ADVANCES IN ENZYMOLOGY

AND RELATED SUBJECTS

Edited by

F. F. NORD
Fordham University
New York, N. Y.

C. H. WERKMAN
Iowa State College
Ames, Iowa

Volume I

With 56 illustrations

1941

INTERSCIENCE PUBLISHERS, INC.

New York

ADVANCES IN ENZYMOLOGY

AND RELATED SUBJECTS

Volume I

CONTRIBUTORS TO VOLUME I

MAX BERGMANN, *The Rockefeller Institute for Medical Research, New York, N. Y.*

HENRY B. BULL, *Department of Chemistry, Northwestern University Medical School, Chicago, Ill.*

JAMES FRANCK, *Department of Chemistry, University of Chicago, Chicago, Ill.*

J. S. FRUTON, *The Rockefeller Institute for Medical Research, New York, N. Y.*

H. GAFFRON, *Department of Chemistry, University of Chicago, Chicago, Ill.*

D. E. GREEN, *Department of Biological Chemistry, Harvard Medical School, Boston, Mass.*

LUISE HOLZAPFEL, *Kaiser Wilhelm Institut fuer Silikatforschung, Berlin-Dahlem*

A. L. KURSSANOV, *Institut für Biochemie der Akademie der Wissenschaften der U.S.S.R., Moscow*

FRITZ LIPMANN, *Department of Biochemistry, Cornell University Medical School, New York, N. Y.*

JAMES B. SUMNER, *Biochemistry Laboratory, Cornell University, Ithaca, N. Y.*

C. B. VAN NIEL, *Hopkins Marine Station of Stanford University, Pacific Grove, Calif.*

H. J. VONK, *Laboratorium voor Vergelijkende Physiologie, Utrecht, Holland*

ADVANCES IN ENZYMOLOGY

AND RELATED SUBJECTS

Edited by

F. F. NORD
Fordham University
New York, N. Y.

C. H. WERKMAN
Iowa State College
Ames, Iowa

Volume I

With 56 illustrations

1941
INTERSCIENCE PUBLISHERS, INC.
New York

Copyright, 1941, by
INTERSCIENCE PUBLISHERS, INC.
215 Fourth Avenue, New York 3, N. Y.

First printing 1941
Second printing (by photo-offset) 1945
Third printing (by photo-offset) 1946

Printed in the United States of America

Preface

This collection of independent monographs is initiated at a time when research is subject to the gravest of interruptions and original thinking liable to the greatest distraction.

ADVANCES IN ENZYMOLOGY may be of service to those investigators who are devoting their efforts to extending our knowledge in this field and related subjects; and to all who are interested in the realm of enzyme behavior.

Meantime, the Editors wish to render tribute, although somewhat belatedly, to the concept of SCHWANN of the order of the cell as a symbol of the harmony and organization of interfacial phenomena, and of organic and physical chemistry.

We express our obligation to the contributors for their unreserved cooperation and to the publishers for their support.

New York, N. Y.
Ames, Iowa,
February, 1941

F. F. N.
C. H. W.

CONTENTS

PAGE

Preface........ v

Contents........ vii

Protein Structure. By HENRY B. BULL, Chicago, Ill........ 1

I. Introduction........ 1
II. Fiber Proteins........ 5
III. Globular Proteins........ 13
Bibliography........ 40

Physikalisch-chemische Gesichtspunkte zum Problem der Virusaktivität. Von LUISE HOLZAPFEL, Berlin-Dahlem........ 43

Einleitung........ 43
I. Einfluss des Ladungscharakters auf die spezifische Viruswirksamkeit........ 45
II. Einfluss von Bestrahlungen........ 48
III. Das Aggregationsproblem........ 50
IV. Gestalt und Teilchengrösse........ 55
Literaturverzeichnis........ 60

The Specificity of Proteinases. By MAX BERGMANN and JOSEPH S. FRUTON, New York, N. Y........ 63

I. Role of Molecular Weight of the Substrate........ 67
II. The Nature of the Linkages Split by Proteinases........ 67
III. Specificity of Gastro-Intestinal Proteinases........ 69
1. Pepsin........ 69
2. Chymotrypsin........ 73
3. Trypsin........ 75
IV. Some General Remarks Concerning the Specificity of Gastro-Intestinal Proteinases........ 76
V. Proteinases and Peptidases........ 78
VI. Specificity of Intracellular Proteinases........ 78
1. Papain........ 78
2. Intracellular Proteolytic Enzymes of Animal Tissues........ 80
VII. The Activation of Intracellular Proteolytic Enzymes........ 85
VIII. Kinetics and Specificity........ 89
IX. Enzymatic Synthesis........ 90
X. Stereochemical Specificity of Proteinases........ 93
Conclusion........ 95
Bibliography........ 96

Metabolic Generation and Utilization of Phosphate Bond Energy. By FRITZ LIPMANN, New York, N. Y........ 99

I. Historical Introduction........ 100
II. Definition of the Term "Group Potential"........ 102
III. Group Potential of Phospho-organic Compounds........ 103
1. Ester Phosphate........ 103
2. Energy-rich Phosphate Bonds........ 105
IV. Chemistry and Distribution of Energy-rich Phosphate Bonds........ 110
1. General Survey........ 110
2. Adenosine Polyphosphate........ 114
3. Phosphoguanidine Linkages (Phosphagens)........ 116

4. Phosphoenol Pyruvic Acid 119
5. Phosphoglyceryl Phosphate 121
6. Acetyl Phosphate 121
V. The Phosphate Cycle 121
1. Primary Phosphorylation 124
2. Transphosphorylation 126
VI. Metabolic Generation of Energy-rich Phosphate Bonds 131
1. Anaerobic Metabolism 132
Pre O/R Transformation Period 133
Post O/R Transformation Period 134
The Oxidation-Reduction Reaction 136
Energy Balance 138
2. Aerobic Metabolism 139
General 139
Pyruvate Oxidation 141
Dicarboxylic Acid Oxidation 142
Resynthesis of Carbohydrate 143
VII. Utilization of Phosphate Bond Energy 148
1. Muscular Contraction 149
2. Absorption 150
3. Transformation of Fructose into Glucose 152
4. Utilizability of Acetyl Phosphate and of Acyl Phosphates for Biosynthesis 152
VIII. Group Transfer as General Metabolic Reaction 154
1. Amination and Transamination 154
2. Transmethylation 155
3. Transaminidation 157
Bibliography 158

The Chemical Nature of Catalase. By JAMES B. SUMNER, Ithaca, N. Y. 163
1. Introduction 163
2. Occurrence of Catalase 164
3. Physiological Role of Catalase 165
4. The Stability of Catalase 165
5. The Inactivation of Catalase by Trypsin 165
6. The Determination of Catalase Activity and Purity 166
7. The Immunochemistry of Catalase 167
8. Can Catalase Be Resynthesized after Dissociation? 167
9. The Absorption Bands of Catalase 168
10. Theories of the Mechanism of Catalase Action 168
11. The Blue Substance Produced from Catalase 170
12. Crystalline Beef Liver Catalase 172
13. Agner's Catalase Preparations 172
14. Is Horse Liver Catalase More Active Than Beef Liver Catalase? 173
15. The Homogeneity of Crystalline Catalase 173
16. Are There Several Catalases, Depending upon the Number of Intact Hemin Residues? 174
Bibliography 175

Enzymes and Trace Substances. By D. E. GREEN, Boston, Mass. 177
Bibliography 196

Photosynthesis, Facts and Interpretations. By J. FRANCK and H. GAFFRON, Chicago, Ill. 199
I. Quantum Efficiency 200
II. Saturation Phenomena 204

1. Saturation in Continuous Light 204
2. The Blackman Period 206
3. Chemical Kinetics of Photosynthesis 209
Theories Using the Hypothesis of the Photosynthetic Unit 209
Theories Explaining the Saturation Phenomena by Back Reaction 212
4. Steady Fluorescence of Green Plants 216
5. Liberation of Oxygen by Illuminated Chloroplasts 219
III. Induction Periods 220
1. Long Induction Periods 220
2. Short Induction Periods 222
3. Fluorescence Outburst During the Induction Period 223
4. Uptake of Carbon Dioxide in the Dark; Experiments with Radioactive Carbon 231
IV. Photo-Oxidation Processes in Plants 232
V. The Metabolism of the Purple Bacteria and van Niel's Theory of Photosynthesis 234
VI. Carbon Dioxide Reduction in the Absence of Oxygen and the "Reduced State" of the Assimilating System in Plants 242
1. Photoreduction with Hydrogen in Algae 243
Adaptation to the Use of Molecular Hydrogen 243
Transition from Photoreduction to Photosynthesis under the Influence of Light 244
The Effect of Oxygen in the Dark and the Oxy-hydrogen Reaction 245
Induction Phenomena 246
The Effect of Organic Substances and of Specific Poisons 247
2. The "Reduced State" in the Absence of Hydrogen 248
3. Theoretical Conclusions 252
VII. The Reduction of Carbon Dioxide in the Dark 255
Bibliography 259

The Bacterial Photosyntheses and Their Importance for the General Problem of Photosynthesis. By C. B. VAN NIEL, Pacific Grove, Calif. 263

I. Introduction 263
II. Early Studies on the Metabolism of the Purple Bacteria 264
III. The Metabolism of Green and Purple Sulfur Bacteria as an "Abnormal" Photosynthetic Process 269
IV. Photosynthesis with Organic Hydrogen Donors and with Molecular Hydrogen 273
V. Objections to the General Concept of Bacterial Photosyntheses, and an Evaluation of the Evidence 282
VI. Consequences 286
VII. Kinetics of the Bacterial Photosyntheses 289
VIII. The Dark Metabolism of the Purple Bacteria 301
IX. The Pigment System of the Purple Bacteria 307
X. The Energetics of the Bacterial Photosyntheses 314
XI. The Chlorophyll-Carbon Dioxide Ratio and Outlook on the Mechanism of Photosynthesis 319
Bibliography 325

Untersuchung enzymatischer Prozesse in der lebenden Pflanze. Von A. L. KURSSANOV, Moscow, U.S.S.R. 329

I. Allgemeine Vorstellungen über die Wirkung der Enzyme in lebenden Zellen 329
II. Bestimmung der Aktivität von Enzymen in lebenden pflanzlichen Geweben 334
III. Bedingungen, von denen die reversible Wirkung der Enzyme in lebenden Zellen abhängt 343

IV. Die physiologische Rolle der Enzyme 347
1. Die enzymatischen Prozesse in den Zellen in ihren Beziehungen zu den Art- und Sortenverschiedenheiten der Pflanzen 348
2. Speicherung von Vorratsstoffen und allgemeine Ertragsfähigkeit der Pflanzen als Resultat reversibler enzymatischer Reaktionen 348
3. Die vorherrschende Richtung der enzymatischen Prozesse im Zusammenhang mit der Entwicklung der Pflanze 352
4. Einfluss der Aschenelemente auf die enzymatischen Prozesse in Pflanzen 356
5. Einfluss der Temperatur auf die Wirkung der Enzyme in lebenden Zellen 361
6. Wasserhaushalt und Dürre-Resistenz 364
V. Zusammenfassung 366
Literaturverzeichnis 367

Die Verdauung bei den niederen Vertebraten. Von H. J. VONK, Utrecht, Holland 371

I. Unterschiede in der Verdauung zwischen Vertebraten und Invertebraten 372
II. Vorbemerkungen über die Anatomie des Verdauungskanals bei den niederen Vertebraten 374
III. Allgemeines über die Verdauung der niederen Vertebraten 378
IV. Die Verdauung bei den Fischen 385
V. Die Verdauung bei den Amphibien 402
VI. Die Verdauung bei den Reptilien 410
VII. Zusammenfassung 412
Literaturverzeichnis 414

Author Index 419

Subject Index 429

PROTEIN STRUCTURE

HENRY B. BULL

Chicago, Illinois

CONTENTS

		PAGE
I.	Introduction	1
II.	Fiber Proteins	5
III.	Globular Proteins	13
	Bibliography	40

The elucidation of protein structure involves the solution of two general problems. The first deals with the quantitative estimation of the amino acids, as well as the non-amino acid portions of proteins. The second has to do with the arrangement of the amino acids in relation to one another and the nature of the forces which maintain this arrangement. We shall present the arguments of this review in three sections. The first is of an introductory nature and contains observations on amino acids and protein prosthetic groups together with a general discussion of the evidence for the peptide linkage being the only important co-valent chemical bond connecting the amino acid residues together. The second part deals with the fibrous protein, while the third part attempts to summarize the knowledge concerning the structure of the globular proteins.

I. Introduction

It hardly seems necessary to enter into the chemistry of the amino acids in any completeness or detail.

Vickery (1) has listed 25 amino acids as having undoubted occurrence in proteins. In addition to these there are 22 amino acids whose status is doubtful. In any case, however, there is a vast range of possible amino acids which can be used for protein synthesis by living tissue. For many of these amino acids there is no reliable method of estimation. Vickery

states that nine amino acids can be satisfactorily determined in a quantitative manner. These are cystine, tyrosine, tryptophane, methionine, aspartic acid, glutamic acid, arginine, histidine, and lysine. There are six amino acids for which more or less satisfactory methods are available, but which have not been extensively employed. There are eight amino acids for which we have no good method, and our knowledge concerning their amounts in proteins is little better than qualitative.

It can be appreciated from this discussion that protein analyses reported in the literature must be evaluated with a critical eye. There are only a few proteins for which more than 70 per cent of the protein has been accounted for, and of the fraction which has been analyzed perhaps not much over half is to be considered as reasonably accurate. Undoubtedly, accurate amino acid analyses could throw much light on protein structure. How much light the analyses now available throw is a matter of speculation.

In passing, a significant fact for protein structure should be pointed out: It has been found that the groups attached to the alpha carbon atom of all naturally occurring amino acids are arranged in the same stero-configuration as they are in *l*-lactic acid (2). There has been a recent controversy (3) on this point. Some workers believed that they were able to show that, in tumor tissue, glutamic acid occurs partly as the unnatural isomer (related to *d*-lactic acid). The paper by Behrens, *et al.* (4), seems to completely refute this claim. It is pointed out in this paper that *l*-glutamic acid is itself racemized in hot hydrochloric acid and, accordingly, the appearance of a small amount of the unnatural isomer is to be expected in the course of hydrolysis of any protein. They conclude that there is no more of the unnatural isomer of glutamic acid in the hydrolysis of tumor protein than can be accounted for on the basis of racemization of the *l*-form.

The question presents itself as to the type of linkage involved in binding the amino acids together in a protein. The early studies of Hofmeister (5) and of Fischer (6) centered attention on the peptide linkage

$$—CH_2—NH—CO—$$

The objection to this type of linkage as being the only important one connecting the amino acids arose from enzymatic studies. The proteolytic enzymes appeared to fall into two well-defined classes: the proteinases which split substances of high molecular weight and the polypeptidases which attack only degradation products of proteins. It is realized now, however, that this does not constitute a valid argument against the importance of the peptide linkage. For example, Bergmann and his associates

(7) have shown that the occasion for the different types of proteolytic enzymes arises, not from different types of linkages, but from the position of the peptide linkage. All proteolytic enzymes hydrolyze the peptide linkage, but it is attacked by a particular proteolytic enzyme only if it occurs in the proper environment of amino acid residues.

The evidence for the peptide linkage being the only important co-valent bond connecting amino acid residues in proteins has been summarized by Vickery and Osborne (8) along the following lines:

1. Native proteins contain little amino nitrogen, while the end-products of protein hydrolysis contain large amounts.

2. The biuret test is characteristically given by substances having the peptide linkage. Upon complete hydrolysis of the protein, the biuret test can no longer be obtained.

3. The peptide linkage is found in other naturally occurring substances, for example, glutathione and hippuric acid.

4. Synthetic polypeptides can be prepared which are hydrolyzable by the enzymes of the digestive tract.

5. Polypeptides have frequently been found in the products of incomplete protein hydrolysis.

6. Hydrolysis of proteins liberates amino acids and carboxyl groups in equivalent amounts.

It is generally accepted at the present time that the peptide linkage is the only important co-valent bond between amino acid residues in proteins. In this connection, the recent theory of Wrinch postulates the existence of rings in the globular proteins which are closed by a different type of co-valent bond. Astbury also postulates a co-valent bond in α-keratin which is different from the peptide bond. These theories will be considered in the course of this review.

While, as we have seen above, a carboxyl group and the α-amino group of each amino acid are involved in the peptide linkage and are thus not reactive as such in the native protein, there are various other groups in the amino acids which are active. These are the side chains of the amino acid residues (the so-called R-groups). The nature of these side chains of the amino acid residues has a profound influence on the chemical and physical properties of a protein. If the free groups of these residues are predominantly carboxyl groups, the protein will have an acid character, while if the amino groups predominate, the protein will have a basic nature. Many of the carboxyl groups have their acid character masked by the formation of an amide. Those carboxyl groups which have not formed an amide together with all amino groups can apparently be titrated by acids and bases.

They are, therefore, free and not involved in co-valent bonds. The nature of the side chains also determines to some extent the hydrophilic properties of a protein; many polar side chains tend to make the protein water-soluble. No doubt, the side chains also play an important role in determining the type of folding of the polypeptide chain or chains in the protein molecule.

While in actual amount amino acids constitute the major portion of any protein, not infrequently there is present a group of a non-amino acid nature. Such groups are called prosthetic groups. These groups are of a diverse kind. Some of them perform an obvious biological function. For example, the hemin of hemoglobin has to do with oxygen transport in blood, while those of the intracellular oxidation-reduction enzymes are the active part of the oxidation-reduction system. On the other hand, the role of other prosthetic groups remains obscure. It has been proposed that some prosthetic groups act as cementing substances holding the protein molecules together in the living cell (9). When the tissue protein is treated with salts, etc., incidental to the preparation of a "pure" protein, the tissue protein is simply being cleaved at the connecting prosthetic groups. An extreme view along somewhat the same line is that there is only one blood serum protein which is called "orosin" and, accordingly, the separation of the serum proteins into several albumins and globulin fractions is an artifact. The reviewer has considerable reservations on this point of view.

It is generally the case that a prosthetic group increases the stability of a protein. Thus, hemoglobin is a much more stable protein than is globin obtained from hemoglobin by the removal of the hemin.

Some prosthetic groups are bound very tightly to their protein. For example, the prosthetic group of egg albumin (10) which is composed of four molecules of mannose and two of glucosamine, together with an unidentified nitrogenous constituent, cannot be removed from the protein without hydrolyzing the protein. On the other hand, the prosthetic group of several of the oxidation-reduction enzymes can be removed with relatively gentle treatment.

The nature of the combination of proteins with phospholipids is an important problem which has not received the attention it deserves. It is, however, a messy one for a chemist to undertake. It is well recognized that an ether extraction of minced tissue or of blood proteins is not capable of removing all the phospholipids present. For example, Sorensen (11) found that only one-fifth of the phospholipids of blood serum could be removed by ether alone. In phospholipid analysis of tissue the ether extract is supplemented by the use of alcohol, but it has not as yet been demonstrated that even the addition of alcohol removes absolutely all of the phos-

pholipid. It is true, however, that by far the largest proportion of the phospholipid-protein association is of a very loose character and does not involve chemical bonding. It is the feeling of the reviewer that it is possible that some of the several protein fractions obtained from blood serum may be traceable to the presence or absence of different amounts of phospholipid bound to the serum proteins.

So far, none of the studies on the natural phospholipid-proteins have differentiated between the phospholipids which might be present. In this connection, Chargaff (12) found that only cephalin formed a water-insoluble complex between *p*H 1.9 and *p*H 11 with the basic protein, salmine. Lecithin and sphingomyelin formed no such complex. Cephalin also formed an isoluble complex at *p*H values 2, 3, and 4 with egg albumin. At *p*H values higher than 4 no complex formed. Lecithin and sphingomyelin did not show complex formation with egg albumin at any *p*H. The behavior of cephalin as contrasted with that of lecithin and sphingomyelin is probably due to the fact that both lecithin and sphingomyelin have isoelectric points around *p*H 7, while that of cephalin is very low and hence cephalin is negatively charged over a very wide range of *p*H and, accordingly, combines with a positively charged protein through a salt linkage.

For any one interested in doing work on the phospholipid-proteins, probably the most readily available protein showing this association is vitellin from egg yolk. There are apparently two well-defined egg yolk proteins: vitellin and livetin (13). The vitellin carries with it the phospholipid and behaves as a globulin. It slowly loses its phospholipid in contact with water, and thereby becomes more and more insoluble in dilute salt solutions.

In order to facilitate the discussion of protein structure, the author pro poses to divide proteins into fibrous and globular as Astbury has done. Under fibrous proteins are included such proteins as collagen, elastin, and the various forms of keratin. Under globular proteins are to be included proteins which do not form fibers. Actually, the distinction comes down to the degree of asymmetry. The fibrous proteins are very asymmetrical and frequently show elasticity, while the globular proteins show much less asymmetry. The distinction is, at best, a matter of a degree and, at worst, does not exist.

II. Fiber Proteins

The animal body makes use of a number of proteins for structural purposes. These are all fibrous proteins and with the exception of muscle

myosin are extracellular. And, again, with the exception of muscle myosin, they are typically water-insoluble. The fiber proteins with the possible exception of the protamines are the simplest proteins which are known. They present, therefore, the best opportunity we have of learning something about protein structure.

In Table I is shown in round numbers the amino acid residues in 100 gm. of dried protein. The accuracy of these analyses is not to be taken too seriously; they are presented simply to show the approximate content of the major component amino acids. The analyses are in no case anywhere near complete.

The first 6 amino acids in Table I give rise to side chain residues which are hydrophobic in nature, *i. e.*, the paraffin chains predominate. The last 5

TABLE I

NUMBER OF AMINO ACID RESIDUES IN 100 GRAMS OF DRY PROTEIN

Amino acid	Silk fibroin	Elastin	Collagen (gelatin)	Keratin (human hair)
Glycine	0.54	0.39	0.35	0.06
Alanine	0.28	...	0.10	0.02
Valine	...	0.12	...	...
Leucine fraction	0.02	0.23	0.05	0.05
Cystine	...	...	...	0.06
Proline	...	0.13	0.15	0.03
Hydroxyproline	...	0.02	0.11	...
Glutamic acid	...	...	0.04	...
Aspartic acid	...	...	0.03	0.05
Arginine	...	...	0.05	0.02
Lysine	...	...	0.04	...

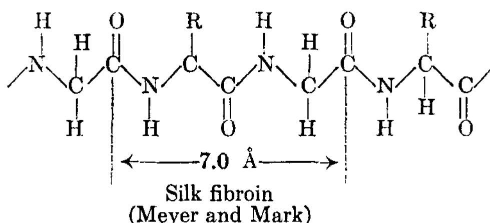
amino acids, on the other hand, have side chain residues which are hydrophilic, *i. e.*, have a polar group in them. It will be seen immediately that, for all the fiber proteins, the hydrophobic groups are more numerous than the hydrophilic ones. In fact, if we sum up the polar residues for a given fiber and divide this into the sum of the non-polar residues, we obtain a kind of a non-polar index of a protein. This has been done using as complete analyses as are available in the literature, and we find the following indices: elastin 29, silk fibroin 19, collagen 3.3, and keratin 2.8. The non-polar character of these fibers is at least part of the reason for the general insolubility in water.

The fact that keratin has the greatest number of polar groups is probably significant in view of the known tendency of keratin to form a well-

defined, folded molecular structure (α-keratin). The presence of cystine (disulfide linkage) is also probably of significance in this connection.

Silk fibroin is the simplest of the fibers with an overwhelming predominance of glycine and alanine. Collagen is characterized by the large amounts of proline and hydroxyproline along with glycine. Elastin may, perhaps, be looked upon as a modified type of collagen. It is very much less polar and, accordingly, it might be anticipated that there would be less attraction between elastin molecules than is the case with collagen, with the consequence that the molecules would have a greater tendency toward random orientation. This is, in fact, substantially what x-ray studies indicate.

Silk Fibroin.—As we have seen, silk fibroin has the simplest chemical composition of the fiber proteins. It also has the simplest structure as revealed by x-ray diffraction studies. It was first shown by Meyer and Mark (14) that the x-ray diffraction pictures of silk fibroin could be interpreted in terms of a stretched polypeptide chain. Silk fibroin is characterized by a spacing along the fiber axis of 7 Å units. This spacing is accounted for on the basis of a unit consisting of glycine and some heavier amino acid residue; the distance between repeating units being 7 Å units (the heavier residue is, for the most part, alanine). The distance of one amino acid residue would, therefore, be 3.5 Å units.



Silk fibroin
(Meyer and Mark)

The distance of the peptide chains from one another in the direction perpendicular to the paper is 4.6 Å units (the so-called "back bone spacing"), and in the direction of the side chain residues of 5.2 Å units.

The diffraction pattern of silk has only a few spots and these spots are not sharp. The amount of crystalline material is not of a high order as compared, say, with cellulose. Mark estimates it to be from 40 to 60 per cent of the material present. The structural picture given above for silk fibroin seems to meet with everyone's satisfaction; it may indeed be regarded as established beyond a reasonable doubt.

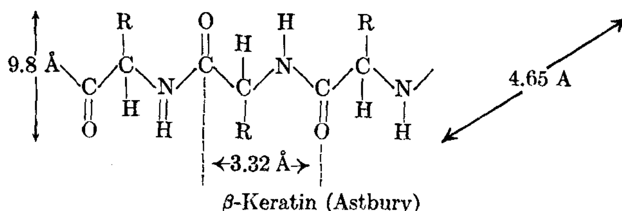
Keratin.—In the keratin group are included all types of hair fibers, the protein of bird feathers, nails, claws, and even porcupine quills. Hair (in-

cluding wools) is distinguished from the other types of keratins by its long-range elasticity. In fact, it may, in general, be stretched by about 300 per cent of its original length and will return to its initial length after the load has been removed. This stretching takes place much more easily in water and in other polar solvents than it does in air or liquids of low dielectric constant (this indicates the importance of electrostatic bonds in maintaining the unstretched keratin in its contracted form). If heat and moisture be applied for a sufficient time to the hair in its stretched condition, the fiber will set and will not return to the contracted state upon release of the force producing the stretching.

The changes associated with the stretching of hair have been investigated in some detail. Astbury has been most active in this field of investigation, and along with his collaborators has published a long series of papers dealing with this topic. Actually, the first paper (15) of this series is the most important one and contains most of the primary information. Astbury was the first to show that the stretching of hair is accompanied by changes in the x-ray diffraction pattern. The normal, unstretched hair is called α-keratin and is characterized by a repeating distance of 5.15 Å units along the axis of the fiber. The back bone spacing is absent, and the side chain spacing is 9.8 Å units. The structure shows a fair degree of orientation (perhaps somewhat less than silk). As the fiber is stretched, the diffraction picture remains unchanged up to about 2 per cent extension, but beyond this extension the diffraction picture of β-keratin begins to appear. The β-diffraction becomes prominent at about 30 per cent extension, and at 60 per cent extension the diffraction picture is entirely of the β-type. The β-keratin is characterized by a repeating distance along the fiber of 3.32 Å units, a back bone spacing of 4.65 Å units, and a side chain spacing of 9.8 Å units. Astbury feels that the stretching of hair cannot be entirely explained on the basis of a transformation of a disorganized, randomly arranged structure to that of an organized, oriented one. Indeed, the repeating space of 5.15 Å units along the unstretched fiber forces the conclusion that there must be an intramolecular folding of a very definite kind. So much seems to be reasonably well established.

Astbury and his collaborators have published structures for α-keratin and for β-keratin which they feel account for the observed x-ray diffraction pictures. The structure for β-keratin is that of a fully stretched peptide chain, the repeating distance of 3.32 Å units being the length of one amino acid residue in direction of the chain.

The R-groups are the side chains of the amino acid residues and their nature depends upon the type of amino acid involved. Note that the R-groups are shown to alternate

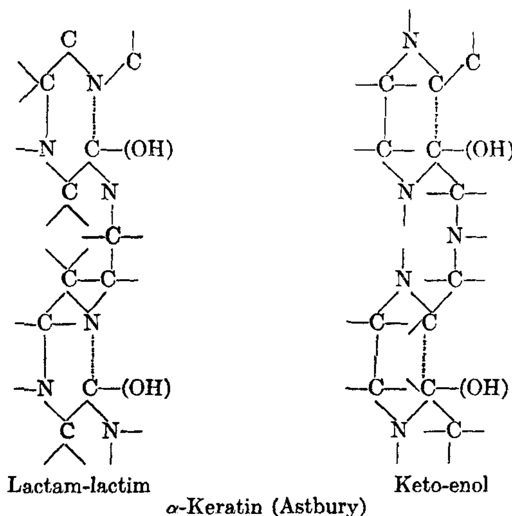


β-Keratin (Astbury)

up and down along the chain; this is a necessary condition in any fully stretched peptide chain and follows from the fact that the amino acids are structurally related to *l*-lactic acid. The distance along the chain of 3.32 Å units per residue is less than that for silk fibroin (3.5 Å units). This discrepancy has not been satisfactorily explained. Corey (91) reports that the distance along a fully stretched chain corresponding to one amino acid residue is 3.67 Å which is greater than that reported for silk fibroin and still greater than that found for β-keratin. Corey states that this fact suggests that in these substances the chain is never fully extended in a truly co-planar configuration but that interactions, steric and otherwise, with its immediate neighbors, cause slight distortions, probably involving rotation about the C—C bond.

The structure proposed by Astbury for α-keratin involves the folding of two amino acids into a six-membered ring. Two types of co-valent linkages were proposed to effect closure of the rings (16). The first type is a lactam-lactim transformation, and the second involves a keto-enol change.

These structures have been criticized by Neurath (17) who has shown rather definitely that they are much too condensed to permit residues other than those of glycine and possibly alanine to occupy the positions called



α-Keratin (Astbury)

for by the carbon atoms of the hexagonal rings to which they are attached.

The source of the embarrassment lies in the valence angle and bond distance requirements. As soon as any peptide chain is folded, all atoms attached directly to atoms in the peptide chain lose their freedom and the R-groups, instead of alternating above and below the chain as they must do in a fully stretched peptide chain (β-keratin), must now all be on the top or bottom of the chains where there is insufficient room to accommodate them.

Huggins (18) has proposed a spiral form of the peptide chain to account for the α-keratin structure. It is claimed that this structure is in accord with all space requirements. The peptide chain is held in its spiral form by means of hydrogen bridges between the oxygen and nitrogen atoms (the hydrogen resonates between the oxygen and nitrogen). This structure has not been published and, accordingly, it seems inappropriate to discuss it at greater length at the present time.

It appears to the reviewer that any structure for α-keratin must recognize two important factors. A definite role must be assigned to the side chains in whatever type of bonding assumed. Why, for example, does silk not show a contracted form analogous to that of α-keratin? This can only be understood when it is remembered that the side chains of silk are typically non-polar, and hence, non-reactive. And, secondly, the effect of neighboring chains must be considered. We, in fact, never deal with the contractions of a single molecular chain, but only with bundles of chains which must be rather closely knit together. A single molecular chain would, if it contained sufficient number of polar groups, probably contract to such an extent that it would approach a spherical form.

In addition to the unstretched and stretched forms of hair keratin, there is still another form of keratin—the supercontracted form. If α-keratin is stretched and then treated with heat and moisture for a short time (not long enough to obtain a set in the β-form) and allowed to contract, it will contract to a shorter length than that of the original unstretched fiber. This is what Astbury calls the supercontracted keratin. The supercontracted keratin shows long range elasticity. The x-ray shows it to be amorphous with little or no orientation. Astbury (19) has proposed a definite molecular folding of a very condensed type. Neurath (17) has shown that such a structure is incompatible with the space requirements of the various groups which are present. The reviewer feels that supercontracted keratin is a case of randomly folded, randomly arranged peptide chains, and that the stretching of supercontracted keratin is closely analogous to the stretching of rubber.

The influence of various groups in keratin on its elastic properties is not, as yet, very well understood, although it is known that deamination of wool

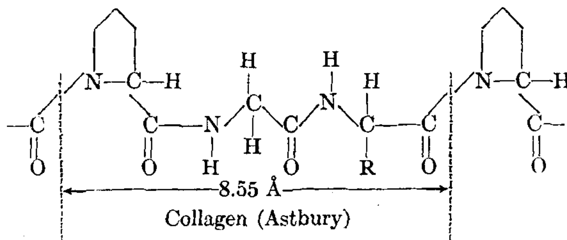
(20) leaves the x-ray diffraction pictures of both α- and β-keratin unchanged which means that the essential framework of the keratin complex remains intact. Deaminated wool will, however, not set in the stretched, β-keratin form. The amino groups are, therefore, involved in the setting of the β-keratin structure.

There has been some speculation (21) about the role of the disulfide linkages in keratin (there is little or no sulfhydryl (SH) present), but the situation remains ambiguous. It is the understanding of the reviewer that Dr. Milton Harris and his collaborators will shortly have some interesting observations to report along this line.

Collagen Group.—Astbury (22) includes in the collagen family the white fibers of connective tissue (ordinary collagen of the leather chemists) tendons, cartilage, the scales and fins (elastoidin) of fishes, the ichthyocol of swim-bladders, the byssus threads of the bivalves such as *Pinna nobilis*, the so-called ova-keratin of egg capsules of the skate, filaments ejected by the sea cucumber, jelly fish, etc. Gelatin, the water-soluble derivative of collagen, is also included. The protein of the red blood cell membrane may also belong to the collagen group.

The essential feature of the collagens is their inelasticity at body temperature, and an x-ray diffraction pattern showing spacings along the fiber of 2.86 Å units (Wyckoff and Corey (23)) give 2.91 Å units as the repeating spacing) as contrasted with the repeating distance of 3.32 Å units of β-keratin. We have already noted that the collagens are characterized by a high content of proline and hydroxyproline.

Astbury has recently proposed a structure for collagen which is given below.



Collagen (Astbury)

The rings are either proline or hydroxyproline and constitute one-third of the total number of residues present. Another third of the residues is glycine. Three residues in this structure will occupy 8.55 Å units along the chain which gives an average value per residue of 2.85 Å units. The reviewer is not clear in his mind as to how the average spacing of a residue could produce a diffraction spacing. It would seem to him that, in analogy

with silk fibroin where two amino acids gave rise to a spacing, the spacing in collagen should be 8.55 Å units instead of 2.86 Å as observed. There are strong spacings along the fiber other than the 2.86 Å unit spacing, but none equal to 8.55 Å units. Some of the Wyckoff and Corey (23) spacings along the fiber are 2.91 Å, 4.03 Å, 7.21 Å, 21.6 Å, etc. The longest spacing they report is 103 Å, while Clark and co-workers (24) have reported a spacing as high as 432 Å.

The above structure for collagen helps to explain the inelasticity of this fiber; it is not easy for this structure to fold, because any folding would result in the side chains swinging over to the side of the back bone where the proline rings lie. The chains, therefore, lie in straight lines, and as such are not capable at ordinary temperatures of becoming elastic.

If collagen be heated to a sufficiently high temperature (about 60° C.), it spontaneously contracts to as little as a quarter of its original length. The contracted fiber, when hot, shows long range elasticity. If it is immersed in cold water after contraction, it spontaneously becomes longer again, but it never recovers completely its initial, uncontracted length, and some of the long range elasticity remains.

Astbury reports that the x-ray photograph of thermally contracted collagen is not simply that of a disoriented crystalline collagen; it is a new photograph altogether—an amorphous pattern. For example, if ordinary collagen be made into a pellet so that the fiber has random orientation in the x-ray beam, it is found that the photograph from this material is much different from that of the contracted collagen, although both show disoriented patterns.

Astbury believes that intramolecular folding has taken place in the contracted collagen. He states that disorientation alone of initially parallel chain bundles cannot account for more than 50 per cent of the observed contraction. It is, however, questionable if any definite and single type of molecular folding is involved. The situation appears to be analogous to that of supercontracted keratin which we have already considered.

Elastin.—Elastin, in contrast to collagen, shows highly elastic behavior at room temperature; biologically, this is as it should be. The diffraction picture of unstretched elastin (*ligamentum nuchae*) is completely amorphous; stretching to 200 per cent extension does not change the type of x-ray photograph (no new spacings appear), but does result in some orientation. Astbury suggests that it is possible that elastin may be a type of collagen which contracts (becomes elastic) at a lower temperature. The chemical analysis gives some color to this suggestion (high proline content), but

nevertheless there remains a considerable difference between the two types of fibers.

The reviewer has found that the walls of the great aorta of a normal human yielded an x-ray diffraction pattern* closely resembling, if not identical with, that obtained by Astbury for the *ligamentum nuchae.* Incidentally, the walls of a highly sclerotic aorta taken from an individual who had died of syphilitic complications gave a diffraction pattern which differed from that of the normal in several important respects; the pathology of the aorta had induced drastic changes in the fiber protein.

Muscle Fiber.—Attempts have been made to study muscles and muscle contraction by means of x-ray diffraction. Myosin, the principal muscle protein, has many of the characteristics of a fiber protein, and Weber (25) has shown that the diffraction pictures of muscle are due to the myosin contained in the muscle. Astbury and Dickinson (26) reported that air-dried myosin fibers could be made to show the α-β transformation and also supercontraction in analogy to keratin. Then later Astbury and Dickinson (27) reported that the α-β transformation could be brought about by stretching muscle itself. No details were given. Still later Astbury (28) states:

"The changes that take place in x-ray photographs of living muscle on isotonic or isometric contraction are essentially similar to the changes observed in the x-ray photographs of keratin in corresponding states. In particular, isotonic contraction, like supercontraction of keratin, results in a disorientation of the grids that is disproportionately small in relation to the contraction observed. In other words, the small disorientation is a secondary effect, the primary cause of contraction being a further folding of the main chains of the grids."

So far as the reviewer is aware, Astbury has published no experimental results on living muscle and, accordingly, it is not possible at the present time to evaluate the role of intramolecular folding in muscle contraction.

III. Globular Proteins

As indicated previously, the term "globular" as applied to such proteins as those of blood serum, egg albumin, and numerous other water-soluble proteins is not free from objection. Many of these proteins are undoubtedly asymmetrical. We shall retain this term, however, since it does convey the idea that this class of proteins shows less asymmetry than do fiber proteins.

The structure of globular proteins is in a much more uncertain state than is that of the fiber proteins; they are much more complicated. Nothing

---

* This work done through the courtesy of Dr. Jack Wilson of the General Electric X-ray Corporation.

is really known as to the arrangement or type of folding of the polypeptide chain or chains in the molecule. We know that within experimental error all of the basic and acidic groups arising from the side chains of the amino acid residues are available for acid-base titration, so that these groups are not involved in any strong co-valent bonds, although they can and, in all probability do, lend stability to the molecule by virtue of their electrostatic charges. Whatever the arrangement of the peptide chain, we have reason to believe that it is a highly specific one and is, in general, unstable; most native proteins in solution, even at room temperatures, are slowly or rapidly undergoing spontaneous denaturation.

It is proposed to discuss the structure of globular proteins under the following four headings: (1) Size (molecular weights); (2) Shape; (3) Possible type of folding and intramolecular arrangement of the peptide chains; (4) Evidence from protein denaturation.

Size.—There has been, and continues to be, a discussion of whether or not one should speak of molecular weights or particle weights in reference to proteins. The reviewer takes the position that the use of the term "molecular weight" in this connection is legitimate.

Without a doubt, the most important tool ever devised for the physical study of proteins has been the ultracentrifuge. This work was pioneered by Svedberg and his co-workers and has since been used by several laboratories in this country and abroad. It would take us too far afield to deal with the technique of the ultracentrifuge, and the reader is referred to the recent book by Svedberg and Pedersen (29). The general theory of the ultracentrifuge seems to be sound enough, and the ambiguities which exist would be expected to give rise to second order errors. The ultracentrifuge gives an anhydrous molecular weight. The reviewer is not yet completely satisfied with the treatment accorded hydration, and it is very possible that hydration does cause small errors. The difficulty, as the reviewer sees it, is involved in the uncertain density of the water of hydration; there has been a volume contraction of the water with the result that the density of such water is greater than that in bulk.

It is almost impossible, by the nature of things, to estimate the actual experimental errors involved in ultracentrifugation determinations. Certainly, the early measurements of the molecular weights of proteins were badly in error. For example, the first determination of the molecular weight of egg albumin was given as 34,500 ± 1000 (30), while the more recent value is 40,500 by the equilibrium ultracentrifugation method and 44,000 by the rate ultracentrifugation method (29). This represents an extreme difference of over 25 per cent between the early and later (rate

sedimentation) determinations. There is still almost 9 per cent difference between the present values of the molecular weight of this protein by the two ultracentrifugation methods. Theoretically, the equilibrium method is on somewhat sounder basis (it is, in part, equivalent to an osmotic pressure method), but the long time required for a determination leaves the equilibrium method open to serious objection. All things considered, the rate sedimentation method seems to be the more reliable. The molecular weight obtained by this method for egg albumin (44,000) and that by the latest osmotic pressure measurements (31) (45,160) are in fairly good agreement and give, if the osmotic pressure method be accepted as the standard, an idea of the errors involved in the more recent ultracentrifuge measurements of the molecular weights of proteins.

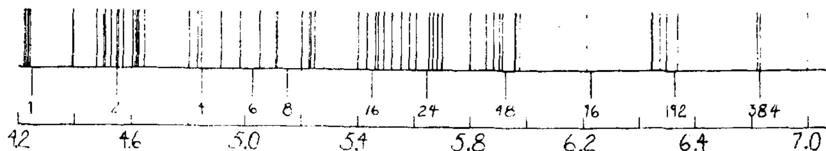


Fig. 1.—A logarithmic "spectrum" of the molecular weights of proteins. The lines of numbers beginning with 1 and ending with 384 are the supposed number of the 17,600 units in the proteins, *i. e.*, the molecular weight classes. The lower line of numbers is an arithmetic scale of logarithms, while the top line indicates the position of the logarithm of the molecular weights taken from "*The Ultracentrifuge*" by T. Svedberg and I. O. Pedersen, Oxford, 1940.

There is one phase of the molecular weight problem which must be critically dealt with. Svedberg and his co-workers (29) have repeatedly claimed that the molecular weights of proteins fall into certain classes. For example, 17,600 is believed to be the unit molecular weight, and the molecular weights of all proteins are supposed to be whole number multiples of this unit. The first class is 1 × 17,600 or 17,600, the second class is 2 × 17,600 or 35,200, the third class is 4 × 17,600 or 70,400 (no proteins are listed which contain 3 units), the fourth class is 6 × 17,600 or 105,000, etc. The most massive protein according to Svedberg contains 384 of the 17,600 units. The idea of molecular weight classes seems to have been more or less generally accepted and has inspired more than one theory of protein structure. It is exceedingly important to decide whether or not the classification of proteins on this basis is a valid one.

Clearly, if the division of the proteins into molecular weights such as outlined above is in accord with reality, there must be some guiding prin-

ciple back of protein synthesis which determines the molecular size irrespective of the type of animal or plant tissue in which they originated. In short, it throws the burden on the chemical nature of the amino acids. On the face of it, it would seem most unlikely that such a molecular weight classification could be true. For example, what guiding principle could exist which would make zein from corn (molecular weight 40,000) and egg albumin from a hen (molecular weight, 45,160), differing as they do in amino acid content, belong to the same molecular weight class?

The author regards the classification of proteins on a molecular weight basis with extreme reservation. In Fig. 1 is shown a logarithmic "spectrum" of the molecular weights of proteins. The molecular weights have been taken from "the ultracentrifuge" by Svedberg and Pedersen (29). The rate sedimentation values have been used where available.

As far as can be told by inspection of the figure there are no molecular weight classes. There is an apparent tendency for certain molecular weights to cluster around 17,000, but no one knows how many "proteins" have smaller molecular weights than 17,000; the clustering may simply indicate that there are a large number of small molecular weight proteins in nature. The reviewer is not impressed by the finding that the higher molecular weights seem to be approximate multiples of 17,600; all numbers, if they are large enough, are approximate multiples of 17,600. It would be most interesting to analyze the molecular weight distribution statistically and see if the distribution departs significantly from a random one. Until this has been done and a significant difference found, the reviewer regards the classification of proteins on a molecular weight basis as unworthy of serious consideration; the burden of proof is on the people who say that such a classification is possible.

The intriguing question arises as to why proteins have definite molecular weights. Why not a distribution of sizes such as one finds in the case of a gold sol? To the reviewer, the answer to this problem is to be found more in physiology than in chemistry. The synthesis of proteins is brought about by enzymes in tissues under highly specific conditions. The recent work of Tiselius and Eriksson-Quensel (32) on the hydrolyzing action of pepsin on egg albumin appeals to the reviewer as being most suggestive. These workers found that the all-or-none principle was involved; the protein molecules were either not attacked or else were broken into fragments of molecular weights of about 1000. It is not inconceivable that in tissue where uniform conditions prevail certain enzymes could synthesize such units, the size and nature of which would be severely controlled. These

units could then be combined by another enzyme system to form a protein which contained a definite number of units.

Bergmann and Niemann (33) bring up the point that if the synthesizing enzymes are proteins (which they probably are), there must be an enzyme capable of synthesizing the synthesizing enzyme or else they must be capable of synthesizing themselves.

Block (34) believes that the basic amino acids, arginine, lysine, and histidine are the controlling amino acids as far as protein synthesis is concerned, and that the ratios of these basic amino acids to each other determine the type of protein which results. Block cites as evidence the similarity of proteins with similar basic amino acid ratios. The basic amino acid complex is called an "Anlage." Bergmann and Niemann (35) find two objections to Block's hypothesis. They point out that it is strange that if the "Anlage" has so decided an effect on the structure of a protein, it is difficult to understand why wool keratin and silk fibroin, which supposedly have the same "Anlage," have, in fact, so different a general structure. Second, the variations noted in the experimental ratios are greater than would normally be ascribed to an experimental error; in general, the constancy of the ratios that would be expected if the "Anlage" theory were correct is apparently not observed.

We wish to consider briefly the association-dissociation of proteins. For many years it has been realized that the serum proteins are capable of considerable interaction (36). This early work led Sorensen to his investigations on this subject and finally to his concept of proteins as reversibly dissociable component systems. Sorensen (37) summarizes his studies in the following manner:

> "Our investigations cover only soluble proteins, *i. e.*, proteins which may be dissolved in water or alcohol with or without the presence of salts, at neutral, acid, or alkaline reaction, without suffering irreversible decompositions. These substances are represented by the ordinary formula $A_x\ B_y\ C_z \ldots$, where A, B, and C, etc., indicate entire complexes, namely polypeptides, whereas the subjoined indices x, y, z, etc., indicate the numbers of the said complexes contained in the whole component system. Within each complex all the atoms or atom groups are interlinked by main valences, whereas the various complexes or components are reversibly interlinked by means of residual valences."

Sorensen was led to his conclusions through solubility studies. He found, for example, that by salt fractionation it was possible to obtain from many times recrystallized serum albumin various crystalline serum albumins of differing physical and chemical compositions which, when combined in the proper proportions, again yielded a crystalline protein having the prop-

erties and composition of the original. It is a question, however, whether these experiments mean what Sorensen thought they meant. In the first place, as far as the experiments with serum albumin are concerned, Hewitt's (38) very careful work has shown that serum albumin can be separated into two definite, well-characterized fractions. One of these fractions is a carbohydrate-rich protein with high solubility which does not crystallize, and the other is a protein with lower solubility and no carbohydrate which can be crystallized easily. Thus, there is no need to regard serum albumin as a co-precipitation system which results from the labile dissociation and rearrangement of the components of a single protein complex. Furthermore, Steinhardt (39) has pointed out that the solubility of proteins can be greatly influenced by a relatively small percentage of a low molecular weight impurity. The extraction of several protein "fractions" from what appears to be a pure, crystalline protein is not to be wondered at and is not to be regarded as a fractionation of various molecular species of the protein.

Quite apart, however, from uncertain indications from solubility experiments, there is a definite evidence from ultracentrifugation studies that proteins can, and do, undergo association and dissociation and with a fair degree of ease. Many proteins have *p*H stability regions outside of which the protein decomposes. In the initial phase, at any rate, this is a definite and stoichiometric affair. In general, the protein splits, if it does split, into halves and then into quarters and then into eighths, etc. Sometimes these dissociations are reversible. Other agents besides protons and the hydroxyl ion will bring about dissociation such as salts, urea, etc. Proteins themselves influence the state of division of other proteins in solution. Dilution also has its influence on the state of aggregation. The book (29) by Svedberg and Pedersen should be consulted for a fuller discussion. It is clear that in some proteins the bonds joining two halves or fourths or eighths, etc., of the molecule together are very weak and easily broken.

There are a number of physical measurements in addition to the ultracentrifuge which can yield protein molecular weights. Among these is osmotic pressure. The reviewer has found that it is possible to make osmotic pressure measurements of protein solutions which are accurate (small experimental variations). If the protein can be isolated in a pure state, it is probably the most unambiguous method for determining the molecular weights of proteins. Not all the osmotic pressure measurements reported in the literature have been done with sufficient care. The results reported by Adair and collaborators are to be recommended. The osmotic pressure

method has been generally neglected, but in the future, with the availability of purer proteins, it will, no doubt, find the extensive usage it deserves.

The size of the protein molecule can also be determined by means of x-ray diffraction studies. The dimensions of the unit cell of the protein crystal can be measured with great accuracy, and if the number of molecules in the unit cell are known along with the density, the hydrated molecular weight can be directly calculated. (It is usually more convenient to make measurements on wet crystals than on dry ones.) The amount of water in the crystal can be determined by drying, and the anhydrous molecular weight estimated. Actually, the order of accuracy is apt to be low; the uncertainty arises from the density measurements. It might be thought that this method would give information concerning the very important question of protein hydration and, as a matter of fact, Sponsler (40) has used x-ray diffraction studies in a very able and convincing manner to estimate the hydration of gelatin (35 per cent water of hydration), but with crystals of globular proteins there is considerable water of crystallization which bears no relation to the hydration of dissolved proteins. The molecular weights of these proteins which have been determined by x-ray diffraction studies (41) (insulin, chymotrypsin, pepsin, hemoglobin, and edestin) are in general agreement with the Svedberg values.

Diffusion measurements alone are not sufficient for molecular weight determinations, as they are complicated by hydration and molecular asymmetry.

On various occasions chemical analyses have been used to estimate protein molecular weights. Quite early, for example, the iron content of hemoglobin molecule was rather accurately determined (42). On the assumption that one hemoglobin molecule contained one iron atom, it was possible to calculate by simple proportion the molecular weight of hemoglobin. It was found that the minimal molecular weight of hemoglobin was about 16,000. It has since been shown that each hemoglobin molecule has four iron atoms and, accordingly, the true molecular weight is four times the minimal molecular weight. By the same token an accurate analysis of any one amino acid would yield a minimal molecular weight. The objection to this type of calculation is clear to be seen; there is no way of knowing by what factor the minimal molecular weight must be multiplied in order to give the true molecular weight.

More recently, Bergmann and his collaborators (43) have attacked this problem with considerable ingenuity and have devised a scheme by which they calculate the total number of amino acid residues in a protein molecule. When the total number of residues is multipled by the average

residue weight, the molecular weight is obtained. The Bergmann theory goes considerably deeper than a mere calculation of the molecular weights of proteins; it deals also with the arrangement of the amino acid residues in relation to one another. It seems appropriate, however, to consider the theory at this point.

The principle of Bergmann's calculation depends upon the assumption of a fixed and invariant periodicity of occurrence of amino acid residues in the peptide chain. It seems unnecessary to give details of their calculations; they can be found in easily available journals. For those proteins investigated, they find the total number of amino acid residues to be 288 or a whole number multiple thereof. Bergmann generalizes this finding and concludes that all proteins have 288 residues or a whole number multiple thereof. He further believes that the molecular weight classes reported by Svedberg are a reflection of the 288 unit. That is, the 35,200 class has 288 residues, while the 70,400 class has 576 residues, etc. The variation of the molecular weight within any given class is caused by variation of the average residue weight.

While the number 288 is not a necessity for Bergmann's theory, it is a great convenience. The individual frequency as well as the individual number of residues of any given amino acid must be exactly divisible into the total number of residues, and it so happens that 288 has the largest number of exact divisors of any number from 0 to 576. By assuming a total number of residues of 288 rather than some other number he has to do less violence to the analytical figures in order to make them fit his theory.

The reviewer has the following comments to make concerning Bergmann's theory:

1. As pointed out previously, Bergmann's theory requires a regular and invariant periodicity of occurrence of the amino acids in a single peptide chain (there can be more than one peptide chain in the molecule providing they are identical). While there is some evidence from x-ray studies which indicates a definite periodicity of amino acid residues in fiber proteins, for globular proteins there is no evidence either direct or indirect for such periodicity.

2. The calculation of the average residue weight is attended with considerable uncertainty. The method used obtains the average residue weight of the amino acids for which analyses are available; there is no way of knowing the average residue weight of the amino acids which have not been determined. The uncertainty over the average residue weight is important in two ways: First, if the average residues weight is in error, the calculated frequencies of the individual residues will be incorrect, which

means in turn that the total number of residues will also be in error. Second, the error in the average residue weight will be compounded in the molecular weight because the molecular weight is obtained by multiplying the total number of residues by the average residue weight.

3. It is very doubtful if the present analytical results for proteins are sufficiently accurate to be used in the fashion in which Bergmann uses them. This point has been clearly brought out by Neuberger (44). Incidentally, the recent analysis of gelatin by Bergmann and Stein (45) for proline and glycine leads to a frequency along the peptide chain of 3 for glycine and 7 for proline. Evidently there would be a conflict between these two acids every 21st position along the chain if the two acids are to have invariant periodicities in the chain.

4. If one uses the molecular weights of proteins as obtained from physical measurements and estimates the average residue weight from analytical results, the total number of residues can be directly calculated. Such calculations can be made for only a comparatively few proteins, because the appropriate data are not available for more. The results from such calculations are shown in Table II.

TABLE II

Protein	Molecular weight	Average residue weight	Total calculated	No. of residues demanded by Bergmann
Lactalbumin	17,400	120	145	144
Egg albumin	43,960*	124	352	288
Insulin	41,000	124	330	288
Horse hemoglobin	65,000†	119	552	576
Edestin	310,000	119	2610	2304

* The carbohydrate prosthetic group has been subtracted.

† The hemin has been subtracted. Cattle globin cannot be given because the molecular weight of cattle hemoglobin has not been determined. This is to be regretted because it will be recalled that cattle globin was one of the proteins which seemed best in accord with Bergmann's treatment.

Recently, Hotchkiss (92) has developed a method for estimating the average residue weight of the amino acids by titrating the acid and basic groups before and after hydrolysis of a given weight of protein. He applied this technique to lactoglobulin and Linderstrøm-Lang (93) has since applied it to insulin. These workers report an average residue weight of 116.7 for lactoglobulin and 123.2 for insulin. Using the rate sedimentation molecular weight as reported by Svedberg for these proteins, it is found that the lactoglobulin molecule contains 349 and insulin molecule 332 amino acid residues.

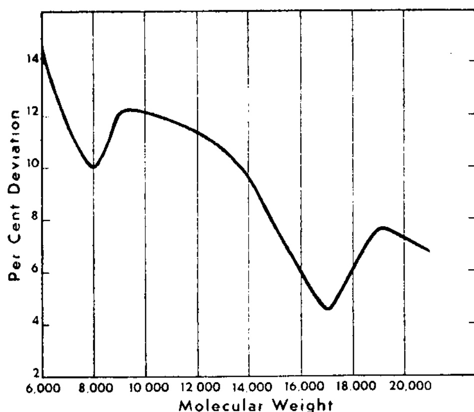


Fig. 2.—The average per cent deviations from whole number residues for lactalbumin plotted against assumed molecular weights of this protein. Calculations made at thousand molecular weight intervals. Data taken from chapter by H. O. Calvery in "*Chemistry of the Amino Acids and Proteins*" by C. L. A. Schmidt, Springfield, Ill., 1938.

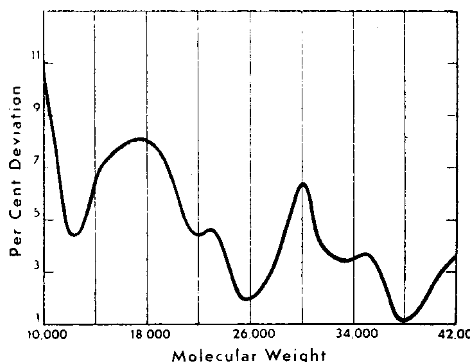


Fig. 3.—The average per cent deviations from whole number residues for insulin plotted against assumed molecular weights of this protein. Calculations made at thousand molecular weight intervals. Data from Vincent du Vigneaud, *Cold Spring Harbor Symposia on Quantitative Biology*, **6**, 275 (1938).

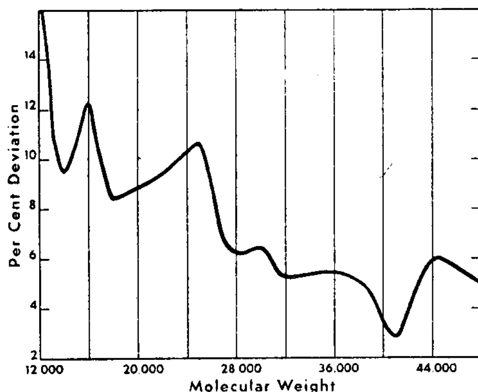


Fig. 4.—The average per cent deviations from whole number residues for hen's egg albumin plotted against assumed molecular weights of this protein. Calculations made at thousand molecular weight intervals. Data from H. B. Vickery, *Symposium on Proteins*, New York Academy of Science, February 2, 1940 (in press).

5. Finally, basic to Bergmann's whole theory is the concept of molecular weight classes and the idea that these classes represent quantized residue numbers. We have already stressed the uncertainty involved in the classification of proteins according to molecular weights.

The reviewer has started out with the simple hypothesis that the only requirement is that the individual amino acid residues must occur as whole numbers (there cannot be fractions of residues). This is equivalent to assuming that there is only one molecular species of protein present. The technique followed is to calculate from chemical analyses the number of residues of the various amino acids present for a series of arbitrary molecular weights. The number of residues for each amino acid for each molecular weight are then changed to the nearest whole number, and the per cent changed is calculated for each amino acid. The per cent deviations from whole number of residues for all the amino acids for each molecular weight are then averaged and these averages plotted against the corresponding assumed molecular weights. Evidently the molecular weight corresponding *to the minimum* average deviation from whole numbers should be the molecular weight best in accord with the chemical analysis. Such plots are shown in Figs. 2, 3, 4, 5, and 6 for lactalbumin, insulin, egg albumin, cattle globin, and horse hemoglobin.

In all cases the molecular weight can only be obtained if it is already approximately known. It is interesting, however, that there is never any doubt as to which minima and, accordingly, which molecular weight to select. Another curious point is that the minima show a certain symmetry of occurrence. For example, lactalbumin has two minima, one at about 8000 and another at 17,000. Insulin has a symmetry of 3 (interestingly enough, x-ray *diffraction studies* also show a symmetry of 3). Cattle globin shows 12 minima (one is not shown on graph). Horse hemoglobin also shows 12 minima (two not shown on graph). Egg albumin shows 6 minima (one not shown on graph). The minima of this protein are, however, very unsymmetrically placed. The reviewer does not propose any theory of protein structure based upon these results—the whole thing may be a curious coincidence. This method of calculation aims to present the results of protein analyses in a graphic manner. When better analytical values are available, it may allow valuable conclusions to be drawn from such data.

Shape.—We know next to nothing regarding the actual shape of protein molecules. The best that we can do at the present time is to make estimates of the asymmetries of the molecules. In most cases we do not even know whether the asymmetry which we estimate takes the form of rods or

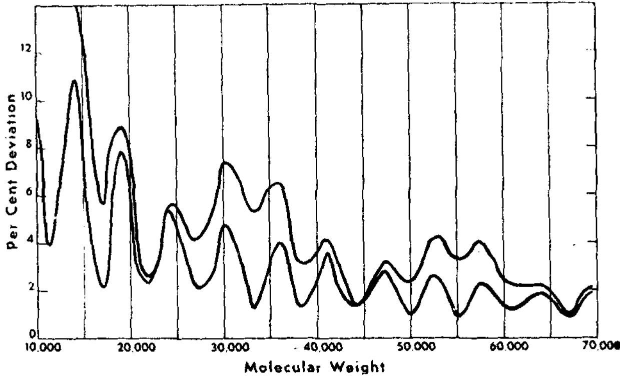


Fig. 5.—The average per cent deviations from whole number residues for cattle globin plotted against assumed molecular weights of this protein. Calculations made at thousand molecular weight intervals. The upper curve includes the cysteine analysis; the lower one does not. Data from Carl Niemann, *Cold Spring Harbor Symposia on Quantitative Biology*, **6**, 58 (1938).

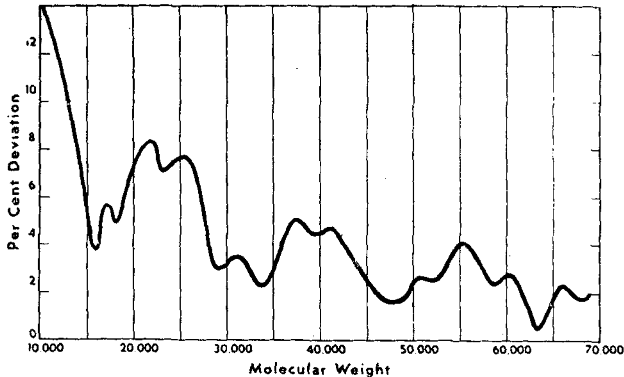


Fig. 6.—The average per cent deviations from whole number residues for horse hemoglobin plotted against assumed molecular weights of this protein. Calculations made at thousand molecular weight intervals. Data from H. B. Vickery, *Symposium on Proteins*, New York Academy of Science, February 2, 1940 (in press).

of disks, although in the case of tobacco mosaic virus protein Lauffer an Stanley (46) have shown rather definitely that this very asymmetric pro tein is in the form of rods. Asymmetry is expressed in terms of the ratio

of the major and minor axis of a prolate or of an oblate ellipsoid of revolution. The protein molecule may not be ellipsoidal at all but, as a first approximation, this is probably a satisfactory assumption.

The problem of asymmetry is complicated by hydration. It is known, for example, that asymmetry of particles increases viscosity and decreases diffusion from what one would expect from a solution of spheres of the same molecular volume. Hydration also increases viscosity and decreases diffusion of molecules in solution. It is thus difficult to know to what extent hydration and asymmetry enter into any particular cases.

Perrin (47) as well as Herzog, Illig, and Kudar (48) have derived equations for the relation between the asymmetry of particles and the ratio of the experimental diffusion constant to that of spheres of the same volume. These equations express the asymmetry as the ratio of the major and minor axis of a prolate or of an oblate ellipsoid. They blame the entire departure of the observed diffusion constant from that of a spherical molecule of the same volume on asymmetry. There is every reason to believe that this is, in fact, not true, but that part of the departure is due to hydration. The question arises as to how much of the departure is due to hydration and how much to asymmetry. Consider the case of egg albumin.

The anhydrous molecular weight of egg albumin is 45,160 (31). It is hydrated in solution to the extent of about 36 per cent (49); accordingly, the hydrated molecular weight of egg albumin is 61,440. The density of egg albumin containing 36 per cent water is 1.202 (49) and, therefore, the hydrated molecular volume is 51,110 cc. If the egg albumin is assumed to be spherical, the radius of the molecule is calculated to be 2.72×10^{-7} cm. Substituting this value in the Sutherland-Einstein diffusion equation for spherical particles, we find the diffusion constant of hydrated, spherical egg albumin molecules in pure water and at 20° C. to be 7.79×10^{-7} sq. cm. per second. By the same method of calculation, the diffusion constant of unhydrated spherical egg albumin molecules (molecular weight of 45,160 and density of 1.34) in pure water and at 20° C. is 8.96×10^{-7} sq. cm. per second. The experimental value for the diffusion constant at 20° C. and pure water is 7.76×10^{-7} sq. cm. (50) per second. The agreement between the diffusion constant calculated for hydrated, spherical molecules and the experimental value is indeed remarkable and indicates that the egg albumin molecule is substantially spherical.

Neurath (51) has used the Perrin equation to calculate the asymmetries of various proteins. He made these calculations with the full realization that they represented first order approximations and are subject to change as knowledge is accumulated. Table III shows the asymmetries calcu-

lated for anhydrous molecules using the technique of Neurath. Also shown are the asymmetries calculated on the assumption of a weight hydration of 35 per cent. In order to carry out these last calculations, it was necessary to make some assumption regarding the density of the water of hydration. This was assumed to be 1.00. This treatment of hydration is admittedly open to serious objection, but it seems more realistic to handle it this way than to neglect it altogether. The assumption of 35 per cent hydration has some basis in fact, thus Sponsler (40) found gelatin to be hydrated to the extent of 35 per cent, and the Neurath-Bull value of 36 per cent hydration for egg albumin has already been discussed. On the other hand, 35 per cent hydration for the higher molecular weight proteins is probably too high unless the molecules actually swell.

TABLE III

ASYMMETRIES CALCULATED FOR ANHYDROUS PROTEINS AND FOR PROTEINS HYDRATED TO THE EXTENT OF 35 PER CENT

(Diffusion data have been used and Perrin's equation for prolate ellipsoids employed.)

Protein	Molecular weight	Asymmetries	
		Anhydrous	Hydrated
Cytochrome c	15,600	5.4	3.1
Lactalbumin	17,400	3.6	1.5
Hordein	27,500	11.5	8.0
Gliadin	27,500	11.0	7.3
Bence Jones	37,000	5.8	3.5
Zein	40,000	30.0	21.0
Insulin	41,000	3.2	1.0
Lactoglobulin	41,500	5.0	2.8
Egg albumin	45,160	3.6	1.2
Hemoglobin (horse)	68,000	4.2	2.6
Serum albumin (horse)	70,000	4.9	2.8
Diphtheria toxin	72,000	4.6	2.6
Serum globulin (horse)	167,000	7.8	4.7
Antipneumococcus serum globulin (man)	195,000	9.5	6.4
Antipneumococcus serum globulin (rabbit)	157,000	9.5	6.2
Catalase	250,000	4.8	3.4
Excelsin	295,000	3.2	1.0
Edestin	310,000	4.2	2.1
Amandin	330,000	5.2	3.1
Urease	480,000	4.0	1.8
Thymoglobulin (pig)	630,000	7.7	4.8
Antipneumococcus serum globulin (horse)	920,000	15.5	11.0

Hydration reduces the apparent asymmetries in all cases. The large molecules show as much reduction as the small ones. As has been pointed out, however, 35 per cent hydration for the larger molecules is probably excessive and, accordingly, the asymmetries of the larger molecules, taking hydration into consideration, are probably closer to the anhydrous figures. The correction of asymmetries, due to hydration, is smaller for more asymmetrical molecules than for the less asymmetrical ones, but in no case can the effect of hydration be neglected.

Any theory of protein structure has to take into account the fact that many of the globular proteins are not symmetrical, notably zein, gliadin, hordein, and the various serum globulin fractions. Also to be noted is variation of asymmetries from protein to protein. Even proteins of approximately the same molecular weights can show many different asymmetries. This indicates to the reviewer that there is no structure common to all proteins, but each has its own individual structure. At the present time it is not possible to interpret the apparent asymmetries in terms of any of the other properties of the proteins, although there does seem to be a small positive correlation between asymmetry and the ratio of the non-polar residues to the polar residues.

The fact that most water-soluble proteins approach a spherical shape is to be understood. Certainly, a long peptide chain containing many polar groups would be expected to fold up into more or less of a globular protein.

Neurath (51) has also considered the asymmetries of protein association and dissociation products. He finds that it is possible to visualize rather clearly the direction in which the protein molecule splits upon dissociation. Some proteins split along the major axis, others along the minor axis. The effect of hydration was not considered, but as can be seen from the results from Table III hydration will decrease all asymmetries, but it will decrease the smaller asymmetries the most, so that taking hydration into account would tend to accentuate Neurath's results and hence make his conclusions more certain.

Viscosity studies have been used to estimate the molecular asymmetry of proteins. Viscosity is, however, inherently a more complicated phenomenon than diffusion. There have been a number of equations derived which relate asymmetry of the particles to the viscosity of a solution of the particles. The latest and, in some respects, the most complete is that of Simha (52). Frankly, the reviewer has considerable reservation in regard to all viscosity equations so far published. It is exceedingly difficult to test these equations experimentally. Polson (53) derived an equation which is empirical to the extent that the equation was formulated by using

the asymmetries obtained from Perrin's diffusion equation (neglecting hydration). He plotted the square of the asymmetries obtained by the use of Perrin's equation against the specific viscosities and found a straight line relation (this indicates protein molecules to be rigid). He formulated an equation to express this straight line relation. Values for asymmetries obtained by the use of Polson's equation are at least consistent with diffusion studies. Polson's equation and Simha's equation give results which are in good agreement for asymmetries greater than about 7 to 1 (80), but for asymmetries less than this the agreement between the two equations becomes progressively worse. A curious fact is that for spherical particles the Polson equation extrapolates to an Einstein coefficient of 4.1 instead of 2.5 as demanded by Einstein's viscosity equation. Such a large discrepancy cannot be explained on the basis of hydration.

There are other ways of estimating asymmetry. If the protein solution shows stream double refraction, the protein molecules are undoubtedly asymmetric. The length of the protein molecule can be estimated from such measurements. Unfortunately, the degree of asymmetry must be considerable before stream double refraction appears, unless one has relatively enormous flow gradients.

A departure of the dispersion curve of the dielectric constant as a function of frequency from the Debye curve can be used to calculate asymmetry, but again we must consider hydration.

X-ray diffraction studies of crystalline proteins give us the dimensions of the unit crystal, although most crystals have more than one molecule per unit cell and, accordingly, it is impossible to determine the molecular asymmetry in such cases without additional data. Fankuchen (41) reports from x-ray studies that horse hemoglobin has the form of a probate ellipsoid with an asymmetry of about 2 to 1. Unfortunately, as Fankuchen points out, proteins with a high degree of asymmetry will not crystallize in a three dimensional crystal, and it is not possible to test the viscosity and diffusion equations involving the appreciable asymmetries by means of x-ray diffraction studies.

Arrangement of the Peptide Chain.—One of the central problems which the protein chemist must solve is that of the arrangement of the peptide chain or chains within the protein molecule. This is a problem for which, at the present time, there is no effective approach. We must content ourselves almost entirely with speculation. As long as we realize that what we say is almost pure speculation, no harm will come to us, but we must be careful not to take ourselves too seriously.

It seems evident that in the globular proteins the peptide chains are