable unit as small as 6,000. This value fits very well into the series in Table 1. The escape curve in 0.1N acetic acid and 20/32 Visking for material purified by countercurrent distribution (CCD) is shown in Fig. 4. The escape curve for a sample of amorphous commercial insulin is also shown. Perhaps, in this latter a certain

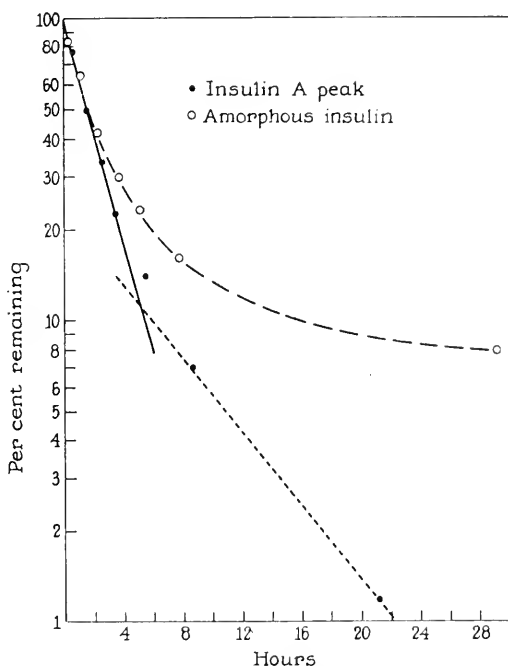


Fig. 4. Escape rate curves with insulin preparations.

percentage of the molecules have already begun to associate preliminary to the formation of insoluble fibrils.¹² Irrespective of the truth of this possibility, when insulin is brought into solution at a concentration of 1% with a minimal amount of ammonia the pH is about 8.0, and when this solution is placed inside the sac of the cell with water outside, the protein will not

112 L. C. CRAIG, W. KONIGSBERG, A. STRACHER, T. P. KING
diffuse through 20/32 Visking. Insulin is well known to aggregate at the higher pH values.

The A chain of insulin prepared by performic acid oxidation and purified by CCD did not give a straight line in the dialysis study at any pH, and scarcely passes the membrane above pH 7. It is apparently completely in the extended form with a high optical activity. On the other hand, the B chain gave a fairly ideal curve much like the purified insulin of Fig. 4 except that it diffused through the membrane at a faster rate.

Very early in the study with proteins, it was discovered that salt in the solution had a retarding effect on the rate of dialysis. As an example, escape curves at 25° and 20/32 Visking with ribonuclease in solutions of different sodium chloride concentrations are shown in Fig. 5. In each case a break in the curve was noted, but at a different point on the curve for each salt concentration. Studies at 40° showed a faster overall escape rate but similar breaks in the curve and were otherwise comparable.

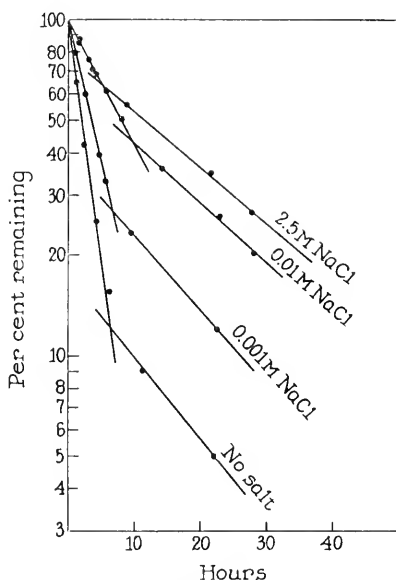


Fig. 5. Escape rate curves of ribonuclease in aqueous solutions with various amounts of sodium chloride.

If the 50% time is plotted against ionic strength the curve shown in Fig. 6 is obtained. When magnesium sulfate was used one-fourth the molar concentration was found to have the same quantitative effect as sodium chloride. Thus the ionic strength of the solution seems to be important rather than the specific ion involved.

An attempt was made to isolate the faster diffusing protein before the break for a run with no salt and also the part emerging after the break.

These two fractions showed no significant difference in enzymatic activity, optical activity, or in their chromatographic behavior on an ion exchange column.¹³ When both fractions were separately re-run in the dialysis procedure there was a noticeable difference in the slope and position of the slower component, but quantitatively much less than would have been expected from the initial run.

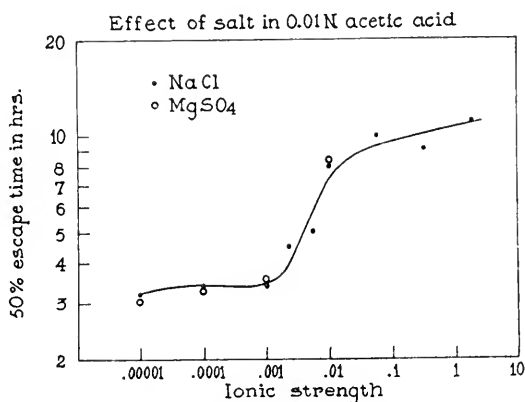


Fig. 6. Effect of ionic strength on 50% escape time.

That the effect noted was due to the solute rather than to some property of the membrane was shown by an experiment with a larger amount of the protein initially. At a time just prior to the break in the curve the inside solution was transferred directly to a new, freshly prepared membrane and cell. The effect from this point on was much more pronounced with a retarded escape rate and with the break much nearer the origin. This demonstrated that the effect was not due to a tendency of the membrane to become 'plugged'. The most likely explanation of the break in the curve would appear to be based on the possibility of more than one configuration, state of aggregation, degree of folding or shape in a single protein and with a sufficiently slow rate of interconversion so that two escape rates are registered in the curves.

Thus far, different membranes from the same roll have shown remarkable uniformity. Even from roll to roll in the same size there has been good uniformity. However, it would appear advisable to standardize each roll since two 20/32 rolls obtained recently have shown a slower escape rate than the earlier ones used to accumulate most of the data presented in this paper.

The escape rate can be reproducibly increased by stretching a wet membrane and decreased by acetylating it. A slower membrane when acetylated gives a slower escape rate for a solute (e.g., with 18/32 bacitracin barely passes) than a faster one (e.g., 20/32 gives a 50% escape time of 3 hours).

Both acetylated membranes readily pass amino acids. The further possibilities here do not require emphasis.

It would appear that the considerable data we have now accumulated all support the view that membrane diffusion with cellophane can be studied in such a way that the size and (or) shape of the molecule is the predominant property being measured. Perhaps, in conclusion, a speculation in regard to the basis for the membrane selectivity noted is in order.

Regardless of how thin a membrane may be, it can be regarded as offering a stationary structure with interstices or passages through which the molecules must diffuse. These passages will not all be of the same size. Thus a cross section might appear to be a mosaic of different sized holes as shown in Fig. 7. Some of the holes will be smaller than the molecular

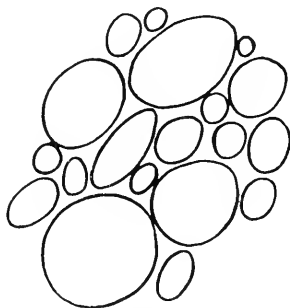


Fig. 7. Schematic drawing of a hypothetical cross section of a membrane.

dimensions of the solutes trying to diffuse through them and thus a small solute would have a larger total cross-sectional free area through which to diffuse than a larger one would have. This differential could be considerably enhanced as the point of total exclusion of passage is approached. If this simple mechanistic theory is correct, the selectivity of a membrane could be a function of the size distribution of the passages through which the solutes must diffuse.

The simple theory briefly outlined is not new. It has been considered before in the extensive studies dealing with the passage of solutes through membranes. No attempt will be made here to discuss this literature. A recent excellent review¹⁴ dealing with capillary permeability covers the subject, including present information about the pore size of certain membranes, the relation of this to the rate of passage of solutes with known effective diameters, etc. The factors controlling restricted diffusion¹⁴ have been treated in a mathematical way.

In spite of these studies, Spandau^{15,16} has presented data to show that dialysis with many smaller solutes can be performed in such a way that the rate seems to correlate better with the inverse of the square root of the molecular weight than it does with molecular volume. Our experience with the

polypeptides and proteins indicates that both the molecular weight and the shape must be considered.

If the molecular weight is known then dialysis can be a useful tool for studying the shape of the molecule in a particular solution. Something about the change in shape brought about by a change in environment can be revealed. Information about the state of aggregation also can be derived. Conversely, if the shape of the molecule is known a good estimate of the molecular weight can be made.

Finally, dialysis as outlined here can be a very useful separation tool which at the same time offers a degree of characterization with respect to size homogeneity.

BIBLIOGRAPHY

1. R. SIGNER, H. HANNI, W. KOESTLER, W. ROTTENBERG and P. VON TAVEL, *Helv. Chim. Acta*, **29**, 1984 (1946).
2. J. H. NORTHROP and M. L. ANSON, *J. Gen. Physiol.*, **12**, 543 (1929).
3. L. C. CRAIG and T. P. KING, *J. Am. Chem. Soc.*, **77**, 6620 (1955).
4. R. E. STAUFFER, in Weissberger, *Technique of Organic Chemistry*, Vol. III, 2nd edition, p. 79, Interscience Publishers, New York, 1956.
5. L. C. CRAIG, T. P. KING and A. STRACHER, *J. Am. Chem. Soc.*; **79**, 3729 (1957).
6. W. HAUSMANN, *J. Am. Chem. Soc.*, **78**, 3663 (1956).
7. S. MOORE and W. H. STEIN, *J. Biol. Chem.*, **211**, 907 (1954).
8. H. NEURATH, J. A. RUPLEY and WM. J. DREYER, *Arch. Biochem. Biophys.*, **65**, 243 (1956).
9. E. NEURATH and WM. J. DREYER, *Dis. Faraday Soc.*, No. 20, 32 (1955).
10. WM. F. HARRINGTON and J. A. SCHELLMAN, *Compt. rend. trav. lab. Carlsberg*, **30**, 21 (1956).
11. E. J. HARFENIST and L. C. CRAIG, *J. Am. Chem. Soc.*, **74**, 3087 (1952).
12. D. F. WAUGH, in Ciba Foundation Colloquia on Endocrinology, Vol. 9, p. 122, *Internal Secretions of the Pancreas*, J. and A. Churchill Ltd., London, 1956.
13. C. H. W. HIRS, S. MOORE and W. H. STEIN, *J. Biol. Chem.*, **200**, 493 (1953).
14. J. R. PAPPENEIMER, *Physiol Rev.*, **33**, 387 (1953).
15. H. SPANDAU, *Angew. Chem.*, **63**, 41 (1951).
16. H. SPANDAU and E. BRUNNECK, *Angew. Chem.*, **65**, 183 (1953).

Quantitative Papierchromatographie der Aminosäuren durch Isotopenverdünnung

F. TURBA

Physiologisch-chemisches Institut der Universität München

Eine Verbindung mit 'anormalem' Isotopengehalt wird durch die üblichen physikalisch-chemischen Methoden des Laboratoriums nicht von ihrem 'normalen' Analogon getrennt (Hevesy¹). Setzt man z.B. einem Proteinhydrolysat mit einem bestimmten Gehalt an Glutaminsäure eine definierte Menge N¹⁵-Glutaminsäure zu, so läßt sich aus der eingetretenen 'Verdünnung' des Isotops auch in einer *nicht* quantitativ isolierten, aber reinen Probe der Glutaminsäuregehalt des Hydrolysats berechnen ('Isotopenverdünnungsmethode' nach Rittenberg²). Man kann die zu bestimmende Verbindung auch mit einem isotope markierten Rest substituieren und einen Überschuß des unmarkierten Derivats zusetzen ('Isotopenderivatmethode' nach Keston, Udenfriend u. Mitarbeitern.^{3,4,5} Wir haben Versuche unternommen, das Isotopenverdünnungsverfahren an C¹⁴-markierten Aminosäuren mit der Papierchromatographie zu kombinieren, um diese Methode dadurch quantitativ zu gestalten.⁶

BESCHREIBUNG DER METHODE

0,5-1,0 mg Protein werden in 5 ml glasdestillierter, 20%iger Salzsäure in einem Reagensglas (mit Schliff) gelöst und die C¹⁴-Analoge der zu bestimmenden Aminosäuren in solcher spezifischer Aktivität zugesetzt, daß das Mengenverhältnis inaktive: aktiver Aminosäure etwa 10 : 1 und die *meßbare* Aktivität etwa 1000-4000 Impulse/min beträgt. Man verschließt die Gläser mit Schliffstopfen und hydrolysiert mindestens 24 Stunden bei 110° im Aluminiumblock. Danach werden die Reagensgläser direkt an einen Rotationsverdampfer angeschlossen, die Lösung im Vakuum mehrmals zur Trockne gebracht und genau neutralisiert. Die 2-dimensionale Papierchromatographie erfolgt wie üblich (z.B.: 1. Dimension: sec.-Butanol/Ameisensäure/Wasser 70 : 15 : 15; 2. Dimension: Methyläthylketon/Pyridin/Wasser 70 : 15 : 15).⁷ Man lokalisiert die zu bestimmenden Aminosäuren im Chromatogramm durch Radio-autographie (24-48 Std. Aufpressen auf

Röntgenpapier) und schneidet danach die betreffenden Stellen des Chromatogramms so aus, daß bei eventuell nicht vollständiger Trennung nur der reine Anteil der Aminosäure zur Elution gelangt. Als Kontrollen werden Ausschnitte aminosäurefreier Stellen des Chromatogramms gleicher Größe nach Trocknung über Phosphorpentoxid bei 0,1 mm und 50° für 12 Std. (zur Entfernung von Ammoniak und flüchtigen Ammoniumsalzen) ebenfalls mit 5 ml Wasser (in 1ml-Portionen) in einer Mikroglassfritte eluiert. Die wenn nötig neutralisierten Eluate (Vermeidung des Ausfällens von Desoxycholsäure, s.u!) werden auf die Meßschälchen überführt.

Messung der Radioaktivität. Es erwies sich als nahezu unmöglich, Aminosäuren ohne spreitenden Zusatz und auf Schälchen mit ebenem Boden in gleichmäßiger Schicht aufzutragen, wie das für reproduzierbare Messungen unentbehrlich ist. Am günstigsten erwiesen sich V2A-Schälchen mit gewölbtem Boden (Abb. 1). Wir fanden weiter, daß bei Zusatz von 200 μg Natrium-

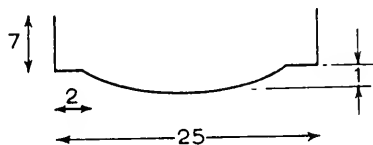


Abb. 1. V2A-Schälchen mit schwach gewölbtem Boden und Rand zur Messung der C^{14} -Aktivität von Aminosäuren. Die Schälchen wurden zwischen einer Matrix und einem Stempel mit einheitlichen Maßen gepresst. Maße in mm.

Desoxycholat zur Probe das Auftragen in Form eines gleichmäßig dünnen Films gelingt. Außerdem wird durch diesen Zusatz erreicht, daß die Änderung der Selbstabsorption, bedingt durch die verschiedenen Mengen an Aminosäuren, bei Vorgabe der relativ großen Menge an organischer Substanz keine Rolle mehr spielt (Abb. 2). Die Messung als C^{14}O_2 im Bernstein-

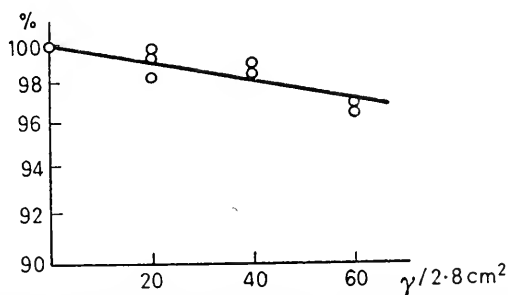


Abb. 2. Einfluß der angewandten Menge der Aminosäuren in $\mu\text{g}/\text{cm}^2$ auf die Selbstabsorption von C^{14} . 200 μg Na-Desoxycholat wurden jeder Probe hinzugefügt.

Ballentine-Zählrohr^{8,9} oder im dynamischen Elektrometer ist umständlicher, da sie eine Veraschung¹⁰ der Substanz voraussetzt. Die Fällung als $\text{BaC}^{14}\text{O}_3$ hat sich bei den vorliegenden kleinen Mengen nicht bewährt. Außerdem hat

die Messung der C^{14} -Aminosäuren als solche den Vorteil, daß man die gesamte, ungeteilte Probe anschließend zur Konzentrationsbestimmung mit Ninhydrin verwenden kann, was die Genauigkeit erhöht; die Desoxycholsäure stört dabei nicht.

Messung der Konzentration. Für die Konzentrationsbestimmung der Aminosäuren kam aus Gründen der Empfindlichkeit und Genauigkeit nur die kolorimetrische Ninhydrinmethode in Frage; sie erfolgte nach der Vorschrift von Moore und Stein,¹¹ aber im evakuierten System (Van Slyke¹²) zur Verhinderung des Entweichens von $C^{14}O_2$ (vgl. Abb. 3). Nach Entwicklung der Farbe (20 min, 100°) wurde 10 min. Luft durch das Reaktionsgefäß in Barytwasser gesaugt.

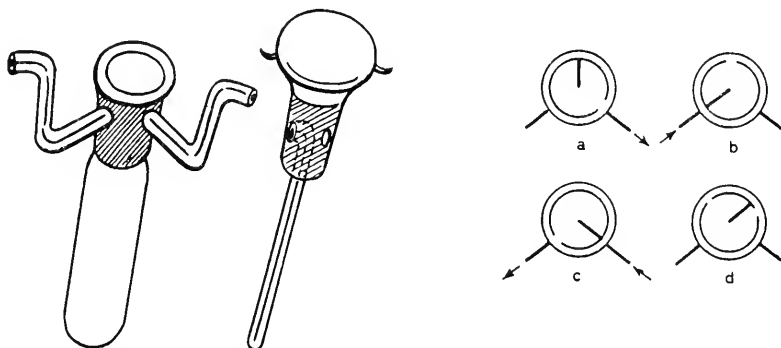


Abb. 3. Reagenzglas mit Schliffstopfen und kapillaren Ansätzen zur kolorimetrischen Ninhydrinbestimmung von C^{14} -Aminosäuren. (a) Stellung des Stopfens zum Evakuieren vor dem Erhitzen. (b) Einsaugen von Luft nach Abkühlen der Analysenprobe. (c) Durchleiten von Luft zur Entfernung von $C^{14}O_2$. (d) Verschluss des evakuierten Röhrchens während der Farbentwicklung.

Berechnung. Diese erfolgte nach $a = b \left(\frac{C_0}{C} - 1 \right)$ (a = inaktive Aminosäure in μg ; b = zugesetzte aktive Aminosäure in μg ; C_0 = Impulse/min: μg für die zugesetzte aktive Aminosäure; C = Impulse/min: μg für die isolierte Analysenprobe). Zur Genauigkeit und Fehlerberechnung vgl. Rittenberg² bzw. Riedel¹³. Unter günstigen Voraussetzungen ($C_0 : C \geq 10$; gemessene Aktivität 1000-4000 Impulse/min) kann man einen Fehler von 2-3% erwarten (reiner Meßfehler der Aktivitätsmessung $< 1\%$,¹⁴ Gesamtfehler der Aktivitätsbestimmung 2-3%; Fehler der Konzentrationsbestimmung $< 1\%$). Vgl. Tab. 1.

DISKUSSION

Der Hauptvorteil der Methode beruht darin, daß an keiner Stelle des Arbeitsgangs quantitative Ausbeuten erforderlich sind (selbst die Übertragung der Substanz vom Meßschälchen nach der Bestimmung der C^{14} -Aktivität zur kolorimetrischen Bestimmung muß nicht vollständig sein, wenn man den verbleibenden Rest der Aktivität zurückmißt). Damit wird das Verfahren

Tab. 1

BESTIMMUNG VON AMINOSÄUREN IM GEMISCH DURCH ISOTOPEN-VERDÜNNUNG MIT C¹⁴-AMINOSÄUREN UND PAPIERCHROMATOGRAPHIE

Hydrolysat	ein- gesetzte Menge Protein	Markierte Aminosäure			Iso- lierte Probe C J/min: µg	a Inak- tive Amino- säure µg (% d.Th)	gefundene g Amino- säure/100 g Protein	Litera- turan- gaben†
		Markie- rungs- ort	b µg	Co J/min: µg				
Künstl. Gemisch*	69,6 µg (Alanin)	1-C ¹⁴ - Alanin	7,5	2666/7,5	1252/37	70,8 (101%)	—	—
β-Lacto- globulin	720 µg	1-C ¹⁴ - Alanin	7,5	2666/7,5	1500/31	51,3	7,12	6,7;‡ 7,1§
β-Lacto- globulin	720 µg	1-C ¹⁴ - Phenyl- alanin	3,2	3026/3,2	1250/11,4	27,8	3,86	3,78‡

* Das Künstliche Gemisch war einem Hydrolysat nachgebildet.

† Wahrscheinlichster Wert fettgedruckt.

‡ G. R. Tristram, *Advances in Protein Chemistry*, 5 (1949).

§ A. S. Keston, S. Udenfriend u. R. K. Cannan, *J. Am. Chem. Soc.*, 68, 1390 (1946).

unabhängig von den Fehlerquellen anderer Analysenmethoden. Bei der Hydrolyse werden manche Aminosäuren teilweise zerstört und entgehen daher normalerweise z.Tl. der Bestimmung, bes. bei Hydrolyse in Gegenwart von Zucker; von Glyko-; Lipo- und Chromoproteiden; von Extrakten aus biologischem Material). Bei der Isotopenverdünnung ist dies nicht der Fall (Tab. 2).

Tab. 2

VERLUSTE BEI DER ANALYSE VON METHIONEN NACH ERHITZEN IN 20% HCl IN GEGENWART VON GLUCOSE

42,2 µg Methionin, 80,0 µg Glucose, 2 ml 20%ige glasdest. HCl, 24 Std., 100°. Chromatographie eindimensional in sec-Butanol/Ameisensäure/Wasser.

C ¹⁴ H ₃ -Methionin zugesezt zu inakt. Methionin	b	Co I/min: µg	C I/min: µg	a	Ausbeute %d. Th.	Verlust bei
vor Erhitzen mit Säure	8.4 µg	4430/8.4	2900/33.8	43,2	102,5	—
nach „ „	8.4 µg	4430/8.4	2620/14.2	15,6	36,2	'Hydrolyse'
nach Elution aus Chromat.	8.4 µg	4430/8.4	4156/15.2	7,8	18.5	Chromato- graphie Elution

Tab. 2 zeigt das Verhalten von *Methionin* bei Erhitzen in Gegenwart von Glucose mit 20% HCl. Setzt man C^{14} -markiertes *Methionin* vor der 'Hydrolyse' zu, so findet man die Aminosäure quantitativ wieder, da die Zerstörung das inaktive wie das aktive *Methionin* gleichermaßen betrifft. Bei Zusatz nach 'Hydrolyse' dagegen sind nur 36,2%, nach der Elution aus dem Chromatogramm nur noch 18,5% bestimmbar (zu Verlusten beim Chromatographieren in versch. Lösungsmitteln, bes. Phenol, Trocknen bei versch. Temperaturen, durch Lichteinwirkung usw. vgl.^{15,16})

Auch in normalen Hydrolysaten erfassen wir nur 40-60% der eingesetzten Mengen bei der Konzentrationsbestimmung am Schluß des Analysenwegs in Substanz (vgl. Tab. 1).

Die zur Kompensation der bei der Papierchromatographie eintretenden Verluste von Fischer u. Dörfel¹⁷ vorgeschlagene Methode, zu jeder Analysenprobe ein auch quantitativ möglichst identisches künstliches Aminosäuregemisch unter genau gleichen Bedingungen mitzuanalysieren und die Analysenwerte auf seine Ausbeuten zu beziehen, ist wegen der notwendigen Vorversuche langwierig und weniger genau als die Isotopenverdünnung. Ein weiterer wesentlicher Vorteil des hier beschriebenen Verfahrens besteht in der genauen und verlustlosen Markierung der Lage der Aminosäuren im Chromatogramm durch Radioautographie. Diese Tatsache ermöglicht die Anwendung der wirksameren 2-dimensionalen Papierchromatographie ohne teilweises Zerstören der Aminosäuren durch Besprühen mit farbbildenden Reagentien,^{18,19} Erzeugung von Fluoreszenz durch Erhitzen²⁰ usw. Auch gelingt die Trennung *aller* Aminosäuren im gleichen Chromatogramm, während beim eindimensionalen Verfahren stets mehrere Versuche mit verschiedenen Lösungsmitteln notwendig sind (McFarren). Schliesslich verdient die quantitative Bestimmung in Lösung nach Elution den Vorzug vor der Kolorimetrie im Papier z.B. nach Besprühen mit Ninhydrin.²¹

Vor den beschriebenen Isotopenderivatmethoden (mit J^{131} und S^{35} -p-Jodphenylsulfonyl-Aminosäuren) hat das entwickelte Verfahren den Vorteil besserer Trennungen bei der Papierchromatographie und der dauernden Verwendbarkeit der markierten Substanzen infolge der hohen Halbwertszeit von C^{14} , vor den N^{15} -Verfahren der einfacheren Messung des radioaktiven Isotops und der Benutzung kleinerer Mengen, wie sie die Papierchromatographie erfordert.

Die einzig ausschlaggebende Fehlerquelle des Verfahrens ist die Verunreinigung der Analysensubstanz mit Ninhydrin-positivem Material, das die Ergebnisse fälschen würde. Da jedes Chromatographiepapier auch nach Auswaschen noch geringe Mengen Verunreinigungen enthält, die mit Ninhydrin reagieren (vgl. Abb. 4), ist es notwendig, von den Analysenwerten entsprechende kleine Korrekturen in Abzug zu bringen. Die Benutzung von reinem Glasfaserpapier zur Chromatographie, das frei von diesem Fehler ist, setzt das Imprägnieren mit Puffersalzen voraus; die C^{14} -Messung kann danach nur noch als $C^{14}O_2$ erfolgen. Das Gleiche gilt für die Anwendung

dieses Verfahrens auf Eluate mit einem Gehalt an nichtsublimierbaren Salzen aus Ionenaustauschersäulen. Ammoniak und Ammoniumsalze werden im Vakuum entfernt (s.o.).

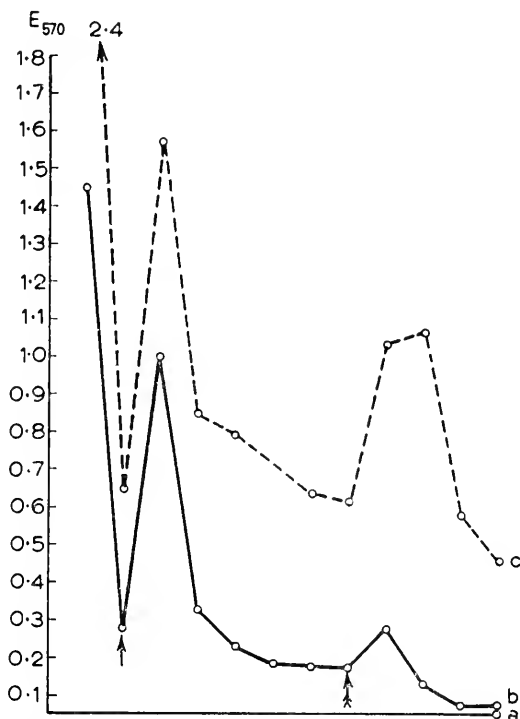


Abb. 4. Auswaschen von Chromatographiepapier. 300 cm² Schleicher-Schüll-Papier 2043 b wurden im Durchlaufverfahren mit Wasser, anschließend (beginnend bei ↑) mit 5% Trichloressigsäure, schliesslich (beginnend bei ↑) mit 5% methanolischer KOH eluiert. Je 10 ml des Eluats wurden im Vakuum eingedampft und nach Ausäthern bzw. Neutralisation die Ninhydrinbestimmung nach Moore und Stein durchgeführt. (a) Nullwert der Reagentien; (b) unhydrolysiertes Eluat; (c) hydrolysiertes Eluat (24 Std., 110°, 20% HCl). Die Extinktion bei 570 mμ von 1,0 entspricht 37,0 μg Alanin.

ANWENDBARKEIT

Das Verfahren hat Bedeutung für die Bestimmung der *genauen Zahl bestimmter Aminosäure-Reste*²² in größeren Peptiden und in Proteinen. Es ist geeignet, Verluste bei der Isolierung von Aminosäuren in biologischer Material (z.B. durch Hydrolyse, Enteiweissung, Entsalzung usw.) auszuschalten. Man darf erwarten, daß mit seiner Hilfe die Genauigkeit von Bestimmungen der Amino- und Carboxylendgruppen in Proteinen, die z.Zt. mit Fehlern von teilweise 50% und mehr belastet sind²³ und umständlicher (gelegentlich auch unsicherer) Korrekturen bedürfen, ähnlich gesteigert werden kann, wie bei der Bestimmung von Aminosäuren in Hydrolysaten; wir sind mit solchen Versuchen beschäftigt.

REFERENCES

1. G. HEVESY, Cold Spring Harbor Sympos. *Quant. Biol.*, **13**, 129 (1948).
2. D. RITTENBERG u. G. L. FOSTER, *J. biol. Chem.*, **133**, 737 (1940).
3. A. S. KESTON, S. UDENFRIEND u. R. K. CANNAN, *J. Am. Chem. Soc.*, **68**, 1390 (1946).
4. A. S. KESTON, S. UDENFRIEND u. M. LEVY, *ibid.*, **69**, 3151 (1947).
5. A. S. KESTON, S. UDENFRIEND u. R. K. CANNAN, *ibid.*, **71**, 249 (1949).
6. vgl. F. Turba, in Hoppe-Seyler-Thierfelder, *Handbuch der physiologischen und pathologisch-chemischen Analyse*, Springer-Verlag Berlin, Göttingen, Heidelberg, 1955, Bd. **3**, 1835.
7. F. TURBA, *Chromat, Methoden in der Proteinchemie*, Springer-Verlag, 1954, S. 175.
8. W. BERNSTEIN u. R. BALLENTINE, *Rev. Scient. Instruments*, **21**, 158 (1950).
9. D. D. VAN SLYKE, R. STEEL u. J. PLAZEN, *J. biol. Chem.*, **192**, 769 (1951).
10. D. D. VAN SLYKE u. J. FOLCH, *J. biol. Chem.*, **136**, 509 (1940).
11. S. MOORE u. W. H. STEIN, *J. biol. Chem.*, **211**, 907 (1954).
12. D. D. VAN SLYKE, D. A. MCFADYEN u. B. P. HAMILTON, *J. biol. Chem.*, **141**, 671 (1941-2).
13. O. RIEDEL, *Angew. Chem.*, **67**, 643 (1955).
14. vgl. dazu: M. CALVIN u.a., *Isotopic Carbon*, John Wiley u. Sons, Inc., New York, 1949, S. 288.
15. F. TURBA, A. LEISMANN u. G. KLEINHENZ, *Biochem. Z.*; im Druck.
16. M. K. BRUSH, R. K. BONTWELL, A. D. BARTON u. C. HEIDELBERGER, *Science*, **113**, 4, (1951).
17. F. G. FISCHER u. H. DÖRFEL, *Biochem. Z.*, **324**, 544 (1953).
18. A. J. LANDUA u. J. AWAPARA, *Science*, **109**, 385 (1949).
19. R. A. BOISSONAS, *Helvet. Chim. Acta*, **33**, 1975 (1950).
20. D. M. P. PHILLIPS, *Nature* (London), **164**, 545 (1949).
21. vgl. z. Bsp. E. HILLER, F. ZINNERT u. G. FRESE, *Biochem. Z.*, **323**, 245 (1952).
22. Zur erforderlichen Genauigkeit vgl. G. R. TRISTRAM, *Adv. Protein Chemistry*, **5**, 124 (1949).
23. R. R. PORTER u. F. SANGER, *Biochem. J.*, **42**, 287 (1948).

SPECIFIC PROTEINS

Haemoglobin and Myoglobin

Crystallographic studies of myoglobin

J. C. KENDREW

Medical Research Council Unit for Molecular Biology,
Cavendish Laboratory, University of Cambridge

This paper is an outline of recent developments in X-ray studies of the structure of myoglobin. These studies have been undertaken in our laboratory and in the Royal Institution at London with a number of colleagues, past and present, including M. M. Bluhm, G. Bodo, H. M. Dintzis, V. M. Ingram, J. Kraut, R. G. Parrish, P. J. Pauling, D. C. Phillips, Mary Pinkerton, Helen Scouloudi, and H. Wyckoff. Several similar projects are under way in other laboratories, all having the object of determining the structure of a globular protein purely by X-ray methods; they are concerned with a variety of proteins, notably haemoglobin, insulin, and ribonuclease. Although the methods they use, and the stage they have reached, vary from project to project, the myoglobin programme may be taken as typical of them all. Until four years ago no one knew, even in principle, how structures as complex as proteins could be solved; but Perutz's successful application of the so-called method of isomorphous replacement to haemoglobin¹ resolved a deadlock in the subject by indicating a secure, though difficult, line of advance. Since then the potentialities of this method in protein crystallography have been explored, and it has been used to obtain several two-dimensional electron density projections of protein crystals (haemoglobin of horse² and ox;³ myoglobin of sperm whale⁴ and seal;⁵ ribonuclease).^{6,7} Gradually it has become clear that the only way of exploiting the new method usefully is to extend it to three dimensions: for a three-dimensional electron density map, whatever its limitations as to resolution and accuracy, would at least completely avoid the overlapping of atoms which makes *projections* of complex structures (including all those just enumerated) impossible to interpret. In most of the crystalline protein projects the possibility of extending the analysis to three dimensions is being explored at the present time. This paper will show that at any rate in myoglobin such extension is possible in principle, although we have not yet collected all the necessary experimental data, nor are we yet certain what resolving power will be needed to reveal recognizable features of the structure.

THE BASIC CRYSTALLOGRAPHIC PROBLEM

It is well known that a three-dimensional Fourier summation, carried out with the amplitudes and phases of all the reflexions in the X-ray pattern of a crystal as terms, gives a representation of the electron density in the unit cell of the crystal; and that the degree of resolution obtained depends on the Bragg spacing of the highest-order reflexions included in the summation, the number of reflexions involved being proportional to the cube of the spacing (e.g. in myoglobin about 400 for 6 Å., 25,600 for 1.5 Å.). Furthermore, if particular sets of reflexions (known as 'zones') are selected and their amplitudes and phases used in a *two*-dimensional Fourier synthesis, the result is a representation of the electron density *projected* along a particular crystallographic direction onto a plane. The amplitudes of the reflexions are the square roots of the observed intensities and are therefore measurable by photographic or other means; but the phases are not directly observable, and in this consists the basic difficulty of X-ray analysis—that experiment gives one just half the information required to solve a structure. The history of the subject is largely a history of attempts to discover ways to get round this difficulty. In simple structures straightforward trial-and-error methods, or sometimes what are in effect sophisticated variants of them, have often sufficed; but in structures as complex as the proteins there is little hope that these will succeed, and indeed in the past they provided only too much free play for the exercise of speculative theories and models of protein structure, without the possibility of proving whether any of the speculations were wrong or right. The method of isomorphous replacement (and its near relative the heavy-atom method) was also developed for simple structures, and, where chemically practicable, it does often lead to an unequivocal solution of the structure. It was the successful application of this method to proteins which first put the structure analysis of their crystals onto a firm basis and it is now the principal tool in all laboratories studying protein crystal structure. To illustrate its use we shall describe its application to myoglobin, first of all in two dimensions.

The reason for limiting the investigation to two dimensions in the first instance, is simply that in most crystals the phases of certain reflexions are restricted by the symmetry of the crystal to values of 0 or π ; or, in other words, these so-called *real* reflexions may be regarded simply as having *signs*, either positive or negative (unlike the generality of reflexions whose phases may assume any values from 0 to 2π). Thus the question what phase is possessed by a reflexion of this kind can have only two answers, '+' or '-', whereas to reflexions in general no such restriction applies. It is consequently much easier to determine the phases of real reflexions; and the real reflexions from a given crystal fall into a group or groups (the real zone or zones of reflexions) such that each zone, when it is used as terms of a Fourier synthesis, gives a projection along a particular crystal axis onto a plane, i.e. a two-dimensional synthesis.

MYOGLOBIN AND THE ISOMORPHOUS REPLACEMENT METHOD IN TWO DIMENSIONS

Let us assume that an isomorphous heavy-atom derivative of crystalline myoglobin has been prepared, containing, for example, one mole of mercuri-iodide ion per mole of protein. This implies that one mercuri-iodide group is specifically attached to a single site on each myoglobin molecule, and that the resulting complex has crystallized in precisely the same arrangement as normal unsubstituted myoglobin, the cell dimensions and packing arrangement of the molecules remaining unchanged. We shall return later to the question how such derivatives can be prepared, and assume for the moment simply that it has been done.

Owing to the presence of the HgI_4^{-2} group the intensities of all the reflexions in the X-ray pattern will be changed. Even as weighty an ion as mercuri-iodide is very small compared with the whole protein molecule, however, and at first sight one might guess that the changes in the diffracted intensities would be too small to be measured. It was Perutz's contribution to realize, in the course of his studies of haemoglobin, that the changes will in reality be very pronounced, in consequence of the low average intensity of the reflexions from a unit cell so uniformly filled with matter as is that of a protein; the average contribution of a single specifically-situated mercury atom is actually about one quarter of the average contribution of the myoglobin molecule as a whole.

The first step in the analysis is to measure the intensities of all the real reflexions before and after the attachment of mercuri-iodide. The square roots of these intensities are the moduli of the amplitudes, $|F|$ and $|F'|$. From them are calculated the so-called 'difference intensities' ($|F| - |F'|$),² and these are used as terms of a Fourier synthesis. A Fourier synthesis using the *intensities* instead of the *amplitudes* of the reflexions from a crystal is known as a vector synthesis, and Patterson showed many years ago that its physical significance is to give a representation of all the interatomic vectors in the structure, transferred to a common origin without regard to the absolute co-ordinates of the atoms at which each vector terminates. If there are n atoms per unit cell there are n^2 interatomic vectors, so the resulting vector map is usually hopelessly confused if the structure is as complex as that of a protein. By the subtractive procedure we have just outlined, however, we virtually cancel out all the atoms of the protein and leave in effect an empty unit cell containing only the heavy atoms; the vector synthesis simply shows the vectors between these atoms (together with a high peak at the origin representing vectors of zero length between each atom and itself). The result is shown in Fig. 1. The origin is in the centre; the unit cell contains two myoglobin molecules related by a screw dyad axis of symmetry, and consequently two mercury atoms A and B. The vectors AB and BA between these atoms are symmetrically disposed about the

origin, and their positions, together with the known symmetry properties of the crystal, enable us to locate the heavy atoms in the structure, at least as far as their x and z co-ordinates are concerned (the projection shown is made along the y axis, and is the only one which can be calculated from real reflexions in the monoclinic space group to which this crystal belongs).

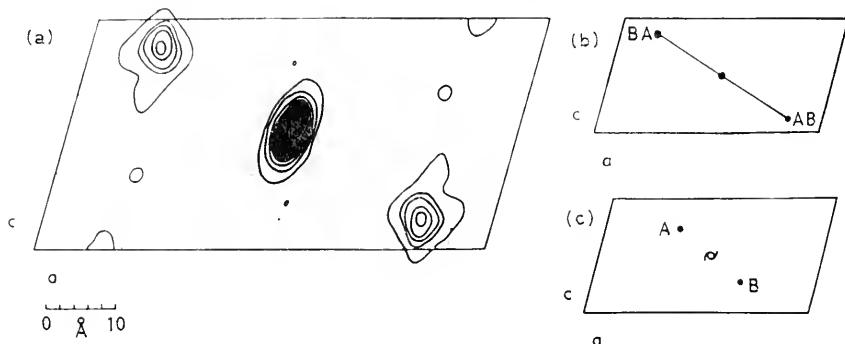


Fig. 1. (a) Difference-vector synthesis of the mercuri-iodide complex of myoglobin (type A, y projection). (b) Schematic representation of vectors between heavy groups. (c) Positions of mercuri-iodide groups in the unit cell. The unit cell contains two myoglobin molecules and therefore two mercuri-iodide groups.

The next step is to compute the contributions—positive or negative—which a heavy atom in the position assigned would make to each reflexion. The signs of these contributions, which are calculated in a straightforward and unambiguous manner, make it possible to allot signs to the protein reflexions by inspection. For if a protein reflexion is positive and the contribution of the heavy atom is likewise positive, the resultant is a stronger reflexion; the reflexion is also strengthened if a *negative* heavy atom contribution is added to a *negative* protein term; on the other hand a *positive* heavy atom contribution together with a *negative* protein contribution, or *vice versa*, leads to a weakening of the reflexion.

The amplitudes of the protein reflexions having been measured and their signs determined, the way is clear to compute a Fourier synthesis, which is a representation of the electron density in the unit cell projected along the y axis. Fig. 2 shows the result. It is computed with a resolution of 6 Å.: we are, however, projecting some 31 Å. thickness of protein and mother liquor onto each point (31 Å. is the length of the y axis along which the projection is made). It is hardly surprising, therefore, that the various features of the structure overlies one another in such confusion that in the resulting projection nothing can be disentangled. Increasing the resolution does not help at all; the same projection computed with a resolution of 4 Å. is almost identical and quite as unintelligible. Of the correctness of the projection there is no doubt: for the whole process of sign determination can be repeated with a different isomorphous substitution in the same unit cell, using, for example, a gold atom which attaches itself to a different

position on the protein molecule, and the signs of the protein reflexions so derived agree perfectly with those obtained from the mercuri-iodide substitution. Similar two-dimensional projections of other protein crystals have been prepared, as has already been mentioned, but in no case can the projection be interpreted in chemical terms. Nevertheless these projections represent an important stage in the development of the subject, since they

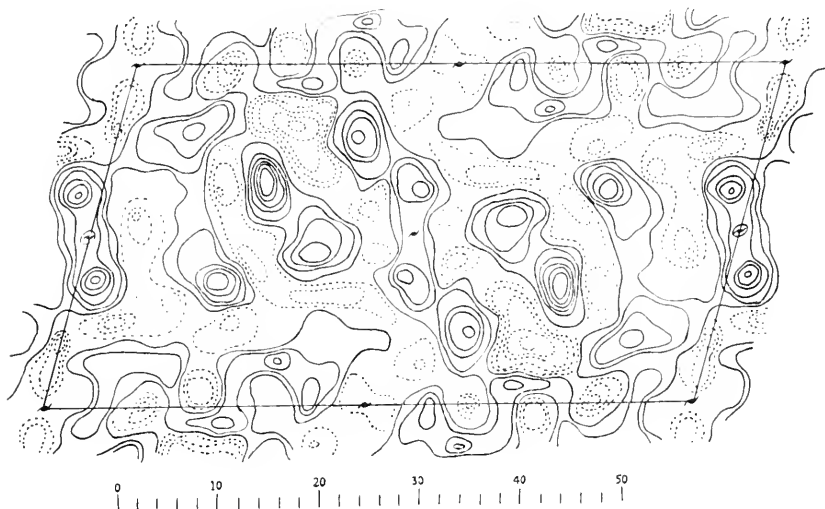


Fig. 2. Fourier projection of Type A myoglobin crystals along y , including all terms of spacing greater than 6.6 \AA .

are the first definitely correct pictures of protein structures to be obtained. It has, however, become clear that isomorphous replacement, powerful method of analysis though it is, will only be exploited usefully if it can be extended to three dimensions in order to circumvent the difficulty of overlap in projections. We shall consider how this might be done after a brief discussion of the chemical problems involved in attaching heavy-atom derivatives to myoglobin.

THE ATTACHMENT OF HEAVY GROUPS TO MYOGLOBIN

In haemoglobin the convenient presence of free sulphhydryl groups makes possible the use of any of the standard $-\text{SH}$ group reagents, such as *p*-chloro-mercuribenzoate or methyl-mercury, for attaching mercury atoms to the molecule. Myoglobin contains no free sulphhydryl groups in any species which we have so far examined, and accordingly less straightforward methods must be adopted. The isomorphous replacement method cannot easily be applied, however, unless the heavy atom is attached specifically to a single site in the molecule, or at the most to a very few sites. In myoglobin the

obvious unique site is the single haem group itself, but we found that although some of the usual haem group reagents (such as isocyanides and nitroso-compounds) could be synthesized with heavy atoms in their molecules, it was very difficult attaching them stoichiometrically to the haem group, partly because of the low solubility of reagents containing heavy atoms such as mercury and partly on account of the very high affinity of myoglobin for oxygen. Unless the most stringently anaerobic conditions are maintained the complexes of myoglobin with, for example, isocyanides rapidly dissociate: and we have not so far been able to carry through the whole process of preparation, crystallization and X-ray photography without an unacceptable amount of decomposition.

The approach which finally proved successful, and which was mainly developed by Drs Bodo and Dintzis, was to crystallize myoglobin in the presence of small amounts of suitable ions containing heavy metals (generally one or two moles of ion per mole of protein); the ions were selected in the light of such general information as is available about the chemical affinities of different protein side-chains. Crystals thus prepared were tested directly by means of X-rays, and not by chemical methods. If their diffraction pattern proved to be identical with that of a normal crystal it was evident that no combination had taken place. But if there were appreciable changes in the intensities of the reflexions, it could be concluded that attachment had taken place. The next stage would be to see whether the combination was at a single site, and to locate that site in the unit cell; for, as already indicated, indiscriminate combination with a large number of sites on the protein surface—with all the lysine amino groups, for example—would be useless for the present purpose. To test this, Fourier methods are used; just as we can prepare a difference vector projection of the heavy-atom/heavy-atom vectors from the changes in diffracted *intensity*, so by using the changes in *amplitude* as terms in a Fourier synthesis (once the signs of the reflexions have been established) we can in effect subtract out the protein and locate the positions of the heavy atoms responsible for the changes in the diffraction pattern. An example of the resulting 'difference-Fourier projection' is given in Fig. 3; this shows the effect of crystallizing myoglobin with *p*-chloro-mercuribenzene sulphonate (PCMBs), and indicates that a single mercury atom has been attached to each of the two myoglobin molecules in the unit cell (the other atoms in PCMBs are too light to show up in the difference-Fourier projection). Of course, PCMBs is a reagent for sulphhydryl groups, and there was no a priori reason to expect that it would combine with myoglobin at all: indeed we discovered by accident that it does so, and we still have no certain knowledge of the chemistry of the process. The same is true of most of the other reagents we have investigated. These include the ions AuCl_4^- , Ag^+ (in silver nitrate), HgI_4^{2-} , $\text{Hg}(\text{NH}_3)_2^{2+}$ (made by dissolving mercuric oxide in concentrated ammonium sulphate), and AuI_4^- . The results of these experiments are summarized in

Fig. 4, which shows the points of attachment (in projection) of the principal ligands which were successfully put on. A number of interesting chemical questions arise; for example, AuI_4^- occupies the same site as HgI_4^{2-} , suggesting that the central atom, and even the charge, may play only a subsidiary role in determining the mode of attachment. Again, Ag^+ and AuCl_4^- are attached to the same site, in spite of their charges being apparently

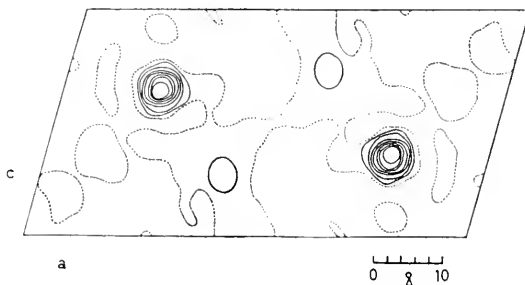


Fig. 3. Difference-Fourier projection of *p*-chloro-mercuri-benzene sulphonate derivative of Type A myoglobin along *y*, showing positions of mercury atoms.

opposite. However, it is found that the gold atom becomes attached only if the crystals are left for many months in their mother liquor before X-ray pictures are taken, suggesting that some secondary chemical process is a necessary preliminary to combination.

There are other examples in which we understand the chemistry rather better. Thus both *p*-iodo-nitrosobenzene and $\text{HO}_3\text{S}\cdot\text{C}_6\text{H}_4\cdot\text{Hg}\cdot\text{S}\cdot\text{C}_6\text{H}_4\cdot\text{NC}$

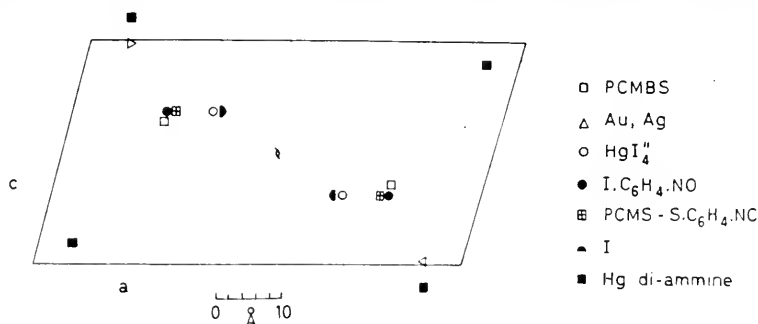


Fig. 4. Heavy-atom complexes of myoglobin type A, showing the positions of heavy atoms in the unit cell in *y* projection.

(PCMBS— $\text{S}\cdot\text{C}_6\text{H}_4\cdot\text{NC}$) are specific haem group reagents, and although it was not possible to use them for determining the signs of the protein reflexions, we were able to use the protein signs established by other methods to work backwards and locate the heavy atoms in these reagents with relation to the crystal axes. In each complex the heavy atom is separated from the iron atom of the haem group by the same number of intermediate atoms, and we do in fact find, as Fig. 4 shows, that they are in almost the same

position in the cell. We conclude that the iron atom of the haem group must be somewhere on the surface of a sphere with centre at the heavy-atom position and radius of some 6 Å. We have thus succeeded in approximately locating the haem group with respect to the axes of the unit cell.

At the other extreme from the specific haem-group reagents are some ions for which, the evidence suggests, combination is interstitial rather than truly chemical; that is to say, the ion lodges in a convenient niche on the protein surface or perhaps in a 'cosy corner' between two neighbouring protein molecules. In such cases the ion may combine specifically with only one particular crystal form, any other form (prepared by different methods, and having the myoglobin molecules packed together in a different manner) either giving no combination at all, or else perhaps combination at several different sites. In general it will be evident that we are very ignorant of the way in which the heavy groups we use are attached to the protein. In most cases it might be quite laborious to discover by conventional chemical methods what mechanism is involved, and it is interesting that we are at present able to locate the heavy atoms far more easily by X-rays than by chemical methods, and this with certainty although the structure of the protein as a whole is still unknown. Even dynamic effects may be observed. Thus on prolonged irradiation of a crystal of the PCMBMS complex with X-rays the mercury atom migrates from its normal position in the unit cell to another one, which turns out (in projection at least) to be identical with the site occupied by the heavy atom in the silver and gold complexes!—the difference-Fourier projection calculated from the amplitudes before and after irradiation is found to contain a region of negative electron density at the former site and a positive peak at the latter. All in all, methods for attaching heavy groups to proteins might well repay study by conventional chemical techniques, and certainly protein crystallographers would be glad to have them investigated in the expectation that analogous and perhaps better reagents might be discovered.

THREE-DIMENSIONAL METHODS

In this final section we shall consider what is involved in making a Fourier synthesis of a protein crystal in three dimensions. All the reflexions whose spacing exceeds the resolution required must be included as terms of such a synthesis, and a large majority of them will have general phases. Their amplitudes, together with the corresponding amplitudes from a crystal containing a heavy atom in a known position in the unit cell, are measured by the usual techniques. Now if we wished to predict the change in the amplitude of a reflexion of general phase which would be caused by a heavy atom at a particular site in the unit cell, it would be necessary to treat the amplitude due to the protein and that due to the heavy atom as vectors, and calculate their resultant by means of a vector diagram. Thus Fig. 5 (*a*) shows how the protein reflexion (with amplitude $|F_p|$ and phase ϕ_p) and

a heavy atom contribution $|f_A|$ (with phase ϕ_A) combine to give a reflexion of amplitude $|F_{PA}|$ and phase ϕ_{PA} in the compound crystal. In practice we are presented with the inverse problem: we are given $|F_P|$ and $|F_{PA}|$ (the square roots of the measured intensities) and $|f_A|$ and ϕ_A (obtained by straightforward computation from the known position of the heavy atoms in the unit cell). We proceed as shown in Fig. 5 (b), drawing two concentric circles whose radii are $|F_P|$ and $|F_{PA}|$ respectively, and marking off intercepts equal in length and direction to the heavy-atom contribution f_A .

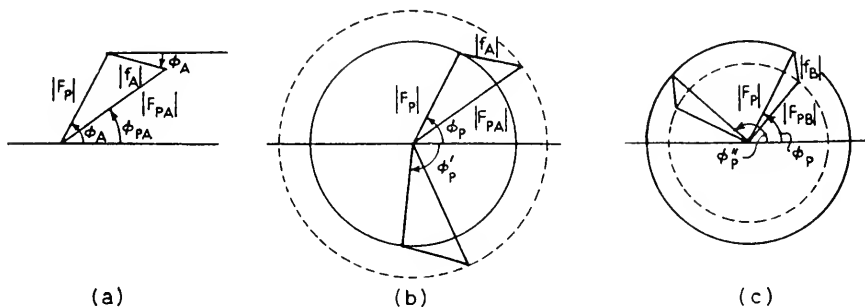


Fig. 5. The vector combination of protein and heavy-atom contributions to a reflexion; for explanation see text.

It is evident that there are in general two solutions which attribute different values ϕ_P and ϕ'_P to the phase of the protein reflexion. This ambiguity cannot be resolved without further information, but as was shown by Bokhoven, Schoone and Bijvoet⁸ in a different connexion, a decision can be reached if two isomorphous replacements can be made available in the same unit cell. A pair of vector diagrams can now be drawn for every reflexion, one for each of the isomorphous replacements, and each gives two alternative values of the protein phase angle. If all is well one value of the angle from the first isomorphous derivative should be identical with one value from the second derivative, and this must be the correct phase angle (see Figs. 5 (b) and (c)). Proceeding in this way one can, theoretically, determine the phase angles of all the reflexions in the diffraction pattern and then compute a three-dimensional Fourier synthesis which should be a correct representation of the electron density throughout the unit cell.

It is for these reasons that we have thought it worth while to develop a number of different methods of attaching heavy groups to various sites on the myoglobin molecule, even though for two-dimensional work a single isomorphous derivative is sufficient to establish nearly all the signs of the real reflexions. Among the complexes illustrated in Fig. 4, we have found that those with HgI_4^{2-} , with AuCl_4^- , with PCMBS, and with $\text{Hg}(\text{NH}_3)^{2+}$, are satisfactory for general phase determination. In addition we have been able to prepare crystals containing two or even three different heavy groups simultaneously, e.g. Au and PCMBS, Au and $\text{Hg}(\text{NH}_3)^{2+}$, PCMBS and

$\text{Hg}(\text{NH}_3)^{2+}$ and finally the triple complex containing Au, PCMBs and $\text{Hg}(\text{NH}_3)^{2+}$ (see for example Fig. 6). Theoretically two separate replacements would suffice, but in practice it is a great advantage to use three or four, since experimental errors are considerable, and for any particular reflexion the contributions of one or more of the heavy atoms may happen to be so small as to preclude their being used for phase determination.

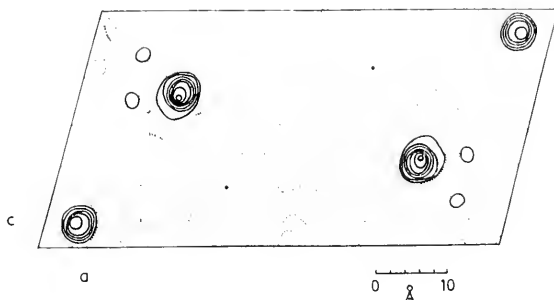


Fig. 6. Difference-Fourier projection of double heavy-atom derivative of Type A myoglobin along y , containing one mole each of PCMBs and $\text{Hg}(\text{NH}_3)^{2+}$ per mole of protein.

In addition, it is a matter of some crystallographic interest that with certain of the complexes a failure of Friedel's Law is observed, in other words, it is found that with these crystals the intensities of two so-called 'equivalent reflexions', 020 and $0\bar{2}0$ for example, are different. It can be shown that this effect will be pronounced, if the wavelength of the X-rays used is in the neighbourhood of that at which one or more of the atoms of the structure have an X-ray absorption band. Pepinsky and Okaya⁹ have shown how it can be used to resolve the ambiguity of phase determination with a single isomorphous replacement; in protein crystals the anomalous effect is not large enough for this to be done, but it can be a valuable confirmation of the phase determined by other methods. This anomalous dispersion effect has also been used by Blow in his work on the x projection of haemoglobin.

In our phase determinations with myoglobin we have so far confined our attention to reflexions of spacing greater than 6 \AA , the intention being to compute in the first instance a three-dimensional synthesis with this resolution. The number of reflexions involved is rather more than 400, of which about one-quarter are real, with signs already determined in the two-dimensional work. We are at the present time calculating the phases of the remaining 300 reflexions. This task is not yet completed, but as a first instalment of the results, a Fourier projection along the z axis of the crystal is shown in Fig. 7; this is computed from 50 terms of which 40 have general phase. This projection is, of course, subject to the same confusions due to overlap as is the y projection illustrated in Fig. 2; it has been reproduced mainly to demonstrate the practicability of general phase determination in a protein crystal. An analogous projection of haemoglobin has recently been calculated by Blow.

Experience gained in preparing this projection suggests that we shall in most cases be able to determine phase angles within $\pm 15\text{-}20^\circ$. We think there is a good chance that a 6 Å. three-dimensional Fourier synthesis calculated using such phases will reveal some recognizable features of the structure. In any case such a synthesis is but a step towards the goal of resolving individual atoms. We see no reason why the methods now being

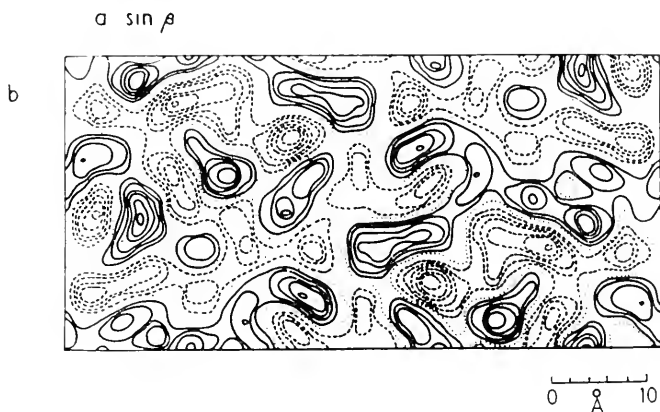


Fig. 7. Fourier projection of Type A myoglobin crystals along z , including all terms of spacing greater than 6 Å

used, or closely related ones, cannot be used at higher resolutions, although the amount of work involved will in any case be considerable. Indeed, to resolve individual atoms would mean measuring at least all the reflexions with spacings exceeding 1.5 Å., that is to say 25,000 reflexions per crystal. It would be premature to guess how far it may prove possible to carry the analysis, although it seems certain that some extension beyond a resolution of 6 Å. will be practicable.

REFERENCES

1. D. W. GREEN, V. M. INGRAM and M. F. PERUTZ, *Proc. Roy. Soc.*, **A 225**, 287 (1954).
2. L. BRAGG and M. F. PERUTZ, *Proc. Roy. Soc.*, **A 225**, 315 (1954).
3. D. W. GREEN and A. C. T. NORTH. Unpublished data.
4. G. BODO, H. M. DINTZIS and J. C. KENDREW. Unpublished data.
5. H. SCOULOUDI and J. C. KENDREW. Unpublished data.
6. D. HARKER, in *Biological and Medical Physics*, Vol. IV, Ed. J. H. Lawrence and C. A. Tobias, Academic Press, New York 1956.
7. C. H. CARLISLE and J. D. BERNAL. Unpublished data.
8. C. BOKHOVEN, J. C. SCHOONE and J. M. BIJVOET, *Acta Cryst.*, **4**, 275 (1951).
9. R. PEPINSKY and Y. OKAYA, *Proc. nat. Acad. Sci.*, Wash., **42**, 286 (1956).

X-ray analysis of haemoglobin

M. F. PERUTZ

Medical Research Council Unit for Molecular Biology, Cavendish Laboratory,
University of Cambridge

When Pauling and Corey first noted the correspondence between the three-dimensional vector structure of haemoglobin and the vector distribution to be expected from the α -helix,¹ and I subsequently found a weak, but detectable, 1.5 Å reflexion in haemoglobin,² I believed that a solution of the haemoglobin structure in terms of parallel polypeptide chains might be found by trial. This led to an attempt by Bragg, Howells and myself³ to calculate the electron density distribution viewed along the suspected chain direction of the molecule, which coincided with the a axis of the haemoglobin crystal. A Fourier projection on the a plane, corresponding to an end-on view of α -helices, was indeed obtained, but its validity rested on the assumption that the interpretation of the structure in terms of parallel α -helices was correct. When the absolute heights of the electron density peaks were measured, they were found to amount to only one-third to a half of the heights to be expected from the initial assumptions, which showed that the model was oversimplified. Crick arrived at the same conclusion by a different argument,⁴ and also pointed out that α -helices are likely to pack together at an angle of about 20°, which speaks against their being parallel in haemoglobin.⁵

The failure of this attempt at analysis by trial pointed to the need for a different approach, and one that was free from any preconceived notion about the structure of the molecule. In practice this meant that the phases of the diffracted rays would have to be determined by direct methods. In a series of papers Bragg, myself and our collaborators have now described four different methods of phase determination. The most useful of these was based on the method of isomorphous substitution described by Dr Kendrew (p. 125). Green, Ingram and Perutz⁶ crystallized compounds of horse haemoglobin in which the sulphhydryl groups were blocked by p -chloro-mercuribenzoate or by silver ions. These compounds proved to be isomorphous with normal horse methaemoglobin. From the changes in the intensities of the X-ray diffraction pattern produced by the presence of the heavy metals, Green, Ingram and Perutz⁷ were able to determine the signs of 87 out of the 94 $h0l$ reflexions covered by their X-ray photograph. From this and other, supplementary, information Bragg and Perutz then calculated

the first maps of the electron density distribution in the haemoglobin molecule in projection along the dyad, or *b* axis.⁷ Two maps were calculated; one showing a projected picture of the volume of hydrated protein into which salt cannot penetrate, from which the external shape of the molecule could be deduced. The other map showed a projected picture of the electron density distribution within a row of haemoglobin molecules suspended in water. Both projections were along the *b* axis of the crystal, i.e. at right angles to *a*, the suspected chain direction along which Bragg, Howells and myself had attempted to calculate the first projected picture of the molecule. The *b* projection is centro-symmetric so that only the signs of the X-ray reflexions have to be determined, whereas the *a* projection is not and the phase angle may have any value.

The maps give the external shape of the molecule as an ellipsoid of $55 \times 55 \times 70$ Å, with a dimple at the centre, or possibly two dimples at the centre on either side of the molecule. They also show that the four free sulphhydryl groups of native haemoglobin⁸ are arranged in two close pairs spaced 30 Å apart and related by the dyad axis of symmetry. Unpublished work on two projections of ox haemoglobin since done by D. W. Green and A. C. T. North suggests that these SH-groups are situated at about 7 Å from the surface of the molecule. Apart from the sulphhydryl groups, which can be located by the heavy atoms attached to them, the projection shows a system of peaks and depressions which has so far defied interpretation and is far too complex to fit into any simple structural theory. One obstacle to interpretation is lack of resolution; this is so low that single atoms are smeared out to a diameter of 9 Å. The main obstacle, however, is the great thickness of matter, corresponding to about 30 atoms, which is projected on to one plane, thus obscuring the individual parts of the molecule by superposition.

At the next stage of the analysis Ann F. Cullis, H. M. Dintzis and myself have now tried both to improve the resolution of the projection along the *b* axis and to prepare some more heavy atom complexes in preparation for a three-dimensional analysis of the structure. At the same time D. M. Blow has made a renewed attempt to solve the projection along the *a* axis, this time making no assumptions about the structure, but relying entirely on isomorphous substitution to determine the phase angles.

An additional pair of sites combining with heavy atoms was found by blocking the sulphhydryl groups with iodoacetamide and then allowing the haemoglobin to react with two equivalents of mercuridiacetate. The positions of these new mercury-combining sites in relation to the sulphhydryl groups are shown in Fig. 1. Their chemical character is still unknown. For calculating the improved Fourier projection along the *b* axis we determined the signs of 420 reflexions instead of the 94 used in the first round, and thus attained a resolution of 2.8 Å instead of the original 6.5. The new map, shown in Fig. 2, contains far more detail than the earlier one, but still fails

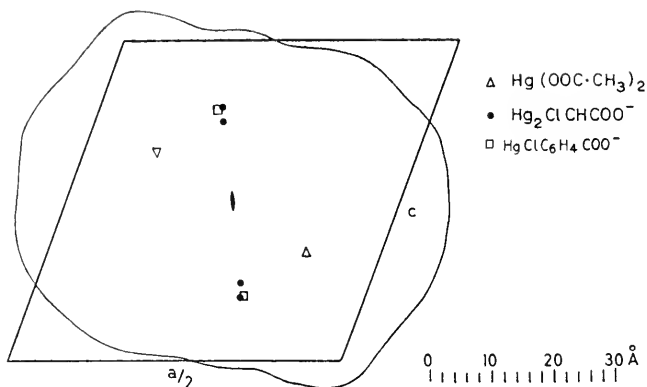


Fig. 1. Positions of mercury atoms in different haemoglobin derivatives. The *p*-chloro-mercuribenzoate and dimercuriacetate groups are attached to the sulphhydryl groups. The mercuri di-acetate groups are attached to a site whose chemical character is not known.

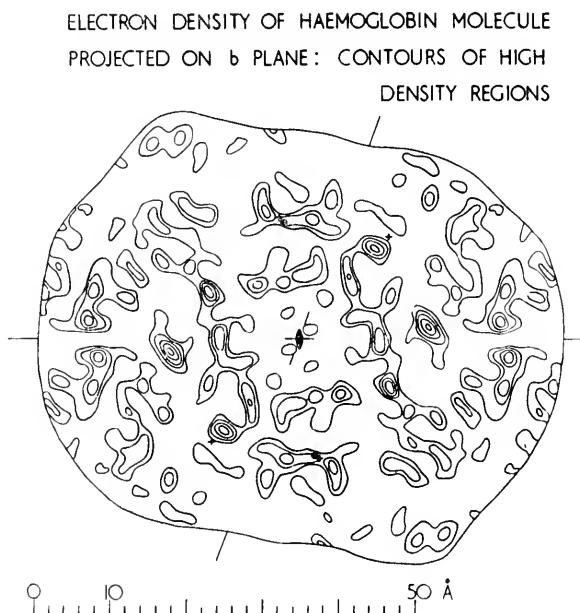


Fig. 2. Electron density distribution in the haemoglobin molecule seen in projection along the dyad axis at a resolution of 2.8 Å. The black dots represent sulphhydryl groups. Contours are drawn at intervals of 2.5 electrons/Å² above the level of the suspension medium, which is 2M ammonium sulphate solution. For the sake of clarity the three lowest contours have been omitted, and only the peaks are shown. The contours near the periphery of the molecule are not exact, being partly modified by the overlapping of neighbouring molecules in the crystal lattice.

to exhibit any features which could be identified with either polypeptide chains or haem groups.

Clearly it is still the great thickness of matter projected on one plane which defeats us, and it looks as though only a three-dimensional analysis, giving the electron density distribution along a series of parallel sections through the molecule, will finally reveal its structure. This will have to be the next step.

Dr Kendrew has already explained to you that this involves determination of the phase angles of each one of many thousands of X-ray reflexions, which is far more difficult than the determination of signs in the projection along the two-fold axis to which our work had been confined so far. It seemed especially doubtful if the phase angles could be measured with sufficient accuracy in a molecule as big as haemoglobin, where the number of sites combining with heavy atoms per unit molecular weight of protein is only half as great as in myoglobin. The projection along the a axis, which lacks a centre of symmetry, was therefore chosen as a test case for a later three-dimensional analysis. Its successful solution by D. M. Blow is most encouraging, though the accuracy of the phase angles so far obtained could be much improved if one or two more sites combining with heavy atoms could be discovered.

In conclusion I should like to mention one interesting piece of information about the structure of the haemoglobin molecule obtained by D. J. E. Ingram, J. F. Gibson and myself.⁹ We determined the orientation of the four haem groups in the haemoglobin molecule by electron spin resonance, using a single crystal of horse haemoglobin. We found the haem groups to be arranged in two pairs, related by the dyad axis of symmetry. The normals to one pair of haem groups lie in the crystallographic a, b plane at an angle of 32° on either side of the a axis. The normals to the other pair have a similar orientation, except that they are tilted by 13° above and below the a, b plane. Knowing the orientation of the haem groups tells us nothing as yet about their positions, but it will make it much easier to find them by X-ray analysis later on.

REFERENCES

1. L. PAULING and R. B. COREY, *Proc. Nat. Acad. Sci. U.S.*, **37**, 282 (1951).
2. M. F. PERUTZ, *Nature*, **167**, 1053 (1951).
3. W. L. BRAGG, E. R. HOWELLS and M. F. PERUTZ, *Acta Cryst.*, **5**, 136 (1952).
4. F. H. C. CRICK, *Acta Cryst.*, **5**, 38 (1952) and **6**, 600 (1953).
5. F. H. C. CRICK, *Acta Cryst.*, **6**, 685 and 689 (1953).
6. D. W. GREEN, V. M. INGRAM and M. F. PERUTZ. *Proc. Roy. Soc. (LONDON)*, **A 225**, 287 (1954).
7. SIR LAWRENCE BRAGG and M. F. PERUTZ, *Proc. Roy. Soc. (LONDON)*, **A 225**, 315 (1954).
8. V. M. INGRAM, *Biochem. J.*, **59**, 653 (1955).
9. D. J. E. INGRAM, J. F. GIBSON and M. F. PERUTZ, *Nature (LONDON)*, **178**, 905 (1956).

Further study on myoglobin*

I. Heterogeneity of human myoglobin

Some properties of Mb I and Mb II

A. ROSSI-FANELLI AND E. ANTONINI

Institute of Biological Chemistry, University of Rome

In connection with the problem of heterogeneity of crystalline proteins, we should like to refer to our work on human crystalline myoglobin (Mb). We have demonstrated that at least three components (Mb I, Mb II, Mb III), differing from one another in their electrophoretic behaviour, are present in preparations of human crystalline Mb.¹ The relative percentages of these components are as follows: Mb I, 74%; Mb II, 19%; Mb III, 7%. It has been possible, by means of preparative zone electrophoresis (on paper and on starch gel²), to isolate Mb I and Mb II and to study some of their properties.

As shown in Fig. 1, the isolated components proved to be homogeneous when subjected again to paper electrophoresis. Spectrophotometric analysis of the various derivatives of Mb I and Mb II (Mb, MbO₂, MbCO, Mb⁺) showed that the two components have absorption curves almost identical with each other both in the visible and Soret region and which correspond with those obtained for unfractionated crystalline human Mb.

A detailed study of the reversible combination with O₂³ revealed that Mb I and Mb II are almost identical as regards oxygen equilibrium, either with respect to the shape of the dissociation curve (Figs. 2 and 3) and O₂ affinity (Fig. 4) at different temperatures. These results seem interesting, particularly as far as the minor component (Mb II) is concerned; this pigment has all the properties of a true myoglobin: absorption spectra, reversible combination with O₂, shape of the dissociation curve and affinity for O₂, so that we must definitely give up the idea that we are dealing with an impurity or an artefact.

We must reject also the hypothesis that Mb II is a residue of foetal Mb, since experiments we have carried out demonstrate that foetal Mb cannot be separated by electrophoresis from adult Mb, and that preparations of foetal Mb show the presence of several components which have electrophoretic characteristics similar to those found in adult Mb.

* Aided by a grant from the Rockefeller Foundation.

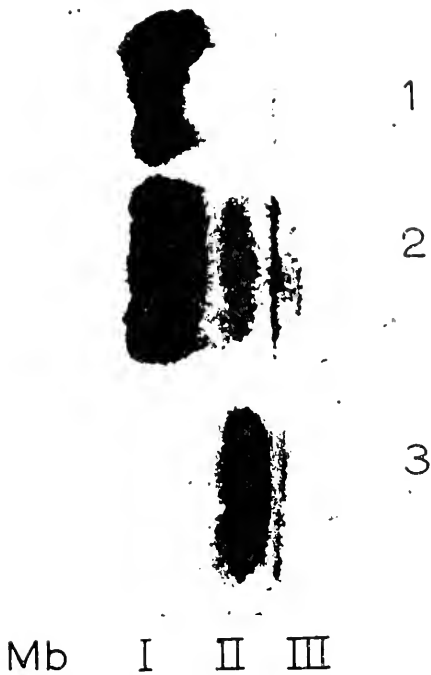
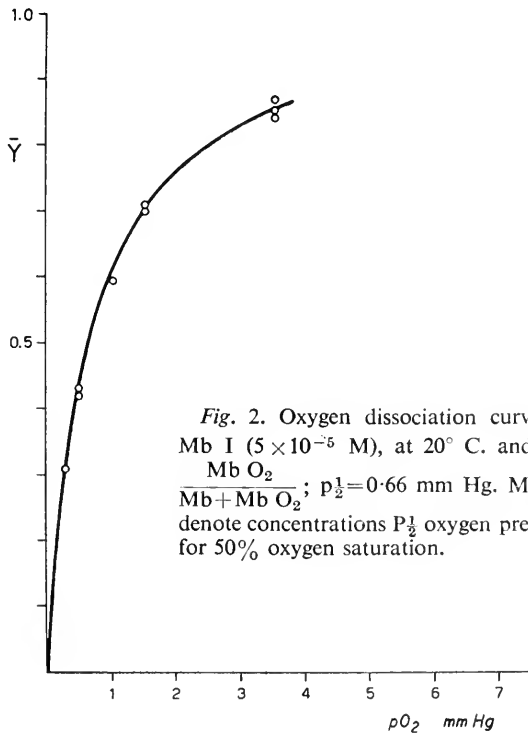


Fig. 1. Paper electrophoresis of homogeneous Mb I,¹ Mb II,³ and non-homogeneous crystalline myoglobin.²



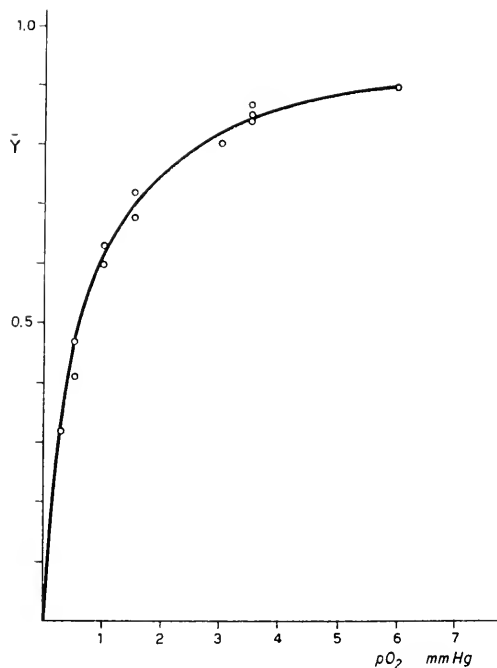


Fig. 3. Oxygen dissociation curve of human Mb II ($4 \times 10^{-5}M$), at $20^{\circ}C$ and pH 8; $\bar{Y} = \frac{MbO_2}{Mb + MbO_2}$; $p_{\frac{1}{2}} = 0.65$ mm Hg.

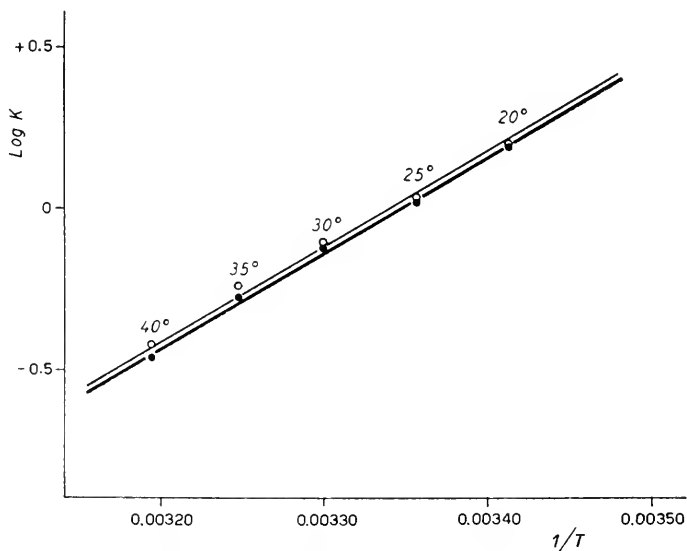


Fig. 4. Influence of temperature on oxygen affinity of human Mb I ● and Mb II ○; pH 8, borate buffer $I = 0.05$. $K = \frac{MbO_2}{Mb \times pO_2}$. Mb I $\Delta H = -13.5$ kcal; Mb II $\Delta H = -13.3$ kcal (calculated from the slopes of the lines.)

From a practical point of view, the above results lead to some limitations in the use of the crystalline pigment in future research on myoglobin. Studies, also with spectrophotometric methods, on the equilibrium between human Mb and O₂, can be properly based on non-homogeneous crystalline pigment, whilst in any research on the structure of myoglobin homogeneous preparations must be used.

Research still in progress in our Laboratory leads us to think that human Mb I and Mb II have different chemical structures.

REFERENCES

1. A. ROSSI-FANELLI and E. ANTONINI, *Arch. Biochem. and Biophys.*, **65**, 587 (1956).
2. O. SMITHIES, *Biochem. J.*, **61**, 629 (1955).
3. A. ROSSI-FANELLI and E. ANTONINI. In press. *Experientia*.

II. Chemical and Biochemical properties of a new type of myoglobin in molluscs

A. ROSSI-FANELLI, E. ANTONINI,
D. POVOLEDO

Institute of Biological Chemistry, University of Rome
and Zoological Station, Naples

The object of this paper is the study of a new type of myoglobin (Mb) isolated and crystallized from the red buccal muscles of two common Mediterranean mollusks: *Aplysia Depilans* and *Aplysia Limacina*.^{*} The purification of this pigment has been previously described.^{1,2} The homogeneity of crystallized *Aplysia* Mb was demonstrated by paper and boundary electrophoresis and by the ultracentrifuge (Fig. 1). The isoelectric point (determined by paper electrophoresis owing to the small quantity of material available) was found to be near pH 4.5 in acetate buffer I=0.05.

The sedimentation constant and the diffusion rate of the pigment were determined by using the Spinco Ultracentrifuge and the Perkin Elmer electrophoresis apparatus.^{3,4} The sedimentation constant S_{20}° was found to be 2.06×10^{-13} ; this is very similar to the figures obtained for myoglobin of different animal species by other authors.^{5,6}

From these data it was possible to establish the molecular weight as approximately 20,000; the minimum molecular weight deduced from the haem content (0.30% iron) is 18,000. These values are comparable with those recently found for pure and homogeneous horse Mb by Theorell and Åkeson.⁶ The accurate spectrophotometric analysis of some derivatives of *Aplysia* Mb (Mb⁺, MbOH, MbO₂, MbCO, Mb) showed that both visible and ultra-violet spectra of this pigment differ widely from those of mammalian Mb (see Table 2).

The reversible combination with O₂ and CO was studied with the spectrophotometric method described elsewhere.⁷ It can be seen from Fig. 2 that the O₂ dissociation curve of *Aplysia* Mb, like the mammalian Mb, is hyperbolic, but the oxygen affinity is lower: $P_{\frac{1}{2}}$ at 20° = 2.7 mm. Hg. The Bohr effect is almost absent ($d \log P_{\frac{1}{2}} / (d \text{pH})_T = 0.02$). The overall heat for the reaction $\text{Mb} + \text{O}_2 \rightarrow \text{MbO}_2$ calculated from the van't Hoff equation between 10° and 30° is -13.6 kcal.

* All the properties analysed showed no significant difference between Mb obtained from *Aplysia Depilans* and *Aplysia Limacina*.



Fig. 1. Sedimentation patterns of *Aplysia Mb* in the Spinco Ultracentrifuge at 59,780 r.p.m. Pictures taken at 8 min. intervals. First 111 min. after reaching full speed.

The values of K for the reaction $\text{MbO}_2 + \text{CO} \rightleftharpoons \text{MbCO} + \text{O}_2$ was found to be 106 at 20° and pH 7.4 and that of y ($y = \frac{\log K}{\text{span}}$) 0.028. This figure is remarkably different from that obtained by Theorell for horse Mb⁸ and by us for human Mb ($y=0.048$).

The difference in the properties of *Aplysia* and mammalian Mb could arise from a substantial difference in the chemical structure of the two proteins. The prosthetic group of *Aplysia* Mb is identical with that of horse

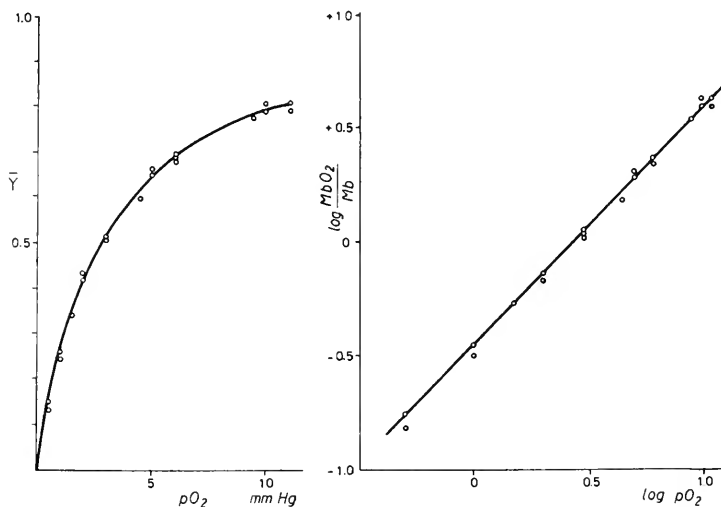


Fig. 2. Oxygen dissociation curve of *Aplysia* Mb ($\sim 3 \times 10^{-4} \text{M}$) at 20°C and pH 7.2 $\bar{Y} = \frac{\text{MbO}_2}{\text{Mb} + \text{MbO}_2}$

Mb: this was demonstrated by spectrophotometry of pyridine haemochromogen and by resynthesis of each pigment from its own globin and the haem split from the other. We then analysed the amino acid composition of crystallized Mb from *Aplysia Depilans* and *Aplysia Limacina*, by ion exchange chromatography according to Moore and Stein.⁹ The results obtained are reported in Table 1, which includes for comparison the amino acid composition of a mammalian (human) Mb.¹⁰ Table 1 shows a noticeable difference in the amino acid content between *Aplysia* and human Mb. The former is richer in aspartic acid, serine, valine, phenyl alanine, arginine, and poorer in threonine, glutamic acid, proline, tyrosine, histidine and lysine. *Aplysia* Mb is characterized by a higher content of dicarboxylic amino acids and by a small content of basic amino acids. Histidine is present in such a quantity as to account for only one mole per mole of protein. This result seems very interesting in connection with the role played by imidazole as haem-linked group.

The most important chemical and physico-chemical properties of Aplysia Mb compared with mammalian Mb, are summarized in Table 2. The data show that the pigment isolated from Aplysia is very similar to mammalian myoglobins with respect to molecular weight, haem content, the shape of the oxygen dissociation curve and the absence of the Bohr effect. On the other hand, Aplysia Mb shows some characteristic features, namely absorption spectra, 'span', relative affinity for CO and O₂, isoelectric point and chemical nature of globin, that strongly differentiate this pigment from all the known myoglobins.

Table 1

AMINO ACID COMPOSITION OF APLYSIA
DEPILANS MYOGLOBIN AND HUMAN
MYOGLOBIN

(Average values of three determinations)
Grams amino acid per 100 g. of protein

	Aplysia Mb	Human Mb
Aspartic acid	12.71	8.27
Threonine	1.70	2.85
Serine	8.29	4.43
Glutamic acid	9.58	16.17
Proline	0.98	5.40
Glycine	4.60	6.08
Alanine	11.66	5.82
Cystine	0.00	0.00
Valine	6.00	4.64
Methionine	1.88	2.69
Isoleucine	4.89	5.27
Leucine	8.73	13.67
Tyrosine	1.48	2.19
Phenylalanine	11.75	8.22
Histidine	1.22	7.79
Lysine	6.89	19.09
Arginine	4.57	2.47

Table 2

CHEMICAL AND PHYSICO-CHEMICAL CONSTANTS OF APLYSIA MYOGLOBIN AND MAMMALIAN MYOGLOBIN

	Aplysia Mb (Depilans and Limacina)	Mammalian Mb (Horse)
Prosthetic group	Ferroprotoporphyrin IX	Ferroprotoporphyrin IX
Molecular weight	18,000-20,000	17,000-19,000
S_{20}°	2.06×10^{-13}	1.97×10^{-13}
Haem/mole	1	1
Shape of the oxygen dissociation curve	hyperbolic	hyperbolic
$P_{\frac{1}{2}}$ (20° pH 7.4)	2.7 mm Hg	0.7 mm Hg
Bohr effect	absent	absent
'Span'	72 Å	30 Å
$K = \frac{[MbCO] \times pO_2}{[MbO_2] \times pCO}$	106	28
$y = \frac{\log K}{\text{span}}$	0.028	0.045
Isoelectric point	~4.5	6.78
Chemical nature of globin	acidic	basic
Auto-oxidation	fairly rapid	fairly rapid
Absorption spectra*		
Mb ⁺ max (mμ)	400 505 640	408 505 630
$\epsilon \times 10^{-3}$	99 13.1 3.8	160 10.2 4.2
MbOH max (mμ)	412 543 580 600	414 540 580
$\epsilon \times 10^{-3}$	91 9 8.7 8.7	90 9.1 8.6
MbCO max (mμ)	424 541 571	423 540 578
$\epsilon \times 10^{-3}$	176 14.2 13.7	178 14 11.8
Mb max (mμ)	438 555	434 555
$\epsilon \times 10^{-3}$	113 13	113
MbO ₂ max (mμ)	416 542 578	
$\epsilon \times 10^{-3}$	108 13.2 13.3	

* Data for horse Mb obtained by the Authors in the present work.

REFERENCES

1. A. ROSSI-FANELLI and E. ANTONINI, Bach commemorative number, *Biochimica (URSS)*, **22**, 336 (1957).
2. A. ROSSI-FANELLI and E. ANTONINI, *Rend. Acc. Naz. Lincei*, **XXII**, 133 (1957).
3. T. SVEDBERG and K. O. PETERSEN, *The ultracentrifuge*. Oxford Press (1940).
4. L. G. LONGSWORTH, *N. Y. Ann. Acad. Sci.*, **41**, 261 (1941).
5. E. N. WYMAN and R. INGALLS, *J. Biol. Chem.*, **153**, 275 (1944).
6. H. THEORELL and Å. ÅKESON, *Annales Academiae Scientiarum Fennicae*, **60**, 303 (1955).
7. A. ROSSI-FANELLI and E. ANTONINI. In press.
8. H. THEORELL, *Biochem. Z.*, **268**, 64 (1934).
9. S. MOORE and W. H. STEIN, *J. Biol. Chem.*, **192**, 663 (1951).
10. A. ROSSI-FANELLI, D. CAVALLINI and C. DE MARCO, *Biochim. et Biophys. Acta*, **17**, 377 (1955).

The chemical effects of gene mutations in some abnormal human haemoglobins

J. A. HUNT AND V. M. INGRAM

Medical Research Council Unit for Molecular Biology,
Cavendish Laboratory, University of Cambridge

It has been known for some years that the inherited haemolytic anaemia, known as sickle cell anaemia, is due to an alteration in the structure of the haemoglobin molecule.¹ This change is thought to be caused by a single mutation of the gene responsible for haemoglobin synthesis.² We can now report that the effect of this mutation is to change only one amino acid residue in polypeptide chains containing nearly 300 residues; one of the glutamic acids in normal haemoglobin has been replaced by valine.

Earlier experiments indicated that the two haemoglobins differed electrophoretically,^{1,3} haemoglobin S—the sickle cell anaemia type—having 2-3 carboxyl groups fewer than haemoglobin A—the normal protein. Deoxygenated haemoglobin S has an abnormally low solubility⁴ which causes the sickling of the red blood cells characteristic of this anaemia. However, determinations of amino acid composition,^{5,6} *N*-terminal^{7,8,9} and *C*-terminal amino acids¹⁰ or sulphhydryl groups^{11,12} did not show any significant differences between them. More recently,¹³ degradation of the two haemoglobins with trypsin showed that the alteration which produces haemoglobin S resides in one very short section of the polypeptide chains. Whilst most of the fragments in the two haemoglobin showed similar behaviour on electrophoresis and chromatography, a pair of peptides, called No. 4, were found in which these properties differed. The complete chemical structure of these two peptides has now been determined by amino acid analyses, end group analyses, some Edman stepwise degradations and the results of partial acid hydrolyses—the last are shown in Fig. 1 (*a*)—which led to the formulation of the two sequences.¹⁴ They each consist of the nine amino acids shown in Fig. 1 (*b*). It can be seen that they differ in only one amino acid; one of the glutamic acids of the normal No. 4 peptide is replaced by valine in the sickle cell No. 4 peptide. To date further investigation of the other peptides which had been obtained by trypsin digestion, has failed to reveal any additional differences. The trypsin resistant cores of the two haemoglobin molecules—about 30% of the molecule—have been digested with

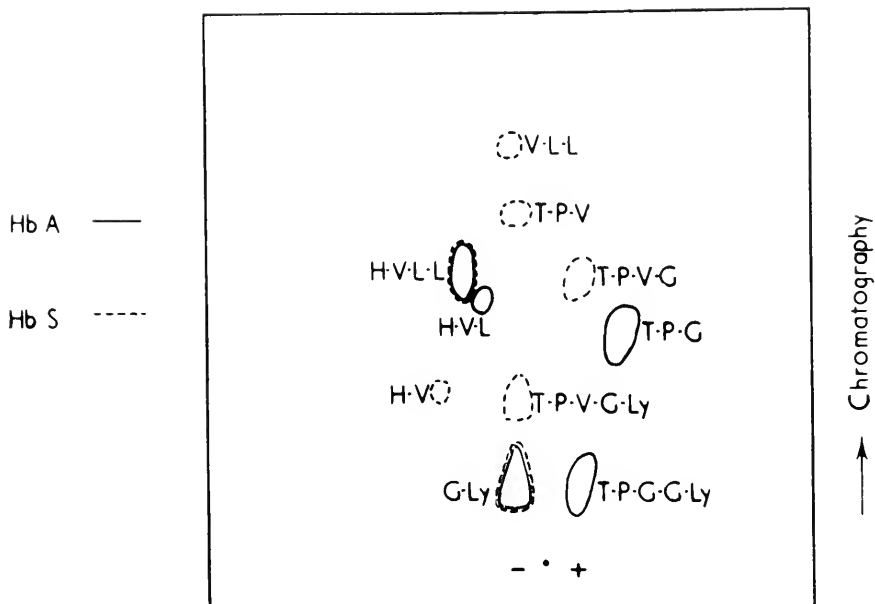


Fig. 1 (a). Acid Degradation of the No. 4 Peptides from Haemoglobins A and S.

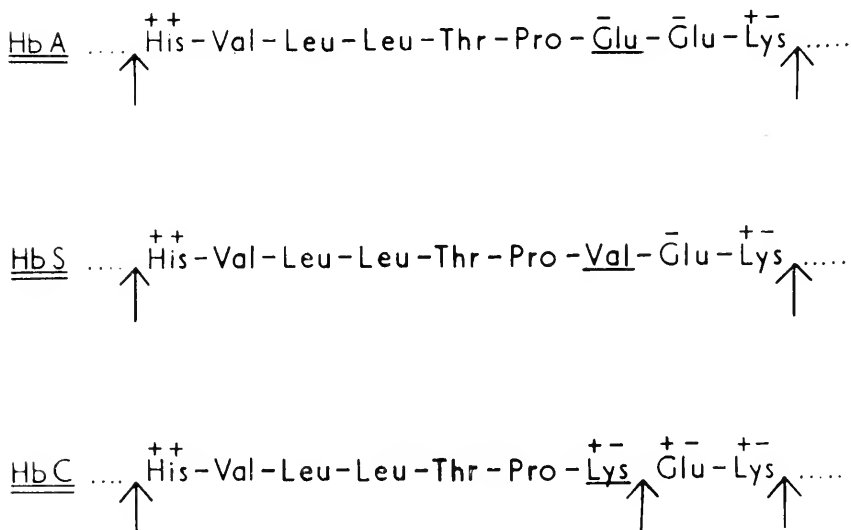


Fig. 1 (b). Structure of the No. 4 Peptides from Haemoglobins A, S and C.

(H=His=histidine, V=Val=valine, L=Leu=leucine, T=Thr=threonine, P=Pro=proline, G=Glu=glutamic acid, Ly=Lys=lysine.)

chymotrypsin and subjected to examination by fingerprinting¹³ and chromatography; no differences could be detected.

This finding is in agreement with such differences between the two haemoglobins as had already been established, when it is remembered that haemoglobin (mol. wt. 66,700) has two identical half-molecules,¹⁵ and that therefore the No. 4 peptide occurs twice in the whole molecule. It still remains to relate this change of amino acids to the solubility differences and to the causation of the haemolytic anaemia. However, it is the kind of change which one would expect.

Human haemoglobin S is a clear case where the mutation of a nuclear gene is inherited in a strictly Mendelian manner. Furthermore, it is well established that this mutation alters the nature of the protein haemoglobin. We can now show for the first time that the effect of such a mutation is a chemical one and that it leads in this case at least to a substitution of only one amino acid in the protein by another one. One may speculate that this very small change in the polypeptide chain of the protein reflects a very small change in the DNA chain of the gene, affecting perhaps not more than one of the nucleotide base pairs of that chain.

Another mutation of the same haemoglobin gene produces another abnormal protein, haemoglobin C.^{16,17} This causes a severe anaemia only when it occurs together with haemoglobin S. The solubility of haemoglobin C is almost normal,^{16,17} and so is its amino acid composition,¹⁸ however, it is easily distinguished electrophoretically, since it has even fewer net negative charges per molecule than haemoglobin S which itself has two fewer than the haemoglobin A molecule.

Haemoglobin C has been degraded with trypsin and, as a second stage, with chymotrypsin, using the methods described for haemoglobin S. The mixtures of peptides obtained were compared with those from haemoglobin A by the 'fingerprinting'¹³ and other methods employed for haemoglobin S. Again, it was found that all peptides behaved identically except the No. 4 peptide in tryptic haemoglobin A digests whose glutamic acid had changed to valine in haemoglobin S. This same peptide was not found in tryptic digests of haemoglobin C. Its place was taken by two new peptide spots, one neutral and the other positively charged at pH 6.4.¹³ Using the methods referred to above, the amino acid sequence of these two peptides has been determined and is shown in Fig. 1 (*b*). Clearly the same amino acid, glutamic acid, of normal haemoglobin has changed also in the haemoglobin C mutation; its place is taken by lysine. In each half molecule a negatively charged side chain has been replaced by a positive group altering the net charge of the protein by four units, twice as many as for haemoglobin S. This interpretation agrees very well with the known electrophoretic series of the haemoglobins which is A—S—C. The additional peptide found in tryptic digests of haemoglobin C is due to the introduction of an additional lysine peptide bond which is hydrolysed by the enzyme.

Genetic evidence shows that the haemoglobin S and the haemoglobin C mutations may be allelic,¹⁷ i.e. that they occur in the same locus. We can now give a chemical interpretation of this statement, not in terms of the chemical constitution of the gene but of the amino acid sequence of the polypeptide chain whose synthesis the gene controls. In the two mutations discussed in this paper the same very small portion of the haemoglobin gene seems to be affected and in turn the identical amino acid in the polypeptide chain is replaced by either valine or lysine in haemoglobin S and C respectively. The two mutations are indeed alleles.

REFERENCES

1. L. PAULING, H. A. ITANO, S. J. SINGER and I. C. WELLS, *Science*, **110**, 543 (1949).
2. J. V. NEEL, *Science*, **110**, 64 (1949).
3. I. H. SCHEINBERG, R. S. HARRIS and J. L. SPITZER, *Proc. U.S. Nat. Acad. Sci.*, **40**, 777 (1954).
4. M. F. PERUTZ and J. M. MITCHISON, *Nature*, **166**, 677 (1950).
5. W. A. SCHROEDER, L. M. KAY and I. C. WELLS, *J. Biol. Chem.*, **187**, 212 (1950).
6. T. H. J. HUISMAN, J. H. P. JONXIS and P. C. VAN DER SCHAAF, *Nature*, **175**, 902 (1955).
7. E. HAVINGA, *Proc. U.S. Nat. Acad. Sci.*, **39**, 59 (1953).
8. T. H. J. HUISMAN and H. DRINKWAARD, *Biochim. Biophys. Acta*, **18**, 588 (1955).
9. H. BROWN, *Arch. Biochem. Biophys.*, **61**, 241 (1956).
10. T. H. J. HUISMAN and A. DOZY, *Biochim. Biophys. Acta*, **20**, 400 (1956).
11. F. A. HOMMES, J. SANTEMA-DRINKWAARD and T. H. J. HUISMAN, *Biochim. Biophys. Acta*, **20**, 564 (1956).
12. V. M. INGRAM, *Biochem. J.*, **65**, 760 (1957).
13. V. M. INGRAM, *Nature*, **178**, 792 (1956).
14. V. M. INGRAM, *Nature*, **180**, 326 (1957).
15. M. F. PERUTZ, A. M. LIQUORI and F. EIRICH, *Nature*, **167**, 929 (1951).
16. H. A. ITANO, *Ann. Rev. Biochem.*, **25**, 311 (1956).
17. W. W. ZUELZER, J. V. NEEL and A. R. ROBINSON, *Progress in Hematology*, (Grune and Stratton, New York), **1**, 91 (1956).
18. W. H. STEIN, H. G. KUNKEL, R. D. COLE, D. H. SPACKMAN and S. MOORE, *Biochim. Biophys. Acta*, **24**, 640 (1957).

Proteolytic Enzymes

Etude de quelques protéolyses limitées du chymotrypsinogène de bœuf

P. DESNUELLE ET M. ROVERY

Laboratoire de Chimie Biologique, Faculté des Sciences, Marseille, France

L'objet de ce mémoire est de décrire certaines expériences effectuées dans le but de mieux comprendre les relations existant entre la structure chimique du chymotrypsinogène et son caractère de précurseur d'un enzyme protéolytique. Comment cette molécule inactive acquiert elle le pouvoir d'hydrolyser d'autres protéines? Nous savons que le phénomène d'activation est en fait déterminé par une protéolyse limitée.¹

I. ACTIVATION 'RAPIDE' DU CHYMOTRYPSINOGENE

Une protéolyse est dite limitée quand le nombre des liaisons coupées par l'enzyme est très faible. Au lieu d'être comme d'habitude transformée en peptides relativement courts, cette protéine donne alors naissance à une ou à plusieurs autres protéines qui ne diffèrent de la première que par l'ouverture de quelques liaisons et éventuellement la perte de quelques résidus. Affectant la structure covalente et fondamentale de la molécule, ces ruptures doivent certainement provoquer des remaniements profonds de la structure secondaire. Or, dans certains cas tout au moins, les remaniements en question n'entraînent pas ce qu'on a coutume d'appeler la 'dénaturation'. Bien au contraire, certaines protéines engendrées par protéolyse limitée manifestent une activité physiologique précise que la protéine-mère ne possède pas. C'est d'ailleurs pourquoi les protéolyses limitées sont si intéressantes puisqu'elles permettent quelquefois d'expliquer en termes chimiques simples ces activations par 'démasquage' dont le principe et les résultats sont connus depuis de nombreuses années. La protéolyse trypsique qui active le chymotrypsinogène ne peut pas être plus limitée puisqu'elle consiste en la rupture d'une liaison unique, reliant un résidu d'arginine à un résidu d'isoleucine (liaison arginyl-isoleucine 1 de la fig. 1).

Il est d'ailleurs tout à fait vraisemblable que le mécanisme de l'activation soit nettement plus compliqué que celui indiqué dans la fig. 1. La variation sensible du pouvoir rotatoire au moment de l'activation² signifie sans doute

que la structure secondaire de la molécule subit quelques remaniements. On peut alors penser que ces remaniements créent le centre actif de l'enzyme, en rapprochant par exemple deux ou plusieurs groupes fonctionnels dont la position initiale dans le zymogène n'est pas favorable. Mais, si l'on s'en tient à la structure covalente de la molécule, la rupture précédemment mentionnée de la liaison arginyl-isoleucine semble être le seul événement déterminant l'activation. La fig. 2 montre d'ailleurs que, tout au long de l'activation, les proportions d'isoleucine N-terminale formées par la rupture de la liaison arginyl-isoleucine sont bien égales au pourcentage d'activation,

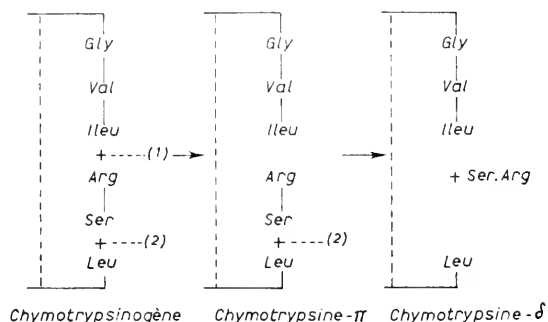


Fig. 1. Schéma de l'activation 'rapide' du chymotrypsinogène. La liaison 1 est coupée par la trypsine. La liaison 2 s'autolyse au sein de la chymotrypsine- π .

c'est-à-dire au nombre de mole de chymotrypsine formées aux dépens de 100 mole de chymotrypsinogène.

La rupture de la liaison arginyl-isoleucine convertit le zymogène en une première chymotrypsine appelée chymotrypsine- π ,³ laquelle est instable et se convertit rapidement par autolyse en une deuxième chymotrypsine douée de la même activité, la chymotrypsine- δ . Cette autolyse intervient au niveau d'une liaison leucyl-sérine (liaison 2) et elle provoque la libération d'un dipeptide, la sérylarginine. La libération de ce peptide n'a d'ailleurs aucune signification particulière. Elle a lieu parce que les liaisons 1 et 2 sont très proches l'une de l'autre. Si ces deux liaisons étaient séparées par un pont disulfure, aucun peptide ne serait libéré. La molécule posséderait simplement une chaîne 'ouverte' de plus. Le fait important est que l'ensemble du phénomène soit très simple. Deux enzymes différents, la trypsine et la chymotrypsine agissent l'un après l'autre. Chaque enzyme coupe une seule liaison choisie, pour une raison qui nous échappe encore, parmi toutes celles qui satisfont aussi bien les exigences structurales des enzymes en question. Rappelons ici⁴ que le chymotrypsinogène contient 4 résidus d'arginine et 12 résidus de lysine en principe favorables à l'action trypsique. Il contient aussi 4 résidus de tyrosine et 6 résidus de tryptophane, sans doute plus favorables à l'action chymotrypsique que le résidu de leucine à côté duquel se produit l'autolyse de la chymotrypsine- π . Il est donc évident que les liaisons

1 et 2 sont particulièrement exposées. Des considérations de forme moléculaire, d'enroulement plus ou moins serré des chaînes peptidiques jouent certainement un rôle déterminant dans la spécificité du phénomène.

Le phénomène de l'activation 'rapide' n'est d'ailleurs pas seulement simple. Il est aussi en quelque sorte pur car les chymotrypsines- π et δ peuvent être obtenues avec des rendements dépassant 95%.⁵ Les chaînes peptidiques sont donc coupées dans ce cas par la trypsine et la chymotrypsine un peu comme on coupe une ficelle avec une paire de ciseaux. Dans la fig. 2, la

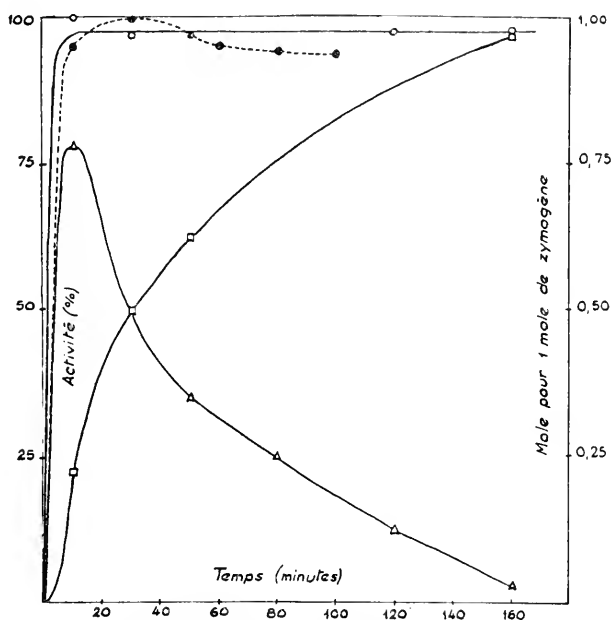


Fig. 2. Cinétique de l'activation 'rapide' du chymotrypsinogène :

● : Activité chymotrypsique (%) vis à vis de l'acétyl-L-tyrosine éthyloester.

○ : Isoleucine N-terminale (mole pour 1 mole de zymogène).

□ : Sérylarginine (mole pour 1 mole de zymogène) ou chymotrypsine- δ (mole pour 1 mole de zymogène).

△ : Chymotrypsine- π (mole pour 1 mole de zymogène).

courbe donnant les quantités de sérylarginine en fonction du temps correspond à la cinétique de la formation de la chymotrypsine- δ . Dans cette même figure, on a tracé la cinétique de la formation de la chymotrypsine- π en retranchant la chymotrypsine- δ des quantités totales de chymotrypsine données par l'isoleucine N-terminale. Les proportions de chymotrypsine- π peuvent devenir beaucoup plus grandes si l'on empêche l'autolyse de cet enzyme par le β -phénylpropionate.⁵

Avant de quitter l'activation 'rapide', notons un fait assez singulier: Il est possible de couper la liaison leucyl-sérine par la chymotrypsine (voir plus loin) sans toucher à la liaison arginyl-isoleucine. Or, la coupure de

cette liaison n'active pas la molécule. Elle respecte simplement dans une certaine mesure son aptitude à être ultérieurement activée.⁶ L'extrême spécificité des mécanismes en jeu pendant l'activation apparaît ainsi clairement. Les remaniements de structure nécessaires à l'activation se produisent quand on ouvre la liaison arginyl-isoleucine. Mais ils ne se produisent pas quand on ouvre une autre liaison séparée de la première par deux résidus seulement.

II. L'ACTIVATION 'LENTE' DU CHYMOTRYPSINOGENE

Malgré leur importance théorique considérable, les chymotrypsines- π et δ ne sont pas encore très connues. La chymotrypsine la plus classique est toujours celle que cristallisèrent Kunitz et Northrop⁷ en 1935. Pour obtenir leur enzyme, les auteurs adoptèrent d'ailleurs des conditions d'activation bien particulières que l'on appelle maintenant les conditions de l'activation 'lente'. Ce mode d'activation a pour caractéristique essentielle d'utiliser des quantités très faibles de trypsine. Il faut donc attendre longtemps (40-48 h. à 5°C) avant que le processus soit terminé. La chymotrypsine- α , qui peut alors être cristallisée avec un rendement d'environ 50%, possède deux résidus terminaux de plus que la chymotrypsine- δ (l'alanine en position N-terminale⁹ et la tyrosine en position C-terminale).¹⁰ Sa formation exige donc certaines protéolyses additionnelles dont il est intéressant de rechercher l'origine. Le tableau I indique les résidus terminaux existant dans le chymotrypsinogène, la chymotrypsine- δ et la chymotrypsine- α . Chaque protéine mentionnée dans le tableau I dérive de la précédente par un processus de protéolyse limitée. Elle contient donc tous les résidus de celle-ci plus les résidus engendrés par la protéolyse qui lui a donné naissance.

Tableau 1

RÉSIDUS TERMINAUX DU CHYMOTRYPSINOGENE ET DES CHYMOTRYPSINES δ ET α

	Résidus N-terminaux	Résidus C-terminaux
Chymotrypsinogène	Cys	Leu
Chymotrypsine- δ	Cys, Ileu	Leu, Leu
Chymotrypsine- α	Cys, Ileu, Ala	Leu, Leu, Tyr

La présence d'un résidu C-terminal de tyrosine dans la chymotrypsine- α suggère que les protéolyses additionnelles de l'activation 'lente' sont dues à la chymotrypsine. La lenteur du phénomène laisse d'ailleurs le temps au chymotrypsinogène d'être attaqué par la chymotrypsine et à la chymotrypsine- δ , produit final de l'activation 'rapide', le temps de s'autolyser. Ces deux phénomènes existent. Ils donnent tous les deux naissance à un résidu de thréonine en position N-terminale.⁶ Ce résidu de thréonine ne fait d'ailleurs qu'une apparition relativement brève. Il cède bientôt la place à un

résidu d'alanine, lequel se trouve justement en position N-terminale dans la chymotrypsine- α (tableau 1). Une telle coïncidence n'est vraisemblablement pas fortuite. Aussi, avons-nous cherché à mieux caractériser les protéines résultant de l'attaque chymotrypsique du chymotrypsinogène. En utilisant les techniques classiques, nous avons réussi à cristalliser deux de ces protéines, l'une contenant la thréonine N-terminale et l'autre, l'alanine. Nous avons également démontré la formation effective de la protéine contenant l'alanine et la sérine (voir fig. 1) N-terminales. Ces protéines n'ont pas été obtenues à l'état pur car la cristallisation ne possède pas ici une efficacité bien considérable.* Mais les protéines en question ont été malgré tout préparées dans un état suffisamment défini pour permettre la réalisation de quelques expériences significatives.

Ces protéines sont inactives. Nous ne nous en étonnerons pas car le chymotrypsinogène n'est pas activé par la chymotrypsine et nous savons déjà que l'on active pas ce zymogène en rompant n'importe quelle liaison. Mais elles sont encore activables par la trypsine et nous les avons pour cette raison appelées 'néochymotrypsinogènes'. Cette activation est identique sur le plan chimique à celle du chymotrypsinogène ordinaire. Elle s'effectue grâce à l'ouverture de la même liaison arginyl-isoleucine. Mais sa vitesse et ses résultats sont bien différents. La fig. 3 permet d'étudier l'apparition de l'activité chymotrypsique quand on active dans les mêmes conditions le chymotrypsinogène (courbe de gauche) et le néochymotrypsinogène à thréonine N-terminale (courbe de droite).

Deux remarques peuvent être faites à propos des courbes de la fig. 3. D'une part, l'activation du néochymotrypsinogène est nettement plus lente que celle du chymotrypsinogène. On sait que la liaison arginyl-isoleucine manifeste au sein du chymotrypsinogène une vulnérabilité vraiment extraordinaire vis-à-vis de la trypsine puisqu'elle est coupée dans toutes les molécules après quelques minutes à 0°C. Cette vulnérabilité est certainement due à un élément particulier de structure que le chymotrypsinogène possède à un très haut degré. L'élément en question est loin de disparaître au moment des ruptures chymotrypsiques. Mais il est atténué de façon sensible. D'autre part, l'activité potentielle du néochymotrypsinogène, c'est à dire l'activité chymotrypsique maximum que peut acquérir cette protéine dans ces conditions déterminées, est plus faible que celle du chymotrypsinogène (2.4 au lieu de 4.0). Les ruptures chymotrypsiques n'empêchent donc pas la formation du centre actif. Mais elles changent vraisemblablement sa structure dans un sens défavorable, en modifiant par exemple les distances entre les divers groupements constituant le centre. En fait, les néochymotrypsinogènes donnent naissance aux chymotrypsines de l'activation 'lente' (chymotrypsines de type α) dont l'activité spécifique est inférieure à celle des chymotrypsines- π et δ . Tout semble ainsi se passer comme si les protéolyses limitées qu'effectue

* La chromatographie ne permet pas non plus, comme nous allons le voir tout à l'heure, de purifier complètement les protéines en question.

la chymotrypsine au sein du chymotrypsinogène ne bouleversent pas suffisamment la structure de la molécule pour lui faire perdre son caractère de zymogène. Mais elles atténuent néanmoins ce caractère à plusieurs points de vue.

La substitution précédemment constatée de la thréonine N-terminale par l'alanine suggère en outre que l'attaque chymotrypsique du chymotrypsinogène comporte à un moment donné la libération de un ou de plusieurs

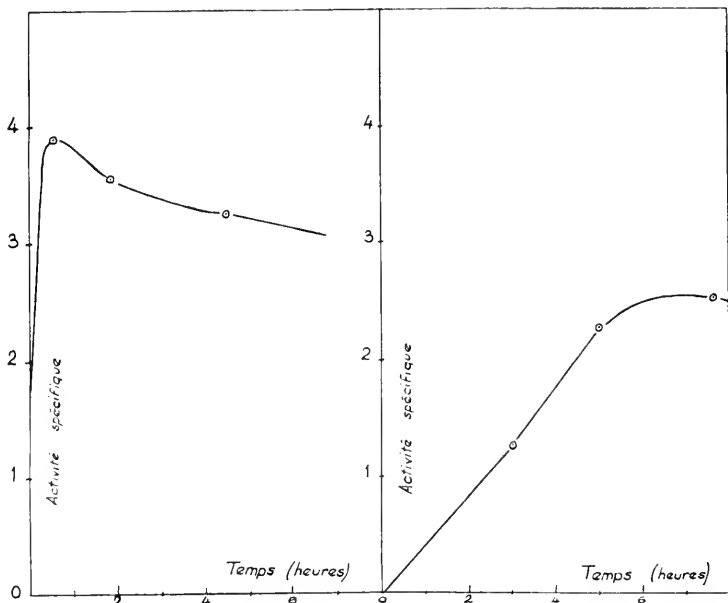


Fig. 3. Activation du chymotrypsinogène et du néochymotrypsinogène à thréonine N-terminale.

Courbe de gauche: activation du chymotrypsinogène.

Courbe de droite: activation du néochymotrypsinogène à thréonine N-terminale.

Ordonnées: activités spécifiques des hydrolysats trypsiques vis à vis de l'acétyl-L-tyrosine éthylester. Conditions de l'activation: zymogène: 5 mg par ml. Rapport pondéral enzyme substrat: 1/40. pH=7.6. Temp.: 0°C.

peptides. La fig. 4 (diagramme du haut) montre que le peptide est unique* et que sa structure est celle de la thréonylasparagine.

L'identification de la thréonylasparagine nous éclaire sur le mécanisme de la genèse des néochymotrypsinogènes. Ce mécanisme est reproduit dans la fig. 5.

Comme on le voit, ce mécanisme est curieusement semblable à celui de l'activation 'rapide' schématisé dans la fig. 1. Il comporte en effet la rupture de deux liaisons et la libération d'un dipeptide. La première liaison rompue

* Une enquête très approfondie a été faite en vue de savoir si la thréonylasparagine est vraiment le seul peptide libéré. Il n'existe ni peptides non-réactifs vis à vis de la ninhydrine, ni peptides contenant de la cystine ou du tryptophane (lesquels, on le sait, migrent mal dans les colonnes de Dowex-50).

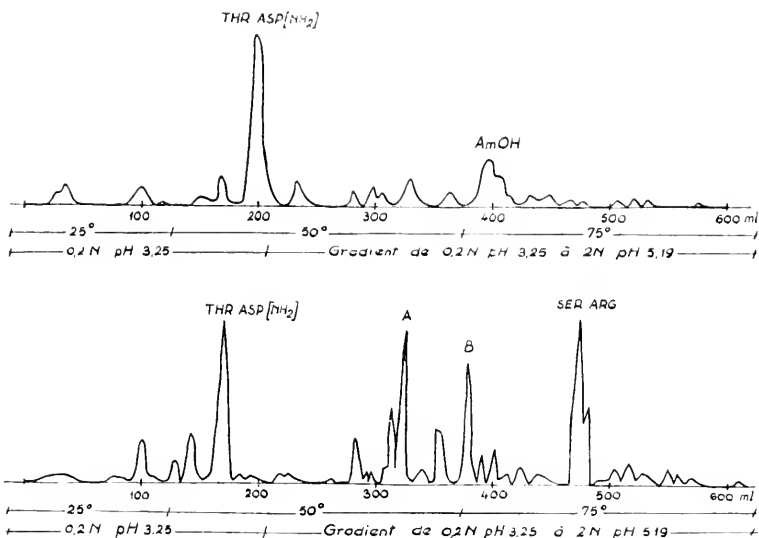


Fig. 4. Peptides engendrés par l'attaque chymotrypsique du chymotrypsinogène et par l'activation 'lente' de ce zymogène (11).

Diagramme du haut: attaque du chymotrypsinogène ($4.4 \times 10^{-3}M$) par la chymotrypsine- α ($0.45 \times 10^{-3}M$) à $0^\circ C$, $pH=7.6$ pendant 46 h.

Diagramme du bas: Activation 'lente' du chymotrypsinogène ($4.4 \times 10^{-3}M$) par la trypsine ($4.8 \times 10^{-7}M$) en présence de SO_4Am_2 $0.3M$ à $6^\circ C$, $pH=7.6$ pendant 46 h.

Technique de Moore et Stein avec gradient de pH et de force ionique. Colonne de 87×0.9 cm. Volume de la chambre de mélange: 330 ml. Débit 8 ml/h. Volume des fractions: 1 ml.

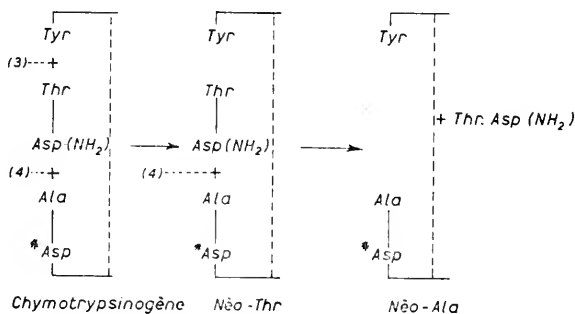


Fig. 5. Schéma des protéolyse chymotrypsiques donnant naissance aux néochymotrypsinogènes. Les protéolyse en question ont été placées dans la partie 'gauche' de la molécule, par opposition à la partie 'droite' où se produisent les protéolyse de l'activation 'rapide' (fig. 1). Cette convention est tout à fait arbitraire. L'essentiel est que les deux groupes de protéolyse se fassent de part et d'autre d'un pont interne matérialisé dans le schéma par un trait en pointillé. Sinon, la molécule se séparerait en plusieurs morceaux.

(liaison 3 : tyrosyl-thréonine) correspond bien à la spécificité de la chymotrypsine. La seconde rupture est un peu plus surprenante puisqu'elle s'effectue à côté d'un résidu d'asparagine pour lequel la chymotrypsine ne semble pas avoir d'affinité particulière. Mais cette rupture est plus lente que la première. Il est en outre intéressant de constater que la présence de sulfate d'ammonium l'accélère beaucoup.* On sait que les ions calcium accélèrent également la protéolyse trypsique d'une liaison lysyl-isoleucine dans le trypsinogène et qu'ils favorisent ainsi l'activation de ce zymogène.¹² L'influence exercée par certains ions métalliques sur les phénomènes de protéolyse est bien connue. Le fait nouveau est que cette influence puisse dans les deux exemples précédents être localisée au niveau d'une liaison déterminée. Une telle spécificité est-elle due à des interactions de l'ion avec le substrat, modifiant sa forme ou attirant l'enzyme vers un point précis? L'ion agit-il au contraire sur l'enzyme en modifiant son affinité et certains de ses propriétés? Des expériences en cours au laboratoire permettront sans doute de répondre bientôt à ces importantes questions.

Le diagramme du bas de la fig. 4 est relatif aux peptides engendrés pendant l'activation 'lente' du chymotrypsinogène. A première vue, ce diagramme semble assez compliqué. Il est en fait aussi simple que le premier car les deux pics centraux A et B correspondent, l'un à l'ammoniaque† et l'autre, à un mélange de plusieurs peptides dont les proportions individuelles sont très faibles. Nous n'avons donc à considérer que deux pics significatifs: le pic de la sérylarginine qui montre que le principe de l'activation 'lente' est le même que celui de l'activation 'rapide' et le pic de la thréonylasparagine, qui montre que les protéolyses additionnelles de l'activation 'lente' sont dues à l'attaque chymotrypsique du chymotrypsinogène ou/et à l'autolyse de la chymotrypsine- δ .

Le caractère mixte du processus de l'activation 'lente' apparaît ainsi en toute clarté. Il comporte la rupture d'une liaison par la trypsine et la rupture de trois liaisons (au maximum) par la chymotrypsine. Quand on coupe successivement quatre liaisons dans un ordre quelconque, $2^4=16$ possibilités sont en principe offertes. La première correspond à l'état initial. Les quinze autres sont trouvées en associant 1 par 1, 2 par 2, 3 par 3 et 4 par 4 les quatre ruptures envisagées. La fig. 6 donne un aperçu de ces quinze possibilités.

Les quinze possibilités mentionnées dans la fig. 6 ne sont d'ailleurs à considérer que si les quatre ruptures interviennent dans un ordre quelconque.

* Il est curieux de noter que Kunitz et Northrop ont obtenu la chymotrypsine- α parce qu'ils ont activé un cake de chymotrypsinogène contenant 50% de sulfate d'ammonium. En activant du chymotrypsinogène dépourvu de sel, ils auraient principalement obtenu une autre chymotrypsine de type α , la chymotrypsine- α_1 , contenant de la thréonine en position N-terminale.

† Comme il vient d'être dit, l'activation 'lente' est réalisée en présence d'une quantité très substantielle de sulfate d'ammonium. Les traitements classiques de dessalage n'arrivent pas à éliminer la totalité des ions ammonium apportés par ce sel.

Pendant l'activation 'rapide', la rupture trypsique de la liaison 1 est quasi-immédiate et elle est suivie à brève échéance par l'autolyse de la liaison 2. Le phénomène est donc simple. Il consiste à convertir le chymotrypsinoïène successivement en deux enzymes (π et δ) dont le dernier jouit d'une relative stabilité. Pendant l'activation 'lente' au contraire, la lenteur des réactions introduit un élément de désordre et de complexité. D'une part,

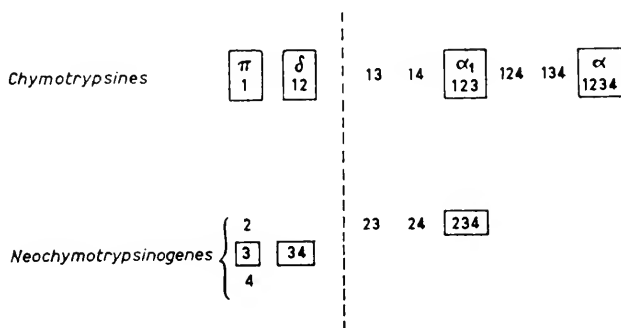


Fig. 6. Possibilités offertes par la rupture de quatre liaisons dans le chymotrypsinoïène. Les cadres indiquent les protéines d'ores et déjà caractérisées: quatre chymotrypsines, dont deux engendrées par l'activation 'rapide' (π et δ) et deux engendrées par l'activation 'lente' (α et α_1); et trois néochymotrypsinoïènes contenant (en plus du résidu de demi-cystine), les résidus N-terminaux suivants: thréonine (3), alanine (34), alanine et sérine (234.)

l'hydrolyse trypsique de la liaison 1 n'est plus forcément terminée avant le début des hydrolyses chymotrypsiques. D'autre part, ces hydrolyses chymotrypsiques ne s'effectuent pas forcément selon un ordre immuable. Les mélanges engendrés sont donc relativement complexes. On sait que la chymotrypsine- α s'obtient avec un rendement assez mauvais et qu'elle n'est pas encore électrophorétiquement homogène après de nombreuses cristallisations.

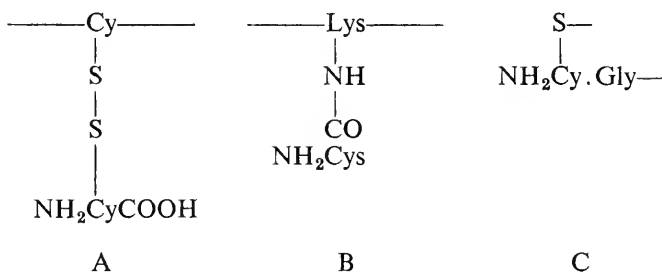
Dès que la liaison 1 est rompue, les molécules sont actives. Les huit protéines de la ligne du haut sont donc des chymotrypsines. Les deux premières sont les chymotrypsines- π et δ de l'activation 'rapide'. Les six autres sont en principe du domaine de l'activation 'lente'. Mais elles n'ont peut-être pas toutes une existence réelle. La dernière est la chymotrypsine- α dont la formation exige le nombre maximum de ruptures (1234). La protéine 123 est la chymotrypsine- α_1 dont l'existence est rendue probable par la présence de thréonine N-terminale dans les hydrolysats de l'activation 'lente'.

Toutes les autres possibilités correspondent en principe à des néochymotrypsinoïènes. Parmi ceux-ci, deux ont été purifiés (3 et 34) et un a été caractérisé (234). Rien n'indique pour l'instant que la vitesse relative des ruptures permet aux autres néochymotrypsinoïènes de se former en quantités décelables.

III. ETUDE PRÉLIMINAIRE DE L'ATTAQUE DU CHYMYOTRYPSINOGENÈ PAR LA LEUCYLAMINOPEPTIDASE

Les protéolyses précédentes sont assez simples pour que leur mécanisme puisse être éclairci de façon complète. Mais, l'endroit précis où elles se produisent, c'est à dire leur position absolue et relative dans la molécule nous restera inconnu aussi longtemps que l'ordre d'enchaînement des résidus de cette molécule n'aura pas été déterminé. En attendant que ce travail ait été réalisé, il paraît intéressant d'attaquer le chymotrypsinogène par des enzymes doués d'une spécificité précise de position et d'observer les répercussions de cette attaque sur l'activité potentielle du zymogène. Nous avons dans ce but utilisé la leucylaminopeptidase du rein récemment purifiée par Smith et *al.*¹³ On sait que cet enzyme agit de préférence sur les peptides. Mais il attaque aussi certaines protéines à partir de leurs extrémités N-terminales.^{14,15}

Le chymotrypsinogène contient un résidu N-terminal de demi-cystine.¹⁰ Ce résidu n'est pas simplement relié au reste de la molécule par son pont disulfure (schéma A) car il suffirait alors d'oxyder le dinitrophényl (DNP) chymotrypsinogène par l'acide performique pour obtenir de l'acide DNP-cystéique libre. Or, ce dérivé n'est obtenu qu'après oxydation et hydrolyse. Le résidu de demi-cystine ne se trouve pas non plus, comme on l'a suggéré,¹⁶ en position latérale sur un groupe ϵ -NH₂ de la lysine (schéma B).



L'expérience suivante montre qu'il est situé à côté d'un résidu de glycolle (schéma C): Du DNP-chymotrypsinogène oxydé par l'acide performique est hydrolysé pendant 30 min. dans HCl 6 N bouillant. Le résidu solide est hydrolysé à nouveau dans les mêmes conditions. Les fractions solubles obtenues après trois opérations semblables sont évaporées et traitées par le Dowex-50 \times 8 forme H⁺.¹⁰ Les substances non-adsorbées sont chromatographiées sur papier dans le système butanol-acide formique. La seule tache jaune obtenue, dont le R_F est identique à celui de l'acide DNP-cystéique, est chromatographiée à nouveau à l'aide de méthyléthylcétone saturée d'eau ou de phénol tamponné à pH 4. On obtient dans les deux cas deux taches jaunes dont l'une migre comme l'acide DNP-cystéique et l'autre

migre un peu plus vite. Cette dernière fournit par hydrolyse des quantités stoechiométriquement égales d'acide DNP-cystéique et de glycolle. Il s'agit donc du peptide DNP-Cys(SO₃H).Gly. Le chymotrypsinogène possède ainsi une véritable extrémité N-terminale Cys.Gly—. Cette extrémité peut en principe constituer un point d'attaque pour l'aminopeptidase. Le tableau 2 résume les principaux résultats obtenus.

RÉSULTATS PRÉLIMINAIRES CONCERNANT L'ATTAQUE DU CHYMOTRYPSINOGENE PAR LA LEUCYLAMINOPEPTIDASE

ChTg: chymotrypsinogène

Aminopeptidase		ChTg cristallisé: (C) ChTg spécialement purifié: (P)	Durée de l'expérience (h.)	Quantités d'acides libérés
Activité spécifique	Nb d'unités par μ -mole de ChTg			
38	4	C	1/4	+++
51	2	C	1/2	+++
51	0.03	C	1/4	++
51	0.8	P	1/4	0
			1	+
			18	+++
92	2	P	2	0
			5	+
90-100	9	P	2	0
			23	+

Les premières expériences ont été réalisées avec des préparations d'aminopeptidase purifiées par précipitations fractionnées. Leurs activités spécifiques vis-à-vis de la L-leucine amide étaient comprises entre 30 et 50. Le chymotrypsinogène était un échantillon commercial, cristallisé trois fois au laboratoire selon les techniques classiques, lyophilisé et conservé à -10°C pendant quelques mois. Les premières expériences réalisées dans ces conditions donnent l'impression que le chymotrypsinogène est bien attaqué par l'aminopeptidase. 0.03 unité d'enzyme par μ mole de substrat libre au bout d'un quart d'heure des quantités déjà considérables d'acides aminés. Toutefois, en analysant ultérieurement notre échantillon de chymotrypsinogène, nous y avons trouvé 0.2 résidu de thréonine N-terminale par mole. Bien qu'il ait été gardé constamment à -10°C , cet échantillon contenait donc 20% de néochymotrypsinogène vraisemblablement formé par un peu de chymotrypsine provenant elle-même d'une trace de trypsine.

Or, cette hétérogénéité grossière pouvait difficilement être mise en évidence par les techniques classiques de la dynamique moléculaire. Au moment où nous avons préparé les néochymotrypsinogènes, nous nous sommes en effet efforcés de séparer ces substances du chymotrypsinogène ordinaire par chromatographie sur Amberlite IRC-50. Les essais ont été infructueux. Dans

toute la zone des pH admissibles, la vitesse de migration a été la même. L'échantillon en cause aurait donc fourni un pic parfaitement symétrique dans une chromatographie de ce type. En outre, l'électrophorèse n'est vraiment efficace dans le cas de protéines aussi voisines qu'à la condition que l'un des constituants du mélange possède une polarité nettement différente des autres. L'unique raison du succès rencontré par Bettelheim et Neurath⁵ dans la séparation électrophorétique des chymotrypsines- π et δ est la présence au sein de la première d'un résidu supplémentaire d'arginine. Dans le cas actuel, le néochymotrypsinogène à thréonine N-terminale possède exactement la même composition que le chymotrypsinogène et le néochymotrypsinogène à alanine N-terminale n'en diffère que par deux résidus (thréonine et asparagine) dont la polarité est très faible. La seule méthode vraiment efficace pour déceler l'hétérogénéité de l'échantillon était donc la méthode chimique de détermination des groupements terminaux. Le fait est digne d'être noté car il peut se reproduire dans d'autres cas.

La présence de néochymotrypsinogènes dans le chymotrypsinogène n'est d'ailleurs pas gênante quand on se propose de déterminer l'ordre d'enchaînement des résidus. Cette étude exige en effet que l'on rompe de nombreuses liaisons et il importe peu que une ou deux liaisons soient déjà rompues dans certaines des molécules utilisées. Mais, quand on veut étudier la structure globale du chymotrypsinogène ou l'action d'une exopeptidase sur cette protéine, une telle hétérogénéité peut provoquer des erreurs très graves. En somme, le chymotrypsinogène est beaucoup plus stable que le trypsinogène et le pepsinogène parce que son activation n'est pas autocatalytique. Mais, il est peu raisonnable d'attendre que le précurseur d'un enzyme protéolytique, si aisément activé par la trypsine et attaqué lentement par son propre enzyme, puisse jouir d'une parfaite stabilité.

Quoi qu'il en soit, nous avons réussi à préparer un échantillon de chymotrypsinogène complètement dépourvu de résidus terminaux parasites (autres que le résidu de demi-cystine) en procédant à plusieurs cristallisations en présence d'une quantité substantielle de diisopropylfluorophosphate. Cet échantillon, dont la stabilité paraît assez satisfaisante, est désigné dans le tableau 2 par la lettre P. Il libère au contact de l'aminopeptidase beaucoup moins d'acides aminés que le premier.

D'autres expériences ont été réalisées avec des préparations d'aminopeptidase spécialement purifiées par électrophorèse de zone dans une colonne d'amidon.¹⁷ Leur activité spécifique atteignait 100, c'est à dire la valeur la plus haute enregistrée jusqu'ici. On voit que cette nouvelle préparation est légèrement moins active que les anciennes sur l'échantillon pur de chymotrypsinogène. Il n'est donc pas toujours prudent de faire agir sur les protéines, comme on le conseille parfois, des préparations d'aminopeptidase dont l'activité spécifique n'excède pas 40.

Enfin, dans les dernières expériences du tableau 2, une aminopeptidase aussi pure que possible a agi sur un chymotrypsinogène spécialement purifié.

9 unités d'enzyme par μ -mole de substrat ont alors engendré au bout de 23 h. environ 0.5 mole/mole de 7 ou 8 amino acides, parmi lesquels se trouvait le glycolle qui occupe la deuxième position dans la séquence N-terminale. On pouvait donc avoir l'espoir d'avoir attaqué le chymotrypsinogène à l'endroit attendu. Toutefois, si la liaison cystyl-glycolle avait été réellement rompue par l'aminopeptidase, le résidu de demi-cystine n'aurait plus été attaché que par son pont disulfure (voir plus haut) et il aurait été susceptible d'être détaché par une simple oxydation. En fait, l'attaque enzymatique ne modifie en rien le comportement de ce résidu.

Le fait surprenant n'est d'ailleurs pas que l'aminopeptidase éprouve quelques difficultés à attaquer la séquence N-terminale du chymotrypsinogène. Cette séquence se compose en effet d'un résidu un peu particulier de demi-cystine et d'un résidu de glycolle pour lequel l'enzyme ne semble pas posséder beaucoup d'affinité. Des expériences ultérieures nous diront si la séquence en question peut être attaquée dans d'autres conditions et si les séquences des néochymotrypsinogènes et des chymotrypsines sont plus accessibles. Mais, malgré leur caractère négatif, nos observations montrent dès maintenant avec quelle prudence il faut interpréter les résultats fournis par l'aminopeptidase, surtout quand il s'agit d'expériences de longue durée, réalisées sur des protéines dont la stabilité n'est pas absolue.

RÉSUMÉ

Le présent mémoire décrit en détail les protéolyses limitées subies par le chymotrypsinogène de boeuf au cours de ses activations 'rapide' et 'lente'. Il donne aussi quelques indications préliminaires concernant l'action de la leucylaminopeptidase sur ce zymogène.

1. L'activation du zymogène résulte de la rupture d'une seule liaison reliant un résidu d'arginine à un résidu d'isoleucine. La chymotrypsine- π ainsi engendrée s'autolyse au niveau d'une liaison leucyl-sérine, en donnant la chymotrypsine- δ dont l'activité est identique et un dipeptide, la séryl-arginine. Le phénomène est très spécifique. Chaque enzyme coupe une seule liaison à concurrence de 95%. La rupture de la liaison leucyl-sérine, laquelle est séparée de la liaison arginyl-isoleucine par deux résidus seulement, ne provoque aucune activation.

2. Pendant l'activation 'lente', deux processus supplémentaires ont le temps de se dérouler: l'attaque chymotrypsique du chymotrypsinogène et l'autolyse de la chymotrypsine- δ . Ces deux processus provoquent la rupture des deux mêmes liaisons (tyrosyl-thréonine et asparaginy-alanine) et la libération d'un dipeptide (la thréonylasparagine). Les nouvelles protéines ainsi engendrées (néochymotrypsinogènes) sont encore activables par la trypsine. Mais leur activation est plus lente et les chymotrypsines formées sont moins actives.

3. Etant le précurseur d'un enzyme protéolytique et étant extrêmement vulnérable vis-à-vis de la trypsine, le chymotrypsinogène ne paraît pas doué

d'une parfaite stabilité. La présence de quantités considérables de néochymotrypsinogènes dans le chymotrypsinogène ne semble pas pouvoir être décelée par les techniques classiques de la dynamique moléculaire.

4. Malgré toutes les précautions prises pour purifier le chymotrypsinogène et la leucylaminopeptidase, cet enzyme libère des quantités très notables d'acides aminés qui ne proviennent pas de la séquence N-terminale de la molécule. Cette séquence est Cys.Gly—.

BIBLIOGRAPHIE

1. M. ROVERY, *Bull. Soc. Chim. Biol.*, **38**, 1101 (1956).
2. J. A. RUPLEY, W. S. DREYER et H. NEURATH, *Biochim. Biophys. Acta*, **18**, 162 (1955).
3. C. F. JACOBSEN, *Compt. rend. trav. Lab. Carlsberg Ser. Chim.*, **25**, 325 (1947).
4. N. M. GREEN et H. NEURATH, *The Proteins*, Vol. II, Part B, H. Neurath et K. Bailey Ed., Academic Press, New York, 1955.
5. F. R. BETTELHEIM et H. NEURATH, *J. Biol. Chem.*, **212**, 241 (1955).
6. M. ROVERY, M. POILROUX, A. YOSHIDA et P. DESNUELLE, *Biochim. Biophys. Acta*, **23**, 608 (1957).
7. M. KUNITZ et J. H. NORTHROP, *J. Gen. Physiol.*, **18**, 433 (1935).
8. M. ROVERY, C. FABRE et P. DESNUELLE, *Biochim. Biophys. Acta*, **12**, 547 (1953).
9. J. A. GLADNER et H. NEURATH, *J. Biol. Chem.*, **205**, 345 (1953).
10. F. R. BETTELHEIM, *J. Biol. Chem.*, **212**, 235 (1955).
11. M. ROVERY, M. POILROUX, A. CURNIER et P. DESNUELLE, *Biochim. Biophys. Acta*, **18**, 571 (1955).
12. C. GABELOTEAU et P. DESNUELLE, *Arch. Biochem. Biophys.*, **69**, 475 (1957).
13. D. H. SPACKMAN, E. L. SMITH et D. M. BROWN, *J. Biol. Chem.*, **212**, 255 (1955).
14. R. L. HILL et E. L. SMITH, *Federation Proc.*, **14**, 226 (1955).
15. R. L. HILL et E. L. SMITH, *Biochim. Biophys. Acta*, **19**, 376 (1956).
16. C. B. ANFINSEN et R. R. REDFIELD, *Advances in Protein Chem.*, **11**, 1 (1956).
17. E. L. SMITH. Communication personnelle.

Studies on the autocatalytic activation of trypsinogen

J. F. PECHÈRE AND HANS NEURATH

Department of Biochemistry, University of Washington, Seattle, Washington

Recent work on the autocatalytic conversion of trypsinogen to trypsin^{1,2} has shown that during the activation, and simultaneously with the appearance of tryptic activity, a lysyl-isoleucine bond is broken in the *N*-terminal region of the trypsinogen molecule, and that correspondingly increasing amounts of a hexapeptide, having the structure valyl-(aspartyl)₄-lysine, can be isolated from the activation mixtures. A similar parallelism has been found to exist between the appearance of enzymatic activity and a decrease in the specific levorotation of the protein.³ While the correlation between these chemical, physical and enzymatic parameters appears to be well established, the interpretation of these data is not self-evident. Thus, the question remains how the removal of a small peptide from the *N*-terminal region of the single polypeptide chain of trypsinogen can account for the appearance of enzymatic activity. In consideration of this problem, attempts were made to establish additional correlations between chemical and physical properties and to look for other events which would help to provide a molecular interpretation for the formation of an enzyme from its inactive precursor.

Such additional correlations have been recently established by investigating the following properties during the activation of trypsinogen: (a) the appearance of trypsin, in the form of soybean-trypsin inhibitor compound, as determined by moving boundary electrophoresis of partial activation mixtures, and (b) measurements of the base consumption during activation, as determined in the Jacobsen-Léonis autoitrator. These were compared with (c) the amount of hexapeptide, as isolated chromatographically from the trichloroacetic acid (TCA)-soluble fraction of activation mixtures,¹ and with (d) the percentage change in optical rotation.³ All these results are summarized in Fig. 1. In each of the four diagrams shown in this figure, the percentage change in property during activation is plotted against the time of activation; the solid line represents in each case the appearance of enzymatic activity (benzoyl-L-arginine ethyl ester as substrate), as calculated by the Kunitz equation⁴ from experimental points (not shown). It is evident that the change in five different chemical or physical properties follows exactly

the same rate law, indicating that they are all manifestations of one and the same molecular event. The correlation between activity and amount of isolated hexapeptide indicates that the lysyl-isoleucine bond is broken simultaneously with activation. The correlation between activity and total amount of trypsin appearing indicates that no intermediate compound having enzymatic activity is formed during the activation process, a conclusion

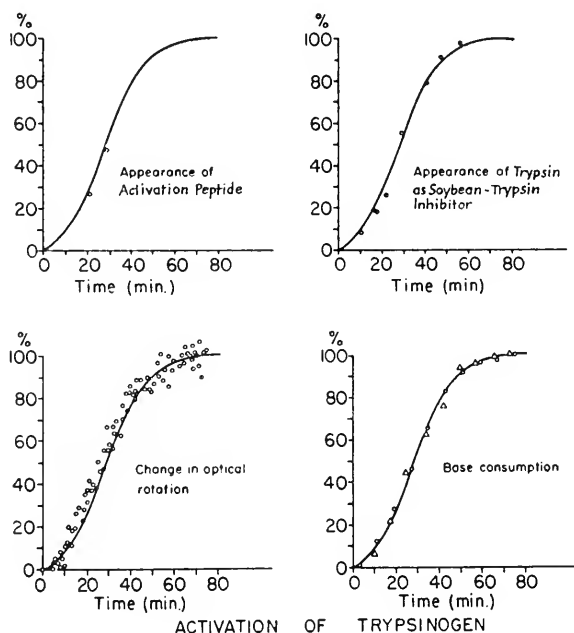


Fig. 1. Correlation between experimental parameters and appearance of enzymatic activity during the autocatalytic activation of trypsinogen. The following conditions applied to all the experiments: trypsinogen concentration approx. 12.5 mg/ml.; trypsinogen/trypsin ratio 20 : 1; 0.1 M Tris buffer (pH 8.0) containing 0.05 M-CaCl₂; temperature approx. 0° C. In all experiments except those shown on the left top quadrant, indole was also present to minimize the action of a slight chymotryptic impurity which was found to be present in the trypsinogen and trypsin preparations used.

which clearly differentiates the activation of trypsinogen from that of chymotrypsinogen.⁵ The correlation between activity and change in optical rotation supports the view that molecular rearrangement (toward a more nearly helical configuration) is an essential aspect of the activation process; and the correlation between enzymatic activity and base consumption is in accord with the assumption that one and only one peptide bond is broken during the reaction. This latter conclusion is dependent in part on an accurate knowledge of the pK of the amino group which is formed during enzymatic hydrolysis. In the present experiments, a pK of 8.1 at 0° was assumed, whereas a pK of 7.6 would have to be assumed if two bonds were broken under these conditions. It is also recognized that the hydrolysis of

a peptide bond involving a more basic amino group, such as the ϵ -amino group of lysine, could not be detected by these measurements.

Trypsinogen contains 16 peptide bonds which, on the basis of the specificity of trypsin, are potentially susceptible to hydrolysis (14 lysyl and 2 arginyl bonds). In order to exclude the possible opening, during activation, of a peptide bond involving the carboxyl group of arginine, acetylated trypsinogen was prepared, by the use of acetic anhydride, and subjected to ion exchange chromatography. There appeared to be an inverse relation between the degree of acetylation and the degree of activation of trypsinogen, suggesting that fully acetylated trypsinogen cannot be activated at all.

Whereas during activation the *N*-terminal valine of trypsinogen is replaced by an isoleucine group,² the *C*-terminal sequence appears to be unaffected during this process. This conclusion is based primarily on the negative evidence that the *C*-terminal group of trypsinogen appears to remain unreactive, before and after activation, toward carboxypeptidase, 'basic carboxypeptidase' or toward hydrazinolysis. The lack of reactivity of the *C*-terminal group in trypsinogen and activated trypsinogen is a structural feature which is shared by chymotrypsinogen as well and which permits various interpretations, as shown in Fig. 2. Structure 1 contains a peptide

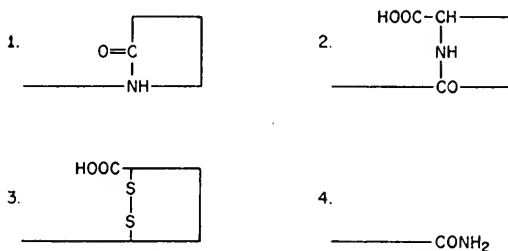


Fig. 2. *C*-terminal structures which would yield negative results by enzymatic (carboxypeptidases) assay or by the hydrazinolysis method.

bond between the *C*-terminal carboxyl group and an ϵ -amino group elsewhere on a side chain. Structure 2 involves a peptide bond between the ϵ -amino group of a *C*-terminal lysine and a side chain carboxyl group elsewhere along the polypeptide chain. Structure 3 has a *C*-terminal cysteine; and structure 4, a *C*-terminal α -amide. All of these structures may be expected to yield negative results by the enzymatic or chemical end-group analyses previously mentioned. The lack of reactivity of activated trypsinogen toward basic carboxypeptidase is to be contrasted with recent work of Gladner, Folk, and Laki⁶ who have reported that crystalline trypsin yields approximately three equivalents of lysine when incubated with this enzyme. Since, in the present experiments, end-group analysis was carried out on freshly-prepared activation mixtures as soon as maximum tryptic activity was reached, it seems likely that during the process of isolation of crystalline

trypsin an additional lysyl-peptide bond had been hydrolyzed by trypsin. Because of the specificity requirements of trypsin, these negative results also seem to exclude the above mentioned possibility of hydrolysis during activation of a peptide bond involving the ϵ -amino group of a lysyl residue.

On the basis of all evidence considered, therefore, the hydrolysis of the lysyl-isoleucine bond in the *N*-terminal region of the single polypeptide chain of trypsinogen seems to be the sole chemical event occurring during autocatalytic activation; the configurational changes, demonstrated by measurements of the optical rotation, must then be considered to be the consequence of this hydrolytic cleavage. It would be difficult to conceive of a mechanism which could account for these observations if trypsinogen were to consist simply of an open polypeptide chain. However, the following considerations suggest that the molecule is not of such a simple structure: (a) the absence of a reactive *C*-terminal group; (b) the presence of 6 intramolecular disulfide bonds; and (c) the relatively low specific levorotation of native trypsinogen (-30°) which suggests a high degree of helical configuration.

A suggested scheme⁷ representing the structural changes involved in activation is represented in Fig. 3, which incorporates the following considera-

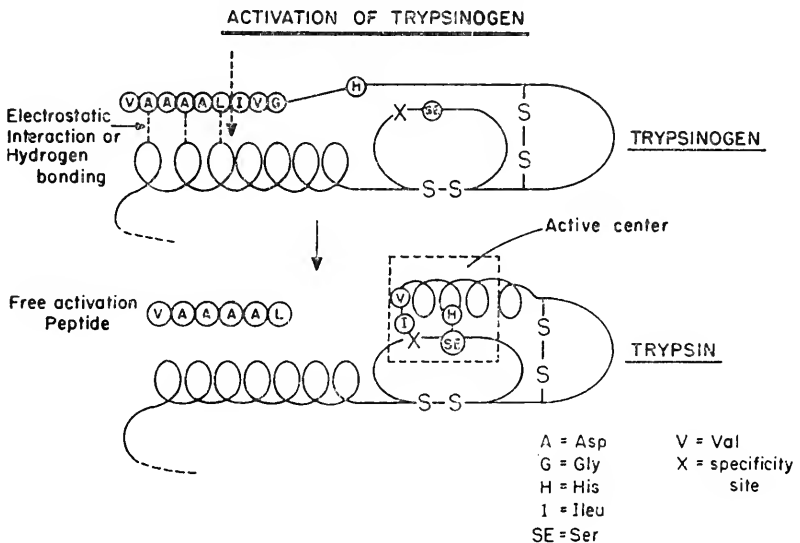


Fig. 3. Proposed scheme for the conversion of trypsinogen to trypsin. For details see the text.

tions: (a) The *C*-terminal region of the molecule is not involved in the activation process and, hence, is omitted from this diagram. (b) A large fraction of the trypsinogen molecule is in a helical configuration, but the *N*-terminal peptide sequence does not follow the helical pattern, presumably because the electrostatic repulsion between the four adjacent aspartic acid residues tends to keep this segment in a more or less extended configuration.

(c) Interaction between the aspartyl carboxyl groups, electrostatically, with positively charged groups, or by hydrogen bonding with groups elsewhere along the chain, provides a loop which is destroyed when the lysyl-isoleucine bond is broken by trypsin, allowing the newly-formed molecule to coil up and to form the active center of trypsin. (d) Two of the six disulfide bridges of trypsinogen are in the region of the molecule which includes the catalytic site. This follows from studies by Dr G. H. Dixon in our laboratory on the tryptic degradation of oxidized diisopropylphosphoryl (DIP)-trypsin, which have shown that the largest DIP-peptide contained five cysteic acid residues among a total of 55 amino-acid residues. (e) The active center, delineated here by the dotted line, includes besides the catalytic site, the grouping conferring side-chain specificity to the enzyme, as denoted by X. (f) The catalytic site involves a histidine and a serine residue. These two groupings are separated from each other in trypsinogen, but upon activation, come sufficiently close to become hydrogen-bonded to one another.^{8,9} This conclusion is based primarily on kinetic studies of the reaction of trypsin and chymotrypsin with acylating agents, such as *p*-nitrophenyl acetate (reviewed in ref. 9).

An essential feature of this scheme is that the splitting of the lysyl-isoleucine bond removes a structural impediment, which allows the *N*-terminal sequence of the newly-formed molecule to coil up, thereby bringing the histidine and serine residues into juxtaposition. According to this picture, cleavage of a bond to the left of the lysine residue might not cause activation, as electrostatic interaction or hydrogen bonding of the lysine side chain would still provide a certain element of structural rigidity and prevent formation of the helix. On the other hand, the splitting of any bond to the right of the lysine side chain would produce the same result as the splitting of the lysyl-isoleucine bond. According to this interpretation, the structural contribution of two or more hydrogen bonds, or salt linkages, is an important factor in the conversion of the zymogen to the active enzyme, and were it possible to rupture these bonds specifically, activation of trypsinogen might ensue. This has not proved possible by the simple expedient of denaturation of trypsinogen by urea; in fact, whereas trypsin is still active in 3M urea, trypsinogen can no longer be activated under these conditions, even by the addition of trypsin, presumably because other structurally important hydrogen bonds are dissolved at the same time.

The highly specific action of trypsin in hydrolysing the lysyl-isoleucine bond may, therefore, be viewed as both a case of limited proteolysis and of limited hydrogen-bond rupture. The question as to whether the hexapeptide is actually liberated during this process or whether it remains attached and is only freed after addition of trichloroacetic acid has not yet been answered in an unequivocal way, but all evidence thus far available suggests strongly that it is liberated as the lysyl-isoleucine bond is broken.

Details of the present results will be published elsewhere.¹⁰

REFERENCES

1. E. W. DAVIE and H. NEURATH, *J. Biol. Chem.*, **212**, 515 (1955).
2. P. DESNUELLE and C. FABRE, *Biochim. et Biophys. Acta*, **18**, 49 (1955).
3. H. NEURATH, J. A. RUPLEY and W. J. DREYER, *Arch. Biochem. Biophys.*, **65**, 243 (1956).
4. M. KUNITZ, *J. Gen. Physiol.*, **22**, 293 (1939).
5. H. NEURATH and G. H. DIXON, *Federation Proc.*; in press.
6. J. A. GLADNER, J. E. FOLK and K. LAKI, *Abstracts, 131st Meeting, Am. Chem. Soc.*, 42c (1957).
7. H. NEURATH and G. H. DIXON, *Federation Proc.*, **16**, 791 (1957).
8. G. H. DIXON and H. NEURATH, *Federation Proc.*, **16**, 173 (1957).
9. L. W. CUNNINGHAM, *Science*, **125**, 1145 (1957).
10. J. F. PECHÈRE and H. NEURATH, *J. Biol. Chem.*; in press.

Comments on the structure of α -chymotrypsin

B. S. HARTLEY

School of Biochemistry, University of Cambridge

I must apologize for presenting somewhat preliminary results on the structure of α -chymotrypsin, but I feel I might be able to supplement some of the information derived from Dr Šorm's peptides.

I have been studying the B-chain of performic acid-oxidized α -chymotrypsin which was first isolated by Meedom (*Acta. Chem. Scand.*, 10, 881 (1956)). It contains about 160 residues—approximately two-thirds of the enzyme—and has a C-terminal tyrosine and an N-terminal Ileu.Val-sequence. It is especially interesting since the active centre of chymotrypsin probably involves a histidine and a unique serine residue, and both these residues are present in this chain.

Both tryptic and chymotryptic digests of the B-chain have been studied. Some of the more profitable results were obtained by following the chymotryptic digestion in a pH-stat and stopping the reaction after 12-15 bonds were split. The fraction of the digest soluble in 50% ethanol at -5° , pH 3.6, consisted almost entirely of 16 neutral and basic peptides, which were purified by paper chromatography and electrophoresis. The composition and sequence in these peptides, which account for about half of the residues in the B-chain, is at present being studied, but some preliminary results may be relevant to this discussion.

The oxidized B-chain contains two arginine residues which appear in the following three peptides after chymotryptic digestion:

1. Ala. Arg. (Thr, Ala, Val). Leu
2. Ala. Asp. Thr. (Asp, Ala, Pro). Arg. Ileu. (Glu₂, Ser₂, Ala, Pro). Leu
3. Ala. Asp. (Asp, Thr, Ala, Pro). Arg. Ileu

It seems probable that peptide 3 arises by further chymotryptic digestion of 2. A tryptic digest of the B-chain yielded Ala. Arg and Ala. (Asp₂, Thr, Ala, Pro). Arg as the only peptides containing arginine, confirming that peptides 1. and 2. above represent the environment of the two arginine residues in this chain.

A further peptide which may be of interest is Ala. (Ala, Asp-NH₂)—isolated in good yield from the chymotryptic digest. This may be another example of an asparaginyl split by chymotrypsin, such as was mentioned by Professor Desnuelle, though one has to be wary of peptides from this oxidized protein because of the unknown fate of the tryptophan residues.

Reaction of trypsin with organophosphorus esters

T. VISWANATHA

Carlsberg Laboratory, Copenhagen, Denmark

The relationship between protein structure and biological activity has been a matter of considerable discussion in recent years. Turning our attention to specific cases of proteolytic enzymes, the opinion as to the nature of the active site(s) appears to be sharply divided. Thus, for example, in the case of α -chymotrypsin, photo-oxidation studies by Weil and co-workers,¹ the dye-binding studies of Hartley and his associates,² the reaction with fluorodinitrobenzene by Whitaker and Jandorf,³ and similar studies by others,⁴ suggest the association of an imidazolyl residue on the enzyme with the active site. Evidence to the contrary have been reported by Neurath and his school⁵ and by Cohen and his associates.⁶ The ability of chymotrypsin to react with diisopropyl fluorophosphonate (DFP) has been used in most of the above-mentioned studies to throw some light on the nature of the active site(s).

Studies on the reaction of trypsin with DFP and diethyl *p*-nitrophenyl phosphate (E600) carried out in the Carlsberg Laboratory have led to the following results and conclusions:

(I) Both DFP and E600 can react with trypsin in the presence of high concentrations of urea resulting in the incorporation of one mole of phosphorus per mole of the enzyme. A nearly stoichiometric yield of *p*-nitrophenol (one mole per mole of enzyme) was obtained during the reaction with E600. The unique feature of the reaction in the presence of urea is that the enzyme maintains its ability to revert to the active form on appropriate dilution. In aqueous medium irreversible inactivation of the enzyme accompanies the reaction.

(II) The reaction of trypsin with DFP in aqueous medium at pH 8.0 results in the incorporation of 2.0 moles of phosphorus per mole of the enzyme. At pH more alkaline than 8.0, more than 2.0 moles of phosphorus per mole of the enzyme are introduced into the protein. While the first 2.0 moles of phosphate introduced appear to be stable to acid treatment, the additional phosphate groups are acid-labile suggesting their attachment to amino groups on the protein.

The presence of 2.0 moles of phosphorus has been investigated further.

It has been found that the incorporation of the first phosphate on the enzyme is a rapid reaction and leads to inactivation. The second phosphate is introduced at a much slower rate than the first.

E600 under identical conditions yields only 1.0 mole phosphorus per mole of the enzyme. It was noted that the reaction between trypsin and E600 occurs at a much lower rate compared with that with DFP. This observation might account for the differences in the phosphorus incorporation data.

(III) Blocking the amino groups of trypsin by acetylation (73% of amino groups blocked as shown by ninhydrin determination) impairs the ability of the enzyme to react with E600 in the presence of urea, whilst the reaction in aqueous medium can still occur.

(IV) The inactive precursor, chymotrypsinogen, could be phosphorylated by DFP either in aqueous medium or in the presence of urea. A polyphosphoryl derivative capable of activation is formed.

(V) A soluble, active diisopropyl phosphoryl trypsin (with 0.75 moles 'P' per mole and 81% activity) has been obtained. All the above-mentioned results suggest that an active enzyme is not an essential pre-requisite for the reaction with DFP and E600. These results and conclusions thereof are in apparent disagreement with those of Harris and Hartley⁷ and Neurath and his associates.⁸ The observations in the present study have been interpreted along the following lines:

The reaction of trypsin with DFP or E600 may be visualized to proceed by two independent phosphorylation mechanisms. The first one is perhaps catalysed by an imidazole group (or some other group in the case of E600) on the protein. This reaction is fast and leads to inactivation. The second phosphorylation, which is a slow process, appears to be mediated by an amino group. A common feature in both mechanisms appears to be a migration of the phosphoryl moiety to an acid-stable acceptor group. Whilst the first mechanism ceases to operate in strong urea solution, the latter is unaffected by such changes in the medium. Acetylation appears to eliminate the second mechanism. The second phosphorylation reaction has no influence on the enzymic activity if carried out under conditions when the first one is not operative as in the presence of urea.

With this information, efforts were next concentrated to obtain some idea of the nature of the active site(s) involved in the reaction.

The results obtained are summarized below:

(I) Histidine or imidazole fails to catalyse the hydrolysis of E600 at pH 7.6, whilst DFP hydrolysis is enhanced by the presence of histidine.

(II) Hydroxylamine is capable of enhancing the rate of DFP and E600 hydrolysis.

(III) None of the amino acids shows any catalytic activity towards E600.

(IV) Studies on the pH optimum for the inactivation of trypsin by the inhibitors show that the rate of inactivation by DFP increases with pH from

5.0 to 7.0, and, above pH 7.0, the rate is independent of pH. On the other hand, in the case of E600 a sharp maximum for the reaction is noticed at pH 6.90. The latter corresponds to the pH optimum for the reactivation of E600 inhibited chymotrypsin reported by Cunningham.⁹

The difference in the behaviour of DFP and E600 recorded above might be interpreted as suggesting the involvement of different groups on the enzyme for the reaction with these organophosphorus esters. It might also reflect on the inadequacy of any one amino acid side chain on the protein molecule to account for the catalytic function of the enzyme. The enzymic activity might be related to the presence of some special structure or spatial configuration of groups in the protein molecule.

A detailed report of this work will appear in *Comptes rendus des Travaux du Laboratoire Carlsberg, Série chimique*.

REFERENCES

1. L. WEIL, S. JAMES and R. BUCHERT, *Arch. Biochem. Biophys.*, **46**, 266 (1953).
2. B. S. HARTLEY and V. MASSEY, *Biochim. Biophys. Acta*, **21**, 70 (1956).
3. J. R. WHITAKER and B. T. JANDORF, *J. Biol. Chem.*, **223**, 751 (1956).
4. T. WAGNER-JAUREGG and B. E. HACKLEY, Jr., *J. Am. Chem. Soc.*, **75**, 2125 (1953).
5. G. H. DIXON, W. J. DREYER and H. NEURATH, *J. Am. Chem. Soc.*, **78**, 4810 (1956).
6. J. A. COHEN, R. A. OSTERBAAN, M. G. P. J. WARRINGA and H. S. JANSZ, *Discussions, Faraday Soc.*, August, 1955.
7. J. I. HARRIS and B. S. HARTLEY, *Biochim. Biophys. Acta*, **21**, 201 (1956).
8. G. H. DIXON and H. NEURATH, *Biochim. Biophys. Acta*, **20**, 572 (1956).
9. L. W. CUNNINGHAM, Jr., *J. Biol. Chem.*, **207**, 443 (1954).

Modification of pepsin by autodigestion

GERTRUDE E. PERLMANN AND
MARY J. MYCEK*

The Rockefeller Institute for Medical Research, New York, U.S.A.

In the work on protein structure one objective has been the discovery and use of enzymes that attack linkages in specific configuration thus leaving the protein substrate relatively intact. Although the action of such enzymes may consist in as little as the cleavage of one bond—or the removal of a certain group of the molecule—considerable changes in some of the properties of the protein may occur.¹ Thus it has been possible to demonstrate that the enzymic removal of the pepsin-phosphorus does not impair the proteolytic activity of the protein although its electrophoretic mobility is altered. In contrast to these results it is of interest that through the action of 'pepsin on pepsin' a modified form of the protein can be obtained. Here, the physico-chemical properties such as the electrophoretic behavior and the molecular weight have not changed appreciably but the relative specific enzymic activity is higher than that of the parent substance.

As shown by Steinhardt,² pepsin, if stored in 1.0 to 6.0 M urea at 3°C for one to fifteen days, retains its enzymic activity. However, a marked loss of proteolysis occurs, if pepsin is exposed to 8.0 M urea at temperatures above 20°C.³ Thus, the activity of a 0.5 per cent pepsin solution kept in urea-acetate of pH 5.3 and 0.07 ionic strength for 24 hours at 25°C, is reduced to 40 per cent of its initial value, whereas at 37°C almost complete inactivation takes place. The rate of inactivation is a function of the urea concentration, the pH and the ionic composition of the solvent. In addition to the loss of enzymic activity the absorption maximum of pepsin in the ultraviolet is shifted from 2780 Å to the shorter wave length of 2760 Å after treatment with urea. It emerges from these results that the action of the urea on pepsin consists in the rupture of hydrogen-bonded phenolic hydroxyl groups of the tyrosine residues. The protein unfolds slightly and a number of peptide bonds hitherto not accessible to enzymic hydrolysis are now exposed and rapidly hydrolyzed. The view that an enzyme-catalyzed auto-lysis is the underlying phenomenon is further supported by the fact that exposure to urea produces an increase in the solubility of pepsin in hot 10 per cent trichloroacetic acid. The formation of non-protein material is most marked in the pH range of 4.0 to 5.5.

* Public Health Service Postdoctoral Fellow of the National Cancer Institute, National Institutes of Health.

Since at any given stage of the urea inactivation not more than 70 per cent of the trichloroacetic acid soluble material is dialysable, samples of pepsin treated with 8.0 M urea at 37°C for 6, 12, 18 and 24 hours, respectively, were dialyzed and analyzed electrophoretically. As indicated in Fig. 1, the

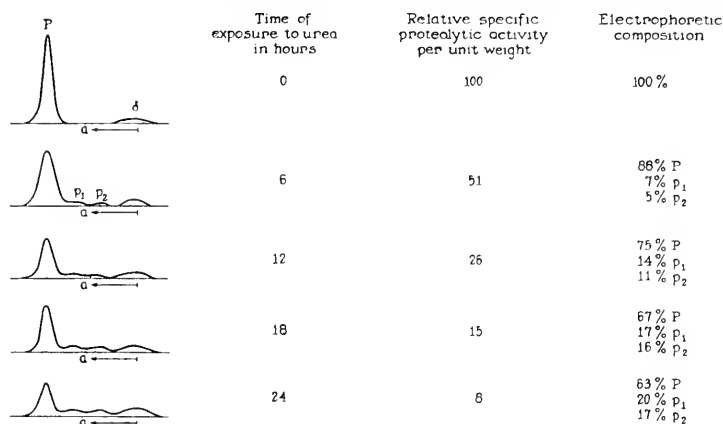


Fig. 1. Tracings of electrophoretic patterns of 0.7% pepsin solutions in sodium acetate buffer of pH 4.6 and 0.1 ionic strength. Electrophoresis was carried out for 5400 sec. at a potential gradient of 6.4 volts per cm.

loss of activity is accompanied by the appearance of two new components, p_1 and p_2 , formed at the expense of the original protein, P . The relative specific activity of the mixture which has been exposed to urea for 24 hours is only 8 per cent of the original. On removal of the low-molecular weight material by dialysis, the relative enzymic activity of the non-dialyzable fraction, however, is similar to that of the original protein. Hence, the question arises whether the residual activity is due to the presence, in the reaction mixture, of some unchanged pepsin.

To test this hypothesis the electrophoretic components of pepsin after treatment with urea at 37°C for 24 hours, were separated by zone electrophoresis on a Geon resin. As shown by Fig. 2, the distribution of the electrophoretic components is similar to the pattern obtained in free electrophoresis (Fig. 1). The relative concentration of p_1 and p_2 , however, had decreased considerably owing to the fact that this sample had been pre-dialyzed against a hydrochloric acid-glycine buffer of pH 2.5 and 0.1 ionic strength prior to the dialysis against the acetate buffer used in the electrophoresis experiment. As indicated in Fig. 2, only the major component of the mixture with a mobility similar to that of pepsin had proteolytic activity.

Although the sedimentation constant of the active component, $s_{20} = 2.8_1 \times 10^{-13}$, is only slightly lower than that of pepsin, $s_{20} = 2.9_6 \times 10^{-13}$,* the following differences have been established: The relative specific activity per unit nitrogen of the modified protein is 40 to 50 per cent higher than

* The authors are indebted to Dr R. Trautman for the ultracentrifuge measurements.

that of the parent substance. If the activity is expressed per unit tyrosine, the increase is 80 to 100 per cent. The hydrolysis of hemoglobin⁴ and of the synthetic substrate acetyl-L-phenylalanyl-diiodotyrosine⁵ are affected to the same extent. Preliminary amino acid analyses, kindly carried out by Dr D. Spackman, revealed differences not only in the tyrosine content but also in that of several other amino acids.

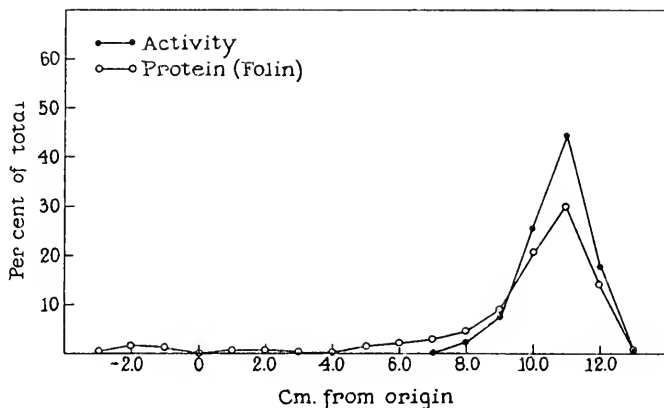


Fig. 2. Zone electrophoresis on Geon resin of pepsin treated with 8.0 M urea at 37°C for 24 hours. Electrophoresis was carried out in acetate buffer, pH 4.6 and 0.1 ionic strength after predialysis in glycine-HCl, pH 2.5, 0.1 ionic strength.

The present observations consist of a few experiments only, and it is not possible to state whether the urea modified pepsin represents a mixture of molecules which differ from each other by small changes in their amino acid composition brought about by autodigestion or whether the original pepsin molecules have been broken down to smaller units which reaggregate to form a complex of a molecular weight of 30,000 to 35,000.

From the results presented here it can be concluded that the apparent loss of the enzymic activity of pepsin after treatment with urea is caused by an enhanced autodigestion. At any given stage, however, during the inactivation enzymically active modifications of the parent enzyme exist which differ in some of their properties. Although an enhancement of activity has been observed it has not yet been possible to change the specificity of the enzyme.

REFERENCES

1. G. E. PERLMANN, *Advances in Protein Chem.*, **10**, 1 (1955).
2. J. STEINHARDT, *J. Biol. Chem.*, **123**, 453 (1938).
3. G. E. PERLMANN, *Arch. Biochem. Biophys.*, **65**, 210 (1956).
4. M. L. ANSON in *Crystalline Enzymes* (J. H. Northrop, M. Kunitz and R. M. Herriott, eds.), p. 305, Columbia University Press, New York, 1948.
5. L. E. BAKER, *J. Biol. Chem.*, **193**, 809 (1951).

Some studies on the structure and activity of papain*

EMIL L. SMITH, ROBERT L. HILL
AND J. R. KIMMEL

The Laboratory for the Study of Hereditary and Metabolic Disorders and the Departments of Biological Chemistry and Medicine, University of Utah College of Medicine, Salt Lake City, Utah

The successful elucidation of the structure of insulins from several species, and of a number of smaller polypeptide hormones and antibiotics has stimulated efforts to understand the relationship between protein structure and biological activity.¹ Enzymes are particularly suited for such investigation inasmuch as their activity and specificity can be studied *in vitro*. The structure of several enzymes is now being investigated in a number of laboratories, and it is likely that these efforts will soon be successfully completed in the case of a few of the smaller molecules. However, a knowledge of amino acid sequence will not by itself solve the problem of correlating structure and function. Much more information will be necessary to achieve real understanding of the catalytic properties of enzymes. First of all, the region or regions of the protein which are involved in specific interaction with the substrates must be identified and a satisfactory explanation of the substrate specificity must be at hand. Secondly, the kinetic behavior and overall reaction mechanism must be explicable in terms of the enzyme structure. It is obviously a difficult task at present to find enzymes of suitable purity and availability which lend themselves to all of the necessary types of study and which are small enough to permit attacking the problem of structure by presently available methods.

In our laboratory, we have chosen a proteolytic enzyme, papain, for intensive study because it possesses a number of desirable features; however, it has some disadvantages as well. Among the many advantages which papain possesses is the fact that the enzyme can be readily isolated in reasonable quantity from the commercially available dried papaya latex.² As with

* The work reported has been supported in large part by grants from the National Institutes of Health, United States Public Health Service, The American Cancer Society and the Rockefeller Foundation.

other proteolytic enzymes, synthetic substrates of known structure can be readily prepared. Because of the good stability of the enzyme under a variety of conditions, kinetic studies with substrates of different structure can be performed and such studies have yielded valuable information concerning the nature of the active site and a possible mechanism of action of the enzyme.^{3,4}

One of the most important features of papain is the finding that a single sulfhydryl group is essential for activity and that this group is intimately concerned in the enzymic mechanism.^{4,5} This presents the opportunity of labeling the specific part of the peptide chain which is directly involved in the interaction with substrates.

This brief consideration of some of the desirable properties of the enzyme should not serve to obscure the fact that a protein containing approximately 180 residues poses difficult problems for study. Although satisfactory methods appear to be available for determining the amino acid sequence, it is obvious to all who are engaged in such studies that the technical problems are formidable indeed and that the labor involved is considerable.

In this presentation we shall discuss some of the studies on crystalline papain which indicate the approaches which are being used in our laboratory. Major emphasis will be given to studies of the overall structure and of attempts to determine the nature of the active region of the molecule by enzymic degradation of the molecule, by labeling the active group, and by kinetic analysis.

GENERAL PROPERTIES AND HOMOGENEITY

Papain has a molecular weight of about 20,500 as estimated by both physical and chemical methods. (Table 1.) The molecule is slightly basic with an iso-

Table 1
SOME PROPERTIES OF PAPAIN

Molecular weight (from sedimentation-diffusion)	20,700
" " (from amino acid content)	20,300
" " (from crystalline mercury complex)	$1/2 \times 41,400$
" " (from maximum mercury binding)	19,900
Isoelectric point (0.1 ionic strength)	8.75
$S_{20, W}$	2.42 S
Diffusion constant ($\times 10^7$ sq. cm. per second)	10.3
Free α -amino groups per mole	1 Isoleucine

electric point at pH 8.75 in buffers of 0.1 ionic strength. The crystalline enzyme, or better, its crystalline mercury derivative, is essentially homogeneous as judged by sedimentation studies in the ultracentrifuge,⁶ by electrophoretic analysis,⁶ and by end-group determinations.⁷ Immunochemical

investigations⁸ show that the crystalline enzyme is a good antigen and the system of protein and rabbit antiserum shows all of the behavior characteristic of a single antigen and a single antibody. Moreover, since the antiserum strongly inhibits the enzymic activity, it is evident that the enzyme itself is the antigen. In essence, all of these studies are consonant with the view that crystalline papain is molecularly homogeneous. Some of the overall properties of the molecule are summarized in Table 1.

THE SULFHYDRYL GROUP AND SULFUR DISTRIBUTION

Papain is inhibited by all of the agents which react with thiol groups, e.g., heavy metal ions, organic mercurials, alkylating agents, and oxidizing agents. The presence of an essential thiol group thus presents a number of disadvantages since the enzyme is readily inactivated by traces of metal ions and by oxidation under aerobic conditions. Full enzymic activity is obtained only when a chelating agent and a reducing agent are present.² Dimercapto-propanol (BAL) serves in both capacities and is a convenient compound to use in kinetic studies of the enzyme.^{4,9}

The strong affinity of the enzyme for heavy metal ions has been useful in permitting the isolation of crystalline mercuripapain which is inactive and is thus not liable to autodigestion.² Moreover, mercuripapain is more homogeneous and more stable than crystalline papain itself. Although crystalline mercuripapain contains one atom of mercury for each two moles of protein, it has been possible to demonstrate that a molecule of reduced papain can bind maximally one mercuric ion.^{4,5}

The spectrophotometric method of Boyer¹⁰ has been employed to estimate the affinity of reduced papain for *p*-chloromercuribenzoate (PCMB).^{5,11} Fig. 1 shows the results of such studies with different preparations of the crystalline enzyme. It is evident that in each case less than one thiol group per mole of papain can react with the mercurial. In addition, the proteolytic coefficient (C_1) for the hydrolysis of benzoyl-L-argininamide, varies with the amount of active SH. It should be noted that Balls and Lineweaver¹² originally observed that different preparations of the crystalline enzyme vary in absolute activity.

The finding that less than one mole of reactive SH is present in papain does not indicate any contamination of the enzyme with other proteins. On the contrary, crystalline preparations of the enzyme with different proteolytic activity manifest the same physical properties, are homogeneous and possess the same amino acid composition. The explanation for the variation in activity and reactive thiol resides in the observations that preparations of the enzyme or its mercury derivative undergo an irreversible inactivation on storage. Indeed, this change appears to occur in the crude latex as well, and freshly prepared samples of crystalline enzyme manifest this variability.⁵

The linear relationship between proteolytic activity and reactive thiol shown in Fig. 1 provides strong evidence that the thiol group is directly involved in the catalytic activity of the enzyme. It may be noted that the thiol group of papain reacts very rapidly with PCMB. Indeed, the reaction is essentially complete in less than two minutes.

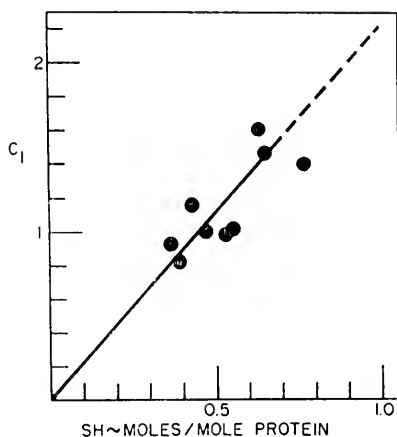


Fig. 1. Proteolytic coefficient (C_1) for hydrolysis of benzoyl-L-argininamide as a function of sulfhydryl reactive with *p*-chloromercuribenzoate for different preparations of column-reduced crystalline papain. (Finkle and Smith, unpublished.)

Less extensive studies of the reaction of iodoacetamide with the enzyme also demonstrate that less than one equivalent of this substance is bound by the enzyme.⁵ Moreover, this reaction produces complete and irreversible inhibition of the enzyme. Reaction of the alkylating agent with a cysteine residue should produce a thiol ether derivative (Fig. 2). Identification of this

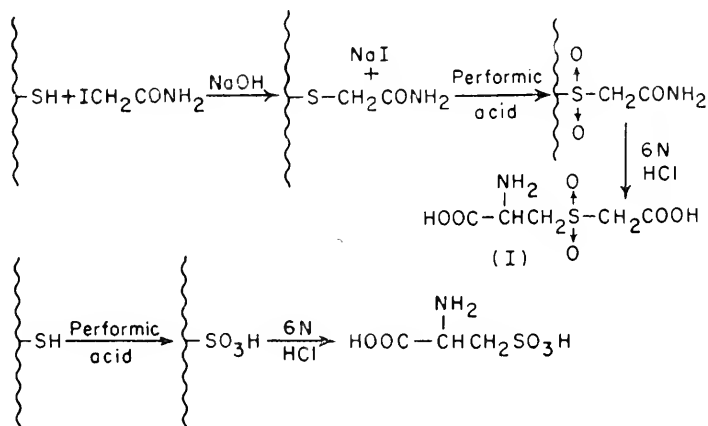


Fig. 2. Reaction of iodoacetamide with a sulfhydryl group and subsequent fate after oxidation and hydrolysis to produce *S*-carboxymethylcysteine sulfone (I).

derivative has been accomplished by the following procedure.⁵ The reduced enzyme was allowed to react with one equivalent of iodoacetamide. The protein derivative was then exhaustively dialyzed and the preparation was oxidized with performic acid. An aliquot was hydrolyzed with 6N-HCl under the usual conditions and worked up in the conventional manner. Chromatography of a sample of the hydrolysate showed the presence of carboxymethylcysteine sulfone which was identical in behavior with that of an authentic, synthetic sample of this compound. Thus, there is no question that the reactive thiol group is present on a residue of cysteine.

The aforementioned studies with PCMB, iodoacetamide and with mercury⁵ are in accord with the view that there is only one reactive thiol group in the enzyme. However, it is essential to ascertain the total number of such groups in the protein. Preparations of papain, treated by various methods, react with more than one mole of PCMB. Under certain conditions, it has been found that the enzyme will react with 6 moles of the mercurial, as judged by mercury analyses.^{4,5} These results indicate that 6 thiol groups are present in the molecule. Inasmuch as hydrolysates of oxidized papain yield 6 moles of cysteic acid,¹³ the agreement between these results suggests that cystine itself, and hence that disulfide bridges, are absent from the papain molecule.

It has been reported^{14,15} that methionine is absent from papain. Nevertheless, elementary analysis reveals that the molecule contains 8 atoms of sulfur^{2,5} and the presence of only 6 cysteine residues leaves two atoms of sulfur to be explained. We have been able to exclude the presence of inorganic sulfate, acid-labile sulfate esters, thiolhistidine, or other known amino acids containing sulfur.⁵ Thompson¹⁶ could not detect lanthionine in hydrolysates of papain. It is our hope that the present systematic investigation of the structure will provide eventually an understanding of the nature of the two missing sulfur atoms.

STUDIES OF STRUCTURE

Amino terminal groups. With the dinitrophenyl method of Sanger,¹⁷ Thompson⁷ demonstrated that mercuripapain contains only one free α -amino group, that of isoleucine. From partial hydrolysates of the dinitrophenyl protein, he isolated the unique tripeptide, DNP-isoleucylprolylglutamic acid. These findings indicate that the papain molecule consists of only a single peptide chain. In addition, Thompson isolated 8 moles of ϵ -DNP-lysine which is in accord with the number of lysine residues found by amino acid analysis.¹⁴

Peptides containing cysteic acid. Inasmuch as at least one cysteine residue is of primary importance in the catalytic activity of papain, a study was made of the strongly acidic peptides which could be isolated from partial acid hydrolysates of papain which had been oxidized with performic acid.¹³ Hydrolysis of the oxidized protein was performed in 12N-HCl at 37° for

7 days. The hydrolysate was worked up in the usual manner and subjected to electrophoresis on paper in acetic acid at pH 2.3 or passed through Dowex-50 \times 4 in the hydrogen cycle. The strongly acidic fraction was obtained by both methods. The peptides containing cysteic acid could then be separated by two-dimensional paper chromatography. The peptides isolated by this procedure, whose structure has been completely or partly elucidated, are listed in Table 2. Obviously, these few peptides do not account for all of the sequences involving cysteic acid which must be present in the oxidized protein, probably because neutral peptides, i.e., those containing cysteic acid and a basic residue, would not be isolated by the methods employed.

It is evident that the fragmentary data of Table 2 give little information by themselves and no clue is obtained as to which one of these sequences, if any, may be concerned in the enzymic activity.

Amino acid composition. The composition of papain is given in Table 3.¹⁴ Estimation of most of the amino acids was performed by chromatography

Table 2

CYSTEIC ACID PEPTIDES DEFINITELY
IDENTIFIED FROM PARTIAL HYDROLYSATES
OF OXIDIZED PAPAIN¹³

Asp. CySO ₃ H	CySO ₃ H. Asp
Ser. CySO ₃ H	CySO ₃ H. Gly
Val. CySO ₃ H	CySO ₃ . Gly. Asp
Gly (Pro, CySO ₃ H)	Ser. CySO ₃ H. CySO ₃ H (?)

Table 3

AMINO ACID COMPOSITION OF PAPAIN^{13,14}

Amino acid	Residues	Amino acid	Residues
Asp	17	Leu	9
Thr	7	Tyr	17
Ser	11	Phe	4
Glu	17	His	1
Pro	9	Lys	8
Gly	23	Arg	9
Ala	13	Tryp	5
Val	15	1/2 Cys	6
Ileu	9	Amide	(19)
		Total	180

on ion-exchange columns by the procedures of Moore and Stein.¹⁸ In our studies,¹⁴ we used different times of hydrolysis in 6N-HCl in order to correct for destruction of labile amino acids and to produce complete liberation of slowly released residues. Tryptophan was estimated by a colorimetric method¹⁹ and cysteine was estimated chromatographically after hydrolysis of the oxidized protein.¹³

The composition of papain shows several noteworthy features: the absence of methionine, the presence of only four residues of phenylalanine and one of histidine, and the high content of glycine and tyrosine.

Tryptic digestion of oxidized papain. The most useful approach to the study of amino acid sequence in large polypeptides and proteins has been to degrade the molecule first into a relatively small number of large peptides. Experience^{17,1} has shown that this type of degradation can best be accomplished with proteolytic enzymes. Selection of the proteolytic enzyme depends upon the nature of the protein being studied and the specificity of the enzyme. With papain, trypsin appeared to be the enzyme of choice for first study. Papain contains 9 arginine and 8 lysine residues, thus providing 17 sites at which trypsin can act and this would yield, theoretically, 18 peptides. The amino acid composition of papain is such that other proteolytic enzymes could be expected to yield a larger number of fragments.

Papain itself, mercuripapain, and papain combined with 6 moles of PCMB are relatively resistant to the action of trypsin.²⁰ Furthermore, each of these materials is potentially capable of being partially activated during tryptic digestion, thus permitting autolysis and inactivation of the trypsin. Autolysis is particularly undesirable because of the broad specificity of papain. The most satisfactory papain derivative for degradation by trypsin is the performic acid-oxidized protein. This derivative has no proteolytic activity and is digested by trypsin at pH 7.5. However, oxidized papain is insoluble at this pH and digestion proceeds slowly for several days. Oxidation of papain results in modification of the 5 tryptophan residues in the protein as well as in oxidation of cysteine to cysteic acid residues.²¹

The procedure has been to digest oxidized papain at pH 7.5 with crystalline trypsin until the ninhydrin color of an aliquot is constant. In general, fresh trypsin has been added every 10 to 20 hours. In no case did the trypsin content of the digestion mixture exceed 2 per cent of that of the oxidized papain. The resulting digest is dried from the frozen state and the dry material is extracted with a 0.2 N formate buffer at pH 2.1. The soluble peptides in the extract are resolved by chromatography. The large amount of insoluble residue must be fractionated by other means.

Chromatography of the soluble peptides has been accomplished on a 150 cm. column of Dowex 50 × 2 with a gradient elution system. The elution pattern and the pH and ionic strength gradients are shown in Fig. 3.²¹ It should be noted that this particular chromatogram was obtained on a

column 7 mm. in diameter and represents the first successful attempt to resolve the tryptic digest of oxidized papain. Subsequent chromatography was performed on a column 30 mm. in diameter having 17 times the capacity of the original column. In general, the transition from the small to the large column has resulted in only minor changes in the elution pattern.²⁰

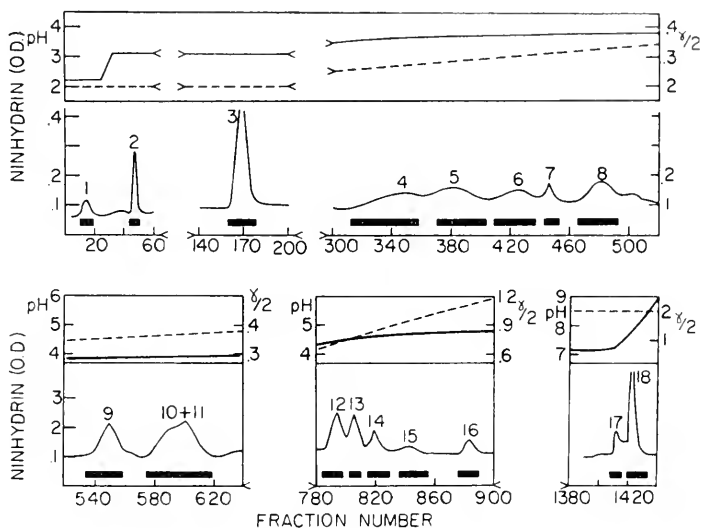


Fig. 3. Chromatographic separation of soluble peptides found in a tryptic digest of oxidized papain. The ninhydrin-reactive components were resolved on a 150 cm. column of Dowex-50 \times 2 with buffers varying in ionic strength and pH. The solid bars show the fractions which were pooled. (Kimmel and Smith, unpublished.) O.D. is optical density; upper part of figure: full lines, pH; broken lines, ionic strength.

Amino acid analyses have been performed on all of the fractions numbered in this figure. Most of the fractions up to Fraction 14 appear to contain single peptides or mixtures of peptides in which each peptide contains only one basic amino acid. Fractions 14 through 18 contain mixtures of large complex peptides with more than one basic amino acid. Presumably, these result from incomplete tryptic digestion, although it is possible that the peptide bonds involving the linked carboxyl groups of lysine and arginine in these fragments are resistant to trypsin, as in a lysylproline sequence.

Fractions 1, 2 and 3 in this chromatogram emerge as sharp symmetrical peaks and the stoichiometry of the amino acid analyses suggested that these fractions were composed of nearly pure peptides. Fraction 1 is particularly interesting since analysis showed 27 amino acid residues including 2 residues of cysteic acid. No basic amino acid was found in this fraction and, therefore, it was assumed that fraction 1 represents the C-terminal portion of papain.²¹ Fraction 2 is a smaller peptide with a high aspartic acid content.

The elution pattern shown in Fig. 3 was obtained on 5 different occasions.

Subsequent patterns (Fig. 4) have shown only traces of ninhydrin-positive material in the regions where Fractions 1, 2, and 3 were expected and analyses of the small peaks in these areas have revealed only small amounts of material with free cysteic acid predominating in the region of Fraction 1 and free aspartic acid in the region of Fraction 2.²⁰ In later portions of the chromatogram reproducibility has been much better. Fractions 5 through 16 have appeared constantly in their expected positions. The complexity of the fractions is, however, quite variable. Fraction 4 (Fig. 3), for example, appears to be a mixture of small amounts of a number of peptides or amino acids, and the shape of the curve shown here is certainly suggestive of this. Fractions 10 plus 11 (Fig. 4) are composed of at least three peptides, and Fractions 12 and 13 overlap with the same large complex peptide. Fractions 5, 6, 8 and 9 appear to be constant in character.

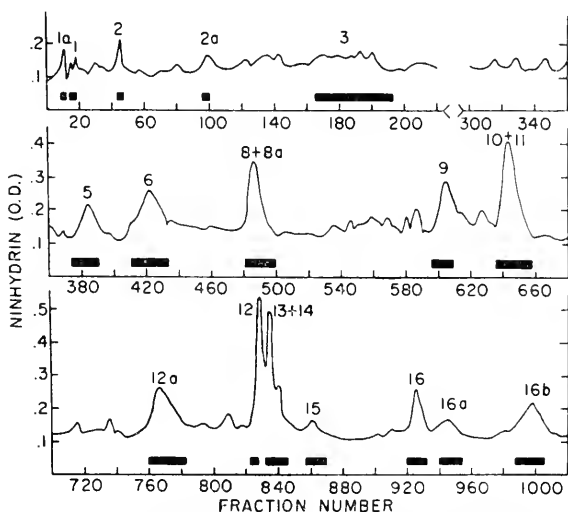


Fig. 4. Chromatographic separation of peptides in a tryptic digest of oxidized papain. Except for the absence of Fractions 1 to 3, the results were similar to those shown in Fig. 3. (From Kimmel and Smith, unpublished.)

Table 4 shows the composition of the peptides that have been isolated and shown to be pure.²⁰ Unknown sequences are enclosed in parentheses. It is interesting that two of the sequences shown have been encountered in other proteins and polypeptides. The sequence serylarginine has been found in chymotrypsin²² and in ribonuclease.²³ The very unusual sequence ala.ala.ala.lys. is also present in ribonuclease.²³

It is obvious that the peptides shown in Table 4 represent less than one third of the papain molecule. The yield of these peptides is not high, ranging from 20 to 40 per cent, but significant enough to indicate that they represent authentic fragments from oxidized papain. Several of the fractions shown in Figs. 3 and 4 are known to be mixtures of peptides, and attempts

Table 4

SOME PURE PEPTIDES ISOLATED FROM A TRYPTIC DIGEST OF OXIDIZED PAPAIN

Fraction	Amino acid composition	Number of residues
1	[(CySO ₃ H) ₂ , Asp ₂ , Thr, Ser ₂ , Glu ₃ , Pro, Gly ₅ , Ala, Val ₂ , Ileu, Leu ₃ , Tyr ₄]	27
5	Gly.Gly.Ileu.Phe.Val. Gly.Pro.CySO ₃ H. Asp.Gly.Lys	11
6	(Asp, Ser ₂ , Glu ₂ , Pro, Gly, Ala ₂ , Val ₂ , Ileu, Leu) Lys	14
8a	(Asp, Thr ₂ , Ser, Glu, Gly ₃ , Ala ₂ , Val ₂) Lys	13
8b	Gly. Ala. Val. Thr. Pro. Val. Lys	7
10a	Ala.Ala.Ala.Lys	4
12	Ser.Arg	2
12a	[(CySO ₃ H) ₂ , Ser, Tyr ₂ , Arg ₂]	7
13	Ileu.Lys	2

are being made to resolve these mixtures by other methods. For example, Fraction 10a has been isolated from Fraction 10 plus 11 by ionophoresis on paper and two peptides (8a and 8b) have been separated by ionophoresis from Fraction 8.

The discussion, thus far, has been concerned with the acid-soluble peptides resulting from tryptic digestion of oxidized papain. The insoluble residue, which is more difficult to fractionate by chromatography, has not been completely resolved. This crude fraction is of great interest since it appears to contain the sulfhydryl group which reacts with iodoacetamide.

A sample of papain has been treated with iodoacetamide, dialyzed free of unreacted reagent, and oxidized with performic acid as before. The oxidized protein was then digested with trypsin and chromatographed with the eluting system shown above. The chromatogram obtained resembled the one shown in Fig. 4. Each fraction from the *soluble* portion of the digest was hydrolyzed with acid and examined for the presence of carboxymethylcysteinesulfone. None of these fractions, pure or impure, known to contain cysteic acid contains this derivative. However, the *insoluble* fraction contained both cysteic acid and the sulfone. These preliminary findings provide strong evidence that iodoacetamide does not react in a random manner with all the sulfhydryl groups of papain. On the contrary, the results indicate that only a single sulfhydryl group in the papain molecule reacts with the alkylating agent. Attempts to isolate and identify the peptide containing the carboxymethylcysteinesulfone have, so far, not been successful. The insoluble residue appears to contain several components and the limited solubility of these peptides in acid solution presents a particularly difficult problem of resolution.

In concluding the present phase of this presentation, it may be noted that the problem of dealing with the tryptic peptides from a protein as large as papain is far from simple. Chromatography by itself has given us very few pure peptides from the tryptic digest of oxidized papain. In fact, we have been badly misled on some occasions in that good stoichiometry has been obtained on amino acid analysis, only to find that several amino end-groups were present as shown by the dinitrophenyl method or that several components could be identified by ionophoresis on paper.

DEGRADATION OF MERCURIPAPAIN BY LEUCINE AMINOPEPTIDASE

The isolation of leucine aminopeptidase in essentially homogeneous form and demonstrably free of other peptidases and of proteinases²⁴ has permitted investigations of the action of this exopeptidase on polypeptides and proteins. In brief, we have been able to show that the aminopeptidase is restricted in its hydrolytic action to peptide bonds adjacent to a free α -amino group, regardless of the size of the substrate. This has been demonstrated not only with small, synthetic, compounds²⁵ but also with the oxidized A and B chains of insulin, with zinc-free insulin, with oxidized ribonuclease and a number of other susceptible proteins.²⁶⁻²⁹ In no case, was it possible to obtain evidence of a hydrolytic action at any locus other than a peptide bond at the amino-terminal end of the substrate molecule. As an illustration of these findings, it may be noted that the aminopeptidase has been successfully used to determine the sequence of the first six residues at the *N*-terminal end of the oxidized B-chain of insulin.²⁹

The known specificity of the aminopeptidase permits use of the enzyme, not only for sequence studies, but for other types of investigations as well. Differentiation between glutamyl and aspartyl residues from their respective ω -amides is simple since the enzyme acts only on α -peptide or α -amide bonds. Moreover, when an optically active amino acid residue is liberated, it must be of the *L* configuration, since the aminopeptidase does not release *N*-terminal *D* residues. This specificity also allows the enzyme to be used for resolution of *DL*- α -amino acid amides. It may be emphasized that we have been unable to find any measurable transferase or synthetic action by the enzyme,³⁰ findings which insure unequivocal results in structural studies.

The sharp specificity of the aminopeptidase permits studies of the relationship between the structure of the amino-terminal end of a susceptible polypeptide or protein and the biological activity of the substance. The present section is devoted to a study of the action of the aminopeptidase on mercuripapain.

Active papain cannot be employed as a substrate for the aminopeptidase, since the latter is rapidly digested and inactivated by the proteinase;³¹ however, mercuripapain is essentially inert. It has already been reported briefly^{26,32}

that the aminopeptidase liberates a large quantity of amino acids from mercuripapain, the extent of degradation depending on the ratio of enzyme to substrate. The data of Table 5 indicate that as much as two thirds of the amino acid residues of mercuripapain can be liberated. What is most striking is that in each case the enzyme, after removal of mercury, still retains its activity towards benzoyl-L-argininamide (Table 5). These results indicate

Table 5

DEGRADATION OF MERCURIPAPAIN BY LEUCINE
AMINOPEPTIDASE

Mg⁺⁺-activated aminopeptidase (*E*) was allowed to act on mercuripapain (*S*) for 24 hours at the pH indicated. The molar ratio of *S* to *E* is indicated. Extent of degradation was estimated by the increment in ninhydrin color expressed as leucine equivalents. Activity was estimated with benzoyl-L-argininamide as substrate as previously described.² Activity of degraded mercuripapain, after activation, given as per cent of control before digestion. From Hill and Smith.³⁵

<i>S</i> to <i>E</i>	pH	Moles of amino acid liberated	Relative activity
150	8.0	18	100
140	8.5	27	88
110	8.0	23	100
110	8.6	36	100
40	8.5	50	104
10	8.5	90	89
5	8.5	120	99

that the active site of papain must be remote from the *N*-terminal region of the molecule.

In documenting the above observations, it is obviously important to demonstrate that only amino acids are liberated and that these are derived entirely from the *N*-terminal end of mercuripapain, as expected from the specificity of the aminopeptidase and from experience with other polypeptides and proteins. Secondly, it is desirable to study the residual mercuripapain itself to determine that the active molecule differs from the original protein. Thirdly, we have attempted to ascertain whether the specificity of the degraded papain is altered. It may be recalled that Perlmann³³ found that the activity of pepsin fragments, obtained by autolysis, differed for protein and peptide substrates as compared with the intact enzyme.

Free amino groups. As already noted above, the DNP amino acids found after hydrolysis of intact DNP-mercuripapain are one equivalent of DNP-isoleucine and 8 of ϵ -DNP-lysine. Table 6 gives the findings with mercuripapain hydrolyzed by the aminopeptidase with different numbers of residues released. It is evident that the mercuripapain must be hydrolyzed at

its *N*-terminal end, both because *N*-terminal isoleucine disappears and because new end groups are found. Inasmuch as the hydrolysis by the aminopeptidase proceeds in a stepwise manner, it is to be anticipated, as the results in Table 6 demonstrate, that a more and more complex mixture of end groups will appear as more breakdown occurs. For this reason, study of end groups formed beyond the level indicated in Table 6 has not been pursued.

Table 6

AMINO END GROUPS OF DEGRADED MERCURIPAPAIN

Preparation	Extent of degradation	Treatment of DNP-protein	DNP-amino acids identified	Moles per mole of protein
I	21	6 <i>N</i> -HCl; 110°; 18 hours	DNP-Asp DNP-Ala DNP-Phe ε-DNP-Lys	* * * 6·9
		Oxidized, 6 <i>N</i> -HCl; 110°, 18 hours	DNP-CySO ₃ H ε-DNP-Lys	0 6·3
		12 <i>N</i> -HCl; 16 hours; 110°	DNP-Gly DNP-Pro	trace 0
II	28	6 <i>N</i> -HCl; 24 hours; 110°	DNP-Asp DNP-Ala DNP-Phe DNP-Arg	0·08 0·04 0·21 0·50
III	36	6 <i>N</i> -HCl; 24 hours; 110°	DNP-Asp DNP-Val DNP-Ser DNP-Glu DNP-Ala	0·38 0·31 0·18 trace trace

* indicates present, but not quantitatively estimated.

Amino acids liberated. At various stages of hydrolysis of mercuripapain by the aminopeptidase, amino acid analyses have been performed. The free amino acids were collected by dialysis and analyzed on ion-exchange columns by the older method of Moore and Stein¹⁸ or by a modification of the newer method of Spackman, Moore and Stein.³⁴ In no case was there any evidence for the presence of unexpected peaks which might be due to liberated peptides. Representative results at two stages of degradation are given in Table 7. Other results are presented in Table 8 and will be discussed below.

It should be emphasized that at no point does the relative composition of the liberated amino acids resemble that of papain or of the aminopeptidase. The data indicate that the composition of the *N*-terminal portion of

Table 7

AMINO ACIDS LIBERATED FROM MERCURIPAPAIN BY
AMINOPEPTIDASE

Preparation I=approximately 19 residues released; Preparation II=approximately 72 residues released

Amino acid	Moles per mole of mercuripapain		Whole papain
	Preparation I	Preparation II	
Cysteic acid	+	1.0	6
Aspartic acid	1.0	4.1	17
Threonine	0.9	2.8	7
Serine and asparagine	2.0	7.1	11
Glutamic acid	1.0	9.2	17
Glycine	2.1	6.7	23
Alanine	2.0	9.2	14
Valine	1.8	4.5	15
Isoleucine	1.0	6.4	9
Leucine	1.3	6.0	9
Tyrosine	1.3	2.7	17
Phenylalanine	0.7	1.2	4
Arginine	1.0	4.4	9
Lysine	1.0	6.6	8
Proline	+	+	9
Tryptophan	0.0	[1.0]	5

the papain molecule is different from that of the C-terminal region. It is striking that much of the threonine, alanine, leucine and isoleucine are in the first 72 residues, whereas most of the tyrosine and valine residues appear to be in the remainder of the molecule. It is particularly noteworthy that only one of the six cysteine residues has been located in the N-terminal region. Obviously this cannot be the enzymically active residue.

It is noteworthy that when 19 residues are released, there appears to be excellent stoichiometry, indicating a rather uniform hydrolysis of all molecules. In contrast to this, when approximately 72 residues are released, the breakdown is no longer uniform in accord with results obtained by the DNP method on residual mercuripapain.

Although we have performed analyses of the liberated amino acids beyond the stage given in Table 7, we are uncertain as to the significance of these results at the present time. There is some suspicion that slight activation of mercuripapain may occur in the presence of the large amounts of aminopeptidase required for such extensive degradation. Moreover, the large amounts of free amino acids may cause also activation of the proteinase. Until such results are adequately controlled we prefer to present only the results of experiments in which it is certain that such complicating factors do not occur.

One of the best methods of checking the results of degradation experiments is to be able to add up, for a given experiment, the amino acids released and the amino acids found in the undegraded residual mercuripapain. Thus far, only two such experiments have been performed. In Experiment I, the liberated amino acids were collected by dialysis and

Table 8

ANALYSIS OF AMINO ACIDS LIBERATED AND THOSE REMAINING IN MERCURIPAPAIN AFTER HYDROLYSIS WITH AMINOPEPTIDASE (LAP)

In Experiment I, the undegraded residual mercuripapain was collected by isoelectric precipitation. In Experiment II, the degraded material was isolated by the chromatographic procedure shown in Fig. 5.

Amino acid	Experiment I		Experiment II		Sum I	Sum II	Composition of Papain
	Liberated by LAP	Left in fragment	Liberated by LAP	Left in fragment			
Cysteic acid	1.1	4.8	0.9	5.0	5.9	5.9	6
Aspartic acid	2.7	13.6	5.7	12.1	16.3	17.8	17
Threonine	1.7	4.7	2.5	4.3	6.4	6.8	7
Serine	2.9	8.2	2.4	7.8	11.1	10.2	11
Glutamic acid	2.4	12.6	6.0	11.2	15.0	17.2	17
Proline	—	—	(2.9)*	6.1	—	(9.0)	9
Glycine	4.4	15.6	5.1	18.3	20.0	23.4	23
Alanine	4.3	9.2	5.1	8.3	13.5	13.4	13
Valine	3.1	12.7	2.8	9.2	15.8	12.0	15
Isoleucine	1.9	7.0	2.4	4.9	8.9	7.3	9
Leucine	2.5	7.9	3.5	5.8	10.4	9.3	9
Tyrosine	1.9	10.3	—	—	12.2	—	17
Phenylalanine	1.0	3.2	1.1	2.6	4.2	3.7	4
Arginine	2.0	7.8	2.0	6.0	9.8	8.2	9
Lysine	1.3	5.6	4.2	5.6	6.9	9.8	8
Histidine	0.0	1.0	0.0	0.8	1.0	0.8	1
Tryptophan	0.7	—	—	—	—	—	5
Total Residues	34	124	47	108	158	155	180
Corrected Total					172†	178†	

* Assumed value.

† For comparison with intact papain, the total number of tyrosine, tryptophan and proline residues has been assumed to be the same as in the original papain.

analyzed as described above. A separate aliquot was oxidized with performic acid for determination of cysteic acid. The residual mercuripapain was then precipitated isoelectrically, oxidized with performic acid, hydrolyzed with 6 N-HCl in the usual manner, and the hydrolysate was analyzed. The results are given in Table 8.

More recently, a method has been developed for separating chromatographically, inactivated aminopeptidase from mercuripapain or from degraded mercuripapain.³⁵ The chromatography is shown in Fig. 5. The main

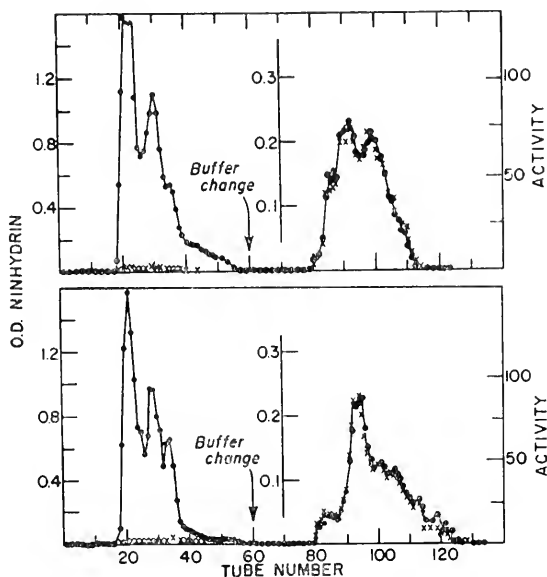


Fig. 5. Chromatographic separation of the proteins in an inactivated preparation of leucine aminopeptidase from mercuripapain (upper figure) or degraded mercuripapain (minus approximately 50 residues) (lower figure). A 1.3×40 cm. column of IRC-50 was used at pH 6.3, initially with 0.1 M phosphate buffer and then with 0.2 M phosphate. (Hill and Smith, unpublished.)

feature of the chromatography on IRC-50 is that inactivated aminopeptidase and contaminating proteins of the aminopeptidase preparation are readily eluted with 0.1 M phosphate buffer at pH 6.3, whereas the mercuripapain or degraded material is quantitatively absorbed. Phosphate buffer at 0.2 M is then used to displace the mercuripapain. Although neither the inert mercuripapain nor the degraded material shows symmetrical peaks, it is evident that there is excellent agreement between the curves for relative activity and for ninhydrin color. This, in effect, demonstrates the absence of significant amount of inert material.

The contents of the tubes containing the degraded material were pooled and worked up by conventional methods. The mixture was then hydrolyzed with 6N-HCl and oxidized with performic acid. Unfortunately, this procedure destroys the tryptophan and much of the tyrosine, but values were obtained for the other amino acids. Values for tyrosine and tryptophan will have to be estimated in other experiments.

In order to compare the results of Experiment II on the hydrolysate of residual papain with the analyses for the liberated amino acids, it has been

198 EMIL L. SMITH, ROBERT L. HILL AND J. R. KIMMEL [11 assumed, as found earlier (Table 7), that only one residue of cysteic acid is found among the oxidized free amino acids and five cysteic acid residues remain (after oxidation) in the residual material. The results of both experiments at two levels of degradation, which are given in Table 8, indicate that, within experimental error, there is excellent agreement between the sum of the residues liberated and those recovered in the residual protein, and with the previously reported¹⁴ analysis of papain. It should be noted that the absence of methionine or its sulfone from oxidized samples, and the expected low recovery of histidine and phenylalanine are consonant with the view that all the materials analyzed are derived from the mercuripapain and not from the aminopeptidase.

Specificity. One of the major reasons why the intensive study of papain was undertaken is that a variety of synthetic substrates of known structure can be studied.² It would appear from general experience that the specificity of enzyme-substrate interaction can be more readily studied under these circumstances than when the enzyme is limited in its action to one or a few compounds.

It was of great interest to study the relative specificity of papain at various stages of degradation.^{32,35} The results with several synthetic substrates of widely differing structure are given in Table 9 and the data obtained with

Table 9
SPECIFICITY OF DEGRADED PAPAIN

A sample of Mg⁺⁺-activated aminopeptidase (21 mg., C₁=60) was incubated at 40° and pH 8.6 with 14.3 mg. of mercuripapain. Aliquots were removed at various intervals and the extent of degradation was estimated by the ninhydrin method. Other aliquots were tested with the substrates shown. Each test mixture contained 0.05 M substrate, 0.005 M cysteine, 0.001 M ethylenediaminetetraacetate and 0.05 M acetate buffer at pH 5.2. Hydrolysis was estimated by titration with 0.01 N-KOH in 90 per cent ethanol. The C₁ was calculated on the basis of initial papain concentration. Optically active substrates were of the L configuration.

Residues released per mole of papain	C ₁				
	Bz.Arg.NH ₂	Cbz.Glu(NH ₂) ₂	Cbz.Leu.NH ₂	Cbz.His.NH ₂	Bz.Gly.NH ₂
0	2.0	0.59	0.048	0.033	0.044
29	2.1	0.56	0.048	0.030	0.042
55	1.8	0.56	0.047	0.024	0.035
70	2.2	0.52	0.048	0.040	0.042

human serum albumin as substrate are shown in Fig. 6. These findings demonstrate that, within experimental error, no change in substrate specificity occurs when as many as 70 of the 180 residues of papain are removed

from the enzyme. In essence, these results support the view, expressed earlier,³⁶⁻³⁸ that substrate specificity and catalytic effectiveness are part of the same phenomenon, namely, that each locus of interaction on the enzyme, which contributes to the specificity, plays a role in the catalytic transformation of the substrate. Perhaps this is just another way of stating that the key to enzymic action resides in the formation of the enzyme-substrate complex, a problem which will be also discussed below.

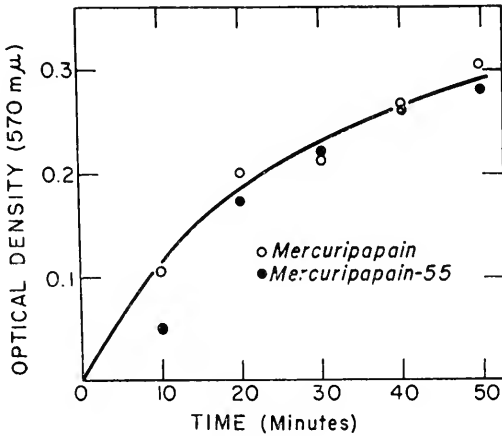


Fig. 6. Action of papain and degraded papain (minus approximately 70 residues) on human serum albumin tested at 40° and in 0.005 M dimercaptopropanol at pH 5.2 in acetate buffer. (Hill and Smith, unpublished.)

AUTODIGESTION OF PAPAINE

Recent studies³⁹ have shown that crystalline mercuripapain or papain can be readily chromatographed on IRC-50 in neutral or slightly acidic 0.2 M phosphate buffers. At pH 7.0 the protein appears to be homogeneous, whereas at pH 6.2 small amounts of material are found which appear in advance of the main active component. On occasion, some of the faster material showed some proteolytic activity, particularly when stored papain was used for the chromatography. Inasmuch as we had known that stored papain undergoes some autodigestion, a solution of active papain was allowed to autolyze for a period. When this solution was chromatographed, as shown in Fig. 7, it was found that the original active peak had diminished in quantity and other, more rapidly eluted, fractions appeared. It is striking, however, that a considerable portion of the proteolytic activity emerged with the papain fragments well in advance of the peak of intact papain.

The yield of active fragments is small, as might be anticipated, since active papain can digest other molecules at a variety of sites. The aforementioned results with the aminopeptidase digestion indicate that the active region is not at the *N*-terminal portion of the papain molecule. Hence it

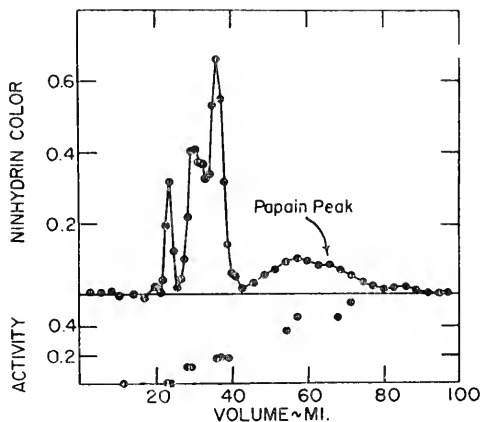


Fig. 7. Chromatography of a partially autodigested preparation of papain in 0.2 M phosphate buffer at pH 6.1 on a 1.3 × 30 cm. column of IRC-50. Activity (in arbitrary units) of individual fractions towards benzoyl-L-argininamide is indicated. (Finkle and Smith, unpublished.)

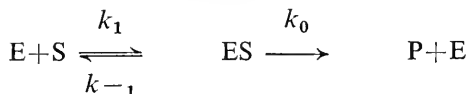
may be assumed that autolytic cleavage at the *N*-terminal region will produce active fragments of the papain molecule, whereas a primary cleavage at or near the *C*-terminal region will result in the formation of inactive fragments. Present evidence suggests that most of the fragments are enzymically inactive.

The results obtained both by digestion with the aminopeptidase and by autodigestion indicate that only a portion of the papain molecule is concerned both in the catalytic activity of the enzyme and in determining specificity. The findings suggest that it will be unnecessary to determine the complete structure of the protein in order to understand the properties and mechanism of action of the enzyme.

KINETIC STUDIES

Evidence concerning the nature of the active region of an enzyme can frequently be derived from a study of the kinetic behavior of an enzyme-substrate system. For this reason, detailed kinetic studies with papain and a variety of synthetic substrates are being performed. Only one phase of these investigations will be discussed here since they have yielded some information on the enzyme groupings which participate in the enzyme-substrate interaction.

The usual formulation of the kinetics of an enzyme reaction is



where E is enzyme, S substrate, P products and ES the enzyme-substrate

complex. K_m and k_0 can be determined by the conventional methods of estimating the initial velocity of hydrolysis over a range of substrate concentrations.

It should be recalled that the complex constant $K_m = (k_0 + k_{-1})/k_1$. It has been shown for some enzymes that $k_0 \gg k_{-1}$. In these cases, $K_m = k_0/k_1$. Our studies⁹ have provided considerable evidence that this is the situation for papain. Therefore, by determining K_m and k_0 , we can calculate the value of k_1 , the velocity constant for the formation of the enzyme-substrate complex. The constants have been determined over a broad pH range at three temperatures for the hydrolysis of benzoyl-L-argininamide.⁹ Let us consider only the calculated values of k_1 which are shown in Fig. 8 as relative values arbitrarily superimposed at their maxima. It is evident that all of the data have the same general form at the acid pH values but the descending curves on the alkaline side have been displaced at the different temperatures.

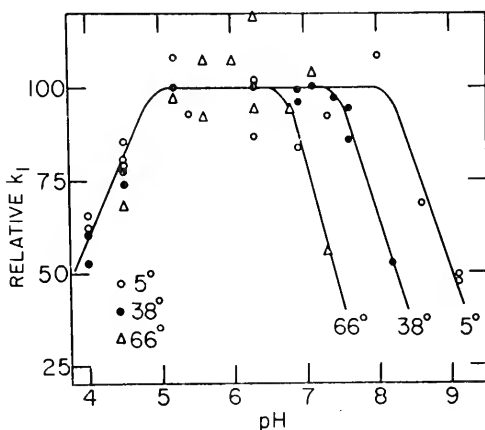


Fig. 8. Effect of pH on k_1 for the hydrolysis of benzoyl-L-argininamide. The average k_1 for the maximum was set equal to 100 for each temperature. (From Stockell and Smith (9).)

These data may be readily explained on the basis of the assumption that the descending portions of the curves represent the titration of groups on the enzyme which are essential for the k_1 step. The pH corresponding to k_1 at half of the maximal value is near pH 3.9, which would represent the pK' of the group being titrated. This pK' value is consistent with that of a carboxyl group. The absence of any appreciable effect of temperature on this pK' value is consistent with the knowledge that the heat of ionization of a carboxyl group is small.⁴⁰

The alkaline portions of the k_1 may be interpreted in a similar manner but it is apparent that the pK' of the group being titrated varies markedly with temperature. The apparent heat of ionization is approximately 9.5 kilocal. per mole at 0°. Titratable groups present in proteins which, at first

sight, might be responsible are an α - or ϵ -ammonium groups, an imidazolium, a phenolic or a sulfhydryl group. It is very unlikely that the strongly cationic substrate, benzoylargininamide, could interact with a cationic site on the enzyme. Moreover, the calculated pK' values are much too low for an ϵ -ammonium or a phenolic group, and too high for an imidazolium group. Papain possesses only one α -amino group, but, as already emphasized, much of the amino-terminal sequence is unessential for the activity of the enzyme. These considerations suggest that the α -amino group does not participate in determining the value of k_1 .

In effect, the only remaining group to be considered is the sulfhydryl group and it is already known that such a group is essential for papain activity. It may be noted that the pK' values are in the range recently found for sulfhydryl groups,⁴¹ but the calculated heat of ionization appears to be rather high. However, these data are rather limited in the critical regions where pH has a marked effect on k_1 , and the precision is rather low. In subsequent studies, the effect of pH on the hydrolysis of other substrates has been determined and, now knowing the important range to be investigated, more complete data have been obtained.

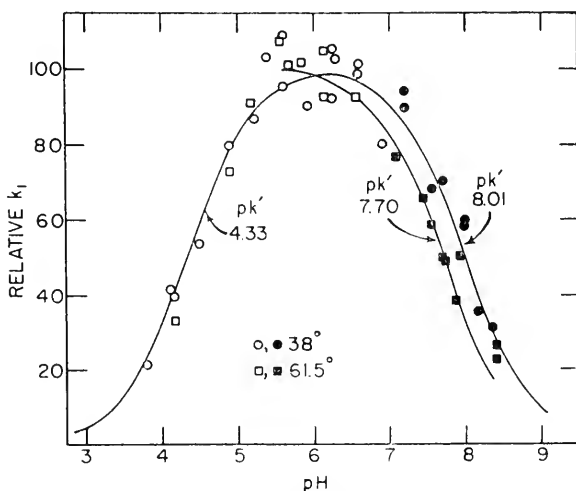


Fig. 9. Effect of pH on k_1 for the hydrolysis of carbobenzyloxy-L-histidinamide with the maximum for each temperature set equal to 100. (From Smith, Chavre and Parker (42).)

Fig. 9 shows the effect of pH on calculated values of k_1 for the hydrolysis of carbobenzyloxy-L-histidinamide.⁴² The data have been fitted by drawing theoretical titration curves for a single titratable group at both sides of the curve. On the acid side of the curve, there is little effect of temperature and the apparent pK' at 38° is near 4.3. On the alkaline side, there is a shift of pK' with temperature and the calculated ΔH , the heat of ionization, is about 4.4 kcal. per mole at 0°. Both the pK' value and the value of ΔH are near

those expected for a sulfhydryl group.⁴¹ Studies at 38° of the effect of pH on k_1 for the hydrolysis of benzoylarginamide⁴² and of benzoyl-L-arginine ethyl ester⁴³ yield curves with the same shape as those shown for carbobenzoxy-L-histidinamide.

The effect of pH on k_1 for another type of substrate, carbobenzoxyglycylglycine, is shown in Fig. 10.⁴² On the alkaline side, the curve has been

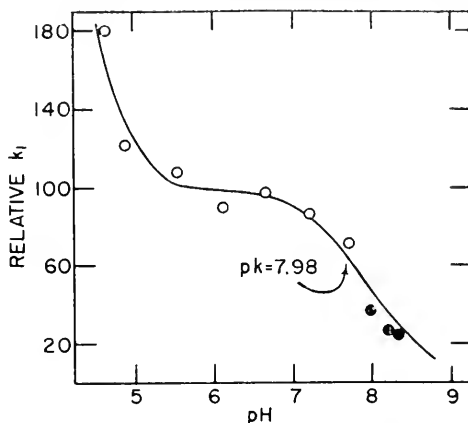


Fig. 10. Effect of pH on k_1 for hydrolysis of carbobenzoxyglycylglycine at 38°. (From Smith, Chavre and Parker (42).)

fitted by a theoretical titration curve with a pK' near 8.0 which is close to that found for other substrates at 38°. However, there is a plateau near pH 6 and k_1 rises sharply at more acid pH values. It should be emphasized that at pH 6, the absolute value for k_1 for the scission of carbobenzoxyglycylglycine is only about 0.1 per cent of that for benzoylargininamide. At pH 6, both the enzyme and carbobenzoxyglycylglycine would have carboxyl groups in ionized form and electrostatic repulsion may be expected. The increasing values of k_1 below pH 5.5 would appear to be due to repression of ionization of one or both carboxyl groups, thus decreasing the degree of repulsion between enzyme and substrate and, in effect, promoting an increase in the value of k_1 .

The kinetic data which have been cited, point to the participation of the essential sulfhydryl group in the k_1 step. Moreover, since the optimal region of k_1 is near pH 6, it is evident that the active form of this group is RSH and not RS^- . In addition, evidence has been obtained that an ionized carboxyl group is involved in the k_1 step. Inasmuch as the kinetic studies have been performed with rather small substrates, it is likely that the carboxyl group must be close to the sulfhydryl group.

A TENTATIVE REACTION MECHANISM

Thus, the evidence from kinetic and other studies indicates that the essential sulfhydryl group participates in the enzyme-substrate interaction and points

to an important role of an ionized carboxyl group in this reaction. A tentative reaction mechanism for the action of papain has been suggested.^{4,9} The main feature of this scheme is that a thiol ester is formed in the enzyme substrate complex.

Briefly, the following types of observations on papain must be considered.

1. The most sensitive known substrates are those which contain the carboxyl group of an arginyl or lysyl residue linked to the sensitive bond; i.e., the compounds have a strongly cationic group on the side chain.
2. Papain can hydrolyze synthetic substrates which possess ester² or thiol ester⁴⁴ bonds, as well as those with peptide or amide bonds.
3. In addition to hydrolytic reactions, papain catalyzes synthetic and transfer reactions.^{3,45}
4. There is one sulfhydryl group which is absolutely essential for the action of papain.⁵

Formation of a thiol ester is consistent with the spatial and chemical specificity of papain which is mainly toward the amino acid residue bearing the carbonyl group of the sensitive peptide bond. The moiety which is displaced from or transferred to the carbon of this carbonyl group may be any one of a variety of compounds, e.g., a peptide, an amino acid, ammonia, aniline, hydrazine, hydroxylamine, an alcohol, a thiol, etc.²

Postulation of a thiol ester intermediate would appear to be necessary to explain the efficiency of papain in catalysing transfer reactions, since the replacement reagent $R-NH_2$ competes with water or the ions of water for the hydrolytic reaction. The expected free energy change for the scission of the thiol ester bond would indicate that the equilibrium would greatly favor hydrolysis. It may also be emphasized, that as in the case of ficin,⁴⁶ we have found that the amide and the ethyl ester of benzoyl-L-arginine are split at about the same rate.⁴³ In fact, the k_0 values are almost identical. This suggests that the cleavage step, in the overall reaction, involves a common intermediate, such as the postulated thiol ester. In view of the usual large difference in free energy of activation for hydrolysis of similar esters and amides, the similar k_0 values are hard to explain except in terms of an identical intermediate.

Before concluding, it may be emphasized that there is one major objection to the view that a thiol ester is formed with a free thiol group of the enzyme. It is very likely that the overall free energy of formation of a thiol ester is much greater than that of the susceptible peptide or ester bond. This, in effect, would make the overall reaction scheme energetically impossible, unless the rate of scission of the thiol ester is sufficient to drive the reaction forward.

A possible explanation is that a free thiol group does not exist as such in active papain. Instead, there may be present either a thiol group bonded as a high energy hydrogen bond to the titrable carboxylate ion or that a thiol ester, formed by coupling of a carboxyl group and a thiol group, exists preformed in the papain itself. Let us consider the latter possibility first. This hypothesis is attractive insofar as the initial reaction between enzyme

and substrate would then involve a transfer reaction with no net energy change to yield an acyl amino acid thiol ester, and this would be followed by a displacement reaction reforming the thiol ester of the enzyme and liberating the acyl amino acid derived from the substrate. Thus, the major part of the enzymic reaction would consist of two transfer reactions involving essentially equivalent thiol ester exchanges resulting in the hydrolysis of the substrate.

This possibility is also interesting since it would explain the catalytic efficiency of the enzyme more readily than does the view that an ordinary thiol group is involved. However, we do not know enough, as yet, to decide whether this is more than an interesting possibility. Some of the evidence at hand does not seem to fit this hypothesis too well, but other evidence would fit very nicely indeed. Just two points may be mentioned. As noted above, the activity of papain is related to the amount of sulfhydryl which reacts rapidly in water with *p*-chloromercuribenzoate (Fig. 1). The rapidly reacting SH may be in a high energy form which is kinetically more reactive than ordinary sulfhydryl.

The other finding, which might be cited, is that for four substrates, whose kinetic constants have been measured, the value of k_0 is related to the value of k_1 . This is shown in the double logarithmic plot of Fig. 11.⁴² If we attach

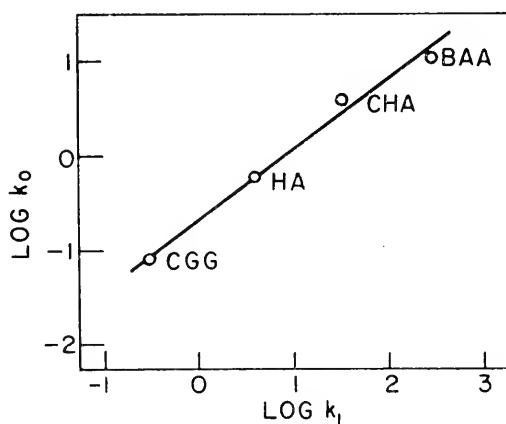


Fig. 11. A hypothetical, tentative scheme for the hydrolytic action of papain. (From Stockell and Smith (9).)

the usual meaning to these constants, the relationship suggests that the more readily ES forms, the more readily it breaks down. Such a view would be in accord with the idea that both reactions, i.e., those involving k_1 and k_0 , involve chemically similar steps, namely transfer reactions of thiol esters. It may be remarked, in passing, that the relationship illustrated in Fig. 11 has also been found for a number of related substrates for carboxypeptidase.⁴⁷

The alternative hypothesis that the enzyme contains RCOO^- linked to R'SH to form a hydrogen-bonded structure, $\text{RCOO}^- \cdots \text{HSR}'$, would readily

206 EMIL L. SMITH, ROBERT L. HILL AND J. R. KIMMEL [11 explain the effect of pH on k_1 , since addition or removal of a proton would open the hydrogen bond. Similarly, if there is a preformed thiol ester, it may be postulated that the linkage is opened by changes in pH, but it would be necessary to assume that refolding of the protein could permit regeneration of the thiol ester.

Either hypothesis is attractive since the presence of a 'high-energy' bond would aid in explaining the catalytic activity of the enzyme. Indeed, it is possible that such 'high-energy' bonds are an essential part of the active sites of enzymes and that their formation depends on the specific folding of the protein. It may be noted that the activity of papain is reversibly lost on denaturation with urea. Thus, papain, like other enzymes, depends for its activity on a specific three-dimensional structure despite the fact that a large part of its amino acid sequence is unessential for the maintenance of its activity.

CONCLUDING REMARKS

It is evident from the above summary of some of our studies with crystalline papain that only a beginning has been made in understanding how the structure of this protein is related to its function. It will take a long time before the mechanism of action and the specificity of this enzyme can be related to its structure. Nevertheless, it is satisfying that, in our approach to the biological activity of the proteins, we now have the general knowledge that considerable modification of structure is possible without alteration of the activity or specificity of antibodies, protein hormones, enzymes, and viruses. Above all else, the observations made on these substances indicate the general validity of the concept of the active site, namely, that only a part of the structure of a protein molecule is responsible for and involved in its specific interactions. Future progress in the isolation and identification of such active sites or regions of proteins will depend as much on our choice of proteins for study, as on the available tools and methods. The proteolytic enzymes will continue to play a dual role in such studies, since, as this symposium illustrates, everyone must use these enzymes in his working methods with other proteins. At the same time, they are among the more favorable materials for investigation.

REFERENCES

1. C. B. ANFINSEN and R. R. REDFIELD, *Advances in Protein Chem.*, **11**, 1 (1956).
2. J. R. KIMMEL and E. L. SMITH, *J. Biol. Chem.*, **207**, 515 (1954).
3. J. R. KIMMEL and E. L. SMITH, *Advances in Enzymol.*, **19**, 267 (1957).
4. E. L. SMITH, B. J. FINKLE and A. STOCKELL, *Faraday Society Discussions*, **20**, 96 (1955).
5. B. J. FINKLE and E. L. SMITH, *J. Biol. Chem.*; in press.
6. E. L. SMITH, J. R. KIMMEL and D. M. BROWN, *J. Biol. Chem.*, **207**, 553 (1954).
7. E. O. P. THOMPSON, *J. Biol. Chem.*, **207**, 563 (1954).

8. R. K. BROWN, B. V. JAGER and E. L. SMITH, unpublished observations.
9. A. STOCKELL and E. L. SMITH, *J. Biol. Chem.*, **227**, 1 (1957).
10. P. D. BOYER, *J. Am. Chem. Soc.*, **76**, 4331 (1954).
11. B. J. FINKLE and E. L. SMITH, *Federation Proc.*, **16**, 180 (1957).
12. A. K. BALLS and H. LINEWEAVER, *J. Biol. Chem.*, **130**, 669 (1939).
13. J. R. KIMMEL, E. O. P. THOMPSON and E. L. SMITH, *J. Biol. Chem.*, **217**, 151 (1955).
14. E. L. SMITH, A. STOCKELL and J. R. KIMMEL, *J. Biol. Chem.*, **207**, 551 (1954).
15. J. CLOSE, S. MOORE and E. J. BIGWOOD, *Enzymologia*, **16**, 137 (1953).
16. E. O. P. THOMPSON, *Proc. Int. Wool Textile Research Conference, Australia, C*, 102 (1956).
17. F. SANGER, *Advances in Protein Chem.*, **7**, 1 (1952).
18. S. MOORE and W. H. STEIN, *J. Biol. Chem.*, **176**, 367 (1948).
19. J. R. SPIES and D. C. CHAMBERS, *Anal. Chem.*, **22**, 1447 (1950).
20. J. R. KIMMEL and E. L. SMITH, unpublished observations.
21. E. L. SMITH and J. R. KIMMEL, *Proc. Int. Wool Textile Research Conference, C*, 199 (1956).
22. W. J. DREYER and H. NEURATH, *J. Biol. Chem.*, **217**, 527 (1955).
23. S. MOORE, C. H. W. HIRS and W. H. STEIN, *Federation Proc.*, **15**, 840 (1956).
24. D. H. SPACKMAN, E. L. SMITH and D. M. BROWN, *J. Biol. Chem.*, **212**, 255 (1955).
25. E. L. SMITH and D. H. SPACKMAN, *J. Biol. Chem.*, **212**, 271 (1955).
26. R. L. HILL and E. L. SMITH, *Federation Proc.*, **14**, 226 (1955).
27. E. L. SMITH and R. L. HILL, *IIIe Congr. intern. biochim.* (Brussels), 1955.
28. E. L. SMITH, *Proc. Int. Wool Textile Research Conference, C*, 407 (1955).
29. R. L. HILL and E. L. SMITH, *J. Biol. Chem.*, **228**, 577 (1957).
30. R. L. HILL and E. L. SMITH, *J. Biol. Chem.*, **224**, 209 (1957).
31. R. L. HILL, unpublished observations.
32. R. L. HILL and E. L. SMITH, *Biochim. et biophys. acta.*, **19**, 376 (1956).
33. G. E. PERLMANN, *Nature*, **173**, 406 (1954).
34. D. H. SPACKMAN, S. MOORE and W. H. STEIN, *Federation Proc.*, **15**, 358 (1956).
35. R. L. HILL and E. L. SMITH, *J. Biol. Chem.*; in press.
36. E. L. SMITH, *Proc. Nat. Acad. Sc.*, **35**, 80 (1949).
37. E. L. SMITH, *Federation Proc.*, **8**, 581 (1949).
38. E. L. SMITH and R. LUMRY, *Cold Spring Harbor Symposia Quant. Biol.*, **14**, 168 (1950).
39. B. J. FINKLE and E. L. SMITH, unpublished observations.
40. E. J. COHN and J. T. EDSALL, *Proteins, Amino Acids, and Peptides*, Reinhold, New York (1943).
41. R. E. BENESCH and R. BENESCH, *J. Am. Chem. Soc.*, **77**, 5877 (1955).
42. E. L. SMITH, V. CHAVRÉ and M. J. PARKER, *J. Biol. Chem.*; in press.
43. E. L. SMITH and M. J. PARKER, unpublished observations.
44. R. B. JOHNSTON, *J. Biol. Chem.*, **221**, 1037 (1956).
45. J. DURELL and J. S. FRUTON, *J. Biol. Chem.*, **207**, 487 (1954).
46. S. A. BERNHARD and H. GUTFREUND, *Biochem. J.*, **63**, 61 (1956).
47. R. LUMRY and E. L. SMITH, *Faraday Society Discussions*, **20**, 105 (1955).

Ribonuclease

Studies on the structure of ribonuclease

C. H. W. HIRS, WILLIAM H. STEIN AND
STANFORD MOORE

The Rockefeller Institute for Medical Research, New York 21, N.Y.

Rather than attempt a review of published information that deals with the structure of ribonuclease (cf. 1, for example), we will in the present discussion emphasize the more recent work and the methods that are currently being employed to bring the structural studies to a conclusion. Some recapitulation is desirable, however, and it can be furnished by reference to the partial structural formula for oxidized ribonuclease shown in Fig. 1, which summarizes the status of our knowledge as of early 1956.² The single chain of 124 residues in the oxidized protein is portrayed in four segments, which should be imagined as joined end to end. The bonds at which the chain is attacked by proteolytic enzymes are indicated by the vertical lines: solid lines for trypsin, dashed lines for chymotrypsin, and dotted lines for pepsin. The residues included within the parentheses are of undetermined sequences.

The derivation of the formula shown in Fig. 1 owes very much to the stimulating and pioneering investigations of Sanger and his colleagues on the structure of insulin (cf. 3 for references). Several of the steps in the derivation were patterned directly after his work; others had to be changed to a greater or lesser extent depending upon the nature of the special problems encountered. A sample of ribonuclease containing 93 per cent of the A component was employed for most of the structural study. The fact that ribonuclease contains but a single peptide chain, and that this chain bears the sequence Lys.Glu.Thr.Ala- at the amino-end, was established by Anfinsen and his colleagues⁴ by the use of the DNP technique. The first step in our own studies was an amino acid analysis that showed that ribonuclease A contained 124 to 126 amino acid residues.⁵ This was followed by oxidation of the single chain at low temperature with performic acid⁶ under conditions chosen so as to avoid the formation of chlorotyrosine.

The four disulfide bonds in the molecule were thus oxidized to eight sulfonic acid groups, and the four methionine residues to the sulfone. The other amino acids were not affected. The oxidized protein was subjected to

tryptic hydrolysis at pH 7 and 25°, the peptides produced in the hydrolysis were separated on a preparative scale on columns of Dowex 50-X2, and their amino acid compositions were determined quantitatively on columns of Dowex 50-X4.⁷ The thirteen principal peptides, formed in yields of between 50 and 100 per cent, accounted for all of the 124 residues in the peptide chain. To determine the order in which these peptides were joined to one another, the products of chymotryptic and peptic hydrolyses were similarly isolated and their amino acid composition determined.^{2,8} From the data so obtained, it was possible to deduce the arrangement shown in Fig. 1 provided three assumptions were made. These were, first, that there is no branching of the peptide chain, and that all of the residues are connected to one another by peptide bonds between α -amino and α -carboxyl groups; second, that in peptides formed by the action of trypsin, the basic amino acids, arginine or lysine, if present, occupy the carboxyl-terminal position; and third, that in peptides formed by the action of chymotrypsin, a tyrosine or phenylalanine residue, if present, occupies the carboxyl-terminal position.

The partial structural formula appears to represent a unique synthesis of the results, in that, within the restrictions imposed by the assumptions made in its derivation, no other arrangement of the peptides has been found possible. Moreover, this formula accommodates all of the data secured thus far. Every peptide that has been characterized appears in the scheme of proteolytic cleavages shown in Fig. 1.

This partial structural formula already reveals much about the structure of ribonuclease, but its main use has been to serve as a point of departure for three relatively independent lines of research. These are: one, the deduction of the sequence of the amino acid residues along the chain; two, the determination of the location of the four disulfide links connecting the eight half-cystine residues; and, three, the localization of the 'active center' of ribonuclease, if indeed, one may hope that such a unique and relatively small catalytic site can be found in this large molecule. The results of all three lines of investigation undoubtedly bear on one another, but the present discussion will be limited to the first two of them.

The problem posed by the sequence of the amino acid residues in ribonuclease can perhaps best be considered in connection with Fig. 2, which shows the sequences as they are known at the present. The location of 90 of the 124 residues in the chain has so far been established (cf. 9), and these are underlined in the formula. Work is currently in progress to fill in the remaining gaps in the sequence.

In the sequence work, it was first necessary to prepare in good yield a series of relatively small peptides which collectively represented all of the residues in the chain of oxidized ribonuclease. The partial structural formula was of great value in pointing the way. Because this formula showed the manner in which the forty-odd peptides formed by enzymatic hydrolyses are

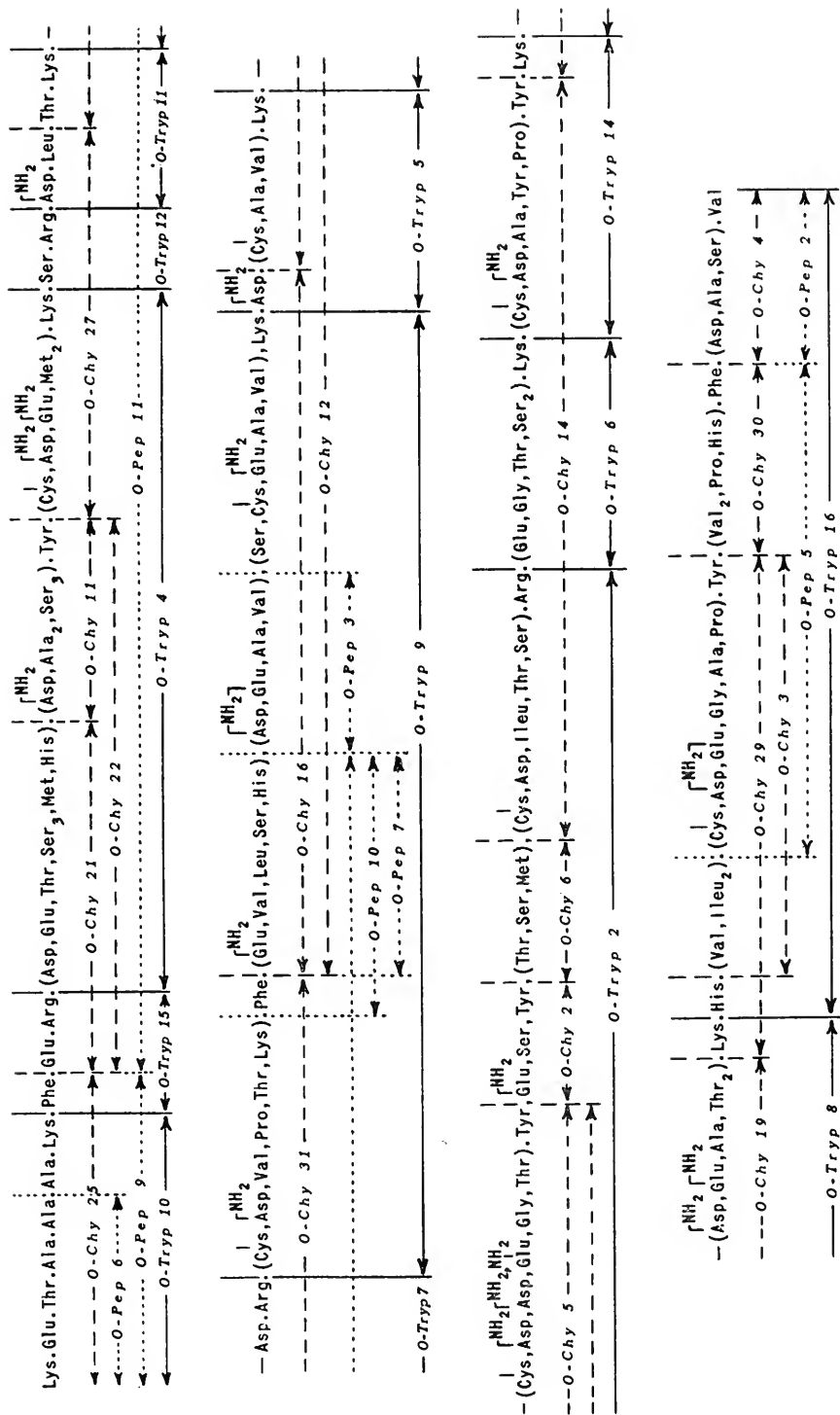


Fig. 1. A partial structural formula for oxidized ribonuclease.² The amino acids are considered to be arranged in a single chain with the segments in the figure joined end to end. The formula is based upon the study of the peptides liberated by tryptic, chymotryptic, and peptic hydrolysis of the oxidized protein. The tentative allocation of amide groups ($-\text{NH}_2$) has been based upon the most recent results, which have permitted definite placement of some of the asparaginyl and glutaminyl residues.⁹ The residues in parentheses are of undetermined sequence.

linked, detailed structural studies of all of these peptides were not required. It was possible to select for sequence work only those peptides of suitable size which do not overlap one another (Fig. 1). The starting point was the series of 13 peptides obtained by tryptic hydrolysis of oxidized ribonuclease. These peptides account for all of the ribonuclease molecule, and can be

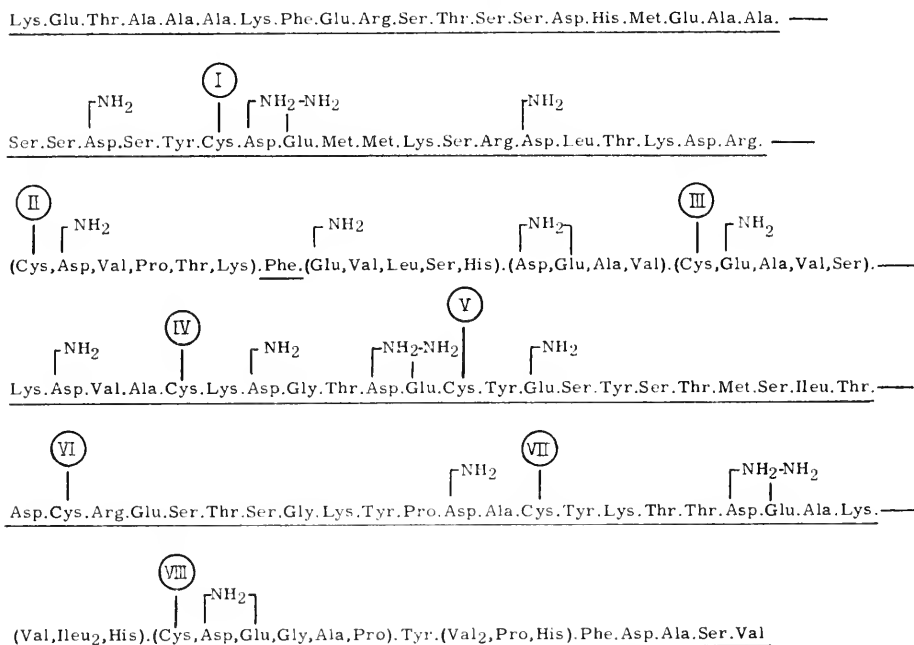


Fig. 2. Arrangement of the amino acid residues in oxidized ribonuclease.⁹ The sequences determined to date are underlined. The residues in parentheses are of undetermined order. The cysteic acid residues have been numbered consecutively, I to VIII, starting from the amino-end of the chain.

prepared in yields ranging from 50 to 100 per cent. Nine of these 13 peptides, varying in size from dipeptides to heptapeptides, account for 42 of the 124 residues in the chain, and are small enough to be studied directly. The structures of all nine have been worked out. The remaining four peptides of the trypsin series, O-Tryp 4, O-Tryp 9, O-Tryp 2, and O-Tryp 16 (Fig. 1), which contain 21, 22, 19, and 20 residues respectively, account for the rest of the molecule, but are too large to be useful, as such, for sequence studies. However, from the knowledge of the points of cleavage by chymotrypsin provided by the partial structural formula, it can be predicted that each of these four large peptides should be split individually with chymotrypsin into fragments of a more manageable size. This has been carried out, and in each instance the expected segments were obtained. For example, the action of chymotrypsin on O-Tryp 4, which contains 21 amino

acid residues, should result in the cleavage of the bonds indicated by the dashed lines with the formation of three peptides. Of these, one should be a peptide already characterized (O-Chy 11, Fig. 1) in chymotryptic hydrolysates of oxidized ribonuclease, while the other two should be new peptides with amino acid compositions in agreement with those predicted by the formula. The top curve in Fig. 3 shows the separation on a 150 cm. column of Dowex 50-X2 of the three peptides formed upon chymotryptic hydrolysis of peptide O-Tryp 4. A comparison of the amino acid composition of these peptides (given adjacent to each of the peaks on the curve) with the partial structural formula in Fig. 1 reveals that the expected fragments have indeed been isolated.

The second curve in Fig. 3 illustrates the corresponding result with peptide O-Tryp 9. In this experiment, the yields (shown in parentheses after each of the peaks on the curve) were not so favorable, because additional cleavage occurred at two bonds which were not markedly attacked when lower enzyme concentrations were used in the experiments with the intact oxidized protein. Finally, the bottom two curves in Fig. 3 show the results obtained upon chymotryptic hydrolysis of peptides O-Tryp 2 and O-Tryp 16. Once again the products predicted from the partial structural formula were formed. These experiments, therefore, not only furnished peptides suitable for further sequence work, but the results also provided strong substantiation for the deductions that led to the derivation of the partial formula.

Once 10 to 20 mg. quantities of 23 peptides of suitable size for sequence work had been prepared from the entire ribonuclease molecule, we were confronted with a more familiar type of problem, namely, that of deriving the amino acid sequence in relatively small peptides. We thus were fortunate in being able to profit from the experience of the many people who have been through this territory before us. A variety of chemical and enzymatic methods has been employed for the determination of the sequences of the amino acid residues in these peptides. The chemical methods have included the DNP procedure of Sanger for *N*-terminal residues, hydrazinolysis for the detection of carboxyl-terminal amino acids (by the Braunitzer and Schramm¹⁰ modification of the Akabori procedure), and, in particular, stepwise degradation from the amino-end by the phenyl-*iso*-thiocyanate method of Edman.¹¹ In the Edman procedure, each step of the degradation has been followed by quantitative amino acid analysis of the peptide remaining after removal of the phenylthiohydantoin.

In our hands, the most effective method for carrying out the reaction of phenyl-*iso*-thiocyanate with the free amino groups of peptides has been the one described by Ottesen and Wollenberger¹² in their work on the peptides liberated from ovalbumin during the formation of plakalbumin. Routinely, the reaction is carried out at 40° in 30 per cent aqueous dioxane, and the pH is maintained at 8.0 with the aid of a pH-stat. In this way the reaction may be followed quantitatively. The rate of the reaction is a function of the

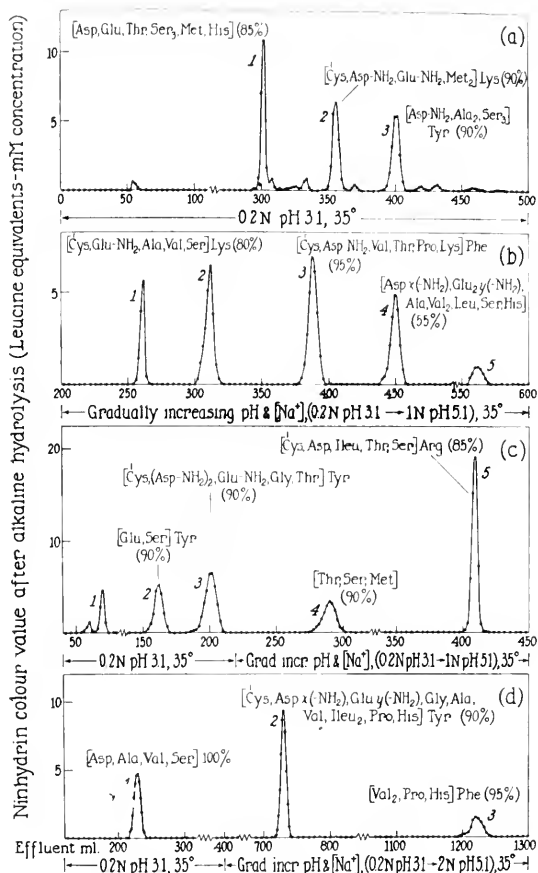


Fig. 3. The peptides in 20-hour chymotryptic hydrolysates of peptides O-Tryp 4, O-Tryp 9, O-Tryp 2, and O-Tryp 16 obtained from performic acid-oxidized ribonuclease. Chromatography of hydrolysates from approximately 5 micromoles of each peptide was carried out on 150×0.9 cm. columns of Dowex 50-X2 in the sodium form. The effluent was collected in 2 ml. fractions. Aliquots (0.1 ml.) were removed for alkaline hydrolysis prior to analysis by the ninhydrin method. The figures in parenthesis represent the yield of each peptide. The sequence of the amino acids within the parentheses is undetermined. (From Hirs, manuscript in preparation.)

(a) O-Tryp. 4. Elution of the column was with 0.2N sodium citrate buffer at pH 3.1.

(b) O-Tryp 9. Elution was with 0.2N sodium citrate buffer at pH 3.1 as far as 150 effluent ml. At this point, the pH and molarity of the influent buffer were gradually increased by the use of a mixing chamber of 610 ml. volume filled with the 0.2N buffer at pH 3.1 into which N sodium citrate-acetate buffer at pH 5.1 was allowed to flow.

(c) O-Tryp 2. The elution procedure followed that given in (b) except that the gradual change of pH and molarity of the buffer was begun after 220 effluent ml.

(d) O-Tryp 16. The elution procedure followed that given in (b) except that the gradual change of pH and molarity of the buffer was begun after 400 effluent ml. After 610 effluent ml., the reservoir on the mixing device was charged with 2N sodium citrate-acetate buffer at pH 5.1.

structure of the peptide, the presence of a large R group, such as leucine, at the amino-end being inhibitory, whereas amino-terminal glycine reacts very rapidly. In the cyclization of the phenylthiocarbamyl peptides, Edman's original method has been used, the reaction being carried out in glacial acetic acid saturated with dry hydrogen chloride at 100°. Attempts to employ aqueous acid in the cyclization step have met with varied success. With peptides containing several residues of serine or threonine not immediately adjacent to the amino-terminal group, the use of dilute aqueous acid led to partial hydrolysis of the peptide. Although the cyclization reaction in glacial acetic acid does not always proceed as rapidly as it does with some peptides in aqueous solution, significant partial hydrolysis in the non-aqueous system has not yet been observed. The rate of the cyclization is dependent upon the structure of the peptide, the presence of an aspartic, glutamic, or cysteic acid residue at the amino-terminal position being markedly inhibitory.

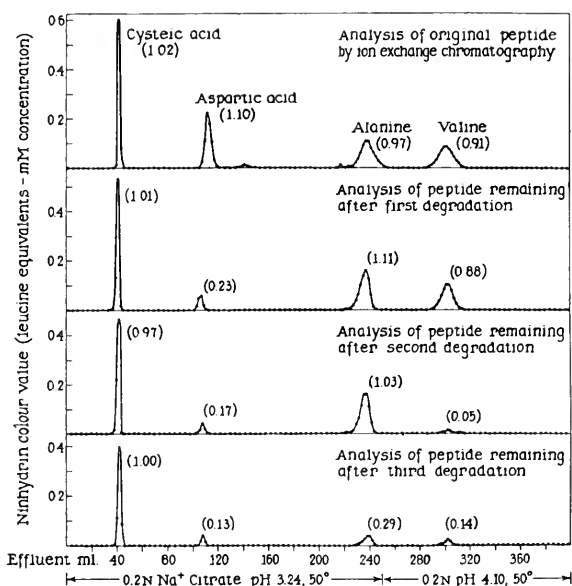


Fig. 4. Edman degradation of Asp-NH₂.Val.Ala.CySO₃H.Lys (O-Tryp 5). The chromatograms show the amino acid composition of acid hydrolysates of the peptides present at each stage of the procedure. The amino acid analyses are now performed on columns of sulfonated polystyrene resins with the aid of automatic recording equipment.¹³

In Fig. 4 is shown an example of how the Edman procedure has been used for the determination of the sequence of the pentapeptide asparaginyl-valylalanylalanylcysteicyllsine (O-Tryp 5). The top curve in Fig. 4 represents a quantitative amino acid analysis of the acidic and neutral amino acids in a hydrolysate of the pentapeptide. A resin, Amberlite IR-120, similar to Dowex 50-X8, was employed. Analyses such as these, in conjunction with Edman degradations, are currently being carried out with automatic recording

218 C. H. W. HIRS, WILLIAM H. STEIN & STANFORD MOORE [12
equipment of the type described by Spackman, Stein, and Moore.¹³ The numbers in parentheses adjacent to each peak are the molar ratios of each component, and it will be noted that the hydrolysate contains nearly equimolar quantities of cysteic acid, aspartic acid, alanine, and valine. Lysine was not eluted in these analyses; however, its presence at the carboxyl-terminal position could be inferred from the specificity of trypsin, and was confirmed by other means, as will be evident presently. After the first application of the Edman degradation, carried through with 2.5 micromoles of the peptide, the quantitative analysis shown on the second line of Fig. 4 was obtained. The molar ratios of the components are still unity, with the exception of aspartic acid, for which the value has fallen from about 1 to 0.23, thus revealing that aspartic acid or asparagine must be the amino terminal residue. After the second application of the degradation, the analysis on the third line of Fig. 4 was obtained, which showed that the molar ratios of both aspartic acid and valine are now low. Valine was thus revealed to be the second amino acid, while the third amino acid was shown to be alanine as a result of a further degradation, as indicated by the last curve of Fig. 4. The degradation, when extended to a fourth stage, indicated a loss of cysteic acid, which established the sequence aspartyl- or asparaginylvalylalanyl cysteiclylysine.

The yield of lysine obtained by acid hydrolysis of peptides that have undergone reaction with phenyl-*iso*-thiocyanate is not quantitative and falls off unaccountably as the number of steps of the Edman degradation is increased. For this reason, analyses for lysine have not been included in Fig. 4. An independent demonstration of the presence of lysine at the carboxyl-terminal position in the peptide was furnished by hydrazinolysis, which gave lysine in a yield of 70 per cent as the only free amino acid. The evidence was now sufficient to reveal completely the sequence of the residues in the peptide.

The final question was whether an aspartyl or an asparaginyl residue is present at the amino-terminal position. Analysis revealed, that on acid hydrolysis, one equivalent of ammonia was released per molecule of peptide, indicating that the amino-terminal residue is probably asparagine and not aspartic acid. This was independently confirmed when it was found that the action of the leucine aminopeptidase of Spackman, Smith, and Brown¹⁴ on the peptide gave asparagine, identified by ion-exchange chromatography, in yields of 27 and 91 per cent after 1 and 8 hours respectively of hydrolysis with the enzyme. In addition, two further peaks were present on the effluent curve; one at the valine position, and another some 25 effluent ml. behind it. They presumably corresponded to the peptides derived from the starting material by the successive removal of asparagine and valine. Similar results have been obtained with some of the other peptides examined when hydrolysis with either leucine aminopeptidase or carboxypeptidase was studied. In addition to providing conclusive evidence for the presence of the amides

of aspartic and glutamic acid, the experiments with these enzymes have in certain instances also furnished information about the sequences of some of the residues in the peptides. The interpretation of the results, however, is not always clear-cut, since, as was noted above, partially degraded peptides may move on the Amberlite IR-120 columns at rates identical with or close to those for free amino acids.

The application of these procedures to 17 of the 23 peptides isolated for sequence studies has given the detailed information shown in Fig. 2, which may now be discussed in more detail. The sequence of the first ten residues in the peptide chain was established earlier; the first four as a result of Dr Anfinsen's end-group analyses on the intact protein⁴ and the rest as a consequence of our analyses of the peptides formed by the action of trypsin and chymotrypsin.^{2,7} The sequence has now been independently confirmed by the use of carboxypeptidase, and by step-wise degradation. More recent work has revealed the sequence from the 11th to the 31st residues shown in Fig. 2. The sequence Ser.Arg.Asp(NH₂).Leu.Thr.Lys.Asp.Arg starting at the 32nd residue was previously worked out by Redfield and Anfinsen,¹⁵ and has been confirmed in the present work. Thus the arrangement of the first 39 residues in the chain is now known. The portion of the chain from the 40th to the 61st residue is accounted for by the tryptic peptide O-Tryp 9, containing 22 residues, on the sequence of which work is in progress. From the sixth lysine residue in the chain (residue 61) the sequence shown in Fig. 2 passes through the pentapeptide O-Tryp 5, which was considered in some detail before, and continues through asparagine and glycine for a total of 44 residues. The latter is the longest sequence of residues worked out to date along the peptide chain of ribonuclease. Finally, the work of Dr Anfinsen and his colleagues¹⁵ has established the sequence of the four last residues in the chain.

Concurrently with these studies, Dr Spackman in our laboratory,¹⁶ and Ryle and Anfinsen at Bethesda,¹⁷ have been working on the positions of the disulfide bonds in ribonuclease. The approach in both laboratories is similar in several respects to the procedures used so successfully by Ryle, Sanger, Smith, and Kitai,¹⁸ in the elucidation of the disulfide structure present in insulin.

Ribonuclease in which the disulfide bonds are intact is not hydrolysed in aqueous solution by either trypsin or chymotrypsin. In the presence of 2 M guanidinium chloride, however, successive hydrolysis at pH 7.5 by trypsin and chymotrypsin, followed by desalting and chromatography on Dowex 50-X2, gave the result shown on the top part of Fig. 5. Enzymatic action was carried out in the presence of *N*-ethylmaleimide to abolish or at least minimize the disulfide interchange reaction. The cystine containing peptides, labeled A to F in the curve, were estimated by the phosphotungstate procedure of Kassell and Brand,¹⁹ the results of which are shown by the open circles. The fractions in the major peaks were pooled, desalted,

220 C. H. W. HIRS, WILLIAM H. STEIN & STANFORD MOORE [12 oxidized with performic acid, and the oxidation mixture was chromatographed as before.

The results from Zone C of the top chromatogram are shown at the bottom of Fig. 5. The cysteic acid peptides formed by oxidation move well ahead of the rest of the mixture originally present in Zone C, and which has not been affected by oxidation. The fractions containing each of these two fast-moving peaks were pooled and their amino acid composition was determined quantitatively. The empirical formulae for the pair of cysteic acid peptides is shown on the curve. In the same manner, a pair of cysteic

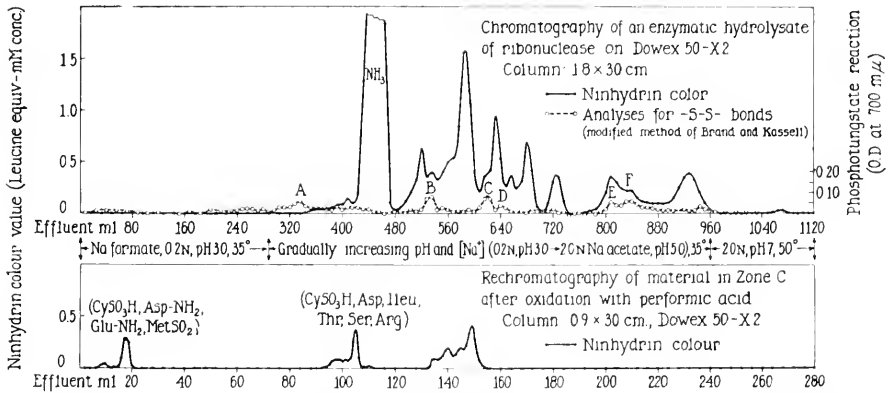


Fig. 5. Determination of the position of the disulfide bonds in ribonuclease. Upper curve: A tryptic and chymotryptic hydrolysate from 200 mg. of ribonuclease was chromatographed. Lower curve: Fractions from Zone C of upper curve were pooled, desalted, oxidized, and chromatographed. (From experiments of Spackman; cf. 16.)

acid peptides of the following composition was obtained from Zone B: (Asp.NH₂,Val,Ala,CySO₃H,Lys) and (Asp.NH₂,Thr,Asp.NH₂,CySO₃H,Glu.NH₂,Gly,Tyr). From the amino acid composition of these two pairs of peptides alone, without the necessity for any sequence work, it could be concluded that a disulfide bond links half-cystine residues I and II, and another links half-cystine residues IV and V.

There is room for improvement in the results thus far, for the yields of the peptides referred to are low, and sufficient quantities of the peptides corresponding to the other two -S-S- bonds presumably present in the molecule have not yet been isolated, although work in this direction is in progress.

With the generous permission of Dr Anfinsen, however, Dr Spackman's results can be considered in conjunction with those obtained by Ryle and Anfinsen¹⁷ on the same subject. By the use of quite different methods, they have (see p. 229) obtained evidence for linkages between half-cystine residues II and VIII, and III and VII, as well as evidence bearing on the probable correctness of the I-VI and IV-V linkages which we have found. Through Dr Anfinsen's courtesy in allowing us to refer to his results, it is possible to

visualize the ribonuclease molecule in the two-dimensional schematic representation shown in Fig. 6. It is quite clear that the protein must exist in a tightly wound folded form. It should be noted that the small disulfide ring around the IV-V disulfide bond, is two residues larger than the small ring already found in insulin by Sanger and in vasopressin and oxytocin by du Vigneaud and his colleagues.

The information contained in Figs. 2 and 6 furnishes as complete a picture of the covalent structure of the ribonuclease molecule as can be drawn

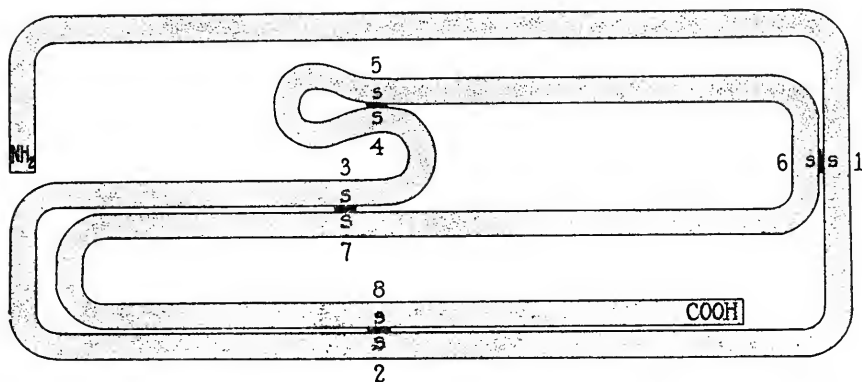


Fig. 6. A two-dimensional diagram indicating the positions of the four disulfide bonds in native ribonuclease. (Based upon the results of Spackman *et al.*¹⁶ and Ryle and Anfinsen.¹⁷)

now. The studies outlined in this communication encourage the hope, however, that the complete elucidation of the structure is not too far off. Although this appears to be a straightforward task, the protein chemist must always be on the look-out for unsuspected types of linkages in proteins. Such linkages have been found in bacitracin by Craig and his colleagues,²⁰ and by Newton and Abraham,²¹ and recently in collagen by Mechanic and Levy.²² It is becoming increasingly clear also that even a complete knowledge of all of the covalent linkages in ribonuclease will almost certainly not mean that the chemical, enzymatic, and physical properties of the protein will automatically be understood. Indeed, it appears as though detailed structural knowledge frequently does not actually *solve* problems, but rather permits the problems to be rephrased in different, albeit more precise, terms. Possibly it is one of the functions of a symposium such as this to draw attention to those areas of protein structure that lie beyond the detailed formulae that we all are so happily constructing at the present time.

REFERENCES

1. S. MOORE and W. H. STEIN, *Harvey Lectures*, Series 52, 119 (1956-57).
2. C. H. W. HIRS, W. H. STEIN and S. MOORE, *J. Biol. Chem.*, **221**, 151 (1956)
3. F. SANGER, *Bull. Soc. chim. Biol.*, **37**, 23 (1955).

- 222 C. H. W. HIRS, WILLIAM H. STEIN & STANFORD MOORE [12
4. C. B. ANFENSEN, R. R. REDFIELD, W. L. CHOATE, J. PAGE and W. R. CARROLL, *J. Biol. Chem.*, **207**, 201 (1954).
 5. C. H. W. HIRS, W. H. STEIN and S. MOORE, *J. Biol. Chem.*, **211**, 941 (1954).
 6. C. H. W. HIRS, *J. Biol. Chem.*, **219**, 611 (1956).
 7. C. H. W. HIRS, S. MOORE and W. H. STEIN, *J. Biol. Chem.*, **219**, 623 (1956).
 8. J. L. BAILEY, S. MOORE and W. H. STEIN, *J. Biol. Chem.*, **221**, 143 (1956).
 9. C. H. W. HIRS, *Federation Proc.*, **16**, 196 (1957).
 10. G. BRAUNITZER and G. SCHRAMM, *Chem. Ber.*, **12**, 2025 (1955).
 11. P. EDMAN, *Acta Chem. Scand.*, **4**, 283 (1950).
 12. M. OTTESEN and A. WOLLENBERGER, *Compt. rend. Lab. Carlsberg, Ser. Chim.*, **28**, 463 (1953).
 13. D. H. SPACKMAN, W. H. STEIN and S. MOORE, *Federation Proc.*, **15**, 358 (1956).
 14. D. H. SPACKMAN, E. L. SMITH and D. M. BROWN, *J. Biol. Chem.*, **212**, 255 (1955).
 15. R. R. REDFIELD and C. B. ANFENSEN, *J. Biol. Chem.*, **221**, 385 (1956).
 16. D. H. SPACKMAN, S. MOORE and W. H. STEIN, *Federation Proc.*, **16**, 252 (1957).
 17. A. P. RYLE and C. B. ANFENSEN, *Biochim. et Biophys. Acta*, **24**, 633 (1957).
 18. A. P. RYLE, F. SANGER, L. F. SMITH and R. KITAI, *Biochem. J.*, **60**, 541 (1955).
 19. B. KASSELL and E. BRAND, *J. Biol. Chem.*, **125**, 115 (1938).
 20. W. HAUSMANN, J. R. WEISIGER and L. C. CRAIG, *J. Am. Chem. Soc.*, **77**, 723 (1955).
 21. G. G. F. NEWTON and E. P. ABRAHAM, *Biochem. J.*, **47**, 257 (1950).
 22. G. L. MECHANIC and M. LEVY, *Federation Proc.*, **16**, 220 (1957).

The structure of ribonuclease in relation to its enzymatic activity and physical properties*

C. B. ANFINSEN

Laboratory of Cellular Physiology and Metabolism,
National Heart Institute, Bethesda, Maryland

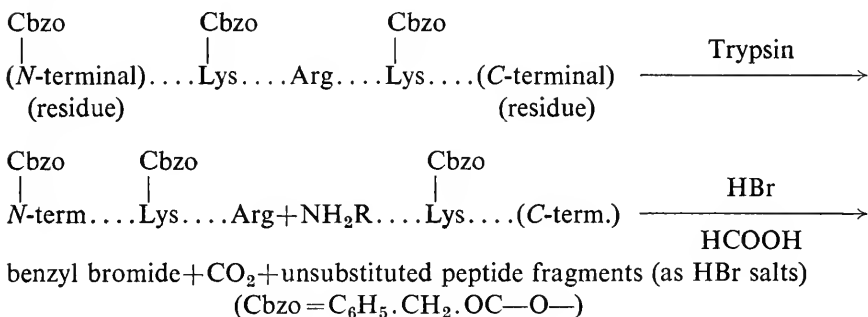
In order to appreciate the present state of protein chemistry we need only consider the material being covered in this symposium from the standpoint of our position ten years ago. Although there was available at that time a large amount of physical data relating to the behavior of proteins in solution and in the solid state, a rational interpretation of these data was essentially impossible because of the almost complete absence of quantitative information on the sequential structure of these molecules. Today, the determination of the amino acid sequence of polypeptide chains has, in general, become a technical problem of only moderate difficulty and the application of available degradative methods is likely to permit the elucidation of the covalent structure of almost any chosen protein whose molecular weight lies within reasonable limits.

Drs Hirs, Moore and Stein have presented to you a summary of their precise studies on the structure of ribonuclease (see p. 211). Detailed reference to their original papers has not been made in the bibliography below but can be found elsewhere in this volume. As they have pointed out, we have also been belabouring this protein in Bethesda and have reached conclusions about the sequence of the molecule which are in good agreement with their findings. In view of this agreement I will only discuss some of the pertinent results which begin to enable us to correlate the activity and physical behavior of this enzyme with its covalent and non-covalent structure.

Before beginning these considerations I would like to refer briefly to two methodological developments which may be of interest to this audience. These methods have to do with the general problem of the specific hydrolysis of a protein molecule to yield reproducible fragments of moderate size, a problem which we all recognize as fundamental in structural analysis.

* Those recent studies discussed in this chapter, which were carried out at Bethesda, are the combined efforts of Drs Michael Sela, W. F. Harrington, F. W. White, and myself. I would like to emphasize my role as spokesman for this group.

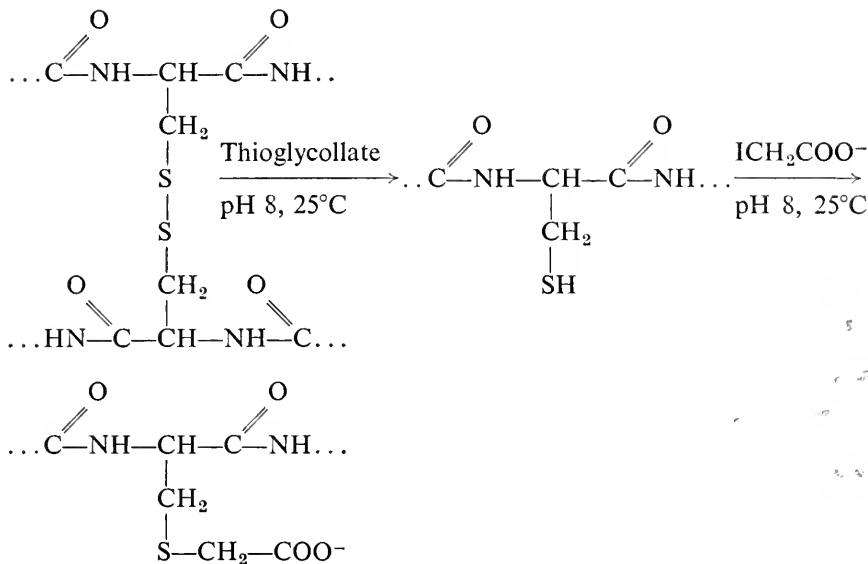
The techniques of hydrolysis just described by Drs Hirs, Moore and Stein may be difficult to apply in the case of polypeptide chains longer than that of ribonuclease since the number of overlapping peptide fragments to be considered might become forbiddingly large. We have recently described a method by which the action of trypsin may be restricted to those peptide bonds in which arginine constitutes the *N*-terminal side.¹ This procedure leads to the formation of peptide fragments, one more in number than the number of arginine residues in the chain. The reactions involved are summarized in the following equation:



As indicated, carbobenzylation of the ϵ -amino groups of lysine prevents the hydrolysis of lysyl bonds by trypsin. Following digestion, the carbobenzyloxy groups are removed to yield a mixture of peptides, which may be separated by one of a variety of methods, for subsequent degradative experiments.

It has been the common experience of many investigators that the trypsin digestion of native proteins proceeds very slowly due, presumably, to the steric hindrances introduced by the presence of disulfide bridges. Hitherto, the rupture of these bridges has been generally accomplished by performic acid oxidation which, unfortunately, causes extensive destruction of tryptophan residues. The full usefulness of the carbobenzylation method depends, therefore, on the use of techniques for the cleavage of S-S bridges under conditions which are not likely to damage portions of the sequence containing tryptophan. As an approach to this problem, the disulfide bridges of ribonuclease (which contains no tryptophan) and of lysozyme (containing 6-7 residues of tryptophan) have been reductively cleaved with thioglycolic acid in 8M urea and the resulting sulfhydryl groups have been converted to their *S*-carboxymethyl derivatives by treatment with iodoacetic acid,² as shown in the equations opposite.

Fraenkel-Conrat *et al.* have previously described a reductive cleavage of lysozyme where SH groups were masked by treatment with iodoacetamide.³ The use of iodoacetic acid, however, appears to have a distinct advantage since the final carboxymethylated derivative is quite soluble over a wide range of pH values in contrast to the amide form.



The quantitative character of the reduction and alkylation procedures has been confirmed by titration of the reduced proteins with *p*-chloromercuribenzoate and by the isolation and determination of *N*-dinitrophenyl-*S*-carboxymethyl-*L*-cysteine from dinitrophenylated hydrolysates of the alkylated derivatives.² Table 1 shows data comparing the determination of SH groups by the two methods in ribonuclease at different stages of reduction. As can

Table 1

ANALYSIS FOR CYSTEINE IN RIBONUCLEASE AT VARIOUS STAGES OF REDUCTION, BY *p*-CHLOROMERCURIBENZOATE (PCMB) TITRATION⁴ AND BY ESTIMATION OF *N*-DINITROPHENYL-*S*-CARBOXYMETHYL-CYSTEINE

SH groups by PCMB titration	DNP- <i>S</i> -carboxymethyl-cysteine*
4.2	5.1
5.2	5.3
6.0	5.2
7.5	7.8
8.0	8.7
8.1	8.4
8.2	8.3

* Determined in dinitrophenylated hydrolysates of reduced, alkylated protein. Corrected for losses due to hydrolysis and chromatography essentially as described by Levy.⁵ Destruction of carboxymethylcysteine during acid hydrolysis is on the order of 5% or less (16 hours in constant-boiling HCl at 110°C.).

be seen here, the chromatographic separation of *N*-dinitrophenyl-*S*-carboxymethylcysteine and its subsequent elution and estimation compares well with the spectrophotometric method. Table 2 deals with similar experiments on

Table 2

ESTIMATION OF CYSTEINE, ASPARTIC ACID,
AND GLUTAMIC ACID CONTENT OF FULLY
REDUCED, CARBOXYMETHYLATED LYSOZYME

	Theoretical*	Found†	
		Exp. 1	Exp. 2
DNP-Asp	20	20	20
DNP-Glu	4.5†	4.8	6.1
DNP-SCMC§	10	10.6	10.0
bis-DNP-cystine	0	0.0	0.2

The reduced, alkylated protein was hydrolyzed at 110° for 16 hours in 6*N*-HCl. The hydrolysate was dinitrophenylated⁵ and, following chromatography according to Levy,⁵ the spot containing the DNP derivatives of glutamic acid, aspartic acid, and *S*-carboxymethyl-cysteine was excized. The mixture of derivatives was separated in the tertiary amyl alcohol system of Blackburn and Lowther,⁶ running the chromatograms in the descending direction for 48 hours.

lysozyme and indicates that the application of the reductive and alkylating techniques to this tryptophan-rich protein furnishes a convenient method both for opening the cross-linked chain and for the subsequent determination of the cystine content of the protein.

It has also been found that, under the conditions of reduction and alkylation employed, histidine, tryptophan, and tyrosine residues in the polypeptide chain are unchanged. Michaelis and Schubert, in 1936,⁷ demonstrated that only SH groups are alkylated by haloacetic acids under sufficiently mild conditions (room temperature and pH values in the neighbourhood of 8). It has recently been reported by Korman and Clarke that more vigorous conditions (40°, pH 9, 4-5 hours) are required for the alkylation of free amino groups and that the ring nitrogen atoms and the phenolic hydroxyl group of tyrosine are even more resistant to reaction.⁸ We have further examined these possible side reactions under the conditions employed by us (room temperature, pH 8, 2 hours) and find that less than 5% of the histidine and tyrosine residues and free amino groups become carboxymethylated under conditions where reaction of SH groups is complete.

* The 'best' values as mentioned by Thompson.⁹

† The value for glutamic acid is undecided. Thompson's value of 7, deduced from sequence studies, is probably high. Lewis *et al.*³² give 4, as do Dose and Caputo,³³ and Thompson⁹ gives 5.1 on the basis of amino acid analysis.

‡ Calculated assuming the presence of 20 residues of aspartic acid in the protein.

§ *N*-dinitrophenyl-*S*-carboxymethylcysteine.

When fully reduced, alkylated ribonuclease was carbobenzoxyated and then subjected to trypsin digestion, results were obtained that completely paralleled those obtained with carbobenzoxyated, *oxidized* ribonuclease. End-group analysis of trypsin digests of these two forms of the extended ribonuclease chain are shown in Table 3. In both cases, the peptide bond

Table 3

N-TERMINAL AMINO ACID ANALYSES OF TRYPTIC DIGESTS OF CARBOBENZOXYLATED OXIDIZED RIBONUCLEASE AND OF REDUCED AND CARBOXYMETHYLATED RIBONUCLEASE

The carbobenzoxy groups were subsequently removed with anhydrous hydrogen bromide.

Oxidized ribonuclease			Reduced and carboxymethylated ribonuclease		
	Found	Theor.		Found	Theor.
DNP-Asp	1.0	1.0	DNP-Asp	1.0	1.0
DNP-Glu	1.8	2.0	DNP-Glu	1.8	2.0
DNP-CySO ₃ H	0.7	1.0	DNP-SCMC*	0.8	1.0
ε-DNP-Lys	10.0	9.0	ε-DNP-Lys	9.0	9.0
bis-DNP-Lys	0.9	1.0	bis-DNP-Lys†	0.4	1.0

which joins arginine to a half-cystine residue in the native protein was the most resistant to digestion, although in one case the half-cystine residue was in the form of cysteic acid and in the other was present as *S*-carboxymethylcysteine.

Data on lysozyme are presented in Table 4. The reduced, alkylated protein was treated with trypsin in this case without prior carbobenzoxylation. Lysozyme contains eleven arginine and six lysine residues and should therefore yield eighteen peptides upon trypsin digestion, if all the lysyl and arginyl bonds are intrinsically susceptible to trypsin action (i.e. not adjacent to proline, etc.). Electrophoretic patterns¹² of such a digest are shown in Figs. 1A and 1B. Fig. 1A shows the pattern obtained in 1.5 hours at 35v/cm. (pH 6.5, pyridine, acetate buffer), and Fig. 1B in 2.5 hours under the same conditions. The end-group data in Table 4 show that approximately fifteen peptide bond cleavages are detectable after digestion and that these may, in most instances, be accounted for in the lysyl and arginyl sequences determined to be present in lysozyme by Thompson,⁹ by Schroeder,¹¹ and by Acher, *et al.*¹⁰ It is worthwhile to point out that the time required for complete cleavage of the reduced, fully alkylated lysozyme chain was less than five minutes at 37°, employing trypsin at a level of 1% that of the substrate.

* *N*-dinitrophenyl-*S*-carboxymethylcysteine.

† The values for bis-DNP-lysine, both in lysozyme and in ribonuclease, are sometimes unaccountably low.

Table 4

DETERMINATION OF N-TERMINAL AMINO ACIDS IN A TRYPSIN DIGEST OF FULLY REDUCED, CARBOXYMETHYLATED LYSOZYME

DIGESTION CONDITIONS: 1% LYSOZYME, 0.01% TRYPSIN, pH 8.0, 25°. PROTEOLYSIS COMPLETE (BY PHSTAT MEASUREMENT) AFTER 5 MINUTES. END-GROUP ANALYSIS BY METHOD OF LEVY⁵

Amino acid	Total in lysozyme	End-groups expected from sequence data*	End-groups found
Aspartic Acid	35	1	4.0
Glutamic acid		3	
S-carboxymethyl cysteine		4	
Serine	9	—	0.7
Threonine	7	—	1.1
Glycine	11-12	3	1.9
Alanine	10-12	—	0.3 (?)
Valine	5-6	1	1.7
Leucine (+isoleucine)	8	2	2.8
Phenylalanine	3	1	1.1
Lysine (N-terminal in lysozyme)	1	1	1.0
Total		12	ca. 15†

In concluding these preliminary remarks about methods I should emphasize that a real measure of the value of the reductive cleavage of disulfide bridges and of the trypsin digestion of carbobenzoxyated polypeptide chains will become apparent only when these methods have been properly applied to tryptophan-containing proteins larger than ribonuclease and lysozyme.

To proceed with the main purpose of this discussion, as indicated in its title, I should like now to consider some aspects of ribonuclease structure in relation to biological and physical properties. As you have seen, the sequential work on this protein is well along and the full sequence should be forthcoming in a relatively short time. As all of us will agree, however, a knowledge of amino acid sequence alone is of little use in the understanding of three-dimensional structure and of the basic reasons for biological activity. It has become increasingly evident that the activity of an enzyme, for example, must ultimately be explained in terms of the secondary and tertiary structure of the molecule, determined and rigidly fixed by covalent cross linkages and by non-covalent bonds of differing strengths.

The Rockefeller Institute group,¹³ and ourselves,¹⁴ have recently undertaken independent investigations of the disulfide bridges of ribonuclease by

* From sequence data of Thompson,⁹ W. Schroeder,¹¹ and R. Acher *et al.*¹⁰ Amino acids listed are those following either lysine or arginine in the sequences listed in the above papers.

† Lysozyme contains 6 lysine and 11 arginine residues. Theoretical maximum number of end-groups in a complete trypsin digest should, therefore, be 18.

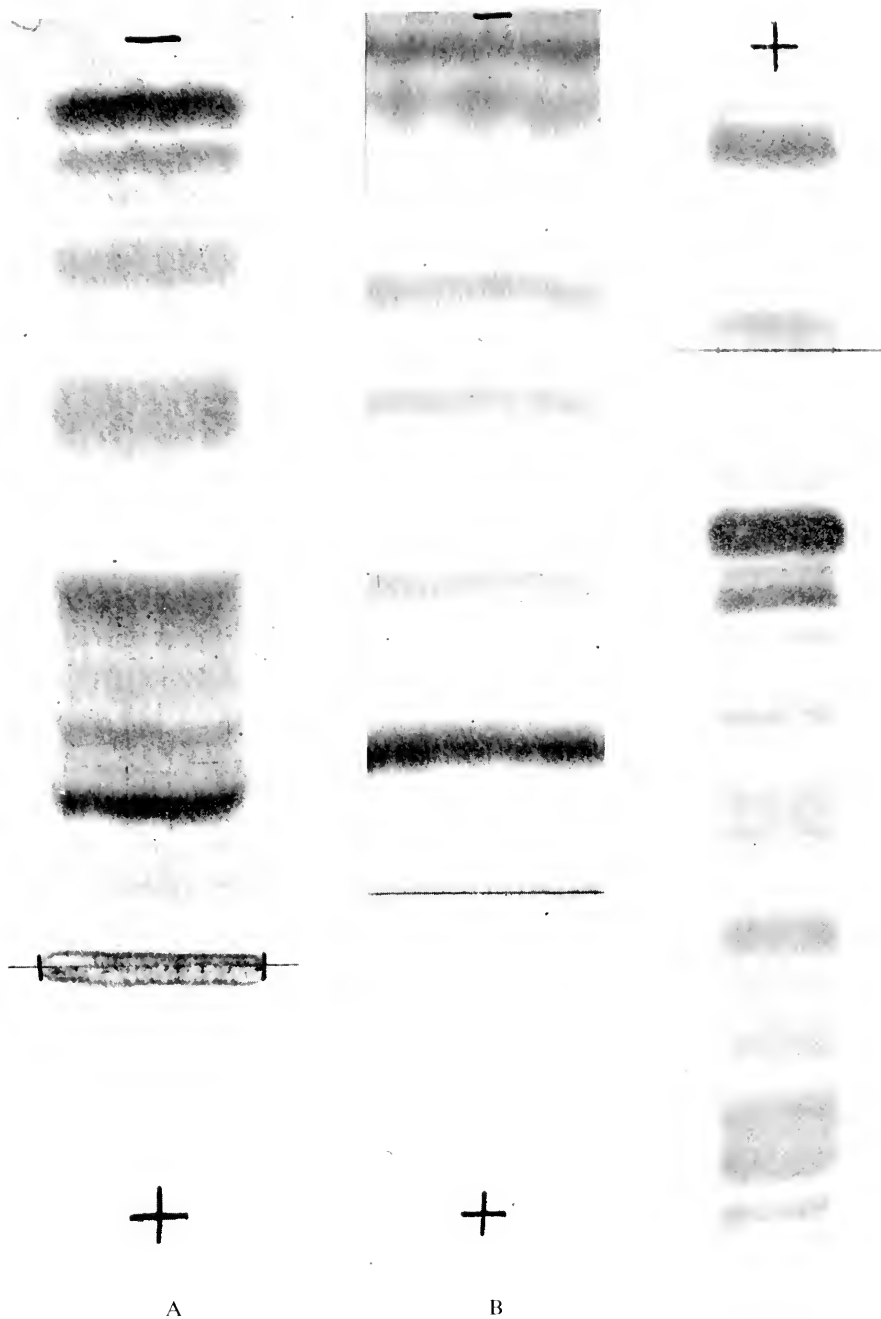


Fig. 1

Fig. 2

Fig. 1. Electrophoretic patterns¹² of trypsin digests of reduced, alkylated lysozyme: (A), 1.5 hours, pH 6.5 pyridine-acetate buffer, 35 V/cm.; (B), 2.5 hours, same conditions.

Fig. 2. Electrophoretic patterns¹² of subtilisin digest of native ribonuclease.¹¹ Conditions as in Fig. 1A.

13] STRUCTURE & ENZYMATIC ACTIVITY OF RIBONUCLEASE 229
 methods essentially the same as those employed by Sanger and his col-
 laborators for insulin.¹⁵ In our studies, subtilisin digests of native ribonu-
 clease were subjected to electrophoresis by the method of Michl¹² and the
 five ninhydrin-positive bands, which also exhibited a positive reaction for
 disulfide bonds,¹⁶ were eluted. A typical ninhydrin pattern is given in Fig. 2.
 One of these cystine-containing components proved to be too large and com-
 plicated for direct study. The other four were oxidized with performic acid and
 the resulting cysteic acid-containing peptides were separated and analysed for
 amino acids. The results of these analyses are given in Table 5. A considera-
 tion of these analyses, taken together with the available data on the amino

Table 5

THE LOCATION OF DISULFIDE BRIDGES IN RIBONUCLEASE

Disulfide- positive Band	Composition of cysteic-acid containing peptides produced by oxidation	Half-cystine residues implicated	Disulfide bridge
A	(1) $\text{CySO}_3\text{-Asp-Arg-Glu-(Ser}_2\text{,Thr,Gly)}$ (2) $(\text{CySO}_3\text{,Asp,Glu})$	No. VI Nos. I, III, V and VIII	Nos. I-VI
C	(1) $(\text{Ser,Glu,CySO}_3)\text{-Lys}$ (2) $(\text{Gly,Asp,CySO}_3\text{,Pro,Ala,Lys,Tyr})$	No. III No. VII	
D	(1) $\text{CySO}_3\text{-[Ala,Tyr]}$ (2) $\text{Asp-Arg-CySO}_3\text{-(Pro,Lys,Asp,Val,Thr)-Phe}$	No. VIII No. II	Nos. II- VIII
E	(1) $(\text{Gly,Asp,CySO}_3\text{,Pro,Ala,Lys,Tyr,Ser})$ (2) $(\text{Ser}_{1.1}\text{,Glu}_{2.0}\text{,Asp}_{2.7}\text{,Lys}_{1.8}\text{,Tyr}_{0.9}\text{,}$ $\text{CySO}_3\text{ }_{3.2}\text{,Gly}_{0.7}\text{,Thr}_{0.6}\text{,Ala}_{1.0}\text{,Val}_{1.0})$	No. VII Nos. III, IV and V together	Nos. IV- V

acid sequence and distribution along the polypeptide chain of ribonuclease, permits the conclusion that the half-cystine residues of the protein are paired as follows: No. I-No. VI, No. II-No. VIII, No. III-No. VII, and No. IV-No. V (numbering the half-cystines from I to VIII, beginning at the *N*-terminal end of the protein). Drs Spackman, Stein and Moore¹³ have independently identified the first and fourth of these bridges as well and are now in the process of examining the remaining two. Their studies are of particular importance since complete agreement in the data from the two laboratories will help to rule out the ever-present possibility that there might have occurred some disulfide exchange or other rearrangement during the course of experimental study. The assignment of the disulfide bridges together with the sequential information we now have enables us to construct the schematic and provisional diagram shown in Fig. 3. In this figure the heavily-encircled residues are arranged in an order which is known (or nearly so—there exists an unresolved disagreement in the sequence adjacent to half-cystine No. VI. This portion of the sequence should probably read,

13] STRUCTURE & ENZYMIC ACTIVITY OF RIBONUCLEASE 231 intermediates. Regeneration of enzymatic activity takes place upon reoxidation of the sulfhydryl groups with molecular oxygen and 'fingerprints' (filter paper electrophoretic patterns of subtilisin digests) of the reaction mixture after reoxidation (Fig. 6) suggest that there occurs a partial reformation of the same disulfide bonds as those occurring in the native molecule.¹⁸ Specific identification of the labile, presumably non-essential, disulfide bonds has

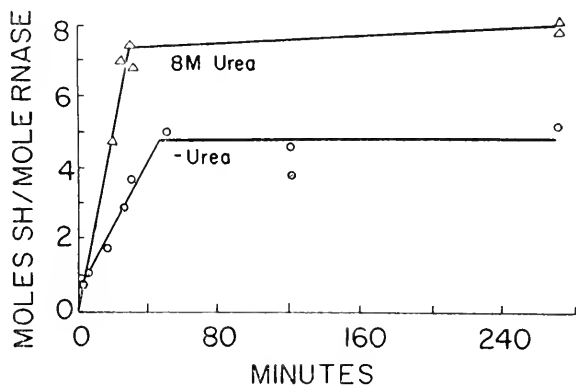


Fig. 4. The kinetics of reduction of the disulfide bridges of ribonuclease in the presence and absence of 8M urea.¹⁸

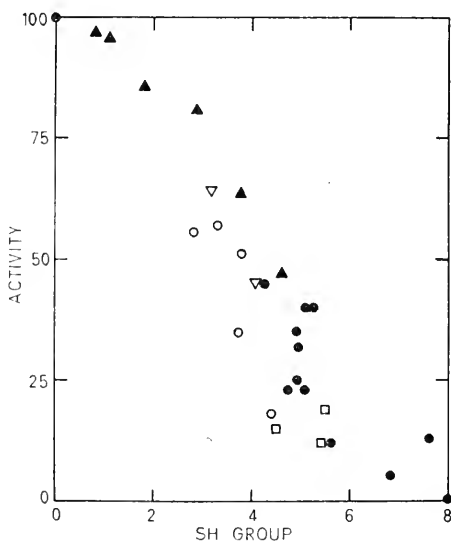


Fig. 5. Activity of ribonuclease at various stages of reduction (expressed as percentage of the specific activity of native ribonuclease) as a function of the number of moles of sulfhydryl per mole of enzyme. ▲, Reduction in absence of urea; ●, reduction in 8M urea; □, reoxidation of fully reduced, inactive ribonuclease; ○, reoxidation of samples containing more than six sulfhydryl groups per average molecule; ▽, reoxidation of samples containing about four sulfhydryl groups per average molecule.²

not yet been achieved. However, it is of interest that the active, partially reduced molecules, free of detectable amounts of native enzyme, are much more susceptible to proteolytic digestion (Fig. 7) than the native protein and that enzyme activity persists even after considerable further degradation with subtilisin.¹⁸ This latter observation is promising as regards the production of a relatively small, catalytically active, 'core'.

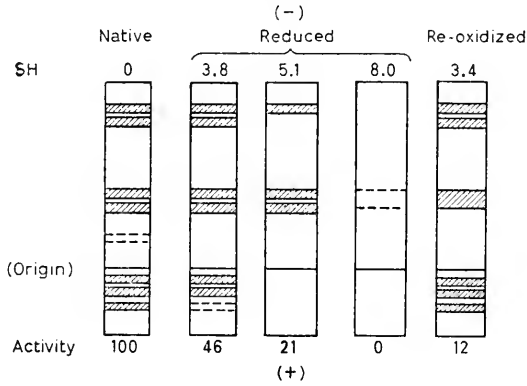


Fig. 6. Drawing of the disulfide-positive bands of the electrophoretic patterns obtained on subtilisin digests of native, reduced, and reoxidized ribonuclease. Same conditions as in Fig. 1A. The sample which was reoxidized with molecular oxygen contained 8.0 SH groups and was completely inactive before oxidation.¹⁸

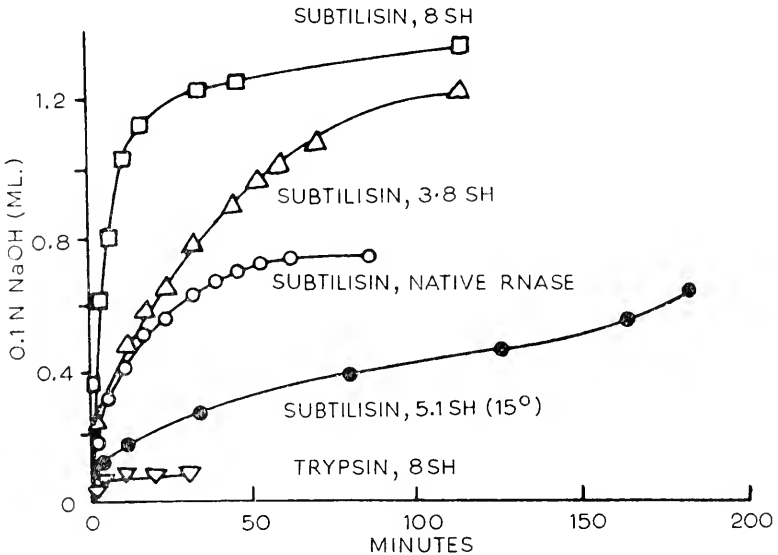


Fig. 7. Digestion of ribonuclease, at various stages of reduction, with subtilisin and trypsin. The reduced derivatives were carboxymethylated, before digestion. Alkali uptake was followed in the Coleman autotitrator at pH 8. Temperature, 37°, except where indicated. Substrate concentration, 10 mg/ml. Proteolytic enzyme concentration, 1% of this level.¹⁸

2. PROTEOLYTIC DIGESTION

As shown by Richards¹⁹ limited subtilisin digestion of ribonuclease causes the cleavage of what appears to be a single peptide bond located within twenty or so residues of the *N*-terminal end of the polypeptide chain.²⁰ The resulting macromolecular derivative has full enzymatic activity. Degradation of the *C*-terminal end of ribonuclease with carboxypeptidase pre-treated with diisopropyl fluorophosphonate (DIP) removes the *C*-terminal valine residue (and a variable amount of the two preceding residues, serine and alanine) without loss of activity.²¹ Non-DIP-treated carboxypeptidase effects a considerably greater number of cleavages and still leaves most of the activity even after fairly extensive attack.²²

Limited pepsin digestion at pH 1.8, on the other hand, causes the complete inactivation of ribonuclease at a rate parallel to the rate of release of the *C*-terminal tetrapeptide sequence.²³ No indications of rupture of other peptide bonds have been found by a variety of chemical and physical methods.²⁴ The inactive derivative contains a *C*-terminal phenylalanine residue as would be expected from the sequence data shown in Fig. 3 above.

3. CORRELATIONS BETWEEN SPECIFIC ASPECTS OF TERTIARY STRUCTURE AND ENZYMATIC ACTIVITY

(a) *Urea*. In an earlier paper it was suggested that enzymatic activity in this protein is not dependent upon an intact system of hydrogen bonds since activity is undiminished in 8M urea solutions.²⁵ Our more recent studies indicate that this conclusion is probably not quantitatively valid. This re-appraisal is necessary in view of the results of experiments in which the marked shift in the spectrum of ribonuclease to lower wavelengths in 8M urea solutions has been studied as a function of the concentration of certain polyvalent anions and polyanions.^{24,26} The curves in Fig. 8, for example, indicate the differences in molar extinction, between native ribonuclease in 0.1 M-KCl and in 8M urea (curve B), and in 8M urea containing 0.003M orthophosphate ions (curve C). Low levels of orthophosphate, arsenate, polymetaphosphate, and uridyate (but not citrate) ions cause a complete reversion of the shifted spectrum to that exhibited by the native enzyme in the absence of urea. This reversion is quantitative at concentrations of ion, as low as 0.05M (i.e., about 60-70 atoms of phosphorous/mole of ribonuclease). A summary of these, and related experiments on ultraviolet absorption characteristics, is given in Table 6.^{17,24,26} The table includes, for comparison, the results of experiments with monovalent anions, which do not show such an effect except at very high ionic strengths. (It should be stated that the most desirable test, namely an estimate of the refolding capacity of ribonucleic acid itself, is not technically feasible because of the

high extinction coefficient of this material.) These findings strongly suggest that the substrate, ribonucleic acid, is capable of refolding the protein in spite of the presence of 8M urea and that the structural configuration responsible for catalytic activity is unchanged by this agent under the conditions of assay.

Table 6

SOME PHYSICAL PROPERTIES OF RIBONUCLEASE AND
MODIFIED RIBONUCLEASE

Ribonuclease	Unfold- ing agent	Anion	$\Delta\epsilon \times 10^{-3}$ (at 285 m μ)	$[\alpha]_D^{20}$	λc^*	η_{sp}/c
Native	—	—	1120	-71.7°	230.7	0.036
Native	8 M urea	—	0	-103.7°	220	0.085
Native	8 M urea	0.50 M acetate	0	—	—	—
Native	8 M urea	0.40 M chloride	110	-98.8°	—	—
Native	8 M urea	0.05 M sulfate	—	-102.5°	—	—
Native	8 M urea	0.25 M sulfate	840	-88.0°	—	—
Native	8 M urea	0.001 M phosphate	112	-103.7°	—	—
Native	8 M urea	0.003 M phosphate	390	-102.9°	—	—
Native	8 M urea	0.01 M phosphate	900	-99.9°	—	0.070
Native	8 M urea	0.02 M phosphate	—	—	—	0.057
Native	8 M urea	0.15 M phosphate	1120	-81.0°	—	0.050
Native	8 M urea	0.15 M arsenate	1120	-80.7°	—	—
Native	8 M urea	0.15 M pyrophos- phate	1120	—	—	—
Native	8 M urea	0.003 M polymeta- phosphate	447	—	—	—
Native	8 M urea	0.055 M polymeta- phosphate	1120	-82.8°	—	—
Native	8 M urea	0.15 M uridylyate	—	—	—	0.052
Native	8 M urea	0.15 M citrate	0	—	—	—
Native	6 M guanidine HCl	—	0	-108.1°	216	0.097
Native	6 M guanidine HCl	0.30 M phosphate	0	-102.3°	217.5	0.104
Oxidized	—	—	0	-91.6°	226	0.116
Reduced and carboxymethy- lated	—	—	0	—	—	0.149
Pepsin inacti- vated	—	—	0	-83.8°	—	0.039
Subtilisin di- gested	—	—	1120	—	—	—
Methyl ester	—	—	0	-84.0°	—	—

The concentration of the proteins was 2.5-2.7 g./l. for spectrophotometric readings, 25-30 g./l. for polarimetric, and 14-15 g./l. for viscometric measurements. All measurements were carried out at pH 6-7.

* See Yang and Doty (1957).

Other data in Table 6 indicate, however, that refolding by anions is not quantitatively complete. Thus measurements of the optical rotatory properties and the intrinsic viscosity of ribonuclease in 8M urea solutions containing orthophosphate or polymetaphosphate ions give values that are compatible with a still partially unfolded form of the protein. The combined

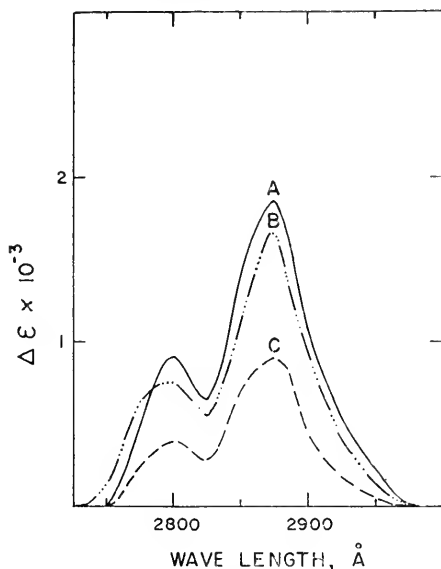


Fig. 8. Decreases in extinction, compared with native ribonuclease in water, of (A), inactive derivative produced by limited pepsin digestion, (B), ribonuclease in 8M urea, and (C), ribonuclease in 8M urea containing 0.003M orthophosphate ions. Curves calculated from individual curves obtained on Cary ultraviolet recording spectrophotometer.²⁶

data suggest that that portion of the molecule which is responsible for the modified light absorption of native ribonuclease is also concerned with catalytic activity and that catalysis is unimpaired in spite of minor structural alterations elsewhere in the fabric of the protein.

(b) *Pepsin digestion.* This conclusion is further supported, in a converse sort of way, by the spectrophotometric and optical rotatory results obtained on the derivative produced by limited pepsin digestion.^{24,26} This derivative, which, as was discussed above, differs from the native enzyme only by lacking the four C-terminal residues, shows a shifted spectrum (curve A, Fig. 8) similar to that seen in 8M urea. Orthophosphate is without effect on this pepsin-induced shift. The specific optical rotation of the derivative is similar to that obtained with native ribonuclease or with ribonuclease in 8M urea solutions containing polyvalent anions. It may be tentatively concluded, therefore, that the secondary structure of the derivative is reasonably 'native', a conclusion also consistent with earlier estimates of the intrinsic viscosity and sedimentation behavior of this material.²³

(c) *Heat, alkali, and methylation.* Ribonuclease may be progressively inactivated, in an irreversible way, by exposure to heat, to alkalinity above pH 12.3-12.7,²⁴ and by esterification of free carboxyl groups. These treatments all lead to a shift in spectrum, the extent of the shift being proportional to the extent of inactivation. The titration studies of Tanford and Hauenstein²⁷ indicate that the irreversible change in spectrum occurring at high pH involves the modification of approximately 3 hydrogen-bonded tyrosine residues, as suggested also by the studies of Harrington and Schellman¹⁷ on the urea denaturation of ribonuclease. In the latter case, however, the shift in spectrum is reversible and is prevented by anions.²⁴

(d) *Acid, subtilisin digestion, oxidation, and reduction.* Scheraga²⁸ has shown that a spectral shift takes place when ribonuclease is dissolved in acid solutions and that the extent of the shift is pH dependent in a manner consistent with the titration of acid groups having pK values in the neighborhood of pH 2. These unusually acid groups may represent the electronegative groups involved in hydrogen bond formation with the tyrosine hydroxyl groups mentioned above.

Performic acid oxidation, which causes complete inactivation,²⁹ produces an irreversible shift in spectrum.²⁴ Partial reduction, however, when carried to a point where approximately one disulfide bridge per molecule has been cleaved, and where no trace of remaining native enzyme can be detected electrophoretically, yields a product with essentially full activity and which shows a 'native' spectrum.²

Subtilisin digestion causes extensive and random digestion of the native enzyme and during this digestion there occur parallel changes in enzyme activity and in the degree to which the spectrum has been shifted to lower wavelengths.

(e) *Guanidine hydrochloride.*²⁶ Guanidine hydrochloride is known to be a much stronger denaturing agent than urea (see, for example,³⁰). It was found that 6M solutions of this salt completely inactivated ribonuclease in contradistinction to the results with urea. The studies with guanidine indicated that the substrate molecule was unable to bring about refolding of the polypeptide chain. Spectrophotometric, optical rotatory, and viscometric measurements bear out this conclusion (see Table 6). Optical dispersion measurements (Table 6 and Fig. 9) also indicate that guanidinium ions cause an even greater disorientation of the secondary structure of ribonuclease than does urea, as shown by a comparison of the rotatory dispersion constants for the protein in these two systems. In view of these results it was of interest to determine the level of guanidine hydrochloride at which ribonuclease is active. Since the activity of ribonuclease is strongly affected by the ionic strength of the assay solution,³¹ and since guanidine hydrochloride is a strong salt, it was first necessary to determine whether or not this agent

had an inactivating effect over and above that shown by equivalent levels of sodium chloride. The data in Fig. 10 show that guanidinium ions exert a specific inactivating effect, not attributable to ionic strength, in the concentration range between 1 and 3 molar. It is precisely in this range of concentrations that low levels of orthophosphate no longer prevent the shift

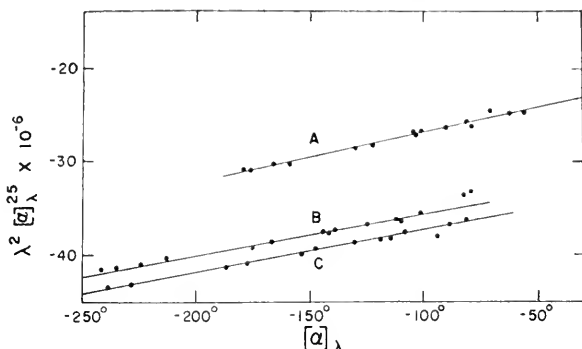


Fig. 9. The optical rotatory dispersion of ribonuclease in: (A), water, (B), 6M guanidine hydrochloride containing 0.3M phosphate buffer, pH 6.5, (C), 6M guanidine hydrochloride.²⁶ The data are plotted according to the suggestion of Yang and Doty.³⁴

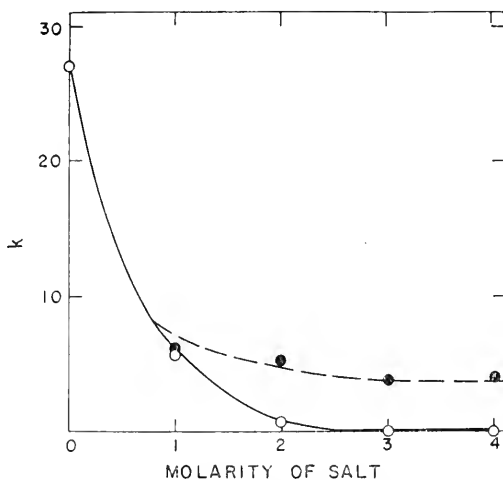


Fig. 10. Activity of ribonuclease in sodium chloride solutions (solid circles), guanidine hydrochloride solutions (open circles). Activity measured by alkali uptake in the Jacobsen-Leonis³⁵ autotitrator at pH 6.0 and plotted as the velocity constant, k , of the first order reaction, in minutes⁻¹.²⁶

in spectrum characteristic of the inactive enzyme. We may conclude, therefore, that the inactivation of ribonuclease by guanidinium ions is also closely correlated with a change in the ultraviolet absorption properties of certain tyrosine residues. The guanidine effect is completely reversible upon dilution of the system or following removal of this material by dialysis.

CONCLUSIONS

It may be stated with reasonable certainty that the polypeptide chain of ribonuclease, free of cross linkages, is not sufficient for catalytic activity since the oxidized and reduced forms of the protein are completely inert. This conclusion is supported by the results of studies which indicate that the methods employed for the cleavage of disulfide bridges do not chemically modify amino acids other than cystine. (One must, however, keep in mind the possibility that the open chain does possess intrinsic enzyme activity but that the formation of the proper catalytic configuration is prevented by the presence of sulfonic acid, carboxymethyl, or sulfhydryl groups.)

The weight of evidence now suggests that there exists in the native molecule a constellation of amino acid residues, considerably less in size than the intact protein, which is responsible for catalytic activity. The results of Richards¹⁹ on the limited degradation of the chain with subtilisin, make it likely that the *N*-terminal portion of the molecule is not involved. Similarly, the last few residues at the *C*-terminal end of the chain appear to be superfluous.^{21*} Partial reduction of disulfide bridges² also seems to be permissible.

The physical studies discussed above have indicated a correlation between activity and the presence of a unique type of spectrum, characteristic of tyrosine residues whose hydroxyl groups seem to be linked through hydrogen bonding to carboxylate groups elsewhere in the molecule. It cannot, of course, be stated that such residues are necessarily implicated in the actual catalytic process. However, it does seem likely that one or more hydrogen-bonded tyrosine residues must be intimately concerned with the 'active center' of ribonuclease and that these exist in such a linkage when the geometry of the 'center' is in its proper configuration.

These chemical and physical observations, when considered together with the fact that limited pepsin digestion causes complete inactivation after the removal of only a small *C*-terminal fragment,²³ lead us to conclude that at least a part of the catalytically active portion of the ribonuclease chain resides near the *C*-terminus. The apparent requirement for at least two intact disulfide bridges suggests that such an area of the molecule is joined in close proximity to one or more other essential amino acid sequences. The size and location of these can obviously not be considered until a great deal of further information is available on the relative importance of the individual disulfide bridges and on the extent to which the ribonuclease molecule can be further degraded without complete loss of function.

* The word 'superfluous' must be used in a very qualified sense, since it appears rather unlikely that 'superfluous' structures would have survived the attrition of mutation and selection that has been at work since Nature first devised ribonuclease molecules if some selective advantage were not involved.

REFERENCES

1. C. B. ANFINSEN, M. SELA and H. TRITCH, *Arch. Biochem. and Biophys.*, **65**, 156 (1956).
2. M. SELA, F. WHITE and C. B. ANFINSEN, *Science*, **125**, 691 (1957), and unpublished results.
3. H. FRAENKEL-CONRAT, A. MOHAMMAD, E. D. DUCAY and D. K. MECHAM, *J. Am. Chem. Soc.*, **73**, 625 (1951).
4. P. D. BOYER, *J. Am. Chem. Soc.*, **76**, 4331 (1954).
5. A. L. LEVY, *Nature*, **174**, 126 (1954).
6. S. BLACKBURN and A. G. LOWTHER, *Biochem. J.*, **48**, 126 (1951).
7. L. MICHAELIS and M. P. SCHUBERT, *J. Biol. Chem.*, **106**, 331 (1934).
8. S. KORMAN and H. T. CLARKE, *J. Biol. Chem.*, **221**, 113, 133 (1956).
9. A. R. THOMPSON, *Biochem. J.*, **61**, 253 (1955); *ibid.*, **60**, 507 (1955).
10. R. ACHER, J. THAUREAUX, C. CROCKER, M. JUTISZ and C. FROMAGEOT, *Biochim. et Biophys. Acta*, **9**, 339 (1952).
11. W. A. SCHROEDER, *J. Am. Chem. Soc.*, **74**, 5118 (1952).
12. H. MICHL, *Monatsh. Chem.*, **82**, 489 (1951).
13. D. H. SPACKMAN, S. MOORE and W. H. STEIN, *Federation Proc.*, **16**, 252 (1957).
14. A. P. RYLE and C. B. ANFINSEN, *Biochim. et Biophys. Acta*, **24**, 633 (1957).
15. A. P. RYLE, F. SANGER, L. F. SMITH and R. KITAI, *Biochem. J.*, **60**, 541 (1955).
16. G. TOENNIES and J. J. KOLB, *Anal. Chem.*, **23**, 823 (1951).
17. W. F. HARRINGTON and J. A. SCHELLMAN, *Compt. rend. trav. lab. Carlsberg Ser. Chim.*, **30**, 13 (1956).
18. F. H. WHITE, Doctoral Dissertation, University of Wisconsin, May, 1957.
19. F. M. RICHARDS, *Compt. rend. trav. lab. Carlsberg Ser. Chim.*, **29**, 329 (1955).
20. F. M. RICHARDS, personal communication.
21. F. M. RICHARDS, M. SELA and C. B. ANFINSEN, unpublished experiments.
22. W. I. ROGERS and G. KALNITSKY, *Biochim. et Biophys. Acta*, **23**, 525 (1957).
23. C. B. ANFINSEN, *J. Biol. Chem.*, **221**, 405 (1956).
24. M. SELA and C. B. ANFINSEN, *Biochim. et Biophys. Acta*, **24**, 229 (1957).
25. C. B. ANFINSEN, W. F. HARRINGTON, A. HVIDT, K. LINDERSTROM-LANG, M. OTTESEN and J. SCHELLMAN, *Biochim. et Biophys. Acta*, **17**, 141 (1955).
26. M. SELA, C. B. ANFINSEN and W. F. HARRINGTON, *Biochim. et Biophys. Acta*, **26**, 502 (1957).
27. C. TANFORD and J. D. HAUENSTEIN, *Biochim. et Biophys. Acta*, **19**, 535 (1956).
28. H. A. SCHERAGA, *Biochim. et Biophys. Acta*, **23**, 196 (1957).
29. C. B. ANFINSEN, R. R. REDFIELD, W. L. CHOATE, J. PAGE and W. R. CARROLL, *J. Biol. Chem.*, **207**, 201 (1954).
30. J. A. SCHELLMAN, R. B. SIMPSON and W. KAUZMANN, *J. Am. Chem. Soc.*, **75**, 5152 (1953).
31. S. R. DICKMAN, J. P. AROSKAR and R. B. KNOPF, *Biochim. et Biophys. Acta*, **21**, 539 (1956).
32. J. C. LEWIS, N. S. SNELL, D. J. HIRSCHMANN and H. FRAENKEL-CONRAT, *J. Biol. Chem.*, **186**, 23 (1950).
33. K. DOSE and A. CAPUTO, *Biochem. Z.*, **328**, 376 (1956).
34. J. T. YANG and P. DOTY, *J. Am. Chem. Soc.*, **79**, 761 (1957).
35. C. F. JACOBSEN and J. LEONIS, *Compt. rend. trav. lab. Carlsberg Ser. chim.*, **27**, 333 (1951).

Immunological approaches to the study of ribonuclease

B. CINADER AND J. H. PEARCE*

Agricultural Research Council, Institute of Animal Physiology
Babraham, Cambridge

IMMUNOLOGICAL CRITERIA OF HOMOGENEITY

The immunological analysis of a protein provides one of the most sensitive methods for the detection of macromolecular impurities. Diffusion and precipitation of antigen and antibody in agar (Oudin, 1952) leads to discrete precipitin zones for each antigen, the total number of independent zones representing a minimum estimate of the number of antigens present.

Since individual animals vary in their responsiveness to the same antigen (Sobey, 1954) and since a massive antibody response occurs more readily to some antigens than to others, it is necessary to examine diffusion and precipitation in agar with antisera from several individuals and with the antisera from the same animals after successive courses of immunization.

Molecules which are indistinguishable as antigens (Heidelberger, 1954) but are of different mobility can be distinguished by electrophoresis in agar and by their subsequent reaction with antibody diffusing from channels placed parallel to the direction of electrophoretic migration (Williams and Grabar, 1953; see also Cinader and Dubert, 1955).

The assay of the nitrogen contents of precipitates of antigen and antibody gives not only information on the combining ratio of reactants, but also on the relative purity of different preparations of the same protein. The shape of a plot of total N precipitated as a function of antigen N added to a constant volume of immune serum (Cohn, 1952) and the quantity of antigen leading to maximum precipitation of antibody (Cinader and Pearce, 1956) can serve as a sensitive index of relative purity.

The supernatants from antigen-antibody precipitates may provide additional information on the heterogeneity of the antigen. The simultaneous presence of antigen and antibody in the same supernatant has usually been accepted as indicating the presence of antigenic contaminants. However, this is not necessarily so; only the examination of supernatants by diffusion and precipitation in agar can show whether the presence of antigen and antibody in the same supernatants is due to soluble complexes of the principal antigen and its antibody or to antigenic contaminants.

* Present address: Microbiological Research Establishment, Porton, Wiltshire.

We can also distinguish between preparations of different purity by attaching the soluble antigen to erythrocytes by means of tannic acid (Boyden, 1951). Antisera to the attached antigen agglutinate such sensitized erythrocytes. The agglutination can be inhibited by the addition of the soluble antigen to the antiserum (Cinader and Dubert, 1955, 1956; Cinader and Pearce, 1956). The amount of protein required for this inhibition depends on the protein impurities present in a given preparation of antigen, that is on the percentage of total protein effective as inhibitor (Cinader and Pearce, 1956).

When the antigen is an enzyme, the precipitation of antigen should run parallel with the disappearance of enzyme activity from the supernatants; when antigen can no longer be demonstrated in the supernatant, enzyme activity should also be minimal. A different sequence of events indicates the presence of protein-impurities without enzyme activity (Cohn, 1954). None of the approaches can be safely interpreted, if used alone; each of them should contribute its information to the final correlation of evidence relating to homogeneity.

The following immunochemical experiments with ribonuclease are concerned with the homogeneity of the antigen, but it also forms the continuation of studies of antibody as an inhibitor of enzyme (Cinader, 1955, 1957).

THE INTERACTION OF RIBONUCLEASE WITH ANTIBODY

Crystallized bovine ribonuclease consists of a major component, the enzyme, and of a minor component which is antigenically distinct from the enzyme. The presence of these two components can be demonstrated by the reaction in agar between antibody to crystallized ribonuclease and chromatographically separated fractions of the crystallized protein (Fig. 1). The arrangement in the agar plates of reactants and the zone of precipitation formed between them is shown in Fig. 2. A common zone corresponding to the enzyme runs between all the cups containing antigen and the cups containing the antibody. One chromatographic fraction, that first eluted from the chromatographic column, contains a second antigen which reacts independently of the zone common to all the fractions and which runs across the main zone formed between ribonuclease A and B and the antiserum.

Electrophoresis in agar of crystallized ribonuclease followed by diffusion of antibody from lateral channels showed the principal antigen to be inhomogeneous (Cinader, Rondle and Pearce, 1955) and to consist of a fast and a slow moving component; these two components of the enzyme are antigenically indistinguishable since antibody reacts with them in agar with the formation of a continuous zone with two maxima. The minor antigen component present in the fraction first eluted from a chromatographic column migrates with the same velocity as the slower of the two electrophoretic components of the enzyme.

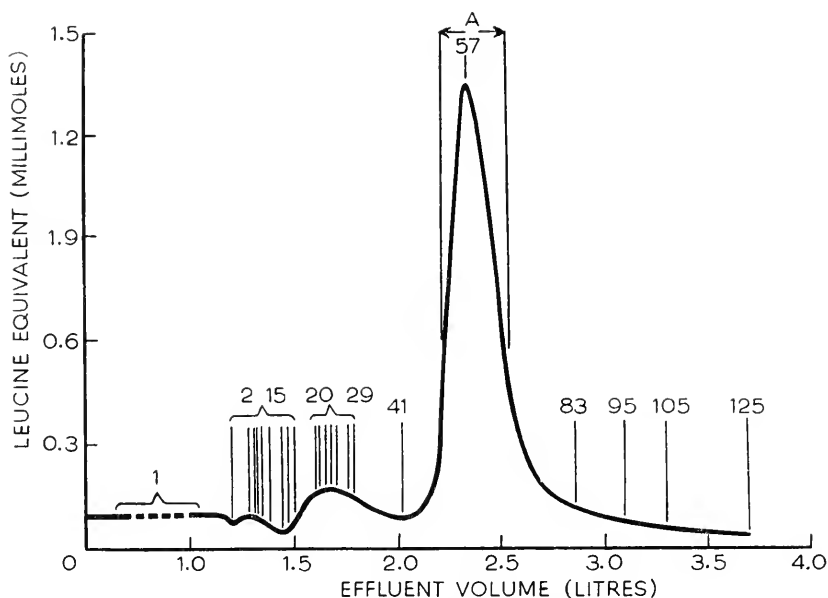


Fig. 1. Chromatographic fractionation of bovine ribonuclease (Armour, 381-59) by the method of Hirs, Moore and Stein (1953). Numbers refer to the contents of individual tubes used in the experiments shown in Fig. 2. 'A' indicates the tubes pooled for the determinations in Tables 1 and 2.

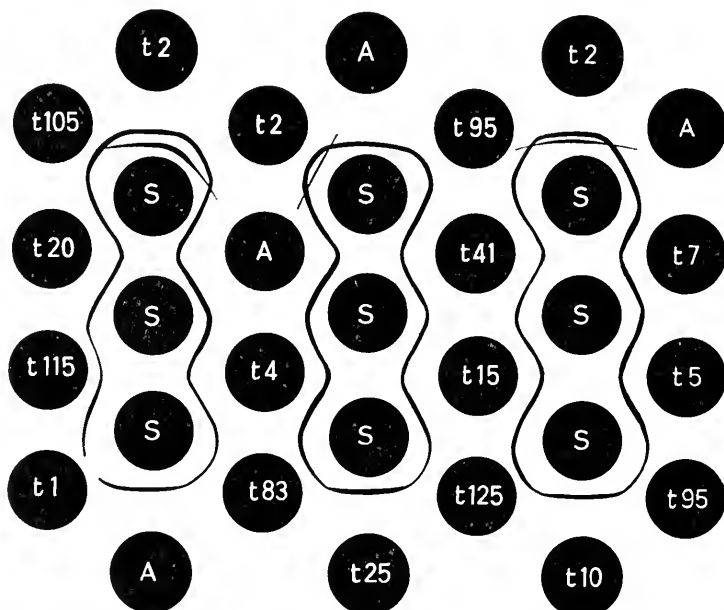


Fig. 2. Drawing of precipitin zones apparent in agar after 3 days of development. Circles denote the position of cups in the agar. S... indicates the position of cups filled with ribonuclease antiserum. t... shows cups containing chromatographic fractions of ribonuclease (Armour, 381-59); numbers refer to tube numbers (Fig. 1).

With some preparations of crystallized ribonuclease a third electrophoretic component is found to move at a greater velocity than the faster of the two components already mentioned and with one preparation of ribonuclease a fourth component of very low mobility was encountered. These two components are antigenically identical with the two principal electrophoretic components of the enzyme.

Ribonuclease A was found to consist of one antigenic species which migrated in agar with the same two mobilities as the two components usually observed in the electrophoresis in agar of the crystallized enzyme. The possibility that these two electrophoretic components were artifacts of electrophoresis could be excluded. Immediately after the electrophoresis was completed, portions of the agar were cut out in positions corresponding to the predetermined areas of highest antigen concentration; these were inserted into fresh agar slides and subjected to electrophoresis. At the end of this second run antiserum was allowed to diffuse through the agar from a median channel; each component of the enzyme was found to react with the formation of a zone showing one maximum only and hence to react as a single electrophoretic component.

It is unlikely that a second electrophoretic component was present through contamination from the neighbouring chromatographic component of ribonuclease B: we have observed two components not only with a pool of the contents of tubes of effluent corresponding to ribonuclease A, but also with the contents of the tube corresponding to the maximum of the chromatographic peak of ribonuclease A (Fig. 1). We cannot exclude the possibility that a second electrophoretic component may have been formed after ribonuclease A had been separated chromatographically.

Supernatants from precipitates formed by varying quantities of ribonuclease A and constant quantities of antiserum were examined by precipitin test and by diffusion and precipitation in agar. In no case could the co-existence of antigen and antibody be shown by precipitin tests, or the presence of an antigen impurity in the equivalence zone by diffusion in agar. However, the co-existence of enzyme and its antibody could be shown in supernatants from precipitates of ribonuclease A and several antisera, when the supernatants were examined for their power to agglutinate tanned and sensitized erythrocytes and for their enzyme activity. Thus antibodies were encountered which gave rise to soluble complexes of enzyme and antibody in a region which would be identified as the equivalence zone by precipitin tests alone (Cinader and Pearce, 1956). Some antibodies formed in response to repeated intravenous injections with bovine ribonuclease gave rise to nearly 50% of soluble complexes of antigen and antibody. It is clear that true combining ratios of antigen and antibody cannot be directly determined with sera of this nature. We therefore selected for the measurement of combining ratios an immune serum which showed an equivalence zone by all the criteria employed. This serum was obtained from a rabbit by immunization

with Freund's adjuvant followed by several courses of intravenous injections. The combining ratios of antibody and antigen were obtained directly in excess antibody (Table 1) and indirectly in excess antigen (Table 2), using a procedure described by Heidelberger and Kendall (1935); see also Kabat and Mayer (1948).

ANTIBODY AS ENZYME-INHIBITOR (see also Cinader, 1957)

The enzyme action of ribonuclease is inhibited by antibody. This is best shown by adding increasing amounts of the antibody to a constant amount of enzyme (Cinader, 1953, 1955). A plot of enzyme activity, as a function of the quantity of antibody added, shows initially a linear decrease and finally levels out in excess antibody at a fairly constant residual activity. This final constant level of activity is not the same for every serum and varies during the course of immunization. Initially only 60-70% of the enzyme activity may be inhibited by excess antibody, but antibody formed after repeated courses of immunization inhibits 95-98% of the activity of the enzyme. The changes in inhibitory power of a series of bleedings from a rabbit are shown in Fig. 3. The rabbit was first injected subcutaneously with Freund's adjuvant and received later successive courses of intravenous injections. The marked change of inhibitory power is not due to a decrease

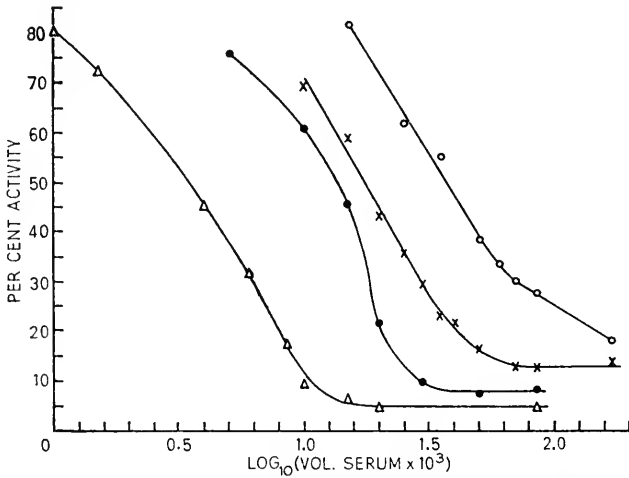


Fig. 3. Changes in inhibiting potency of antibody produced in response to repeated courses of immunization.

The effect of the addition of increasing volumes of immune serum to constant quantities of bovine ribonuclease.

Ribonuclease antisera from rabbit 619

Symbol	Date of bleeding
○	10.7.54
×	2.8.54
●	22.9.54
△	1.6.55

Table 1

PRECIPITATION OF RIBONUCLEASE BY RABBIT ANTIBODY

Addition of increasing amounts of ribonuclease to 0.05 ml. of serum 619
Rn = Ribonuclease

Antigen N ($\mu\text{g.}$)	Total N precipitated ($\mu\text{g.}$)	Ratio antibody N to antigen N in precipitate	Molecular ratio* (antibody to antigen) in:		Tests on supernatant			
			Precipitate	Super-natant	Antigen		Antibody	
					Flocculation	Enzyme activity as $\mu\text{g. N RnA}$ per 3 ml.	Flocculation	Agglutination titre
1.0	34.4	33.4	2.86	—	—	—	+	
2.0	66.8	32.6	2.79	—	—	—	+	200
3.05	101.0	32.1	2.74	—	—	—	+	200
3.95	118.9	29.1	2.49	—	—	—	+	200
4.9	135.2	26.4	2.26	—	—	—	+	50
5.9	148.2	24.0	2.05	—	—	—	—	50
7.4	177.6	22.9	1.96	—	—	—	—	10
8.9	185.7	19.9	1.70	—	—	—	—	1
9.8	191.4	18.5	1.58	—	—	—	—	< 1
10.2	188.4	17.7	1.51	—	+	0.5	—	< 1
13.2	138.5	15.9	1.36	0.87	+	3.8	—	< 1
15.3	106.4	16.7	1.43	0.75	+	7.2	—	< 1
17.3	83.5	17.8	1.52	0.71	+	10.2	—	< 1
19.3	69.2	17.0	1.45	0.65	+	13.6	—	< 1

* Calculated assuming identical N content for ribonuclease and antibody, a molecular weight 13,680 for ribonuclease (Hirs, Moore and Stein 1956) and of 160,000 for rabbit antibody.

Table 2

MOLECULAR RATIO (RABBIT ANTIBODY/RIBONUCLEASE A) IN REGION OF ANTIGEN EXCESS; SERUM 619

Rn = Ribonuclease A

Rn N added to 0.16 ml. of serum ($\mu\text{g.}$)	Total N in ppt. ($\mu\text{g.}$)	Specific N in supernatant ($\mu\text{g.}$)	Vol. anti-serum added to supernatant (ml.)	Total N in ppt.* from analysis of supernatant ($\mu\text{g.}$)	Antibody N precipitated† from serum added to supernatant ($\mu\text{g.}$)	Per cent Rn N in 2nd ppt.	Rn N in entire supernatant ($\mu\text{g.}$)	Rn N in ppt. ($\mu\text{g.}$)	Molecular ratio of antibody to antigen in ppt.
32.6	602.7	11.0	0.028	9.0	1.7	2.2	0.4	32.3	1.51
42.4	443.1	180.4	0.12	350.2	229.9	3.1	19.3	26.3	1.36
48.9	340.5	289.5	0.188	594.6	401.5	3.4	30.65	18.3	1.43
55.4	267.2	369.3	0.228	796.6	550.4	3.45	41.2	14.2	1.52
61.9	221.5	421.5	0.252	936.7	655.7	3.5	49.6	12.3	1.45

* All values determined on 2/3 of supernatant.

† From serum added to supernatant.

in soluble antigen-antibody complex but to a decrease in the enzyme activity of the precipitates of antigen and antibody.

So far little is known from direct experiments about the specificity of antibodies to native proteins. Experiments by Landsteiner and van der Scheer (1934) and of Landsteiner (1945) with peptide azo proteins have indicated that the specificity of the antibody may correspond to at least 5 amino acids, but this is a minimum estimate in a system which may not represent an adequate model for a native protein. The work of Porter presented in this symposium (p. 290) and observations recently published by Lapresle (1955) thus constitute a valuable advance of this problem.

The studies of Moore and of his colleagues are rapidly unravelling the detailed structure of ribonuclease, thus presenting us with known polypeptides which might be used to inhibit the reaction between ribonuclease and its antibody. The enzyme nature of ribonuclease would permit studies in which antibody action could not only be inhibited in terms of precipitation between enzyme and antibody, but also in terms of the effect of antibody on enzyme action.

Thus the collaboration of enzyme-chemist and immuno-chemist may not only advance our knowledge of antibody specificity but may even result in some information on the nature of the catalytic site of enzymes.

REFERENCES

- BOYDEN, S. V. (1951), *J. exp. Med.*, **93**, 107.
CINADER, B. (1953), *Immunochemistry*, Biochemical Society Symposia No. 10, Cambridge University Press, Cambridge, p. 16.
CINADER, B. (1955), *Bull. Soc. Chim. biol., Paris*, **37**, 761.
CINADER, B. (1957), *Ann. Rev. Microbiol.*, **11**, 371.
CINADER, B. and DUBERT, J. M. (1955), *Brit. J. exp. Path.*, **36**, 515.
CINADER, B. and DUBERT, J. M. (1956), *Proc. roy. Soc. B.*, **146**, 18.
CINADER, B., RONDLE, C. and PEARCE, J. H. (1955), 3rd Int. Congr. Biochem., 14-43.
CINADER, B. and PEARCE, J. H. (1956), *Brit. J. exp. Path.*, **37**, 541.
COHN, M. (1952), *Methods in Medical Research*, **5**, Year Book Publ. Inc., Chicago.
COHN, M. (1954), *Serological Approaches to Studies of Protein Structure*, p. 38, ed. by William H. Cole; Rutgers University Press, New Brunswick, New Jersey.
GRABAR, P. and WILLIAMS, Jr., C. A. (1953) *Biochim. biophys. Acta*, **10**, 193.
HEIDELBERGER, M. and KENDALL, F. E. (1935), *J. exp. Med.*, **62**, 697.
HEIDELBERGER, M. (1954), *Serological Approaches to Studies of Protein Structure*, p. 10. Ed. by William H. Cole; Rutgers University Press, New Brunswick, New Jersey.
HIRS, C. H. W., MOORE, S. and STEIN, W. H. (1953), *J. biol. Chem.* **200**, 193.
HIRS, C. H. W., MOORE, S. and STEIN, W. H. (1956), *J. biol. Chem.*, **219**, 623.
KABAT, E. A. and MAYER, M. M. (1948), *Experimental Immunochemistry*, Charles C. Thomas, Springfield, Ill.
LAPRESLE, C. (1955), *Ann. Inst. Pasteur*, **89**, 654.
LANDSTEINER, K. and VAN DER SCHEER, J. (1934), *J. exp. Med.*, **59**, 769.
LANDSTEINER, K. (1945), *The Specificity of Serological Reactions*, p. 310, Harvard University Press, Cambridge, Mass.
OUDIN, J. (1952), *Methods in Medical Research*, **5**, 335, Year Book Publ. Inc., Chicago.
SOBEY, W. R. (1954), *Aust. J. biol. Sci.*, **7**, 111.

Tobacco Mosaic Virus

Degradation and structure of tobacco mosaic virus

H. FRAENKEL-CONRAT AND K. NARITA

Virus Laboratory, University of California,
Berkeley, California, U.S.A.

Thanks to electron microscopy the shape of the tobacco mosaic virus rod is now so well known, as to require hardly any recapitulation. The 300 m μ long particle consists to approximately 94% of protein and to 6% of nucleic acid. The protein is generally regarded as forming the outer shell around the nucleic acid cylinder, although recent X-ray data indicate that the nucleic acid is deeply embedded in the protein, and may actually approach the periphery where the surface is deeply grooved. Within the past two years it has clearly been demonstrated that the nucleic acid alone can initiate infection and that it carries all the necessary 'information' for the production of complete progeny virus.¹⁻³ The efforts of virologists are, therefore, now centering on this crucial minor constituent of the virus. However, the role played by the protein moiety in various important regards must not be overlooked. In quantitative respects, the protein contributes to viral infectivity, since it potentiates the action of the nucleic acid 20-50 fold, possibly by giving it protection and rigidity.^{4,5} The protein also accounts for the serological activity of viruses and thus gives us a tool with which to combat them. And finally, the protein probably influences host range specificities and virulence. It is of particular interest that the protein components have been shown to differ in specific and characteristic manner from one virus strain to another.^{6,7} The protein may, therefore, be regarded as a phenotypic expression of the chemically still obscure genotypic characteristics of viruses. This symposium has clearly documented our advances towards the goal of complete structural analysis of proteins. The possibility exists, therefore, that the structure of the virus protein may be elucidated within the next few years. The virus may then become the Rosetta stone of biochemical genetics, in supplying the bilingual record needed to decipher nucleic acid in terms of protein structure.

While various estimates of the virus particle weight range from 40 to 50 million, it has been recognized for several years that each rod represents

an aggregate of much smaller protein and nucleic acid subunits or molecules. The minimal molecular weight of the protein subunit, based on amino acid analyses, is about 18,000. One cysteine and two lysine residues occur in all strains of TMV, and one histidine in the H.R. strain, in a unit of about 18,000 gms.^{4,6,7} That this approximates the true molecular weight of the smallest unit was strongly suggested by the discovery of Harris and Knight that carboxypeptidase releases only threonine and no other amino acid from the virus and that the rapidly reached maximum amount of threonine corresponds to one residue per 17,300^{8,9} or about 2900 residues per virus particle. It has since been established beyond doubt that this threonine residue represents a C-terminal group,^{10,11} and that the virus is actually composed of about 2900 separate peptide chains, held together only by secondary forces. Since 'macro-structure' is the main function of this protein, it is not surprising that those secondary binding affinities are much in evidence, so that the subunits are hard to separate, and even harder to keep separated under conditions favorable for physicochemical characterization. Thus, only rough approximations of the subunit weight have as yet been made on the basis of sedimentation studies,¹² but these support the concept that single peptide chains supply the bricks building up the supermolecule.

DEGRADATION OF VIRUS AND ISOLATION OF PROTEIN

A variety of methods has been used for the degradation of the intact virus and for the isolation of disaggregated nucleic acid-free protein. Detergents, such as sodium dodecyl sulfate (SDS), represent the most efficient agents for the disruption of all interunit bonds, but the protein obtained from such reaction mixtures shows many undesirable features, some of which will be referred to later. It appears that these agents break not only inter-chain but also intra-chain secondary bonds. Thus the product is not the expected mixture of sticky bricks, i.e., subunits of definite shape, but a mixture of sticky threads, i.e., denatured peptide chains, with a tendency to unlimited random aggregation into a most intractable and insoluble material.

Gentler methods for the isolation of the protein make use of alkali which was found by Stanley to degrade the virus.¹³ A commonly used procedure was proposed by Schramm *et al.*¹⁴ in which the virus is solubilized at about pH 10.5 in the cold. Carbonate, glycine, and more recently alkanolamine buffers have been used for this purpose.¹⁵ Harrington and Schachman have studied the kinetics of this degradation, and have demonstrated the great tendency for reaggregation of split products in the alkaline reaction mixtures, particularly at higher temperatures.¹⁶ Ammonium sulfate can conveniently be employed to separate the protein cleanly from the nucleic acid after all types of degradation. The final protein preparation obtained from alkali-degraded virus appears 'native' in various respects, to be discussed

below, and gives water-clear solutions except in the isoelectric region (pH 3.5-6.5). In that pH range there occurs a remarkable reaction: The solution becomes opalescent and its 'virus-like' appearance is borne out by electron micrographs which show that most of the material is again in the form of rods of great variability in length, but exactly the same diameter as the original virus. The shortest rods often stand upright and appear as circular disks with a central perforation. These have been referred to as doughnuts. Thus, the prime biological activity of this particular protein, namely its tendency to controlled aggregation resulting in a particular macro-structure, has not been destroyed in the course of degradation and separation of the nucleic acid. The sedimentation coefficient of this material in solution (usually determined at about pH 9) resembles that of Schramm's electrophoretically separated A-protein (about 4 *S*) and suggests a molecular weight of about 100,000 corresponding to six-chain aggregates. It appears quite possible but by no means proven that units of such a size are required for the proper functioning of this protein.

A convenient method for the breakdown of the virus and isolation of the protein has recently been found which has the particular advantage of great simplicity.¹⁷ The virus is degraded by means of cold 67% acetic acid, the precipitate which flocculates in a few minutes is centrifuged off (largely nucleic acid), and the protein which alone remains in the water-clear solution is isolated after dialysis. The resulting protein preparation appears indistinguishable in all respects from that obtained by alkali splitting and ammonium sulfate fractionation. This is at present the method of choice in our laboratory.*

TERTIARY STRUCTURE OF VIRUS PROTEIN

From the preceding description it is clear that the protein as isolated from dilute alkali or from acetic acid must retain a definite shape which enables it to bridge the gulf between a 100,000 molecular weight 'pseudoglobulin' and a virus rod particle with a mol. wt. of 50 million. What type of bonds are involved in maintaining its structure? This question has repeatedly been attacked in our laboratory. One amino acid side chain which is very probably implicated in this structure is the cysteine-SH group. This group was found to exist in a form which protects it from many —SH reagents of various types, including oxygen, and thus renders it as stable as the masked —SH groups of many other proteins.^{19,20} On the other hand the —SH group can readily be titrated after unmasking it by the usual denaturing agents (such as guanidine hydrobromide), and represents a convenient means of estimating as to what extent a protein preparation has retained its original native character. It was found to be quantitatively present (one per 18,000

* It has only recently come to our attention that acetic acid was used by Bawden and Pirie in 1937 for the degradation of TMV, although not as a means of preparing the pure virus protein.¹⁸

mol. wt.) in the two types of soluble protein described above,¹⁷ but absent, presumably autoxidized, in the detergent type of preparation.

Further studies of this group have shed new light on its mode of linkage, and probably of masked protein —SH groups generally.²¹ For it has become evident that the H atom of this masked group can be replaced by CH₃Hg— or by an iodine atom without apparent disturbance of protein structure. The substitution by iodine is particularly surprising since sulfenyl-iodides are normally quite unstable and decompose within a few seconds in aqueous solvents. In the native virus, in contrast, the sulfenyl-iodide group is stable, but decomposes rapidly if the virus is denatured by agents such as guanidine hydrobromide.

It must be concluded from these results that the masked —SH groups (1) occur as such in the protein and (2) probably represent the acceptors rather than donors of hydrogen-bonded structures involving other as yet

unknown groups ($\begin{array}{c} \text{H} \\ \diagdown \\ \text{---S---H---X} \end{array}$). This linkage must be of unusual firmness in the present case since the masked —SH groups of ovalbumin and a few other proteins that were tested did not give stable sulfenyl-iodide derivatives, although their reactivity otherwise resembled that of TMV.

Another type of tertiary bond, which is at present receiving much attention from protein chemists, is the hydrogen-bonded interaction of the phenolic groups of tyrosine residues with the carboxylate ion groups of glutamic or aspartic acid residues. Evidence for the presence of such groups is often based on the spectral shift to longer wavelengths observed for tyrosine so bonded. If this spectral shift is largely reversed in acid solution, and if the pK of this effect is within the range of that of carboxyl groups, then a tyrosine-carboxylate interaction is postulated.^{22,23} In the case of the native TMV protein the spectral divergence from that of an *ad hoc* prepared amino acid mixture has already been pointed out⁴ as well as the correction of this divergence by the use of an acidic solvent. Difference spectra have now been plotted for the protein at pH 7 as compared to pH 2.2, and these were found to resemble closely those obtained with ribonuclease, insulin, etc. (Fig. 1). Thus it appears probable that bonds of the tyrosine-carboxylate type do occur in the isolated virus protein. It must be noted, however, that the making and breaking of these bonds appears readily reversible and that they are not stable under some, if any, of the conditions required for the degradation of the virus prior to isolation of the protein. Thus it would appear impossible to attribute to these bonds any major role in maintaining the original folding and shape of the protein. Nor can they play part in the inter-unit bonding, unless originally intermolecular bonds of this type are presumed to have exchanged partners during degradation of the virus to give the possibly weaker intra-molecular bonds observed in the final preparation. A search for spectral differences between the intact virus and that

degraded by acid (+SDS) require correction for the light scattering by the intact virus. Comparison of the spectra of virus which was degraded by SDS either in an acid solution or at pH 8 need no such correction. In both cases a shift in the opposite direction from that given by the protein alone was obtained, the acid solution giving the positive difference spectrum with a maximum about 290 $m\mu$. However, virus nucleic acid alone was found to show a similar behavior, and its presence makes such reaction mixture too complex to attempt any interpretation of the spectral changes.

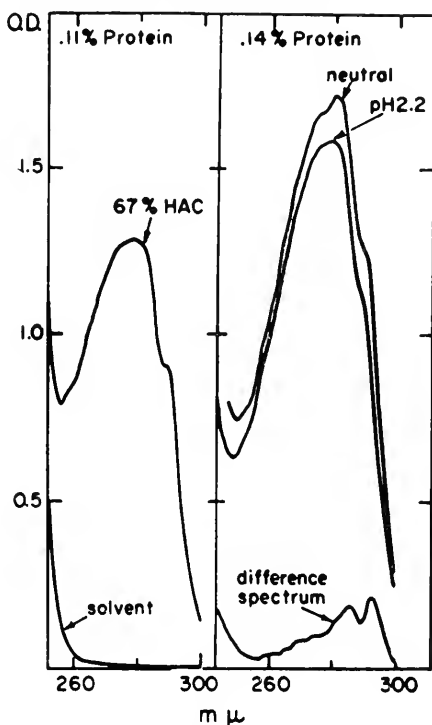


Fig. 1. Ultraviolet absorption of TMV-protein in 67% acetic acid, HCl (pH 2.2) and in water at neutrality.

NATURE OF INTER-UNIT BONDS

Previous studies by Harrington, Levy and Schachman²⁴ have indicated that OH^- ions are rapidly consumed when TMV is degraded in weakly alkaline solutions. These studies have now been extended to other experimental conditions leading to the splitting of the virus. One type of experiment is based on the new finding that the virus is degraded rapidly at room temperature in 1% SDS adjusted to pH 3.5 with dilute HCl. (Nucleic acid of similar or slightly higher infectiousness as the standard preparations can be isolated from such reaction mixtures, as well as nucleic acid-free, though denatured,

protein.) When the reaction mixture containing the degraded virus was subsequently titrated back to the original pH (usually pH 7.0), more alkali was required than that equivalent to the acid used to reach the pH at which the virus was degraded. The difference was similar to the alkali consumption observed in the earlier experiments,²⁴ and corresponded to about 1.5 μ mole per 18 mg virus. Subsequent titrations between pH 7.0 and 3.5 required equivalent amounts of acid and alkali, and followed the same curve as the first back titration (Fig. 2). An alternative procedure is to perform

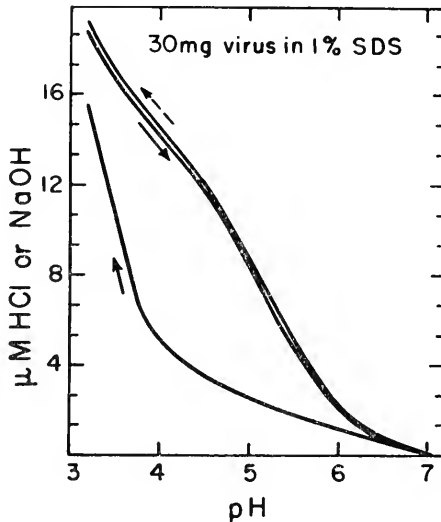


Fig. 2. Alkali consumption during degradation of TMV by 1% sodium dodecyl sulfate below pH 4, as indicated by back titration to original pH.

the customary degradation with SDS at about pH 7.5 and 40° in a pH-stat and note the amount of alkali required to maintain a constant pH. Such experiments have given an average alkali consumption of 1.9 μ mole/18 mg virus (Fig. 3).

A working hypothesis which could account for these observations is as follows: A certain number of carboxyl groups of the virus (probably 2 per subunit) are hydrogen bonded in the undissociated form to some other,

possibly tyrosine, groups ($\begin{array}{c} \text{OH} \cdots \text{X} \\ | \\ \text{---C=O} \cdots \text{H} \end{array}$). This bond is so stable as to prevent the carboxyl groups from dissociating until disaggregation of the macro-molecule re-establishes their normal pK and titrability. Carboxyl-tyrosine bonds have been proposed by Edelhoich for pepsin,²⁵ and are in that case believed to be stable up to about pH 6, at which point pepsin becomes denatured. In the case of TMV one would have to assume an extension of the pH stability range of such groups to almost pH 9. Inter-

actions between tyrosine and undissociated carboxyl groups have been observed also in model experiments of Laskowski,²⁶ and there are indications that they may produce much smaller spectral shifts than does the carboxylate-tyrosine interaction. However, such an effective masking of protein carboxyl groups has not previously been demonstrated and will have to be supported by much more experimental evidence than is available at present.

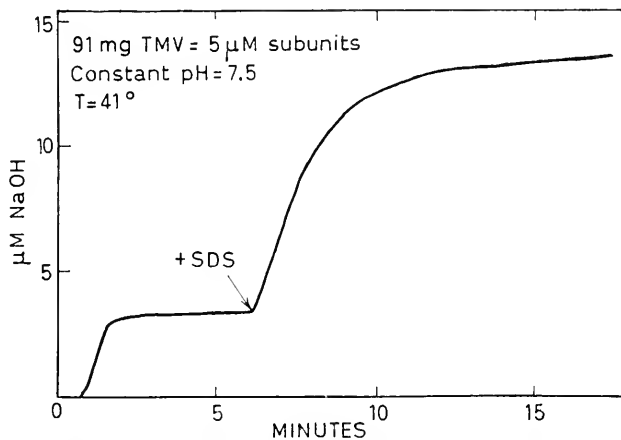
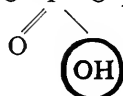


Fig. 3. Alkali consumption during degradation of TMV in 1% sodium dodecyl sulfate at a constant pH (7.4).

A recent series of observations appears to both confirm and extend the information concerning particular inter-subunit carboxyl binding sites in the virus. The X-ray diffraction data of Caspar²⁷ have shown that lead is bound by the virus stoichiometrically, and in two distinct localizations. Unpublished observations from this laboratory have indicated that both these sites differ from the SH group which is not substituted by lead. It has also been observed that the addition of lead acetate (pH 6) to virus at pH 6 produces a sharp drop in pH which is compensated by an equivalent amount of alkali. This could be regarded as evidence that lead can replace the two non-dissociated carboxyl hydrogen atoms of each subunit. Support for this concept comes from the finding that no alkali is consumed upon the subsequent addition of SDS, or at pH 9 (40°). The supposed replacement of the hydrogen bond by a lead atom does not weaken, but rather strengthens, the structure of TMV as indicated by the extent of splitting of protein and nucleic acid under standard conditions of degradation. Other divalent metals giving stable acetates (pH 5.5-6.5) such as zinc and magnesium, react similarly to lead, although with decreasing affinity. This was indicated by slower and less stoichiometric interaction (approaching only one site with Mg^{++}), and less stabilization upon splitting. Also subsequently added lead produced a further pH drop, while other metal salts added after lead acetate had no such effect. The particular affinity of lead for carboxyl groups is in accord with the

finding of Gurd and Murray,²⁸ but this appears to be the first case where lead and to a lesser extent other divalent ions can enter into protein structure by replacing a normally non-dissociable carboxyl hydrogen atom.

One alternative possibility which could account for the presented observations is that about 5000 secondary phosphate—OH groups ($-\text{R}-\text{O}-\text{P}-\text{O}^-$),



rather than carboxyl groups, prevented from dissociation by their involvement in structural stabilization, might acquire their normal dissociability upon degradation of the virus. Since each virus particle contains only about 10,000 nucleotides, most if not all phosphate-diester linked, this explanation appears to be ruled out. Unusual bonds as numerous as the titrations indicate can only be located in the protein. And unusual bonds of some kind or other will probably be required to explain the unusual stability of the intact virus as compared with its major component, the protein. Simple aggregation, as observed with so many proteins, does not generally evoke major changes in character. Yet the virus is stable at temperatures above 60°, while the isolated protein becomes denatured below 40°.²⁹ Differences in resistance towards enzymes and other agents may be expressions of the same phenomenon. The fact that the ease of degradation of the virus differs for different strains, as well as for chemical derivatives of the same strain, also points to the existence of a definite determining and limiting factor in the inter-unit bonding. It has also been noted that the electrophoretic mobilities of the virus, as well as of various virus protein aggregates, are often higher, i.e., the protein is less acidic than one would expect from the amino acid composition. An unexplained shift in the opposite direction was observed for the enzymatically dethreoninated virus.⁸ It would appear that carboxyl-masking could be a determining factor in the observed electrophoretic differences, as well as in immunological and other particle surface properties.

STUDIES OF THE PRIMARY STRUCTURE OF TMV

During the past years various studies have been initiated by several investigators in the University of California Virus Laboratory to solve structural and peptide sequence problems connected with the TMV protein. The approach is by now almost conventional. The steps involved are, (1) enzymatic degradation of the protein; (2) separation of pure peptide fragments; (3) elucidation of their composition and sequence; and (4) fitting together of the pieces on the basis of overlaps supplied by the two or more sets of peptides obtained through the use of two or more different enzymes.

(1) Considerable effort has been made to use the SDS-denatured protein, which is a by-product of the nucleic acid preparation, for these studies. However, the susceptibility of this material to enzymatic digestion has been

found to be somewhat erratic. Digestion of the acetic-acid prepared protein with 1% trypsin (Worthington, cryst.) at 37°, i.e., near its point of heat denaturation, for about 8 hours has given the most reproducible results. The ninhydrin analyses indicate the splitting of a maximum of about 12 bonds (leucine equivalents). The C-terminal groups found in hydrazinolysis experiments in the acid-soluble fraction of the digest were, besides the original C-terminal threonine, in average about 9 arginine, 1.3 lysine, and 0.25 residues of leucine and serine. This indicates a high degree of specificity for the enzyme used, but falls slightly short of quantitative digestion [expected: 11 arginine, 2 lysine groups³⁰]; but the high correction factors, particularly for arginine,³¹ lessen the strictly quantitative significance of these values. The N-terminal groups found^{30,32} are about one each of aspartic and glutamic acid, serine, threonine, phenylalanine, and tyrosine, possibly two of glycine, and less than one of alanine, valine and leucine, in general quantitative accord with the sum of C-terminal residues.

Chymotrypsin digestion liberated about 14 leucine equivalents. C-terminal residues were (besides the original threonine) phenylalanine (6 residues), leucine and isoleucine (3), tyrosine (3), tryptophan (about 1) [total of these amino acids: 8, 13+9, 4 and 2³⁰], and traces (0.2) of serine, glycine, valine and proline. The N-terminal residues found in such digests were aspartic and glutamic acid (about 4 and 2), glycine and serine (2), and about one each of threonine, valine, alanine, arginine, leucine and phenylalanine.

(2) For separation of the peptides of TMV digests Dowex 50-X-2 columns, as advocated by Hirs, Moore and Stein,³³ have been used with some success.³² Several sharp peaks, however, could be resolved into two or three components by subsequent paper chromatography. At present Dr D. Gish is subjecting the digests to countercurrent distribution and appears to get good resolution of the mixture, although here also seemingly homogeneous peaks may occasionally contain more than one component. A two-dimensional paper electrophoresis—paper chromatography technique is being applied to the problem by Knight and Woody.³⁴ A reproducible pattern of spots indicating about 12 peptides is revealed by the ninhydrin and/or the chlorine-iodide test.³⁵ This technique has been used to survey a number of virus strains in comparative fashion.³⁴ Indistinguishable peptide patterns were obtained for strains of very similar amino acid composition (masked strain vs. common TMV), while differences were observed for strains showing different composition.

Most of the peptides obtained from similar digests by the various techniques of isolation discussed above have been cross-identified.

(3) The amino acid composition of most peptides has been determined by the Levy fluorodinitrobenzene (FDNB) method^{36,37} after desalting the preparations according to Hirs *et al.*,³³ or more recently by the method of the Haugaards.³⁸ The N-terminal amino acids were determined by standard procedures.³⁷ The C-terminal amino acid was generally established by the

enzyme specificity, but was occasionally ascertained by hydrazinolysis^{31,39} or carboxypeptidase. Sequences have been determined in only a limited number of peptides, ranging from dipeptides to hexapeptides.^{30,32} One hexapeptide isolated from chymotryptic digests as the DNP derivative appeared of particular interest because it proved to represent the *C*-terminal peptide of the protein.³⁹ The same hexapeptide was obtained from the masked and YA strains, but one of slightly different composition was obtained from the H.R. strain.

(4) The identified fragments are devoid of key amino acids and too small to permit even a beginning of the final stage of the work, the fitting together in proper sequence. The disturbing possibility must also be faced that all peptide chains need not necessarily be identical. An appreciable heterogeneity within that population would render our aim, as now approached, unrealizable. At present, one can only be sure that all or almost all peptide chains have the same *C*-terminal sequence. It is also definitely established and agreed^{40,41} that none or almost none of the chains has a free *N*-terminal group. Several preliminary notes by Braunitzer and Schramm discussing the nature of the 'hidden' *N*-terminal group and describing a crystalline *N*-terminal trypsin fragment of the protein have not been confirmed. Some of the claims and ideas were most attractive, but they had to be discarded for lack of factual support.

The absence of an *N*-terminal group raises the possibility that digestion of the protein by enzymes other than trypsin might produce peptides which are devoid of any basic group. A search for such peptides was therefore initiated by one of us (K.N.)⁴² by passing various digests through the acid form of Dowex-50-X2 columns. The same main component of this acid peptide fraction was isolated from either chymotryptic or peptic digests, and

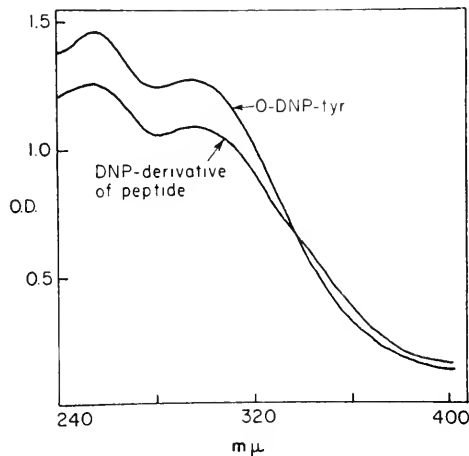


Fig. 4. Ultraviolet absorption spectrum of DNP-derivative of acidic peptide, and of *O*-DNP-tyrosine.

further purified by Dowex-2-X10 chromatography. It proved to contain only two amino acids, having the composition ser,tyr. Both hydrazinolysis and carboxypeptidase proved that the tyrosine was C-terminal which also is in accord with known enzyme specificities. However, the status of the N-terminus of this simple peptide represented a considerable mystery. As stated, its mode of isolation precluded the presence of a basic group. Furthermore, the peptide did not give any ninhydrin color, nor a yellow DNP-derivative, although it did give a colorless ether-soluble derivative which showed the typical spectrum of O-DNP tyrosine (Fig. 4). Previous suggestions that the missing N-terminus could be due to a loop peptide had to be abandoned in view of the absence of aspartic or glutamic acid. It seemed that only the presence of an acyl substituent on the amino group could account for the observed properties of the acidic peptide. To search for such a component, the peptide was hydrazinolyzed and the resultant mixture of hydrazides (+tyrosine) analysed by paper chromatography in search for any neutral acyl-hydrazides. As shown on Table 1, the main component, as detected with ammoniacal silver nitrate, showed R_f values, in two solvents, which were identical with those of acetyl-hydrazide.

Table 1
 R_f VALUES OF SEVERAL HYDRAZIDES*

A: Solvent: Pyridine·Aniline·Water (9 : 1 : 4, by vol.)			
	Standard substance	Hydrazinolysate of sample	
Formic hydrazide	0·58	—	
Acetic "	0·66	0·66	
Propionic "	0·75	—	
Serine "	0·51	0·51	
Hydrazine "	0·39	0·40	

B: Solvent: Collidine·Water (10 : 2, by vol.)			
	Standard substance	Hydrazinolysate of sample	Hydrazinolysate of seryl-tyrosine
Formic hydrazide	0·28	—	—
Acetic "	0·35	0·33	—
Propionic "	0·50	—	—
Serine "	0·074	0·073	0·074
Hydrazide	0·00	0·00	0·00
Decomposition product of serine or its hydrazide	—	0·28	0·27

* Detecting reagent: Ammoniacal silver nitrate reagent (0·1 N—AgNO₃, 0·5 N—NH₄OH, 1 : 1 by vol.).

Direct chromatography of a hydrolysate of the acidic peptide also indicated the presence of acetic acid. It thus appears established, that the structure of the acidic peptide is acetyl-ser-tyr. Thus acetylation may well be responsible for the absence of *N*-terminal groups in TMV. The above peptide probably represents the acetyl-*N*-terminal sequence of the virus protein, although its occurrence as a branch chain is not excluded. This is believed to be the first instance* where an acetyl group has been definitely shown to be a component of a simple protein, and has been allocated to a specific site in that protein.

In conclusion, then, TMV protein seems to be composed of peptide chains of a molecular weight of about 18,000 which instead of a free amino end terminate in acetyl-ser-tyr- and which have a thr-gly-ser-pro-ala-thr sequence at the carboxyl end. About six polypeptide chains are united into a definite aggregate (A-protein) resembling native proteins in having masked —SH groups and tyrosine-carboxylate bonds. The A-protein tends to aggregate to particles resembling in shape the original virus, but lacking its stability.

It appears probable that the —SH group (one per chain) is hydrogen-bonded, playing the role of an acceptor, since its H atom is free. In the degradation of the virus, about 5000 new acid groups appear and become titrable which may be originally present as masked carboxyl groups, hydrogen bonded in their undissociated form. To these groups may be attributed the remarkable stability and other surface properties which differentiate the intact virus from the isolated protein. These same groups appear to exchange their hydrogen for lead and with lesser affinity other divalent metals.

REFERENCES

1. H. FRAENKEL-CONRAT, *J. Am. Chem. Soc.*, **78**, 882 (1956).
2. A. GIERER and G. SCHRAMM, *Z. f. Naturforsch.*, **11b**, 138 (1956).
3. H. FRAENKEL-CONRAT, B. SINGER and R. C. WILLIAMS, In: W. D. McElroy and B. Glass, McCollum Pratt Symp., *The Chemical Basis of Heredity*. Baltimore: Johns Hopkins Press, p. 501 (1957). Also: *Biochim. et Biophys. Acta*, **25**, 87 (1957).
4. H. FRAENKEL-CONRAT and B. SINGER, *Biochim. et Biophys. Acta*, **24**, 540 (1957).
5. H. FRAENKEL-CONRAT and B. SINGER, *Spec. Pull. N.Y. Acad. Sci. V*, 217 (1957).
6. C. A. KNIGHT, *J. Biol. Chem.*, **171**, 297 (1947).
7. F. L. BLACK and C. A. KNIGHT, *J. Biol. Chem.*, **202**, 51 (1953).
8. J. I. HARRIS and C. A. KNIGHT, *Nature*, **170**, 613 (1952).
9. J. I. HARRIS and C. A. KNIGHT, *J. Biol. Chem.*, **214**, 215 (1955).
10. G. BRAUNITZER, *Z. f. Naturforsch.*, **9b**, 675 (1954); *Ber.* **88**, 2025 (1955).
11. C.-I. NIU and H. FRAENKEL-CONRAT, *Biochim. et Biophys. Acta.*, **16**, 597 (1955).
12. H. K. SCHACHMAN, unpublished.
13. W. M. STANLEY, *Ergebnisse d. Physiol., Biol. Chemie und Exp. Pharmakol.*, **39**, 294 (1937).
14. G. SCHRAMM, G. SCHUMACHER and W. ZILLIG, *Z. f. Naturforsch.*, **10b**, 481 (1955).

* See, however, article by J. I. Harris (p. 335).—ED.

15. P. NEWMARK and R. W. MYERS, *Federation Proc.*, **16**, 226 (1957).
16. W. F. HARRINGTON and H. K. SCHACHMAN, *Arch. Biochem. and Biophys.*, **65**, 278 (1956).
17. H. FRAENKEL-CONRAT, *Virology*, **4**, 1 (1957).
18. F. C. BAWDEN and N. W. PIRIE, *Proc. Royal Soc.*, London, Series B, **123**, 274 (1937).
19. W. M. STANLEY and M. A. LAUFFER, *Science*, **89**, 345 (1939).
20. M. L. ANSON and W. M. STANLEY, *J. Gen. Physiol.*, **24**, 679 (1941).
21. H. FRAENKEL-CONRAT, *J. Biol. Chem.*, **217**, 373 (1955).
22. M. LASKOWSKI, JR., J. M. WIDOM, M. L. MCFADDEN and H. A. SCHERAGA, *Biochim. et Biophys. Acta.*, **19**, 581 (1956).
23. H. A. SCHERAGA, *Biochim. et Biophys. Acta.*, **23**, 196 (1957).
24. W. F. HARRINGTON, A. L. LEVY and H. K. SCHACHMAN, personal communication.
25. H. EDELHOCH, *Biochim. et Biophys. Acta*, **22**, 401 (1956).
26. M. LASKOWSKI, JR., *Am. Chem. Soc. Abstracts*, 47C, Miami, (1957).
27. D. L. D. CASPAR, *Nature*, **177**, 928 (1956).
28. F. R. N. GURD and G. R. MURRAY, JR., *J. Am. Chem. Soc.*, **76**, 187 (1954).
29. P. STARLINGER, *Z. f. Naturforsch.*, **10b**, 339 (1955).
30. L. K. RAMACHANDRAN, in press.
31. C.-I. NIU and H. FRAENKEL-CONRAT, *J. Am. Chem. Soc.*, **77**, 5882 (1955).
32. C.-I. NIU and H. FRAENKEL-CONRAT, unpublished.
33. C. H. W. HIRS, S. MOORE and W. H. STEIN, *J. Biol. Chem.*, **219**, 623 (1956).
34. C. A. KNIGHT, in *Ciba Foundation Symp. on The Nature of Viruses*. London: Churchill Press, p. 69 (1956).
35. H. N. RYDON and P. W. G. SMITH, *Nature*, **169**, 922 (1952).
36. A. L. LEVY, *Nature*, **174**, 126 (1954).
37. H. FRAENKEL-CONRAT, J. I. HARRIS and A. L. LEVY, in *Methods of Biochem. Analysis*. New York: Interscience Publishers, **2**, 359 (1955).
38. N. HAUGAARD and E. S. HAUGAARD, *Compt. rend. trav. lab. Carlesberg Sér. chim.*, **29**, 347 (1955).
39. C.-I. NIU and H. FRAENKEL-CONRAT, *Arch. Biochem. and Biophys.*, **59**, 538 (1955).
40. H. FRAENKEL-CONRAT and B. SINGER, *J. Am. Chem. Soc.*, **76**, 180 (1954).
41. G. SCHRAMM, J. W. SCHNEIDER and A. ANDERER, *Z. f. Naturforsch.*, **11b**, 12 (1956).
42. K. NARITA, *Biochim. et Biophys. Acta.*, in press.

Strukturuntersuchungen am Protein des Tabakmosaikvirus

G. SCHRAMM, G. BRAUNITZER,
F. A. ANDERER, J. W. SCHNEIDER
UND H. UHLIG

Max-Planck-Institut für Virusforschung, Tübingen

Bei der Infektion einer Tabakpflanze mit der Protein-freien Ribonucleinsäure (RNS) aus Tabakmosaikvirus (TMV) wird in der Zelle das spezifische Virusprotein produziert.¹ Es müssen also mehr oder weniger direkte Beziehungen zwischen der Struktur der RNS und dem gebildeten Protein bestehen. Es ist anzunehmen, dass auch die Struktur anderer in Organismen vorkommender Proteine durch spezifische Ribonucleinsäuren bestimmt wird. Aber nur im Falle des TMV gelingt es, sowohl die RNS als auch das Protein leicht in reiner Form zu gewinnen, so dass sich hier ein Beispiel bietet, das von allgemeiner Bedeutung für die Biosynthese der Proteine sein dürfte. Leider fehlt es bisher noch an chemischen Methoden zur Strukturaufklärung der RNS, so dass wir uns vorerst darauf beschränken müssen, den Proteinanteil näher zu analysieren.

Zu Beginn der Strukturuntersuchung eines Proteins erhebt sich stets die Frage nach der Einheitlichkeit des Untersuchungsmaterials. Das TMV lässt sich verhältnismässig leicht von den Proteinen der Wirtspflanze abtrennen. Wir prüften durch Komplementbindung die Reaktion eines Antiserums gegen Normalprotein mit dem Virus und auch umgekehrt die Reaktion von normalem Pflanzenprotein gegen Antiserum, das gegen unser Viruspräparat hergestellt worden war. In beiden Fällen ergab sich übereinstimmend, dass die auf dem üblichen Wege (Ammonsulfatfällung, wiederholte Ultrazentrifugation) hergestellten Viruspräparate nicht mehr als 0,2% Normalprotein enthielten.²

Schwieriger ist die Frage zu beantworten, wieweit die Virusteilchen unter sich einheitlich sind. Da beim TMV zahlreiche Mutanten bekannt sind, die sich in ihren biologischen Eigenschaften unterscheiden, müssen wir annehmen, dass jedes Viruspräparat eine Population mehr oder weniger nahe verwandter Stämme darstellt. Auch durch Einzelherd-Isolierung lassen sich nicht mit

Sicherheit reine Stämme gewinnen, da bei der nachfolgenden Vermehrung des isolierten Stammes die Möglichkeit zu Mutationen gegeben ist.

Gewisse Versuche deuten darauf hin, dass auch ein einzelnes TMV-Stäbchen hinsichtlich seiner Struktur nicht homogen ist. Bei der Behandlung des TMV mit mildem Alkali wird die äussere Proteinhülle abgestreift und es entsteht das sogenannte A-Protein. Schramm und Mitarbeiter^{3,4} fanden, dass 70% des Proteins verhältnismässig leicht von der Nucleinsäure getrennt werden können, bei 30% des Materials ist hingegen die Bindung des Proteins an die RNS so fest, dass es auch bei längerer Behandlung bei pH 10,3 nicht abgespalten werden kann. Dieser Anteil wurde als stabile Fraktion bezeichnet. Die Ergebnisse wurden inzwischen von Harrington und Schachman⁵ bestätigt. Die Autoren geben eine Erklärung, die uns sehr plausibel erscheint. Danach wird von einem Ende her das A-Protein von dem Nucleinsäurestrang abgestreift, bis etwa 2/3 entfernt sind, das Rest-Segment liefert die stabile Fraktion. Die unterschiedliche Stabilität könnte darauf beruhen, dass die Nucleinsäurezusammensetzung sich entlang der Achse ändert; es können aber auch Unterschiede in den Peptiduntereinheiten vorliegen. Bei anderen Stämmen des TMV findet sich ebenfalls eine stabile Fraktion, der mengenmässige Anteil ist jedoch unterschiedlich, bei dem Stamm Dahlemense beträgt er nur 12%. Aus diesen Gründen dürfen wir nicht ohne weiteres voraussetzen, dass wir völlig eindeutige Resultate bei der Bestimmung der Aminosäure-Sequenzen erhalten. Es erscheint uns jedoch sinnvoll aufzuklären, wieweit die Übereinstimmung geht.

DIE GRÖSSE DER PEPTID-UNTEREINHEITEN BEIM TMV

Eine genaue Bestimmung des Molgewichts des TMV ist sehr schwierig. Für die folgenden Darlegungen wird der von Schramm und Bergold⁶ bestimmte Wert von $40 \cdot 10^6$ zugrunde gelegt. Das A-Protein, das beim alkalischen Abbau des TMV erhalten wird, hat nach Diffusions- und Sedimentationsmessungen ein Molgewicht von etwa $90\,000 \pm 10\%$. Es handelt sich hier um ein natives Protein, das unter geeigneten Bedingungen wieder zu Stäbchen von der Form und Grösse des Virus aggregiert werden kann. Wird bei der Darstellung pH 10,6 überschritten, so denaturiert das A-Protein und wird unlöslich. Nach einer persönlichen Mitteilung von P. Newmark^{3,4} lassen sich beim TMV auch native Untereinheiten mit Molgewichten unter 90 000 darstellen, wenn man die Salzkonzentration sehr niedrig hält und bis zur äussersten Grenze des erlaubten pH-Werts geht. In Gegenwart von Netzmitteln wurden von Schachman und Hersh⁷ Einheiten erhalten, die ein Molgewicht zwischen 10 und 20 000 besaßen.

Eine genaue Grösse der Peptiduntereinheiten liess sich durch Endgruppen-Bestimmung erhalten. Harris und Knight⁸ fanden, dass bei Einwirkung von Carboxypeptidase ein Mol Threonin je 17 300 g Virus freigesetzt wird. Etwa

264 SCHRAMM, BRAUNITZER, ANDERER, SCHNEIDER, UHLIG [15 den gleichen Wert fanden Schramm und Braunitzer⁹ an dem von ihnen untersuchten TMV-Stamm. Gegen die enzymatische Methode können nun gewisse Bedenken erhoben werden. Braunitzer¹⁰⁻¹² bestimmte daher die Carboxylendgruppe auf chemischem Wege mit der Hydrazinmethode von Akabori. Als einzige Endgruppe wurde ebenfalls Threonin und zwar in einer Menge von 1 Mol je 15 000 g Virus festgestellt. Bei kurzzeitiger Einwirkung von Hydrazin (3-4 Stunden bei 100°C) gelang es auch, das Dipeptid Ala-Thr und das Tripeptid Pro-Ala-Thr in Form ihrer Dinitrophenyl(DNP)-Derivate zu isolieren. Nach Abspaltung des Threonins mit Carboxypeptidase trat als einzige Endgruppe Alanin auf und zwar wiederum in einer Menge von ein Mol je 16 000 g Virus. Zu etwa den gleichen Ergebnissen kamen unabhängig hiervon Niu und Fraenkel-Conrat.¹³

Die Bestimmung der N-terminalen Konfiguration in der Peptid-Untereinheit des TMV bietet Schwierigkeiten. Im nativen Virus sind keine endständige Aminogruppen nachweisbar^{14,15}. Nach Behandlung des Virus mit Trichloressigsäure (pH 1,8, 30 Minuten bei 85°C) konnte jedoch Prolin als Amino-endständige Gruppe nachgewiesen werden. Die Bestimmung erfolgte nach den Methoden von Sanger¹⁶ und Edman.¹⁷ Mit beiden Methoden wurden Werte gefunden, die der Menge der Carboxylgruppen entsprechen. Die Schwierigkeiten bei dem Verfahren nach Sanger liegen darin, dass DNP-Prolin während der sauren Hydrolyse zu Dinitrophenol und einigen in geringen Mengen auftretenden Nebenprodukten gespalten wird.³³ Dieser Abbau kann teilweise unterdrückt werden, indem man die Hydrolyse in nahezu wasserfreiem Medium¹⁸ durchführt. Nach Kontrollversuchen liefern die anderen DNP-Aminosäuren bei der sauren Hydrolyse kein Dinitrophenol, so dass die Menge an endständigem Prolin ziemlich genau sich aus der Bestimmung von Dinitrophenol + DNP-Prolin ergibt.

UNTERSUCHUNGEN ÜBER DEN EDMANSCHEN ABBAU

Auch bei der Bestimmung der Endgruppen nach Edman an dem mit Trichloressigsäure behandelten Protein ergeben sich gewisse Komplikationen. Das Thiohydantoin des Prolins ist unter den Bedingungen der Cyclisierung instabil und liefert in kleinen Mengen ein Beiprodukt mit einem etwas geringeren Rf-Wert. Die quantitative Auswertung der beiden Flecken des Chromatogramms ergab F2400 endständige Prolin-Moleküle je Molekül TMV. Daneben fanden wir in einer Menge von etwa 20% noch andere Thiohydantoine, diese entsprechen teilweise Endgruppen, die bei der Trichloressigsäure-Behandlung im Innern der Peptidketten freigesetzt wurden. Ausserdem reagiert aber das Phenylthioisocyanat (PTI) mit gewissen nicht-endständigen Aminosäuren unter Bildung von Thiohydantoinen, wie wir an reinen synthetischen Peptiden feststellten. Bei einstündiger Reaktionsdauer mit PTI bei 40° entstanden aus Leu-Gly-Gly und Ala-Gly-Gly 3-7%

Thiohydantoine des nicht-endständigen Glycins.¹⁹ Da eine Hydrolyse der Peptidkette unter diesen Bedingungen nicht erfolgt, handelt es sich um einen direkten Angriff des PTI auf mittelständige Iminogruppen. Auch bei anderen Eiweisstoffen fanden wir stets neben den Banden der Endgruppen in einer Menge von 2-12% Nebenbanden, die auf diese Reaktion zurück gehen. Auch Rohrlich und Schlüssler²⁰ fanden, dass Gliadin aus Weizen in kleiner Menge Thiohydantoine des Glycins und Alanins liefert, selbst wenn die Endgruppen vorher mit dem DNP-Rest substituiert wurden. In der folgenden Tabelle 1 sind die Ergebnisse der Endgruppenbestimmung am TMV

Tabelle 1

ENDGRUPPENBESTIMMUNG NACH EDMAN BEI
VERSCHIEDENEN PROTEINEN

Protein	Endgruppen/Mol		Nebenbanden in % der Gesamtextinktion
	Pro	2300	
TMV*	Pro	2300	20
Hämoglobin	Val+Meth	3,9	12
Rind	Val	3,8	8
Pferd	Val	3,8	2
Mensch	Lys	0,9	10

und einigen anderen Eiweisstoffen zusammengefasst. Es ist darauf hinzuweisen, dass die Ergebnisse beim Rinder-Globulin und beim Lysozym sehr gut mit den in der Literatur angegebenen Werten übereinstimmen. Abweichend von Porter und Sanger²¹ fanden wir jedoch beim Hämoglobin des Pferdes und des Menschen die gleiche Anzahl von Endgruppen wie beim Rind, was inzwischen von anderer Seite²² bestätigt wurde.

Durch Verwendung substituierter Phenylthioisocyanate konnte die Nebenreaktion teilweise unterdrückt werden, Beim Abbau von Ala-Gly-Gly und Leu-Gly-Gly mit ortho-Nitrophenylthioisocyanat entsteht in der ersten Stufe kein Glycin-Thiohydantoin, während die Reaktion der terminalen Gruppe schneller und vollständiger verläuft als mit Phenylthioisocyanat. Jedoch ist auch die Reaktionsgeschwindigkeit mit Wasser erhöht, so dass sich grosse Mengen an Harnstoff-Derivaten bilden, die die Chromatographie stören²³ (vergleiche Tabelle 2). Wir halten daher das Phenylthioisocyanat für das geeignetste Reagenz für den Edmanschen Abbau. Im allgemeinen ist die Menge der Nebenprodukte so klein, dass die Identifizierung der Endgruppen nicht erschwert wird und die grosse Bedeutung des Edmanschen Abbaus,

* Mit Trichloressigsäure behandelt.

266 SCHRAMM, BRAUNITZER, ANDERER, SCHNEIDER, UHLIG [15
den wir ständig für unsere Untersuchungen benützen, hierdurch nicht be-
einträchtigt wird.

Tabelle 2

REAKTIONSFÄHIGKEIT DER PHENYLTHIOISOCYANATE

Die verschiedenen substituierten Thioisocyanate worden mit Leucin umgesetzt und die Menge an Phenylthiohydantoin des Leucins (PTH-Leu) und an Nebenprodukten chromatographisch bestimmt.

Thioisocyanate	Gesamtausbeute in Mol% bez. auf Leucin	PTH-Leu-Deriv. in Mol%	Nebenbanden in Mol% bez. auf PTH-Leu
Phenyl-	90	82	8
o-Nitrophenyl	100	80	20
m-Nitrophenyl	120	80	40
p-Nitrophenyl	120	70	50
o-Äthoxyphenyl	45	33	12
p-Äthoxyphenyl	83	62	21
p-Methoxyphenyl	78	53	25
o-Chlorphenyl	70	58	12
2,5-Dichlorphenyl	67	54	13
p-Azobenzolphenyl	20	14	6

VERSUCHE ZUR BESTIMMUNG DER N-TERMINALEN KONFIGURATION IM NATIVEN TMV-PROTEIN

Braunitzer²⁴ versuchte aufzuklären, warum im nativen TMV die Prolin-Endgruppe nicht erfassbar ist. Es ist bekannt, dass Amide des γ -Carboxyls der Glutaminsäure und des β -Carboxyls der Asparaginsäure durch Säuren sehr viel leichter gespalten werden als α -Peptidbindungen. Es war daher wahrscheinlich, dass die Blockierung auf einer derartigen β bzw. γ -Peptidbindung der Seitenkette beruht und diese durch Trichloressigsäure gespalten werden. Derartige Peptidbindungen werden auch durch Hydroxylamin leicht unter Bildung der entsprechenden Hydroxamsäuren gelöst.²⁵ In der Tat wurden bei Behandlung von TMV mit 3n NH₂OH—Lösung bei pH 7 (24 Stunden bei 60°C) 50-80% der theoretischen Menge an endständigem Prolin in Freiheit gesetzt, ohne dass eine Spaltung von α -Peptidbindungen in grösserem Ausmass bemerkbar war. Einige weitere Versuche, die jedoch noch nicht abgeschlossen sind, sprechen dafür, dass die Blockierung durch die β -Carboxylgruppe des Asparaginsäurerestes zustande kommt.

Die Ergebnisse der nach vier verschiedenen Methoden durchgeführten Endgruppenbestimmung sind in Tabelle 3 nochmals zusammengefasst. Die Existenz definierter Peptid-Untereinheiten, die anfangs von der Arbeitsgruppe in Berkeley stark angezweifelt wurde, dürfte damit gesichert sein. Als wahrscheinlichster Wert für das Molgewicht dieser Untereinheiten ist

Tabelle 3

ZAHL DER ENDGRUPPEN BEIM TABAKMOSAIKVIRUS

Material	Methode	Endgruppen Mol/40·10 ⁶ g TMV			Autor (Ref.)
		C-terminal	N-terminal		
natives TMV	Carboxypeptidase Hydrazinolyse	Thr		2300	8,9 10-12
		Thr		2700	
TMV beh. mit Trichloressigsäure	Sanger Edman		Pro	2300	14 19
			Pro	2300	

16 500 anzusehen. Dieser steht in bester Übereinstimmung mit den von Knight²⁶ durchgeführten Aminosäure-Analysen. Wenn man annimmt, dass jede Kette ein Molekül Cystein enthält, ergibt sich ein Molgewicht von 16 300 und eine Zahl der Aminosäuren von circa 150. Dieses Molekulargewicht stimmt auch gut mit der Grösse der Untereinheiten überein, die R. Franklin aus Röntgenuntersuchungen erhielt.²⁷ Nach elektronenmikroskopischen Messungen²⁸ beträgt die Länge des TMV-Stäbchens 2980 ± 10 Å. Auf eine Länge von 69 Å kommen nach Franklin 49 Untereinheiten, auf das ganze Stäbchen also 2 100. Bei der Unsicherheit, mit der Längenmessung und Molgewichtsbestimmungen beim TMV behaftet sind, ist diese Übereinstimmung als befriedigend anzusehen.

BESTIMMUNG DER AMINOSÄURE-SEQUENZ IN
DER PEPTIDUNTEREINHEIT

Die Bestimmung der Sequenz der Aminosäuren innerhalb der Peptiduntereinheit steht in unserem Laboratorium erst am Anfang, ich möchte mich daher darauf beschränken, die allgemeinen Methoden zu beschreiben, die wir hierbei verwenden. Im Nucleinsäure-freien Protein des Virus wurde zunächst durch Behandlung mit Trichloressigsäure das N-terminale Prolin in Freiheit gesetzt und anschliessend das Protein mit Dinitrofluorbenzol umgesetzt. Hierbei entstehen in der Peptidkette drei chromophore Gruppen, eine des endständigen DNP-Prolins und die beiden mittelständigen ϵ -DNP-Lysylreste.



Durch anschliessende enzymatische Spaltung des substituierten Proteins wurde zunächst versucht, die Reihenfolge der Aminosäuren in der Nachbarschaft der drei chromophoren Gruppen festzulegen, um hierdurch 3 Fixpunkte zu schaffen, die die gegenseitige Zuordnung der übrigen Spaltpeptide erleichtern.²³ Die enzymatische Spaltung des substituierten Proteins

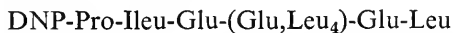
268 SCHRAMM, BRAUNITZER, ANDERER, SCHNEIDER, UHLIG [15 erfolgte mit Chymotrypsin, Subtilisin und einem Proteinasegemisch aus *Aspergillus oryzae*. Das letztere bewährte sich vor allem zur Herstellung kürzerer Spaltstücke. Pepsin eignet sich weniger, da das DNP-Protein im sauren Bereich zu wenig löslich ist.

Erwartungsgemäss erhielten wir drei Gruppen gelber Peptide, die entweder DNP-Prolin oder ϵ -DNP-Lysin enthielten. Die Trennung erwies sich zum Teil als ausserordentlich schwierig. Wie Sanger²⁹ zeigte, treten bei der enzymatischen Spaltung von Polypeptiden häufig auch Pyrrolidincarboxylpeptide auf, die in Essigester leicht löslich sind und dadurch die Trennung der DNP-Peptide erschweren. Die farblosen Pyrrolidincarboxylpeptide lassen sich jedoch leicht durch die Fluoreszenz-Auslöschung im UV-Licht in den Chromatogrammen erkennen. Als Trennungsmethoden benützten wir: 1.) Verteilungschromatographie an Kieselgursäulen, wobei wir als mobile Phase hauptsächlich Essigester-Butanol-Gemische und als stationäre Pufferlösungen mit verschiedenem pH verwandten. 2.) Hochspannungselektrophorese. Wir arbeiteten meist bei Spannungen von 5- 10 000 V in der von Michl³⁰ angegebenen Anordnung in einem Toluolbad, das durch Kühlung auf etwa 10°C gehalten wurde. 3.) Papierchromatographie, wobei wir besonders Pyridinhaltige Lösungen und reine Puffer, z.B. 2*m* Natriumacetat benützten.

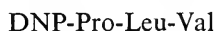
Bei der Konstitutionsermittlung der erhaltenen DNP-Peptide bewährte sich neben den bekannten Methoden auch das von Justisz und Ritschard³¹ beschriebene Verfahren, bei dem zunächst der DNP-Rest durch Hydrierung zu einem Dihydrochinoxalon abgespalten wird.

Das Restpeptid kann entweder durch Wiederholung der Operation oder nach Edman weiter abgebaut werden. Ein ähnliches Verfahren wurde unabhängig von Uhlig³² ausgearbeitet. Die Cyclisierung tritt bereits während der Hydrierung in Eisessig oder in 1*n* HCl-Methanol mit Pt oder Pd ein. Bei der Umsetzung des Reaktionsgemisches mit Dinitrofluorbenzol liefern die Chinoxalone Bis-DNP-Derivate, die näher charakterisiert wurden. Sie sind in Chloroform oder Eisessig leicht löslich und können chromatographisch identifiziert werden.

Zu Beginn unserer Untersuchungen wurde bei der Spaltung mit Chymotrypsin haltigem Trypsin ein kristallisiertes DNP-Prolylpeptid isoliert, das bei der Analyse sehr konstante Werte lieferte und von dem wir annahmen, dass es aus 21 Aminosäuren besteht. Mit den verbesserten Trennmethoden erwies sich jedoch dieses Kristallisat als nicht einheitlich. Es gelang schliesslich, eine Reihe von DNP-Prolylpeptiden in reiner Form zu isolieren, von denen sich die meisten der Sequenz

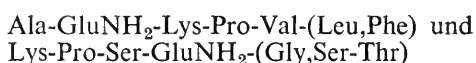


zuordnen liessen. Daneben fanden wir aber auch ein Peptid, dem die Sequenz



zukommt und das mit der eben angeführten Reihenfolge nicht zu vereinbaren ist. Beide Sequenzen wurden auch nebeneinander bei der partiellen

Säurehydrolyse aufgefunden. Das Auftreten dieser unterschiedlichen Sequenzen kann einmal darauf beruhen, dass das benützte TMV-Material genetisch nicht einheitlich war. Es wäre aber auch möglich, dass in einer und derselben TMV-Partikel Polypeptidketten mit verschiedenen Sequenzen vorkommen. Eine dritte Möglichkeit wäre, dass eine Prolyl-Sequenz aus der Mitte der Kette stammt und durch die Behandlung mit Trichloressigsäure freigelegt wurde. Da eine quantitative Bestimmung der Ausbeuten an den beiden Peptidsequenzen wegen des komplizierten Trennungsganges nicht möglich war, kann die Entscheidung zwischen diesen Möglichkeiten zur Zeit noch nicht getroffen werden. Die Peptide, die einen DNP-Lysylrest enthalten, lassen sich zu zwei verschiedenen Sequenzen zusammenstellen:



Es gelten jedoch hier die gleichen Einschränkungen wie für das Prolylpeptid. Es besteht ebenfalls die Möglichkeit, dass die beiden Lysyl-Sequenzen nicht aus derselben Polypeptidkette stammen. Zur Zeit sind Untersuchungen im Gang, durch Spaltung am nicht substituierten TMV-Protein Polypeptide zu isolieren, um zu sehen, ob sich eine lückenlose Verbindung zwischen den beiden Lysylpeptiden herstellen lässt.

Bei diesen Untersuchungen hoffen wir auch weitere Aufschlüsse über die N-terminale Konfiguration am nativen Virus zu erhalten.

VERZEICHNIS DER LITERATUR

1. A. GIERER u. G. SCHRAMM: *Z. f. Naturforsch.*, **11b**, 138 (1956); *Nature*, **177**, 702 (1956).
2. G. SCHRAMM u. B. V. KEREKJARTO, unveröffentlicht.
3. G. SCHRAMM, G. SCHUMACHER u. W. ZILLIG: *Z. f. Naturforsch.*, **10b**, 481 (55).
G. SCHRAMM, W. ZILLIG: *Z. f. Naturforsch.*, **10b**, 493 (1955).
4. G. SCHRAMM, G. SCHUMACHER, W. ZILLIG: *Nature*, **175**, 549 (1955).
5. W. F. HARRINGTON, H. K. SCHACHMAN: *Arch. Biochem. Biophys.*, **65**, 278 (1956).
6. G. SCHRAMM u. G. BERGOLD: *Z. f. Naturforsch.*, **2b**, 108 (1947).
7. H. K. SCHACHMAN u. HERSH, zitiert nach *J. Am. Chem. Soc.*, **76**, 180 (1953).
8. J. I. HARRIS, C. A. KNIGHT, *Nature*, **170**, 613 (1952); *J. Bio. Chem.*, **214**, 215 (1955).
9. G. SCHRAMM, G. BRAUNITZER, J. W. SCHNEIDER: *Z. f. Naturforsch.*, **9b**, 298 (1954).
10. G. BRAUNITZER: *Z. f. Naturforsch.*, **9b**, 676 (1954).
11. G. BRAUNITZER: *Naturwissenschaften*, **42**, 371 (1955).
12. G. BRAUNITZER: *Chem. Berichte*, **88**, 2025 (1955).
13. C. I. NIU, H. FRAENKEL-CONRAT: *Biochem. Biophys. Acta*, **16**, 597 (1955).
14. G. SCHRAMM u. G. BRAUNITZER: *Z. f. Naturforsch.*, **8b**, 61 (1953).
15. G. SCHRAMM, G. BRAUNITZER, J. W. SCHNEIDER: *Nature*, **176**, 456 (1955).
16. F. SANGER: *Biochem. J.*, **39**, 507 (1945).
17. P. EDMAN: *Acta Chem. Scand.*, **4**, 277 (1952).
18. C. S. HANES, F. J. R. HIRD, F. A. ISHERWOOD: *Biochem. J.*, **51**, 25 (1952).

- 270 SCHRAMM, BRAUNITZER, ANDERER, SCHNEIDER, UHLIG [15
19. G. SCHRAMM, J. W. SCHNEIDER, F. A. ANDERER: *Z. f. Naturforsch.*, **11b**, 12 (1956).
 20. M. ROHRLICH, H. J. SCHLÜSSLER: *Naturwissenschaften* **44**, 37 (1957).
 21. R. R. PORTER, F. SANGER: *Biochem. J.*, **42**, 287 (1948).
 22. H. OZAWA, K. SATAKE: *J. of Biochemistry* (Japan), **42**, 641 (1955).
 23. F. A. ANDERER: *Dissertation Tübingen* (1957).
 24. G. BRAUNITZER: *Biochim. Biophys. Acta*, **19**, 574 (1956).
 25. W. J. WILLIAMS, C. B. THORNE: *J. Biol. Chem.*, **210**, 203 (1954), u. **211**, 631 (1954).
 26. C. A. KNIGHT: *J. Biol. Chem.*, **171**, 297 (1947).
F. L. BLACK, C. A. KNIGHT: *J. Biol. Chem.*, **202**, 51 (1953).
 27. A. KLUG, R. E. FRANKLIN: *Biochim. Biophys. Acta*, **23**, 199 (1957).
 28. R. C. WILLIAMS, R. L. STEERE: *J. Am. Chem. Soc.*, **73**, 2057 (1951).
 29. F. SANGER, E. O. THOMPSON, R. KITAI: *Biochem. J.*, **59**, 509 (1955).
 30. H. MICHL: *Mts. hefte der Chemie*, **82**, 489 (1951).
 31. M. JUTISZ, W. RITSCHARD: *Biochim. Biophys. Acta*, **17**, 548 (1955).
 32. H. UHLIG, *Diplomarbeit Tübingen* (1956).
 33. F. S. SEANES, B. T. TOZER: *Biochem. J.*, **63**, 282 (1956).
 34. P. NEWMARK, R. W. MYERS: *Fed. Proc.* **16**, 226 (1957), inzwischen veröffentlicht.

X-ray diffraction studies of the structure of the protein of tobacco mosaic virus

ROSALIND E. FRANKLIN

Crystallographic Laboratory, Birkbeck College,
University of London

A systematic application of the methods of X-ray diffraction to the study of the structure of tobacco mosaic virus (TMV) has provided some information concerning the general structural characteristics of the protein of the virus.

Well-orientated specimens of concentrated TMV solutions can be obtained in fine glass capillaries, and give excellent X-ray fibre-diagrams (Bernal and Fankuchen, 1941; Franklin, 1955*a*). Fibre-diagrams which are similar but significantly different may be obtained from heavy-atom derivatives of TMV (Caspar, 1956; Franklin and Holmes, 1956), from the repolymerized nucleic acid-free virus protein (Franklin, 1955*b*) and from different strains of the virus (Franklin, 1956*a*). Structure analysis is based on a detailed quantitative comparison of the X-ray diagrams obtained from all these substances. The analysis so far carried out is based on only a small part of the X-ray diagram, and the work is therefore still at a relatively early stage.

It has been shown that the virus protein is in the form of structurally equivalent sub-units set in helical array about the particle axis (Watson, 1954; Franklin, 1955*a*) and (Franklin, Klug and Holmes, 1956) that these structural sub-units may be identified with the chemical sub-units established by amino acid and end group analysis (e.g. Harris and Knight, 1952, 1955; Schramm and Braunitzer, 1953; Niu and Fraenkel-Conrat, 1955).

The protein sub-units lie on a helix of pitch 23 Å, and there are 49 sub-units to every 3 turns of the helix (Franklin and Holmes, 1956). A large part of the results so far obtained is concerned with the *shape* of these sub-units. If the TMV particle were a simple uniform-density rod, of circular cross-section, the simplest possible shape of the protein sub-units would be one resembling a slice of cake. We find, however, that the density of the TMV particle is far from uniform, and that the shape of the protein sub-unit is not simple. The rod-shaped particle has a hollow core, of radius about 20 Å (Caspar, 1956; Franklin, Klug and Holmes, 1956) and the shape of each sub-unit is such that it protrudes at both the inner and outer surfaces. The helical array of protuberances on the outer surface has already been described (Franklin and Klug, 1956). That on the inner surface has been

made apparent by recent work in which a *helical projection* of the virus structure was calculated from X-ray data. This projection (which will be described in detail elsewhere) shows that the internal and external protuberances do not occur opposite to one another, thus suggesting that the general lie of the protein sub-units is not strictly perpendicular to the particle axis, but somewhat skew to it.

Independent evidence from three different parts of the X-ray diagram indicates that the shape of the protein sub-units is such that their helical packing results in a set of *holes* in the virus particle at a radial distance of 55-60 Å from the particle axis. The mean radius of the particle is about 75 Å, and the maximum radius about 90 Å. The nucleic acid lies at a radius of 40 Å (Franklin, 1956*b*). The X-ray results therefore suggest that the nucleic acid might well be more accessible to reagents via the external groove and the 55-60 Å holes than via the hollow central core of the particle.

The X-ray results, considered in conjunction with the ratio of nucleic acid to protein, make it appear probable (though not yet certain) that the nucleic acid is in the form of a single molecular chain following a rather flat helical path, and fitting into the protein structure in a rather compact manner. The shape of the protein sub-units must therefore be such as to allow this RNA molecular chain to pass between adjacent sub-unit layers.

Thus, although it is not yet possible to define the shape of the protein sub-units, we know that it must be such as to favour their helical packing and yet to leave holes between neighbouring sub-units at a distance of 55-60 Å from the particle axis, to leave a hollow core of radius about 20 Å in the particle, to provide some kind of gap through which the RNA molecule may pass at a radius of 40 Å and to give a protuberance at both the inner and outer surfaces of the particle.

Our X-ray diffraction results also indicate that the direction of the polypeptide chains within the sub-units tends to be perpendicular to the particle axis, and in a tangential rather than a radial direction.

It may be of interest to compare the information obtainable from X-ray fibre-diagrams of TMV and from X-ray diagrams of single crystals of globular proteins. Because of the strongly uniaxial character of the TMV particle and also, in part, because of its great stability, a limited amount of information concerning the inner structure of the particle can be obtained with relatively little effort. By a much more laborious quantitative analysis of X-ray fibre-diagrams of TMV and its heavy-atom derivatives, analogous to that being carried out by Perutz, Kendrew and co-workers on globular proteins, it should be possible to determine in some detail not only the shape of the particle and of its sub-units, but also the direction of the back-bone chains of both the protein and nucleic acid components. Some progress has already been made in this direction. On the other hand, the three-dimensional diffraction picture of TMV, which must be derived from fibre-diagrams only, will never be as accurate as that obtained from single crystals

of protein, and the resulting structural information will therefore never be as detailed. We cannot hope, in TMV, to find the positions of *atoms* by X-ray methods. Thus, while some general structural characteristics of the protein of TMV can be determined much more easily than in the simpler structures of crystals of globular proteins, the detailed configuration of the polypeptide chains of globular proteins must be studied in single crystals. It is hoped, however, to find out enough about the structure of the protein (and also of the nucleic acid) of TMV to give some indication of the nature of the nucleoprotein complex in the intact virus.

REFERENCES

- BERNAL, J. D. and FANKUCHEN, I. (1941), *J. Gen. Physiol.*, **25**, 111, 147.
CASPER, D. L. D. (1956), *Nature*, London, **177**, 928.
FRANKLIN, R. E. (1955a), *Nature*, London, **175**, 379.
FRANKLIN, R. E. (1955b), *Biochim. Biophys. Acta*, **18**, 313.
FRANKLIN, R. E. (1956a), *Biochim. Biophys. Acta*, **19**, 203.
FRANKLIN, R. E. (1956b), *Nature*, London, **177**, 929.
FRANKLIN, R. E., KLUG, A. and HOLMES, K. C. (1956), *Ciba Foundation Symposium of the Nature of Viruses*, pp. 39-52.
FRANKLIN, R. E. and HOLMES, K. C. (1956), **21**, 405.
FRANKLIN, R. E. and KLUG, A. (1956), *Biochim. Biophys. Acta*, **19**, 403.
HARRIS, J. I. and KNIGHT, C. A. (1952), *Nature*, London, **170**, 613.
HARRIS, J. I. and KNIGHT, C. A. (1955), *J. Biol. Chem.*, **214**, 215.
NIU, C. I. and FRAENKEL-CONRAT, H. (1955), *J. Amer. Chem. Soc.*, **77**, 5882.
SCHRAMM, G. and BRAUNITZER, G. (1953), *Z. Naturf.*, **8b**, 61.
WATSON, J. D. (1954), *Biochim. Biophys. Acta*, **13**, 10.

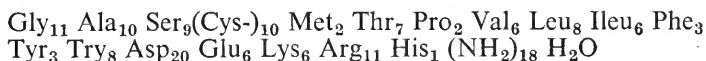
Other Proteins and Peptides

Nouvelles données concernant la structure du lysozyme d'œuf de poule

P. JOLLÈS, J. JOLLÈS-THAUREAUX ET
C. FROMAGEOT

Laboratoire de Chimie Biologique de la Faculté des Sciences
Paris

Avant de parler des résultats nouvellement acquis concernant la structure chimique du lysozyme, résultats qui font l'objet essentiel de cet exposé, il paraît utile de rappeler brièvement quelques-unes des données précédemment établies, ou proposées, se rapportant à la structure de la protéine en question. D'après toute une série de déterminations^{1,2} effectuées soit par voie chimique et notamment par chromatographie, soit par voie microbiologique, le lysozyme d'œuf de poule présente une composition en acides aminés correspondant approximativement à la formule brute suivante:



le poids moléculaire étant d'environ 14.800.

La recherche des acides aminés N-terminaux, recherche effectuée soit par la méthode de Sanger, soit par la méthode d'Edman, a montré de façon indiscutable que le lysozyme ne possédait qu'un seul acide aminé N-terminal, la lysine.^{3,4} D'autre part, la recherche des acides aminés C-terminaux, recherche effectuée soit par l'utilisation de la carboxypeptidase,^{5,6} soit par la méthode de réduction par AlLiH_4 ,⁷ a montré de façon non moins indiscutable que le lysozyme ne possédait qu'un seul acide aminé C-terminal, la leucine. Le lysozyme d'œuf de poule est donc, selon toute vraisemblance, constitué par une seule chaîne peptidique non ramifiée. En outre, il apparaît que cette chaîne ne contient aucun groupe $-\text{SH}$.¹ Les cinq résidus de cystine correspondent donc à cinq ponts $-\text{S}-\text{S}-$ qui doivent former dans la molécule autant de boucles. On sait également que l'ouverture des ponts $-\text{S}-\text{S}-$ et leur remplacement par des groupes $-\text{S}.\text{CH}_2.\text{COOH}$ ne provoque aucune scission de la molécule,⁸ ce qui confirme le caractère uncaténaire de cette dernière.

La formule brute qui vient d'être donnée présente une utilité évidente

comme guide pour les recherches concernant la structure de la molécule; mais elle ne peut être considérée comme rigoureusement exacte: la formule définitive ne pourra être établie, sans aucune ambiguïté, qu'après que la structure chimique complète, ou tout au moins l'ordre complet d'enchaînement de tous les acides aminés constitutifs, aura été déterminé.

Cette formule montre néanmoins déjà quelques particularités intéressantes ici: la molécule de lysozyme renferme seulement un résidu d'histidine, deux de proline, deux de méthionine, trois de phénylalanine et trois de tyrosine. Ces résidus d'acides aminés, assez caractéristiques, peuvent donc être utilisés avec avantage comme marqueurs des fragments de la protéine qui les contiennent; en outre, la lysine N-terminale et la leucine C-terminale peuvent également être considérées comme marqueurs. L'isolement des peptides correspondant à ces fragments et leur caractérisation apparaissent donc comme un stade en quelque sorte préliminaire, indispensable à l'étude de la structure chimique du lysozyme.

Cette formule montre, d'autre part, que le lysozyme contient un nombre relativement élevé de résidus de tryptophane. C'est là un caractère qui, du point de vue qui nous occupe, différencie cette protéine des autres protéines telles que l'insuline, la ribonucléase ou autres, dont la structure chimique a déjà été établie ou est en passe de l'être et qui, soit ne renferment pas cet acide aminé, soit n'en contiennent qu'un résidu. La présence de tryptophane est en effet une cause de difficultés supplémentaires, qui seront considérées plus loin.

Au cours d'une première série d'investigations, faites à notre laboratoire, un certain nombre de peptides marqués par des résidus d'acides aminés peu fréquents ont pu être isolés. Ces peptides ont été obtenus en soumettant le lysozyme à divers procédés d'hydrolyse: hydrolyse partielle acide, hydrolyse chymotrypsique, hydrolyse pepsique; leur isolement a été fait essentiellement à l'aide de la chromatographie sur papier et de l'ionophorèse sur papier. Ils ont néanmoins permis d'établir les enchaînements présentés dans le tableau 1.

On voit que ces enchaînements permettent déjà de caractériser l'entourage de l'unique résidu d'histidine, et les entourages respectifs de chacun des trois résidus de phénylalanine et des trois résidus de tyrosine, mais seulement l'entourage complet d'un seul résidu de proline. Parmi les enchaînements comportant un résidu de phénylalanine, l'enchaînement Lys.Val.Phe.Gly.Arg, complétant l'enchaînement Lys.Val.Phe.Gly obtenu antérieurement par Schroeder,³ est particulièrement intéressant, puisque c'est l'enchaînement de l'extrémité N-terminale de la protéine. En outre, quelques peptides contenant du tryptophane ont été caractérisés (tableau 2).

D'autre part, A. R. Thompson¹³ a obtenu par hydrolyse acide ménagée un assez grand nombre de peptides; ceux-ci ont été séparés par chromatographie sur Dowex 50 × 4, puis purifiés par chromatographie sur papier. L'auteur australien, par des tentatives de recombinaison, en a conclu à

Tableau 1

PEPTIDES 'MARQUÉS' PAR HIS, TYR,
PHE, PRO OU MET, ISOLÉS SUR
PAPIER ET RECONSTITUÉS APRÈS
HYDROLYSE ACIDE, TRYPSIQUE,
PEPSIQUE OU CHYMOTRYPSIQUE

- | |
|--|
| 1. Arg.His.Lys ⁹ |
| 2. Asp.Tyr.Arg.Gly ¹⁰ |
| 3. Arg.Gly.Tyr.Ileu.Leu ¹⁰ |
| 4. Asp(NH ₂).Ala.Tyr.Gly.Ser.Leu.Asp(NH ₂) ¹⁰ |
| 5. Lys.Val.Phe.Gly.Arg ⁴ |
| 6. Glu.Ser.Phe.Asp ¹⁰ |
| 7. Ala.Lys.Phe.Glu ¹⁰ |
| 8. Leu.Pro ¹⁰ |
| 9. Thr.Pro.Gly.Ser.Arg ¹¹ |
| 10. Ala.Ala.Met ¹⁰ |

Tableau 2

PEPTIDES CONTENANT DU TRYPTOPHANE,
ISOLÉS PAR CHROMATOGRAPHIE SUR PAPIER
APRÈS HYDROLYSE CHYMOTRYPSIQUE¹²

- | | |
|----------------|--------------------------|
| 1. Asp.Ala.Try | 3. (Gly,Arg).Try |
| 2. Val.Ala.Try | 4. (Gly,Ala,Try,Arg).Leu |

Tableau 3

ENCHAÎNEMENTS PROPOSÉS PAR A. R. THOMPSON¹³

- | | |
|----------------------------|-----------------------------|
| 1. Thr.Asp.Val.Glu.Ala | 14. Thr.Asp.Arg.Arg |
| 2. Ileu.Glu.Leu.Ala.Leu | 15. Thr.Gly.Asp.Val |
| 3. Asp.Glu.Ala | 16. Ser.Val.Cys.Ala.Lys.Gly |
| 4. Leu.Thr.Ala. | 17. Gly.Cys.Asp |
| 5. Glu.Asp.Ileu | 18. Leu.Gly.Ala.Val |
| 6. Thr.Glu.Ala.Gly | 19. Asp.Ileu.Pro.Cys |
| 7. Ser.Asp.Gly.Met.Asp | 20. Arg.Cys.Lys.Gly |
| 8. Asp.Ala.Met.Lys.Cys.Arg | 21. Ser.Val.Asp.Cys.Ala |
| 9. Val.Thr.Pro.Gly.Ala | 22. Asp.Leu.Cys.Asp |
| 10. Ser.Asp.Arg | 23. Arg.Asp.Cys.Ileu |
| 11. Lys.Phe.Glu.Gly | 24. Ser.Arg.Leu |
| 12. Arg.Cys.Glu.Ala | 25. Ser.Asp.Cys.Arg.Leu |
| 13. Ser.Phe.Asp.Glu | |

l'existence, dans le lysozyme, des enchaînements suivants (tableau 3). Malheureusement, certains de ces enchaînements présentent encore un caractère quelque peu hypothétique.

D'autre part, enfin, Ohno,¹⁴ après avoir soumis le lysozyme à une hydrazinolyse ménagée, a été conduit à proposer comme enchaînement C-terminal de la protéine: Asp.Gly.Ala.Asp.(NH₂).Leu. On verra cependant plus loin que cet enchaînement ne paraît pas correspondre à la réalité.

Les diverses techniques d'obtention ou de séparation des peptides à partir du lysozyme, telles qu'elles ont été utilisées jusqu'ici, ne permettaient pas, à elles seules, d'aller beaucoup plus loin dans ce genre de travail. D'autre part, la difficulté d'établir des enchaînements avec certitude nécessitait l'utilisation de nouvelles investigations dont les résultats devaient permettre de confirmer, ou éventuellement d'infirmer les conclusions tirées jusqu'alors. Aussi, avons-nous poursuivi nos investigations d'une part en utilisant des procédés de scission de la molécule protéique aussi spécifique que possible, et d'autre part, en mettant en oeuvre, pour la séparation et l'analyse des peptides résultant de cette scission, l'élégante méthode de Hirs, Moore et Stein.¹⁵ Cette méthode permet, en principe, de recueillir et de séparer la totalité des peptides provenant de la scission d'une protéine.

HYDROLYSE DU LYSOZYME PAR LA TRYPSINE

Pour avoir un nombre de peptides raisonnable, nous nous sommes adressés tout d'abord à l'hydrolyse enzymatique par la trypsine. La spécificité de cet enzyme est maintenant bien établie; le fait qu'il ne coupe que les liaisons peptidiques dans lesquelles sont impliquées, par leur groupe carboxylique, l'arginine et la lysine, faisait prévoir que le nombre maximum de peptides différents que l'on pouvait obtenir à partir du lysozyme était de 17 ou 18. Mais la présence, dans la protéine, d'une série de boucles dues à l'existence de ponts —S—S— fait que certains des peptides détachés par l'action de la trypsine doivent avoir une structure compliquée et risquent de ne pas pouvoir être séparés convenablement par chromatographie. Aussi l'action de la trypsine a-t-elle été appliquée ici, non seulement sur le lysozyme n'ayant pas subi de traitement autre qu'une dénaturation par la chaleur, mais aussi sur du lysozyme dont les ponts —S—S— ont été rompus par oxydation et transformés ainsi en un nombre double de groupes sulfoniques.

Action de la trypsine sur le lysozyme non préalablement oxydé. Plusieurs observations^{16,17} ont montré que, comme d'autres protéines, le lysozyme 'natif', c'est-à-dire obtenu dans des conditions telles qu'il ait conservé toute son activité enzymatique, est à peu près complètement résistant à l'action de la trypsine, tout au moins dans les conditions habituelles. Pour que la protéine puisse être attaquée par la trypsine, il convient donc de la dénaturer, ce que l'on obtient facilement ici par chauffage d'une heure à 100° en solution aqueuse. Cette manière de faire a l'avantage de conserver intacts les

résidus de tryptophane. On fait agir la trypsine sur le lysozyme ainsi dénaturé, pendant 6 heures, à 37°, à pH 7,8, le rapport enzyme/substrat étant de 1/100. Une première séparation des peptides produits par la trypsine est effectuée sur colonne de Dowex 50 \times 2. Les détails de cette séparation et les résultats qu'elle a permis d'obtenir sont indiqués dans la fig. 1.¹⁸

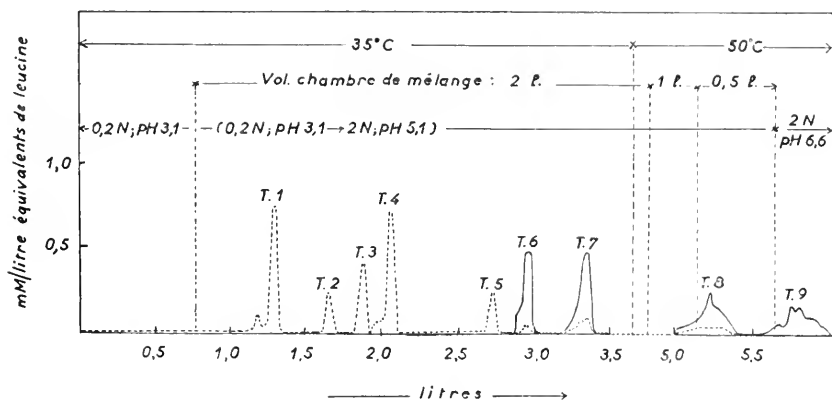


Fig. 1. Séparation des produits obtenus par action de la trypsine sur 25 μ M de lysozyme dénaturé non oxydé.

— coloration de la réaction à la ninhydrine après hydrolyse alcaline.

----- coloration de la réaction à la ninhydrine effectuée directement.

On constate la présence de 9 seulement des 17 ou 18 pics que l'on pouvait espérer obtenir. La différence correspond, comme on le verra plus loin, aux peptides qui contiennent de la cystine et qui sont énergiquement retenus sur la colonne. Après cette première séparation, les substances ont été soumises à une série de contrôles concernant leur pureté et ont été, lorsque cela s'est révélé nécessaire, purifiés par diverses méthodes: nouvelles chromatographies sur colonne, et, après dessalification, chromatographies sur papier et électrophorèses sur papier.¹⁹ Il a été ainsi possible d'étudier de façon précise chacune des substances, ammoniac, acides aminés et peptides constituant les pics représentés par la fig. 1. La nature des acides aminés et la composition des peptides est donnée dans le tableau 4.

En plus des acides aminés libres (leucine, lysine, arginine) et de l'ammoniac, quatre peptides seulement, T.6 à T.9, ont été obtenus avec un rendement supérieur à 10 %. Leur structure a été établie²⁰ en mettant en oeuvre les techniques habituelles: * détermination des acides aminés N-terminaux par la méthode de Sanger et la méthode d'Edman, cette dernière étant éventuellement utilisée également pour l'établissement des enchaînements N-terminaux, hydrolyses partielles acides ou hydrolyses chymotrypsiques fournissant des peptides plus courts étudiés à leur tour d'une part à l'aide des méthodes

* Le détail des opérations ayant permis d'établir les structures indiquées ci-dessous est donné dans 19.

Tableau 4

COMPOSITION DES PICS SÉPARÉS APRÈS HYDROLYSE TRYPSIQUE DU LYSOZYME NON PRÉALABLEMENT OXYDÉ¹⁸

Les rendements de chaque peptide sont indiqués entre parenthèses

T.1	Leucine libre (50 %) (correspond à leucine terminale).
T.2	Peptide à très faible rendement, contenant Try, etc.
T.3	NH ₃
T.4	Lysine libre (92 %) (correspond à lysine initiale).
T.5	Arginine libre (40 %).
T.6	(Asp(NH ₂), Arg) (45 %).
T.7	(Gly, Ser, Pro, Thr, Arg) (36 %).
T.8	(Gly, Ala, Ser, Leu, Tyr ₂ , Asp ₃ , Glu, Lys, Arg, His) (17 %).
T.9	(Ala, Ser, Thr, Phe ₂ , Asp ₂ , Glu ₂ , Arg) (25 %).

précédentes, et d'autre part au moyen de la carboxypeptidase, etc. Les résultats trouvés, combinés à ceux qui avaient été précédemment obtenus dans ce laboratoire (tableau 1) ont permis d'attribuer à ces peptides les structures indiquées dans la tableau 5.

Tableau 5

STRUCTURE DES PEPTIDES ISOLÉS APRÈS HYDROLYSE TRYPSIQUE DU LYSOZYME NON OXYDÉ²⁰

T.6	Asp(NH ₂).Arg
T.7	Thr.Pro.Gly.Ser.Arg
T.8	His.Lys.Glu.*Asp*.Ala.Tyr.Gly.Ser.Leu.Asp.Asp(NH ₂).Tyr.Arg
T.9	Phe.Glu.*Ser.Phe.Asp.Glu.*Ala.Thr.Asp.*Arg

Les données des tableaux 4 et 5 montrent que les peptides, isolés après action de la trypsine sur le lysozyme simplement dénaturé par la chaleur, présentent tous les quatre l'arginine comme acide aminé C-terminal. Aucun ne contient de cystine; un seul renferme de la lysine. On peut donc s'attendre à ce que ce soit les peptides, qui, renfermant de la cystine, sont retenus dans les conditions actuelles sur la colonne, et qui possèdent la lysine comme acide aminé C-terminal. Cette observation montre l'intérêt d'oxyder le lysozyme avant de le soumettre à l'action de la trypsine pour pouvoir obtenir la totalité des peptides libérés par hydrolyse trypsique.

Action de la trypsine sur le lysozyme oxydé. La transformation des ponts —S—S— du lysozyme en groupes sulfoniques a été faite ici au moyen de l'acide performique, dans les conditions particulièrement douces décrites par

* Il n'a pas été déterminé si ce résidu est amidé ou non.

du Vigneaud et col.²¹ On doit toutefois remarquer qu'il est pratiquement impossible d'éviter lors de l'oxydation en question la destruction de la majeure partie du tryptophane, qui se trouve essentiellement transformé en kynurénine¹⁹; il semble toutefois que, au moins dans la majorité des cas, la transformation subie par les résidus de tryptophane n'entraîne pas la rupture des liaisons peptidiques auxquelles ils prennent part. D'autre part, il semble également que, dans les conditions actuelles, la destruction des résidus de tyrosine soit à peu près complètement évitée. Néanmoins, il convient d'interpréter avec prudence les données obtenues à partir du lysozyme ainsi oxydé.

Le lysozyme oxydé a été soumis à l'action de la trypsine dans les mêmes conditions que le lysozyme non oxydé. Les produits résultant de cette action ont été séparés par chromatographie sur colonne de Dowex 50 \times 2. Les conditions de cette séparation et les résultats qu'elle a fournis sont indiqués dans la fig. 2.¹⁸

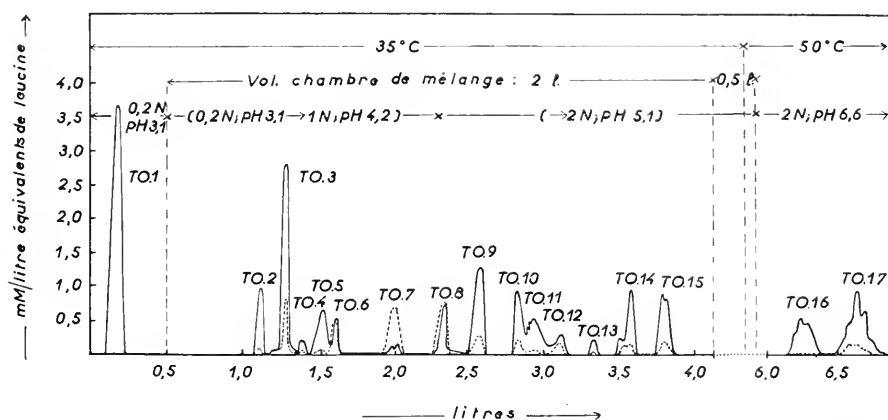


Fig. 2. Séparation des produits obtenus par action de la trypsine sur 25 μ M de lysozyme oxydé.

— coloration de la réaction à la ninhydrine après hydrolyse alcaline.

----- coloration de la réaction à la ninhydrine effectuée directement.

On constate ici la présence de 17 pics; après cette première séparation, les substances correspondant à chacun des pics ont été, comme dans le cas du lysozyme non oxydé, soumises à divers contrôles concernant leur homogénéité et ont été purifiées par les méthodes habituelles. Il a été ainsi possible de caractériser de façon précise chacune de ces substances: la nature des acides aminés et la composition des peptides sont données dans le tableau 6.

En plus des acides aminés libres (leucine, lysine, arginine) et de l'ammoniac, 12 peptides ont été obtenus avec un rendement suffisant pour permettre l'étude de leurs structures. Ces structures ont été établies pour 9 d'entre eux. Les résultats trouvés, combinés à ceux précédemment obtenus (tableau 1) ont permis d'attribuer à ces peptides les constitutions indiquées dans le tableau 7.

Tableau 6

COMPOSITION DES PICS SÉPARÉS APRÈS HYDROLYSE TRYPSIQUE DU LYSOZYME PRÉALABLEMENT OXYDÉ¹⁸

Les rendements pour chaque peptide sont indiqués entre parenthèses

TO.1	(Ala ₂ , Ser ₂ , CySO ₃ H, Thr ₁₋₂ , Pro, Val, Leu(?), Ileu, Asp ₂₋₃ , Lys).
TO.2	même composition que TO.1 (plus amidé) (6 %).
TO.3	{ Leucine libre (15 %) (correspond à leucine terminale) (Cf T.1), (Ala ₃ , CySO ₃ H, (MetO ₂), Leu, Glu, Lys) (35 %).
TO.4	non étudié (5 %).
TO.5	(Gly ₂ , Ser ₁ , Thr ₁₋₂ , Leu, Ileu ₂ , Tyr, Asp ₂₋₃ , Glu).
TO.6	(CySO ₃ H, Lys) (35 %).
TO.7	NH ₃ (Cf. T.3).
TO.8	Lysine libre (78 %) (correspond à lysine initiale) (Cf. T.4).
TO.9	(Gly, CySO ₃ H, Leu, Arg) (58 %) (peptide C-terminal).
TO.10	(Ala ₃ , CySO ₃ H, (MetO ₂), Leu, Glu, Lys, Arg) (17 %).
TO.11	(Gly ₂ , Ser ₂ , Thr ₁₋₂ , Leu, Ileu ₂ , Tyr, Asp ₄ , Glu, Arg) (7 %).
TO.12	(Gly, Val, Phe, Arg) (16 %).
TO.13	non étudié.
TO.14	{ Arginine libre (Cf. T.5). (Asp(NH ₂), Arg) (79 %).
TO.15	(Gly, Ser, Thr, Pro, Arg) (60 %).
TO.16	(Gly, Ala, Ser, Leu, Tyr ₂ , Asp ₃ , Glu, Lys, Arg, His) (19 %).
TO.17	(Ala, Ser, Thr, Phe ₂ , Asp ₂ , Glu ₂ , Arg) (25 %).

Tableau 7

STRUCTURE DES PEPTIDES ISOLÉS APRÈS HYDROLYSE TRYPSIQUE DU LYSOZYME OXYDÉ¹⁹

TO.3	CySO ₃ H.Glu.*Ala.Leu.Ala.Ala.MetO ₂ .Lys
TO.6	CySO ₃ H.Lys
TO.9	Gly.CySO ₃ H.Arg.Leu
TO.10	CySO ₃ H.Glu*.Ala.Leu.Ala.Ala.MetO ₂ .Lys.Arg
TO.12	Val.Phe.Gly.Arg
TO.14	Asp(NH ₂).Arg (Cf. T.6).
TO.15	Thr.Pro.Gly.Ser.Arg (Cf. T.7).
TO.16	His.Lys.Glu*.Asp*.Ala.Tyr.Gly.Ser.Leu.Asp.Asp (NH ₂). Tyr.Arg. (Cf. T.8).
TO.17	Phe.Glu*.Ser.Phe.Asp.Glu*.Ala.Thr.Asp*.Arg (Cf. T.9).

Les données du tableau 7 montrent tout d'abord que les peptides TO.14, TO.15, TO.16 et TO.17 isolés après action de la trypsine sur le lysozyme oxydé sont respectivement les mêmes que les peptides T.6, T.7, T.8 et T.9 résultant de l'hydrolyse trypsique du lysozyme non oxydé. On voit également que, parmi les peptides nouveaux obtenus, quatre correspondent à des enchaînements du lysozyme contenant de la cystine et deux possèdent la

* Il n'a pas été déterminé si ce résidu est amidé ou non.

lysine comme acide aminé C-terminal; on constate d'autre part que le peptide TO.10 correspond au même enchaînement que le peptide TO.3 avec seulement un résidu d'arginine supplémentaire, en position C-terminale. Enfin, le peptide TO.9 est particulièrement intéressant: il se différencie de tous les autres peptides obtenus par hydrolyse trypsique et dont la structure a été établie, du fait que l'acide aminé qu'il possède en position C-terminale n'est ni la lysine, ni l'arginine, mais la leucine. On verra plus loin que le peptide TO.9 correspond à l'enchaînement C-terminal du lysozyme.

REMARQUES

1. Le peptide TO.5 mérite une mention particulière; il ne contient en effet aucun résidu d'acide aminé basique; sa production correspond donc à une coupure aberrante, peut-être due à la présence d'un enzyme protéolytique étranger dans l'échantillon de trypsine utilisé. De toute façon, le rendement de ce peptide est très faible.

2. La liste des peptides fournis par les tableaux 6 et 7 ne rend pas compte de l'ensemble des résidus d'acides aminés du lysozyme; il manque notamment les fragments de la protéine qui contiennent les résidus de tryptophane. Il est intéressant à ce sujet de noter qu'il a été possible d'isoler sur papier, après action de la trypsine sur le lysozyme non oxydé, un peptide dont la structure est Val.Ala.Try.Arg. Ce peptide ne se retrouve pas après chromatographie de l'hydrolysats trypsique sur colonne de Dowex 50 \times 2; comme probablement les autres peptides renfermant du tryptophane, il reste fortement adsorbé sur la résine et ne pourrait être élué qu'en milieu très alcalin.

HYDROLYSE DU LYSOZYME PAR LA PEPSINE

Il paraît utile de donner ici les résultats d'une première étude des produits résultant de l'action de la pepsine sur le lysozyme.²² Comme dans le cas de l'hydrolyse trypsique, l'hydrolyse par la pepsine a porté d'une part sur le lysozyme non oxydé, ici le lysozyme natif, et d'autre part sur le lysozyme préalablement oxydé par l'acide performique. Dans les deux cas, la pepsine exerce son action pendant 24 heures, à 37°, à pH 2,2, le rapport enzyme/substrat étant de 1/100. Les premières séparations des produits formés sont effectuées par chromatographie sur colonne de Dowex 50 \times 2, comme précédemment. Les détails de ces séparations et leur résultat sont indiqués dans la fig. 3 pour le lysozyme natif et dans la fig. 4 pour le lysozyme oxydé.

Les fractions correspondant à un pic donné sont réunies, dessalifiées puis soumises aux opérations habituelles de purification. On constate ainsi que la plupart des pics renferment deux et quelques-uns même trois peptides différents. La structure de quelques uns seulement des peptides ainsi obtenus a été jusqu'ici établie:

PA. 6 ou PO. 6	Val.Glu(NH ₂).Ala	PA. 16 b	Gly.Tyr.Ileu.Leu
PA. 7 ou PO. 7	Ileu.Thr.Ala	PA. 17 ou PO. 17	Ala.Lys.Phc
PA. 12 ou PO. 12	Gly.Ileu.Leu	PA. 18c ou PO. 18c	Tyr.Arg.Gly

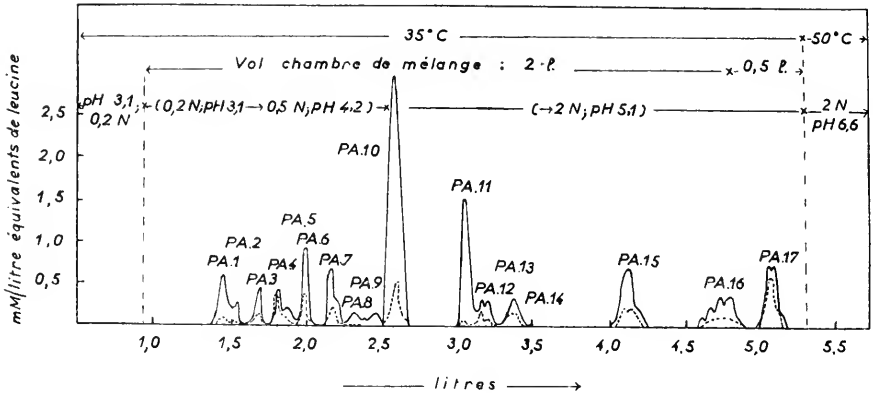


Fig. 3. Séparation des produits obtenus par action de la pepsine sur 20 μ M de lysozyme non oxydé.

— coloration de la réaction à la ninhydrine après hydrolyse alcaline.
 ----- coloration de la réaction à la ninhydrine effectuée directement.

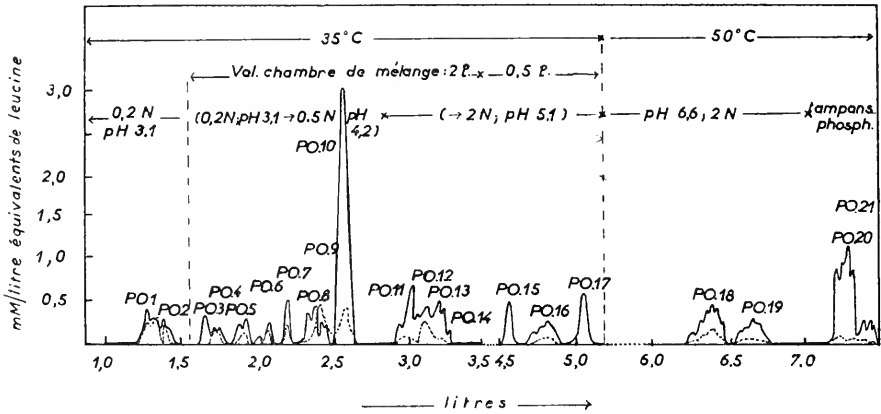


Fig. 4. Séparation des produits obtenus par action de la pepsine sur 20 μ M de lysozyme oxydé.

— coloration de la réaction à la ninhydrine après hydrolyse alcaline.
 ----- coloration de la réaction à la ninhydrine effectuée directement.

ETAT DES CONNAISSANCES ACQUISES DANS LE PRÉSENT TRAVAIL SUR LA STRUCTURE DU LYSOZYME

Enchaînement N-terminal et enchaînement C-terminal. L'enchaînement N-terminal a été, il y a déjà quelques années⁴, démontré être: H.Lys.Val.Phe. Gly.Arg. Il est donc inutile d'y revenir ici. L'enchaînement C-terminal a fait récemment l'objet d'une étude détaillée.²³ De nombreux arguments montrent qu'il est constitué par Gly.Cys.Arg.Leu.OH. Les principaux de ces

arguments sont les suivants: 1) la leucine C-terminale dans le lysozyme est immédiatement précédée d'un résidu d'acide aminé basique, lysine ou arginine, étant donné la spécificité de la trypsine: 2) le peptide TO.9, obtenu avec un rendement de 60 %, est le seul peptide qui possède la leucine comme acide aminé C-terminal, cette leucine étant immédiatement précédée par l'arginine: 3) l'action de la carboxypeptidase sur le DNP-lysozyme conduit, après un temps suffisamment long, à la libération d'une quantité faible mais nette d'arginine, alors que la leucine terminale a déjà été détachée.

Enchaînements intérieurs. Tenant compte de ce que la molécule de lysozyme renferme seulement trois résidus de tyrosine, trois résidus de phénylalanine, un résidu d'histidine, et utilisant les peptides caractérisés antérieurement dans ce laboratoire (Tableaux 1 et 2), il est possible de représenter l'état actuel de nos connaissances sur la structure du lysozyme comme il est indiqué dans le tableau 8.

Tableau 8

ENSEMBLE DES RÉSULTATS OBTENUS DANS
LE PRÉSENT TRAVAIL SUR LA STRUCTURE DU
LYSOZYME

<p>H.Lys.Val.Phe.Gly.Arg. [.Arg.His.Lys.Glu*.Asp(NH₂).Ala.Tyr.Gly.Ser.Leu.Asp.Asp(NH₂). Tyr.Arg.Gly.],[Ala.Lys.Phe.Glu*.Ser.Phe.Asp.Glu*.Ala.Thr.Asp*. Arg.],[Cys.Glu*.Ala.Leu.Ala.Ala.Met.Lys.Arg.],[Thr.Pro.Gly. Ser.Arg.],[Val.Ala.Try.Arg.],[Asp.Ala.Try],[Ileu.Thr.Ala.],[Gly.Ileu.Leu.], [.Val.Glu(NH₂).Ala.],[Asp(NH₂).Arg.],[Leu.Pro.],[Cys.Lys.], [.Arg.Gly.Tyr.Ileu.Leu.(Gly,Ser₂,Thr,Ileu,Asp₄*,Glu*).Arg.], Gly.Cys.Arg.Leu.OH.</p>
--

Les données du tableau 8 appellent les commentaires suivants: 1) L'existence du peptide PO.17 (ou PA.17) Ala.Lys.Phe prouve que l'enchaînement Ala.Lys doit précéder immédiatement le peptide TO.17 (ou T.9); en effet les places des trois résidus de phénylalanine sont connues et le peptide TO.17 (ou T.9) est le seul présentant la phénylalanine comme résidu N-terminal. 2) L'existence du peptide PO.18 Tyr.Arg.Gly permet d'allonger d'un résidu d'acide aminé le peptide TO.16 (ou T.8) du côté C-terminal; le peptide TO.16 (ou T.8) est en effet le seul qui, obtenu par hydrolyse trypsique, possède l'enchaînement Tyr.Arg comme enchaînement C-terminal, et les places des deux autres résidus de tyrosine sont connues. 3) La formation par hydrolyse trypsique du peptide TO.16 (ou T.8), présentant un résidu d'histidine comme résidu N-terminal, implique que dans le lysozyme ce résidu soit lié à un résidu de lysine ou d'arginine. L'isolement du peptide Arg.His.Lys

* Il n'a pas été déterminé si ces résidus sont amidés ou non.

(1. tableau 1) montre que ce résidu est l'arginine. 4) Les enchaînements Ala.Lys.Phe.Glu (7. tableau 1) et Glu.Ser.Phe.Asp (6. tableau 1) font partie d'un enchaînement plus important auquel participe également le peptide TO.17 (ou T.9). 5) Les enchaînements 2 et 4 du tableau 1, tous deux marqués par la tyrosine, se rencontrent l'un à la suite de l'autre dans le peptide TO.16 (ou T.8). Il convient toutefois de remarquer qu'une correction est ici nécessaire: le résidu asparagine C-terminal dans l'enchaînement 4 du tableau 1 n'est pas précédé immédiatement d'un résidu de leucine, mais en est séparé par un résidu d'acide aspartique. D'autre part, deux résidus de tyrosine se trouvant ainsi localisés dans le peptide TO.16 (ou T.8), le seul résidu de tyrosine restant disponible est celui qui se trouve dans le peptide TO.11. Il doit donc faire partie de l'enchaînement 3 du tableau 1. On est donc conduit à penser que le peptide TO.11 fait partie d'un ensemble dont la structure partiellement connue est celle indiquée dans le tableau 8. 6) Le peptide TO.10 correspondant à l'enchaînement Cys.Glu.Ala.Leu.Ala.Ala.Met.Lys.Arg est en bon accord avec le tripeptide Ala.Ala.Met (10. tableau 1), mais il est en contradiction avec l'enchaînement Asp.Ala.Met.Lys.Cys.Arg proposé par Thompson (8. tableau 3). D'autres contradictions se présentent entre les résultats de cet auteur et ceux du présent travail: l'enchaînement Ala.Lys.Phe.Glu.Ser.Phe.Asp (7 et 6. tableau 1) n'est pas compatible avec l'enchaînement Lys.Phe.Glu.Gly (11. tableau 3) et l'enchaînement Thr.Pro.Gly.Ser.Arg (T.7) n'est pas compatible non plus avec l'enchaînement Val.Thr.Pro.Gly.Ala (9. tableau 3) de l'auteur australien. Dans ce dernier cas, la vadilité de l'enchaînement proposé ici est confirmée par les observations suivantes: le lysozyme ne contient que deux résidus de proline, or l'un est engagé dans l'enchaînement Leu.Pro (8. tableau 1) ou Ileu.Pro (19. Tableau 3), l'autre dans l'enchaînement Thr.Pro.Gly trouvé également par Thompson. Du côté N-terminal, la spécificité de l'action de la trypsine suggère que Thr. doit être précédé d'un résidu d'acide aminé basique. Du côté C-terminal, la présence d'un résidu d'alanine lié à un résidu de glycocolle n'a jamais pu être constatée. 7) Il convient enfin de signaler ici l'incompatibilité totale entre l'enchaînement C-terminal proposé par Ohno et celui qui fait l'objet du présent exposé.

Pour expliquer partiellement les contradictions qui viennent d'être signalées, on peut remarquer que les peptides étudiés ici ont été obtenus essentiellement par voie enzymatique; ils sont donc relativement longs et peu nombreux et assez faciles à purifier, alors que Thompson a obtenu, après hydrolyse acide partielle, un mélange complexe de di- et de tripeptides correspondant à des enchaînements partiels dont la position à l'intérieur de la molécule de lysozyme est très difficile à fixer avec certitude.

REMARQUES SUR LA SPÉCIFICITÉ DE LA TRYPSINE

L'action de la trypsine a permis, comme on vient de le voir, d'isoler un certain nombre de peptides dont la structure a pu être déterminée. En outre

l'étude de cette action a permis de préciser certaines particularités concernant la spécificité de l'enzyme. On constate ainsi que la trypsine n'a pas scindé la liaison Lys.Glu dans l'enchaînement du peptide TO.16 (ou T.8). On peut attribuer la résistance de cette liaison à la présence du groupement carboxylique libre du résidu glutamique; on retrouve d'ailleurs la même résistance de la liaison Lys.Glu vis-à-vis de la trypsine dans le cas de la ribonucléase.¹⁵ Il est également intéressant de noter que l'enchaînement Arg.His.Lys (1. tableau 1) a été scindé par la trypsine uniquement entre Arg et His. Il n'a été trouvé ni histidine libre ni peptide His.Lys. On pourrait toutefois penser que si le résidu lysine dont il s'agit ici n'avait pas été engagé dans la liaison Lys.Glu difficile à rompre, une certaine quantité du peptide His.Lys eut été trouvée. Enfin, il convient de souligner les différences dans les rendements de leucine libre (leucine C-terminale) provenant du peptide TO.9: le rendement obtenu après hydrolyse du lysozyme simplement dénaturé est de 50 %; ce rendement tombe à 15 % dans le cas du lysozyme oxydé. Ce fait peut être expliqué par la présence, à côté de l'arginine, du groupement sulfonique du résidu acide cystéique, qui par sa forte acidité doit inhiber l'action de la trypsine.

BIBLIOGRAPHIE

1. H. L. FEVOLD, *Adv. Prot. Chem.*, **6**, 187 (1951).
2. A. R. THOMPSON, *Biochem. J.*, **60**, 507 (1955).
3. W. A. SCHROEDER, *J. Am. Chem. Soc.*, **74**, 5118 (1952).
4. R. ACHER, U. R. LAURILA, J. THAUREAUX et C. FROMAGEOT, *Biochim. Biophys. Acta*, **14**, 151 (1954).
5. J. I. HARRIS, *J. Am. Chem. Soc.*, **74**, 2944 (1952).
6. A. R. THOMPSON, *Nature*, **169**, 495 (1952).
7. L. PÉNASSE, M. JUTISZ et C. FROMAGEOT, *Bull. Soc. Chim. biol.*, **35**, 376 (1953).
8. H. FRAENKEL-CONRAT, *J. Am. Chem. Soc.*, **73**, 625 (1951).
9. R. ACHER, J. THAUREAUX, C. CROCKER, M. JUTISZ et C. FROMAGEOT, *Biochim. Biophys. Acta*, **9**, 339 (1952).
10. R. ACHER, U. R. LAURILA et C. FROMAGEOT, *Biochim. Biophys. Acta*, **19**, 97 (1956).
11. J. THAUREAUX et R. ACHER, *Biochim. Biophys. Acta*, **20**, 559 (1956).
12. U. R. LAURILA, expériences inédites.
13. A. R. THOMPSON, *Biochem. J.*, **61**, 253 (1955).
14. K. OHNO, *J. Biochem. (Japan)*, **42**, 615 (1955).
15. C. H. W. HIRS, S. MOORE et W. H. STEIN, *J. Biol. Chem.*, **219**, 623 (1956).
16. G. ALDERTON, W. H. WARD, et H. L. FEVOLD, *J. Biol. Chem.*, **157**, 43 (1945).
17. L. GORINI, F. FELIX et C. FROMAGEOT, *Biochim. Biophys. Acta*, **12**, 283 (1953).
18. P. JOLLÈS et J. THAUREAUX, *Compt. rend.*, **243**, 1685 (1956).
19. J. JOLLÈS-THAUREAUX, P. JOLLÈS et C. FROMAGEOT, *Biochim. Biophys. Acta*, **27** (1958).
20. J. THAUREAUX et P. JOLLÈS, *Compt. rend.*, **243**, 1926 (1956).
21. J. M. MUELLER, J. G. PIERCE et V. DU VIGNEAUD, *J. Biol. Chem.*, **204**, 857 (1953).
22. P. JOLLÈS, J. JOLLÈS-THAUREAUX et C. FROMAGEOT, *Biochim. Biophys. Acta*, **27** (1958).
23. P. JOLLÈS, J. THAUREAUX et C. FROMAGEOT, *Arch. Biochem. Biophys.*, **69**, 290 (1957).

The isolation of an immunologically active fragment of bovine serum albumin

R. R. PORTER

National Institute for Medical Research, Mill Hill, London, N.W.7

It has been recognized for about thirty years that proteins may possess intrinsic biological activity as enzymes, hormones, antigens and antibodies, and Northrop (1932) was perhaps the first to formulate and try to answer the next question: Is the whole protein molecule essential for these activities or is it a property of only part of the molecule? Northrop attempted to hydrolyse crystalline pepsin without loss of activity but failed. Success was soon achieved with antibodies (horse anti-diphtheria toxin serum) by Parventjev (1936), Petermann and Pappenheimer (1941), and Northrop (1942), the last two isolating approximately half molecules which would still combine with the antigen. Later a fifth part of rabbit anti ovalbumin antibody was obtained, by papain digestion, which retained the specific combining power (Porter, 1950).

Recently the partial degradation of several enzymes without loss of activity has been reported, e.g. pepsin (Perlman, 1954), papain (Hill and Smith, 1956) and ribonuclease (Richards, 1955; Kalnitsky and Rogers, 1956; Rogers and Kalnitsky, 1957). The degradation, in several cases, appeared to be very considerable, although the isolation of the fragment and proof that the active fragments were as small as other evidence appeared to suggest has not yet been described. Some terminal amino acids may be removed from the hormones insulin (Harris and Li, 1952), adrenocorticotrophin and growth hormone (Harris, Li, Condliffe and Pon, 1954) and also from tobacco mosaic virus (Harris and Knight, 1952) without loss of activity.

The combining centre of a protein antigen might appear to offer the easiest opportunity of isolating a specifically active fragment of protein, as Landsteiner's work with artificially conjugated antigens showed that the specificity is directed, in these compounds, to the small attached groups—the haptenes. This he showed by using the hapten to specifically inhibit the combination of the conjugated antigen with its antiserum. However, a very considerable excess of hapten is required to achieve inhibition (a molar

ratio of haptene to antigen of more than 1000/1) suggesting that the haptene may be only a small part of the actual combining site. The only attempt to isolate an antigenic combining site from a protein was carried out by Landsteiner (1942). He used silk fibroin as antigen rendered soluble by treatment with acid and increased its antigenic power by adsorption onto charcoal before injecting it into rabbits. The silk fibroin was then degraded by strong sulphuric acid and the digest was tested for its ability to inhibit the combination of antigen and antiserum. Evidence was obtained showing that apparently small peptides would inhibit this combination and that they must retain therefore some of the features of the combining sites of the original antigen. With the methods available at the time it was not possible to characterize the active fragments further and as the original antigen was partially degraded fibroin protein it was not clear what significance these results bore relative to the nature of the combining centres of native globular protein antigens. Other work has suggested that native globular proteins will retain the power to combine with specific antisera after at least slight enzymic hydrolysis (Landsteiner and Chase, 1933; Holiday, 1939; Kleczkowski, 1945; Lapresle, 1955*a, b*).

We have therefore investigated the problem further and have found that, if crystalline bovine serum albumin is hydrolysed by chymotrypsin in a dialysis sack suspended in water brought to pH 8 with NH_3 at 25° , then material appears in the outer liquor (the diffusate) which will specifically inhibit the combination of bovine serum albumin with its antiserum. Full details of this work have been given elsewhere (Porter, 1957).

It was found that the active material (subsequently referred to as the inhibitor) would not pass through a different sample of dialysis tubing and hence after concentration in a rotary still it was redialysed in this second sack with a smaller pore size. About half the weight was lost without loss of activity. It was subsequently purified by ethanol precipitation and zone electrophoresis and finally by precipitation with 5% trichloroacetic acid. The precipitate redissolved in water and there was no loss of immunological activity, but small adsorbed peptides were removed.

The material now appeared to be pure. Electrophoresis and ultracentrifugation showed only one component and end group assay showed the presence of only *N*-terminal amino acid—phenylalanine—present as 1 mole per 14,000.

Estimation of molecular weight by sedimentation gave a value of about 24,000—much higher than had been expected—and it was found that the purified inhibitor would no longer pass through the large pore size dialysis tubing as it had done in the first stage of preparation. It was eventually shown that dimerization had occurred during purification and that this could be reversed by cysteine at pH 4. Under these conditions the sedimentation coefficient corrected to water at 20° fell from 2.13 to 1.42 and the calculated molecular weight from 24,000 to 12,000. The latter figure is close to the

minimum molecular weight found by end group assay of 14,000. The monomer readily dimerizes on being brought to neutral or slightly alkaline reaction in the absence of a reducing agent. The dimerization of human and bovine serum albumin on addition of mercuric ion or a divalent organic mercurial has been extensively studied at Harvard since the original observation of the phenomenon by Hughes (1950). It was shown that 70-80% of the molecules possess a single SH group which is apparently unable to interact with another SH group for steric reasons but which will cause dimerization in the presence of Hg. It seems probable that the 12,000 fragment which we have isolated possesses this single sulphhydryl group but owing to the removal of four-fifths of the original molecule the steric hindrance has been eliminated and these fragments can now dimerize freely. Amino acid analysis shows big differences (Table 1) between the inhibitor

Table 1

AMINO ACID ANALYSIS OF INHIBITOR
AND OF SERUM ALBUMIN

Results are expressed as g./100 g. of protein

Amino acid	Inhibitor	Serum albumin*
Aspartic acid	9.0	10.9
Threonine	7.2	5.8
Serine	2.0	4.2
Glutamic acid	16.3	16.5
Proline	7.3	4.8
Glycine	1.9	1.8
Cystine	2.7	6.0
Alanine	5.6	6.3
Valine	9.3	5.9
Methionine	0.8	0.8
Leucine	11.6	12.3
Isoleucine	2.8	2.6
Tyrosine	1.2	5.1
Phenylalanine	6.1	6.6
Tryptophan	0†	0.58
Lysine	12.9	12.8
Histidine	2.1	4.0
Arginine	5.4	5.9

and the whole molecule and it may be noted that there are 1.4 molecules of cystine present suggesting that there is one disulphide bond and a free SH group.

The immunological activity of the fragment is lost if it is dissolved in

* The amino acid content of serum albumin has been taken from Stein and Moore (1949).

† Tryptophan could not be detected spectroscopically with the Holiday photographic method (Beaven and Holiday, 1952).

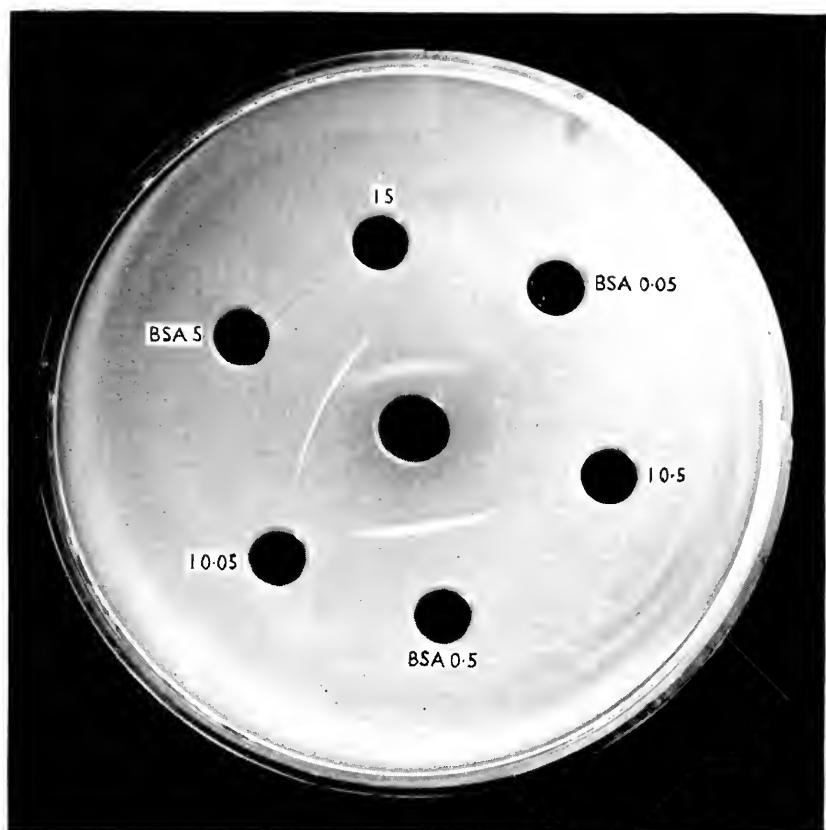


Fig. 1. Comparison of behaviour of bovine-serum albumin (BSA) and inhibitor (I) on diffusion into antiserum in agar gel by the method of Ochterlony (1953). Concentrations, 5.0.5 and 0.05 mg./ml. with both reagents. Centre cup contains rabbit anti-bovine serum albumin. Photograph was taken after the plate had been standing for 5 days at 2°.

a strongly reducing medium such as neutral M sodium thioglycollate—the solution turning into a gel on standing at room temperature. Apparently when the internal disulphide bond is broken, the configuration of the molecule is disrupted with loss of solubility and biological activity. Further evidence of the importance of configuration was obtained. If a solution of inhibitor is heated to 100° for 2 min. a visible precipitate forms and no activity is left. If the inhibitor is dissolved in 3M urea at room temperature then the activity is lost slowly—it is reduced to about half in 24 hours. It seems clear that the combining site of the fragment is a relatively large section of the peptide chain, and its ability to combine with antibody is dependant on the integrity of the native configuration. This is lost if the disulphide bond is broken or if more labile bonds are broken by heat or urea. At the same time the inhibitor is stable to ethanol at room temperature and also to pH 2 and pH 10·5—the latter observation suggesting that electrostatic bonds are not important in preserving the native configuration.

When the immunological properties of this fragment were investigated it became clear that only some, perhaps only one, of the combining sites of the original antigen were present in the inhibitor. For example, with anti-serum from rabbits after a short course of injections with bovine serum albumin, the inhibitor would combine and subsequently prevent the precipitation with serum albumin of about one-third of the antibody present. This power to combine with only part of the antibodies could be demonstrated clearly when the behaviour of whole antigen and inhibitor were compared by precipitation in agar gel according to the technique of Ochterlony. In Fig. 1 it can be seen that a typical Y type of interaction occurs showing that the inhibitor is combining with part of the antibody, the remainder diffusing through until it meets the serum albumin front. This presumably means that antibodies are present in the serum which are directed to antigenic binding sites situated in the four-fifths of the molecule which has been lost.

It was interesting to observe that in serum, taken from animals which had been given a prolonged series of injections of serum albumin, the inhibitor could combine with a much larger proportion of the antibodies present and, in some cases, with all antibodies present giving complete inhibition of precipitation. This is taken to mean that, as immunization continues, the antibodies produced are directed against an increasing number of the potential antigenic sites on a protein. In the example quoted where complete inhibition was achieved the molar ratio required for this was 1·8 moles inhibitor to 1 mole serum albumin. This low value makes it improbable that the inhibition is competitive and other evidence such as the inability of excess serum albumin to reverse the inhibition supports this view. This low ratio is also in contrast to the results of Landsteiner using haptenes where a very large molar excess is required for inhibition of precipitation with the conjugated antigen. This difference may arise because

the inhibitor contains the whole surface of the combining centre which binds more firmly to the antibody site than the smaller haptenic groups. The importance of configuration in the former suggests that a relatively large surface area is involved. With a small haptenic group the secondary valence forces would be less and binding correspondingly less firm.

The dimerization of the inhibitor had no influence on its immunological behaviour showing that the surface adjacent to the sulphhydryl group, which would be blocked in the dimer, could not be a part of the combining site. As the work on the dimerization of the whole molecule suggests that the sulphhydryl group is sunk into the surface and is not readily accessible, it is perhaps not surprising that it is not part of a combining site. Unless an antigen is partly broken down prior to its participation in stimulating antibody formation, it would be expected that only the surface of the molecule determines the complementary character of the antibody combining site. It should also be noted that the part of the inhibitor surface adjacent to the sulphhydryl group could play an essential part in determining the configuration of the combining centre, though not itself part of it.

The inhibitor will cause anaphylactic shock when injected into a guinea pig positively immunized with rabbit anti-bovine serum albumin serum and at about the same level as that required to cause shock by injection of the whole antigen. Landsteiner and Van der Scheer (1933) showed that azo dye haptenes would cause anaphylactic shock in sensitized guinea pigs, if the dyes used had two haptenic groups. This type of hapten will also cause precipitation of antisera, but with the inhibitor molecule there does not appear to be any correlation between precipitating power and ability to cause anaphylactic shock.

SUMMARY

A 12,000 mol. wt. fragment of bovine serum albumin has been isolated from a chymotryptic digest.

This fragment retained the power to combine with the specific antibody as demonstrated by inhibition and precipitation studies and also by its power to provoke anaphylactic shock in sensitized guinea pigs.

The fragment readily dimerizes at neutral or alkaline reaction and evidence is presented which suggests that this dimerization is caused by interaction of SH groups, probably the same groups which lead to the formation of a Hg dimer mercaptalbumin with the original molecule.

REFERENCES

- BEAVEN, G. H. and HOLIDAY, E. R. (1952), *Advanc. Protein Chem.*, **7**, 319.
HARRIS, J. I. and KNIGHT, C. A. (1952), *Nature*, London, **170**, 613.
HARRIS, J. I. and LI, C. H. (1952), *J. Am. chem. Soc.*, **74**, 2945.
HARRIS, J. I., LI, C. H., CONDLIFFE P. G. and PON, N. G. (1954), *J. biol. Chem.* **209**, 133.

- HILL, R. L. and SMITH, E. L. (1956), *Biochim. Biophys. Acta*, **19**, 376.
- HOLIDAY, E. R. (1939), *Proc. Roy. Soc.*, **A**, **170**, 79.
- HUGHES, W. L. (1950), *Cold. Spr. Harb. Symp. quart. Biol.*, **14**, 79.
- KALNITSKY, G. and ROGERS, W. I. (1956), *Biochim. Biophys. Act.*, **20**, 378.
- KLECZKOWSKI, A. (1945), *Brit. J. exp. Path.*, **26**, 24.
- LANDSTEINER, K. (1942), *J. exp. Med.*, **75**, 269.
- LANDSTEINER, K. and VAN DER SCHEER, J. (1933), *J. exp. Med.*, **57**, 633.
- LANDSTEINER, K. and CHASE, M. W. (1933), *Proc. Soc. exp. Biol.*, N.Y., **30**, 1413.
- LAPRESLE, C. (1955a), *Ann. Inst. Pasteur*, **89**, 654.
- LAPRESLE, C. (1955b), *Bull. Soc. Chim. biol.*, Paris, **37**, 969.
- NORTHROP, J. H. (1932), *J. gen. Physiol.*, **16**, 33.
- NORTHROP, J. H. (1942), *J. gen. Physiol.*, **25**, 465.
- OCHTERLONY, O. (1953), *Acta path. microbiol. Scand.*, **32**, 231.
- PARVENTJEV, I. A. (1936), *U.S. Patent*, 2065, 196.
- PERLMAN, G. E. (1954), *Nature*, London, **173**, 406.
- PETERMANN, M. L. and PAPPENHEIMER, A. M. (1941), *J. Phys. Chem.*, **45**, 1.
- PORTER, R. R. (1950), *Biochem. J.*, **46**, 479.
- PORTER, R. R. (1957), *Biochem. J.*, **66**, 677.
- RICHARDS, F. M. (1955), *C.R. Lab. Carlsberg. Ser. Chim.*, **29**, 329.
- ROGERS, W. I. and KALNITSKY, G. (1957), *Biochim. Biophys. Act.*, **23**, 525.
- STEIN, W. H. and MOORE, S. (1949), *J. biol. Chem.*, **178**, 79.

Serum protein changes during differentiation

EARL FRIEDEN

Department of Chemistry, Florida State University,
Tallahassee, Florida, U.S.A.

Biochemists have long cherished the hope that the phylogenetic classification of animals, based essentially on morphology, would find a corresponding chemical expression of evolutionary differences between animals. The discussion here today has also served to emphasize the relation between species specificity and protein structure. I now wish to present some recent data which illustrate the variation in proteins that occur during developmental changes in a given species.

For some years we have been studying the chemical changes that are associated with amphibian metamorphosis.¹ Recently, in collaboration with Albert Herner, Lloyd Fish, and Mrs E. J. Casson Lewis, we have observed significant alterations in serum proteins during normal and induced metamorphosis.² Marked changes in both the amount and the distribution of the serum proteins were noted. The distribution of the serum proteins was estimated from paper electrophoresis studies of serum obtained from freshly clotted tadpole and frog blood.

As shown in Table 1, the Florida swamp frog has virtually no detectable protein in the electrophoretic fraction corresponding to serum albumin. As metamorphosis proceeds, the albumin fraction appears and increases until a normal complement of 45% is observed for the frog of this species. The common bullfrog tadpole has a small fraction of serum albumin which also increases until the albumin is about one-half of the total serum protein. The very young bullfrog and the adult bullfrog possess identical proportions of serum albumin. In early tadpoles of both of these species, the albumin fraction can be appreciably increased by the administration of thyroid hormone. The non-albumin fraction of tadpole serum protein seems to be predominantly globulin in character.

The implications of these observations for comparative biochemistry are too numerous and lengthy to be fully discussed here and will be presented elsewhere.² We think that it is not unlikely that the tadpole-frog system is reflecting a serum protein alteration typical of the change from aquatic to terrestrial forms.³

Certain subtle changes also occur in the hemoglobins of the metamorphosing bullfrog tadpole as reported previously by McCutcheon⁴ and by

Table 1

CHANGES IN SERUM ALBUMIN DURING AMPHIBIAN METAMORPHOSIS

	Serum albumin/total serum protein	
	Young Tadpole	Frog
Florida swamp frog (<i>R. hechsherii</i>)	<0.02	0.45
Common bullfrog (<i>R. catesbiana</i>)	0.10 (\pm 0.03)	0.49 (\pm 0.05)
„ „ injected with triiodothyronine	0.20	

Riggs.⁵ Our laboratory is in the process of accumulating sufficient amounts of these various blood proteins so that the structural basis for their differences may be investigated.⁶

REFERENCES

1. E. FRIEDEN and B. NAILE, *Science*, **121**, 37 (1955); J. L. DOLPHIN and E. FRIEDEN, *J. Biol. Chem.*, **217**, 735 (1955).
2. E. FRIEDEN, A. HERNER, L. FISH and E. J. CASSON LEWIS, *Science*, **126**, 559 (1957).
3. See H. F. DEUTSCH and M. B. GOODLOE, *J. Biol. Chem.*, **161**, 1 (1945), for an example of a comparative study of serum proteins.
4. F. H. MCCUTCHEON, *J. Cell. and Comp. Physiol.*, **8**, 63 (1936).
5. A. RIGGS, *J. Gen. Physiol.*, **35**, 23 (1951-2).
6. This work was aided by grants from the United States Public Health Service and the National Science Foundation.

La variabilité allotypique de certaines protéines du serum

JACQUES OUDIN

Laboratoire d'Immunochimie analytique, Institut Pasteur, Paris

Un exemple, connu depuis peu, de protéines dont la structure est susceptible de variations individuelles à déterminisme héréditaire, est celui que nous avons signalé récemment chez le lapin.^{1,2} Ces variations de structure se manifestent par des différences de spécificité antigénique, d'un individu à l'autre au sein d'une même espèce, entre des protéines qu'on croyait identiques. Le fait qu'il puisse exister, chez deux lapins, deux formes, de spécificité antigénique différente—ou allotypes—d'une même protéine, explique qu'on puisse immuniser l'un des deux lapins contre une ou plusieurs protéines du sérum de l'autre.

Le procédé d'immunisation dont nous nous sommes servi consiste à injecter à un lapin selon la technique décrite par Freund et ses collaborateurs³ du précipité spécifique formé par la réaction de l'immunsérum d'un autre lapin avec l'antigène homologue. La nature de l'antigène injecté ne semble pas intervenir. Lorsque l'injection entraîne la formation d'anticorps, ceux-ci réagissent avec le sérum qui a servi à l'immunisation et avec un certain nombre d'autres sérums de lapins. La quantité d'anticorps formés peut être telle que le poids d'azote précipité par 1 ml. d'immunsérum dépasse 0,35 mg.

La nature du matériel injecté tendrait à suggérer que l'antigène immunisant est le constituant de beaucoup le plus abondant dans le précipité spécifique, c'est à dire l'anticorps. Nous avons pu vérifier dans plusieurs cas qu'il en était ainsi. Un sérum précipitable, contenant l'antigène, était un immunsérum agglutinant S. Typhi (suspension 0) à la dilution 1/3200. Le même sérum agglutinant a été additionné de 20 fois son volume d'un immunsérum antilapin, et le précipité formé éliminé par centrifugation. Le sérum anti S. Typhi ainsi débarrassé de son antigène précipitable par l'immunsérum antilapin n'agglutinait plus S. Typhi qu'à la dilution 1/800. Nous avons observé dans des conditions analogues la diminution du titre H d'un autre sérum agglutinant de lapin anti S. Typhi. La diminution du titre agglutinant était évidemment due dans les deux cas à l'élimination préalable d'une partie des anticorps anti S. Typhi dans la réaction où ils avaient été précipités en tant qu'antigène.

Il n'est nullement nécessaire que le sérum précipitable soit celui d'un animal hyperimmunisé; dans la plupart des réactions effectuées, l'antigène se trouvait dans des sérums de lapins normaux, parfois même dans des sérums de lapins nouveau-nés. Ces réactions ont été effectuées pour une part en milieu liquide; mais nous avons été amené à utiliser de plus en plus la précipitation en milieu gélifié selon la technique de diffusion simple à une dimension, en tubes.^{4,5,6} Dans les réactions de plusieurs sérums, contenant l'antigène, avec un même immunosérum, cette technique permet en effet de mieux apprécier l'égalité ou la différence d'intensité de la réaction selon la densité de précipité de la zone, de déceler, même à intensité égale, des différences d'aspect, et naturellement de démontrer, par la multiplicité des zones de précipitation, que plusieurs antigènes réagissent.

Parmi les immunosérums, choisis après un grand nombre d'immunisations comme donnant avec les mêmes sérums de lapins des réactions de types différents, les quatre immunosérums dont les réactions sont représentées au tableau 1 peuvent servir d'exemples simples. Ces immunosérums ne donnent chacun, en effet, qu'une seule zone de précipitation, d'intensité constante. Le

Tableau 1

RÉACTIONS DE 4 IMMUNOSÉRUMS DE LAPINS AVEC 50 SÉRUMS DE LAPINS NON IMMUNISÉS

Immunosérums (Nos des lapins)	Réactions (gel) de 50 sérums de lapins non immunisés								Nombre total de réactions positives avec chaque immunosérum
420	+	+	+	+	+	+	o	o	48
432	+	+	+	o	o	o	+	o	22
194	+	o	o	+	o	o	+	+	15
287	o	+	o	o	+	o	o	o	7
Nombre de sérums don- nant chaque type de réaction	3	3	15	10	4	13	1	1	—

tableau indique donc seulement, par + ou o, la présence ou l'absence de la réaction manifestée par une zone de précipitation. On peut admettre que chacun de ces immunosérums contient des anticorps contre une structure spécifique qui se trouve être présente dans chacun des sérums avec lesquels il précipite, et absente des autres sérums. Il est commode d'appeler 'motifs allotypiques' les structures de ce genre, sans préjuger de leur nature chimique. On voit que la fréquence des motifs allotypiques responsables des réactions du tableau 1 est très variable: 48/50, 22/50, 15/50, 7/50. Pour qu'une réaction ait lieu, il faut que le motif allotypique qui en est responsable soit présent

dans le sérum immunisant et non dans le sérum de l'animal immunisé. Aucune réaction ne peut évidemment intéresser un motif qui est, soit présent dans l'un et l'autre, soit absent de l'un et de l'autre. Lorsque la fréquence d'un motif allotypique est très faible, la réalisation de ces conditions, nécessaires (mais non suffisantes) à l'apparition d'anticorps, devient très peu probable. Il en est de même pour un motif de fréquence très élevée. Or les fréquences indiquées au tableau 1 nous permettent, au prix d'une extrapolation vraisemblable, de supposer que d'autres motifs allotypiques existent, soit plus fréquents, soit moins fréquents que ceux dont il s'agit, dans la population étudiée. La probabilité que ces motifs rares ou fréquents échappent aux recherches peut donc être très grande. D'autre part, l'impression que donnent des motifs allotypiques différents, de ne pas susciter aussi aisément les uns que les autres la formation d'anticorps, est une raison supplémentaire de supposer que les motifs allotypiques décelés ne représentent qu'une très petite partie de ceux qui existent.

Si l'on répartit les individus en classes telles qu'à l'intérieur de chaque classe tous les sérums donnent exactement les mêmes réactions avec tous les immunosérums dont on dispose, le nombre des classes théoriquement possibles croît exponentiellement en fonction du nombre des motifs allotypiques qu'on est capable de déceler. Encore faudrait-il, pour qu'on puisse observer réellement le nombre ainsi prévisible de classes différentes (à supposer qu'on examine les sérums d'un nombre suffisant d'individus), que la présence d'un motif n'entraîne nécessairement ni la présence ni l'absence d'aucun autre motif. Dans le cas où la présence d'un motif entraînerait obligatoirement la présence d'un autre motif donné, on ne pourrait distinguer ces deux motifs qu'à la condition qu'ils soient portés par deux molécules différentes. La réaction en milieu gélifié pourrait alors déceler leur réaction sous forme de deux zones de précipitation distinctes, mais toujours jumelées. Certaines observations que nous avons faites pourraient s'expliquer de cette façon. Dans l'ensemble, la plupart des réactions observées décèlent des motifs qui peuvent exister l'un sans l'autre. Mais il semble qu'il y ait parfois, pour deux motifs allotypiques donnés, une certaine tendance à se trouver plus souvent, soit présents tous deux à la fois, soit même présent pour l'un et absent pour l'autre.

Du fait que deux motifs allotypiques peuvent exister l'un sans l'autre, il ne s'ensuit pas nécessairement qu'ils soient hérités séparément. L'expérience montre qu'ils peuvent l'être: le tableau 2 donne l'exemple des réactions des lapins de deux portées, comparées à celles des parents. Les conclusions qui peuvent être tirées d'autre part des croisements effectués sont actuellement limitées par l'ignorance où nous sommes de l'ascendance, et partant du génotype, des parents.

Rien ne permet jusqu'à présent de supposer qu'un état pathologique quelconque puisse être la conséquence des variations allotypiques que nous

Tableau 2

 RÉACTIONS DES SÉRUMS DES LAPINS DE DEUX PORTÉES,
 ET DES SÉRUMS DES PARENTS

Immunsérums Nos	Parents		Portée A					Parents		Portée B		
	10 ♂	32 ♀	60 ♀	61 ♀	62 ♀	63 ♂	64 ♂	44 ♂	52 ♀	101 ♂	102 ♂	109 ♀
420	+	+	+	+	+	+	+	+	+	+	+	+
432	o	+	o	+	+	o	+	+	+	+	+	+
194	o	o	o	o	o	o	o	o	+	o	+	+
287	+	o	+	+	o	+	o	+	o	+	+	o

avons observées, comme c'est le cas pour les variations individuelles de structure de l'hémoglobine.⁷

Il est vraisemblable que le phénomène dont il s'agit pourra être observé dans beaucoup d'autres espèces que le lapin. Nous avons pu déjà déceler, dans les mêmes conditions que chez le lapin, des réactions intraspécifiques analogues, quoique d'intensité plus faible, chez le cobaye, donnant à penser que dans cette espèce aussi une ou plusieurs protéines du sérum sont susceptibles de variabilité allotypique.

BIBLIOGRAPHIE

1. J. OUDIN, Réaction de précipitation spécifique entre des sérums d'animaux de même espèce, *Compt. rendus Acad. Sci.*, **242**, pp. 2489-90 (1956).
2. J. OUDIN, L' 'Allotypie' de certains antigènes protéidiques du sérum, *Compt. rendus Acad. Sci.*, **242**, pp. 2606-8 (1956).
3. J. FREUND et M. V. BONANTO, The effect of paraffin-oil, lanolin-like substances and killed tubercle bacilli on immunization with diphtheria toxoid and *Bact. typhosum*, *J. Immunol.*, **48**, 325-334 (1944).
4. J. OUDIN, Méthode d'Analyse Immuno-chimique par précipitation spécifique en milieu gélifié. *Compt. rendus Acad. Sci.* 2 janvier 1946, **222**, pp. 115-16.
5. J. OUDIN, *Specific Precipitation in Gels and Its Application to Immunochemical Analysis*, Methods in Medical Research, vol. 5, pp. 335-78; Year Book, publishers, Chicago (1952).
6. J. OUDIN, L'analyse immuno-chimique par la méthode des gels. Moyens et techniques d'identification des antigènes, *Ann. Inst. Pasteur*, **89**, pp. 531-55 (1955).
7. L. PAULING. Factors affecting the structure of haemoglobins and other proteins, *Colloque sur la structure des protéines*, Paris, 25 Juillet 1957.

Species variation and structural aspects in some pituitary hormones

CHOH HAO LI

The Hormone Research Laboratory and the Department of Biochemistry,
University of California, Berkeley, California

The pituitary body is distinguished from all the other endocrine glands by its structural division into three main parts. The posterior lobe, connected with the brain by the infundibulum, is the secretory site of the two peptide hormones, vasopressin and oxytocin. The intermediate lobe, which controls the dispersion of pigment granules, lies between the anterior lobe and the posterior lobe. It is, however, the glandular or anterior lobe that is responsible for the physiological dominance exercised by the pituitary over the other glands of internal secretion. Fig. 1 presents a diagrammatic summary of the biological properties of the pituitary hormones.

Despite the intensive investigations into the various biological functions of the anterior pituitary which have been carried out in recent years, there is as yet no conclusive evidence for attributing the functions of this lobe to any new hormone other than the six already established;^{1,2} namely, growth hormone (GH, or somatotropin (STH)), corticotropin (ACTH), thyrotropin (TSH), lactogenic hormone (luteotropin, prolactin), follicle-stimulating hormone (FSH), and interstitial-cell stimulating hormone (ICSH, or luteinizing hormone (LH)). These hormones are protein or peptide in nature, and three of them (TSH, FSH, and ICSH) appear to contain carbohydrates in addition to amino acids.³ These three carbohydrate-containing hormone proteins have not been isolated in pure form and hence no investigations of their structure have been undertaken.

The isolation, structure and synthesis of the posterior pituitary hormones have recently been accomplished by du Vigneaud and his collaborators.⁴ This brilliant achievement is well known and shall not be discussed here, except to note that an interesting species difference was observed between pig and beef vasopressins, the former peptide hormone containing lysine instead of the arginine found in the latter. On the other hand, the oxytocin

molecules from beef and pig pituitaries manifest no difference with respect to amino acid composition.

The speculation that the melanocyte-stimulating hormone (MSH, intermedin) of the intermediate lobe and ACTH might be identical^{5,6} has given a new impetus to the studies on the former hormone. This has led to recent

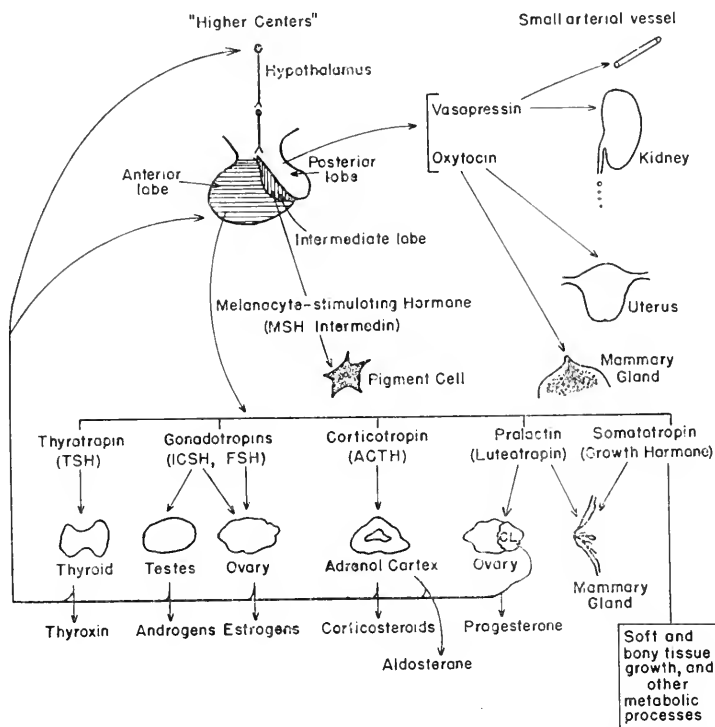


Fig. 1. A diagrammatical summary of the biological properties of pituitary hormones. CL.=corpus luteum.

successes in the isolation and structural elucidation of the hormone from both pig and beef pituitary glands.⁷ In this paper, we shall discuss some recent data derived from structural investigations of MSH, the corticotropins, and two anterior pituitary hormones (somatotropin and prolactin), especially with regard to species variations.

I. MELANOCYTE-STIMULATING HORMONE (MSH, INTERMEDIN)

In 1916, Smith⁸ and Allen⁹ independently reported a striking change in the pigmentation of the tadpole following hypophysectomy. According to Smith,^{10,11} the removal of the hypophysis in the tadpole is followed by a decrease in the amount of intracellular melanin as well as in the number

of the epidermal melanophores. Subsequently, Allen¹² and Swingle¹³ showed that transplantation of the intermediate lobe of adult frogs into normal or hypophysectomized tadpoles induces a marked expansion of the melanocytes. It was further demonstrated^{11,14} that extracts of the intermediate lobes of the bovine pituitary exercised the most marked influence on the pigmentation of the hypophysectomized tadpole, although suspensions of the anterior lobes were also effective in this regard. These earlier investigations have left no doubt that the pituitary gland possesses a principle or principles which regulate the pigmentation of amphibia.

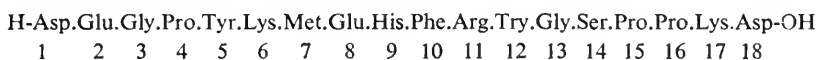
The main sources of starting material for the purification of the melanocyte-stimulating hormone have been the posterior-intermediate lobes of beef and pig pituitary glands. The pioneer studies of Hogben, Zondek, Jores, Stehle, Landgrebe, and their collaborators, have shown the hormone to be a polypeptide of comparatively low molecular weight. Since 1955, four groups of investigators^{15,16,17,18,19,20,21,22} have published methods for the purification and isolation of MSH from beef and pig pituitary glands. All the published procedures for the preparation of MSH concentrates from either pig or beef posterior-intermediate pituitaries incorporate the acetic acid extraction technique first introduced by Kamm *et al.*²³ for posterior pituitary hormones and later developed by Payne *et al.*²⁴ for ACTH. In addition, oxycellulose is employed to adsorb the MSH activity from the acetic acid extract according to the procedure commonly used for the corticotropins.²⁵ From these MSH concentrates, highly purified or pure preparations may be obtained with zone electrophoresis¹⁹ or countercurrent distribution¹⁸ or a combination of these two techniques.^{15,20,21,22}

Lee and Lerner¹⁵ reported the presence in hog pituitary glands of two substances possessing MSH activity; they observed two active peaks when pituitary powder prepared from either posterior or anterior lobes was submitted to countercurrent distribution in a 2-butanol-0.5% aqueous trichloroacetic acid system. The preparation with a high partition coefficient was designated as α -MSH, and the other component as β -MSH. These investigators believe that α -MSH represents the main MSH component of the hog pituitary gland because it accounts for the greatest percentage of the total MSH activity. They ascribe the fact that other workers have obtained β -MSH as the main component to the step in the fractionation procedure of Landgrebe and Mitchell^{26*} involving precipitation with acetone, during which they postulate that the α -MSH is separated from the β -MSH and discarded in the lyophilized filtrate. However, none of the other investigators^{18,19,20,21} have detected the presence of α -MSH in the course of the purification of their MSH preparations.

The structure of the pig β -MSH has been elucidated independently in

* This explanation cannot be applied to the preparation of Geschwind and Li;²¹ in their procedure the supernatant fluid was lyophilized instead of being precipitated with acetone.

two laboratories.^{27,20,28} The data obtained by Harris and Roos²⁷ and Geschwind *et al.*^{20,28} are in complete agreement. Using techniques developed for structural studies on insulin²⁹ and on the corticotropins,³⁰ both groups of investigators obtained the following amino acid sequence for β -MSH:



It has been shown that neither β -aspartyl, nor γ -glutamyl, nor ϵ -amino linkages occur in β -MSH.²⁸ The two residues of aspartic acid that occur in the peptide occupy the *N*- and *C*-terminal positions in the chain. Hydrolysis of the hormone peptide with trypsin and chymotrypsin has been found to proceed in accordance with the known specificity of these proteolytic enzymes, with the one exception that the Lys.Asp linkage (positions 17-18) is resistant to cleavage by trypsin.

Although preparations of active MSH extracts from *bovine* pituitary have been reported,^{31,18} the isolation of a bovine MSH in pure form has only

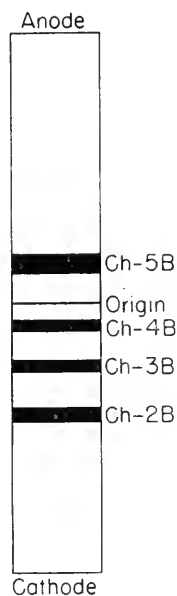


Fig. 2. Resolution of chymotryptic digest of beef β -MSH by paper electrophoresis in γ -collidine-acetic acid buffer of pH 7.0; 11 v/cm.

recently been achieved.^{22*} The complete amino acid sequence of the bovine MSH has now also been elucidated by various methods, as shown in Table 1. It may be seen that the structure proposed for the bovine hormone differs from that proposed for the pig β -MSH only at position 2, where a seryl

* Very recent investigations with I. I. Geschwind indicate that about 80% of the MSH component isolated from sheep posterior-intermediate pituitaries is identical with the bovine β -MSH in both composition and the amino acid sequence, whilst the remaining active material is identical with the pig hormone.

Table 1

THE SEQUENCE OF AMINO ACIDS IN BEEF MELANOTROPIN ('SERYL- β -MSH')

Method*	Substrate	Product†	Sequence
FDNB	H†		Asp.
LAP	H		Asp.Ser.
PITC	H		Asp.Ser.Gly.Pro.Tyr.Lys.Met.
Chymotrypsin	H	Ch-5B§	Asp(Ser,Gly,Pro)Tyr
Chymotrypsin	H	Ch-3B§	Lys(Met,Glu)His.Phe
LAP	Ch-3B		Lys.Met(Glu,His,Phe)
Chymotrypsin	H	Ch-2B§	Arg.Try
Chymotrypsin	H	Ch-4B§	Gly(Ser,Pro,Pro, Lys)Asp
LAP	Ch-4B		Gly
PITC	Ch-4B		Gly.Ser.Pro.Pro(Lys,Asp)
Carboxypeptidase	H		Asp

Complete Sequence: Asp.Ser.Gly.Pro.Tyr.Lys.Met.Glu.His.Phe.Arg.Try.Gly.Ser.Pro.Pro.Lys,Asp
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

* FDNB, fluorodinitrobenzene procedure of Sanger; LAP, leucine aminopeptidase of Smith; PITC, phenylisocarbonyl procedure of Edman.

† See Fig. 2. ‡ H = hormone.

§ The N-terminal amino acid was obtained by the FDNB procedure, and the C-terminal amino acid or sequence by the carboxypeptidase method.

residue replaces a glutamyl. This replacement of an uncharged residue (serine) by a charged one (glutamic acid) has not to our knowledge been encountered previously in connection with species variation among biologically active peptides. Because of this structural difference, the isoelectric point of the beef hormone (pH 7.0) is higher than that of pig β -MSH (pH 5.8). The latter was found to migrate more slowly on starch in a pyridine-acetate buffer of pH 4.93 than does the former (Fig. 3). Because of the presence in the bovine MSH of serine at position 2 instead of glutamic acid, it would be expected, on the basis of the known specificity of leucine aminopeptidase (LAP),³² that the rate of release of amino acids from this hormone by the enzyme would be considerably slower than it would be from the pig β -MSH. That this was indeed found to be the case can be seen from Fig. 4.

According to the *in vitro* frog skin assay,³³ beef MSH possesses a considerably lower potency than does pig β -MSH. It will be of interest to find out whether this difference in activity between the hormones of the two species is due mainly to the difference in the amino acid at position 2 of the sequence, or whether other factors, such as secondary structure, are involved. This question can be resolved only by the synthesis of the two melanocyte-stimulating peptides.* Table 2 summarizes some physical and chemical characteristics of the pig and beef hormones.

II. ADRENOCORTICOTROPIC HORMONE (ACTH, CORTICOTROPIN)

It has been well established that hypophysectomy, in almost all the species that have been studied, brings about a pronounced atrophy of the cortex of the adrenal glands, whereas the medulla is affected scarcely at all.^{2,34} On the other hand, the implantation of anterior pituitary tissue or the injection of pituitary extracts into normal animals have as their chief effect a marked hypertrophy of the adrenal cortex. Hence, the name adrenocorticotrophic hormone (ACTH) has been employed to designate the active principle in pituitary extracts that accomplishes repair of adrenal-cortical atrophy in hypophysectomized animals,³ and in 1951 the term corticotropin was adopted as the formal designation for preparations possessing ACTH activity.³⁵

From unhydrolyzed ACTH concentrates, three groups of investigators have succeeded independently in isolating the active peptide in pure form. In 1953, White,³⁶ in the Armour Laboratories, described the preparation of

* For consonance with the pattern of terminology employed for the other pituitary hormones, it was proposed that the melanocyte-stimulating hormones be called *melanotropins*.⁷ Following the precedent set by du Vigneaud⁴ for the vasopressins, the beef MSH could be differentiated from the pig β -MSH by designating the former seryl-melanotropin and the latter glutamyl-melanotropin. These designations could be abbreviated as seryl- β -MSH and glutamyl- β -MSH.

an apparently pure peptide from pig pituitary glands, which they designated Corticotropin-A. The following year Li and co-workers³⁷ at the University of California reported the isolation, from a *sheep* ACTH concentrate that had not been submitted to peptic digestion, of a peptide which behaves like a pure substance. The California group designated this unhydrolyzed ACTH obtained from sheep pituitaries α -corticotropin, since some of its chemical

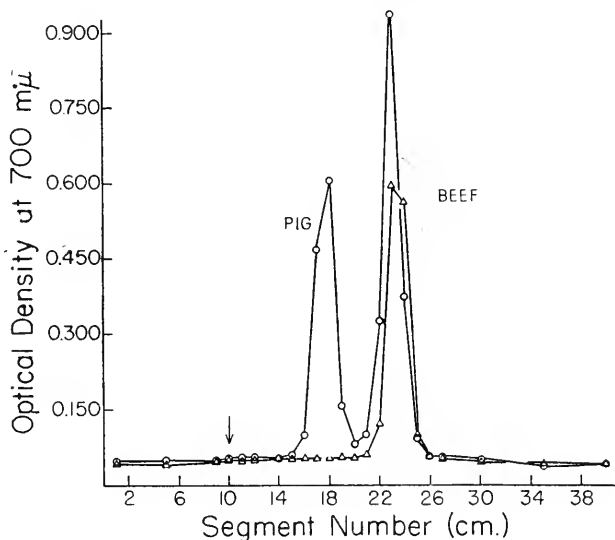


Fig. 3. Zone electrophoresis of beef β -MSH (Δ - Δ) and a mixture of beef β -MSH and pig β -MSH (o-o) on starch; pyridine-acetate buffer of pH 4.83 and $I=0.1$; 200 v for 24 hours.

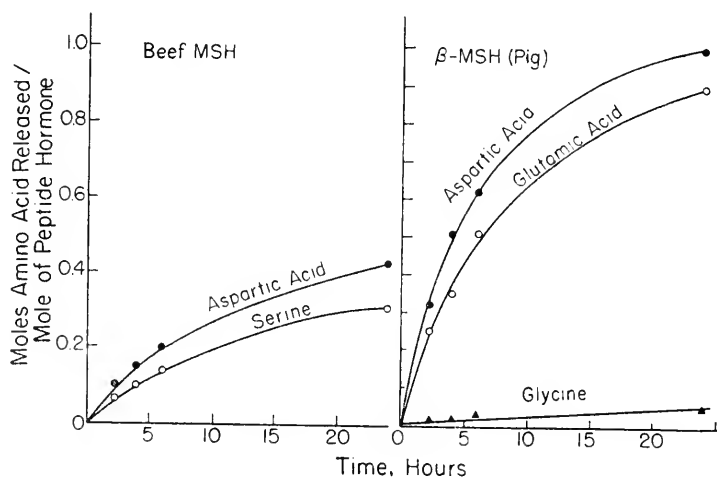


Fig. 4. Action of leucine aminopeptidase on beef and pig β -MSH; 0.6 mg. enzyme/1.4 mg. peptide; pH 8.1, 0.02 M Tris buffer containing 0.002 M-Mg⁺⁺ at 40°C.

Table 2

SOME PROPERTIES OF PIG β -MSH AND BEEF MELANOTROPIN

Properties	Pig β -MSH (Glutamyl- β -MSH)	Beef Melanotropin (Seryl- β -MSH)
Amino acid sequence	Asp. Glu. Gly. Pro. Tyr. Lys. Met. Glu. His. 1 2 3 4 5 6 7 8 9 Phe. Arg. Try. Gly. Ser. Pro. Pro. Lys. Asp. 10 11 12 13 14 15 16 17 18	Asp. Ser. Gly. Pro. Tyr. Lys. Met. Glu. His. 1 2 3 4 5 6 7 8 9 Phe. Arg. Try. Gly. Ser. Pro. Pro. Lys. Asp. 10 11 12 13 14 15 16 17 18
Amide-N	none	none
Molecular Weight	2177	2135
Isoelectric point	5.8	7.0
Isoionic pH	6.0-6.1	
Partition coefficient (2-butanol/0.5% trichloroacetic acid)	0.60	0.52
Specific Rotation	-97.50	----
Dissociation constants: (pK')		----
α -NH ₂	8.4	----
Phenolic	10.13	9.94
Melanocyte-stimulating activity (U/mg.)	2 \times 10 ⁷	0.2-0.4 \times 10 ⁷

and physicochemical characteristics differ from those described for corticotropin-A. In the same year, Bell³⁸ and his colleagues of the American Cyanamid Company announced the purification of β -corticotropin from ACTH concentrates of pig pituitaries. Since both the Armour and American Cyanamid groups employed unhydrolyzed pig ACTH concentrates as their starting material, it cannot be assumed that corticotropin-A and β -corticotropin are different peptides. The isolation of a polypeptide possessing adrenocorticotropic activity from bovine pituitaries in pure form has only recently been described.³⁹ Physicochemical properties of the bovine hormone are identical in every respect with that of the sheep α -corticotropin. The corticotropins from all three species are made up of 39 amino acids in a single polypeptide chain with serine as the *N*-terminal amino acid and phenylalanine at the C-terminus.

The complete amino acid sequences of the pig corticotropin-A and of the sheep α -corticotropin have been elucidated.^{40,41} Preliminary structural investigations on the beef corticotropin⁴² indicated that its amino acid sequence is identical with that of the sheep α -ACTH.⁴⁰ It may be seen in Table 3 that the complete sequence was derived from the identification of over 40 different peptide fragments obtained by chymotryptic, tryptic, peptic and partial acid hydrolysis, as well as by submitting the whole molecule to stepwise degradation by the phenylisothiocyanate (PITC) procedure of Edman and to kinetic investigations with leucine aminopeptidase (LAP) and carboxypeptidase, and from *N*- and *C*-terminal residue analyses by the fluorodinitrobenzene (FDNB) procedure of Sanger and hydrazinolysis according to the procedure of Akabori, respectively.

There are two amide groups in sheep α -corticotropin. The location of one of these groups in position 33 was recently established in the course of a kinetic investigation carried out with J. Léonis and D. Chung on the effect of chymotryptic digestion on the hormone peptide. The location of the other amide is yet to be determined, but there have been no indications that it occurs in position 30 as it does in the case of pig corticotropin-A (Cyanamid).⁴⁰ In this connection, it is of interest to note that pig corticotropin-A (Armour) possesses no amides at all.³⁶ Moreover, there are some variations, reported by three groups of investigators,^{40,41,43} with respect to the amino acids in positions 25-28, as shown in Table 4. It may be noted that the component amino acids in this segment are the same in all cases; only the order differs. Moreover, the only difference in amino acid composition between the peptide hormones isolated from sheep (beef) and pig glands appears to be one more serine and one less leucine in the former,^{30,39} owing to an . . . Ala.Ser . . . sequence in the sheep peptide in place of a . . . Leu.Ala . . . sequence in the pig peptide at amino acid positions 31-32.

Despite their differences in amino acid composition, and variations in certain portions of their amino acid sequence, both sheep α -corticotropin and pig corticotropin-A possess identical specific ACTH activity. Consequently,

it may be inferred that those portions of the molecule containing these differences are not essential for biological potency. Indeed, limited hydrolysis of both corticotropins with pepsin releases 11 amino acid residues from the C-terminus without loss of adrenal-stimulating activity.^{41,44,45,46} Further hydrolysis of this active core with dilute acid did not result in any loss of ACTH activity.³⁰ Recently, the synthesis of the first 20 amino acids from the N-terminus of the corticotropins has been achieved by Boissonnas

Table 4

VARIATIONS IN AMINO ACID SEQUENCES AMONG DIFFERENT PREPARATIONS OF ACTH

Preparation	Species	Residue No.								
		25	26	27	28	29	30	31	32	33
α -Corticotropin Corticotropin-A (Cyanamid) Corticotropin-A (Armour)	sheep	Ala.Gly.Glu.Asp.Asp.Glu						Ala.Ser.Glu.NH ₂		
	pig	Asp.Gly.Ala.Glu.Asp.Glu.NH ₂						Leu.Ala.Glu		
	pig	Gly.Ala.Glu.Asp.Asp.Glu						Leu.Ala.Glu		

et al.,⁴⁷ and the synthetic icosapeptide has been demonstrated to possess ACTH activity. Thus, both the sheep and pig corticotropins appear to have an identical active core that is responsible for their hormonal function.

The behavior of corticotropin-A and α -corticotropin in countercurrent distribution reveals marked differences in the partition coefficients of these peptides in various solvent systems. For example, the partition coefficient of corticotropin-A in a 2-butanol/0.2% TCA system is 1.82,³⁶ whereas that of α -corticotropin in the same system is 0.71.⁴⁸ In a solvent system made up of 6% acetic acid containing 3.5% NaCl/*n*-BuOH, corticotropin-A (Cyanamid) has a K value of 0.95,³⁸ whereas that of α -corticotropin in the same system is less than 0.1.⁴² It is unlikely that these differences are due to variation in amide content among the various corticotropin preparations, since, even though there is a difference of 2 amide groups between two preparations of α -corticotropin, they cannot be differentiated by their behavior in countercurrent distribution in a 2-butanol-aqueous TCA system.⁴⁸

There are also some indications that corticotropin-A and α -corticotropin may be distinguished from each other according to their behavior in enzymatic hydrolysis. The peptide fragment Lys.Arg.Arg.Pro.Val.Lys occurs in the tryptic digest of corticotropin-A (Cyanamid)⁴⁹ but it is not encountered in similar digests of α -corticotropin.⁴¹ Moreover, the sheep hormone appears to be more labile toward pepsin than is the pig hormone. It is not improbable that two polypeptides possessing similar amino acid sequences may differ in the spatial arrangement (helices) of the residues which compose it,

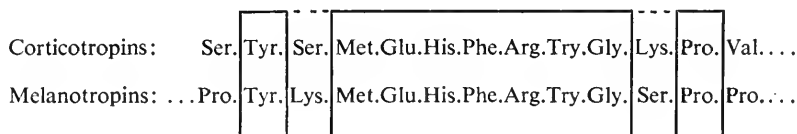
and yet may have identical biological activity. Table 5 summarizes some properties of pig corticotropin-A and sheep α -corticotropin.

Table 5

SOME PROPERTIES OF CORTICOTROPINS

	α -Corticotropin (Sheep)	Corticotropin-A (Pig)	
		Armour	Cyanamid
Molecular weight			
Analytical data	4541	3500	4567
Ultracentrifuge	5363*	—	4390†
Isoelectric point, pH	6.6	7.8	—
Sedimentation constant, S_{20}	0.73	—	0.54
Diffusion constant, $D_{20} \times 10^7$	12.2	—	11.3
Partition coefficients			
2-butanol/0.2% TCA:	0.72	1.82	—
n-butanol/6% HOAc-3.5% NaCl:	0	—	0.95
ACTH activity (units/mg.)	150	125	80-100
MSH activity (units/mg.)	1×10^5	1.7×10^5	‡

A comparison of the amino acid sequences of both β -MSH and the bovine MSH with that of the corticotropins reveals that there is a central heptapeptide sequence that is common to both molecules:



It may also be noted that were it not for a transposition of the order of the serines and lysines on either side of the heptapeptide core, the area of identity between the two hormones would extend to 11 amino acids. It appears likely that the presence of this heptapeptide sequence in the corticotropins explains their intrinsic melanocyte-stimulating activity.^{30§} Thus, it can be postulated that it is by virtue of an arrangement of a different sequence of amino acids on each side of the heptapeptide core that the molecule acquires adrenal-stimulating as well as melanocyte-stimulating activity.

* As trichloroacetate.

† As acetate.

‡ Minimum effective dose, 3-5 γ /100 g. frog.

§ Melanocyte stimulation is not the only function of α -ACTH in the absence of adrenal glands. Recent studies of Menkin⁵⁰ have shown that the hormone peptide exerts a direct local effect of repressing in an inflamed area the increased capillary permeability in adrenalectomized animals. In addition, it has been shown that the adipokinetic activity of α -ACTH can be demonstrated in adrenalectomized mice maintained with corticoid therapy.⁵¹

III. LACTOGENIC HORMONES (PROLACTINS, LUTEOTROPINS)

The first indication that a hormone of the anterior pituitary might effect lactation in mammals was given by the experiments of Stricker and Grueter⁵² in 1928. Several years later growth, to a macroscopically observable extent, of the crop glands of the pigeon after the administration of pituitary extract was reported by Riddle and Braucher;⁵³ this effect in the bird has been shown by subsequent work to be due to the same factor that elicits lactation in mammals. A number of terms have since been proposed for this factor, including lactogenic hormone, prolactin, galactin, mammatropin and lactogen. After the discovery that the hormone is capable of maintaining luteal function in hypophysectomized rats,⁵⁴ the term luteotropin was also introduced.

Highly purified preparations of prolactin have been obtained from both beef and sheep pituitary glands.^{3,55} Recently a new simplified procedure has

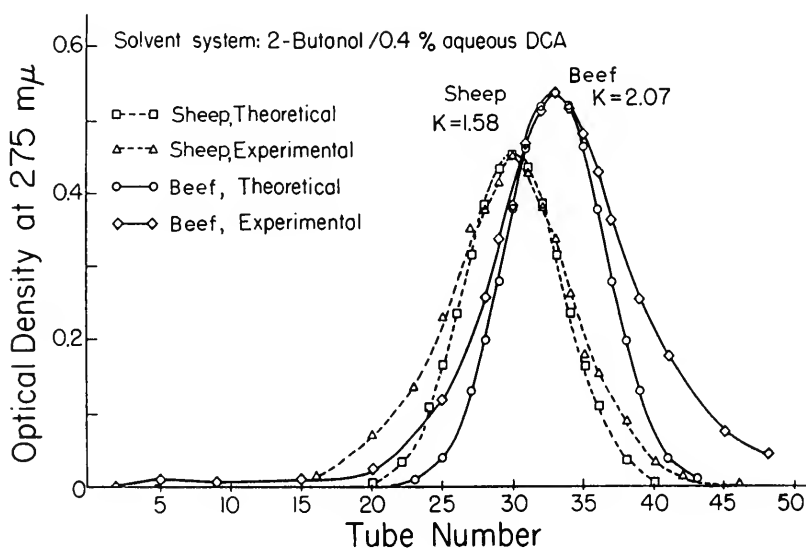


Fig. 5. Countercurrent distribution of sheep and beef prolactin at 20°C in the 2-butanol-0.4% aqueous dichloroacetic acid system.

been described⁵⁶ for the isolation of sheep prolactin; the same method has also been employed for the preparation of prolactin in a highly purified form from bovine pituitaries.⁵⁷ Earlier observations that *sheep* prolactin can be differentiated from the *beef* hormone by differences in solubility behavior and in tyrosine content,³ have now been confirmed.⁵⁷ It may be seen in Fig. 5 that the partition coefficient of beef prolactin is higher than that of the sheep hormone in the 2-butanol-0.4% dichloroacetic acid system.

This appears to be consistent with the earlier findings of Fleischer⁵⁸ that in alcoholic solution the bovine hormone is more soluble than the ovine.

Table 6 presents a summary of the known physical and chemical properties of prolactin from beef and sheep pituitary glands. The tyrosine content as determined by the spectrophotometric method is definitely higher

Table 6

SOME PHYSICAL AND CHEMICAL PROPERTIES
OF PROLACTIN FROM SHEEP AND BEEF
PITUITARY GLANDS

Physical and Chemical Properties	Sheep	Beef
Molecular weight		
Sedimentation-diffusion	24,200	
Osmotic Pressure	26,500	26,000
Analytical data	24,100	—
Diffusion coefficient* (D_{20})	8.44×10^{-7}	—
Sedimentation constant* ($s_{20, w}$)	2.19	—
Partial Specific Volume (V_{20})	0.739	—
Isoelectric point† (pH)	5.73	5.73
Specific Rotation	-40.5°	-40.5°
Partition Coefficient		
(2-butanol/0.35% aqueous tri-chloroacetic acid)	1.58	2.07
Tyrosine, % (by wt.)	5.26	6.62
Tryptophan, % (by wt.)	1.69	1.75
Cystine, residue/mole	3	3
<i>N</i> -terminal amino acid	Threonine	Threonine
<i>C</i> -terminal amino acid	none	none

for the beef hormone than for the sheep hormone. Both hormones have a molecular weight of 26,000 and an isoelectric point at pH 5.73; they both consist of a single peptide chain with 211 amino acid residues⁵⁹ and the sequence Thr.Pro.Val.Thr.Pro... at the *N*-terminus.⁶⁰ Studies⁷ in which prolactic preparations of both species were subjected to reaction with carboxypeptidase and to hydrazinolysis, have given no indication that either hormone possesses a *C*-terminal residue at all. Prolactin that has undergone oxidation with performic acid has been found to behave as a single component in electrophoresis, sedimentation, and chromatography, as well as according to the results of *N*-terminal residue analysis.⁶¹ From these data, as well as from the absence of evidence of any *C*-terminal residue, it was postulated tentatively that the prolactin molecule has a cyclic configuration at the *C*-terminus.⁷ This tentative structure has now been verified by the

* Determined in 0.1 M-NaHCO₃ solutions.

† Estimated in acetate buffers of 0.05 ionic strength.

identification of cysteine as the C-terminal residue after reductive cleavage of the disulfide bridges* of the hormone protein.⁵⁷

In earlier investigations,⁶³ difficulties were encountered in attempting to cleave the disulfide bonds in the hormone by reductive procedures. Even in the presence of 6M urea at pH 9.4, the disulfide bridges in prolactin are only partially reduced.⁵⁷ For the complete reductive cleavage of the disulfide bonds, the following procedure has been shown to be effective: 100 mg. of sheep or beef prolactin were dissolved in 10 ml. of 10M urea containing 3% mercaptoethanol, and 4 ml. of conc. NH_4OH were added to bring the pH of the solution to 9.8. The whole solution was placed in a cellophane bag and dialyzed against 125 ml. of the reducing mixture (i.e., 10M urea, 3% mercaptoethanol, pH 9.8) in a glass-stoppered cylinder. After 3 days, the contents of the bag were placed in a beaker containing 1 g. of iodoacetamide; 10 minutes later, the whole solution was dialyzed against cold running tap water. After 2 hours of dialysis, 200 mg. of iodoacetamide were added to the bag to insure the complete alkylation of the $-\text{SH}$ groups. The reduced hormone (prolactin $-\text{SCH}_2\text{CONH}_2$) solution was thoroughly dialyzed and lyophilized. If prolactin $-\text{SH}$ was desired, the reduced solution was not reacted with iodoacetamide, but instead was dialyzed against 3% mercaptoethanol for 5 days, lyophilized and stored *in vacuo*.

In electrophoretic experiments on starch, prolactin $-\text{SCH}_2\text{CONH}_2$ migrated essentially as a single component with a mobility higher than that of

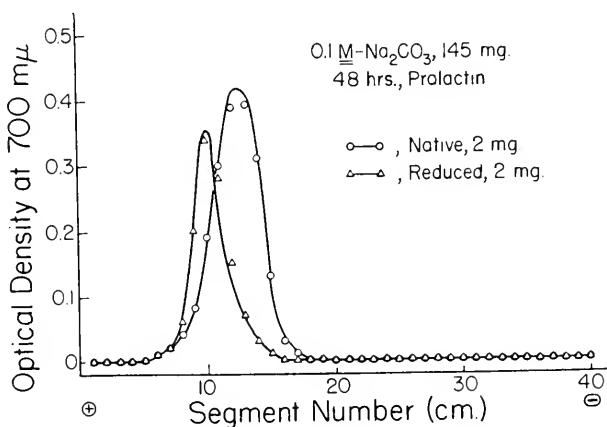


Fig. 6. Zone electrophoresis on starch of prolactin $-\text{S}-\text{CH}_2\text{CONH}_2$; 0.1 M Na_2CO_3 , 145 v, 48 hours.

the native hormone in 0.1M- Na_2CO_3 (Fig. 6). In no case was there separation of reduced prolactins into more than one component after the reductive cleavage of the disulfide bonds in the hormone protein. It was further

* It should be recalled that both sheep and beef lactogenic hormones have 3 residues of cystine per mole of hormone protein.^{62,69}

noted that the amount of material originally applied to the starch trough could be accounted for by the protein recovered from the peak.

Biological assays of prolactin $-\text{SCH}_2\text{CONH}_2$ or prolactin $-\text{SH}$ according to either its crop sac-stimulating or its luteotropic activity indicate that reductive cleavage of the disulfide bridges causes the complete destruction of the hormonal activity, indicating that the integrity of the disulfide bridges is essential for the biological function of lactogenic hormones.

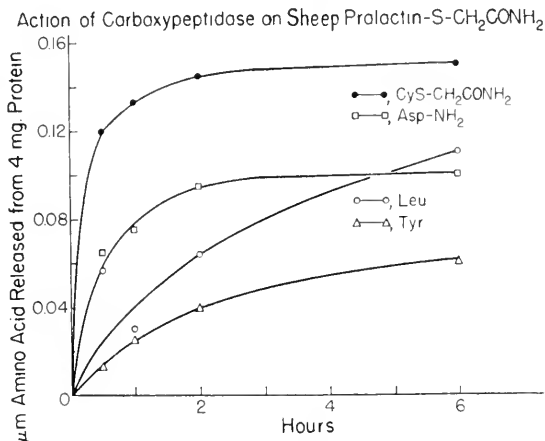


Fig. 7. Action of carboxypeptidase on prolactin $-\text{S}-\text{CH}_2\text{CONH}_2$; enzyme-substrate ratio, 1 : 35; pH 8.5 at 40°C.

When either prolactin $-\text{SCH}_2\text{CONH}_2$ or prolactin $-\text{SH}$ was allowed to react with carboxypeptidase, it was found that $\text{CyS}-\text{CH}_2\text{CONH}_2$ (or CySH), asparagine, and leucine are released successively (Fig. 7). The liberation of $\text{CyS}-\text{CH}_2\text{CONH}_2$ was nearly complete in 4 hours under the conditions of digestion used. From these studies, it was concluded that prolactin $-\text{SCH}_2\text{CONH}_2$ possesses the sequence $\dots\text{Leu.AspNH}_2.\text{CyS}-\text{CH}_2\text{CONH}_2$ at the C-terminus. That this C-terminal sequence is derived from reductive cleavage of a disulfide bridge which forms a loop at the C-terminus of the untreated prolactin molecule is further supported by the identification⁵⁷ of C-terminal cysteic acid in performic-acid-oxidized prolactins, by means of the hydrazinolysis method of Akabori. Hence, it may be concluded that sheep and beef prolactins consist of a single polypeptide chain with the sequence Thr.Pro.Val.Thr.Pro at the N-terminus and with an intrachain disulfide loop at the C-terminus (Fig. 8). In this connection, it is of interest to recall that both insulin²⁹ and the posterior pituitary hormones⁴ have similar intrachain disulfide loops, involving only 20-atom rings. There remains to be investigated the size of the loop and the location of the remaining two disulfide bridges in prolactin molecules. In the case of insulin, the only differences between pig, sheep and beef hormones are found among the amino acid residues in the ring.²⁹ So far no differences have been found in

the amino acids which are released by carboxypeptidase from the reduced form of either beef or sheep prolactin.⁵⁷

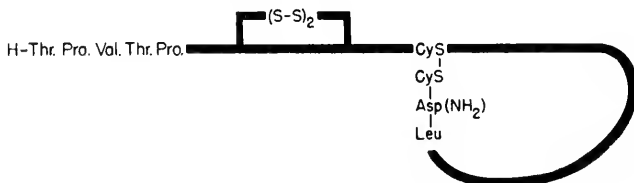


Fig. 8. Proposed gross structure for the prolactin molecule. (S—S)₂ indicates schematically the existence of two loops containing disulfide bridges. This model is somewhat speculative and structures other than that shown here are possible.

IV. GROWTH HORMONES (SOMATOTROPINS)

The relation of the pituitary gland to growth was first given experimental demonstration by Cushing and his collaborators⁶⁴ in their classical experiments on the hypophysectomy of dogs in 1910, and by 1930 investigations with various species, including fish, amphibia, reptiles, birds and mammals, had left little doubt that the anterior pituitary contains a hormone which initiates resumption of growth in hypophysectomized animals. During the period between 1944 and 1948, methods for the isolation of growth hormone in a highly purified form from the anterior lobes of *bovine* pituitaries were described by two groups of investigators.^{65,66} During the past few years, the many attempts^{67,68,69,70} that have been made to purify further the products isolated by the published procedures^{65,66} have not met with success. The failure of these attempts has lent support to the belief that the protein is the hormone.^{30,71}

Bovine Growth Hormone. The bovine somatotropin is a protein with a molecular weight of 45,000 and an isoelectric point of pH 6.85; the protein nitrogen can be accounted for completely by its amino acid and amide content.³⁰ The hormone protein contains a total of 396 amino acids; structurally, it is a branched polypeptide chain with two *N*-terminal residues (phenylalanine and alanine) and only one *C*-terminal residue (phenylalanine),* as shown in Fig. 9.

When the bovine growth hormone from which two residues of phenylalanine had been removed by the action of carboxypeptidase was assayed for biological activity, the *C*-terminal phenylalanine was found not to be essential for the biological function of the hormone.⁷² Chymotryptic digestion, performed at a temperature of 25°, in a buffer of pH 9.5 and with an

* Recent work with H. Papkoff has shown that growth hormone isolated from whole sheep pituitary glands by the same procedure as that described for beef glands has a similar structure to that of the bovine somatotropin, with two *N*-terminal residues (phenylalanine and alanine) and one *C*-terminal residue (phenylalanine). Its molecular weight is 47,800 and its isoelectric point is at pH 6.8.

enzyme/hormone ratio of 1/300, did not result in inactivation if hydrolysis were allowed to proceed up to about 25%; however, the biological potency was abolished by longer digestion. Thus, it should be noted that if an active digest is desired, the extent of hydrolysis must not be permitted to exceed 30%. The non-protein nitrogen can be separated from the whole digest either by dialysis or by treatment with a 5% solution of trichloroacetic acid. It was demonstrated by biological assay that the growth-promoting activity

Proposed Structure of the Beef Growth-Hormone Molecule

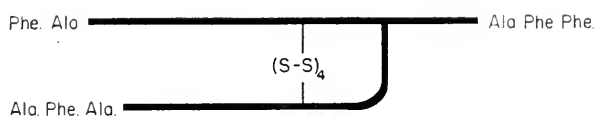


Fig. 9. Proposed gross structure for the bovine somatotropin molecule. This model is somewhat speculative and structures other than that shown here are possible.

resides in the remaining components (the core), which are non-dialyzable and insoluble in the trichloroacetic acid solution.⁷³

Boundary electrophoresis of the core has not revealed any component that behaves like the untreated bovine hormone. *N*-Terminal analysis of the core reveals a number of new residues (threonine, serine, tyrosine, lysine, etc.), in addition to phenylalanine and alanine; if we assume that the terminal phenylalanine and alanine residues have come from the undigested bovine hormone, it can be estimated that the native hormone in the core amounts to less than 20%, a percentage certainly not sufficient to account for the biological activity of the core. From these results, it may be concluded that the activity does not depend upon the integrity of the bovine protein, and that the growth-promoting activity resides in a center (or centers) of activity in the molecule.

Human and Monkey Growth Hormones. In recent years, many attempts have been made to apply growth hormone clinically to humans as an effective therapeutic agent, but these attempts have met with disappointment.⁷⁴ One of the obvious explanations for this failure is that the bovine growth hormone is chemically different from the hormone derived from human glands. Indeed, it has been shown that the growth hormone prepared from fish pituitaries is active in fish,⁷⁵ but not in rats;⁷⁶ and that monkeys are not responsive to bovine growth hormone, whereas they are responsive to growth hormone prepared from pituitaries of their own species.^{77,78} Our recent investigations with human and monkey pituitary glands indicate that the human and monkey growth hormones are similar to each other in structure and properties, but that they both differ completely from the hormone molecule isolated from the pituitaries of cattle.^{79,80}

Table 7

A PROTOCOL FOR THE ISOLATION OF HUMAN GROWTH HORMONE

Fraction	Weight (mg.)	Procedure	Growth-promoting activity* (micra)	N-terminal residues†				
				Phe	Glu-Asp	Ala	Tyr	Val
A	1000	Lyophilized whole pituitaries	—	—	—	—	—	—
AS	530	Extracted with pH 10.5 Ca(OH) ₂ solution; fractionated with (NH ₄) ₂ SO ₄	206	0.39	0.16	0.16	0.17	0.19
B	210	1.9M-(NH ₄) ₂ SO ₄ precipitate, dialyzed and lyophilized	213	0.53	0.14	0.09	0.10	0.14
—	—	Extracted with 0.45M-(NH ₄) ₂ SO ₄ in pH 5.1 phosphate buffer	—	—	—	—	—	—
C	140	Soluble fraction submitted to chromatography on IRC-50 resin column	221	0.65	0.12	0	0.11	0.12
D	75	Active peak; recovered by (NH ₄) ₂ SO ₄ precipitation as B	235	0.87	0.13	0	0	0
—	—	pH and ethanol fractionation	—	—	—	—	—	—
E (Growth Hormone)	40	5-20% (v/v) ethanol precipitate; dissolved and lyophilized	248	1.00	0	0	0	0

* Assayed at a total dose of 20 μ g. in 4 days by the Tibia Test; 6 hypophysectomized rats in each assay; the tibia width for controls is 160 μ .

† DNP-amino acid/total DNP-amino acids.

Table 7 outlines the procedure for the isolation of growth hormone from either monkey or human pituitary glands.^{79,80} The primate somatotropins have been assayed for their growth-promoting activity according to the stimulation of the width of uncalcified tibial cartilage in hypophysectomized rats, and their specific activities were found to be comparable to that of the bovine hormone.^{79,80} As is to be expected, human and monkey growth hormones cause a marked retention of nitrogen and phosphorous in patients with mild hypopituitarism and in hypophysectomized male subjects.^{82,83}

From sedimentation and diffusion data and from computed values for partial specific volume,⁷⁹ the molecular weights for human and monkey somatotropins were calculated to be 27,100 and 25,400 respectively. These values differ greatly from the molecular weight of the bovine hormone.⁸⁰ Recently, Ehrenberg and Heijkenskjöld⁸⁴ reported similar differences between the molecular weights of human and bovine growth hormones.

The electrophoretic mobility of both human and monkey somatotropins as a function of pH was determined in buffers of 0.1 ionic strength, and it was found that the isoelectric point of the primate hormones is more acidic than that of the bovine hormone.⁸⁰ Table 8 summarizes some of the physicochemical properties of somatotropins isolated from monkey and human pituitary glands. It may be seen that the monkey hormone contains four residues of cystine, whereas the human contains only two. In this respect, the monkey hormone is more like the bovine somatotropin.

Table 8

PHYSICOCHEMICAL CHARACTERISTICS OF
GROWTH HORMONE ISOLATED FROM HUMAN
AND MONKEY PITUITARY GLANDS

Physicochemical characteristics	Monkey	Human
Sedimentation constant ($s_{20, w}$)	1.88	2.47
Diffusion constant ($D_{20} \times 10^7$)	7.20	8.88
Molecular weight	25,400	27,100
Isoelectric point (pH)	5.5	4.9
Tyrosine (Residues/mole)	10	8
pK _i *	10.2	9.8
Tryptophan (Residues/mole)	1	1
Cystine (Residues/mole)	4	2
Methionine (Residues/mole)	6	4
Total amino acid residues	241	245

From investigations of the *N*- and *C*-termini by both enzymic and chemical methods,^{80,85} it was postulated that both human and monkey somatotropins are composed of a single polypeptide chain with phenylalanine as

* pK of phenolic group.

both C- and N-terminal residues. Since the human and monkey hormones have two and four cystine residues⁷⁹ respectively, it might be anticipated that all —S—S— bridges would be located intramolecularly along the polypeptide chain. This was found to be the case when performic acid-oxidized somatotropins were submitted to electrophoresis and ultracentrifugation. Hence, it was proposed⁸⁰ that human and monkey growth hormone have the generalized gross structure shown in Fig. 10.

Proposed Structure of the Human and Monkey Growth-Hormone Molecule

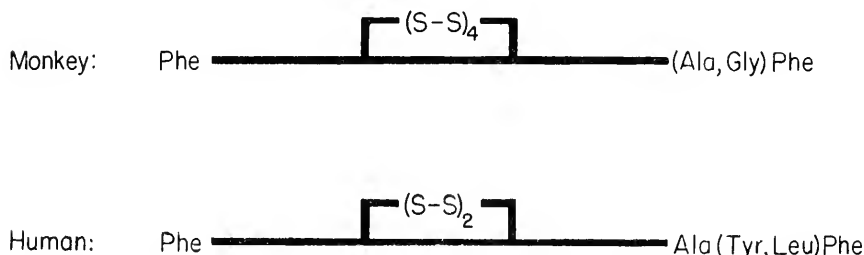


Fig. 10. Proposed gross structure for the human and monkey somatotropins. This model is somewhat speculative and structures other than that shown here are possible.

It has been demonstrated⁸⁰ that the removal of C-terminal phenylalanine from either human or monkey somatotropin by the action of carboxypeptidase does not result in any loss of the growth-promoting activity. It is also possible to show that limited digestion of both somatotropins with chymotrypsin does not lead to inactivation.⁸⁰ Human somatotropin was not inactivated by the action of chymotrypsin only if hydrolysis did not exceed 10%; longer digestion did diminish the biological activity. On the other hand, the monkey hormone could be digested to 19% without any change in the growth-promoting potency; as expected, digestion with the enzyme to any further extent was found to result in the loss of the biological activity. As in the case of bovine somatotropin,³⁰ these results seem to indicate that the integrity of the intact protein is not essential for the activity of the primate growth hormone.

Whale Somatotropin. There has been only one report in the literature in connection with growth-promoting activity of extracts of whale pituitaries. In 1934, Valsö⁸⁶ reported that the content of growth-promoting activity in the anterior pituitary of blue whale approached that found in bovine pituitaries. In recent investigations with H. Papkoff, growth hormone from anterior lobes of whale pituitary glands* has been successfully isolated in

* The anterior pituitaries ranged in weight from 6-40 g., the average being about 21 g. The whales were all of the humpback variety, which are commonly found off the California coast and which attain an average length at maturity of around 40-50 feet. The average interval post mortem before removal of the pituitary was about 20 hours. Once removed, the gland was frozen as rapidly as possible and stored at -15° C. until used.

a highly purified state; this whale somatotropin was demonstrated to be structurally more similar to primate hormones than to sheep or beef somatotropin.

The procedure employed for the purification and isolation of whale somatotropin is essentially the same as those described for the bovine⁸¹ and the primate^{79,80} hormones. Table 9 presents a summary of the average yield

Table 9

YIELD AND ACTIVITY OF FRACTIONS OBTAINED FROM THE ISOLATION PROCEDURE OF WHALE GROWTH HORMONE

Fraction	Amount	Activity*		
		Total dose	No. of rats	Tibia width
	(g.)	(mg.)		(μ)
Wet weight of anterior lobes	95	—	—	—
Lyophilized tissue	16	0.40	4	228 \pm 5
0.2-0.4 saturated (NH ₄) ₂ SO ₄	3	0.20	9	232 \pm 3
IRC-50 column chromatography	1	0.04	8	219 \pm 4
5-20% ethanol (somatotropin)	0.3	0.04	15	238 \pm 3

and activity of fractions obtained from each step of the isolation procedure. The pattern obtained when the somatotropin concentrate was submitted to ion exchange chromatography can be seen in Fig. 11; the activity was found in tubes 140 to 170. The purified product behaves as a homogeneous substance when submitted to chromatographic analysis on the same resin (Fig. 11). In addition, no evidence of heterogeneity was observed in the ultracentrifuge or in free boundary electrophoresis. Terminal group analyses indicate that whale somatotropin consists of *one* *N*-terminal phenylalanine and *one* *C*-terminal phenylalanine, in amounts consistent with a molecular weight of 39,900. The isoelectric point has been determined in buffers of 0.1 ionic strength and found to be located at pH 6.2. As is the case in the other somatotropins, the integrity of the whole protein molecule is not essential for the growth-promoting activity of whale somatotropin (Table 10).

Highly purified preparations of growth hormones from five different mammalian species are now available for study and comparison. Table 11 summarizes some physicochemical properties of these five somatotropins. The molecular weights range from 25,000 for monkey growth hormone to 48,000 for the ovine hormone. Beef, sheep and whale growth hormones have isoelectric points at pH 6.85, 6.8 and 6.2 respectively, while human and monkey somatotropins are more acidic, with isoelectric points at pH 4.9 and 5.5 respectively. Both beef and sheep growth hormones are composed

* Total dose in 4 days; tibia width in mean \pm S.E.; tibia width for the controls, 160 μ .

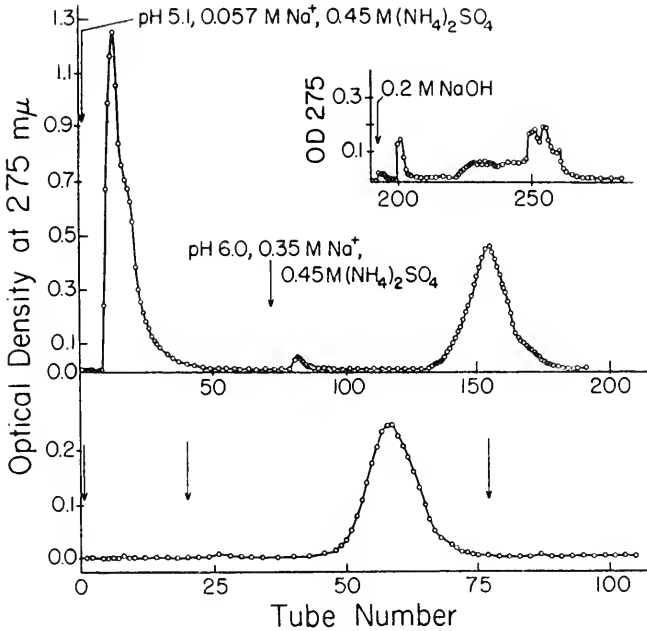


Fig. 11. Upper: Chromatography on a cation exchanger, Amberlite IRC-50 resin, of whale somatotropin concentrate (0.2-0.4 SAS fraction); 3 cm. diameter column containing 220 ml. pH 5.1 buffer applied to the column; 15.5 ml./tube.

Lower: Chromatography on a cation exchanger, Amberlite IRC-50 resin, of highly purified whale somatotropin; 0.9 cm. diameter column containing 20 ml. resin; 15 mg. protein in 10 ml. pH 5.1 buffer applied to the column; arrows indicate same sequence of buffer as upper diagram; 3.0 ml./tube.

Table 10

ACTION OF CHYMOTRYPSIN ON
WHALE GROWTH HORMONE

Time of digestion*	Nonprotein nitrogen	Bioassay†	
		No. of rats	Tibia width
(min.)	(%)		
0	0	7	230 ± 3‡
80	26	9	223 ± 4
150	38	9	214 ± 6
240	48	5	184 ± 5

* pH 9.5, enzyme/substrate, 1/300 (w/w), 25°C.

† A total dose of 40 micrograms in 4 days.

‡ Mean ± standard error; tibia width for the controls, 160 μ.

of a branched polypeptide chain having two *N*-terminal residues (phenylalanine and alanine) and only one *C*-terminal residue (phenylalanine), whereas whale, monkey and human somatotropins are made up of only one polypeptide chain with phenylalanine as both the *N*-terminal and the *C*-terminal residues. As is evident from Table 12, somatotropins isolated from pituitary glands of various species can easily be differentiated by the *N*- and *C*-terminal

Table 11

SOME PHYSICO-CHEMICAL PROPERTIES OF VARIOUS SOMATOTROPINS

Physicochemical characteristics*	Beef	Sheep	Whale	Monkey	Human
$s_{20, w}$	3.19	2.76	2.84	1.88	2.47
$D_{20, w} \times 10^7$	7.23	5.25	6.56	7.20	8.88
\bar{V}	0.76	0.733	0.737	0.726	0.732
Molecular wt.	45,000	47,800	39,900	25,400	27,100
f/f_0	1.31	1.68	1.45	1.57	1.23
P_I	6.85	6.8	6.2	5.5	4.9
Cystine	4	5	3	4	2
<i>N</i> -terminal Residue	Phe,Ala	Phe,Ala	Phe	Phe	Phe
<i>C</i> -terminal Residue	Phe	Phe	Phe	Phe	Phe

* $s_{20, w}$ in *S*; $D_{20, w}$ in $\text{cm}^2/\text{sec.}$; \bar{V} in cc./g. ; f/f_0 , dissymmetry constant; P_I , isoelectric point; cystine in residues per mole.

Table 12

N- AND *C*-TERMINAL SEQUENCES OF SOMATOTROPINS FROM VARIOUS SPECIES

Somatotropins	Terminal sequences	
	Amino*	Carboxyl†
Beef‡	Phe.Ala.Thr.... Ala.Phe.Ala....Ala.Phe.Phe
Sheep	Phe.... Ala....Tyr.Ala.Phe
Whale	Phe(?)Lys....Leu.Ala.Phe
Monkey	Phe(?)Thr....Ala.Gly.Phe
Human	Phe.Ser.Thr....Tyr.Leu.Phe

* By the phenylisothiocarbonyl procedure of Edman; the quantitative estimates of the *N*-terminal residue were obtained by Sanger's fluorodinitrobenzene procedure.

† As revealed by the action of carboxypeptidase; the *C*-terminal residue was verified by the hydrazinolysis procedure of Akabori.

‡ *N*-terminal amino acid sequences were established by identification of DNP-peptides in partial acid hydrolysates of the bovine DNP-somatotropin.

sequences in the polypeptide chains. Moreover, recent investigations with T. Hayashida showed that whale, monkey and human somatotropins are immunologically distinct from the bovine hormone (Table 13). On the other hand, bovine somatotropin cannot be differentiated from the ovine hormone by the precipitin test.

Table 13

PRECIPITIN TEST OF VARIOUS GROWTH HORMONES WITH RABBIT ANTI-SERUM* TO BOVINE GROWTH HORMONE

Somatotropin	Antigen, mg.						
	0.5	0.125	0.032	0.008	0.004	0.002	0.001
Beef	+	+	+	+	+	+	0
Sheep	+	+	+	+	+	0	0
Whale	0	0	0	0	0	0	0
Monkey	0	0	0	0	0	0	0
Human	0	0	0	0	0	0	0

All somatotropins could be hydrolyzed partially with chymotrypsin without inactivation. The extent to which hydrolysis of these five somatotropins could be carried out with no loss of growth-promoting activity decreases in the following order: Bovine, sheep, whale (25%) > monkey (20%) > human (10%). Studies on the rate of chymotryptic digestion of these hormones of the various species indicate that human somatotropin is the most resistant to the enzyme action whereas bovine hormone is the most susceptible.⁸⁰ These differences in kinetic behavior in the course of enzymic hydrolysis may also be taken as an indication of dissimilarity in composition and structure. These observations provide an additional instance of the fact that somatotropin molecules from various species differ completely from one another with respect to chemical and structural characteristics. The hormones from all species are equally active, however, in promoting growth in rats, whereas only monkey and human hormones are effective in man.† Does this mean that the rat is capable of degrading the bovine or ovine hormone to the active core, whereas humans do not possess this ability? If this were true, one might then assume that all somatotropins possess an identical core

* Anti-serum to bovine growth hormone was obtained by immunizing rabbits as follows: Rabbits were injected subcutaneously with 5 mg. antigen in 1 ml. adjuvant (Bayol-Arlacel) and 5 mg. antigen in 1 ml. saline once weekly for 3 weeks. After no injections for 3 weeks, the same injection schedule was followed for another 3 weeks. After no injections for another 3 weeks, the rabbits were bled by cardiac puncture. For precipitin tests, 0.1 ml. anti-serum was employed together with 0.5 ml. saline containing the antigen.

† Whether or not whale somatotropin exercises growth-promoting activity in man remains to be investigated. From structural considerations, the whale hormone might be expected to be active in human subjects.

representing the structural center responsible for the growth-promoting activity.

Acknowledgements. Experimental work mentioned herein was aided in part by grants from the National Institutes of Health, U.S. Public Health Service (Nos. G-2907 and RG-4097), the American Cancer Society upon recommendation of the Committee on Growth of the National Research Council, and the Albert and Mary Lasker Foundation.

REFERENCES

1. P. E. SMITH, *J. Am. Med. Assoc.*, **88**, 158 (1927).
2. P. E. SMITH, *Am. J. Anat.*, **45**, 205 (1930).
3. C. H. LI and H. M. EVANS, *The Hormones*, **1**, 665 (1948).
4. V. DU VIGNEAUD, *Harvey Lectures*, **L**, 1 (1956).
5. F. G. SULMAN, *Nature*, **169**, 588 (1952).
6. S. JOHNSON and B. HÖGBERG, *Nature*, **169**, 286 (1952).
7. C. H. LI, *Adv. Prot. Chem.*, **12**; in press.
8. P. E. SMITH, *Science*, **44**, 280 (1916).
9. B. M. ALLEN, *Science*, **44**, 755 (1916).
10. P. E. SMITH, *Proc. Soc. Exptl. Biol. and Med.*, **16**, 74 (1919).
11. P. E. SMITH and I. P. SMITH, *Anat. Rec.*, **25**, 150 (1923).
12. B. M. ALLEN, *Science*, **52**, 274 (1920).
13. W. W. SWINGLE, *J. Exptl. Zool.*, **34**, 119 (1921).
14. H. WARING and F. W. LANDGREBE, *The Hormones*, **2**, 427 (1950).
15. T. H. LEE and A. B. LERNER, *J. Biol. Chem.*, **221**, 943 (1956).
16. A. B. LERNER and T. H. LEE, *J. Am. Chem. Soc.*, **77**, 1066 (1955).
17. B. J. BENFEY and J. L. PURVIS, *J. Am. Chem. Soc.*, **77**, 5167 (1955).
18. B. J. BENFEY and J. L. PURVIS, *Biochem. J.*, **62**, 588 (1956).
19. J. PORATH, P. ROOS, F. W. LANDGREBE and G. M. MITCHELL, *Biochim. et Biophys. Acta*, **17**, 598 (1955).
20. I. I. GESCHWIND, C. H. LI and L. BARNAFI, *J. Am. Chem. Soc.*, **78**, 4494 (1956).
21. I. I. GESCHWIND and C. H. LI, *J. Am. Chem. Soc.*, **79**, 615 (1957).
22. I. I. GESCHWIND, C. H. LI and L. BARNAFI, *J. Am. Chem. Soc.*, **79**, 1003 (1957).
23. O. KAMM, T. B. ALDRICH, I. W. GROTE, L. W. ROWE and E. P. BUGBEE, *J. Am. Chem. Soc.*, **50**, 573 (1928).
24. R. W. PAYNE, M. S. RABEN and E. B. ASTWOOD, *J. Biol. Chem.*, **187**, 719 (1950).
25. E. B. ASTWOOD, M. S. RABEN, R. W. PAYNE and A. B. GRADY, *J. Am. Chem. Soc.*, **73**, 2969 (1951).
26. F. W. LANDGREBE and G. M. MITCHELL, *Quart. J. Exptl. Physiol.*, **39**, 11 (1954).
27. J. I. HARRIS and P. ROOS, *Nature*, **178**, 90 (1956).
28. I. I. GESCHWIND, C. H. LI and L. BARNAFI, *J. Am. Chem. Soc.*, **79**, 620 (1957).
29. F. SANGER, *Currents in Biochemical Research* (David E. Green, ed.), Interscience Publishers, 1956, p. 434.
30. C. H. LI, *Adv. Prot. Chem.*, **11**, 101 (1956).
31. R. L. STEHLE, *Brit. J. Pharmacol. and Chemotherap.*, **8**, 439 (1953).
32. D. SPACKMAN, E. L. SMITH and D. M. BROWN, *J. Biol. Chem.*, **212**, 255 (1955).
33. K. SHIZUME, A. B. LERNER and T. B. FITZPATRICK, *Endocrinology*, **54**, 553 (1954).
34. P. E. SMITH, *Anat. Rec.*, **32**, 221 (1926).
35. see *J. Am. Med. Assoc.*, **147**, 326 (1951).
36. W. F. WHITE, *J. Am. Chem. Soc.*, **75**, 503 (1953).
37. C. H. LI, I. I. GESCHWIND, A. L. LEVY, J. I. HARRIS, J. S. DIXON, N. G. PON and J. O. PORATH, *Nature*, **173**, 251 (1954).

38. P. H. BELL, *J. Am. Chem. Soc.*, **76**, 5565 (1954).
39. C. H. LI and J. S. DIXON, *Science*, **124**, 934 (1956).
40. R. G. SHEPHERD, K. S. HOWARD, P. H. BELL, A. R. CACCIOLA, R. G. CHILD, M. C. DAVIES, J. P. ENGLISH, B. M. FINN, J. H. MEISENHEDER, A. W. MOYER and J. VAN DER SCHEER, *J. Am. Chem. Soc.*, **78**, 5051 (1956).
41. C. H. LI, I. I. GESCHWIND, R. D. COLE, I. D. RAACKE, J. I. HARRIS and J. S. DIXON, *Nature*, **176**, 687 (1955).
42. J. S. DIXON and C. H. LI, unpublished experiments.
43. W. F. WHITE and W. A. LANDMANN, *J. Am. Chem. Soc.*, **77**, 1711 (1955).
44. W. F. WHITE, *J. Am. Chem. Soc.*, **76**, 4194 (1954).
45. R. D. COLE, C. H. LI, J. I. HARRIS and N. G. PON, *J. Biol. Chem.*, **219**, 903 (1956).
46. P. H. BELL, K. S. HOWARD, R. G. SHEPHERD, B. M. FINN and J. H. MEISENHEDER, *J. Am. Chem. Soc.*, **78**, 5059 (1956).
47. R. W. BOISSONNAS, ST. GUTTMANN, J. P. WALLER and P. A. JAQUENOUD, *Experientia*, **12**, 446 (1956).
48. C. H. LI, I. I. GESCHWIND, J. S. DIXON, A. L. LEVY and J. I. HARRIS, *J. Biol. Chem.*, **213**, 171 (1955).
49. R. G. SHEPHERD, S. D. WILLSON, K. S. HOWARD, P. H. BELL, D. S. DAVIES, S. B. DAVIS, E. A. EIGNER and N. E. SHAKESPEARE, *J. Am. Chem. Soc.*, **78**, 5067 (1956).
50. V. MENKIN, *Science*, **123**, 527 (1956).
51. C. H. LI, P. FÖNSS-BECH, I. I. GESCHWIND, T. HAYASHIDA, G. HUNGERFORD, A. J. LOSTROH, W. R. LYONS, H. D. MOON, W. O. REINHARDT and M. SIDEMAN, *J. Exptl. Med.*, **105**, 335 (1957).
52. P. STRICKER and F. GRUETER, *Comptes Rend. Soc. Biol.*, **99**, 1978 (1928).
53. O. RIDDLE and P. F. BRAUCHER, *Am. J. Physiol.*, **97**, 617 (1931).
54. H. M. EVANS, M. E. SIMPSON and W. R. LYONS, *Proc. Soc. Exptl. Biol. and Med.*, **46**, 586 (1941).
55. A. WHITE, *Vitamins and Hormones*, **7**, 253 (1949).
56. R. D. COLE and C. H. LI, *J. Biol. Chem.*, **213**, 197 (1955).
57. C. H. LI and J. CUMMINS, unpublished experiments.
58. J. A. FLEISCHER, *J. Biol. Chem.*, **147**, 525 (1943).
59. C. H. LI, R. D. COLE and M. COVAL, *J. Biol. Chem.*, **229**, 153 (1957).
60. R. D. COLE, I. I. GESCHWIND and C. H. LI, *J. Biol. Chem.*, **224**, 399 (1957).
61. C. H. LI, *J. Biol. Chem.*, **229**, 157 (1957).
62. C. H. LI, *J. Biol. Chem.*, **178**, 459 (1949).
63. H. FRAENKEL-CONRAT, *J. Biol. Chem.*, **142**, 119 (1942).
64. S. J. CROWE, H. CUSHING and J. HOMANS, *Bull. Johns Hopkins Hosp.*, **21**, 127 (1910).
65. C. H. LI and H. M. EVANS, *Science*, **99**, 183 (1944); C. H. LI, H. M. EVANS and M. E. SIMPSON, *J. Biol. Chem.*, **159**, 353 (1945).
66. A. E. WILHELMI, J. B. FISHMAN and J. A. RUSSELL, *J. Biol. Chem.*, **176**, 737 (1948).
67. H. CLAUSER and C. H. LI, *J. Am. Chem. Soc.*, **76**, 4337 (1954).
68. J. G. PIERCE, *Biochem. J.*, **57**, 16 (1954).
69. E. REID and A. E. WILHELMI, *Proc. Soc. Exptl. Biol. and Med.*, **91**, 267 (1956).
70. S. ELLIS and M. E. SIMPSON, *J. Biol. Chem.*, **220**, 939 (1956).
71. C. H. LI, *Harvey Lectures*, **41**, 181 (1952).
72. J. I. HARRIS, C. H. LI, P. G. CONDLIFFE and N. G. PON, *J. Biol. Chem.*, **209**, 133 (1954).
73. C. H. LI, H. PAKKOFF, P. FÖNSS-BECH, and P. CONDLIFFE, *J. Biol. Chem.*, **218**, 33 (1956).
74. See for instance, E. SHORR, A. C. CARTER, R. W. SMITH, JR, B. J. KENNEDY, R. J. HAVE, T. N. ROBERTS, L. L. SONKIN and E. T. LIVINGSTONE, in *The Hypophyseal Growth Hormone, Nature and Actions*, R. W. Smith, Jr., O. H. Gaebler and C. N. H. Long, eds., New York, Blakiston, 1955, p. 522.

75. G. E. PICKFORD, *Endocrinology*, **55**, 274 (1954).
76. A. E. WILHELMI, cited by Pickford.⁷⁵
77. E. KNOBIL and R. O. GREEP, *Fed. Proc.*, **15**, 111 (1956).
78. E. KNOBIL, R. C. WOLF, R. O. GREEP and A. E. WILHELMI, *Endocrinology*, **60**, 166 (1957).
79. C. H. LI, and H. PAPKOFF, *Science*, **124**, 1293 (1956).
80. C. H. LI, *Fed. Proc.*, in press.
81. C. H. LI, *J. Biol. Chem.*, **211**, 555 (1954).
82. D. M. BERGENSTALL, private communication, quoted in.⁸⁰
83. J. C. BECK, E. E. MCGARRY, I. CYRENFURTH and E. H. VENNING, *Science*, **125**, 884 (1957).
84. A. EHRENBERG and F. HEIJKENSKJÖLD, *Acta chem. Scand.*, **10**, 1675 (1956).
85. C. H. LI, A. PARCELLS and H. PAPKOFF, unpublished experiments.
86. J. VALSÖ, *Klin. Wchnschr.*, **132**, 1819 (1934).

Intervention sur l'exposé du Professeur Li

A propos des groupes C-terminaux de l'hormone de croissance et de l'hormone lactogène

M. JUTISZ

Laboratoire de Morphologie Expérimentale et Endocrinologie,
Collège de France

Le Professeur C. H. Li m'a demandé d'apporter, à la fin de son exposé, quelques précisions au sujet d'un travail que Mlle D. M. Meyer et moi-même avons fait en collaboration avec lui.¹ Il s'agit de la détermination des groupes C-terminaux de l'hormone de croissance et de l'hormone lactogène par réduction au moyen de l'hydrure double de lithium et d'aluminium.

Nous avons utilisé dans ces recherches le procédé publié par Fromageot et Coll. en 1950.² Ce procédé, je le rappelle brièvement, consiste à réduire les groupes C-terminaux des protéines et des peptides estérifiés ou non en alcools primaires et à isoler, après l'hydrolyse complète du produit réduit, les aminoalcools correspondants.

Cette méthode a l'inconvénient de ne pas toujours respecter l'intégrité du produit traité et de conduire à des ruptures de certaines liaisons peptidiques particulièrement fragiles.^{1,3,4} C'est donc pour nous entourer de quelques garanties que nous avons contrôlé, à chaque stade de l'opération, les groupements N-terminaux des produits obtenus, par la méthode de Sanger.

Hormone de croissance (somatotropine). Cette hormone a été estérifiée par la méthode de Fischer et Speier,⁵ dans le méthanol chlorhydrique 0,1 N, à 25° C, pendant 24 heures.

Les résultats de détermination des groupes N-terminaux et C-terminaux à des différents stades du traitement de l'hormone ont été rassemblés dans le tableau 1.

On peut tirer de l'examen du tableau 1 les conclusions suivantes:

1) la réduction de l'hormone native à 35° C fait apparaître un groupement N-terminal nouveau du glycocole, en plus de deux résidus d'alanine et de phénylalanine, existant à l'origine dans la molécule;

2) l'estérification de l'hormone native à 25° C engendre également un groupement N-terminal supplémentaire du glycocole;

3) la réduction de l'hormone estérifiée donne naissance à 1,8 résidus du glycocole;

4) en ce qui concerne les groupes C-terminaux, la réduction de l'hormone

Tableau 1

LES CHIFFRES EXPRIMENT LES RÉSIDUS DES DNP-AMINOACIDES
OU DES AMINOALCOOLS PAR 47·000 G DE SOMATOTROPINE

	Résidus N-terminaux: (après correction)	Aminoalcools: (après correction)
Somatotropine native	DNP-alanine 1,00 DNP-phénylalanine 1,02	
Somatotropine native réduite à 35° C	DNP-alanine 1,13 DNP-phénylalanine 1,13 DNP-glycocolle 1,00	colamine 0,94 phénylalaninol traces
Somatotropine ester (25° C, pendant 24 H)	DNP-alanine 0,92 DNP-phénylalanine 0,98 DNP-glycocolle 0,87	
Somatotropine ester réduite à 35° C	DNP-alanine 0,85 DNP-phénylalanine 1,00 DNP-glycocolle 1,80	colamine 1,50 phénylalaninol 0,86

native fait apparaître 0,9 résidu de colamine et seulement des traces de phénylalaninol;

5) on trouve, dans l'hormone estérifiée réduite, 1,5 résidus de colamine et 0,86 résidu de phénylalaninol.

Il semble logique d'admettre, à la lumière de ces résultats, que les résidus de colamine proviennent d'une coupure de la liaison Gly-Gly dans la chaîne peptidique de l'hormone, puisqu'ils ont leurs correspondants sous forme de glycocolle du côté aminé, et que 0,86 résidu de phénylalaninol est dû à la réduction du groupe C-terminal estérifiée de phénylalanine.

Hormone lactogène (prolactine). Le procédé utilisé dans le cas de cette hormone a été le même que précédemment, sauf pour l'estérification, qui a été poursuivie pendant 70 heures.⁶ Les résultats obtenus sont donnés dans le tableau 2.

Tableau 2

LES CHIFFRES EXPRIMENT LES RÉSIDUS DES DNP-AMINOACIDES OU DES AMINOALCOOLS PAR 26·550 G DE PROLACTINE

	Résidus N-terminaux (sans correction)	Aminoalcools: (sans correction)
Prolactine native	DNP-thréonine 0,46	
Prolactine ester (25° C, pendant 70 H)	DNP-thréonine 0,57	
Prolactine ester réduite à 35° C	DNP-thréonine 0,58 DNP-X 0,31	colamine 0,27

Il apparaît d'après ces résultats que :

1) l'estérification ne provoque pas de coupures dans la chaîne peptidique de l'hormone lactogène, contrairement au cas de la somatotropine;

2) au cours de la réduction de la prolactine estérifiée il y a apparition de quantités à peu près équivalentes de DNP-aminoacide non identifié, et de colamine; ces deux produits provenant probablement d'une dégradation partielle de la molécule.

On peut admettre, suivant ces résultats, que la prolactine ne possède pas de groupement C-terminal libre, ou bien que ce groupement, s'il en existe un, soit résistant à la réduction par les moyens utilisés.

BIBLIOGRAPHIE

1. D. M. MEYER, M. JUTISZ et C. H. LI, *Bull. Soc. Chim. Biol.*, **38**, 615 (1956).
2. CL. FROMAGEOT, M. JUTISZ, D. MEYER et L. PÉNASSE, *Biochim. Biophys. Acta*, **6**, 283 (1950).
3. M. JUTISZ, *Bull. Soc. Chim. Biol.*, **36**, 109 (1954).
4. M. JUTISZ, D. M. MEYER et L. PÉNASSE, *Bull. Soc. Chim. France*, 1087 (1954).
5. E. FISCHER et A. SPEIER, *Ber. Chem. Ges.*, **28**, 3252 (1895).
6. C. H. LI et H. FRAENKEL-CONRAT, *J. Biol. Chem.*, **167**, 495 (1947).

The structure and activity of melanocyte-stimulating and adrenocorticotropic peptides

J. IEUAN HARRIS*

Department of Biochemistry, University of Cambridge

The isolation of highly purified melanocyte stimulating substances from pig posterior pituitary extracts has been reported from several different laboratories during the past two years. (Lerner and Lee, 1955; Lee and Lerner, 1956; Porath, Roos, Landgrebe and Mitchell, 1955; Benfey and Purvis, 1955; Geschwind, Li and Barnafi, 1956.)

The material isolated by Lerner and Lee (1955) was found to be a highly basic polypeptide with an isoelectric point of 11.0-11.5, whilst both Porath *et al.* (1955) and Geschwind *et al.* (1956) reported isoelectric points of 5.2 and 5.8, respectively, for the melanocyte stimulating hormone (MSH) preparations which they had isolated. This apparent discrepancy was later resolved when Lee and Lerner (1956) showed that both forms of the hormone were present in the same extracts of pig posterior pituitaries. Consequently they proposed that the basic polypeptide which they had originally isolated be called α -melanocyte-stimulating hormone (α -MSH), and that the acidic polypeptide hormone isolated in the other laboratories be called β -melanocyte stimulating hormone (β -MSH).

THE β -MELANOCYTE-STIMULATING HORMONE (β -MSH)

The complete structure of β -MSH was elucidated by Harris and Roos (1956) using material prepared by Roos according to the isolation procedure of Porath *et al.* (1955). The amino acid sequence of the peptide hormone was deduced from the results of 'stepwise' degradation studies on the intact molecule and from the characterization of peptide fragments isolated after digestion with trypsin and chymotrypsin, as shown in Fig. 1. In this way β -MSH was shown to consist of eighteen amino acid residues in a single polypeptide chain, and to contain a sequence of seven amino acids, Met.Glu.His.Phe.Arg.Try.Gly. (positions 7-13), which had previously been shown to occur in the corticotropins (e.g. Bell, 1954); in addition, the tyrosine and proline residues, in positions 5 and 15, respectively, also occur in

* Member of the Scientific Staff of the Medical Research Council.

corresponding positions in the corticotropin molecule. The structural similarity between β -MSH and corticotropin is of particular interest since it provides a chemical basis for the observation (Bell, 1954; Dixon, 1956a) that pure corticotropin preparations possess intrinsic melanocyte stimulating activity in addition to their primary adrenocorticotrophic action.

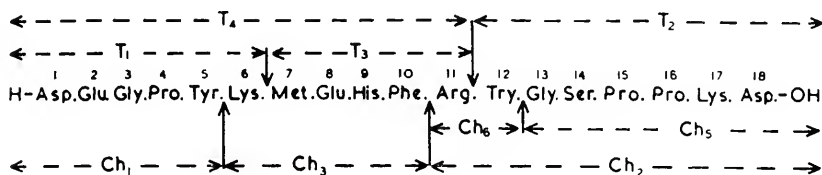


Fig. 1. Amino-acid sequence of β -melanocyte-stimulating hormone (pig β -MSH); arrows indicate points of cleavage by trypsin (T) and chymotrypsin (C).

Geschwind, Li and Barnafi (1956, 1957a) have reported similar conclusions concerning the melanocyte stimulating peptide which they isolated from pig posterior pituitary extracts, showing that it was identical in all respects with the material isolated by Porath *et al.* (1955) and characterized by Roos (1956). Geschwind, Li and Barnafi (1957b) have also reported the isolation and structure of a β -type melanocyte-stimulating hormone from ox pituitaries; β -MSH (ox) was found to differ from β -MSH (pig) only in position 2 (Fig. 1), where the glutamic acid residue in pig β -MSH is replaced by a serine residue in the ox hormone.

THE α -MELANOCYTE-STIMULATING HORMONE (α -MSH)

The structure of α -MSH has been determined by Harris and Lerner (1957), using material prepared by Dr Teh Lee according to the procedure described by Lee and Lerner (1956). Although it did not contain either a free *N*-terminal or *C*-terminal group, the amino acid sequence of α -MSH could be deduced from the structure of peptide fragments derived from it by digestion with trypsin and chymotrypsin. The results are summarized in Fig. 2.

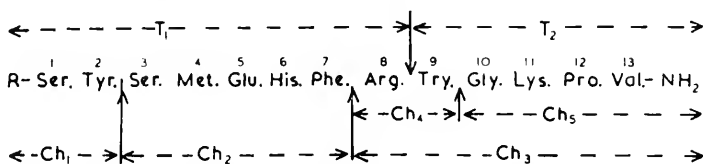
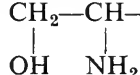


Fig. 2. Amino-acid sequence of α -melanocyte-stimulating hormone (pig α -MSH); arrows indicate points of cleavage by trypsin (T) and chymotrypsin (C).

α -MSH is shown to consist of thirteen amino acid residues in a single peptide chain, and the amino acid sequence is identical with the *N*-terminal tridecapeptide sequence in corticotropin. However, it differs from the related portion of the corticotropin molecule in that the *N*-terminal residue,

of α -MSH now reveals that the structure essential for melanocyte-stimulating activity must reside within the *N*-terminal tridecapeptide fragment of corticotropin; and from a comparison of the structures of the two melanocyte-stimulating hormones it would appear that the structure essential for biological activity is in both cases to be found within a sequence of eleven amino acid residues (i.e. positions 2-12 in α -MSH, and 5-15 in β -MSH), since the *N*-terminal tetrapeptide and *C*-terminal tripeptide sequences in β -MSH can be replaced by the *N*-substituted serine, and valine amide moieties, respectively, in α -MSH.

Although the entire amino acid sequence of α -MSH is contained within the corticotropin molecule, its actual melanocyte-stimulating activity is considerably less than that of α -MSH. This suggests that the additional structural features which are necessary for adrenocorticotropic activity may at the same time inhibit the potential melanocyte-stimulating properties of the corticotropin molecule. For example, the *N*-terminal serine residue has been shown to be essential for the adrenocorticotropic activity, but not for the melanocyte-stimulating activity, of corticotropin (Dixon, 1956*b*). α -MSH, on the other hand, in which the *N*-terminal serine occurs as an *N*-substituted derivative, acquires enhanced melanocyte-stimulating activity, and it could be inferred that the unsubstituted $\text{CH}_2\text{—CH—}$ structure may, at least in



part, be responsible for the reduced melanocyte-stimulating activity of corticotropin. In accord with this hypothesis, it has now been shown (Dixon and Lerner, unpublished results) that when the *N*-terminal serine residue in corticotropin is oxidized with periodate, loss of adrenocorticotropic activity is accompanied by a significant increase in the melanocyte-stimulating activity of the periodate modified corticotropin. Similarly, Shepherd *et al.* (1956) have shown that when corticotropin is subjected to mild alkaline hydrolysis (which results in the cleavage of peptide bonds in the *N*-terminal Ser.Tyr.Ser. structure), loss of adrenocorticotropic activity is again accompanied by an increase in melanocyte-stimulating activity.

Although a conclusive evaluation of these results must await the isolation and characterization of the periodate and alkali degradation products of corticotropin, it is clear that corticotropin can be changed into a predominantly melanocyte-stimulating substance, and that the *N*-terminal serine plays an important role in determining which of the two biological properties shall predominate.

The other outstanding question concerns the nature of the smallest active adrenocorticotropic peptide. It is already clear that, in addition to the common heptapeptide sequence, a free *N*-terminal serine, and in all probability the entire Ser.Tyr.Ser sequence, is essential for adrenocorticotropic activity. Preliminary studies on the selective removal of the *N*-terminal substituent in α -MSH show that the unsubstituted tridecapeptide does not

possess adrenocorticotropic activity. This indicates that a specific sequence of an as yet undetermined number of amino acids must also be added to the C-terminal residue of the common sequence in order to produce an active adrenocorticotropic peptide.

REFERENCES

- BELL, P. H. (1954), *J. Amer. Chem. Soc.*, **76**, 5565.
BENFEY, B. J. and PURVIS, J. L. (1955), *J. Amer. Chem. Soc.*, **77**, 5167.
DIXON, H. B. F. (1956a), *Biochim. Biophys. Acta*, **19**, 392.
DIXON, H. B. F. (1956b), *Biochem. J.*, **62**, 25P.
GESCHWIND, I. I., LI, C. H. and BARNAFI, L. (1956), *J. Amer. Chem. Soc.*, **78**, 4494.
GESCHWIND, I. I., LI, C. H. and BARNAFI, L. (1957a), *J. Amer. Chem. Soc.*, **79**, 620.
GESCHWIND, I. I., LI, C. H. and BARNAFI, L. (1957b), *J. Amer. Chem. Soc.*, **79**, 1003.
HARRIS, J. I. and ROOS, P. (1956), *Nature*, **178**, 90.
HARRIS, J. I. and LERNER, A. B. (1957), *Nature*, **179**, 1346.
LERNER, A. B. and LEE, T. H. (1955), *J. Amer. Chem. Soc.*, **77**, 1066.
LEE, T. H. and LERNER, A. B. (1956), *J. biol. Chem.*, **221**, 943.
PORATH, J., ROOS, P., LANDGREBE, F. W. and MITCHELL, G. M. (1955), *Biochim. Biophys. Acta.*, **17**, 598.
ROOS, P. (1956), *Acta. Chem. Scand.*, **10**, 1061.
SHEPHERD, R. G., WILLSON, S. D., HOWARD, K. S., BELL, P. H., DAVIES, D. S., DAVIS, S. B., EIGNER, E. A. and SHAKESPEARE, N. E. (1956), *J. Amer. Chem. Soc.*; **78**, 5067.
SHIZUME, K., LERNER, A. B. and FITZPATRICK, T. B. (1954), *Endocrinology*, **54**, 553.

Some observations on the structural basis of enolase activity

BO G. MALMSTRÖM

Institute of Biochemistry, Uppsala, Sweden

Studies on pepsin,¹ ribonuclease,²⁻⁴ lysozyme,⁵ and papain⁶ have demonstrated that the intact protein molecule is not necessary for the catalytic activity of these enzymes. Recent experiments with yeast enolase have shown that this enzyme also can be considerably modified without affecting the enzymic activity. It has been found⁷ that the crystalline enzyme, prepared according to Warburg and Christian,⁸ contains varying amounts of several active forms (enolase *a*, *b*, and *c*), characterized by different electrophoretic mobilities. The *b* and *c* enzymes are believed to arise from enolase *a* by proteolytic action during the yeast autolysis [see Malmström⁷], and this has prompted a study of the effect of limited proteolysis on the activity of enolase. Some preliminary results of these experiments, together with a number of other observations on the structural basis of enolase activity, will be reported in the present communication.

The enzyme was purified by zone electrophoresis and specific adsorption on to the Mg^{++} form of a cation exchanger, sulphomethyl cellulose;⁷ this last step takes advantage of the high affinity between enolase and its activating ions. By the new purification procedure it is possible to prepare gram quantities of enolase relatively easily. The method also allows separation of the different active forms of the enzyme, and purified enolase *a* shows the presence of only one component both in zone electrophoresis and in ultracentrifuge experiments over a wide range of conditions.

Qualitative end-group studies have been made. Edman's method⁹ was used for the *N*-terminal residues, and DFP (diisopropyl fluorophosphate) treated carboxypeptidase¹⁰ for the *C*-terminal groups. No differences in either *N*- or *C*-terminal amino acids could be found between enolase *a* and *b*, showing that the difference between these forms is not in the end groups.

The effect of digestion with trypsin, carboxypeptidase and aminopeptidase on the activity of the enzyme has been studied. The trypsin and carboxypeptidase used were commercial preparations (Worthington), while aminopeptidase was prepared according to Smith and Hill [personal communication (see also Ref.¹¹)]. Both exopeptidases were treated with DFP prior to use. Trypsin digestion led rather rapidly to a considerable drop in activity, and

no attempts to isolate active fragments have as yet been made. However, with both carboxypeptidase and aminopeptidase a large number of amino acids can be removed from the enolase molecule without changing the activity. Table I summarizes some results on the liberation of amino acids

Table I
THE LIBERATION OF AMINO ACIDS FROM
ENOLASE ON DIGESTION WITH EXOPEPTIDASES
AT pH 8.5 AND 22°

Time of incubation (hr.)	Leucine equivalents/mole of enolase	
	Carboxypeptidase	Aminopeptidase
0.6	7	13
3.1	33	47
10.3	67	73
22.5	80	92

as a function of time of incubation with these two enzymes. In both cases, the molar ratio between enolase and exopeptidase was 30 : 1. The incubations were carried out at pH 8.5; aliquots of the incubation mixtures were taken out at different times and the reaction terminated by precipitating the proteins with trichloroacetic acid to a final concentration of 5 per cent. Amino acids were determined by the ninhydrin method of Moore and Stein,¹² and the results, expressed as leucine equivalents, have been corrected for small amounts of ninhydrin-positive material liberated also in the absence of added peptidase. It is apparent that both peptidases liberate close to 100 residues in about 24 hours. In connection with the rather extensive degradation achieved with carboxypeptidase, it is of interest to note that enolase contains no proline. The removal of this number of amino acids from either the C- or N-terminal end causes no appreciable loss of activity. The digestion with carboxypeptidase was found to cause a large change in the electrophoretic mobility of the enzyme.

The experiments described show that about 17 per cent of the enolase molecule is not essential for maintaining the active site. This is similar to the amount of digestion achieved with ribonuclease without loss of activity,⁴ but considerably less than in the case of papain,⁶ where close to 70 per cent of the molecule can be removed.

While limited removal of the N- and C-terminal chains does not affect the activity, the specific steric configuration of the active site appears dependent on non-covalent cross-linkages, since the enzyme is completely inactive in 8 M urea. This is in contrast to findings with ribonuclease² but similar to the behavior of many other enzymes.¹³ With enolase, the unfolding is reversible during the first few hours of incubation with urea, since the

activity is regained on dilution. However, after 24 hours in urea solution the inactivation is irreversible.

The active enolase molecule remaining after proteolytic digestion still contains about 500 amino acid residues, which is much too large to allow structural studies. Thus, as with most other enzymes, information about the active site is of an indirect nature. The influence of pH¹⁴ and solvent composition¹⁵ on the kinetics and chelating properties implicates the involvement of histidine both in metal binding and proton transfer. However, it is hoped that a more direct chemical attack will be possible by utilizing the strong binding of metal ions to enolase compared with most proteins and peptides.¹⁶ It is possible that, even after the activity has disappeared, the chelating site is left intact, and attempts to isolate metal-binding fragments after trypsin digestion are now in progress.

I wish to thank Professor A. Tiselius for his stimulating interest and helpful discussions. Mr. O. Nylander has cooperated in most of the experimental work, and a detailed report will later be published jointly with him.

REFERENCES

1. G. E. PERLMANN, *Nature*, **173**, 406 (1954).
2. C. B. ANFINSEN, N. F. HARRINGTON, A. HVIDT, K. LINDERSTRØM-LANG, M. OTTESEN and J. SCHELLMAN, *Biochim. Biophys. Acta*, **17**, 41 (1955).
3. S. M. KALMAN, K. LINDERSTRØM-LANG, M. OTTESEN and F. M. RICHARDS, *Biochim. Biophys. Acta*, **16**, 297 (1955).
4. G. KALNITSKY and W. I. ROGERS, *Biochim. Biophys. Acta*, **20**, 378 (1956).
5. J. I. HARRIS, C. H. LI, P. G. CONDLIFFE and N. G. PON, *J. Biol. Chem.*, **209**, 133 (1954).
6. R. L. HILL and E. L. SMITH, *Biochim. Biophys. Acta*, **19**, 376 (1956).
7. B. G. MALMSTRÖM, *Arch. Biochem. and Biophys.*, **70**, 58 (1957).
8. O. WARBURG and W. CHRISTIAN, *Biochem. Z.*, **310**, 384 (1942).
9. P. EDMAN, *Acta Chem. Scand.*, **4**, 283 (1950).
10. J. I. HARRIS, in D. Glick (ed.), *Methods of Biochemical Analysis*, **2**, 397 (1955).
11. D. H. SPACKMAN, E. L. SMITH and D. M. BROWN, *J. Biol. Chem.*, **212**, 255 (1955).
12. S. MOORE and W. H. STEIN, *J. Biol. Chem.*, **192**, 663 (1951).
13. J. I. HARRIS, *Nature*, **177**, 471 (1956).
14. B. G. MALMSTRÖM and L. E. WESTLUND, *Arch. Biochem. and Biophys.*, **61**, 186 (1956).
15. E. W. WESTHEAD and B. G. MALMSTRÖM, *J. Biol. Chem.*, **228**, 655 (1957).
16. B. G. MALMSTRÖM, *Arch. Biochem. Biophys.*, **46**, 345 (1953).

Insulin crystals: zinc atoms in the unit-cell, nucleation, growth and shape

JØRGEN SCHLICHTKRULL

Novo Terapeutisk Laboratorium, Copenhagen, Denmark

The number of metal atoms (Zn, Cd, Co or Ni) present in the unit-cell ($M = ca. 35,000$) of rhombohedral insulin crystals is an important detail in the structural concept. The various values reported for metal contents in the literature are quite confusing in contrast to the original figure,^{1,2,3} of 3 atoms/cell, which is also the number of dimer elements in the cell. Since the metal content of insulin crystals is actually subject to gross variations⁴ depending upon many conditions, attempts have been made to determine the number of zinc atoms/unit cell which is most relevant in structural considerations.

A quantitative examination showed some interesting features of the role played by zinc or the other metal ions in the crystallization. Insulin was crystallized with zinc in a phosphate buffer and recrystallized in ammonium acetate. The crystals contained 3 atoms of Zn/unit-cell (0.51% Zn), in perfect agreement with Scott and Fisher's experimental results.^{1,2} However, when the crystals were recrystallized once more, the zinc content dropped to 2 atoms/cell and remained constant at that value through several subsequent recrystallizations in ammonium acetate. If insulin is recrystallized from a citrate buffer containing zinc ions—which are strongly complexed to the citrate—then the crystals contain 2 atoms of zinc/unit-cell.⁵ When such crystals are recrystallized from a citrate buffer without the addition of zinc ions, the new crystals are again found to contain 2 atoms/unit-cell, but in the course of several such recrystallizations the content decreases a little and at the same time the crystals deteriorate in shape and amorphous particles appear in the precipitates. This slight decrease in the zinc content is not observed with recrystallizations from ammonium acetate and can presumably be accounted for by competition between insulin and citrate with respect to the binding of zinc ions. It is noteworthy that the zinc content of the citrate-recrystallized insulin (2 atoms/unit-cell) is not dependent on the pH in the interval over which crystallization takes place ($5.5 < \text{pH} < 7$). It is concluded from these experiments that among the many combinations between zinc and insulin crystals, the combination containing 2 atoms/unit-cell is a very strong complex and represents the minimum content of zinc ions sufficient for crystallization.

This value, 2 atoms/unit-cell, is also significant in quite other investigations, viz. in the ultracentrifugal studies of Cunningham, Fischer and Vestling on insulin solutions.⁶ These authors found that the sedimentation constant was independent of the zinc content when the zinc concentration varied from 0.2 to about 0.6-0.8 moles of zinc per 12,000 grams of insulin. However, a rise in the zinc concentration above the latter value brought about a linear rise in the sedimentation constant. The minimum amount sufficient for crystallization, 2 atoms of Zn/unit-cell, is equivalent to 0.7 moles of zinc per 12,000 grams of insulin and it is therefore likely that the sudden rise in the sedimentation constant at this value corresponds to the fact that the condition for crystallization is reached at this Zn content.

The shape of the insulin crystals depends on the species.⁷ Recrystallized pig-insulin crystals are perfectly shaped single rhombohedrons, but beef insulin crystals are twin-like bodies of a very different, star-like appearance. (It should be noted that the internal structure is rhombohedral in all cases considered here.) If beef-insulin crystals are recrystallized in one of the usual buffers and sodium chloride is added to the buffer, it can be observed that the shape of the crystals formed depends on the concentration of sodium chloride. If the concentration of sodium chloride exceeds approx. 6%, the crystals attain the perfect rhombohedral shape characteristic of pig-insulin. This rather abrupt change in shape is associated with a change in the minimum amount of zinc ions sufficient for crystallization. If the buffer contains so much sodium chloride, e.g. 7%, that the shape of beef insulin crystals becomes rhombohedral, the minimum adequate zinc concentration—and this applies also to pig-insulin—is no longer 2 but 4 atoms/unit-cell. Other experiments have shown that the chloride ion is responsible for these transformations and that the other halogen ions, with the exception of fluoride, have the same effect. Nitrate, sulphate and phosphate proved to have no effect. The transformations are not contingent upon the sodium ions since the latter can be replaced by potassium, ammonium and calcium.

Microscopical, goniometric measurements on a large (1.3 mm.) perfect insulin rhombohedron⁷ gave a value ($114^{\circ} 22' \pm 4'$ (S.E.)) which was in agreement with the X-ray value⁸ ($114^{\circ} 16'$) for the unit-cell in wet insulin crystals.

The kinetics of insulin crystallization have been studied in detail. It was found that the star-shaped or twin-like beef-insulin crystals have a tendency to form monodisperse suspensions. It means that the nuclei are formed almost exclusively in the initial stage of crystallization where the insulin concentration is greatest. The sharp and perfect pig-insulin rhombohedrons and the identically shaped beef-insulin crystals formed in saline are very polydisperse. The rate of nucleation was deduced from the very peculiar observation that the ratios between the total volume, total surface, total length and number of the crystals remain constant in the course of crystallization from the moment when the first few crystals had formed. It is possible to account for the nucleation in the most simple way if—and apparently

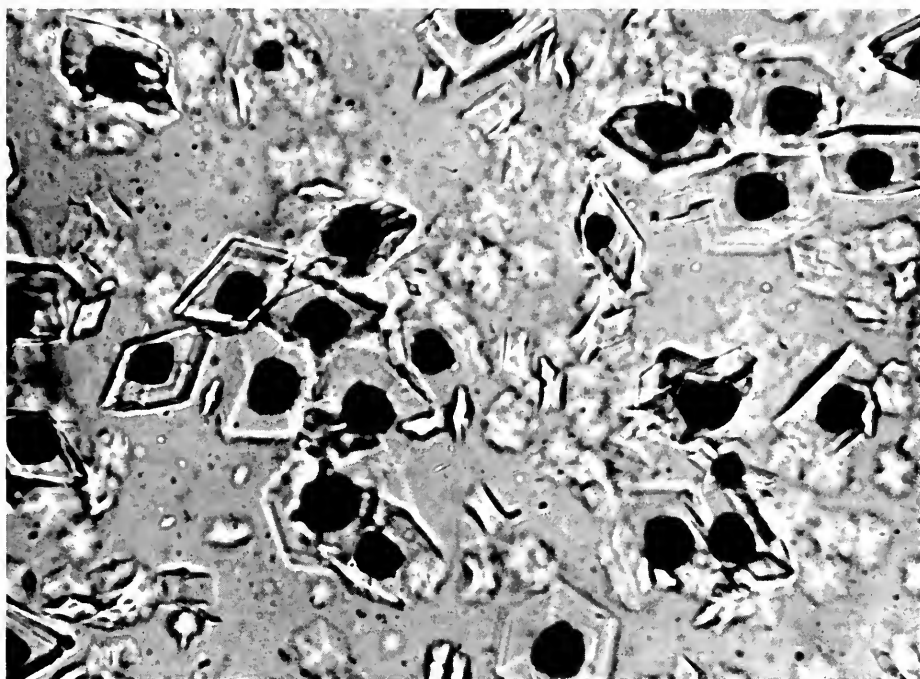


Fig. 1. Stained crystals after growth in colourless medium (300x).

only if—the *ad hoc* hypothesis is made that during crystallization the crystals throw off nuclei or crystal fractions at a rate proportional to their surface area and to their linear growth rate. This hypothesis finds support in the microscopical appearance of the suspension, since the very small ($<4\mu$) crystals are predominantly found on the vertices or faces of the bigger crystals. It was furthermore observed that the nucleation is enhanced if the surface of the crystallization system is increased, and suppressed by Tween-80 or by a surface layer of paraffin. The linear rate of deposition of insulin on the crystal faces was approx. 1μ /minute at the maximum. It was shown to be a function of the insulin concentration. In the absence of added halogenide, it was proportional to the third power of the supersaturation and in 7% NaCl to the second power.

The deposition of insulin on the various crystal faces was studied by letting the crystals first grow in a buffer containing Indian ink and then in a non-coloured buffer. The Indian ink became incorporated in the crystal bodies, and after some further growth in the colourless medium, it was observed that insulin is only deposited on the three faces meeting in one of the obtuse vertices. The same phenomenon was also observed in seeding experiments with non-coloured crystals. Since the unit-cells are identically orientated in the crystals, an 'up and down' with respect to growth in the unit-cell can be concluded.

If it is true that the rate of deposition on a crystal face is the same on the mirror image of that face, then it can be concluded from these observations that the unit-cell of the rhombohedral insulin crystals has no center of symmetry. This conclusion is in agreement with the lack of centro symmetry in the unit-cell of the rhombohedral insulin crystal.

REFERENCES

1. D. A. SCOTT, *Biochem. J.*, London, **28**, 1952 (1934).
2. D. A. SCOTT and A. M. FISHER, *Biochem. J.*, London, **29**, 1048 (1935).
3. D. A. SCOTT and A. M. FISHER, *Trans. Roy. Soc. Can.* V, **32**, 55 (1938).
4. K. HALLAS-MØLLER, K. PETERSEN and J. SCHLICHTKRULL, *Science*, **116**, 394 (1952).
5. J. SCHLICHTKRULL, *J. Acta Chem. Scand.*, **10**, 1455 (1956).
6. L. W. CUNNINGHAM, R. L. FISCHER and C. S. VESTLING, *J. Amer. Chem. Soc.*, **77**, 5703 (1955).
7. J. SCHLICHTKRULL, *Acta Chem. Scand.*, **10**, 1459 (1956).
8. D. CROWFOOT and D. RILEY, *Nature*, **144**, 1011 (1939).

Subject Index



- ACTH*, see Corticotropin
- Absorption spectra, ultra-violet*
of myoglobin, 144, 147
of ribonuclease, 233-234, 236
of tobacco mosaic virus, 252-253
- Acetyl groups in proteins*, 259-260, 335
- Albumin*, see Serum albumin
- Amide group, configuration of*, 18-21
localization in peptides, 192, 218-219
- Amino acids*
desalting of, 257
quantitative estimation of,
by paper chromatography, 91, 116-121
by isotope dilution, 116-121
as DNP derivatives, 257
- Amino groups, terminal, determination* (see also Peptides, structure determination)
by dinitrofluorobenzene, 91, 264, 267-268
by Edman degradation, 215-218, 264-266
by leucinaminopeptidase, 192, 218-219
- Aminopeptidase*, see Leucinaminopeptidase
- Antibodies*
as enzyme inhibitors, 244-246
enzymic digestion of, 290
to growth hormones, 326
to ribonuclease, 241-246
to serum proteins, 298-301
to silk fibroin, 291
to tobacco mosaic virus, 262
- Antigen-antibody reaction*, 246, 293-294
as criterion of homogeneity, 240-241
by diffusion in gels, 241-243, 292, 299-301
inhibition of, 291-294
- Asparagine, localization in peptides*, 192, 218-219
- Bacitracin, dialysis rate*, 108-110
- Carbobenzoxylation of -NH₂ groups*, 224-232
- Carboxyl groups, terminal, determination* (see also Peptides, structure determination)
by carboxypeptidase, 171, 218-219, 233, 250, 263-264, 287, 315, 318, 322
by hydrazinolysis, 171, 215, 218, 258-260, 264, 310, 315, 317
by LiAlH₄ reduction, 277, 330-332
- Carboxypeptidase, action on*
enolase, 338-340
growth hormone, 318, 322
lysozyme, 277, 285
poly-DL-alanine, 33
prolactin, 315, 317
ribonuclease, 233
tobacco mosaic virus, 250, 263-264
- Carboxypeptidase B, action on tryptophan*, 171
- Chromatography of*
amino-acids on paper, 91, 116-121
peptides
combined with paper electrophoresis, 148-150, 257
on ion exchange resins, 161, 186-192, 212-221, 257, 281, 283, 285-286
proteins
on calcium phosphate, 96-99
on cellulose derivatives, 100-103, 338
on ion exchange resins, 100-103, 197, 199, 200, 324
- Chymotrypsins, α , α_1 , δ , π* , 155-167
action on
asparaginyl bonds, 161-162, 175, 213

- Chymotrypsins*, action on—*cont.*
 chymotrypsinogen, 159-163
 corticotropin, 310, 311
 growth hormone, 318-319, 322, 324, 326
 melanotropin, 305, 334
 ribonuclease, 212-213, 219
 ribonuclease peptides, 214-216
 serum albumin, 291
 tobacco mosaic virus, 257-258, 268
 dialysis rate, 108, 110
 oxidized, enzymic hydrolysis of, 175
- Chymotrypsinogen*
 action of chymotrypsin on, 159-163
 action of leucinaminopeptidase on, 164-168
 activation, 155-167
 amino-acid composition, 81
 basic peptides from, 86-87, 175
 cysteic acid peptides from, 84
 dialysis rate, 108, 110
 reaction with DFP, 177
- Collagen*
 ϵ -NH₂ groups, covalent linkage, 221
 viscosity, 38
- Conalbumin*
 configuration, 55-56
 physico-chemical properties, 55-56
 viscosity, 38
- Corticotropins*
 biological activity and structure, 312-313, 335-337
 enzymic hydrolysis, 310-312
 partial synthesis, 312
 periodate oxidation, 336
 relation to melanotropins, 313, 333-337
 species differences, 73, 307-313
- Counter current distribution*, 257, 304, 312, 314
- C-terminal groups*, see Carboxyl groups
- Cysteic acid peptides*, isolation of, 84-85, 186-187
- Cytochrome C*, 66-72
 dialysis rate, 108, 110
 haemopeptide from, 70
 species differences in, 71
 structure, 70-72
- Denaturation*, 44, 45, 49-62, 233-238, 250-254, 280, 339-340
- Deuterium exchange in*
 insulin A chain, 32
 β -lactoglobulin, 32
 poly-DL-alanine, 23-34
 kinetics of, 23-34
 poly-DL-serine, 33
 ribonuclease, oxidized, 32
 triglycine, 30-31
- Dialysis*
 rate of, and ionic strength, 112-113
 and pH, 110
 and temperature, 109
 and urea, 110
 separation of proteins, by, 104-115, 291
- Diffusion coefficient*, see under individual proteins, physico-chemical properties of
- Dinitrophenylation* (see also Amino terminal group determination)
 microtechnique, 91
 of proline peptides, 264
- Dinitrophenyl protein*, from tobacco mosaic virus, 267
 enzymic degradation of, 268
 removal of DNP residue from, 268
- Disulphide bonds*
 and stability of proteins, 44, 45, 62
 oxidation, 175, 211, 220, 224-232, 236, 282-284, 316, 322
 reductive cleavage, 224-232, 316
 in ribonuclease, position of, 219-221, 228-230
- Edman degradation*, 215-218, 264-266
- Electrophoresis in*
 agar, 241-243
 cellulose columns, 94-97, 291
 multicompartament apparatus, 90
 paper, 90, 141, 148-150, 227-229, 232, 268, 305
 starch gel, 140, 307-308, 316
- Electrostatic interaction in*
 haemoglobin, 44
 polymethacrylic acid, 47-49
 protein models, 39-41, 46
- Enolase*, 338-340
- Esterification of proteins*, 236, 330-332

- Fibrinogen*, viscosity, 38
- Fibroin*, α -helix in, 17
antibodies to, 291
- Fluorescence*, polarization of, 54, 59
- Gliadin*, dialysis rate, 108
- γ -Globulin
optical rotation, 57-58
viscosity, 57-58
- Glucagon*, dialysis rate, 108
- Glutamine*, localization in peptides, 192, 218-219
- Growth hormone*
enzymic modification and activity
chymotrypsin, 318-319, 322, 324, 326
carboxypeptidase, 318, 322
LiAlH₄ reduction, 330-332
physico-chemical properties, 318, 321, 325
purification, 320
species differences, 318-327
- Guanidine-acetate ion interaction*, 42-43
- Guanidine hydrochloride*, 219, 236, 251
- Haemoglobin*
basic peptides from, 87-89
C, 150-151
configuration, pH dependence of, 54-55
denaturation, 44, 55
dissociation, 61, 89
haem groups, orientation of, 139
peptides from tryptic hydrolysates, 148-151
genetic control of structure of, 148-151
shape of molecule, 137
sickle cell, 17-18, 74, 148-151
sulphydryl groups, 136-138
X-ray analysis, 136-139
- Haptenes*, 293-294
- α -Helix
in polypeptides, 17, 24, 28-29, 31-33, 36
in proteins, 17, 32, 36-38, 136, 172-173
- Histidine*
in active centre of chymotrypsin, 175
cytochrome C, 72
myoglobin, 145
trypsin, 172-173, 175
and organophosphorus ester hydrolysis, 177
- Hydrazides*, chromatography of, 259
- Hydrazinolysis*, 171, 215, 218, 258-260, 264, 267, 310, 315, 317
- Hydrogen bonds*
between peptide groups, 17-34, 41, 42, 47 (see also α -helix and deuterium exchange)
breakage mechanism, 28, 29, 32
linearity of, 19-21
carboxyl-carboxyl, 41, 42
carboxyl-guanidine, 42-43
carboxyl-tyrosine, 41, 42 (see also Tyrosine, ionization of)
in pepsin, 179
in ribonuclease, 236-238
in tobacco mosaic virus, 252-255
in serum albumin, 53-54
and sulphydryl groups, 252
in trypsin, 173
in urea solutions, 41
in water, 39, 41
- Hydrophobic bonds*, 33, 39, 47, 49
- Hydroxylamine*,
reaction with organophosphorus esters, 177
tobacco mosaic virus, 266
- Hypertensin*, species differences, 73
- Insulin*
A chain, deuterium exchange, 32
dialysis rate, 111, 112
B chain, α -helix, 17
dialysis rate, 108, 112
configuration in solution, 59
deuterium exchange, 32
polymerization, 60, 111
species differences, 73
zinc, role in crystallization of, 341-343
- Intermedin*, see Melanotropin
- Ionization of groups in proteins*, 43-62, 201-203, 236, 252-254
- Isotope dilution*, amino acid estimation by, 116-121
- Kinetics*
of deuterium exchange, 23-34
of papain action, 200-203

- Lactogenic hormone*, see Prolactin
- β -Lactoglobulin*
 configuration and pH, 57
 deuterium exchange, 32
 polymerization, 61
 viscosity, 38
- Leucinaminopeptidase*
 action on chymotrypsinogen, 164-167
 enolase, 338-340
 β -melanotropin, 307, 308
 mercuripapain, 192-199
 poly-DL-alanine, 33
 localization of amide groups in peptides by, 192, 218-219
- Lysozyme*
 action of pepsin on, 285-288
 chymotrypsin on, 280-285, 289
 trypsin on, 227-228, 280-285, 289
 configuration and pH, 59-60
 dialysis rate, 108, 110
 disulphide bonds, oxidation, 282-285
 reduction and carboxymethylation, 226-228
- α -Melanotropin*
 structure, 334-335
 relation to corticotropin, 335-337
- β -Melanotropin*
 structure, 303-307, 333-334
 properties, species differences in, 309
 relation to corticotropin, 313, 335-337
- Myoglobin*
 amino acid composition, 145-146
 electrophoretic heterogeneity, 140-143
 heavy atom derivatives, 129-132
 molluscan, 144-147
 oxygen dissociation curves, 141-142, 145
 physico-chemical properties, 144-147
 X-ray analysis, 127-135
- Neochymotrypsinogens*, 159-163
- Nucleic acid*, 234, 250, 256, 262, 272
- Optical rotation*, 17
 and chymotrypsinogen activation, 155
- γ -globulin, 58
 ribonuclease, 234-235
 effect of guanidine on, 236-237
 and trypsinogen activation, 169, 170
- Ovalbumin*
 configuration and pH, 59
 dialysis rate, 108
 disulphide bonds in, 45
 viscosity, 38
- Ovomucoid*, dialysis rate, 108
- Papain*
 action on antiovalbumin, 290
 amino acid composition, 187
 autolysis, 199, 200
 chromatography of, 197, 200
 cysteic acid peptides from, 186-187
 homogeneity, 183-184
 kinetics, 200-206
 leucinaminopeptidase on, 192-199
 mechanism of action, 203-208
 mercury derivative, 184
 oxidized, action of trypsin on, 188-192
 pH, effect on activity, 200-203
 physico-chemical properties, 183-184
 sulphydryl groups and activity, 184-186, 191, 202-206
- Pepsin*
 action on corticotropin, 310-312
 lysozyme, 285-286
 ribonuclease, 212-213, 233, 235
 autolysis, effect of urea on, 179-181
 configuration and pH, 55, 57
 dialysis rate, 108
 viscosity, 38, 55, 57
- Peptidases*, see Carboxypeptidase and Leucinaminopeptidase
- Peptide bond*, configuration of, 18-21
- Peptides* (see also under individual proteolytic enzymes and proteins)
 basic, 86-89
 comparison of composition from different proteins, 79-84
 cysteic acid, 84-85, 186-187, 200, 229-230, 283-284

Peptides—cont.

- fractionation by electrophoresis, 78, 79, 90, 227-229, 268, 281
- ion-exchange chromatography, 161, 186-192, 212-221, 257, 281, 283, 285-286
- paper chromatography, 78, 79, 90, 268, 281
- partition chromatography, 268
- two-dimensional electrophoresis chromatography, 148-150, 257
- structure determination, 148, 215-229, 257-258, 281-282, 305-312, 333-335 (see also Amino groups, terminal and carboxyl groups, terminal)
- Performic acid oxidation*, 175, 186, 191, 211, 220, 224-232, 236, 282-284, 316, 322
- Peroxidase, chromatography of*, 100-103
- pH and protein configuration*, 45-60
- Phenolic groups in proteins*, see Tyrosine
- Phosphatase, chromatography of*, 100-103
- Pituitary hormones*, 302-327, 333-337 (see also under individual hormones)
- Poly-DL-alanine*, 24-34
 - action of peptidases on, 33-34
 - deuterium exchange in, 26-33
 - molecular species in solution, 31
 - physico-chemical properties, 25
- Polymethacrylic acid*, 47-49, 51
- Polypeptides, synthetic*
 - α -helix in, 17, 24, 33, 36
 - hydrogen bonds in, 19-20, 23-34
- Poly-DL-serine, deuterium exchange*, 33
- Prolactin*, 314-318, 330-332
 - chemical modification and activity, 317
 - oxidation, 315
 - physico-chemical properties, 315
 - purification, 314
 - reduction, 316
 - structure, partial, 317-318
- Protamine, fractionation by dialysis*, 107-108

Proteins

- chemical modification with retention of activity
 - antibodies, 290
 - chymotrypsinogen, 159-163
 - corticotropin, 312, 335-336
 - growth hormone, 318-319, 322, 324, 326
 - insulin, 290
 - papain, 193, 197-200
 - pepsin, 179-181
 - serum albumin, 290-294
 - tobacco mosaic virus, 251, 290
 - trypsinogen, 171
- configuration, 25-62
- homogeneity, 93-99, 140-141, 144, 183, 262-263, 304, 308, 314, 316, 323-325
 - immunological criteria, 184, 240-244
- purification,
 - chromatography, 96-103, 197, 199, 241-243, 324
 - countercurrent distribution, 304, 314
 - electrophoresis, 94-97, 140-141, 291, 307-308, 316
 - species specificity, 66-75, 144-147, 305-327
 - structure, genetic control of, 148-151, 247, 262-263, 298-301
 - surface, polarity of, 43
- Pyrrolidone carboxyl peptides*, 268
- Random coils*, in denatured proteins, 48-59
 - in polyelectrolytes, 35-38, 47-49
- Ribonuclease*
 - action of carboxypeptidase on, 233
 - chymotrypsin on, 212, 219-220
 - pepsin on, 233-235
 - subtilisin on, 229, 233, 236
 - trypsin on, 211-220
 - amino acid sequence, 213, 230
 - carbobenzoylation, 224
 - dialysis rate, 108, 110
 - disulphide bonds, oxidation, 211, 219-221, 229, 236
 - reduction, 224-228, 230-232, 236
 - immunological heterogeneity, 241-246

Ribonuclease—cont.

- inhibition by alkaline pH, 236
 - antibodies, 244, 246
 - esterification, 236
 - guanidine salts, 236, 237
 - oxidation, 236, 238
 - pepsin digestion, 233, 235, 238
 - reduction, 230-232
 - optical rotation, 234, 235
 - oxidized, deuterium exchange, 32
 - dialysis rate, 108, 110
 - enzymic hydrolysis of, 211-215, 219, 227
 - reduced and carboxymethylated, 225-227, 230-232
 - enzymic hydrolysis of, 227, 232
 - viscosity, 38, 49, 58, 60, 234-235
- Ribonucleic acid*, 234, 250, 256, 262, 272

'Salt links' in proteins, 42-43*Sedimentation coefficients*, see under individual proteins, physico-chemical properties*Serum albumin*

- amino acid composition, 292
- species variations in, 73, 79
- configuration and pH, 52-54
- diffusion constant, 52
- disulphide bonds, 45, 62
- immunologically active fragment of, 290-294
 - amino acid composition, 292
 - dimerization, 291-293
 - physico-chemical properties, 291-292
 - stability, 293
- polymerization, 62
- sedimentation constant, 52
- tyrosine ionization in, 53
- viscosity, 38, 52

Serum proteins

- in amphibian metamorphosis, 296-297
- as iso-antigens, 298-301

Subtilisin

- action on DNP-tobacco mosaic virus protein, 268
- ribonuclease, 229, 231-233, 236

Sulphydryl groups

- in haemoglobin, 136
- in papain, 184-186, 202-206

- reaction with p-chloromercuribenzoate, 130, 136, 184-186, 205
- iodine, 252
- iodoacetamide, 185, 186, 316
- iodoacetate, 224-232
- methyl mercury, 252
- in serum albumin, 292-293
- in tobacco mosaic virus, 251-252

Titration curves, see Ionization*Tobacco mosaic virus*

- hydroxylamine, reaction with, 266
- metal ion binding by, 255-256
- peptides from proteolysis of, 256-260, 267-269
- protein subunits, 249-251, 263, 271-273
 - bonding of, 253-256
 - molecular weight of, 250-251, 267-269
- stability, 250-251, 256, 263
- strains, differences between, 73, 247, 262-263
- sulphydryl groups, 251-252
- trichloroacetic acid on, 266, 269
- X-ray analysis, 271-273

Trypsin

- action on carbobenzoxyated proteins, 224, 227-228
- chymotrypsinogen, 155-159, 162-163
- corticotropin, 310-311
- enolase, 338-339
- lysozyme, 227-228, 280-285, 289
- melanotropin, 305, 333-334
- papain, oxidized, 188-192
- ribonuclease, 211-221, 227
- tobacco mosaic virus, 257
- trypsinogen, 169-173
- dialysis rate, 108, 110
- reaction with organophosphorus esters,
 - effect of acetylation on, 177
 - effect of urea on, 176-178
- specificity of, 288-289
- C-terminal configuration, 171

Trypsinogen

- activation of, 169-174
 - acetyl derivatives of, 171
 - effect of urea on, 173
- amino acid composition, 81
- basic peptides from, 85-87

- Trypsinogen—cont.*
cysteic acid peptides from, 84
dialysis rate, 108
- Tryptophan in proteins, oxidation of,*
224-227, 283, 285
- Tyrosine in proteins*
hydrogen bonding of, 41, 42, 179,
236-238, 252-255
ionization of, 43, 50, 53, 55, 57, 59,
236, 252
- Ultra-violet absorption spectra*
myoglobin, 146-147
ribonuclease, 59
effect of chemical modification
on, 233-238
tobacco mosaic virus, 252-253
- Urea*
effect on dialysis rate of proteins,
110
enolase activity, 339
pepsin autolysis, 179-181
ribonuclease activity, 233-235
serum albumin, 62
trypsinogen activation, 173
hydrogen bonding of, in solution,
41
- Vasopressin, species differences,* 73
- Viscosity*
of polyelectrolytes, 48
of proteins, 37-39, 52-60
- Water, hydrogen bonding and protein
configuration,* 39, 41
- X-ray analysis*
haemoglobin, 136-139
myoglobin, 125-135
isomorphous replacement
method, 126-132
three-dimensional Fourier syn-
thesis, 132-135
tobacco mosaic virus, 255, 271-273





