

INTRODUCTION TO



Protein Chemistry

FOSTER

GEORGE WILLIAMS



COLLEGE

LIBRARY

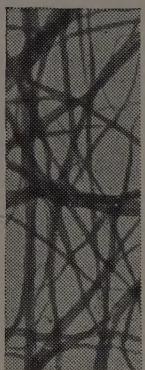
Downers Grove, Illinois

GEORGE WILLIAMS COLLEGE
LIBRARY
5315 DREXEL AVENUE
CHICAGO 15, ILL.

INTRODUCTION TO

Protein Chemistry

INTRODUCTION TO



New York · John Wiley & Sons, Inc.

London · Chapman & Hall, Limited



Protein Chemistry

Sidney W. Fox

*Professor in the Chemistry Department
Florida State University
and
Director of the Oceanographic Institute*

and

Joseph F. Foster

*Professor of Chemistry
Purdue University*

THE MASTER'S COLLEGE
POWELL LIBRARY
SANTA CLARITA, CA 91321

GEORGE WILLIAMS COLLEGE
LIBRARY
5315 DREXEL AVENUE
CHICAGO 15, ILL

547.75
F 8334

36623

Copyright © 1957 by John Wiley & Sons, Inc.

All Rights Reserved. This book or any part thereof must not be reproduced in any form without the written permission of the publisher.

Library of Congress Catalog Card Number: 57-13442

Printed in the United States of America

Preface

The need, in a course in protein chemistry, for a readable, integrated, and relatively comprehensive introduction to the field was in part responsible for the initiation of this book. Our aim has been to write a text that would provide an outline of knowledge, freeing time for more intensive treatment of selected subject matter in class.

Although the objective in the present book is primarily the presentation of the rudiments of knowledge in the field, we have attempted to avoid superficiality, and it is hoped that those specializing in the field may find much material of value. Some reported findings must necessarily be excluded in a book of this size. This is no reflection on the particular work omitted. More and more, advances from the frontiers of protein chemistry depend upon tentative interpretations which are offered by the investigator as subject to agreement or disagreement with later results and the intellectual products from other laboratories. This approach in the accumulation of knowledge is, in fact, increasingly a part of scientific methodology. Some of the eventually most significant discoveries and generalizations are to be found in this "wait and see" limbo. In the present book, an attempt has been made to present tentative or highly controversial concepts with much qualification, or to omit them altogether.

A difficulty of another sort in the preparation of a book in a fast-moving field such as protein chemistry is the time lag in publication, which outdates the knowledge on particular points with almost hopeless rapidity. In the time that is required to execute a final revision of such a book, new data make another revision desirable. If one insists on perfection in this regard, the book may never appear and whatever value it has will be lost.

We hope this book fulfills the purpose of being accurate as well as understandable. As a check on clarity, a number of graduate students without previous orientation in the subject have read the text. For the purpose of accuracy, a number of specialists have read individual chapters. We wish especially to thank Dr. Robert Baker, Dr. Arthur Cherkin, Dr. John T. Edall, Dr. Dexter French, Dr. John Gowen, Dr. Stanley Marcus, Dr. Richard E. Maxwell, Dr. J. L. Oncley, Dr. E. L. Powell, Dr. E. G. Samsa, Dr. Emil Smith, Dr. Pearl Swanson, and Dr. Milton Winitz. The entire manuscript was not so reviewed, however, and further corrections will be appreciated.

In accord with the primary goal in writing this book, it has been deemed advisable to document all statements with specific literature citations. The reading references at the end of each chapter might be more accurately titled "References with Which the Authors Happen to Be Acquainted or Have Located for This Book, and Which Appear to Be Appropriate to Clarifying or Extending the Subject Matter." The repetition of such a title at the end of each chapter would be too lengthy in an economic period in which prices of books are more than the publishers, authors, or readers would like them to be. The authors have therefore chosen to use instead the heading Selected References.

Aside from these references, a number of books suitable for advanced reading in the field can be recommended. These include: Cohn and Edsall, *Proteins, Amino Acids and Peptides*; Haurowitz, *Chemistry and Biology of the Proteins*; Greenberg, *Amino Acids and Proteins*; and Neurath and Bailey, *The Proteins*.

Particular thanks are due to Dr. Felix Haurowitz for his generous suggestions of emphases to be followed in this book.

SIDNEY W. FOX
JOSEPH F. FOSTER

Tallahassee, Florida

Lafayette, Indiana

September 1957

Contents

1	Scope of Protein Chemistry	1
2	Description of the Amino Acids	5
3	Physical Properties of the Amino Acids	20
4	Chemical Properties of the Amino Acids	49
5	Preparation of the Amino Acids	65
6	Assay of the Amino Acids	85
7	Metabolic and Nutritive Significance of the Amino Acids	105
8	Naturally Occurring Peptides and Their Fractionation	131
9	Determination of Structure of Peptides	144
10	Synthesis of Peptides	160
11	Amphoteric Properties of Proteins	175
12	Electrophoresis of Proteins	191
13	Physical Methods for Investigation of Proteins	207
14	The Solubility Behavior of Proteins	234
15	Preparation and Purification of Proteins	260
16	Protein Structure	277
17	Denaturation of Proteins	304
18	Blood Proteins	327
19	Some Notable Protein Systems	353
20	Hormonal Proteins	367

CONTENTS

21	Enzymes	379
22	Peptides and Proteins as Substrates	399
23	Additional Proteins with Fundamental Biological Functions	414
24	Origin, Evolution, and Biosynthesis of Protein	429
	Index	441

Scope of Protein Chemistry

The outer door to the vault containing the secrets of the chemistry of living material was opened in 1828 by Wöhler, who first synthesized a product of nature (urea) from an inorganic compound (ammonium cyanate). This key experiment greatly weakened the concept of vitalism. Vitalism had ascribed to living entities a mystic nature precluding the imitation in the laboratory of any of the processes or products found in the biological realm.

A little more than a century later, the significance of Wöhler's experiment was enhanced by the isolation and crystallization of tobacco mosaic virus by Stanley. The substance of tobacco mosaic virus was shown to be a (nucleo-) protein. This substance could be studied in the laboratory by the numerous techniques of protein investigation developed especially during the twentieth century. This same chemical material, when introduced into the plant on which it is a natural parasite, exhibits an outstanding characteristic of life, that of replication. In nature this virus is by the criterion of replication alive; in laboratory glassware it is another of the substances with which the protein chemist typically deals. Stanley was then responsible for erection of a laboratory intensively and extensively devoted to virus chemistry. In this same laboratory in 1956, Fraenkel-Conrat and Williams succeeded in isolating functional nucleic acids and proteins from several viruses and in hybridizing them. Through these and other intensive studies of the chemistry of protein and nucleic acids, biology is being reduced step by step to its chemical components.

In the century between the first negation of vitalism and the isolation and characterization of self-replicating substance, organic chemistry awoke to its destiny and biochemistry flourished. Myriads of

chemical substances were separated by biochemists from nature's prolific store of organisms. Many if not most of these compounds were synthesized by organic chemists. These materials frequently served as patterns for the hundreds of thousands of artificial chemical structures prepared by chemists. From all of this, and with the work of Wöhler, Northrop, Stanley, and many others as signposts, the fundamental role of proteins in life became clearer. The innermost vault contains proteins. The central secrets of life are predominantly protein secrets. Viruses, antibodies, the master hormones, enzymes, essential animal tissues, special epidermal tissues, and membranes are proteins or, in some of these cases, consist largely of proteins. Evidence is mounting that the genes, the bearers of specific heritable characteristics, are (nucleo-) proteins. Although the total available evidence indicates that the nucleic acids are most intimately involved, the proteins evidently function at least as lieutenants over the ordering of processes that underlie the fingerprint diversity of inheritance.

Knowledge of the intimate structure of protein is yielding to intensive attack in many laboratories. In 1945 a review of knowledge revealed numerous chemical means for elucidating the entire structure of a protein in terms of the sequence of the amino acid units, despite widespread pessimism for such studies. Although the methods were already abundantly available, only one worker, Felix, had attempted to reconstruct the structure of an entire protein molecule, that of clupein, from sequence studies of fragments. Within a decade after 1945, Sanger had proposed a complete sequence for the considerably more complex molecule of insulin. The methods of sequential analysis had also begun to invade many other of the formal divisions of protein chemistry. Du Vigneaud successfully orchestrated the many methods of analysis and synthesis available in 1954, to produce artificially the protein-type hormone oxytocin.

Proteins are associated not only with the most intimate affairs of life; they are widespread in the less specialized tissues of animals. The growth and maintenance of such structures as animal tissues, membranes, hair, and nails, as well as the replenishment of the more specialized proteins which are hormones, enzymes, antibodies, and replicative substances, engage the attention of workers in nutrition and metabolism. These fields of study are concerned with what happens to proteins in feeding, and the quantity and quality that are available for these purposes. Since 1940 laymen such as ranchers and farmers, without technical training, have come to understand much of the gross significance of protein feeding.

The development on a chemical basis of crops, dairy products, and

packing-house materials for practical nutrition has made possible the beginning of chemical industries based on proteins. Revolutionary changes in the age-old arts centering about fibers are actualities. Not only fibers but plastics and films prepared from proteins in the laboratory and factory are receiving increasing attention. The amino acids which are derivable from proteins are raw materials of chemical industry, as are the proteins themselves.

Medicine has likewise found amino acids and proteins to be of great value. Hormones which organize body processes, enzymes which give them their characteristic living tempo, and antibodies which counteract pathogenic foreign proteins such as some of the viruses, are mainstays of curative and preventive medicine. In protein therapy by the administration of balanced amino acid solutions, nutrition and medicine have again found common ground.

All of these developments have been built upon the groundwork laid by the labors of a variety of chemists, in most cases specialists. Biochemists, organic chemists, physical chemists, nutritionists, physiologists, geneticists, embryologists, physicians, and numerous others have contributed, each with his characteristic approach. Much of value has been incorporated into the structure by individual workers and by groups of workers employing two or more types of techniques simultaneously. Since problems involving proteins are often comparatively complex and ramified, studies employing combined techniques have been especially fruitful. In the literature repose increasing multitudes of isolated facts that will be correlated by individuals and groups of individuals who possess the interest and understanding necessary to ferret them out.

Some workers in the field of protein chemistry choose to study only the large units which are proteins. Yet others find time to specialize only in the study of the derivable amino acids. In the early days of protein chemistry this latter activity properly constituted the principal effort of such pioneers as Emil Fischer. Other workers are learning much by confining their investigations to peptides, which are compounds of a few amino acids but are not as structurally diversified as the larger protein entities. Although there are many kinds of justification for such specialization by research workers, the student will benefit by a broad outlook on the subject. If he studies only amino acids or peptides, he may lose sight of the fact that amino acids are junior to proteins themselves. He cannot, on the other hand, truly understand modern knowledge of the proteins without a thorough appreciation of the behavior of the amino acids and peptides. The subject obviously can be developed logically through the amino acids

and peptides consecutively, after which the proteins may be treated and understood in terms of what they are—complex bodies that exhibit, simultaneously, diverse reactions of amino acids and peptides as well as peculiar behavior of their own.

It is along this line of development of the subject that the present book is organized.

Description of the Amino Acids

Terminology

Criteria of acceptability

Tabulation of acceptable amino acids

Classification of amino acids

Other amino acids

History of discovery

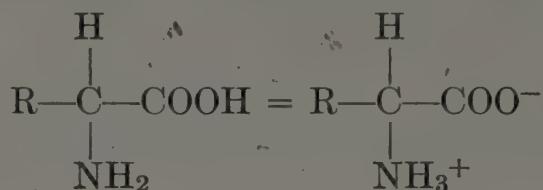
TERMINOLOGY

The amino acids are in a sense the alphabet of protein chemistry. One manifestation of this analogy is the great diversity of the various combinations, in proteins on the one hand, and in words on the other. An understanding of proteins without a knowledge of the amino acid alphabet is perhaps even more difficult than an understanding of the construction of English words without a knowledge of the letters from which they are constructed.

Strictly speaking, the term *amino acid* refers to what its name suggests, that is, an organic compound containing amino and acid groups. From the standpoint of biological and protein chemistry the meaning is restricted to units which are obtained upon hydrolysis of protein; these units are invariably found to contain an amino group in a position alpha to a carboxyl group.*

* An exception is *p*-aminobenzoic acid, which is found as part of the vitamin folic acid. On the basis of a strictly organic chemical definition *p*-aminobenzoic acid is an amino acid.

The prototypic structure* of the amino acid is

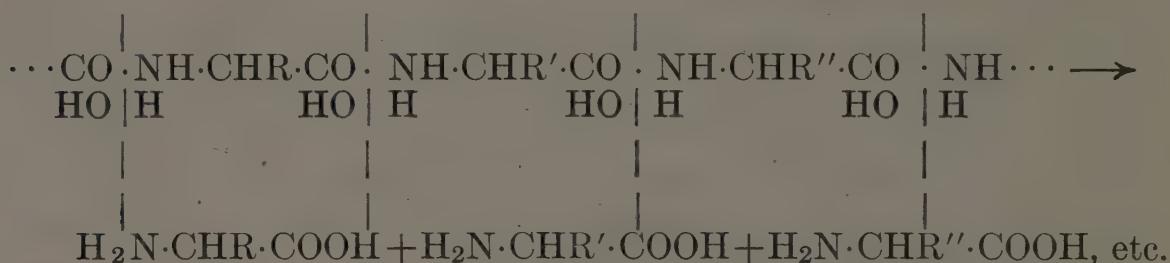


The carboxyl and amino groups are sometimes designated as α -groups, inasmuch as there are in some of the amino acids side chains with an additional amino or carboxyl group, denoted by ω - . The hydrogen atom is similarly referred to as the α -hydrogen atom. The α -H, α -NH₂, —COOH, and R-groups are bonded to the α -carbon atom in a relationship that can be expressed by a tetrahedron, with the carbon atom at the center and the bonded groups at the apices. The α -carbon is an asymmetric carbon atom in all known amino acids except glycine. In glycine, R is simply hydrogen, and there are thus two hydrogen atoms bonded to one carbon, which condition precludes asymmetry. All amino acids except glycine are accordingly capable of existence in optically active forms (p. 41).

In proteins the various amino acids are found to be linked schematically through loss of water from α -carboxyl and α -amino groups:



The combination of carbonyl and imino groups, —CO·NH—, is known as the *peptide linkage*. The repeating unit, —NH·CHR·CO—, is an *amino acid residue*. A combination of two or more amino acids in peptide linkage through their respective amino and carboxyl groups represents a *peptide*. All except the smallest peptides may sometimes be called *polypeptides*, which term has no meaning otherwise distinct from that of peptides. Some peptides found in nature are called proteins. There is, however, no unequivocal criterion of distinction between peptides and proteins (p. 131). Upon hydrolysis of a peptide molecule, the peptide linkages are broken.



*The ionized form predominates under most conditions (Chap. 3). Some reactions, however, funnel through the undissociated form, and many reactions are written with this form on the basis that such a mechanism is presumed in the individual case.

The amino acids differ in the nature of their R-groups or side chains. The virtually limitless theoretical diversity of protein molecules is based predominantly on slightly less than two dozen amino acids. Experimental investigation of this group of substances has made it desirable to establish criteria for "acceptable" amino acids.

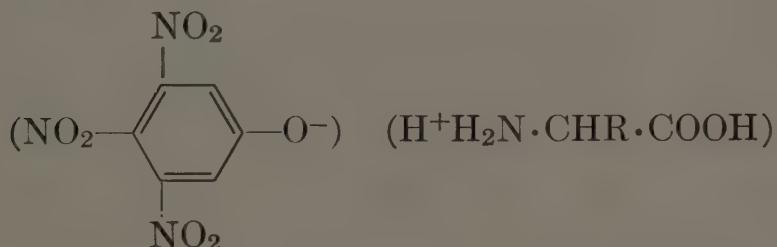
CRITERIA OF ACCEPTABILITY

These criteria, as set out by Vickery and Schmidt (1931), are of three main types:

1. Necessary criteria.
 - a. The amino acid must be liberated by hydrolysis from a protein of demonstrated purity.
 - b. The amino acid must be characterized by analysis of salts and of typical derivatives.
2. Arbitrary, but justifiable, criteria.
 - a. The amino acid must be isolated by the discoverer and one other worker as well.
 - b. The constitution must be established by synthesis of the amino acid.
 - c. The racemized natural compound must be shown to be identical to the synthetic amino acid, or one form of the resolved acid must be shown to be identical with the natural product.
3. Desirable criteria.
 - a. A synthetic peptide containing the amino acid should be attacked by a proteolytic enzyme.
 - b. The amino acid should be oxidized in the animal body.

The above criteria, published in 1931, have in general proved to be quite definitive. Experimental difficulties arise, however, in the demonstration of purity of a protein (Chap. 15), and if absolute purity of the protein were required, probably no amino acid would qualify as acceptable. In reality the term "protein of demonstrated purity" need be interpreted as implying only the absence of nonprotein material.

Several derivatives are typically employed for the characterization of amino acids. These include the N-benzoyl compounds, $C_6H_5CO \cdot NH \cdot CHR \cdot COOH$; the picrate salts,



INTRODUCTION TO PROTEIN CHEMISTRY

TABLE 2-1. The Amino Acids

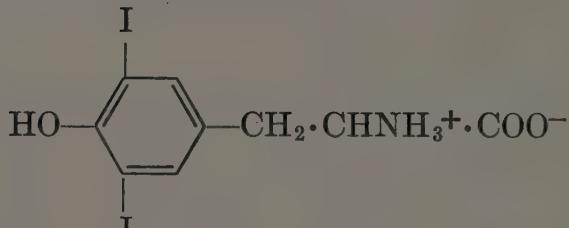
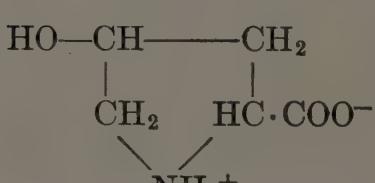
Name	Structure
Alanine	$\text{CH}_3 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Arginine	$\begin{array}{c} \text{NH}_2 \\ \\ \text{HN}=\text{C} \\ \\ \text{NH} \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^- \end{array}$
Aspartic acid	$\text{HOOC} \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Cystine } Cysteine }	$\begin{array}{c} -\text{OOC} \cdot \text{CHNH}_3^+ \cdot \text{CH}_2 \cdot \text{S}=\text{S} \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^- \\ \hspace{10em} \text{HSCH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^- \end{array}$
Diiodotyrosine	
Glutamic acid	$\text{HOOC} \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Glycine	$\text{H}_3\text{N}^+ \cdot \text{CH}_2 \cdot \text{COO}^-$
Histidine	
Hydroxylysine	$\text{H}_2\text{N} \cdot \text{CH}_2 \cdot \text{CHOH} \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Hydroxyproline	

TABLE 2-1. (Continued)

Distinguishing Chemical Characteristic	Biological Significance
Monoaminomonocarboxylic	Glycogenic
Basic in reaction due to guanidyl group	Essential for dogs and rats; participates in urea cycle
$\begin{array}{c} \text{NH} \\ \\ \text{H}_2\text{N}\cdot\text{C}\cdot\text{NH}- \end{array}$	
Acidic in reaction; dicarboxylic acid	Glycogenic
Sulfur-containing	Detoxicant, glycogenic
Phenolic; iodine-containing	Precursor of thyroxine
Acidic in reaction; dicarboxylic acid	Detoxicant, glycogenic, participates in transamination
Monoaminomonocarboxylic. No asymmetric carbon atom	Purine and porphyrin intermediate
Basic in reaction due to imidazolyl group	Essential for growing animals; convertible to histamine
$\begin{array}{c} \text{CH} \\ \diagdown \quad \diagup \\ \text{HN} \quad \text{N} \\ \qquad \\ \text{HC} = \text{C} \end{array}$	
Basic in reaction	
Pyrrolidine ring representing aliphatic R condensed with α -amino group	Glycogenic

(Continued)

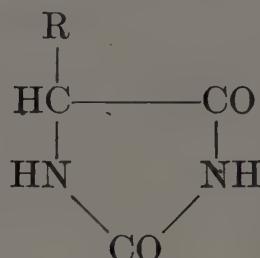
TABLE 2-1. (Continued)

Name	Structure
Isoleucine	$\begin{array}{c} \text{C}_2\text{H}_5 \\ \\ \text{CH} \cdot \text{CHNH}_3^+ \cdot \text{COO}^- \\ \\ \text{CH}_3 \\ \\ \text{CH}_3 \end{array}$
Leucine	$\begin{array}{c} \text{CH}_3 \\ \\ \text{CH} \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^- \\ \\ \text{CH}_3 \end{array}$
Lysine	$\text{H}_2\text{NCH}_2 \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Methionine	$\text{CH}_3 \cdot \text{S} \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Phenylalanine	$\text{C}_6\text{H}_5-\text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Proline	$\begin{array}{c} \text{CH}_2-\text{CH}_2 \\ \quad \\ \text{CH}_2 \quad \text{HC} \cdot \text{COO}^- \\ \quad \\ \text{NH}_2^+ \end{array}$
Serine	$\text{HOCH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Threonine	$\text{CH}_3 \cdot \text{CHOH} \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Thyroxine	$\text{HO}-\text{C}_6\text{H}_3(\text{I})_2-\text{O}-\text{C}_6\text{H}_3(\text{I})_2-\text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Tryptophan	$\begin{array}{c} \text{C}_6\text{H}_5-\text{C} \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^- \\ \\ \text{NH} \end{array}$
Tyrosine	$\text{HO}-\text{C}_6\text{H}_4-\text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$
Valine	$\begin{array}{c} \text{CH}_3 \\ \\ \text{CH} \cdot \text{CHNH}_3^+ \cdot \text{COO}^- \\ \\ \text{CH}_3 \end{array}$

TABLE 2-1. (Continued)

Distinguishing Chemical Characteristic	Biological Significance
Monoaminomonocarboxylic	Essential
Monoaminomonocarboxylic	Essential
Basic in reaction; diamino acid	Essential
Sulfur-containing	Essential; methyl group donor
Phenyl group	Essential
Pyrrolidine ring representing ali- phatic R condensed with α -amino group	Glycogenic
Hydroxyl group; monoaminomonocarboxylic	Glycogenic
Hydroxyl group; monoaminomonocarboxylic	Essential
Phenolic; iodine-containing	Active principle of thyroid
Indole ring	Essential, related to β -indolylacetic acid
Phenolic; weakly acidic	Precursor of epinephrine and natural pigments
Monoaminomonocarboxylic acid	Essential

the solid ester hydrochlorides, $-\text{Cl} + \text{H}_3\text{N}\cdot\text{CHR}\cdot\text{COOC}_2\text{H}_5$ (almost all known esters are liquids which decompose at their atmospheric boiling points); and the hydantoins

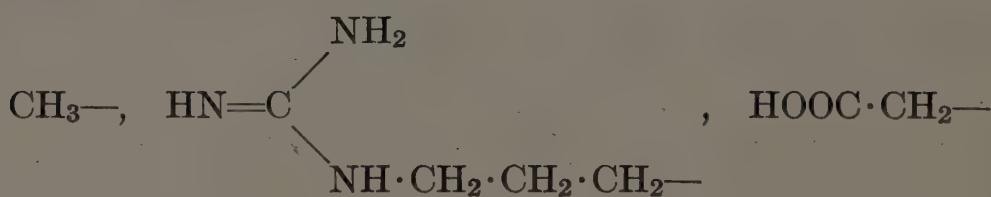


Criteria 3a and 3b each relate to enzymic behavior that in actuality is relatively specific, in 3a for hydrolysis of peptides, and in 3b for oxidation of amino acids. Since the publication of Vickery and Schmidt's criteria, enzymes which catalyze the oxidation of amino acids have been isolated. It is worth pointing out here that only appropriate proteolytic enzymes will hydrolyze a given peptide, or only the appropriate amino acid oxidase will oxidize a given amino acid; for example, not all proteolytic enzymes will hydrolyze all peptides. This point is clarified in Chap. 22.

TABULATION OF ACCEPTABLE AMINO ACIDS

In Table 2-1 the amino acids which are acceptable by the preceding criteria are listed alphabetically by their most common names. The chemical structures follow in the next column. The differences in properties of proteins can be explained to a large extent in terms of the side chains that characterize the amino acids. The proteins vary in their contents of various amino acids and therefore in the proportions of side chains having various properties. It is well to bear in mind that in the protein nearly all of the functional α -amino and α -carboxyl groups are combined in peptide linkage, masking their reactivity. This arrangement leaves the side chains partially free, and these side chains therefore contribute significantly to the reactivity of the whole protein. In the last column of Table 2-1 are presented some of the more significant biological manifestations of the amino acids.

The R-group or side chain in all cases represents all except the $-\text{CHNH}_3^+ \cdot \text{COO}^-$ portion of the amino acid $\text{R} \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$. Thus, the R-groups in alanine, arginine, and aspartic acid are respectively:



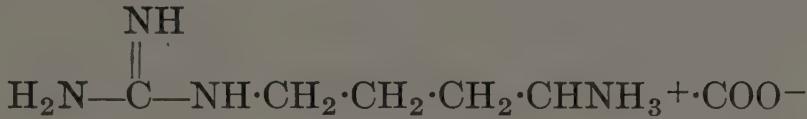
The side chain in alanine, as in other monoaminomonocarboxylic acids, is hydrocarbon in nature. The exceptions to this rule are hydrogen in glycine and the alcoholic side chains in serine and threonine. The side chain in each of these amino acids is therefore neutral. One may consider that the amino and carboxyl groups approximately neutralize each other; therefore, inasmuch as the side chain is neutral, the whole amino acid is neutral. The monoamino-monocarboxylic acids are in fact sometimes designated as neutral amino acids. In the case of arginine the guanidine radical



is basic, the side chain



is therefore basic, and the whole amino acid



is accordingly basic. Similarly, $\text{HOOC}\cdot\text{CH}_2-$, the R-group of aspartic acid, is acidic, and the entire amino acid is acidic in reaction. The same principles apply to the other amino acids.

Glycine was historically known as glycocoll.* Diiiodotyrosine is known as iodogorgoic acid because of its original discovery in the coral *Gorgonia cavolini*. Cysteine (cyst-ā-ene) and cystine are grouped together because they are readily interconvertible. Evidence exists for both the cystine residue and the cysteine residue as constituents of proteins. In the air, cysteine oxidizes readily in solution to cystine; consequently, typical mineral acid hydrolysis of protein yields only cystine. When, however, hydrolysis is executed under reducing conditions, as in the presence of a metal evolving hydrogen with the mineral acid, cysteine may be obtained:



The glycogenic amino acids are those which are metabolically convertible to carbohydrates by mammals. The *essential* or *indispensable*

* A toxic chemical used as a photographic developer is frequently designated "Glycine," and this has led to some confusion commercially.

pensable amino acids are those which mammals must ingest in their foods; these substances are necessary for both growth of the young and maintenance of the adult. The amino acids not designated essential are in all cases essential to life. They are not listed as "essential" by definition, because they are synthesized within the body from other amino acids, for example, from phenylalanine in the case of thyroxine. Thyroxine is not required as such in the diet.

Most of the amino acids are found in most proteins. A few amino acids are highly restricted in distribution. These latter include diiodotyrosine, thyroxine, hydroxylysine, and hydroxyproline.

CLASSIFICATION OF AMINO ACIDS

In memorizing structures of amino acids, it is helpful to observe that in the aromatic amino acids (diiodotyrosine, histidine, phenylalanine, thyroxine, tryptophan, and tyrosine) the α -carbon atom is

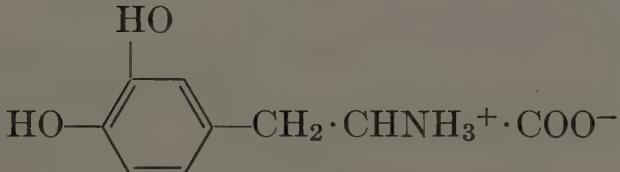
TABLE 2-2. Classification of Amino Acids

Classification	Acid
Monoaminomonocarboxylic acids	Glycine Alanine Valine Leucine Isoleucine
Hydroxyamino acids	Serine Threonine
Acidic amino acids	Aspartic acid Glutamic acid
Basic amino acids	Histidine Arginine Lysine Hydroxylysine
Sulfur-containing amino acids	Cystine-cysteine Methionine
Benzoid amino acids	Phenylalanine Tyrosine Diiodotyrosine Thyroxine Tryptophan
Pyrrolidyl amino acids	Proline Hydroxyproline

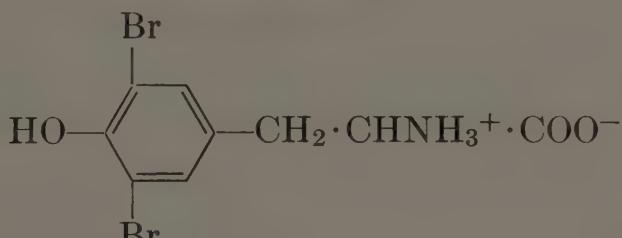
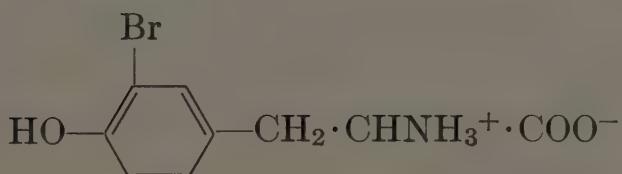
found in all cases in a 3-carbon radical, $-\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$. It may also be noted that each of the basic amino acids contains a total of six carbon atoms. Correlation of the physical and chemical properties of amino acids is facilitated by the grouping in Table 2-2.

OTHER AMINO ACIDS

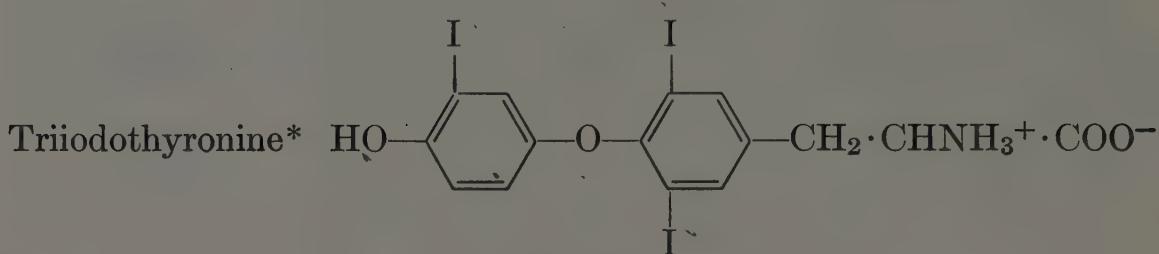
Amino acids which have been reported as found in nature but which do not qualify for inclusion in the "standard" list include the following:

α -Amino adipic acid	$\text{HOOC}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$
α -Amino butyric acid	$\text{CH}_3\cdot\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$
Canaline	$\text{H}_2\text{N}\cdot\text{OCH}_2\cdot\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$
Canavanine	$\begin{array}{c} \text{NH}_2 \\ \\ \text{HN}=\text{C} \\ \\ \text{NH}\cdot\text{OCH}_2\cdot\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^- \end{array}$
Citrulline	$\text{H}_2\text{NCONH}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$
Diaminopimelic acid	$-\text{OOC}\cdot\text{CHNH}_3^+\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$
Djenkolic acid	$-\text{OOC}\cdot\text{CHNH}_3^+\cdot\text{CH}_2\cdot\text{S}\cdot\text{CH}_2\cdot\text{S}\cdot\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$
Dihydroxyphenylalanine	
Ornithine	$\text{H}_2\text{NCH}_2\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$

Also, a number of halogenated tyrosines and monoiodohistidine have been reported. The halogen derivatives include, from sponge,

Dibromotyrosine	
Monobromotyrosine	

and from thyroglobulin of mammalian origin,



diiodothyronine, monoiodotyrosine, and the monoiodohistidine earlier mentioned.

Many of these have not been isolated by two investigators, or have been found in the free state only. α -Aminoadipic acid is reported as a metabolic intermediate and as a hydrolytic product of corn protein. The occurrence of α -aminobutyric acid as a product of the hydrolysis of casein and silk fibroin has been reported, but has been questioned. Hydroxyglutamic acid appeared in some tables but has finally been withdrawn. Norleucine, $\text{CH}_3 \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CHNH}_3^+ \cdot \text{COO}^-$, has been reported as a protein constituent by Abderhalden. Hydrolyzates of spinal-cord protein, originally claimed to be a source, have yielded no norleucine by the sensitive procedure of paper-strip chromatography (Chap. 6). On the other hand, this same method revealed the presence of the previously unrecognized diaminopimelic acid in diphtheria bacilli.

Citrulline and ornithine, with arginine, are intermediates of the physiologically important urea cycle. By chromatography each of these has been found in hydrolyzates of algal protein. The relationship of canavanine (from soybean or jackbean) to cananine is similar to that of arginine to ornithine. Djenkolic acid was discovered as a crystalline component of urine of Javanese natives who eat djenkol nuts. The chemical structure is very closely related to that of cystine, as are the physical properties. Each amino acid forms hard crystals which may separate from physiological fluids. Djenkolic acid was first recognized by the pain it caused on crystallizing in the gastrointestinal tract. Cystine has been found as a predominant constituent of some gallstones.

Dihydroxyphenylalanine is also known as "dopa," a name derived from its German designation, *dioxypheylalanine*. It is an intermediate in the biological conversion of tyrosine to natural pigments, the latter exemplified by the brown color of a sliced potato or the color of human skin.

* Gross and Pitt-Rivers (1953).

HISTORY OF DISCOVERY

The history of the discovery of the amino acids is presented in Table 2-3.

This table serves principally to place in proper historical setting the discoveries of amino acids from protein. As in other situations involving priority, several laboratories might, if they chose, claim some

TABLE 2-3. History of Discovery of the Amino Acids
as Protein Hydrolytic Products

Amino Acid	Date Discovered	Discoverer(s)	Source
Glycine	1820	Braconnot	Gelatin
Leucine	1820	Braconnot	Wool and muscle
Tyrosine	1849	Bopp	Casein
Serine	1865	Cramer	Silk
Glutamic acid	1866	Ritthausen	Gliadin
Aspartic acid	1868	Ritthausen	Conglutin, legumin
Phenylalanine	1881	Schultze and Barbieri	Lupine sprouts
Alanine	1881	Weyl	Silk fibroin
Lysine	1889	Drechsel	Coral
Arginine	1895	Hedin	Horn
Diodotyrosine	1896	Drechsel	Coral
Histidine	1896	Kossel, Hedin	Sturin, casein
Cystine	1899	Mörner	Horn
Valine	1901	Fischer	Casein
Proline	1901	Fischer	Casein
Tryptophan	1901	Hopkins and Cole	Casein
Hydroxyproline	1902	Fischer	Gelatin
Isoleucine	1904	Ehrlich	Fibrin
Thyroxine	1915	Kendall	Thyroid
Methionine	1922	Mueller	Casein
Hydroxylysine	1925	Schryver, Biston, and Mukherjee	Isinglass
Threonine	1935	McCoy, Meyer, and Rose	Casein

of the initial findings. In the case of tyrosine, for example, Liebig obtained this amino acid in 1846 from the alkaline fusion of casein, and applied the name of tyrosine to the product. Liebig did not, however, determine the structure. In the case of a more recently recognized amino acid, hydroxylysine, Van Slyke and Hiller first suggested the compound from analytical studies of hydrolyzates in 1921, and Van Slyke and co-workers also proved the nature of the compound somewhat more conclusively than the generally acceptable proof of Schryver, Biston, and Mukherjee. Many of the other dis-

coveries represent the accumulated efforts of many laboratories; a search of the literature on any one of these is likely to prove rewarding for the perspective it provides in the development of a type of chemical thought.

The earliest amino acid to be recognized under any conditions was cystine, isolated in the crystalline state from a urinary stone in 1810 by Wollaston. It was 1899, however, before cystine was recognized as a hydrolytic product of protein. The number of amino acids discovered or characterized by Emil Fischer is testimony to the energy of this chemist and to the power of the ester distillation method which he perfected (p. 57). Newer methods came into prominence later in the twentieth century, and these were applied widely enough that for many common proteins it is now possible to list virtually 100 per cent of the amino acids which may be obtained upon their hydrolysis.

SELECTED REFERENCES

Reviews

"History of the Discovery of the Amino Acids," H. B. Vickery and C. L. A. Schmidt, *Chem. Revs.*, **9**, 169 (1931).

"The Constitution and Synthesis of Amino Acids," M. S. Dunn and C. L. A. Schmidt, *Chemistry of the Amino Acids and Proteins*, p. 21, Charles C Thomas, Springfield, Illinois, 1945.

"Neue natürliche Aminosäuren," H. Musso, *Angew. Chem.*, **68**, 313 (1956).

Original Articles

"Baldriansäure und ein neuer Körper aus Käsestoff," J. Liebig, *Ann. Chem. Liebigs*, **57**, 127 (1846).

"Zur Kenntnis des Cystins," E. Fischer and U. Suzuki, *Hoppe-Seyler's Z. physiol. Chem.*, **45**, 405 (1905).

"A New Sulfur-Containing Amino Acid Isolated from the Hydrolytic Products of Protein," J. H. Mueller, *J. Biol. Chem.*, **56**, 157 (1923).

"Chemistry of Thyroxine. II. Constitution and Synthesis of Desiodothyroxine," C. R. Harington, *Biochem. J.*, **20**, 300 (1926).

"The Physiology of Ergothioneine," B. A. Eagles and H. M. Vars, *J. Biol. Chem.*, **80**, 615 (1928).

"A Note on the Crystallization of Free Lysine," H. B. Vickery and C. S. Leavenworth, *J. Biol. Chem.*, **76**, 437 (1928).

"Isolierung des Citrullins (δ -Carbamidoornithin) aus tryptischen Verdauungsprodukten des Caseins," M. Wada, *Biochem. Z.*, **257**, 1 (1933).

"Die Djenkolsäure, eine neue schwefelhaltige Aminosäure," A. G. Van Veen and A. J. Hyman, *Rec. trav. chim.*, **54**, 493 (1935).

"The Carbamido Diacetyl Reaction: A Test for Citrulline," W. R. Fearon, *Biochem. J.*, **33**, 902 (1939).

"The Chemistry of Melanin. III. Mechanism of the Oxidation of Dihydroxyphenylalanine by Tyrosinase," H. S. Mason, *J. Biol. Chem.*, **172**, 83 (1948).

"The Isolation of L-Pipecolinic Acid from *Trifolium repens*," R. I. Morrison, *Biochem. J.*, **53**, 474 (1953).

"3:5:3'-Triiodothyronine. 1. Isolation from Thyroid Gland and Synthesis," J. Gross and R. Pitt-Rivers, *Biochem. J.*, **53**, 645 (1953).

"Synthesis and Biochemistry of 5- and 7-Hydroxytryptophan and Derivatives," A. Ek and B. Witkop, *J. Am. Chem. Soc.*, **75**, 500 (1953).

"The Synthesis of 3-(4'-Dimethylamino-3':5'-Dinitrophenyl) Hydantoin Derivatives of Various Amino Acids, and Their Use for the Determination of N-Terminal Amino Acids," G. G. Evans and W. S. Reith, *Biochem. J.*, **56**, 111 (1954).

"Observations on the Iodo-Amino-Acids of Marine Algae Using Iodine-131," R. Scott, *Nature*, **173**, 1098 (1954).

"The Combined Amino Acids in Several Species of Marine Algae," D. G. Smith and E. G. Young, *J. Biol. Chem.*, **217**, 845 (1955).

"The Nature of D-Alanine in Lactic Acid Bacteria," E. E. Snell, N. S. Radin, and M. Ikawa, *J. Biol. Chem.*, **217**, 803 (1955).

3

Physical Properties of the Amino Acids

Theory of acid-base equilibria

Dissociation of amino acids

The dipolarionic form of amino acids

The titration behavior of amino acids and peptides

Melting point

Solubility

Configuration and optical rotatory power

The amino acids, containing as they do both carboxylic and amino groups, are amphoteric; that is, they can act both as acids and as bases. An understanding of the amphoteric properties of the amino acids facilitates the study of their over-all physical and chemical properties. Furthermore, such knowledge is essential to an understanding of many of the properties of proteins, particularly their electrochemical and solubility properties and their complex interactions.

THEORY OF ACID-BASE EQUILIBRIA

Advantages of the Brønsted Treatment. It seems to be advisable to preface the consideration of the dissociation properties of amino acids with a brief treatment of the theory of dissociation of weak acids and bases. For purposes of understanding the amphoteric properties of amino acids, and particularly of proteins, the Brønsted point of view is preferable to the classical Arrhenius treatment. Insofar as aqueous systems are concerned,* the merit of the Brønsted

* The most obvious advantage of the Brønsted treatment is that by considering only protons (hydrogen ions), and not hydroxyl ions, one may apply the concept

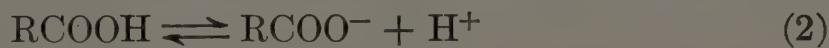
formulation arises from the universal use of a single *pH* scale for expressing both the acidity and alkalinity of aqueous solutions. Thus one thinks of an alkaline solution as being one poorer than 10^{-7} molar in hydrogen ions and not one rich in hydroxyl ions. By focusing attention on the ability of groups to dissociate or to combine with protons (hydrogen ions), the Brønsted theory is readily adaptable to this *pH* system, and it also somewhat simplifies the consideration of complex systems.

This perspective is particularly advantageous in the interpretation of the titration curves in cases in which there is uncertainty as to what dissociating groups are involved, as in peptides and proteins. Furthermore, all dissociating groups can be ordered according to acidity on one unidirectional logarithmic scale, a procedure somewhat analogous to the use of the *pH* scale. These advantages will become more apparent after consideration of the dissociation behavior of amino acids and especially of proteins (Chap. 11).

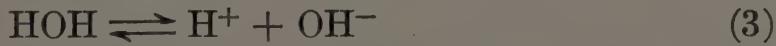
Brønsted Acids and Bases. According to the Brønsted formulation, an acid is defined as any substance which can give rise to protons. Furthermore, for every acid there is a conjugate base, the body which results from the acid as a consequence of dissociation of the proton.* More generally, a base may be defined as any compound or ion which can bind protons. This definition is illustrated by the schematic reaction



in which AH^+ is an acid and B its conjugate base. Thus all undissociated carboxylic acids are acids in the Brønsted sense; on the other hand, their anions are bases, as we see from the reaction



Water can act either as a base or as an acid, and hence is amphoteric. Thus in the reaction



water is an acid, and hydroxyl ion is its conjugate base. In the reaction



water is a base, and oxonium ion H_3O^+ is the acid.† In this case

to nonaqueous systems in which the solvent base is other than hydroxyl ion. Since we are interested primarily in aqueous systems, this advantage is not a real one here.

* Conversely, corresponding to every base there is a conjugate acid.

† This equilibrium is far enough to the right that most of the protons in aqueous solution exist in the hydrated form. For simplicity in our discussion we will usually indicate protons simply by H^+ .

the acid is positively charged, the base is neutral. It should be noted that there is no limitation on the charge possessed by an acid or base except that the acid must always possess one more positive charge, or one less negative charge, than the conjugate base, owing to the loss of the proton.

On the basis of equations 3 and 4, the acid-base equilibria in aqueous solution can be pictured as a competition between added base and the omnipresent bases OH^- and H_2O for the available protons.

Brønsted Dissociation Constant. It is of interest now to write the equilibrium constant for the dissociation of a Brønsted acid according to equation 1:

$$K_a = \frac{[\text{B}][\text{H}^+]}{[\text{AH}^+]} \quad (5)$$

The terms in brackets refer to activities* of the species in question. K_a is the acidic dissociation constant of the acid AH^+ . As examples we might write the equilibrium constants for acetic acid

$$K_a = \frac{[\text{CH}_3\text{COO}^-][\text{H}^+]}{[\text{CH}_3\text{COOH}]} \quad (6a)$$

and for water

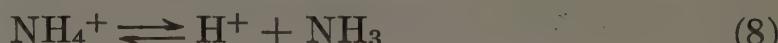
$$K_a = \frac{[\text{H}^+][\text{OH}^-]}{[\text{H}_2\text{O}]} \quad (6b)$$

In aqueous solution, water usually occurs in such great excess that its activity can be considered constant. This constant is furthermore taken as unity by convention, so that the acidic dissociation constant of water is usually written

$$K_w = [\text{H}^+][\text{OH}^-] \quad (7)$$

The value of K_w is about 10^{-14} at 25° ,† but increases with increasing temperature. It becomes about 10^{-12} at the boiling point.

Another type of acid is represented by the ammonium ion NH_4^+ , which dissociates according to the reaction



Here NH_3 is the conjugate base. One commonly thinks of NH_4OH

* The term activity means *effective* concentration and is in general somewhat different from actual concentration, being equal to it only in the so-called ideal solution.

† Temperature indications throughout this book are to be read as degrees Centigrade.

as the base, and indeed this is just a hydrated form of NH_3 . Thus, from the Brønsted point of view, one considers the basicity of NH_3 as due to its tendency to combine with protons. As might be expected, there is a very simple mathematical relationship between the K_a of the ammonium group and the K_b of ammonium hydroxide as usually formulated. Thus

$$K_a \text{ (of } \text{NH}_4^+ \text{)} = \frac{[\text{NH}_3][\text{H}^+]}{[\text{NH}_4^+]} \quad (9a)$$

$$K_b \text{ (of } \text{NH}_4\text{OH} \text{)} = \frac{[\text{NH}_4^+][\text{OH}^-]}{[\text{NH}_4\text{OH}]} \quad (9b)$$

But by equation 7, $[\text{H}^+][\text{OH}^-] = K_w$ or $[\text{OH}^-] = K_w/[\text{H}^+]$. By substituting for $[\text{OH}^-]$ in equation 9b,

$$K_b = \frac{[\text{NH}_4^+]K_w}{[\text{NH}_4\text{OH}][\text{H}^+]} \quad (10)$$

and by replacing $[\text{NH}_4\text{OH}]$ by its unhydrated form $[\text{NH}_3]$,

$$K_b = \frac{[\text{NH}_4^+]K_w}{[\text{NH}_3][\text{H}^+]} = \frac{K_w}{K_a} \quad (11)$$

Thus it is possible to express the basic strengths of Arrhenius bases in terms of the acid strengths of their conjugate acids. This makes possible the arrangement of all proton-dissociating and proton-combining groups on a single scale.

In Table 3-1 are summarized the Brønsted dissociation constants of some common acids. In addition there are included the logarithmic dissociation constants, under the column headed pK_a . The symbol p is used in the same sense as in $p\text{H}$, that is, to denote "negative logarithm of." Thus the stronger the acid, the larger is K_a and the smaller is pK_a . The acidic groups included in this table are seen to cover the pK_a range 2.1 to 17. The strong mineral acids have even lower pK_a values, too low to measure with accuracy. Alcohols are included as acids even though they are not usually thought of as possessing an acidic function. To be sure, as compared to water they are very weak acids (pK of about 17 as compared to 14, or K_a 1000 times smaller) and so will not be appreciably ionized as long as water is present.

Titration Behavior and Buffering. Returning to the generalized form of the equilibrium constant, equation 5, it is of interest to solve for $[\text{H}^+]$, the activity of hydrogen ions. This is given by

$$[\text{H}^+] = \frac{K_a[\text{AH}^+]}{[\text{B}]} \quad (12)$$

By taking the common logarithm of both sides,

$$\log [H^+] = \log K_a + \log \frac{[AH^+]}{[B]} \quad (13)$$

We now make use of the symbol p , denoting "negative logarithm of," and obtain

$$pH = pK_a + \log \frac{[B]}{[AH^+]} \quad (14)$$

Thus, in the *buffer system* consisting of acid A and its conjugate base B, the pH is related through this equation to the pK_a of the acid, and the ratio of the concentration of form B to the concentration of form AH^+ existing in the solution.

TABLE 3-1. The Dissociation Constants of Some Important Acids

Acid	Formula	K_a	pK_a
Acetic	$CH_3\cdot COOH$	1.75×10^{-5}	4.76
Propionic	$C_2H_5\cdot COOH$	1.34×10^{-5}	4.87
Chloroacetic	$ClCH_2\cdot COOH$	1.40×10^{-3}	2.86
Lactic	$CH_3\cdot CHO\cdot COOH$	1.40×10^{-4}	3.86
Malonic K_1	$HOOC\cdot CH_2\cdot COOH$	1.40×10^{-3}	2.85
K_2		2×10^{-6}	5.7
Succinic K_1	$HOOC\cdot CH_2\cdot CH_2\cdot COOH$	6.4×10^{-5}	4.2
K_2		2.7×10^{-6}	5.6
Phosphoric K_1	H_3PO_4	7.58×10^{-3}	2.12
K_2		6.2×10^{-8}	7.21
K_3		4.8×10^{-13}	12.32
Boric	H_3BO_3	5.8×10^{-10}	9.24
Ammonium ion	NH_4^+	5.5×10^{-10}	9.26
Ethylammonium ion	$C_2H_5NH_3^+$	2.5×10^{-11}	10.6
Guanidinium ion	$C(NH_2)_3^+$	2×10^{-14}	13.7
Imidazolium ion	$ \begin{array}{c} HC \text{ --- } NH^+ \\ \qquad \\ HC \qquad \text{CH} \\ \qquad \\ \text{NH} \end{array} $	8×10^{-8}	7.1
Phenol		1.3×10^{-10}	9.89
Water	HOH	1.0×10^{-14}	14.0
Aliphatic alcohols	ROH	<i>ca.</i> 10^{-17}	<i>ca.</i> 17
Mercaptans	RSH	<i>ca.</i> 10^{-11}	<i>ca.</i> 11

In Fig. 3-1 the dissociation curves are plotted for several acids of interest, as calculated from equation 14 by using the pK values obtained from Table 3-1. Several interesting and important points are illustrated in this figure.

- (1) The sigmoidal shape of the curve is independent of the acid involved.
- (2) The half-titration point occurs at a pH equal to the pK of the acid, since at that point $[AH^+] = [B]$ and $\log [B]/[AH^+]$ is zero.

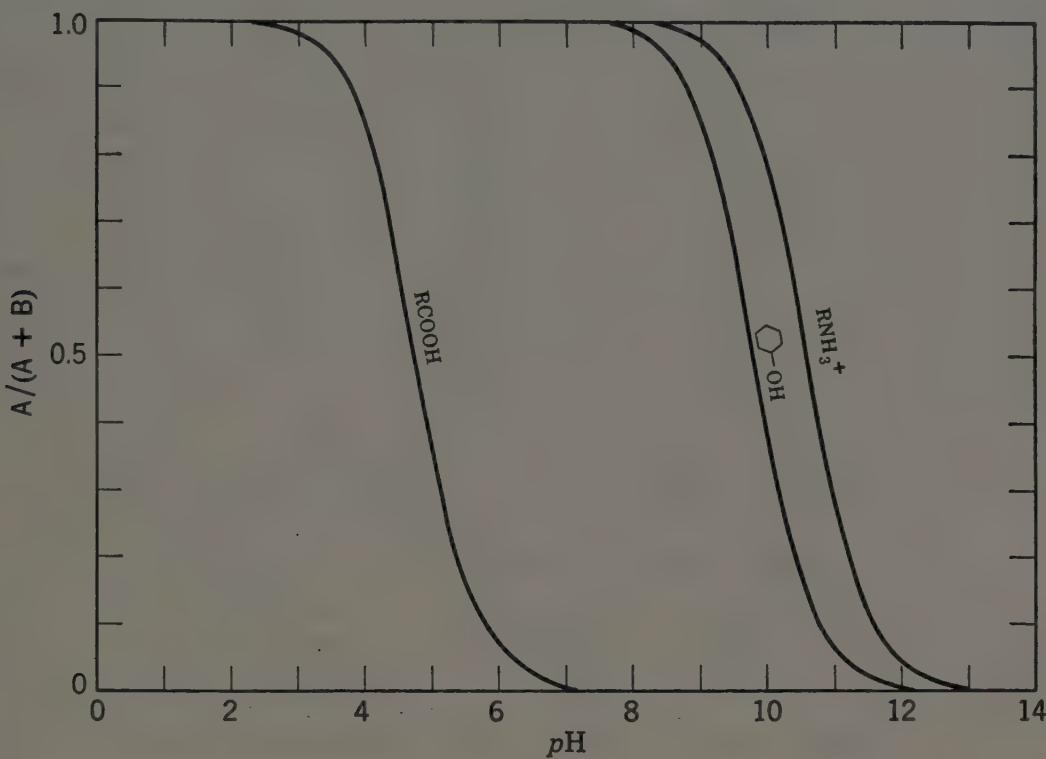


Fig. 3-1. Theoretical titration curves for an aliphatic carboxylic acid, phenol, and a primary aliphatic amine. The ordinate can be regarded as that fraction of the total (acidic plus basic form) which is in the acidic form A.

(3) The maximum buffering power also occurs at the mid-point, since at that point the change in pH with change in the ratio of the two forms is minimal.

(4) At a pH one unit below the pK , approximately 90 per cent of the buffer is in the acid form; and at a pH of two units below the pK , approximately 99 per cent. Similar considerations apply to the proportion of the basic form on the alkaline side of the mid-titration point.

DISSOCIATION OF AMINO ACIDS

In explaining the dissociation properties of an amino acid it is obviously necessary to consider two or more different dissociable groups. The dissociation constants of these groups can be measured directly by titration, as seen in the preceding section. Examples of the titration curves of some amino acids are shown in Fig. 3-5 and are discussed in a later section. Having evaluated the two or more

TABLE 3-2. The Experimentally Measured Logarithmic
Dissociation Constants of Some Amino Acids¹

SIMPLE AMINO ACIDS				Isoionic Point, <i>pI</i>
Amino Acid	<i>pK</i> ₁ (—COOH)	<i>pK</i> ₂ (—NH ₃ ⁺)		
Glycine	2.34	9.60		5.97
Alanine	2.34	9.69		6.00
Valine	2.32	9.62		5.96
Leucine	2.36	9.60		5.98
Isoleucine	2.36	9.68		6.02
Serine	2.21	9.15		5.68
Proline	1.99	10.60		6.30
Phenylalanine	1.83	9.13		5.48
Tryptophan	2.38	9.39		5.89
Methionine	2.28	9.21		5.74

AMINO ACIDS CONTAINING ALSO A PHENOLIC OR SULFHYDRYL GROUP

	<i>pK</i> ₁ (COOH)	<i>pK</i> ₂	<i>pK</i> ₃	<i>pI</i>
Tyrosine	2.20	9.11 (NH ₃ ⁺)	10.97 (OH)	5.66
Diiodotyrosine	2.12	6.48 (OH)	7.82 (NH ₃ ⁺)	4.29
Cysteine	1.96	8.18 ²	10.28 ²	5.07

AMINO ACIDS CONTAINING TWO CARBOXYL GROUPS

	<i>pK</i> ₁ (α-COOH)	<i>pK</i> ₂ (COOH)	<i>pK</i> ₃ (NH ₃ ⁺)	<i>pI</i>
Aspartic acid	1.88	3.65	9.60	2.77
Glutamic acid	2.19	4.25	9.67	3.22

AMINO ACIDS CONTAINING TWO BASIC GROUPS

	<i>pK</i> ₁ (COOH)	<i>pK</i> ₂	<i>pK</i> ₃	<i>pI</i>
Histidine	1.82	6.00 (IM ⁺)	9.17 (NH ₃ ⁺)	7.59
Arginine	2.17	9.04 (NH ₃ ⁺)	12.48 (Guan. ⁺)	10.76
Lysine	2.18	8.95 (α-NH ₃ ⁺)	10.53 (—NH ₃ ⁺)	9.74

¹ From a more complete compilation by J. T. Edsall in Cohn and Edsall, *Proteins, Amino Acids and Peptides*, Reinhold, New York, 1943.

² See text for a discussion of the assignment of these *pK*'s.

dissociation constants experimentally, one is faced with the problem of assigning the observed constants to the various known chemical groupings.

Table 3-2 summarizes the dissociation constants, expressed as *pK*'s, of the acidic groups in various amino acids. In the first group, the simple amino acids having only two titratable groups, it is observed

that in all cases there is one relatively strong group having a pK somewhere around 2, and another relatively weak acid group having a pK around 10. As was seen in the preceding section, carboxyl groups usually have pK 's around 4 to 5, whereas the substituted ammonium groups have pK 's in the neighborhood of 9 to 10. It thus seems most probable that the stronger acid function pK_1^* is to be associated with the $-\text{COOH}$, and the other with the $-\text{NH}_3^+$ group. The fact that the pK 's of the carboxyl groups are somewhat lower than those found in simple carboxylic acids can be accounted for, at least qualitatively, by the influence of the adjacent NH_3^+ grouping. In the case of the other acidity constants in the more complicated amino acids, assignment to the various groups is somewhat more difficult. For example, the dissociation of protons from the NH_3^+ group and from the $-\text{OH}$ group in tyrosine occurs at nearly the same pH value. The assignments made in Table 3-2 are based on rather involved interpretation of experimental evidence such as heats of dissociation or spectroscopic data.

In the case of cysteine it appears probable that the pK 's of the $-\text{NH}_3^+$ and $-\text{SH}$ are identical within experimental error, so that there is an equal probability of either group ionizing first. When one has ionized, the pK of the other is raised by about 2.1 units. Hence it is not always possible to assign observed pK values to individual groups.

The Dipolarionic Form of Amino Acids. Considering again the simple case of an amino acid having only a single carboxyl group and amino group, we see that if the lower pK is to be assigned to the carboxyl group, the amino acid in neutral solution must exist predominantly in the form $\text{H}_3^+\text{N}\cdot\text{CHR}\cdot\text{COO}^-$. This point can be clarified by reference to Fig. 3-2. The completely uncharged form could not exist unless the lower pK were assigned to the NH_3^+ group. It thus becomes apparent that amino acids must exist predominantly, at least in aqueous solution, in the doubly charged or dipolarionic form. This form has more commonly been termed *zwwitterionic* after the German term for hybrid or amphoteric. This term and the noun *zwwitterion* have been in common usage for so long that it is difficult to replace them with the somewhat more modern terms *dipolarionic* and *dipolarion*, and the even more descriptive noun

* Since from here on we will be concerned with several pK values in a given molecule, it is necessary to use subscript numbers to distinguish between them. It is usual to number them in the order of decreasing acidic strength, pK_1 referring to the strongest acidic group. For convenience the subscript a will be omitted henceforth.

dipolar ampholyte. The term *amphion* has also been used in the literature.

The complete equilibrium system involved in a solution of an amino acid in passing from strongly acidic media to basic media could theoretically involve the four equilibria and four species shown in Fig. 3-2. The two forms occurring in the center of the diagram carry no *net charge*, that is, they are neither anionic nor cationic. They are said to be *isoelectric* or *isoionic* (a distinction between these terms will be made in a later section).

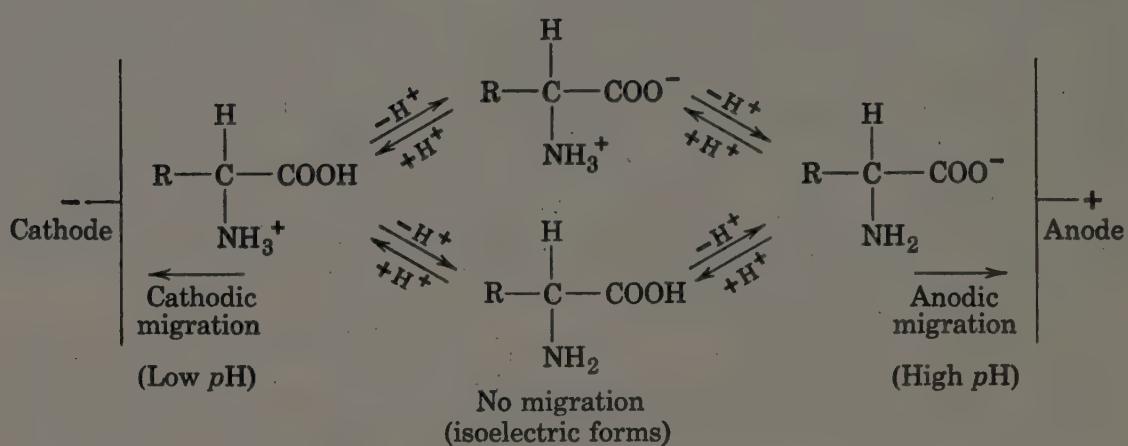


Fig. 3-2. Schematic representation of the dissociation equilibria of an amino acid.

If the uncharged form exists to any significant extent in the isoelectric state, the actually observed dissociation constants would be related to both of the dissociation equilibria pathways shown in Fig. 3-2. A complete analysis of this situation involves information about the dissociation of the ammonium group with an alpha uncharged carboxyl group, or of a carboxyl group with an alpha uncharged amino group. Estimates of the magnitude of the dissociation constant for an ammonium group with an uncharged carboxyl in the alpha position can be made from a study of the esterified amino acids. Such studies have led to the conclusion that the ratio of the dipolarionic form to the molecular form in the isoelectric state, K_Z , is of the order 100,000. In other words, the molecular form can be considered as practically nonexistent in aqueous solution.

It must not be assumed that the molecular form, because it is so far outweighed by the dipolar form, is not of importance in chemical reactions. Actually there exists a rapidly reversible equilibrium, and chemical reactions can funnel through either form. Furthermore, it must be remembered that in nonaqueous media the relationship between the two forms may be quite different. Thus it is calculated

that in 90 per cent aqueous ethanol, the value of K_z is only about 500 to 1000.

Evidence for the Dipolarionic Form. The postulation of the dipolarionic state is necessary for the assignment of reasonable dissociation constants to the chemical groups involved. Considerable evidence from other sources indicates that the dipolar form is indeed the predominant one, both in aqueous solution and also in the crystalline state.

(1) Spectroscopic studies, principally by the Raman method, demonstrate the absence of the characteristic vibration frequencies due to the —COOH group and the —NH_2 group, not only in isoelectric solutions but also in the crystals of the amino acids. Figure 3-3

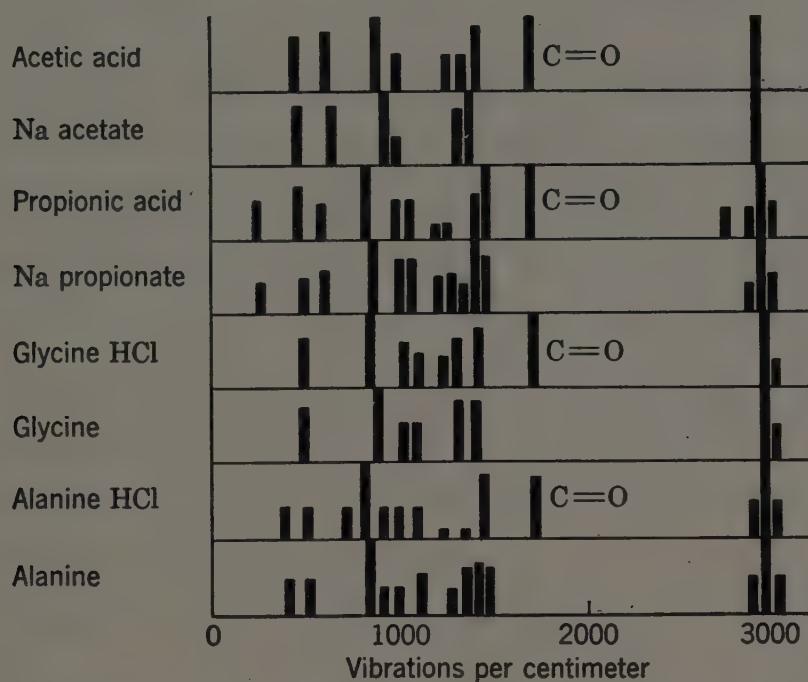


Fig. 3-3. Raman frequencies of fatty acids and amino acids. The Raman frequencies are plotted in reciprocal wavelengths (per centimeter). The height of each line is roughly proportional to its relative intensity. From J. T. Edsall, *Cold Spring Harbor Symposium Quant. Biol.*, 6, 40 (1938).

shows Raman spectra of some amino acids, both in the isoelectric form and in the acid form, as compared with carboxylic acids and salts of carboxylic acids. There is a strong Raman line corresponding to the carbonyl group in the uncharged acid. This line is seen to be absent in the charged or salt form. It will be noted, too, that this line is strongly present in the acidic form of the amino acids, but absent in the isoelectric amino acids.

(2) Solutions of isoelectric amino acids have a very high dielectric constant, as would be expected if the amino acids exist in the dipolarionic form.

(3) The process of solution of isoelectric amino acids in water is accompanied by a strong negative volume change. This is best explained on the basis of electrostriction, that is, contraction due to aggregation of water around the charged groups, which would not be expected if the amino acid existed in the uncharged form.

(4) The high melting point of amino acid crystals, together with their ready solubility in water, is more nearly characteristic of the behavior of ionic than of molecular crystals. In other words, it appears that the predominant force binding the crystals together is electrostatic in nature.

(5) Structural studies of glycine crystals by means of the method of X-ray diffraction indicate that the distribution of the bonds about the nitrogen is tetrahedral, as would be expected if it exists in the cationic form.

(6) Measurements of the enthalpy change (ΔH) of ionization, from studies of the effect of temperature on the titration curve, are in agreement with the dipolarionic concept. The second dissociation is accompanied by a ΔH about ten times as large as that of the first. Simple ammonium compounds have ΔH values of about 10 to 12 kcal. per mole as compared to about 1 kcal. or less for simple carboxylic acids.

(7) The greater reactivity of amino acid derivatives than of the free amino acids, as in the readier formation of diketopiperazines from amino acid esters (Chap. 4), is in accord with the dipolarionic concept.

There are other lines of evidence for the overwhelming importance of the dipolar structure of the isoelectric amino acids. But these, the most important pieces of evidence, should suffice to prove the point beyond all question of doubt.

Distinction between the Isoelectric and Isoionic Points. We have previously used the term isoelectric as applying to the amino acid under conditions in which it carries no net positive or negative charge. Strictly speaking, the term isoelectric is meant to imply that condition in which a substance does not migrate toward either the anode or the cathode in an electric field. Insofar as the total net charge on a molecule is expressed by the binding of protons, the term isoelectric is applicable to the form in which the net binding of protons is zero. Long ago, however, it was pointed out by Sørensen that ions other than protons may be bound or adsorbed. In this case, the pH at which there is no net charge, with respect to the binding of protons alone, may not necessarily be the same as the pH at which there is no migration in an electric field. Sørensen used the term *isoionic* to

refer to the *pH* at which there is no net charge provided no ions are bound other than protons. Thus the term *isoionic point* is more appropriate than the term *isoelectric point* in a discussion of the simple proton-binding of amino acids. Actually, in the case of amino acids the binding of other ions is, in general, so slight that it is a matter of little importance whether or not one distinguishes between *isoelectric* and *isoionic* points. The distinction is very important in the case of proteins (Chap. 11).

Calculation of the Isoionic Point. It is often of interest to calculate the *pH* at which an amino acid is isoionic. This *pH* is usually denoted by *pI*. The calculation can be made in a relatively simple fashion for the monoaminomonocarboxylic acids, and in most cases for the more complex amino acids. In the case of peptides the calculation becomes more complicated, and in the case of proteins even more so. It will be sufficient for our purposes to consider only the simple case of an amino acid with two dissociable groups, one uncharged in the acid form (COOH), and the other cationic (NH_3^+). The two equilibria can then be written in the generalized form (U represents the amino acid)



and



and the corresponding equilibrium constants

$$K_1 = \frac{[\text{U}^\pm][\text{H}^+]}{[\text{U}^+]} \quad (16a)$$

$$K_2 = \frac{[\text{U}^-][\text{H}^+]}{[\text{U}^\pm]} \quad (16b)$$

By solving for the concentrations $[\text{U}^+]$ and $[\text{U}^-]$,

$$[\text{U}^+] = \frac{[\text{U}^\pm][\text{H}^+]}{K_1} \quad (17a)$$

and

$$[\text{U}^-] = \frac{K_2[\text{U}^\pm]}{[\text{H}^+]} \quad (17b)$$

The *isoelectric condition* corresponds to the situation in which the amino acid is neutral with respect to net charge. We can find the desired *pH* by equating the concentrations of U^+ and U^- since

they must contribute to an equal extent.* Hence

$$\frac{[U^\pm][H^+]}{K_1} = \frac{K_2[U^\pm]}{[H^+]} \quad (18)$$

By solving,

$$[H^+]^2 = K_1 \cdot K_2 \quad (19a)$$

$$[H^+] = \sqrt{K_1 \cdot K_2} \quad (19b)$$

or

$$pH = pI = \frac{1}{2}pK_1 + \frac{1}{2}pK_2 \quad (20)$$

It should be emphasized that these equations apply only when the equilibrium constants are expressed in accordance with the Brønsted formulation. Using equation 20 it is possible to calculate the isoionic point in cases in which dissociation constants can be inferred from titration data. In the case of amino acids, such values agree satisfactorily with *pI*'s determined by migration experiments.

In Table 3-2 are included the isoelectric points expressed as *pI*'s for most of the amino acids. In the case of those amino acids containing three or more dissociating groups it will be noticed that the isoelectric point can be approximated by applying equation 20 to two of the *pK*'s. This is more nearly true the wider the spread between the various dissociation constants. The question as to which pair of *pK*'s to average can be answered by a simple consideration of the charge situation. For further details an excellent discussion by J. T. Edsall (1943) is recommended.

In the case of proteins the situation is more involved (Chap. 11). However, in a qualitative manner it can be seen that the isoelectric point depends on the relative number of acidic and basic amino acids present. The protamines, having a preponderance of basic amino acids, have high isoelectric points, in the range of 12. Pepsin, on the other hand, is rich in the acidic amino acids, aspartic acid, and glutamic acid, but very poor in the basic amino acids, and consequently has a very low isoelectric point. In most proteins there is a fairly close balance of these types of amino acids, and the isoelectric points are in general in a *pH* range of about 4 to 7. The amphoteric properties of proteins will be discussed in Chap. 11.

* From another point of view, while a given amino acid molecule exists most of the time in the dipolar state, the chance of finding it at a given instant in the anionic form must be precisely equal to the chance of finding it in the cationic form.

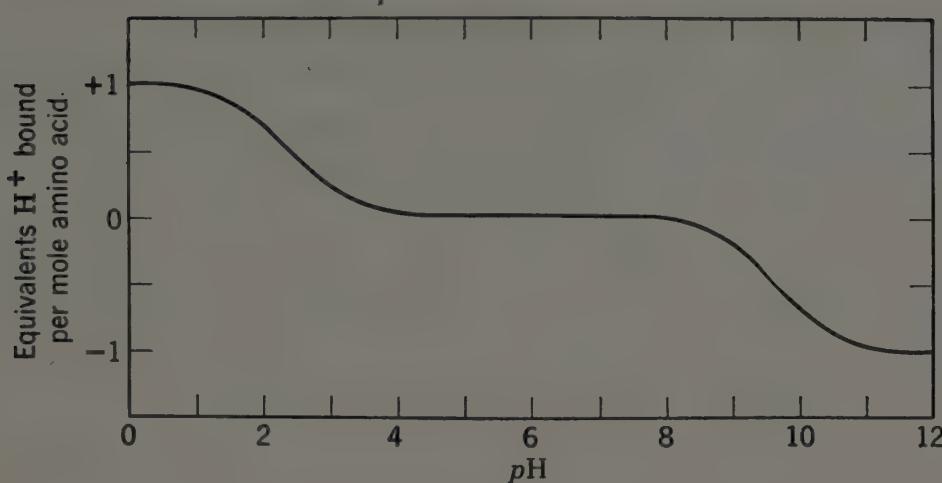
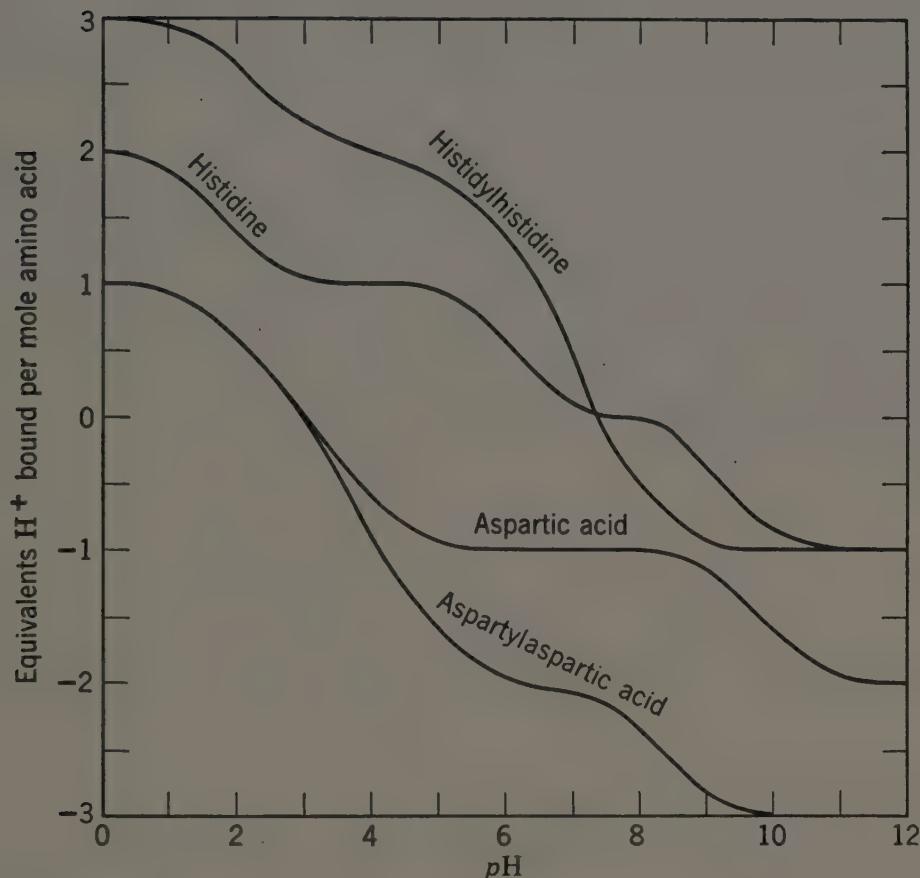


Fig. 3-4. Theoretical titration curve of glycine.

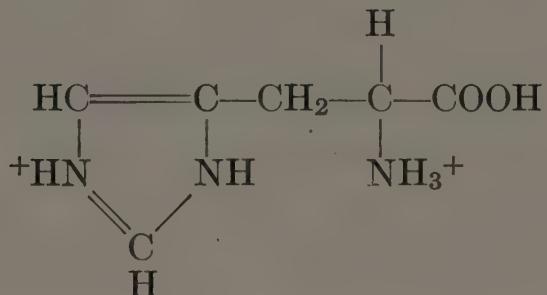
Fig. 3-5. Theoretical titration curves for the amino acids histidine and aspartic acid and the dipeptides histidylhistidine and aspartylaspartic acid. Drawn on the basis of the known titration constants. Data in excellent agreement with these curves have been published by J. P. Greenstein, *J. Biol. Chem.*, 93, 479 (1931).

THE TITRATION BEHAVIOR OF AMINO ACIDS AND PEPTIDES

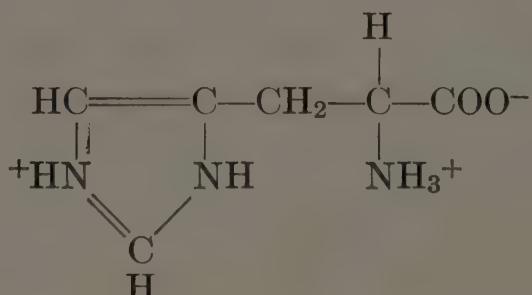
In Figs. 3-4 and 3-5 are plotted the titration curves of some amino acids and dipeptides. In the case of glycine, Fig. 3-4, we see the two titration regions corresponding to the titration of carboxyl groups over the pH range of about 1.0 to 4.0, and ammonium groups over the range 8 to 12. The ordinate in this curve is equivalents of H⁺ bound per

mole of amino acid, and is analogous to the ordinate used in Fig. 3-1, except for the fact that two separate equilibria are involved. At the point where H^+ bound per mole is 1.0, the amino acid is in the cationic form $H_3^+N\cdot CHR\cdot COOH$, which is the acidic form with respect to the first titration. At zero on the ordinate the amino acid is in the isoionic form $NH_3^+\cdot CHRCOO^-$, the basic form with respect to the first dissociation reaction. This is, however, the acidic form with respect to the second dissociation reaction, which leads to the base $NH_2-CHR-COO^-$. The fact that the two pK 's are so widely separated permits the interpretation in terms of two separate titration curves, dissociation of the $-COOH$ groups being substantially complete before dissociation of protons from NH_3^+ begins. Also as a consequence of this separation of pK 's, the isoelectric condition, from the practical standpoint, exists over a broad zone extending from about pH 4.5 to 7.5.

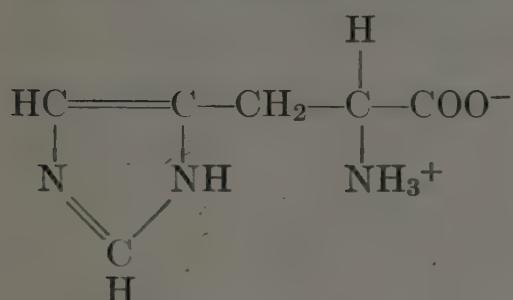
The titration curve of histidine shown in Fig. 3-5 is somewhat more complex in that we have an additional group, the imidazolium ion, which has a pK of about 6.0. The ordinate in this case is again equivalent to that in Figs. 3-1 and 3-4 although it extends from +3 to -3 protons bound per molecule. The situation at the left-hand (low pH) side, where the curve approaches +2 on the ordinate, corresponds to the form



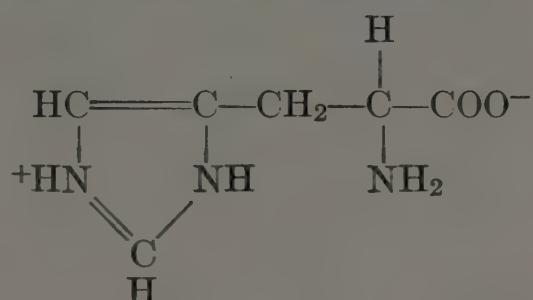
The first titration region again corresponds to the loss of a proton by the $-COOH$ group, and is complete at a pH of about 4. Although there is not the flat portion which was seen in Fig. 3-4, the carboxyl dissociation is substantially complete before dissociation of the imidazolium group begins. The state of the amino acid at pH 4 can then be considered as largely



It will be seen that the curve does not level off following the imidazole titration but overlaps with the titration of the ammonium groups. When two protons have been released corresponding to a pH of about 7.6, the amino acid will exist as a mixture of the forms

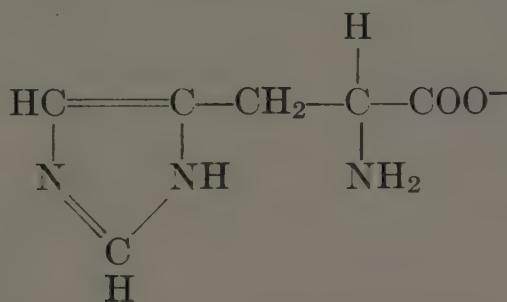


(A)

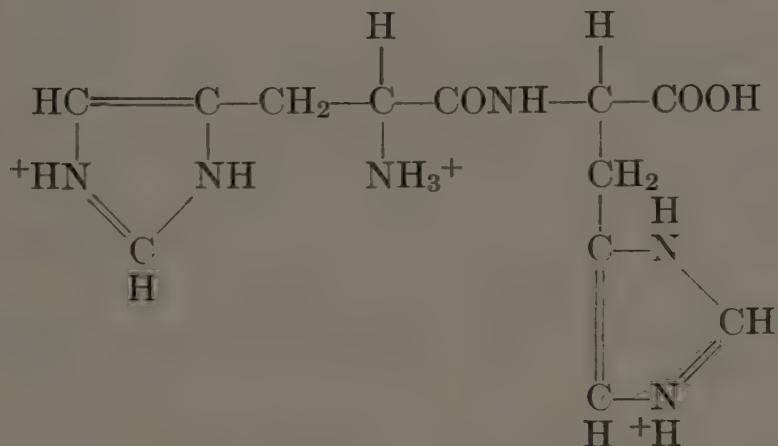


(B)

with form A predominating. Nevertheless, on the average the molecules will carry the same number of positive as negative charges, and are thus isoionic. In this case the isoionic point is seen to be indeed a point rather than a zone. Finally, at a pH of about 11 a total of 3 protons have been removed from the amino acid cation, and it exists in the anionic form



In the case of the dipeptide histidylhistidine, the acid form of which is



(shown also in Fig. 3-5), there is some overlapping between the carboxyl and imidazole regions, and extensive overlapping between the two imidazole and the one ammonium titration ranges. The

isoionic point would correspond in this case to the loss of a proton from the carboxyl to yield a negative group, plus the loss of two more protons, so that only one positive group remained. This situation exists at pH about 7.3.

Similar considerations apply to the titration behavior of aspartic acid and aspartylaspartic acid, shown in Fig. 3-5.

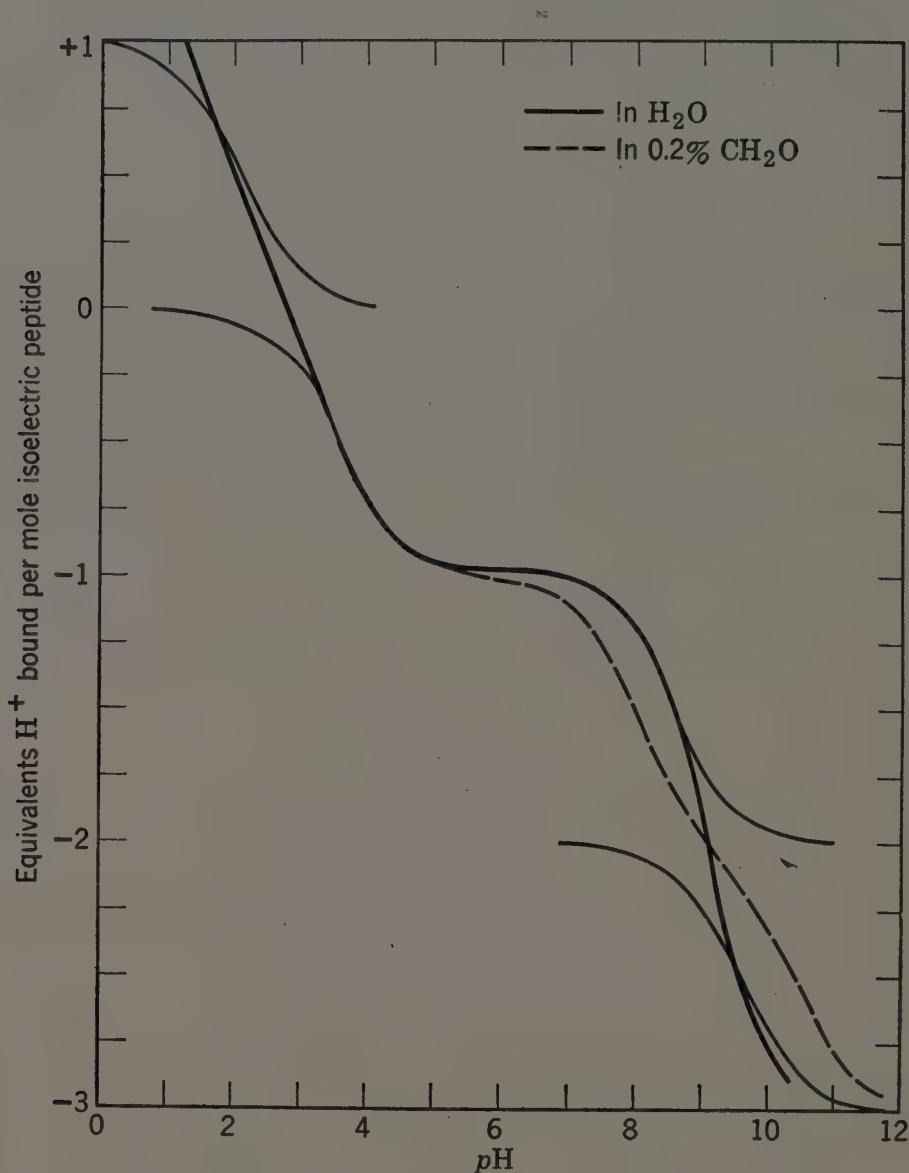


Fig. 3-6. Titration curve of reduced glutathione in water and in aqueous formaldehyde. Adapted from N. W. Pirie and K. G. Pinhey, *J. Biol. Chem.*, 84, 321 (1929). The continuous solid line is the titration curve in water, the dotted curve was obtained in 0.2 per cent formaldehyde. The four sigmoid curves show the manner in which the over-all curve can be fitted with four component titrating groups.

Careful analysis of the titration behavior of peptides and proteins can give some indication as to their composition and structure. A classical example of structural evidence obtained in this way is seen in the analysis of glutathione carried out by Pirie and Pinhey (1929).

Their results on the titration of reduced glutathione are reproduced in Fig. 3-6. Fitting of the curve with typical sigmoid titration curves, as indicated, shows the presence of four titratable groups, having the pK values 2.12, 3.53, 8.66, and 9.62. The first two of these certainly correspond to carboxyls, the other two to the amino and thiol groups. On the basis that the 2.12 pK is typical of a carboxyl group with a free amino group adjacent, and that 8.66 is similar to the value expected for an amino group with a free α -carboxyl group, it was concluded that the glutamic acid residue must be linked through its γ -carboxyl. This finding, which was unexpected at the time, has since been confirmed in several ways.

In the case of proteins, one is limited to conclusions as to the number of various classes of dissociating groups present in the molecule. Detailed structural interpretation is not feasible because of the complexity. The titration behavior of proteins is considered further in Chap. 11.

The Effect of Ionic Concentration and Solvent on Titration Curves. Variation in the ionic concentration, as for example by adding NaCl or other salts, exerts some influence on the titration curves of amino acids and peptides. These effects are a consequence of the influence of ionic concentration on the "activity coefficients"—that is, the effectiveness—of other ions. Such effects are usually fairly small in the case of amino acids, amounting to a shift in the apparent pK 's of a few hundredths or, at most, a few tenths of a unit. In the case of amino acids and proteins the theory is complex, involving the extension of the Debye-Hückel theory of ionic activities to the case of dipolar ions. This problem is discussed somewhat further in Chaps. 11 and 14.

Variation in the chemical nature of the solvent produces large effects on titration behavior, as would be expected. For example, in 70 to 80 per cent ethanol the pK 's of α -carboxyl groups are raised a full unit or more over those in water. Amino groups are affected only slightly. The end point in the titration of amino groups is somewhat sharper in such media, a fact which is utilized in the quantitative determination of certain amino acids by titration in aqueous alcohol, dioxane, or acetone.

Formaldehyde combines reversibly with amines and the amino groups of amino acids. The presence of formaldehyde thus has the effect of reducing the activity of the basic form by combining with it, thereby promoting the dissociation of the $R-NH_3^+$ group. Titration in the presence of fairly high concentrations of formaldehyde (formol) is frequently used in the quantitative determination of amino acids and of amino groups in proteins. This is the basis of the so-called

formol titration of Sørensen. The apparent reduction of the pK of the amino groups permits their titration to a sharp end point with phenolphthalein. An example is seen in Fig. 3-6, where the typical shift of pK in the presence of increasing concentration of HCHO was used to aid in identification of the titration of the amino groups in glutathione.

MELTING POINT

As is well known, one of the most useful physical properties of organic compounds is that of the melting point, inasmuch as this generally provides a characterization of the compound. This criterion is not of such value with the amino acids. In accordance with their saltlike character, they do not have low melting points; in fact, all of them (the imino acids are exceptions) decompose before or during melting. By special means, the melting points can be approximated, and are found in general to be above 200° . The melting points of analogous substituted acids or amines in which the substituent is not a basic group or an acidic group, respectively, are very much lower. This fact can indeed be quoted as partial substantiation of the dipolar-ionic structure of the crystalline amino acids, as we have seen. In line with this behavior, if one considers either the amino acid esters or the acylamino acids, in which the dipolarion structure cannot exist, he finds low melting points. Almost all of the ethyl esters of amino acids, for example, are liquid at room temperature.

SOLUBILITY

The solubility behavior of the amino acids resembles that of the inorganic salts more than that of nonpolar organic compounds. Most of the amino acids are soluble in water; hardly any are appreciably soluble in nonaqueous organic solvents. Proline and hydroxyproline are here again exceptions because of their atypical structure, and are found to be soluble in alcohol. Many amino acids are, however, soluble in alcohols such as butanol saturated with water, a property which has been utilized for the separation of amino acids into groups (Chap. 6).

The solubility of an amino acid is ordinarily quoted for one of two standard solvents, (1) pure water and (2) water adjusted to the isoelectric point of the amino acid. The isoelectric point does in fact represent the pH of minimum solubility. A few of the amino acids are found to be virtually insoluble in water at their isoelectric points. These are cystine, tyrosine, iodogorgoic acid, thyroxine, glutamic acid, and aspartic acid. The alteration of pH by the addition of acid, or especially of alkali, increases the solubility of these and other amino

acids by forming salts. Other electrolytes, especially other amino acids, provide a solubilizing effect in aqueous solution.* For these reasons the solubility of amino acids in protein hydrolyzates is often found to be markedly different from their solubility in water.

TABLE 3-3. Solubilities of the Amino Acids in Grams per 100 Grams of Water at 25°¹

Amino Acid	Solubility	Amino Acid	Solubility
DL-Alanine	16.7	L-Leucine	2.4
L-Alanine	16.7	DL-Methionine	3.4
DL-Aspartic acid	0.78	DL-Norleucine	1.1
L-Cystine	0.011	DL-Phenylalanine	1.4
DL-Diiodotyrosine	0.034	L-Phenylalanine	3.0
L-Diiodotyrosine	0.062	L-Proline	162
DL-Glutamic acid	2.1	DL-Serine	5.0
L-Glutamic acid	0.86	L-Tryptophan	1.1
Glycine	25.0	DL-Tyrosine	0.035
L-Histidine	4.2	L-Tyrosine	0.045
L-Hydroxyproline	36.1	D-Tyrosine	0.045
DL-Isoleucine	2.2	DL-Valine	7.1
DL-Leucine	0.99	L-Valine	8.9

¹ From M. S. Dunn, F. J. Ross, and M. P. Stoddard in *Handbook of Chemistry and Physics*, Chemical Rubber Publishing Co., Cleveland, Ohio.

The solubilities of the amino acids in water at 25° are presented in Table 3-3. Close study of this table will reveal interesting relationships between structure and solubility. For example, phenylalanine is much more soluble than tyrosine.

CONFIGURATION AND OPTICAL ROTATORY POWER

The physical property which fundamentally distinguishes most compounds of natural origin from substances prepared in the laboratory is that of optical activity. This is in turn related to a manifestation of chemical makeup, that is, configuration of the constituent atoms. The occurrence of optically active compounds in nature can be ascribed to the fact that they are synthesized through the agency of enzymes, which are themselves optically active proteins.

Optical Isomers. Optical activity is observed when appropriate substances, in solution or as crystals, are subjected to the action of plane-polarized light. A solution of optically active molecules will rotate the plane of polarization of the light. The direction and extent of this rotation can be determined in a polarimeter. If the plane of

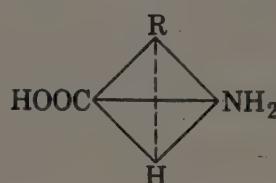
* The principles involved, as applied to proteins, are discussed in Chap. 14.

polarization of the light is rotated clockwise, the direction of rotation is said to be plus (+); it is minus (-) if the rotation is counter-clockwise. With the exception of glycine, each of the amino acids can exist in two optical forms, either of which has a rotatory power equal and opposite to that of its *enantiomorph*, *optical antipode*, *optical isomer*, or *mirror image* (these terms are equivalent in this use).

The most common structural arrangement resulting in optical activity is the presence on one carbon atom of four different attached atoms or groups. For the α -carbon atom of amino acids, this requirement is met by all members except glycine. Molecules are of course three-dimensional, and two-dimensional representations are not adequate for picturing the true state of affairs. One should study this point with actual models or with projections. When either of these is employed, all of the two-dimensional arrangements which may be written will fall into only two actual three-dimensional projections, expressible by



Another apparently different arrangement of groups

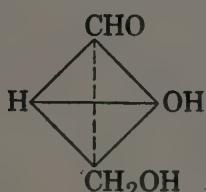


is found to be identical with the D-arrangement if the tetrahedron is merely upset in such a way that any two of the groups occupy the same relative position as those in the D-form. There are thus only two configurational isomers in the simple case.

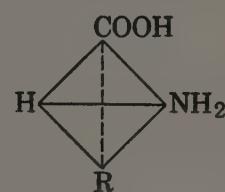
By referring to these antipodes again, we can see that these two forms are not superposable one upon the other, and that in fact each is the mirror image of the other. When examined in the polarimeter, the two forms show equal and opposite rotations of polarized light. All other physical and chemical properties are identical for the two forms, with the exception of the reactivity towards other optically active agents, including enzymes. Our knowledge of configuration of amino acids permits the drawing of many interesting conclusions. It is well, however, to bear in mind that the sign of optical activity does not bear a fully predictable relationship to configuration.

The actual development of the subject has gone through confusing stages of terminology. The modern convention permits the designation of one configuration as D- the other as L-. Optical activity is no longer designated as d- or l-, but only as plus or minus. These symbols are not included in prefixes to amino acids, although they were at one time. L-Arginine, for instance, has been known as d-arginine (optical activity), l-(+)-arginine, and now officially as L-arginine.

Configurations are each determined by relation to configuration of a standard substance D-glyceraldehyde, to which the arrangement of atoms was arbitrarily assigned. The studies of Bijvoet, particularly, suggest that this arbitrary assignment is correct, and that the arrangements presented are the ones that exist. The glyceraldehyde and



D-Glyceraldehyde



D-Amino acid

amino acid diagrammed are both D- because the aldehyde group is convertible to the carboxyl and the hydroxyl substituent is replaceable by amino. These facts are established by chemical interconversion. The actual sign of optical activity of the D-amino acid will vary, being different with different R-groups and with the conditions used for the determination. If the D-form proves to have a plus rotation, all succeeding samples in the same solvent at the same temperature will have the same rotation. The L-enantiomorph must have the opposite or minus rotation.

Effect of pH on the Optical Rotation of Amino Acids. Temperature, the nature and concentration of added electrolytes, concentration of amino acid, and pH each affect the specific rotation of amino acids in aqueous solution. Of these factors by far the most critical, as might be anticipated from the complex amphoteric properties of amino acids, is the pH of the solution. This is brought out strikingly in the curves, constructed from the classical data of Lutz and Jirgensons, shown in Fig. 3-7. Not only does the magnitude of the rotation change markedly with the addition of alkali or acid, but even the sign is frequently pH-dependent. This emphasizes the fact that the sign of rotation, per se, cannot be used as an indication of configuration.

It will be noted that all of the curves reproduced in Fig. 3-7 have certain features in common. In particular, there is in all cases a pronounced increase in plus rotation (or decrease in minus rotation)

upon adding acid to the isoelectric form. This is common to all known amino acids of the L-series. The changes with addition of alkali, while somewhat less characteristic, do have in common a rise in plus rotation,

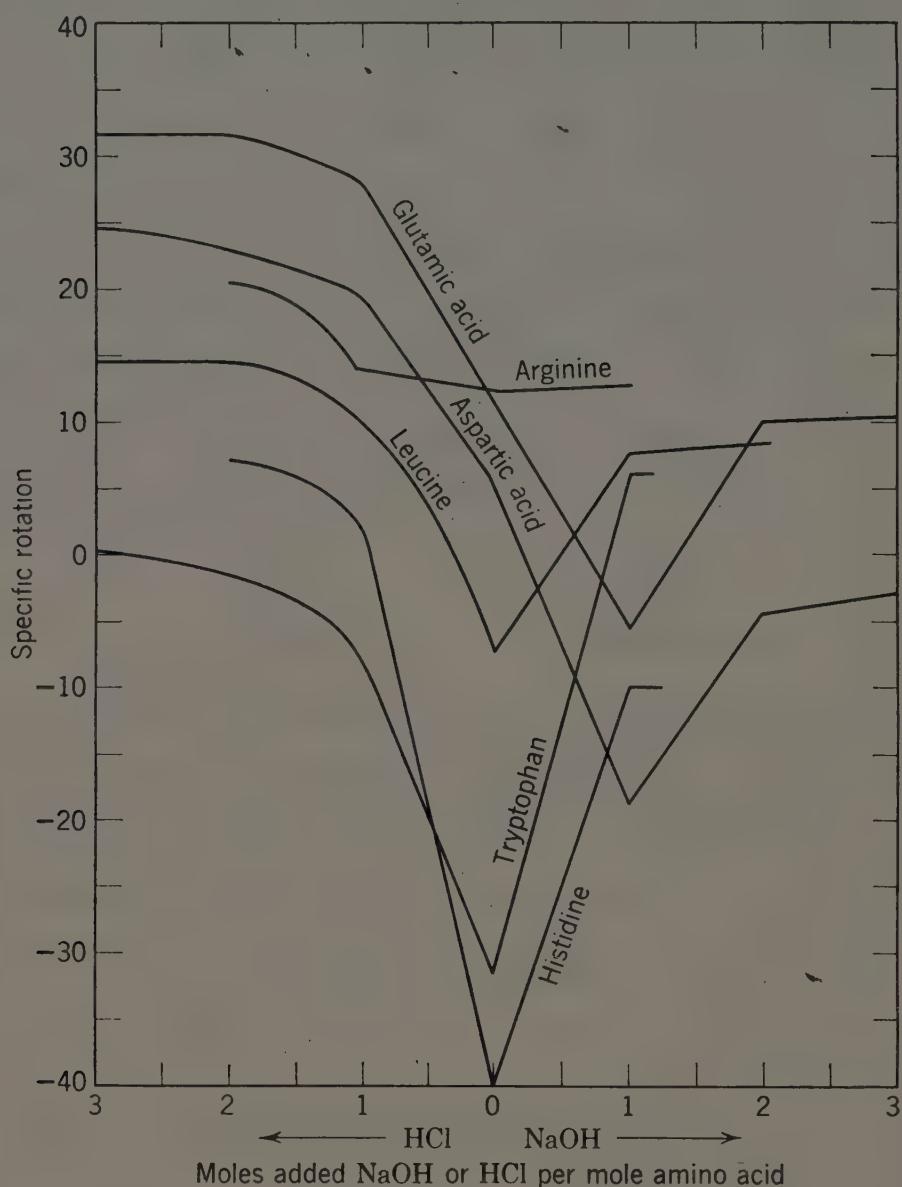


Fig. 3-7. Effect of acid and alkali on the specific rotation of some representative amino acids
Compiled from data of O. Lutz and B. Jirgensons, Ber., 63B, 448 (1930); 64B, 1221 (1931)

beginning either at the isoelectric point or after the addition of one mole of alkali. A study of the direction of these changes provides the most convenient known method for determining the configuration of amino acids.

These changes in rotation can be explained on the basis of the varying proportion of the various ionic forms of the amino acid with changing pH. Each form must have its own characteristic specific rotation, and the net rotation would be the sum of the contributions of each form. The most detailed study of the rotation of any amino

acid is that of Pertzoff on glutamic acid. Fig. 3-8, drawn from some of his data, gives rotation as a function of the net charge per molecule of glutamic acid. (This method of plotting is to be preferred over that used in Fig. 3-7.) The straight lines with the sharp breaks

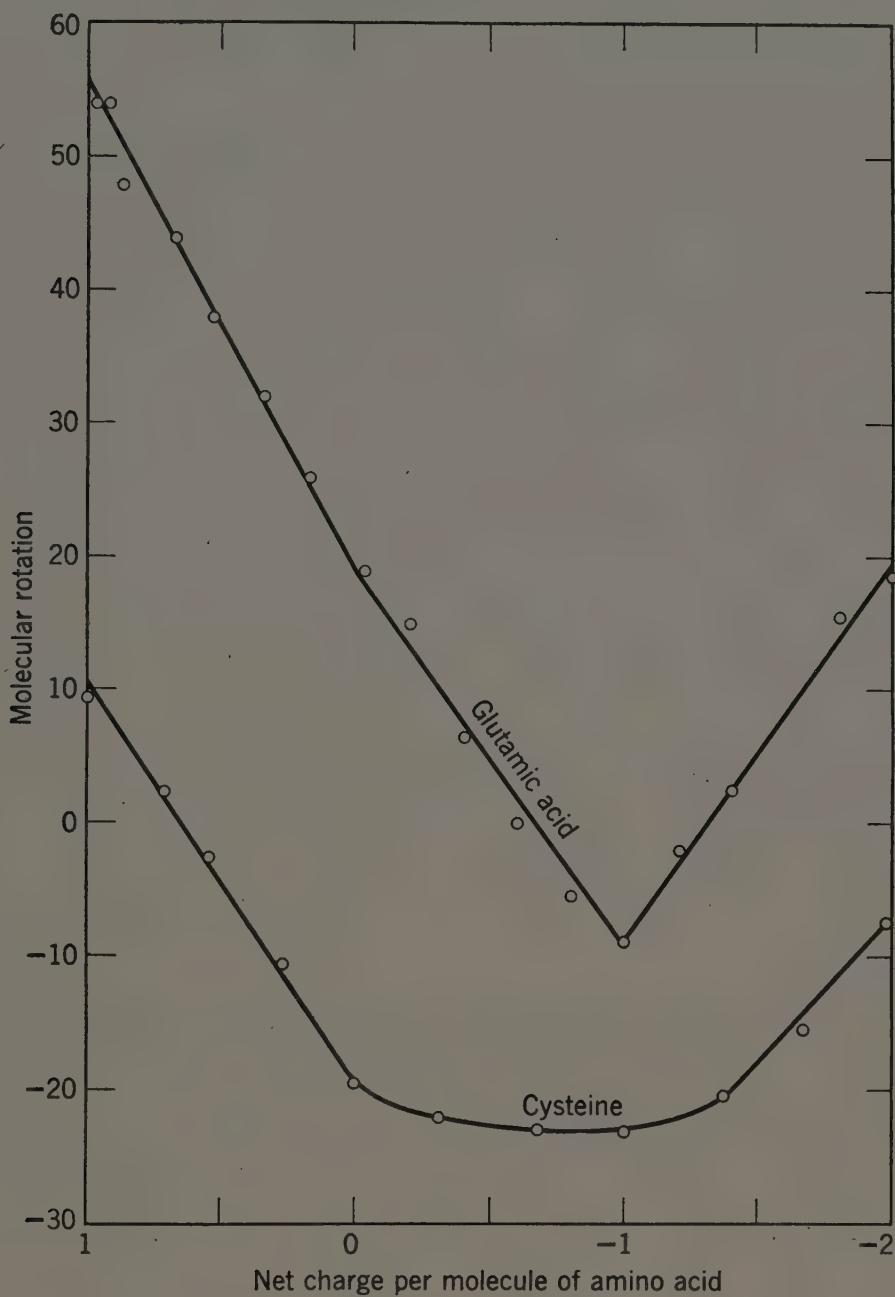


Fig. 3-8. Molecular rotations of glutamic acid and cysteine as a function of net charge. Plotted from data of Pertzoff, Ph.D. thesis, Montpelier University, 1937 and unpublished data of Dr. Dexter French.

occurring at the end points of the various titration regions are in accord with the idea that the various ionic forms are acting independently. It is possible to estimate the specific rotations of the various ionic species, even though they never exist alone, on the basis of this additivity. For L-glutamic acid they are: cationic form (G^+) +

37.55° ; G^\pm (α -carboxyl dissociated) + 12.51° ; $G^+ = -6.55^\circ$; $G^- + 12.35^\circ$.

In the practical use of polarimetry for assessing purity of an amino acid, it is clearly necessary to control in the same way the ionic state. This can be done by buffering, but it is more common to add a strong acid (less commonly alkali) in large molar excess, usually above a ratio of 3 : 1.

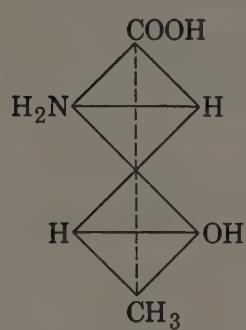
Another fruitful correlation from the above study is the observation that all amino acids which could otherwise be related as possessing the same configuration have curves of the same general shape. The opposite family of configurations, furthermore, has curves which are mirror images of the first. All of the amino acids isolated from proteins and studied in the above way were found to belong to the L-family, and the Lutz and Jirgensons study presented a dominant part of the evidence. The evidence is also in part chemical, and is supported by many earlier studies involving such transformations as reduction of L-serine to an optical form of alanine



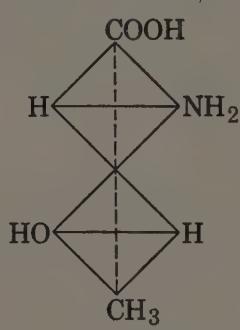
The product of this conversion was found to be identical with natural alanine and otherwise relatable to L-glyceraldehyde.

It should be immediately stated that although the forms of amino acids occurring naturally are predominantly L, D-forms have been isolated, notably from antibiotic substances. Many organisms have been shown to contain an enzyme, D-amino acid oxidase, which indicates strongly that there has been a need in nature to deal with D-amino acids.

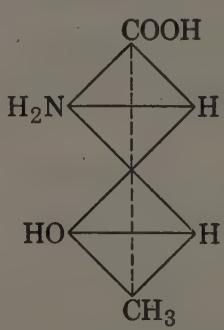
Amino Acids with Secondary Asymmetric Centers. Some amino acids, such as cystine, threonine, isoleucine, and hydroxyproline, contain two asymmetric carbon atoms in their internal structure. Only one form of the configurational pair arising from the non- α -carbon is found naturally. The term *allo* is employed to designate those forms that are not represented by the natural stereoisomer or its mirror image. This relationship is defined for threonine as follows:



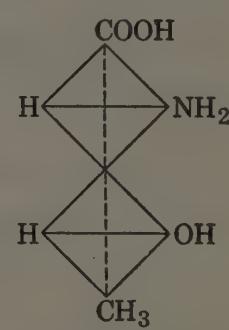
L-Threonine



D-Threonine



L-Allothreonine



D-Allothreonine

INVERSION. Some reagents (nitrosyl halides, ammonia, etc.) are capable of causing a complete change in configuration of an optically active compound. This process is known as *inversion*. When an optically active amino acid is carried through a series of reactions and inversion has been found to occur, it is not always easy to decide at what step inversion occurred or whether it happened once or thrice. An intermediate acid may, for instance, be inverted by a given reagent, whereas the corresponding ester is not.

RACEMIZATION. This term refers to the change of an optical form to an equal mixture of the two forms, called a *racemate*. Racemates frequently possess physical properties differing from those of the antipodes. Racemization of amino acids can be brought about by hot alkali, or by acylating reagents used in proportions to exceed merely the formation of acyl derivatives. Racemization by alkali occurs most rapidly when the amino acids are in peptide linkage.

Whether a racemate exists as a mixture of D- and L-forms or as a compound of the two is not clearly inferential in each case. Theoretically, a mixture of D- and L-amino acids would have twice the solubility of either form, whereas a compound might be less soluble or more soluble than either form. All of these relations may be observed in Table 3-3.

RESOLUTION. The process of separating a DL-mixture or compound of amino acids into the component L- and D-forms is known as *resolution*. This operation is of great preparative significance, and is treated more fully in Chap. 5.

Physiological Manifestations of the Dissymmetry of Amino Acids. While the physical and chemical properties of enantiomeric amino acids are identical except for signs of rotation (which are equal and opposite), many biochemical properties of each antipode are different. This is explained on the basis that enzymes, being themselves composed from L-amino acids, affect one form differently than the other. As an example of biochemical differentiation, L-leucine tastes bitter, D-leucine is sweet.

D-Amino acids have been isolated from the capsular substances of various microorganisms, such as the anthrax bacillus, from a peptide-like constituent of ergot, and most strikingly from a considerable number of antibiotics. Among the antibiotics that possess a structure which entitles them to be classed as D-amino acid derivatives (some of them being true peptides) are: gramicidin, gramicidin S, tyrocidine, the penicillins, bacitracin, and the polymixins. In penicillin the amino acid residue is one critical structural feature, since penicillin constructed from the L-enantiomorph is without antibacterial activity.

The activity of such antibiotic molecules is undoubtedly a reflection of the specific structure in each case, but the fact that nature has employed this principle so widely suggests that the D-amino acid unit may possess some unique property, either structural or functional, or both. Antibiotics may be considered weapons in the competitive struggle for existence among microorganisms, and the survivability of various microorganisms can be attributed at least in part to the evolutionary survival value of the antibiotics which they produce. The amino acid unit may contribute in turn to the survival value of such antibiotics (for example, by inhibiting enzymic breakdown). Some compositional studies, however, have shown that D-amino acids occur as part of the make-up of microorganisms more often than they do in higher forms. The antibiotics that are D-amino acid derivatives may thus be in part, or wholly, a reflection of the circumstance that such molecules were present.

A hypothesis developed by Kögl, which has received widespread attention, states that tumor tissue is characterized by a relatively large proportion of D-forms of amino acids, which are absent in a normal tissue. This attractively simple criterion of distinction is based upon experiments in which some D-amino acids have been isolated from tumor-tissue protein in relatively substantial proportions; this has been claimed particularly for glutamic acid. While this finding has not been entirely abandoned, it has been severely criticized in many laboratories, especially on the basis that the isolative techniques employed in the original experiments would lead to sufficient racemization to account for the quantities recovered.

Other findings with D-amino acids include the occurrence of D-leucine in hair of the horse of increasing age, and antibacterial activity for some of the unsubstituted D-amino acids. This activity is not comparable to that of the antibiotics; the property is absent, however, in the L-isomers at the same concentration. As might be expected, this point has received study in nutritional situations (Chap. 7).

The D-forms of amino acids have often been referred to as "unnatural." They have now been found in so many sources, notably a considerable number of antibiotics, that the term "atypical" appears to be more appropriate.

SELECTED REFERENCES

Reviews

"Amphoteric Properties of the Amino Acids and Proteins," D. Hitchcock, Chap. 11 in Schmidt's *Chemistry of the Amino Acids and Proteins*, Charles C Thomas, Springfield, Illinois, 1938.

"Acids and Bases," H. B. Bull, Chap. 7 in *Physical Biochemistry*, 2nd ed., John Wiley & Sons, New York, 1951.

"Dipolar Ions and Acid-Base Equilibria," J. T. Edsall, Chap. 4 in Cohn and Edsall's *Proteins, Amino Acids and Peptides*, Reinhold, New York, 1943.

"Rules for the Nomenclature of the Natural Amino Acids and Related Substances," H. B. Vickery, *J. Biol. Chem.*, **169**, 237 (1947).

"Stereochemistry of Amino Acids," A. Neuberger, *Advances in Protein Chem.*, **4**, 298 (1948).

"Do Tumor Proteins Contain D-Amino Acids? A Review of the Controversy," J. A. Miller, *Cancer Research*, **10**, 65 (1950).

Original Articles

"Zur Kenntnis der Waldenschen Umkehrung," I. E. Fischer, *Ber.*, **40**, 489 (1907).

"The Titration Curve of Glutathione," N. W. Pirie and K. G. Pinhey, *J. Biol. Chem.*, **84**, 321 (1929).

"Über eine neue Methode der Zuteilung optisch-aktiver α -Amino-Säuren zur Rechts-oder Linksreiche," O. Lutz and B. Jirgensons, *Ber.*, **64B**, 1221 (1931).

"Decomposition Points of the Amino Acids," M. S. Dunn and T. W. Brophy, *J. Biol. Chem.*, **99**, 221 (1932).

"Untersuchungen über das konfigurative Verhalten von Isoleucin bei der Waldenschen Umkehrung," E. Abderhalden and F. Schweitzer, *Hoppe-Seyler's Z. physiol. Chem.*, **206**, 116 (1932).

"Über die Konfiguration des natürlichen (-) Tyrosins," S. Goldschmidt and G. Freyss, *Ber.*, **66B**, 784 (1933).

"The Solubility of the Amino Acids in Water," M. S. Dunn, F. J. Ross, and L. S. Read, *J. Biol. Chem.*, **103**, 579 (1933).

"Pouvoir rotatoire des ions de l'acids D-glutamique," V. A. Pertzoff, thesis, Montpelier University (France) 1937.

"Benzoylation of Amino Acids," H. E. Carter and C. M. Stevens, *J. Biol. Chem.*, **138**, 627 (1941).

"The Occurrence of D-Amino Acids in Gramicidin and Tyrocidine," F. Lipmann, R. D. Hotchkiss, and R. J. Dubos, *J. Biol. Chem.*, **141**, 163 (1941).

"The Dissociation of Certain Amino Acids in Dioxane-Water Mixtures," E. L. Duggan and C. L. A. Schmidt, *Arch. Biochem.*, **1**, 453 (1943).

"Apparent Dissociation Constants of Homocystine and Homocysteine," L. R. Ryklan and C. L. A. Schmidt, *Arch. Biochem.*, **5**, 89 (1944).

"Untersuchungen über den optischen Reinheitsgrad von Aminosäuren in natürlichen Eiweisstoffen," K. Weil and W. Kuhn, *Helv. Chim. Acta*, **27**, 1648 (1944).

"Antipodal Specificity in the Inhibition of Growth of *Escherichia coli* by Amino Acids," Y. Kobayashi, M. Fling, and S. W. Fox, *J. Biol. Chem.*, **174**, 391 (1948).

"The Nature of Bacitracin," G. T. Barry, J. D. Gregory, and L. C. Craig, *J. Biol. Chem.*, **175**, 485 (1948).

"Infrared Absorption of Aqueous Solutions of Organic Acids and their Salts," R. C. Gore, R. B. Barnes, and E. Petersen, *Anal. Chem.*, **21**, 382 (1949).

"Preparation of L-Threonine. Interconversion of the Four Stereoisomeric α -Amino- β -Hydroxybutyric Acids," D. F. Elliott, *J. Chem. Soc.*, **1950**, 62.

"Raman Spectra of Amino Acids and Related Compounds. VII. Glycylglycine, Cysteine, Cystine and Other Amino Acids," J. T. Edsall, J. W. Otvos, and A. Rich, *J. Am. Chem. Soc.*, **72**, 474 (1950).

"Determination of the Absolute Configuration of Optically Active Compounds by Means of X-Rays," J. M. Bijvoet, A. F. Peerdemann, and A. J. van Bommel, *Nature*, **168**, 271 (1951).

"The Acid Strength of the -SH Group in Cysteine and Related Compounds," R. E. Benesch and R. Benesch, *J. Am. Chem. Soc.*, **77**, 5877 (1955).

"Étude spectrophotométrique de la dissociation de la fonction sulfhydrique et structure moléculaire de la cysteine," R. DeDeken, J. Broekhuysen, J. Bechet, and A. Mortier, *Biochim. et Biophys. Acta*, **19**, 45 (1956).

Chemical Properties of the Amino Acids

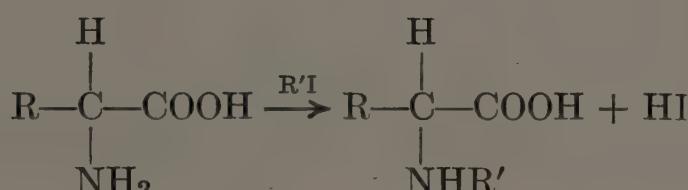
Reactions of amino acids

Reactions of amino acid residues in proteins

In each of the reactions depicted in this chapter that form of the amino acid which is believed to be the predominant participant is chosen to illustrate the reaction. It is to be understood that the other forms are in equilibrium with the one shown, and as the given reaction proceeds, more of the depicted form is generated from the others. Unless otherwise stated, the reactions occur in aqueous solution. When the reaction is known or believed to be reversible, arrows are written in both directions.

REACTIONS OF AMINO ACIDS

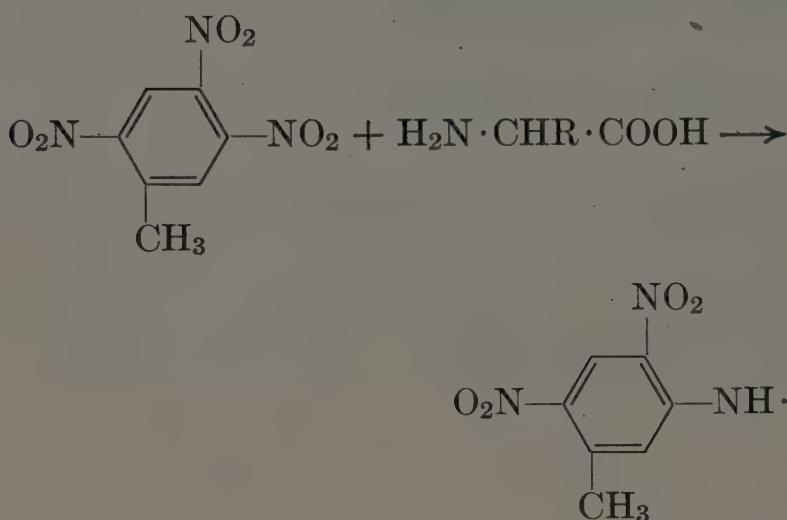
Alkylation.



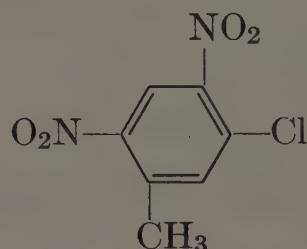
Monoalkylamino acids are obtained with alkyl iodides. The dimethyl derivatives may be made smoothly by use of formaldehyde and catalytic hydrogen.

The trimethylated glycine, known as betaine, is $(\text{CH}_3)_3\text{N}^+\cdot\text{CH}_2\cdot\text{COO}^-$. This substance is found naturally, as is the betaine of 2-thiolhistidine, known as ergothioneine.

Amino acids may also be arylated by 2,4-dinitrochlorobenzene or 2,4-dinitrofluorobenzene. These derivatives have utility in peptide sequence analyses. Such arylation was first performed by Barger by warming an amino acid with asymmetric trinitrotoluene in alcoholic solution. Nitrous acid was eliminated:

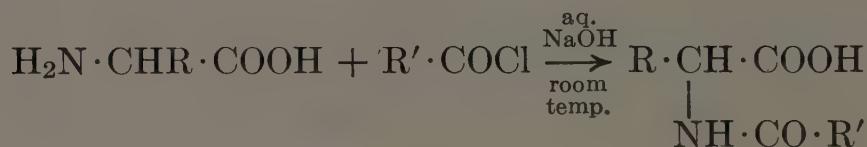


Abderhalden subsequently employed



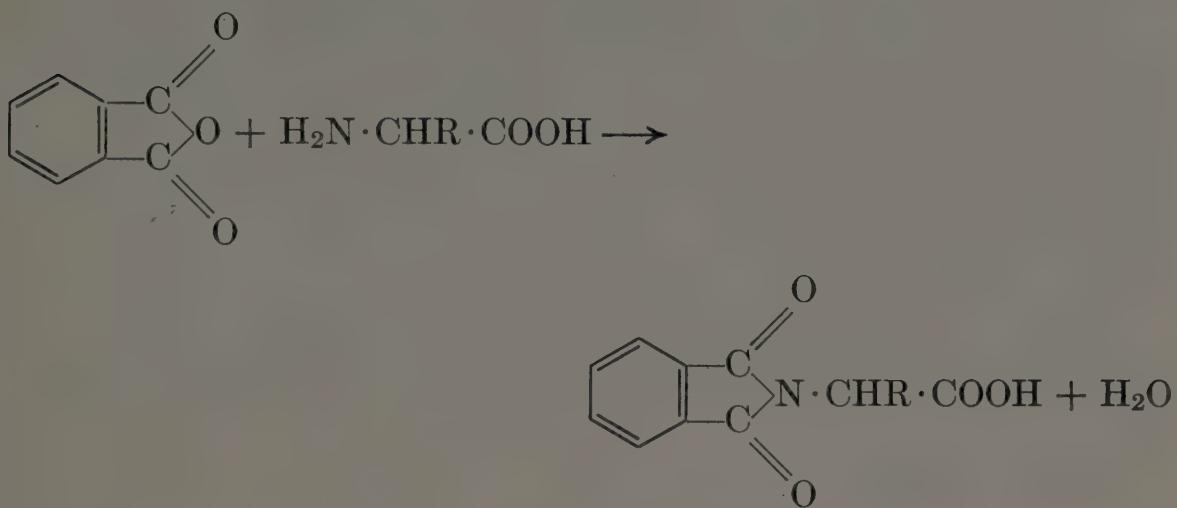
for the same general purpose. Sanger replaced the chlorine by fluorine, and thus supplied a reagent which was more rapidly reactive.

Acylation.



The acylamino acid has many of the properties of an organic acid. Whereas the amino acids exhibit decomposition points rather than melting points, the acylated products frequently possess sharp and characteristic melting points. The benzoyl derivatives are especially useful for characterization.

One important variation in the use of acylating agents is that involving phthaloylation:

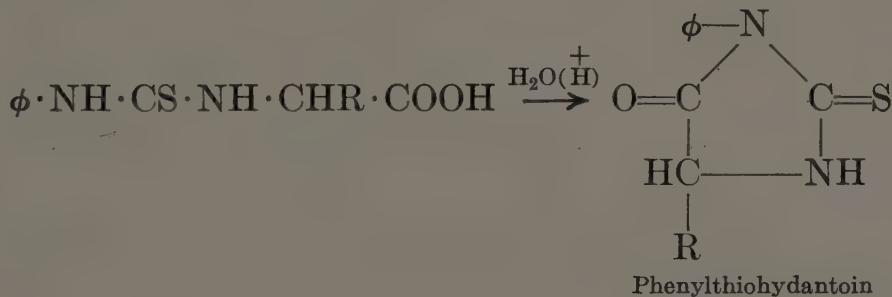


These compounds are of value in peptide synthesis.

Isocyanates and isothiocyanates also react with amino acids to form hydantoic acids:



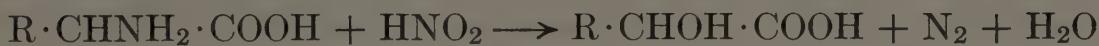
which are then cyclizable under conditions which lead to hydrolysis of ordinary peptide bonds:



Compounds of this type are of value in assignment of residue sequences.

Acetylation may be performed with acetyl chloride, acetic anhydride, or ketene, $\text{CH}_2=\text{C}=\text{O}$.

Nitrous Acid Reaction.



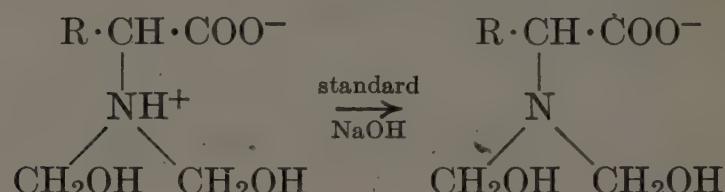
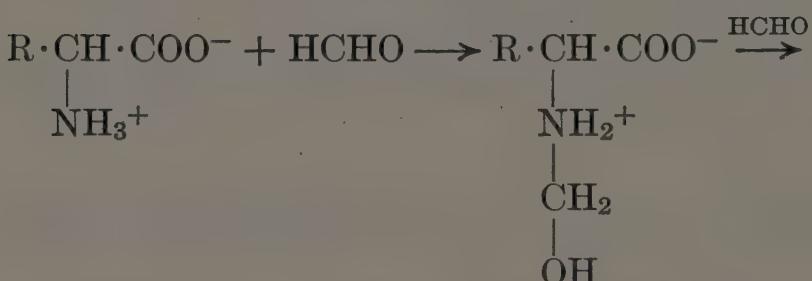
In the reaction with nitrous acid, one mole of nitrogen gas is formed per mole of amino acid. At room temperature the α -amino group reacts completely in less than ten minutes. Under standardized condi-

tions, this reaction is employed for its analytical utility in a gasometric apparatus developed by Van Slyke. The reaction and the appropriate apparatus are accordingly designated by the name of Van Slyke (1929).

The Van Slyke reaction is used for the identification and assay of purity of amino acids. Its greatest utility is in the determination of the extent and rate of protein hydrolysis by various reagents. As hydrolysis of protein proceeds, more primary amino groups are liberated from peptide linkages. These amino groups can be determined by Van Slyke analysis.

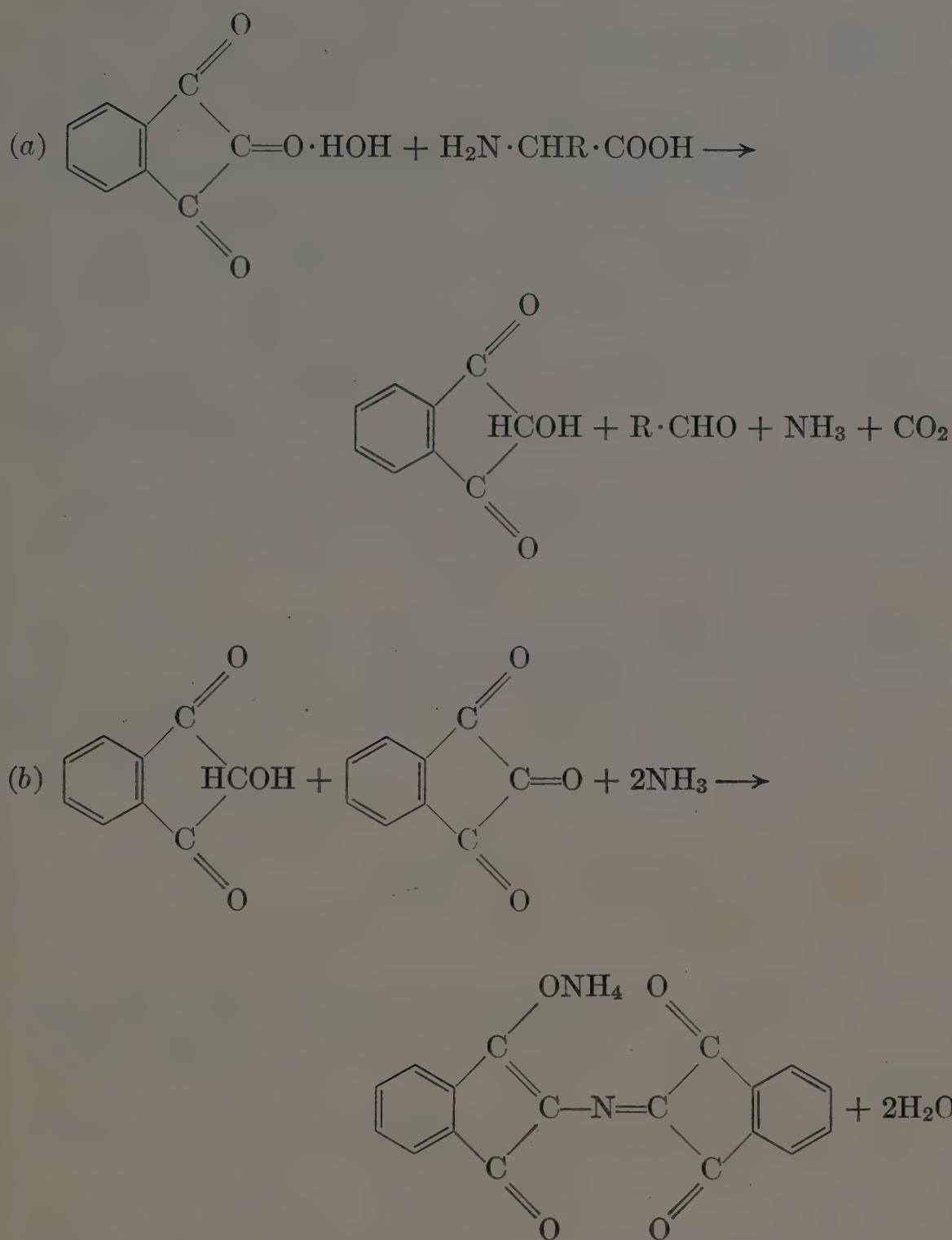
The ϵ -amino groups of lysine react in thirty minutes. When protein hydrolysis is followed by this procedure, it becomes necessary to carry out a blank correction for the ϵ -amino groups of the lysine side chains.

Formol Reaction.



The formol reaction has the same analytical utilities as the nitrous acid reaction. These two analytical procedures have at times been used concurrently, one to check the results obtained by the other. When employed analytically, the reaction is the basis of Sørensen's formol titration method. The titration may be carried out on solutions of whole proteins, such as milk. For the formol reaction, room temperature is employed. In proteins, higher temperatures of reaction lead to further products, which are indicated under the reactions of amino acid residues (p. 62). Aldehydes other than formaldehyde react with amino acids in a manner similar to that of formaldehyde.

Ninhydrin Reaction. A number of polycarbonyl compounds undergo extensive reaction with amino acids. The most studied of these reagents is ninhydrin, triketohydrindene hydrate, which participates in a deaminative oxidative decarboxylation (step a):



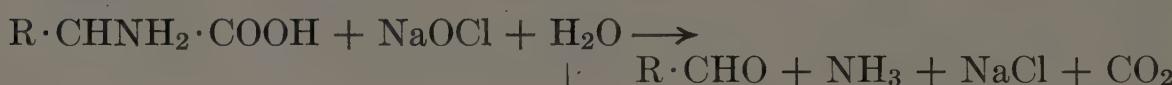
In step *b* the reagent condenses further, to give a blue pigment. This pigment may be used for qualitative or quantitative recognition of amino acid following separation either on paper strip or column (p. 93).

The evolved carbon dioxide has also been measured manometrically. This method therefore supplements the nitrous acid and formol reactions.

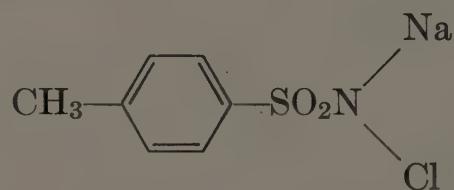
Oxidation to Keto Acid.

The deamination of amino acids to keto acids is a reaction of considerable biological significance. The biological reaction may be simulated chemically by such reagents as H_2O_2 . The catalysts for the reactions in organisms are L-amino acid oxidase and D-amino acid oxidase. The L- and D-amino acid oxidases catalyze oxidation of the typical and atypical forms of amino acid, respectively.

The keto acids which result from deamination of amino acids are identical with some of the keto acids which result from carbohydrate metabolism. The keto acid is therefore sometimes alluded to as the missing link between protein and carbohydrate metabolism. The reactions involved can help to explain, for instance, how protein is converted to carbohydrate in the body.

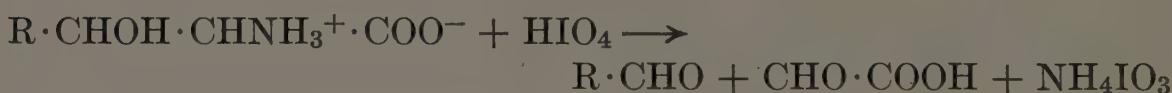
Oxidation to Aldehyde.

Instead of $NaOCl$, chloramine-T



may be employed. This reaction is of interest almost entirely for purposes such as the preparation of aldehydes otherwise obtained with great difficulty.

Periodic acid, when employed as an oxidizing agent on β -hydroxy amino acids, may be used analytically.

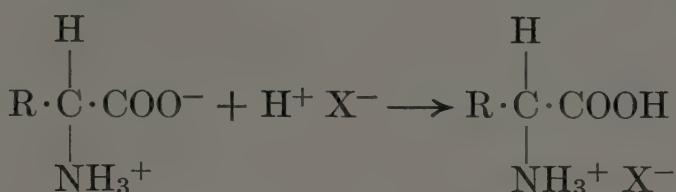


The reaction of amino acid with ninhydrin also leads to aldehyde formation; this reaction has been discussed separately (p. 52) because of its unique value.

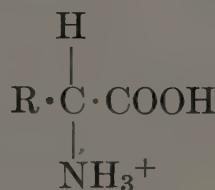
Oxidation to Nitrile.

This reaction differs from the preceding in that a nitrile instead of an aldehyde is obtained when 2 moles of chloramine-T are used as the oxidizing agent.

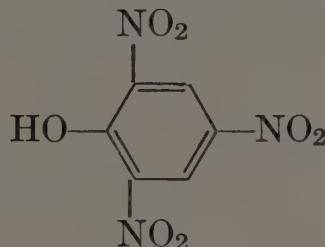
Reaction with Organic Acids.



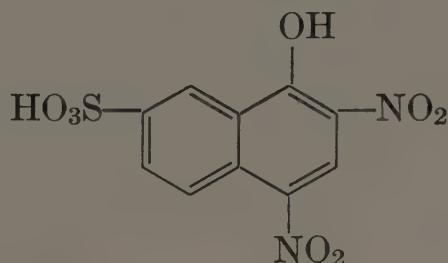
As seen in Chap. 3, strong acids convert the isoelectric amino acids to cationic forms such as



In the case of certain complex organic acids there is an affinity of the acid anion for the amino acid cation, with the result that an insoluble salt precipitates. Examples of such acids are picric acid (2,4,6-trinitrophenol)



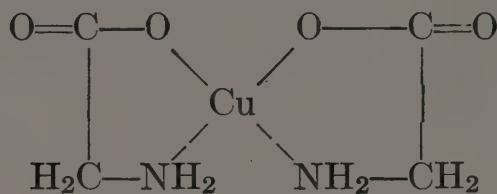
and flavianic acid (2,4-dinitro-1-naphththol-7-sulfonic acid)



Because such salts frequently have definite melting points, they are of value as derivatives in the characterization of amino acids. They are also used for separation of amino acids by selective precipitation. In the basic amino acids, both basic groups can react to yield salts containing two equivalents of, for example, flavianate ion. The flavianate has been used, in the case of arginine, both in preparation and in analysis.

The precipitability of such organic acids with amino acids is found to extend to the proteins (Chap. 14). For these larger molecules it is of course the basic side chains and terminal α -amino groups which react. The reaction product, like the dipolarion, may be looked upon as a substituted ammonium salt.

Reaction with Metal Ions and Complex Salts. The heavy-metal ions, such as Hg^{++} , Cu^{++} , and Ag^+ , react readily with amino acids, forming insoluble precipitates in many cases. Evidence is at hand that some of these are not merely metal salts, but complexes involving the amino group. A typical complex of copper and glycine is considered to be



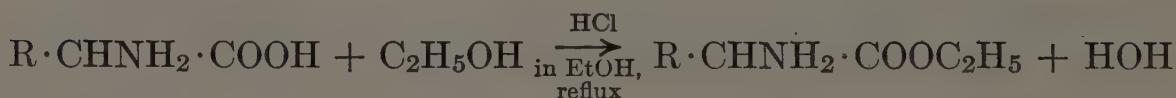
Part of the evidence for complex formation is the deep blue color observable in these copper compounds. This characteristic hue is very similar to that of the complexes of copper and ammonia.

The proclivity for reaction with heavy metal ions is frequently of preparative value; the heavy metal can be removed by decomposition of the complex with hydrogen sulfide, and the amino acid recovered. These same metal ions also precipitate proteins (Chap. 14).

Inorganic complex salts have also been used for selective precipitation of amino acids from protein hydrolyzates. Potassium trioxalatochromate will, for instance, selectively precipitate glycine. The compound obtained has the formula $[Cr(C_2O_4)_3]_9K_{18}(C_2H_6O_2N)_9 \cdot HCl$. Complex salts of the Werner coordination type are used to precipitate selectively hydroxyproline and alanine, as well as glycine.

Useful complex reagents are ammonium reineckate, $[(NH_3)_2-Cr(CNS)_4]NH_4$, which precipitates proline and hydroxyproline, and ammonium rhodanilate, $[Cr(CNS)_4(C_6H_5\cdot NH_2)_2]NH_4$, which can be employed for selective precipitation of proline.

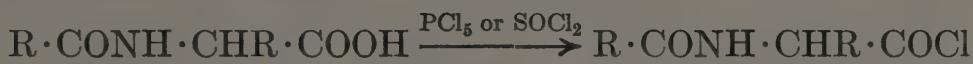
Esterification.



Esterification is an amino acid reaction of considerable historical importance. The esters do not possess the dipolarion structure found in the unsubstituted amino acids, and are therefore more volatile. The amino acids cannot be practically separated by distillation, since

extensive decomposition occurs in such attempts. If one esterifies the amino acids or mixtures of amino acids, however, it is possible to distil the esters without extensive decomposition. The separation of amino acid esters by fractional distillation under reduced pressure is a technique which was developed to a high degree by Emil Fischer and his students. The utilization of this technique required considerable skill and experience, because decomposition occurred readily during distillation under reduced pressure. In the hands of the most skilled operators, a recovery of two thirds of the amino acids in a protein hydrolyzate was considered satisfactory. In spite of these difficulties, for many years the best data on amino acid composition of proteins were obtained by the Fischer ester-distillation method.

Acyl Halide Formation.



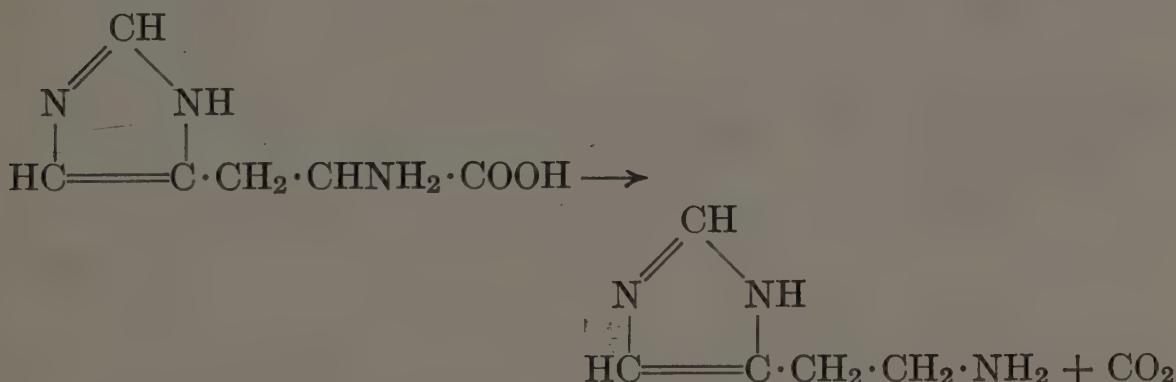
It is not in general feasible to convert the carboxyl group of an amino acid directly to an aminoacyl halide, because of the lability of the amino group in the presence of halogenating agents. The aminoacyl halides are desirable intermediates, however, and some of the purposes of making them can be met by first protecting the amino group, as in the equation given. The protecting group is frequently benzoyl.

Decarboxylation.



Decarboxylation of amino acids is a significant reaction biologically. There are specific enzymes which catalyze the decarboxylation of specific amino acids. In the laboratory, decarboxylation may be induced by heat, acids, bases, or by special reagents such as ninhydrin.

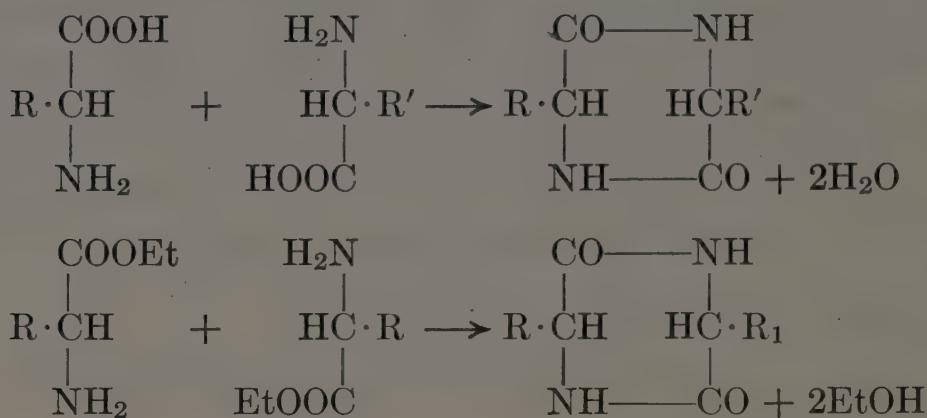
The decarboxylation most studied physiologically is that of histidine to histamine:



Histamine is a direct causative agent in allergy, shock, and similar

states, and is believed to arise by decarboxylation, although the exact precursor is uncertain. Histamine has related pharmaceutical significance. It is manufactured in part by chemical decarboxylation of histidine.

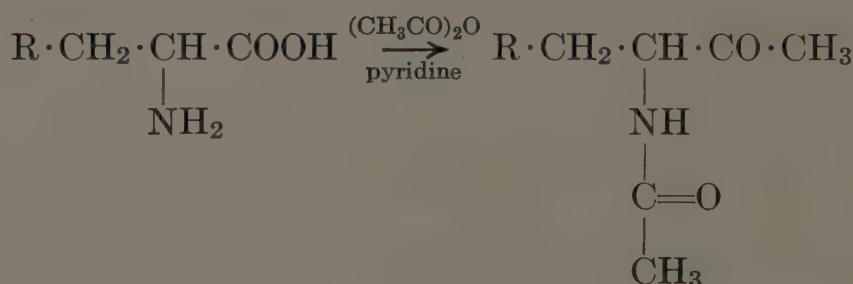
Cyclodehydration.



Two molecules of an amino acid may react with the loss of water to form an amino acid anhydride (or R,R'-substituted diketopiperazine). This kind of amino acid anhydride is to be distinguished from N-carboxyamino acid anhydrides (Chap. 10). The reaction may be brought about by heating the amino acid in high-boiling solvents such as ethylene glycol. The esters of amino acids give the same products more readily, with splitting out of two molecules of alcohol.

The diketopiperazines are of interest because they have frequently been found as products of hydrolysis of proteins. In many cases the anhydride isolated is a mixed anhydride, that is, the two amino acid residues composing it are different. The diketopiperazine structure is of historical interest because of the earlier belief that proteins were fundamentally constructed from the anhydride unit. This hypothesis is not generally honored any longer (Chap. 16).

The Dakin-West Reaction. As an illustration of unique reactions of amino acids, that involving pyridine and acetic anhydride is presented here. Dakin and West found that warming an amino acid with these reagents would lead to an acetyl derivative of an aminoketone, containing one carbon atom more than the amino acid:



The Browning Reaction. Not all of the equations of this reaction or sequence of reactions can be written. The "reaction" is recognized, however, as occurring in at least two stages. In the first stage, the aldehyde group of the sugar combines with the amino group of the amino acid, as in the formol reaction. In later stages, the product is converted to a complex colored material referred to as a *melanoidin*.

The interaction of amino acids and sugars is of much interest in food technology. The melanoidins are responsible for coloration and flavor in fruit juices, beer, etc. The reaction of sugar with free amino groups of protein is also important in food technology. If sugar is removed from eggs before drying, for instance, the proteins of the dried egg retain the ability to redissolve, or reconstitute, which would otherwise be lost. Discoloration is also avoided.

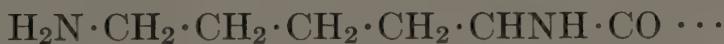
REACTIONS OF AMINO ACID RESIDUES IN PROTEINS

In the reactions given, the behavior of the R-group has been almost ignored. Many of the R-groups are quite functional. The corresponding reactions are of special significance in the protein molecule in which the α -amino and α -carboxyl groups are for the most part shielded by incorporation into peptide linkages.

Reactions of Amino Groups. The two main types of amino group in proteins are (a) the α -amino groups of residues found at the ends of peptide chains in protein.



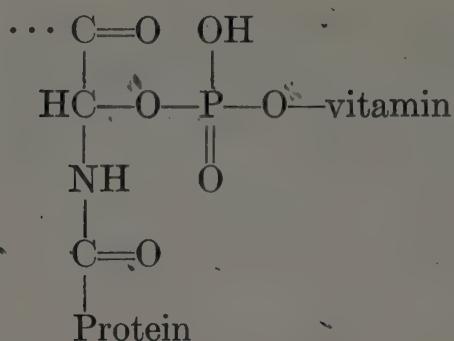
and (b) the ϵ -amino groups of lysine side chains



|

These are capable of undergoing the reactions of the amino groups of amino acids as presented earlier in the chapter. They are, for instance, a primary site of reaction in proteins treated with formaldehyde.

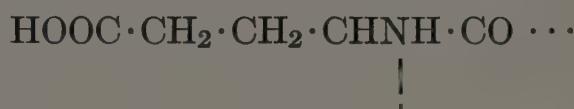
Reactions of the Hydroxyl Groups. Serine, threonine, hydroxyproline, tyrosine, diiodotyrosine, and thyroxine each contain hydroxyl groups. The first three include aliphatic hydroxyl groups, and are capable of participation in ester linkages. Some enzymes are believed to contain linkages of the type



binding a vitamin to a protein through a phosphate ester linkage, or in the case of carboxylase, by a pyrophosphate residue.

The phenolic hydroxyl of the tyrosine type of amino acid exhibits its reactivity in imparting acid properties to proteins. Like the aliphatic hydroxyls, it can be acetylated.

Reaction of Carboxyl Groups. Glutamic acid and aspartic acid residues, by virtue of their extra carboxyl groups, contribute acidic property to proteins:



In many but not all cases, these carboxyl groups are covered as amides. In proteins which liberate much glutamic acid on hydrolysis, a considerable amount of amide nitrogen is also found. The cereal proteins are of this nature.

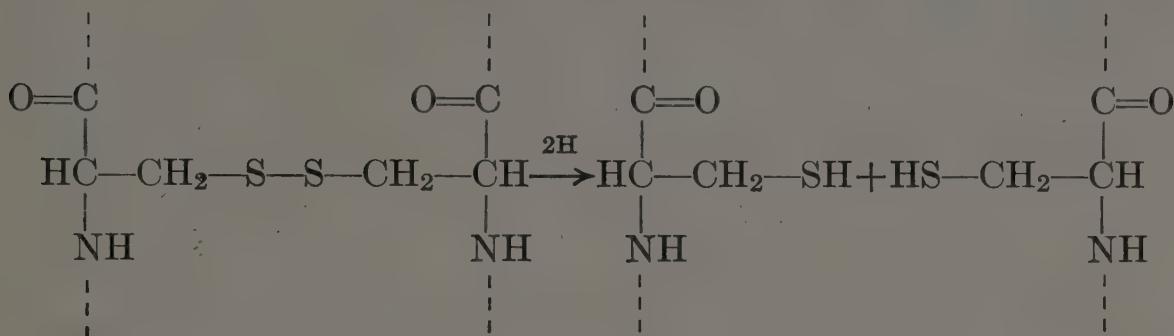
Reaction of Basic Groups. Besides the amino groups already discussed, the basic groups are the imidazole ring of histidine and the guanidino group of arginine. The aromatic nature of the imidazole structure explains why it reacts like the benzenoid amino acids in functioning as a site of substitution reactions.

The guanidino group is highly basic; accordingly the arginine-rich protamines, as cations, readily form salts with acid groups of other proteins.

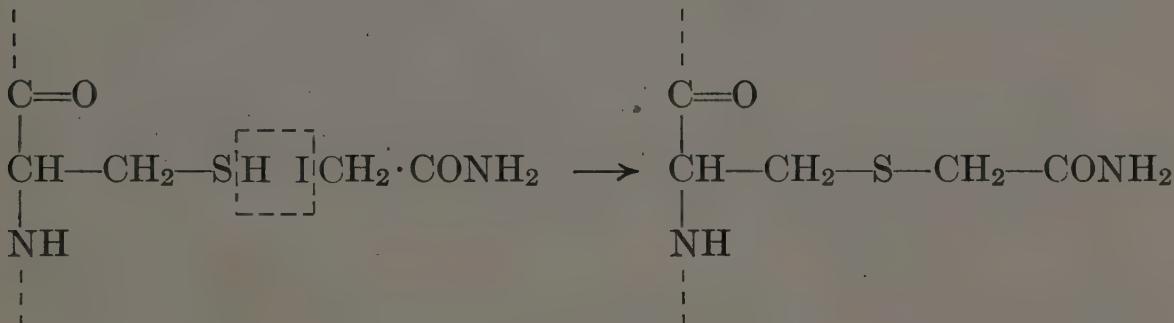
Reactions of the Cystine and Cysteine Residues. Proteins found in nature are known as *native proteins*. Those that have undergone certain changes to be discussed later are known as *denatured proteins* (Chap. 17). The sulfhydryl, $-\text{SH}$, groups of protein are chemically more evident following treatments which cause denaturation. Because the process of denaturation has much biological and technological importance, this group has received much attention.

Inasmuch as the $-\text{SH}$ group is a reducing group, its presence can be recognized by its effect on appropriate oxidized compounds. For example, the dye porphyrindin, which has been employed in such

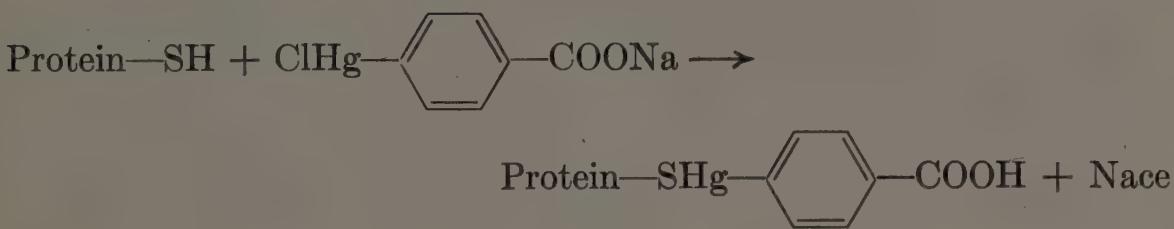
studies, undergoes an observable color change. Disulfide groups also are more evident on denaturation, and can be recognized chemically after first being reduced to sulfhydryl.



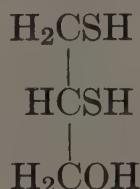
In many chemical studies, the sulfhydryl residues are irreversibly alkylated. A popular reagent for this purpose is iodoacetic acid or its amide:



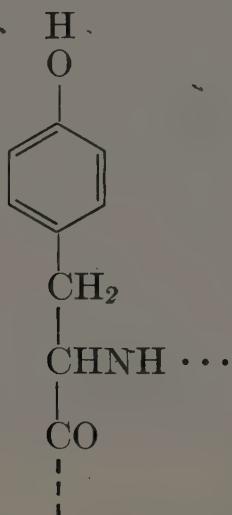
Sulfhydryl groups are also capable of reaction with metal ions. The reagent *p*-chloromercuribenzoate has been useful in studying sulfhydryl groups.



The sulfhydryl enzymes may be protected by added sulfhydryl compounds, such as British Antilewisite (BAL):



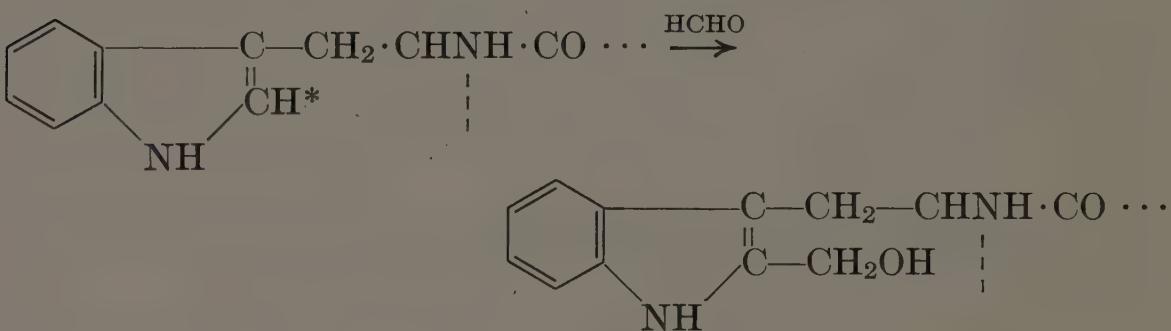
Reactions of the Benzenoid Residues. The typical substitution reactions of benzenoid compounds are encountered in proteins because of the content of aromatic amino acids. The site which has received the most study is the locus ortho to the phenolic hydroxyl of the tyrosine residue.



An often observed reaction of this type results from the spilling of nitric acid on the skin. The reaction involved is essentially nitration of the tyrosine residues of the protein of skin.

The nitrated tyrosine has been isolated from hydrolyzed nitrated protein. This locus is also important in the production of altered proteins for immunochemical experimentation, in which diazonium compounds have reacted at this position.

Reaction of the Indole Nucleus. The H marked with an asterisk in tryptophan



is a reactive hydrogen, and has been shown to take an important part in the condensation of proteins with formaldehyde.

Effects of Substitution on Biological Activity. Table 4-1 gives the results of substitution on the activity of some biologically potent proteins. These substitutions have been made through reactions such as those described in this section.

TABLE 4-1. Essentiality of Groups for Biological Activity
(N = necessary, U = unnecessary)

Protein	Phenol	Amino	Sulfhydryl	Disulfide
Insulin	N	U		N
Diphtheria toxin		N		
Tobacco mosaic virus	N	U	U	
Trypsin	U	U		U
Pepsin	N	U		
Beta amylase (barley)		U		N
Amylase (pancreas)		N		

As arranged from H. Fraenkel-Conrat in D. M. Greenberg's *Amino Acids and Proteins*, p. 543, Charles C Thomas, Springfield, Illinois, 1951.

SELECTED REFERENCES

Reviews

"The Reactions of Formaldehyde with Amino Acids and Proteins," D. French and J. T. Edsall, *Advances in Protein Chem.*, **2**, 278 (1945).

"Reactions of Native Proteins with Chemical Reagents," R. M. Herriott, *Advances in Protein Chem.*, **3**, 169 (1947).

"Chemical Reactions of Proteins," H. Fraenkel-Conrat in Greenberg's *Amino Acids and Proteins*, p. 532 Charles C Thomas, Springfield, Illinois, 1951.

"Zinc and Metalloenzymes," B. L. Vallee, *Advances in Protein Chem.*, **10**, 317 (1955).

"Protein-Iodine Interaction," L. K. Ramachandran, *Chem. Revs.*, **56**, 199 (1956).

Original Articles

"Über die Hydrolyse des Caseins durch Salzsäure," E. Fischer, *Hoppe-Seyler's Z. physiol. Chem.*, **33**, 151 (1901).

"Oxidation of Amino Acids and of Related Substances with Chloramine T," H. D. Dakin, *Biochem. J.*, **11**, 79 (1917).

"Manometric Determination of Primary Amino Nitrogen and Its Application to Blood Analysis," D. D. Van Slyke, *J. Biol. Chem.*, **83**, 425 (1929).

"The Inactivation of Crystalline Insulin by Cysteine and Glutathione," V. du Vigneaud, A. Fitch, E. Perkarek, and W. W. Lockwood, *J. Biol. Chem.*, **94**, 233 (1931).

"Studies on the Constitution of Insulin. III. The Acetylation of Insulin by Ketene," K. G. Stern and A. White, *J. Biol. Chem.*, **122**, 371 (1938).

"Aromatic Sulfonic Acids as Reagents for Amino Acids," D. G. Doherty, W. H. Stein, and M. Bergmann, *J. Biol. Chem.*, **135**, 487 (1940).

"Zur Chemie des Tabakmosaikvirus. Über die Einwirkung von Keten und Phenylisocyanat," G. Schramm and H. Müller, *Hoppe-Seyler's Z. physiol. Chem.*, **266**, 43 (1940).

"Méthode simple de synthèse des 2,5-diketopiperazines (Anhydrides des acides α -amines)," C. Sannie, *Bull. soc. chim. France*, **9**, 487 (1942).

"Effect of Acylating Agents on the Sulfhydryl Groups of Crystalline Egg Albumin," H. Fraenkel-Conrat, *J. Biol. Chem.*, **152**, 385 (1944).

"On the Mechanism of Histamine Formation," E. Geiger, *Arch. Biochem.*, **17**, 391 (1948).

"Amino Acid Impairment in Casein Heated with Glucose," A. R. Patton, E. G. Hill, and E. M. Foreman, *Science*, **107**, 623 (1948).

"Essential Groups for the Interaction of Ovomucoid (Egg White Trypsin Inhibitor) and Trypsin, and for Tryptic Activity," H. Fraenkel-Conrat, R. S. Bean, and H. Lineweaver, *J. Biol. Chem.*, **177**, 385 (1949).

"The L-Amino Acid Oxidases of Snake Venom. II. Isolation and Characterization of Homogeneous L-Amino Acid Oxidase," T. P. Singer and E. B. Kearney, *Arch. Biochem.*, **29**, 190 (1950).

"On the Mechanism of the Reaction of Ninhydrin with α -Amino Acids," D. A. MacFadyen and N. Fowler, *J. Biol. Chem.*, **186**, 13 (1950).

"Electrostatic Effects in the Combination of Cobalt with Amino Acids," C. Tanford and W. S. Shore, *J. Am. Chem. Soc.*, **75**, 816 (1953).

"Observations on the Composition and Activity of Partially Arylated Insulin," G. L. Mills, *Biochem. J.*, **53**, 37 (1953).

"Mikrobestimmung primärer Aminogruppen in aliphatischen und aromatischen Verbindungen, in Aminosäuren und Peptiden, sowie in Carbonsäureamiden und Sulfonsäureamiden," G. Kainz, *Mikrochim. Acta*, **1953**, 349.

"Chemical Interactions of Amino Compounds and Sugars. VIII. Influence of Water," M. L. Wolfrom and C. S. Rooney, *J. Am. Chem. Soc.*, **75**, 5435 (1953).

"Deuterium Studies on the Mechanism of Enzymatic Amino Acid Decarboxylation," S. Mandel, R. Koppelman, and M. E. Hanke, *J. Biol. Chem.*, **209**, 327 (1954).

5 Preparation of the Amino Acids

Methods of synthesis

Resolution

Inversion

Hydrolysis of proteins

Isolation of groups of amino acids

Isolation of individual amino acids

General consideration of preparative methods

The various modes of preparation of amino acids yield either the racemic form or the optically active form. The synthetic DL-amino acids have frequently been of direct experimental use; in other cases, the L-forms, because of their closer relationship to the units isolated from protein, have been required or desired. For this latter purpose, isolation from protein hydrolyzates or resolution of the synthetic amino acid has been employed. The general choice between the overall process of synthesis and resolution on one hand, or isolation on the other, depends upon many factors, such as the availability of intermediates and the experience of the chemist.

The discussion of synthesis of amino acids in this chapter will be of interest primarily to organic chemists, but it has practical implications for other technical fields. As a phase of organic chemistry, the synthesis of amino acids offers broad experience in the practice of preparative techniques. As a field of endeavor, it has historical and cultural values, and scope enough to challenge investigators such as Emil Fischer, who made it one of his principal activities.

Progress in many other areas of protein science was restricted until pure amino acids were available in quantity, as witness our present knowledge of protein nutrition. When it became increasingly clear that the nutritional value of protein would ultimately be expressed largely in terms of amino acid content, the preparation of amino acids of high purity, in quantity, had first to be improved. Conclusions could not safely be drawn from diets made up of amino acids contaminated by other amino acids carried along in the processing of protein hydrolyzates. In furnishing amino acids of high purity in quantity, synthesis has played a leading role.

As mentioned in Chap. 2, methods of synthesis are also critical in establishing the constitution of amino acids, a prerequisite to clarification of the structure of proteins.

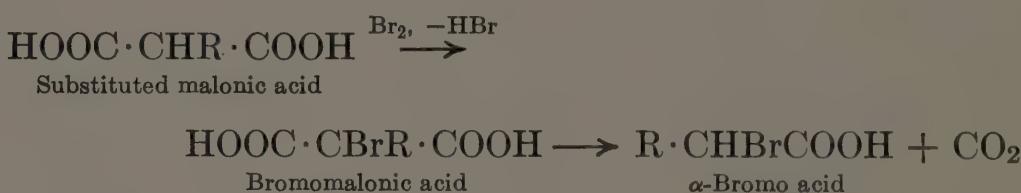
In the synthesis of amino acids, those methods which are best for a number of amino acids of one kind are frequently employed. Although special methods are sometimes, and increasingly, applied for individual amino acids, the general syntheses described here are most commonly used.

METHODS OF SYNTHESIS

α -Halogen Acid Synthesis. The reactions involved in this type of synthesis are essentially



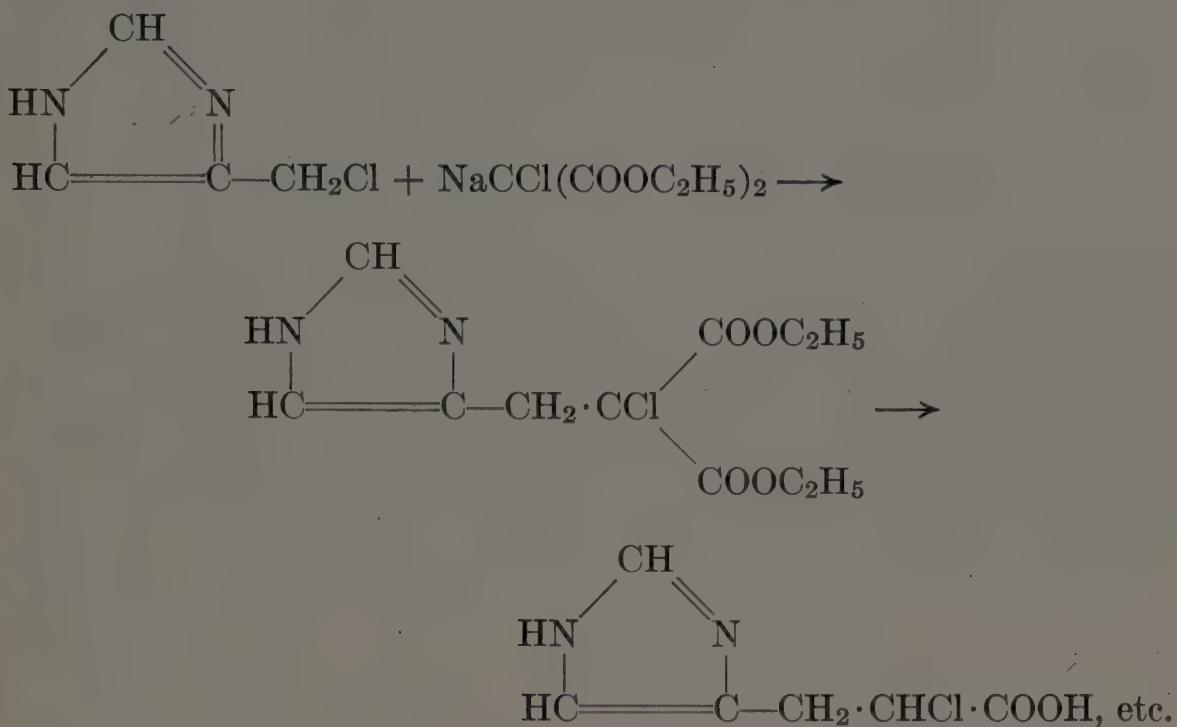
Inasmuch as most of the alkanoic acids corresponding to the mono-aminomonocarboxylic acids are relatively easily obtainable, this method is much favored for these amino acids. In the bromination of acids in the α -position, only a small amount of phosphorus halide is needed as a catalyst. The catalytic reagent introduced may be red phosphorus, which is converted to the halide in the presence of the halogen. The α -bromo acid may also be synthesized through the appropriately substituted malonic acid:



These malonic acids are available through the classical malonic ester syntheses of the organic chemist. It is possible with this method to

employ acids containing heterocycles which would themselves be brominated under the conditions used for aliphatic acids.

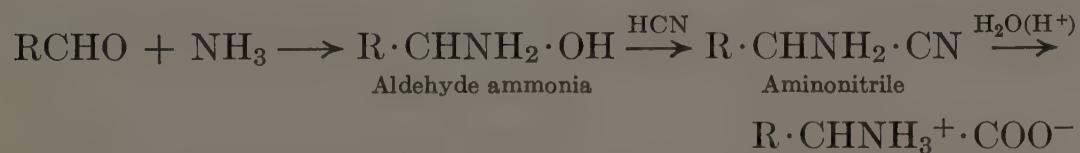
In a further modification, the already halogenated malonic ester may be employed. This route has been used for the synthesis of histidine (Pyman, 1911):



The substitution of halogen by ammonia usually is brought about by treating with an excess of concentrated ammonium hydroxide at room temperature for 2 to 3 days. The substitution may be hastened by operating at higher temperatures with solutions of ammonium carbonate. The ammonia may also be partially substituted as in phthalimide. This latter is more properly described under the malonic ester type of syntheses discussed later.

The α -bromo acids have also been of especial interest for the synthesis of peptides.

Strecker Synthesis. This method bears similarity to classical procedures in carbohydrate chemistry. Named after the worker who first used it to synthesize alanine, the Strecker cyanohydrin synthesis requires the appropriate aldehyde, ammonia, and hydrocyanic acid:



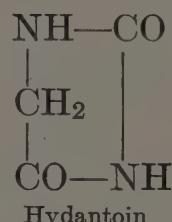
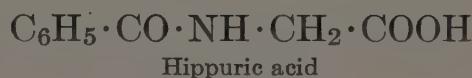
In a modification, an alkali cyanide and ammonium chloride are used instead of the more dangerous liquid hydrocyanic acid.

This synthesis is of value for monoaminomonocarboxylic acids, methionine, and phenylalanine. The starting aldehydes are in general more difficult to obtain than the acids employed in the α -halogen acid method.

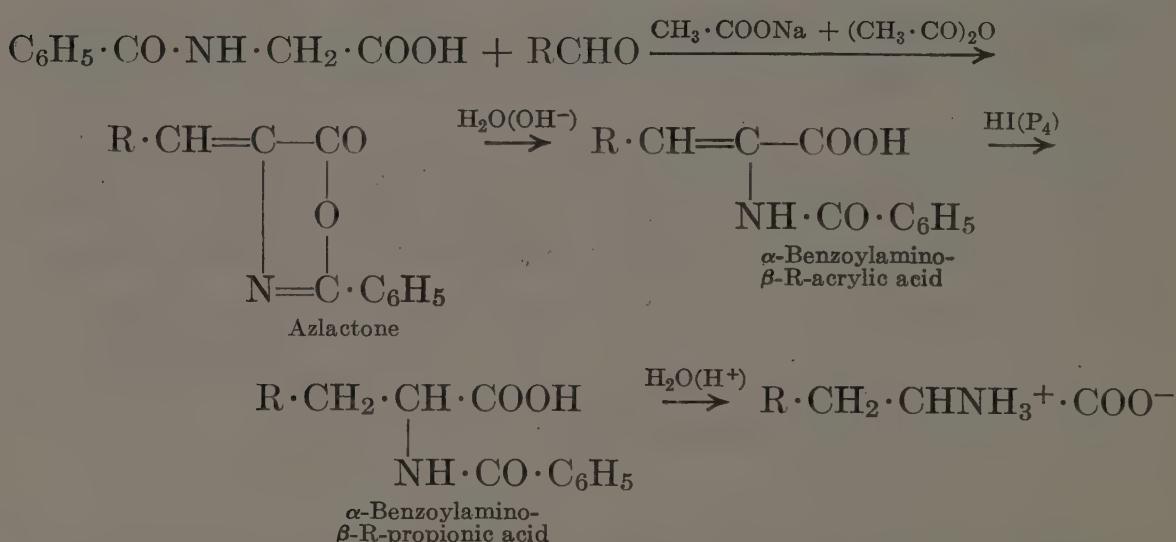
Aldehyde Condensations with Glycine Derivatives. Structurally, all of the amino acids may be considered as derivatives of glycine in which one of the two α -hydrogen atoms of glycine becomes substituted by the side chain:



It is possible to introduce the side chain by reacting aldehydes with the α -hydrogens of glycine. In order to carry out such reactions in accordance with the synthesis described, it is necessary to mask the α -amino group, which would otherwise condense first. This is usually accomplished by employing one of three derivatives of glycine. Glycine is fairly inexpensive, and these three derivatives are also accessible. They are

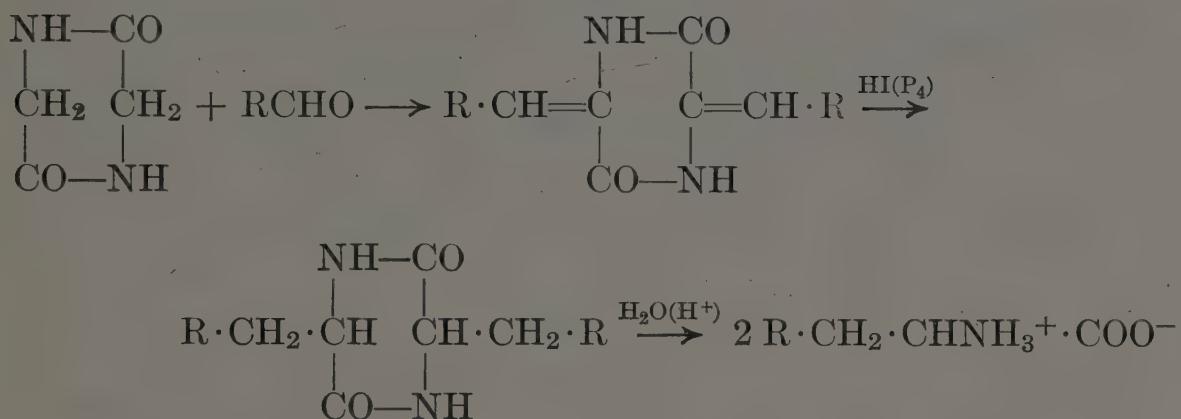


The method employing hippuric acid involves the following steps:



The syntheses through the other two glycine derivatives involve

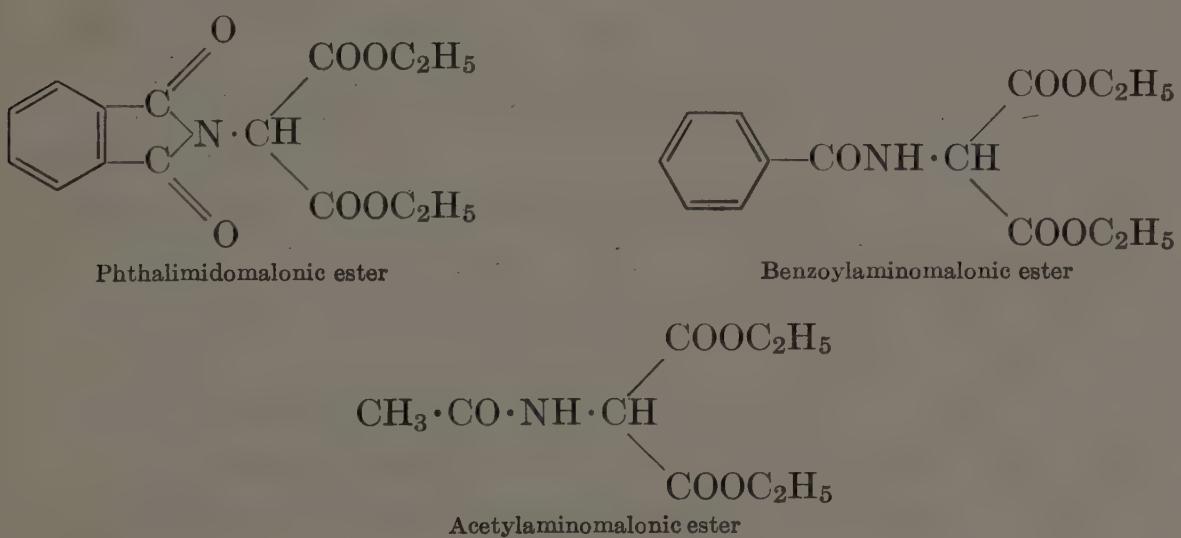
similar reactions (intermediates from only diketopiperazine are illustrated):



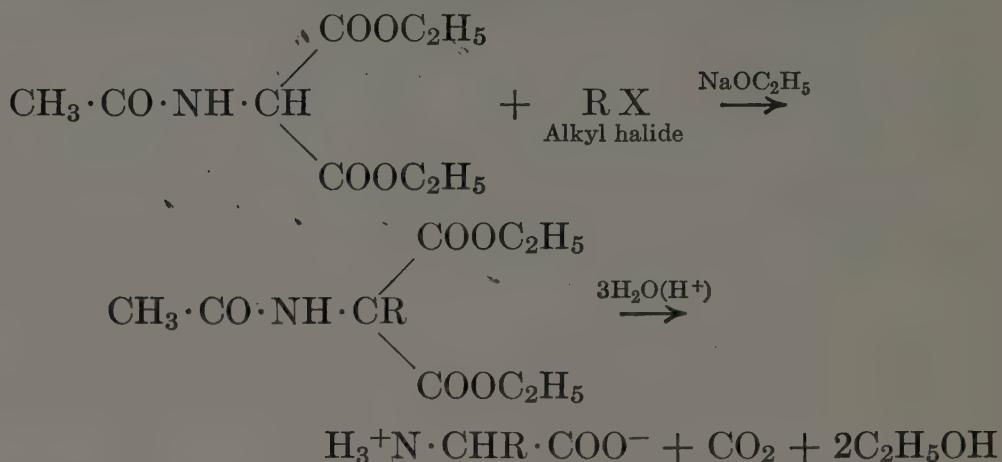
As has been pointed out in Chap. 2, many of the amino acids, such as histidine and thyroxine, can be considered to be made up of a heterocyclic group and a 3-carbon radical, $-\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$; that is, they are β -substituted alanines, $\text{R}\cdot\text{CH}_2\cdot\text{CHNH}_3^+\cdot\text{COO}^-$, the structures obtained as the end products of the above glycine derivative syntheses. The method has little value for the simpler amino acids, but is of considerable worth in the synthesis of the aromatic amino acids.

Malonic Ester Syntheses. One type of malonic ester synthesis has been illustrated under the preparation of α -halogen acids. Although this method holds considerable elegance in the eyes of the organic chemist, its preparative value is overshadowed by that of methods which are less expensive of time and materials.

A considerable improvement can be effected by starting with a malonic ester in which the amino group has already been introduced. This can be brought about by employing one of the following three compounds:



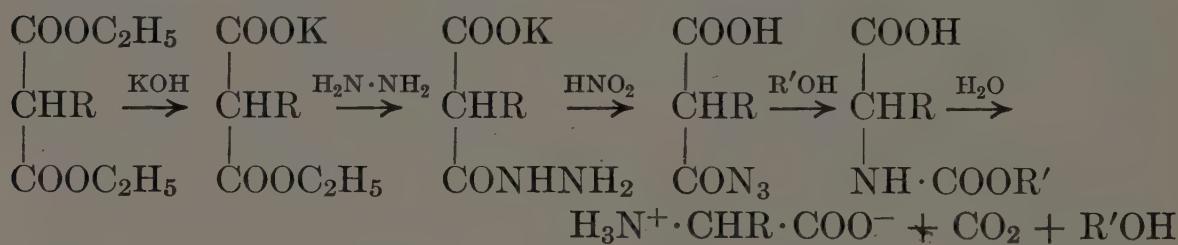
With acetylaminomalonic ester the reactions would be



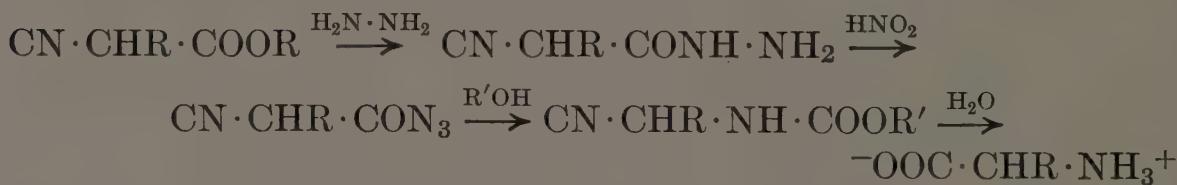
For the other derivatives of malonic ester the reactions are similar, the blocking group being finally removed by hydrolysis in each case.

These procedures are preferable to the simpler malonic ester type because of higher yields; there is less opportunity for polysubstitution when the R-group is introduced. These methods have been used for the preparation of a variety of types of amino acid. Considerable effort, however, is generally required in preparation of the starting materials.

In Curtius's modification, the R-group is introduced originally, and the amino group is introduced last:

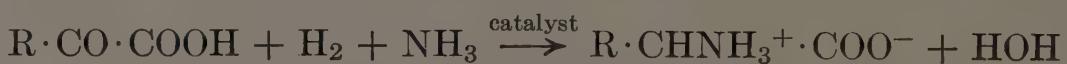


In yet another synthesis, related to the malonic ester type and developed largely by Gagnon, the initial intermediate is a cyanoacetic ester:



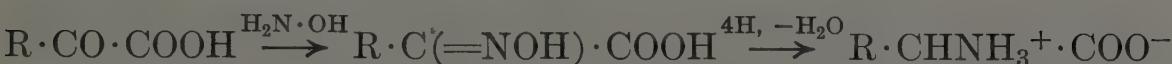
This method has been applied to the synthesis of a large number of amino acids.

α -Keto Acid Reactions. The appropriate α -keto acid may be reductively aminated:



This method has had special value for metabolic studies, by making it possible to obtain amino acids marked with isotopic nitrogen. The

corresponding oxime or hydrazone may be reduced:



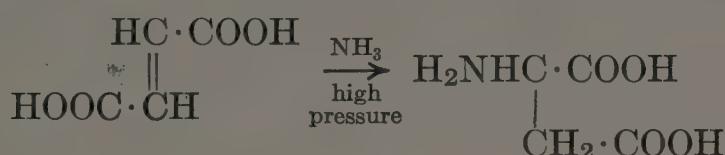
A limitation of these methods is the general unavailability of keto acids.

Oxidation of Amino Alcohols. Comparable to the manner in which alcohols may be oxidized to acids, amino alcohols may be oxidized to amino acids, provided the labile amino group is first protected by benzoylation:

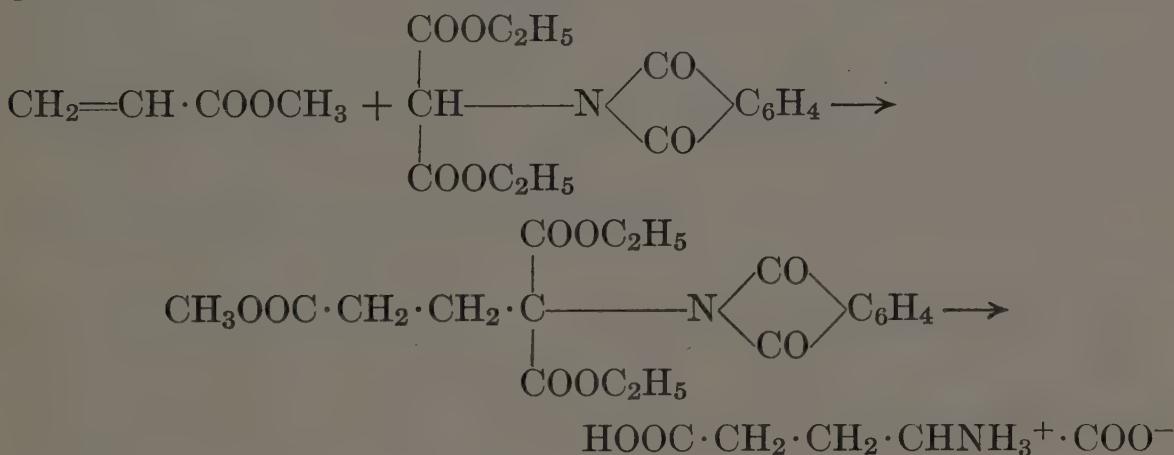


This method is of interest for the monoaminomonocarboxylic acids since the original amino alcohols are increasingly available commercially.

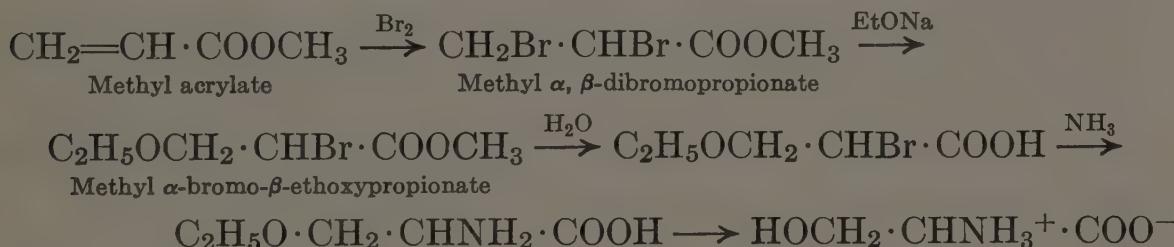
Unsaturated Acid Syntheses. The amino acids with side chain carboxyl or hydroxyl groups may all be synthesized from the appropriate unsaturated acids. For example, commercially available fumaric acid may be converted in one step to aspartic acid.



A modification of this procedure has been used for the synthesis of glutamic acid:

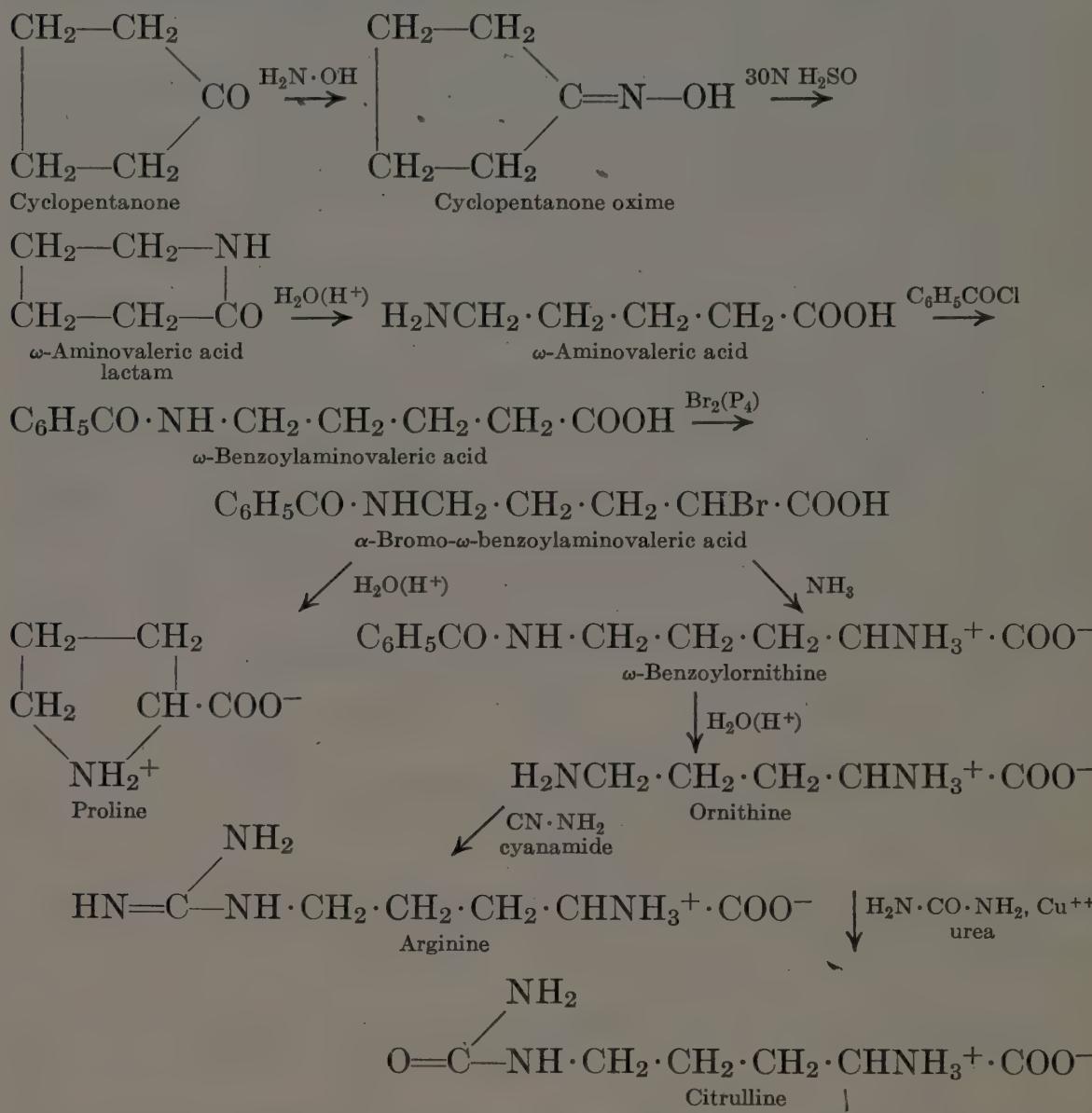


Serine may also be synthesized from methyl acrylate:



Threonine has been synthesized by a similar procedure.

Synthesis through the Beckmann Rearrangement. The Beckmann rearrangement is a useful tool for the synthesis of ω -amino acids, which include arginine, ornithine, proline, citrulline, and lysine.



By using cyclohexanone instead of cyclopentanone, one may obtain lysine instead of ornithine. The Beckmann rearrangement involves the redistribution of the nitrogen atom into such position that it will subsequently be part of a terminal amino group on the side chain.

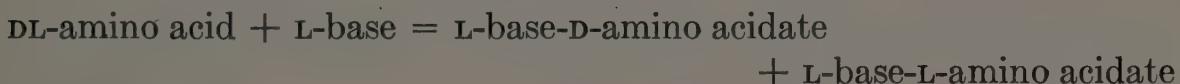
RESOLUTION

The process of separating a DL-mixture or compound of an amino acid into the corresponding L- and D-forms, or resolution (p. 45), has been of the greatest use in providing D-isomers of amino acids for study. Resolution has been comparable to isolation for utility in obtaining L-amino acids. Most resolutions, in terms of necessary effort, have been as demanding as the original synthesis of the DL-

amino acid; the combined synthesis and resolution, however, has the advantage, over isolation, of providing more certain freedom from other amino acids as impurities. Although methods of resolution of amino acids possessing theoretical interest are numerous, enzymic and chemical types have been used predominantly.

The disadvantages of the use of synthetic amino acids without resolution relate to the presence of the D-isomer, which cannot safely be considered to be inert, although this assumption has frequently been made. In some studies, the D-isomer is needed, for example, to check on whether in an experimental project the D-component of a DL-amino acid can be safely considered to be inert. In such instances the chemist may resort to resolution, or less frequently to inversion, of a natural form.

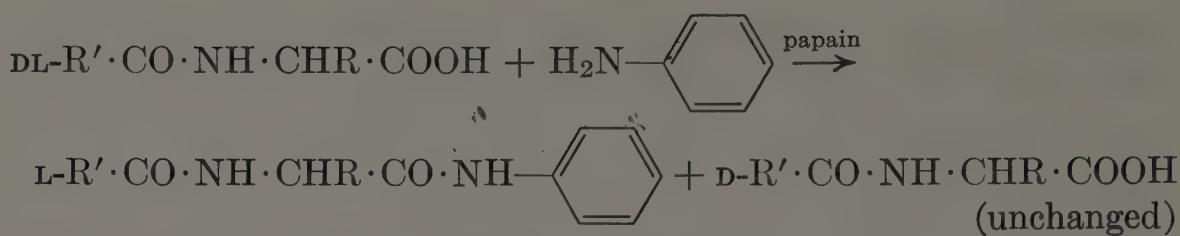
The principle upon which the chemical method of resolution is based is that of the formation of a *diastereoisomer*, or *diastereomer*. The D- and L-forms of an amino acid have identical properties, such as solubility; but if they are combined with one optical form of, for example, a base, they will then participate in forming diastereoisomeric organic salts:



It may be seen from this equation that although the amino acid molecules are optical and configurational opposites, the diastereoisomeric conversion products into which a second center of asymmetry has been introduced are not mirror images. Because of this fact, the two diastereoisomers do not have equivalent properties; their different solubilities frequently make it possible to separate the two forms and to recover the purified amino acid enantiomorphs.

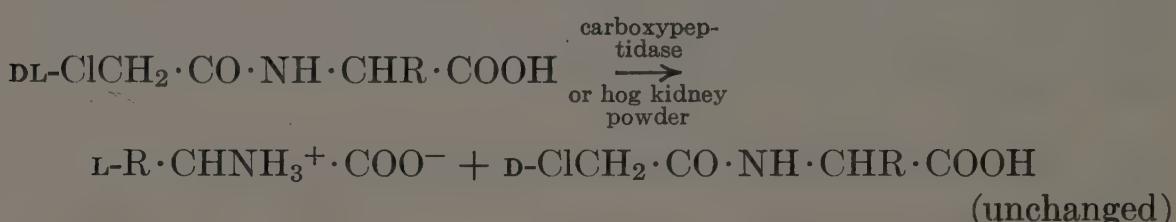
Alkaloids obtained from plants frequently possess the requisite properties, including basicity and single steric configuration. Brucine and strychnine are examples of popular reagents. Such alkaloids as these, however, are weaker bases than the corresponding amino acids. It therefore becomes necessary to mask the relatively strong basic group of an amino acid before it is reacted with the optically active alkaloid. Favorite blocking groups are benzoyl and formyl. The latter, especially, is readily removed by brief hydrolysis with mineral acid; subsequent racemization of the separated optically active form is thus held to a minimum.

In the enzymic method, advantage is taken of the fact that enzymes that act upon peptides show *antipodal specificity*. Papain has been used as in the following series of reactions:



Because antipodal specificity exists for the L-form, only the anilide of this form is obtained, and the acyl-D-amino acid is left unchanged. The two products have markedly different solubility properties, generally more different than those observed with the diastereoisomers obtained by the chemical method. Both the anilide and the acyl-D-amino acid may be hydrolyzed, and the details of a number of resolutions of amino acids have been worked out on these principles.

Another type of enzymic procedure which has been applicable to large quantities of amino acids involves the action of pure or crude carboxypeptidase preparations upon chloroacetyl or other acyl amino acids:

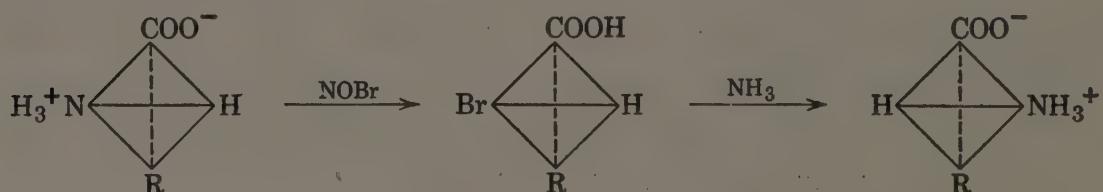


These effective procedures have been developed by Greenstein and co-workers (Levintow *et al.*, 1950). The chloroacetyl-L-amino acid is readily hydrolyzed by the enzyme preparation to yield the L-amino acid, and the D-form can be recovered by acid hydrolysis of the unchanged acyl-D-amino acid after extraction.

The enzymes are capable of participating in resolution by virtue of the fact that they are themselves composed of optically active units. There is good reason to believe that at the moment of the enzyme-substrate interaction, a type of diastereoisomer is formed.

INVERSION

Inversion (Walden inversion, p. 45) has been used to a minor extent to provide a not readily available amino acid isomer, such as D-leucine, from its available enantiomorph. For this purpose, such reactions as represented here have been employed:



The nature of the reagent, the type of substitution in the treated molecule, the type of solvent, and other factors, determine whether or not inversion will occur and at which step in a sequence of reactions.

HYDROLYSIS OF PROTEINS

The hydrolytic agent is in all cases water. The catalyst employed is usually mineral acid, alkali, or proteolytic enzyme.

Acidic Hydrolysis. The most commonly used catalyst is a mineral acid, such as hydrochloric acid or sulfuric acid. Typical hydrolysis with hydrochloric acid employs three to five parts of 20 per cent hydrochloric acid per part of protein, under reflux for 16 to 24 hours. The time may be shortened by hydrolysis with concentrated hydrochloric acid, which loses hydrochloric acid fumes until a constant-boiling mixture results. By this procedure the hydrolysis is completed in 8 to 10 hours, except for occasionally resistant bonds in some proteins. The hydrolysis may also be accelerated by operating under pressure at a more elevated temperature, such as 120° . At these higher temperatures, ordinarily attained in an autoclave, more dilute acids can be used.

When hydrochloric acid is the hydrolytic catalyst, the amino acids bind an amount of the mineral acid equivalent to the amino groups that are freed by hydrolysis. The excess hydrochloric acid may be removed by evaporation under reduced pressure. It is seldom necessary to remove the bound hydrochloric acid. When sulfuric acid, typically 6 to 8 N, is employed, the acid may be completely removed by barium hydroxide or carbonate, or almost completely by the corresponding calcium compounds. A principal disadvantage of a mineral acid catalyst is that most or all of the tryptophan is destroyed under the conditions ordinarily employed. Lesser proportions of cystine, serine, and threonine are also destroyed.

In acid hydrolysis of protein a black solid is produced. This material, known as *humin*, is believed to result particularly from condensation of tryptophan and aldehydes formed during hydrolysis. These latter may be molecules such as glucose, or decomposition products of hydroxyamino acids.

Alkaline Hydrolysis. Sodium hydroxide and barium hydroxide solutions, such as 6 N under reflux for 6 hours, are occasionally used for protein hydrolysis. No humin results, but arginine, cystine, and in part lysine are destroyed. A principal disadvantage of alkaline hydrolysis is that it causes substantial racemization of amino acids, which are then less useful nutritionally. Acid hydrolysis causes very little racemization.

Enzymic Hydrolysis (Enzymolysis). The proteolytic enzymes ordinarily employed are those which carry out proteolysis in the digestive systems of many mammals. Some plant enzymes are also used. Examples include: *pepsin* (stomach), *trypsin* (pancreas), *erepsin* (intestines), and *papain* (papaya tree). The temperature is typically 37° , and optimum pH's are about 1.5 for pepsin, 8.0 for trypsin, 7.0 for erepsin, and 5.0 to 7.0 for papain. The pH optimum is dependent in part upon the substrate. Strictly speaking, pH optima should be designated for enzyme-substrate systems rather than for the enzymes alone (p. 386).

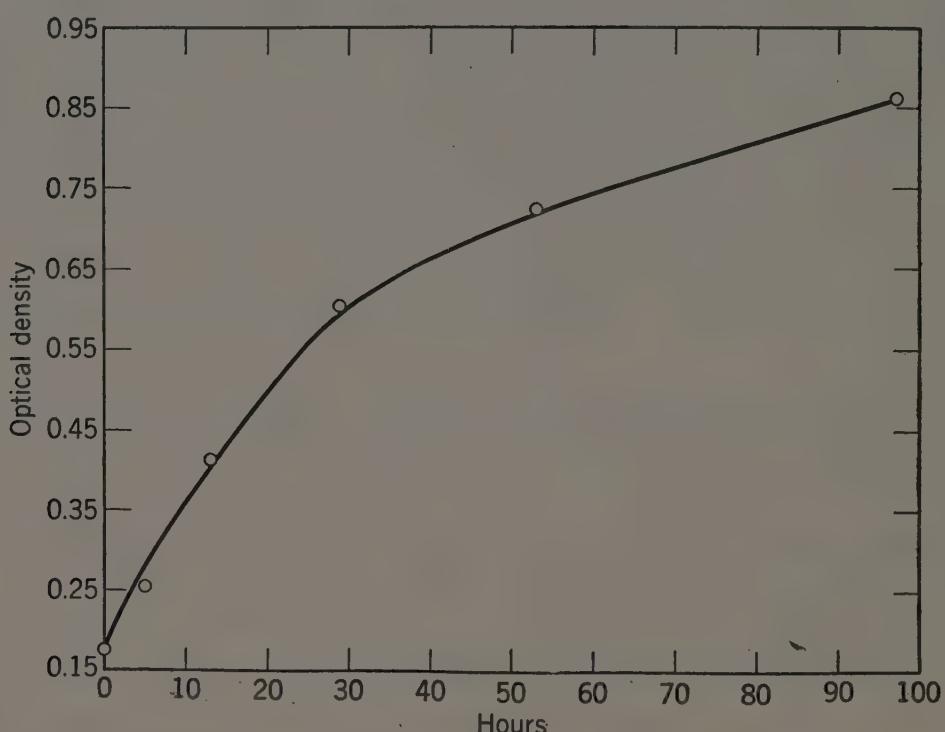


Fig. 5-1. Change in optical density with hydrolysis of lysozyme catalyzed by chymotrypsin. From Ph.D. thesis of T. L. Hurst, Iowa State College, 1953.

Enzymic proteolysis requires protracted action of at least several days for hydrolysis that is not complete. Difficulties in removal of enzymes are also encountered. Tryptophan, however, is not significantly destroyed in this way. Incomplete hydrolysis at 60° with dilute hydrochloric acid simulates enzymic hydrolysis fairly well in conserving much of the contained tryptophan and in bringing about hydrolysis that is not entirely complete. These partial hydrolysates are useful in nutritional medicine, for patients who cannot properly ingest protein in the usual manner.

Estimation of Rate and Extent of Hydrolysis. The rate or extent of hydrolysis may be followed in various ways: by the Van Slyke nitrous acid reaction (Chap. 4), by formol titration (Chap. 4), by titration

of liberated carboxyl in nonaqueous solvents, or by observing changes in optical density following precipitation, with trichloroacetic acid, of the larger residual molecules (Fig. 5-1). It is also possible to assess the release of specific amino acid residues during hydrolyses (Chap. 9).

It is important to recognize that the rate of hydrolysis of a given peptide linkage varies over a wide range with the nature of the component residues. This fact is particularly pertinent to some of the methods employed in assigning sequence of residues in peptides (Chap. 9). The results of a systematic study carried out by Synge are presented in Table 5-1.

TABLE 5-1. Rates of Hydrolysis of Some Peptides

Peptide	Half life, days ¹
Glycylglycine	2.8
Glycyl-DL-alanine	4.5
DL-Alanylglycine	4.5
Glycyl-DL-leucine	7.0
Glycyl-L-tryptophan	8.0
Glycyl-DL-valine	9.0
DL-Leucylglycine	12
DL-Leucyl-DL-leucine A ²	58
L-Leucyl-L-tryptophan	68
DL-Valylglycine	190

¹ In acetic acid-HCl at 37°.

² A is one of the two possible racemates, D-D and L-L, or D-L and L-D.

Adapted from R. L. M. Synge (1945).

It may be seen that the rate of hydrolysis varies widely with the peptide. The position of the residue is important, as can be seen in the comparison of the rates for glycylvaline and valylglycine. The order of rates was found to be similar in alkaline hydrolysis. The slow hydrolysis of valine-containing peptides is believed to be due to steric hindrance. In conformity with this, the more highly hindered glycyl- α -aminoisobutyric acid hydrolyzes under mildly alkaline conditions with infinite slowness.

ISOLATION OF GROUPS OF AMINO ACIDS

The subject of isolation of amino acids is not only of preparative interest, but also logically precedes much of the rationale of assay. In isolating amino acids from protein, the methods described for groups of amino acids are frequently more of academic than of practical interest.

The Fischer Ester-Distillation Method. The first comprehensive method of value was the ester-distillation method developed by Emil

Fischer (p. 56). This method has much historical utility and importance, but is not used in modern investigations.

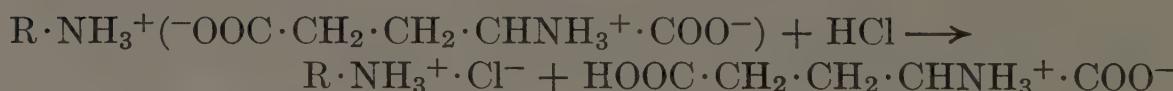
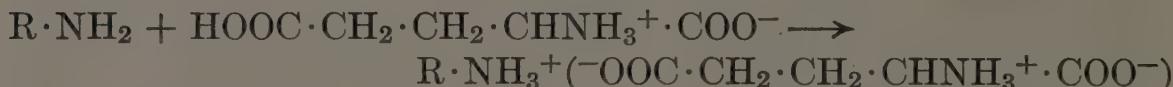
Dakin's Butanol Method. The amino acids may be separated into neutral and other amino acids by butanol-1 saturated with water. The neutral amino acids in aqueous solution are extracted with this solvent. It is of interest that the neutral amino acids are not appreciably soluble in pure alcohol, but are solubilized by saturation of the alcohol with water. The acidic and basic amino acids are left behind.

This method is useful especially in conjunction with supplementary procedures. Tryptophan, for example, may be extracted by the Dakin procedure and subsequently separated from other neutral amino acids by selective precipitation with mercuric salts. The extraction is generally continuous, and the apparatus is simple.

Electrical Transport. This procedure takes advantage of electrical imbalance of the charges on amino acids at pH values other than the isoelectric point (see Fig. 3-2). The original group separation is typically carried out at pH 5.5. Under such conditions the positively charged basic amino acids, lysine, arginine, and histidine, are attracted to the negative electrode (cathode) when a current is passed through the transport apparatus. The acidic amino acids, such as aspartic acid and glutamic acid, are attracted to the anode and are isolated from the solution in the anode compartment. The neutral amino acids are left behind in the central compartment. The solution of the basic amino acids may then be resubjected to electrical transport at pH 7.5, whereupon lysine and arginine migrate into the cathode compartment and histidine remains behind, the isoelectric points of arginine, histidine, and lysine being respectively 10.8, 7.6, and 9.7.

The electrical transport method has been rather widely used, as in the commercial preparation of glutamic acid. The fundamental principles involved are treated in Chap. 3.

Separation on Ion-Exchange Columns. The reactions involved in the isolation of amino acids by ion-exchange columns may be exemplified by the recovery of glutamic acid from a mixture, on an anion exchanger:



The ion-exchange material, represented by $\text{R}\cdot\text{NH}_2$, is insoluble and can be formed into a column. It may be an anion exchanger, as repre-

sented, or it may possess numerous —COOH or —SO₃H groups, in which case it is a cation exchanger. The factors that determine which substances will be held by the column include the relative affinities and the concentrations. In actual operation, the conditions and successful manipulation are attained empirically.

The most practical use of resins has been in the separation of acidic, neutral, and basic groups. It is of interest that Block has observed tryptophan to separate with the basic amino acids in some cases.

Kossel's Silver Salt Method. The utility of this method is limited to the basic amino acids which precipitate selectively with silver. Silver complex salts are formed. The silver complexes of the basic amino acids are insoluble under the conditions employed. The individual basic amino acids can then be separated.

Foreman's Dicarboxylic Amino Acid Method. Foreman's method is applicable only to dicarboxylic amino acids. The barium or calcium salts of an amino acid mixture are prepared in aqueous solution, and several volumes of alcohol are added. The calcium or barium salts of aspartic acid and glutamic acid then precipitate.

ISOLATION OF INDIVIDUAL AMINO ACIDS

For the practical preparation of amino acids, isolation of groups is of limited value. The methods which may be applied to individual amino acids sometimes follow group isolations. Some of the methods for individual amino acids have wide applicability; others are highly specific.

The selective precipitation of insoluble complex salts, and selective adsorption by chromatography of individual amino acids and their derivatives, are methods of particular worth. The latter procedure has been intensively developed. Among other features, it is adaptable to microanalysis. Physical separation of fourteen amino acids from as small a quantity as 0.3 mg. of protein hydrolyzate has been conducted on a paper-strip chromatogram. Although selective adsorption has more value as a microanalytical tool than as an isolative method, special practical utilities are beginning to emerge from the application of this technique. A large number of reagents have been studied and proposed as selective precipitants of amino acids from protein hydrolyzates. By judicious choice of reagents it is possible to isolate a number of amino acids from the same sample of hydrolyzate.

Alanine. Alanine may be isolated from the hydrolyzate of silk fibroin with azobenzene-*p*-sulfonic acid.

Arginine. One of the best selective precipitants is flavianic acid

(2,4,-dinitro-1-naphthol-7-sulfonic acid). It selectively precipitates arginine from protein hydrolyzates. Flavianic acid also precipitates other basic amino acids, but when the material hydrolyzed is gelatin, the flavianate of arginine can be obtained in high purity.

Aspartic Acid. Aspartic acid is often obtained by brief acid hydrolysis of asparagine, in turn readily prepared from lupines.

Cystine. Cystine may be isolated as such by careful adjustment of the pH of the hydrolyzate to the isoelectric range for cystine. Hair, wool, and feather are especially rich in this amino acid, and serve as good sources.

Glutamic Acid. Glutamic acid is best isolated as the hydrochloride from an aqueous hydrolyzate of wheat or corn gluten saturated with hydrochloric acid. If other proteins containing much less than the high content of glutamic acid in wheat gluten are used, this method is not appropriate. Acid hydrolysis of glutamine, concentrated in the liquors of beet-sugar manufacture, also is used to obtain glutamic acid.

Glycine. Glycine may be isolated selectively as the trioxalato-chromiate (p. 56).

Histidine. Histidine may be isolated from hydrolyzates of blood fractions by selective precipitation with 3,4-dichlorobenzenesulfonic acid.

Hydroxyproline. Hydroxyproline is isolated from gelatin hydrolyzates after removal of arginine as the flavianate and proline as the rhodanilate (see proline). For hydroxyproline, ammonium reineckate $[(\text{NH}_3)_2 \text{Cr}(\text{CNS})_4] \text{NH}_4$ is the reagent employed.

Leucine. After removal of tyrosine from a hydrochloric acid hydrolyzate of casein (see tyrosine), it is possible to concentrate the partly neutralized mother liquor to a solution having a low pH and a high salt content. Under these conditions leucine precipitates. Leucine obtained commercially contains methionine, which cannot be removed by simple recrystallization. Methionine can be removed, however, by recrystallization of the formyl derivatives or hydrochloric acid salts, and in other ways.

Lysine. For the isolation of lysine from gelatin hydrolyzate, a preliminary step involves formation of the copper salts in arginine-free solution. Owing to the ϵ -amino group, which does not form a copper complex in the absence of adjacent carboxyl, benzoylation of the copper salt solution leads to an insoluble copper salt (α -complex) of ϵ -benzoyllysine, which is reconverted to lysine.

Methionine. Methionine is isolated from casein hydrolyzate by extraction with butanol, precipitation as mercury complex in mercuric acetate solution, and removal of mercury ion as sulfide.

Proline. After removal of arginine, proline may be isolated from gelatin hydrolyzates by use of the complex acid rhodanilic acid, H [Cr(CNS)₄ (C₆H₅NH₂)₂].

Tryptophan. Enzymic digestion of casein leads to tryptophan, which may be selectively precipitated with mercuric sulfate from a dilute sulfuric acid solution.

Tyrosine. Many proteins yield part of their tyrosine by adjustment of their hydrolyzates to the pH range of 5 to 7 and deposition of precipitates of tyrosine from the solution.

GENERAL CONSIDERATION OF PREPARATIVE METHODS

In order to evaluate the various methods of preparation, it is necessary to test them all on a fairly large scale. Some comparison has been made at the University of Illinois in the course of preparation of many organic chemicals. A study of amino acid synthesis and isolation, with special emphasis on the purity necessary for nutritional research, has been conducted under the supervision of Dunn in another nonprofit organization, Amino Acid Manufacturers, at the University of California at Los Angeles. Following these and other pioneer efforts, many chemical and drug houses now market amino acids. Based on such experience, some economical methods for obtaining amino acids are listed in Table 5-2.

Of the "nonacceptable" amino acids, norleucine and norvaline may be prepared from the α -halogen acids, citrulline and ornithine through the Beckmann rearrangement and other special methods, and dopa through the glycine derivative or from velvet beans.

Many of the methods which have been tested and developed in the academic institutions previously mentioned are now employed in commercial manufacture. It should be emphasized, in connection with a list of preferred methods, that choices of method will vary with the intermediates available, with mother liquors that may be available from other isolations, and with the past experience and preferences of the chemist.

The history of the commercial production of amino acids has emphasized, as for other fine chemicals, a decreasing cost of production. The largest decreases in price have occurred with the onset of large-scale manufacture of a given amino acid, as in the case of methionine, which has been used as a supplement in feeds. Prior to large-scale use, amino acids typically cost, when available, up to several dollars per gram. After large-scale manufacture began in the 1940's, prices of a few dollars per pound began to appear for individual amino acids such as methionine and aspartic acid.

TABLE 5-2. Some Economical Methods for Obtaining Amino Acids

Amino Acid	Method
Alanine	Synthesis from α -halogen acid
Arginine	Isolation as flavianate from glue or fish sperm
Aspartic acid	Synthesis from fumaric acid or ester; purchase
Cystine	Isolation from hair or feather
Diodotyrosine	Iodination of tyrosine
Glutamic acid	Isolation from wheat gluten as hydrochloride; purchase
Glycine	Synthesis from chloroacetic acid
Histidine	Isolation from blood paste as 3,4-dichlorobenzenesulfonate
Hydroxyproline	Isolation from glue as reineckate
Isoleucine	Synthesis from α -halogen acid
Leucine	Synthesis from α -halogen acid; isolation from many proteins
Lysine	Synthesis by Beckmann rearrangement
Methionine	Several special syntheses; purchase
Phenylalanine	Synthesis through glycine derivative
Proline	Isolation from glue as rhodanilate
Serine	Strecker synthesis
Threonine	Synthesis from methyl acrylate
Thyroxine	Synthesis through glycine derivative; special methods
Tryptophan	Syntheses through gramine
Tyrosine	Isolation from many protein hydrolyzates, with glutamic acid and cystine
Valine	Synthesis through α -halogen acid

To cite an important example, glutamic acid, because of the use of its sodium salt as a flavor enhancer, has been available at less than \$2.00 per pound in the 1950's. Original isolation methods, employing protein (gluten) hydrolysis, met competition from the beet-sugar industry, in which glutamine is found in the juice of the sugar beet and appears as glutamic acid in what was otherwise waste liquor.

The era of tonnage production of individual amino acids has begun, and is destined to see increased industrial activity.

SELECTED REFERENCES

Reviews

"Constitution and Synthesis of Amino Acids," M. S. Dunn in Schmidt's *Chemistry of the Amino Acids and Proteins*, pp. 21, 1035, Charles C Thomas, Springfield, Illinois, 1943.

"Natural Amino Acids," H. T. Clarke, Chap. 14 in Gilman's *Organic Chemistry*, Vol. II, 2nd ed., John Wiley & Sons, New York, 1943.

"Synthesis and Isolation of Amino Acids," H. E. Carter and I. R. Hooper in Sahyun's *Outline of the Amino Acids and Proteins*, p. 94, Reinhold, New York, 1948.

"Hydrolysis of Proteins," H. Bull, *Record Chem. Progr.*, **10**, 195 (1949).

"The Preparation of Amino Acids and Polypeptides," S. Archer in Greenberg's *Amino Acids and Proteins*, Charles C Thomas, Springfield, Illinois, 115 (1951).

"Isolation of Amino Acids," M. S. Dunn and L. B. Rockland in Greenberg's *Amino Acids and Proteins*, Charles C Thomas, Springfield, Illinois, 213 (1951).

Original Articles

"Gegenseitige Umwandlung der optisch-aktiven Brombernsteinsäure und Asparaginsäure," E. Fischer and K. Raske, *Ber.*, **40**, 1951 (1907).

"Isomere Leucyl-leucine und deren Anhydride," E. Fischer and A. H. Koelker, *Ann.*, **354**, 39 (1907).

"Derivate des *l*-Phenylalanin," E. Fischer and W. Schoeller, *Ann.*, **357**, 1 (1907).

"The Synthesis of Histidine," F. L. Pyman, *J. Chem. Soc.*, **99**, 1386 (1911).

"On Hydantoins. A Synthesis of Phenylalanine and of Tyrosine," H. L. Wheeler and C. Hoffman, *Am. Chem. J.*, **45**, 368 (1911).

"The Reaction of Chloroacetic Acid with Ammonia and the Preparation of Glycine," G. R. Robertson, *J. Am. Chem. Soc.*, **49**, 2889 (1927).

"The Preparation of Lysine, Histidine, and Arginine from Hydrolyzed Blood Corpuscle Paste by Electrical Transport," G. J. Cox, H. King, and C. P. Berg, *J. Biol. Chem.*, **81**, 755 (1929).

"Experiments on the Synthetic Preparation and Isolation of Some of the Simpler Amino-Acids," W. Cocker and A. Lapworth, *J. Chem. Soc.*, **1931**, 1391.

"A Convenient Synthesis of *dl*-Lysine," J. C. Eck and C. S. Marvel, *J. Biol. Chem.*, **106**, 387 (1934).

"Synthesis of α -Amino- β -Hydroxy-*n*-Butyric Acids. III. A Simple Method of Preparing a Mixture of the Two Forms," H. D. West and H. E. Carter, *J. Biol. Chem.*, **119**, 103 (1937).

"Synthesis of α -Amino- β -Hydroxy-*n*-Butyric Acids. IV. Separation of Mixtures of the Two Forms and Preparation of *d*(-)- and *l*(+)-Threonine," H. D. West and H. E. Carter, *J. Biol. Chem.*, **119**, 109 (1937).

"A Convenient Synthesis of *dl*-Glutamic Acid," C. S. Marvel and M. P. Stoddard, *J. Org. Chem.*, **3**, 198 (1938).

"The Synthesis of Dipeptides from α -Keto Acids," D. Shemin and R. M. Herbst, *J. Am. Chem. Soc.*, **60**, 1951 (1938).

"Aromatic Sulfonic Acids As Reagents for Amino Acids," D. G. Doherty, W. H. Stein, and M. Bergmann, *J. Biol. Chem.*, **135**, 487 (1940).

"Preparation of Histidine by Means of 3,4-Dichlorobenzenesulfonic Acid," H. B. Vickery, *J. Biol. Chem.*, **143**, 77 (1942).

"Amino Acids. I. Glycine," J. H. Billman and E. A. Parker, *J. Am. Chem. Soc.*, **65**, 761 (1943).

"*l*-Tryptophane," G. J. Cox and H. King, *Organic Syntheses*, Collective Vol. 2 (A. H. Blatt, ed.), p. 612, John Wiley & Sons, New York, 1943.

"Convenient Synthesis of *dl*-Tryptophan," H. R. Snyder and C. W. Smith, *J. Am. Chem. Soc.*, **66**, 350 (1944).

"The Kinetics of Low Temperature Acid Hydrolysis of Gramicidin and of Some Related Dipeptides," R. L. M. Synge, *Biochem. J.*, **39**, 351 (1945).

"Crystalline Soybean Trypsin Inhibitor (estimation of extent of protein hydrolysis by optical density measurements)," *J. Gen. Physiol.*, **30**, 307 (1947).

"Ethyl Formylaminomalonate: An Intermediate in the Synthesis of Amino Acids," A. Galat, *J. Am. Chem. Soc.*, **69**, 965 (1947).

- "The Biological Synthesis of Radioactive Adrenalin from Phenylalanine," S. Gurin and A. M. Delluva, *J. Biol. Chem.*, **170**, 545 (1947).
- "Séparations chromatographiques d'acides, amines et de peptides," C. Fromageot, M. Jutisz, and E. Lederer, *Biochim. et Biophys. Acta*, **2**, 487 (1948).
- "A Resolution of Histidine, Cystine, and Alanine by Asymmetric Enzymatic Hydrolysis of the Racemic Amides," L. Levintow, V. E. Price, and J. P. Greenstein, *J. Biol. Chem.*, **184**, 55 (1950).
- "The Resolution of Amino Acids by Asymmetric Enzymatic Syntheses," D. G. Doherty and E. A. Popenoe, Jr., *J. Biol. Chem.*, **189**, 447 (1951).
- "Resolution into Optical Isomers of Some Amino Acids by Paper Chromatography," M. Kotake, T. Sakan, N. Nakamura, and S. Senoh, *J. Am. Chem. Soc.*, **73**, 2974 (1951).
- "The Resolution of Amino Acids. I. Phenylalanine and Valine," L. R. Overby and A. W. Ingersoll, *J. Am. Chem. Soc.*, **73**, 3363 (1951).
- "Resolution of Glutamic Acid with 1-Hydroxy-2-Aminobutane," F. H. Radke, R. B. Fearing, and S. W. Fox, *J. Am. Chem. Soc.*, **76**, 2801 (1954).
- "Pyrosynthesis of Aspartic Acid and Alanine from Citric Acid Cycle Intermediates," S. W. Fox, J. E. Johnson, and M. Middlebrook, *J. Am. Chem. Soc.*, **77**, 1048 (1955).
- "Production of Some Organic Compounds under Possible Primitive Earth Conditions," S. L. Miller, *J. Am. Chem. Soc.*, **77**, 2351 (1955).
- "Hydrolysis of Proteins and Dipeptides by Ion-Exchange Resin Catalysis," J. R. Whitaker and F. E. Deatheraze, *J. Am. Chem. Soc.*, **77**, 3360 (1955).
- "On the Sublimation of Amino Acids and Peptides," D. Gross and G. Grodsky, *J. Am. Chem. Soc.*, **77**, 1678 (1955).
- "Preparation of 5-Hydroxy-L- and D-Tryptophan," A. J. Morris and M. D. Armstrong, *J. Org. Chem.*, **22**, 306 (1957).

6

Assay of the Amino Acids

Effects of hydrolysis

Types of procedure

Types of estimation

Criteria of reliability

Protein composition

Until the fifth decade of this century nearly all of the information about the amino acid composition of proteins was obtained through methods involving distillation, under reduced pressure, of the esters of the amino acids; through colorimetry; or by a few methods of miscellaneous types. Gravimetric procedures which were used were based on weighing those amino acids which could be induced to precipitate from hydrolyzates in relatively quantitative recoveries. In the early 1940's a number of new types of amino acid assay appeared. Pre-eminent among these was the microbiological type of estimation, but the seeds of other procedures were sown about the same time. This new group of methods filled a considerable part of the demand for sensitive and convenient ways of carrying out the desired analyses.

Many methods of assay for any one given amino acid were now available. This situation did not lead, as might have been expected, to a competition between methods, yielding finally to one supreme method. It has been found instead that the suitability of a type of procedure depends upon many factors, such as whether the analyst is interested in carrying out a comprehensive assay of many or all the amino acids present, or whether he seeks to evaluate only one or a

few. Microbial assay, for example, proves to be highly efficient when many samples are to be assayed for one or a few amino acids simultaneously, and when the analyst proposes to evaluate his determinations statistically. Assay by separation on columns offers its greatest economy of time when the objective is complete analysis of a single protein. It has also proved useful to check results obtained in one of these two ways with those obtained by the other method.

The analysis of proteins depends upon the assay of all of the individual amino acids. The protein may thus be said to be analyzed, the amino acids to be assayed.

The development of assay methods and their dependable application to biological materials require chemical ability of the highest order, coupled with well-balanced judgment as to just how short of perfection one should stop. Absolute dependability in assay is seldom attained. In the usual case, assurance is lacking that some as yet unrecognized constituent is neither contributing to the analytical response nor masking it. Particularly is this true when assays which have been developed for pure substances, or with biological material, are applied to another biological material. Definite procedures are available, however, for assessing the validity of assays.

As the first step in a working procedure, it is well to review in *Chemical Abstracts* the recent literature of the assay of the substance under consideration. Under the listing of the substance, the appropriate subheadings to be searched are *analysis (of)*, *assay (of)*, *detection*, *estimation*, and sometimes other subheadings. For the application of specific methods, the original literature should be consulted, as well as some of the more specialized reviews and books in the field (Block and Bolling, 1951; Olcott, 1951). In such work as this, hours in the library can save days and weeks in the laboratory.

The following discussion of the various types of amino assay stresses principles rather than methods for specific amino acids.

EFFECTS OF HYDROLYSIS

The majority of methods in the field of protein analysis are applied to hydrolyzates rather than to whole proteins. Decomposition of a few amino acids occurs to a marked degree under acid hydrolytic conditions, for example, tryptophan or cystine, and there is likelihood of some decomposition of many amino acids when proteinaceous material is hydrolyzed in the presence of carbohydrates and other impurities. Other amino acids for which there is evidence of decomposition during hydrolysis of the protein include: serine, threonine, tyrosine, and phenylalanine. The rate of destruction is probably

influenced not only by nonprotein constituents but also by the nature of the protein molecule in which the amino acid occurs.

Both gross destruction and racemization may take place. There is considerable evidence that a small but appreciable degree of racemization occurs for some amino acids during the usual period and conditions of hydrolysis with mineral acid. This may not be apparent when an amino acid is isolated, because a small amount of antipode or racemate formed will tend to remain in solution. For some reagents, such as acetic anhydride and alkali, the interior amino acid residues are more susceptible to racemization than the terminal residues or the free amino acids. Alkaline hydrolysis is sometimes used for preparing amino acid solutions for assay, particularly of tryptophan. Typical hydrolytic conditions would involve 6 hours of refluxing with 6 N sodium or barium hydroxide solution. Under these conditions, racemization is considered to be close to complete.

In view of these and other difficulties, methods for determination of amino acid residues in protein prior to hydrolysis are desired and are in fact slowly being developed. These are not generally at a very advanced stage. The aromatic amino acids may, however, be determined by ultraviolet spectrophotometry. This technique has been most successful for tyrosine and tryptophan. Successful application of some principle to determine amino acid residues in intact protein would serve to eliminate much of the doubt that attends those assays which now require hydrolytic preparation. A procedure for correcting for hydrolytic decomposition involves assaying after lengthening periods of hydrolysis and employing the resultant data to correct the value obtained, by extrapolating to zero time. Isoleucine and valine are generally released slowly in hydrolysis. The most careful analyses have employed prolonged hydrolytic periods. Redistillation of hydrochloric acid for hydrolysis diminishes decomposition in some cases.

TYPES OF PROCEDURE

Assay procedures for amino acids are sometimes classified as standard or routine, or as comprehensive or individual. These are distinctions of degree and practice rather than distinctions of kind. For example, the isotope-dilution method of assay is considered to be always a standard method, one which may be used for a reference, and never routine. It does not follow, however, that the microbiological assay type of method, which is often routinely used, may never be taken as standard. Some methods of this kind have become increasingly standard as they have been more widely used and as their

dependability has been evaluated. Assay methods developed for individual amino acids may sometimes be inapplicable to other amino acids, but in other cases such methods may be applied, with little or no modification, to a number of amino acids. The method then becomes a comprehensive one. Chemical methods, such as colorimetric types, tend to be individual, that is, of limited applicability. Methods involving the use of organisms tend to be comprehensive, since these organisms frequently have multiple amino acid requirements.

TYPES OF ESTIMATION

Gravimetric Procedures. Isolation and weighing have been used for cystine and for tyrosine, but colorimetric methods for these amino acids are more dependable. The isolation of amino acids can be brought about in a more quantitative fashion by formation of derivatives, salts, and complexes of specific types.

Simple salts, such as the calcium or barium salts of the dicarboxylic amino acids, can be separated by precipitation from an aqueous solution with alcohol. This method is known as Foreman's dicarboxylic acid method.

The Fischer ester-distillation method for analysis of proteins has been previously described. Since this method requires the weighing of the separated acids after saponification of the esters, it is in essence a gravimetric method. Historically, the Fischer distillation analysis was a favored method, and it was for a long time of prime value. Recoveries were low, however, in common with gravimetric procedures in general. A recovery of two thirds of the amino acids in a protein was considered to be evidence of skillful execution of the procedure.

Selective precipitation of insoluble salts or complexes of amino acids has often been used in the past for analytical purposes. The precipitation of arginine as the flavianate is one of the better known procedures. Bergmann and co-workers developed a large number of inorganic complex salts and organic acids which, under proper conditions, could be used to precipitate selectively various amino acids and thus lead to a series of quantitative determinations. These procedures, as analytical devices, have been largely superseded by some of the others described later in this chapter.

Isotope Dilution. In the isotope-dilution method, one adds to the compound present in the biological material a small but measured amount of the same substance into which has been incorporated, by synthesis, a radioactive or stable isotope. When this compound is an amino acid or a simple derivative, the element incorporated into the tracer substance is frequently N¹⁵ or C¹⁴. The "heavy" nitrogen can

be determined in the final isolated product by use of a mass spectrograph, and the carbon isotope can be assayed with radioactivity counters. On the basis of the warranted assumption that both the normal and the tracer compounds have virtually the same chemical properties, one can determine the amount of isotope recovered and calculate therefrom the per cent of recovery of the substance in the sample. This may be done with aid of the formula

$$\frac{x}{x+y} = \frac{c}{c_0} \quad \text{or} \quad y = x \left(\frac{c_0}{c} - 1 \right)$$

in which x is the amount of compound added, c_0 is the isotope content (per cent) of compound added, y is the amount of compound originally present, and c is the isotope content (per cent) of isolated diluted compound.

Inasmuch as in most such experiments the amount of compound added is negligible in contrast to the amount of compound present, the left-hand expression reduces to x/y , and the corresponding equation is $x/y = c/c_0$. In other words, the per cent of isotope recovered is equivalent to the per cent of compound recovered. For this procedure complete recovery is unnecessary, but purification of the isolated compound is a prerequisite to successful use of the method. Although errors can be reduced to ± 1 per cent by the isotope-dilution method, the procedure has not seen much use because of the labor involved. As a standard procedure, isotope dilution is unexcelled.

Selective Adsorption. As the name implies, selective adsorption refers to the separation of the components of a mixture through preferential adsorption on a suitable adsorbent. In its simplest form this technique consists in stirring the adsorbent in a solution of the mixture to be separated, and then filtering or centrifuging.

A much more refined technique based on selective adsorption is the method known commonly as *chromatography*. Here a solution of the mixture to be separated is allowed to percolate slowly through a column of adsorbent packed in a long vertical tube. The migration of various component solutes is retarded to an extent dependent on their tendency toward adsorption. The more strongly adsorbed a given component, under the conditions used, the higher will be the corresponding adsorption band in the column. After such a separation, the column may be removed and sectioned, and the various components removed from the appropriate sections by *elution* (extraction of adsorbed substances). This technique is particularly useful in those cases in which the components being separated are colored so that they may be readily located in the column. Indeed, the technique

was originally developed on such compounds, hence the name chromatography. The method is applied to colorless substances by using staining techniques or fluorescence. The term chromatography is therefore inaccurate for many of the applications of adsorption on columns.

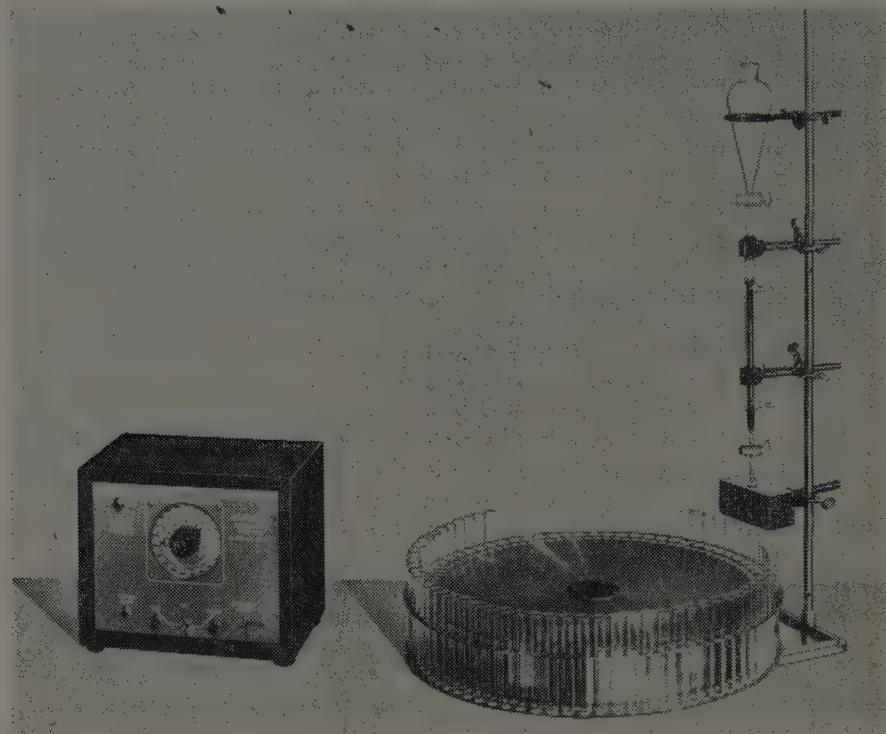


Fig. 6-1. An automatic fraction collector. Courtesy of Packard Instrument Co.

Alternatively, the components can be separated by direct elution from the intact column. This may be accomplished by passing large volumes of eluting solvent through the column and collecting and analyzing successive "cuts." Analysis of the "cuts" can be made colorimetrically after treatment with a suitable reagent—for example, ninhydrin—or by measurement of some other physical property, such as refractive index or optical rotation. The components may be identified on the basis of the volume of solvent required for their elution under definite conditions, or by specific chemical tests in some cases. If, for example, a plot is made of the refractive index of eluate versus volume passed, a series of peaks is obtained, each representing a separate component. The position on the abscissa (volume) characterizes the component, while the area under each peak is proportional to the amount of that component present. The labor of this method is reduced by employing an automatic fraction collector (Fig. 6-1).

The modification of this technique which has been most popular and useful is the Stein-Moore assay. Developed originally by Stein and Moore (1948) for starch columns, it has been used principally with Dowex resins. These resins are made from sulfonated poly-

styrene; their properties have been shown to vary considerably with the proportion of cross-linking in the resin. Elution is conducted most efficiently with occasional changes of buffer from one pH to another. In a more recent technique the buffer is continuously changed by feeding the solution through a train of two vessels, with mixing in the second vessel (the one next to the column). Temperature is also varied during development of the column, in order to heighten the differences in rate of elution (Fig. 6-2).

Among the difficulties likely to be encountered with a protein hydrolyzate is incomplete separation of the individual components. In chromatography in general, it is possible to vary the order of elution of seventeen amino acids by varying the composition of the column and of the solvent. In any standard system, the order of elution will typically be the same from one analysis to another, but the rate of removal of each band may show some variation. This variation can of course be minimized by standardization of technique.

In another closely related technique, instead of eluting with solvent the solution itself is allowed to flow continuously through the adsorbent bed until all components are present in the outflow. Components are now not resolved into separate fractions; instead, there is a series of solution fronts, a new solute component appearing at each front but existing then in all solution following this front. It is convenient to measure the refractive index of the outflowing solution as a measure of total solute present. Instruments have been developed for doing this continuously and automatically, a tracing in the form of a series of steps of refractive index versus volume being obtained directly. Just as in the elution method, the volume passed at the time of appearance of a component characterizes that component, but the amount of component is proportional to the height of the step it produces. For obvious reasons, this technique is termed *frontal analysis*, and it has been applied to amino acid mixtures with considerable success. Its use has not been widespread, probably because of the relatively elaborate equipment required.

Another technique, very closely related to adsorption chromatography, is treated here for convenience even though it does not depend primarily on adsorption. In this case, separation is accomplished by partition of the various solutes between two phases, one normally an aqueous phase and the other a hydrophobic phase, one of the phases being immobilized and the other flowing. Usually the aqueous phase is provided by the water of hydration on the substrate comprising the column. Hydrated starch has been one of the most useful substances for this purpose. It is difficult to say to what extent

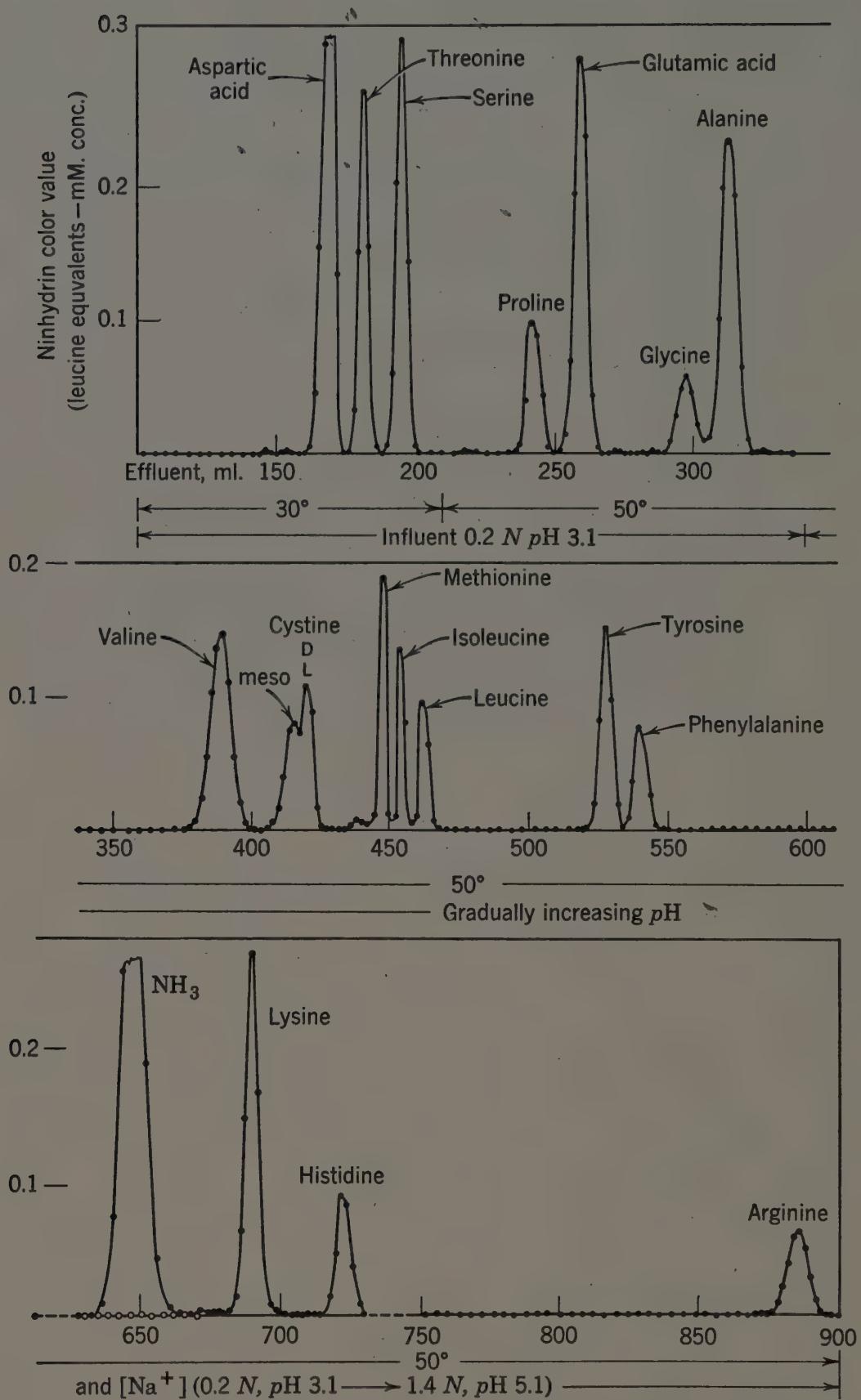


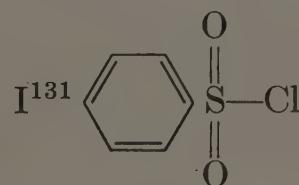
Fig. 6-2. Elution pattern of amino acids from ribonuclease hydrolyzate. From C. H. W. Hirs, W. H. Stein, and S. Moore, *J. Biol. Chem.*, **211**, 941 (1954).

the separation results from simple partition and to what extent adsorption plays a role. This technique is sometimes referred to as *partition chromatography* or *partography*.

In an even more advanced technique, instead of a packed column a sheet of moist paper is used as the separating medium. A drop of the solution to be separated is placed at either the top or the bottom of a strip of moistened paper contained in a tight chamber that is saturated with respect to the vapor of the solvent used. The liquid then moves either downward or upward by capillarity, the dissolved constituents resolving as zones in accordance with their relative mobilities.* Usually it is necessary to stain with some reagent to give visible spots following the development of the chromatogram. In some cases two-dimensional paper chromatography is used, a first separation being carried out with one solvent system as previously outlined, the paper then being turned through 90 degrees and a second separation being made, usually with another solvent system.

Paper chromatography has been of more use qualitatively than quantitatively, but it is receiving increasing application in the latter respect. For quantitative analysis it is necessary to determine in some manner the relative amount of material in each spot. Numerous modifications of the essential principle have appeared in the literature. Most of the methods employ determination of color intensity resulting from reaction with ninhydrin (p. 52). The resultant colored spots may be estimated directly by a planimeter or by scanning with a densitometer. All of the amino acids may be separated on one-dimensional strips, providing a few strips are used with different sets of solvents. In other methods the amino acids are eluted and determined by ninhydrin reaction, copper-complex formation, titration in glacial acetic acid, or in other ways. A principal uncertainty which applies to the use of such methods is the question of whether or not a discrete spot represents only one amino acid.

Although the paper-strip method can be applied to small amounts of material such as a few tenths of a milligram of protein hydrolyzate, it can be made much more sensitive when radioactive derivatives are employed. For this purpose the reagent pipsyl chloride



* These mobilities are usually designated as R_f values, defined as the ratio of the rate of movement of zone to the rate of movement of developing liquid.

has been introduced. The reagent is named from initialization of paraiodophenylsulfonyl chloride. A protein hydrolyzate may be exhaustively pipsylated with pipsyl chloride through use of the Schotten-Baumann reaction, and then separated on a paper chromatogram. By cutting standard transverse pieces of the paper strip and measuring the radioactivity of these transverse strips, one can observe a peak of radioactivity wherever a pipsyl amino acid has banded out. The activity count, furthermore, can be used to calculate the proportion of the given amino acid in the band. This method makes it possible to operate below the millimicrogram range for a given amino acid. Strictly speaking, the method involves "isotope swamping" rather than isotope dilution.

Colorimetry. A large volume of analytical literature in the protein area has been built up in the past from colorimetric assays. The basic principle is the formation of a colored compound, the intensity of the color being related in some regular way to the concentration of the substance under assay. Some of these colors are stable. Others must be measured immediately or evaluated until a maximum reading is obtained. The structure of the compound responsible for the color is seldom known.

In the simple colorimeter, white or, more commonly, filtered light is shone through the solution being assayed and through a standard solution with which it is matched. The photoelectric colorimeter or *photometer* substitutes a photoelectric device for the human eye, thus eliminating subjective factors. Light filters are employed. The *spectrophotometer* is similar in principle but permits use of narrower ranges of wave length of light, which results in more specific freedom from interfering compounds. The general equation which applies to the use of the spectrophotometer is known as Beer's law:

$$\log \frac{I_0}{I} = klc$$

in which I_0 is incident light; I is transmitted light; k is a constant, the extinction coefficient; l is the thickness of the light-absorbing solution; and c is the concentration of the absorbing substance.

Colorimetric assays which have been much used include those for tryptophan, tyrosine, and cystine. Tryptophan may be estimated by any of various aldehydes that react to form colored derivatives. Glyoxylic acid, $\text{OHC}\cdot\text{COOH}$, and *p*-dimethylaminobenzaldehyde



Ehrlich's reagent, have especially been used. Tyrosine may be determined with Millon's reagent after separation from tryptophan. Cystine is lost rapidly on alkaline hydrolysis, and gradually on acidic hydrolysis. The effect of the latter may be mitigated by employing a 1:1 mixture of hydrochloric acid (20 per cent) and formic acid (90 per cent) as hydrolytic agent. Following removal of acids, the cystine may be determined by the color produced on reaction with phospho-18-tungstic acid of the cystine reduced to cysteine.

Some colorimetric assays yield results frequently higher than the true values, because other compounds often enhance the color.

Microbiological Assay. The original development of knowledge of the substances required for growth necessitated the use of animals. This seldom yields a convenient method of assay, but the use of microorganisms frequently does, and at the present time this type of analysis offers a most generally satisfactory combination of precision and convenience. Most microbiological assay procedures permit the determination of many amino acids with the same analytical organism.

The principles of microbiological assay may best be introduced by explaining the use of mold mutants, which are also of value in elucidating the pathway of synthesis of amino acids. The common bread mold *Neurospora crassa* requires for its nutrition water, carbohydrate as a source of energy, inorganic elements, biotin, and inorganic nitrogen. With these nutrients in its medium the mold is able to synthesize the various organic nitrogen compounds (amino acids, and vitamins except biotin) which it requires. When irradiated with ultra-violet light or X-rays, the organism may lose its power to synthesize one of these organic nitrogen compounds. Such mutants have been catalogued in detail by Beadle and co-workers, and mutants for nearly all of the amino acids have been found.

A mutant which is unable to synthesize tryptophan, for example, will not grow in the simple basal medium unless tryptophan or compounds convertible to tryptophan are supplied. When tryptophan is supplied in graded amounts, the mutant will grow to an extent that can be correlated with the amount of tryptophan added (as in Fig. 6-3). If now a material with an unknown amount of tryptophan is added, the quantity may be determined by comparing the growth with the equivalent growth on a standard curve constructed from known amounts of tryptophan. Such a mutant is of value not only potentially for bioassay, but also to test the ability of the mutant to grow on various postulated intermediates in the natural synthesis. Inasmuch as a unique mutant may be found for each step in a

synthesis, one may learn how tryptophan and other nitrogenous compounds are synthesized in nature (Chap. 7).

For microbiological assay various normal bacteria have received far more use than have the mutant microorganisms. These bacteria, predominantly dairy bacilli such as *Lactobacillus arabinosus*, *L. casei*, *Streptococcus faecalis*, and *Leuconostoc mesenteroides*, have nutritional requirements more like those of man than like those of

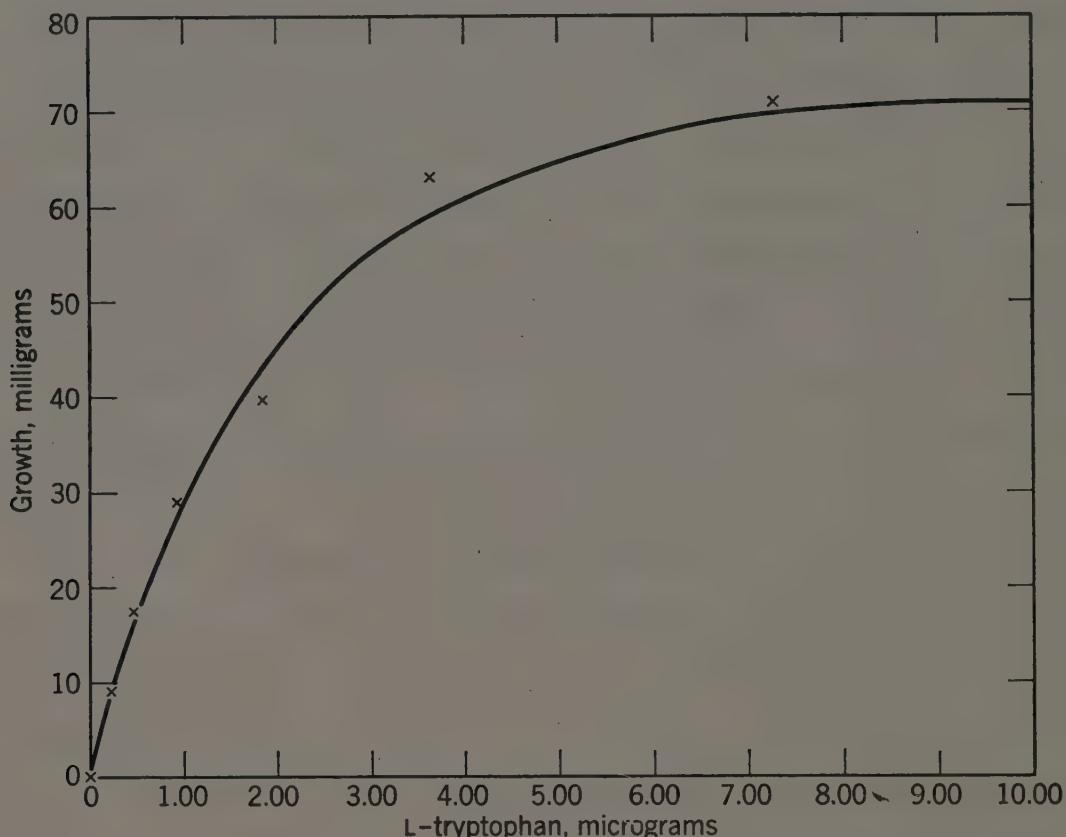


Fig. 6-3. Growth of tryptophanless *Neurospora* mutant as a function of tryptophan.

Neurospora (Table 6-1). In such cases all of the nutrient requirements, with the exception of the amino acid under study, are included in the medium. The standard curve is then constructed in the same way as for *N. crassa* mutants (Fig. 6-3), and the determinations are carried out by measuring the growth of the bacteria in the presence of the material to be assayed, after it has been hydrolyzed. This can be done turbidimetrically; or in the case of the lactic acid bacteria, the acid produced may be titrated, inasmuch as its amount is closely related to cellular proliferation.

CRITERIA OF RELIABILITY

In inorganic analysis it is possible to bracket the element under investigation with the aid of the periodic table. It is then feasible to

TABLE 6-1. Amino Acid Requirements of Some Common
Microbiological Assay Organisms¹

(E = essential, D = dispensable, A = accessory)

Amino Acid	Assay Organisms			
	<i>Lactobacillus arabinosus</i> 17-5	<i>Streptococcus faecalis</i>	<i>Lactobacillus brevis</i>	<i>Leuconostoc mesenteroides</i>
Alanine	D,A	E,D,A	D	D,A
Arginine	A	E	E	E
Aspartic acid	A	E	E	E
Cystine	E	D	E	E
Glutamic acid	E	E	E	E
Glycine	D	E	E	E
Histidine	D,A	E	E	E
Isoleucine	E	E	E	E
Leucine	E	E	E	E
Lysine	E,A	E	E	E
Methionine	E,A	E,A	E,A	E
Phenylalanine	E,A	E,A	E	E
Proline	D,A	D	E	E,A
Serine	D,A	E	D	A
Threonine	A,E	E	A	A
Tryptophan	E	E	E	E
Tyrosine	E,A	E	E	E
Valine	E	E,A	E,A	E

¹ From data by C. Warner. It may particularly be noticed that different strains of the same organism are found, in different laboratories, to have different requirements.

test other related elements for their effect in contributing to or masking analytical responses of whatever kind they may be. The comparable procedure of checking all organic compounds in an organic analysis, especially when there are substances of unknown constitution in biological material, is virtually impossible. Amino acid assay is a type of organic analysis, and such a limitation is particularly applicable in this field. This section describes a number of criteria for reducing to a minimum the uncertainty that attends methods for assay of organic substances found in biological materials.

1. Determination of Specificity of Structure of the Compound Assayed.

Do other compounds behave similarly in colorimetric or microbiological procedures? Do other compounds mask the reaction? In order to decide where to seek and "shoot" trouble of this type, it is necessary to understand the reactions of various chemical groups, the pathways of metabolism, and what information is available on *anti-metabolites* (substances which interfere specifically in the synthesis or utilization of metabolites; see also Chap. 21). As more compounds

are tested and found to be without influence on the assay under consideration, more dependence can be placed upon the method.

This type of investigation is occasionally carried out prior to the announcement of a new method. Compounds usually tested first are those of structure similar to the one being assayed. Other leads may be found if the pathways of metabolism of the pertinent compound are known. For instance, some microorganisms known to require tryptophan are also known to be able to synthesize tryptophan from indole. In an assay for tryptophan it is therefore important that all indole which may be present is first removed. At times compounds of unrelated structure may cause difficulty. If cysteine, for instance, is to be assayed in terms of its reducing power, other reducing agents such as glucose or ascorbic acid may need to be eliminated or accounted for.

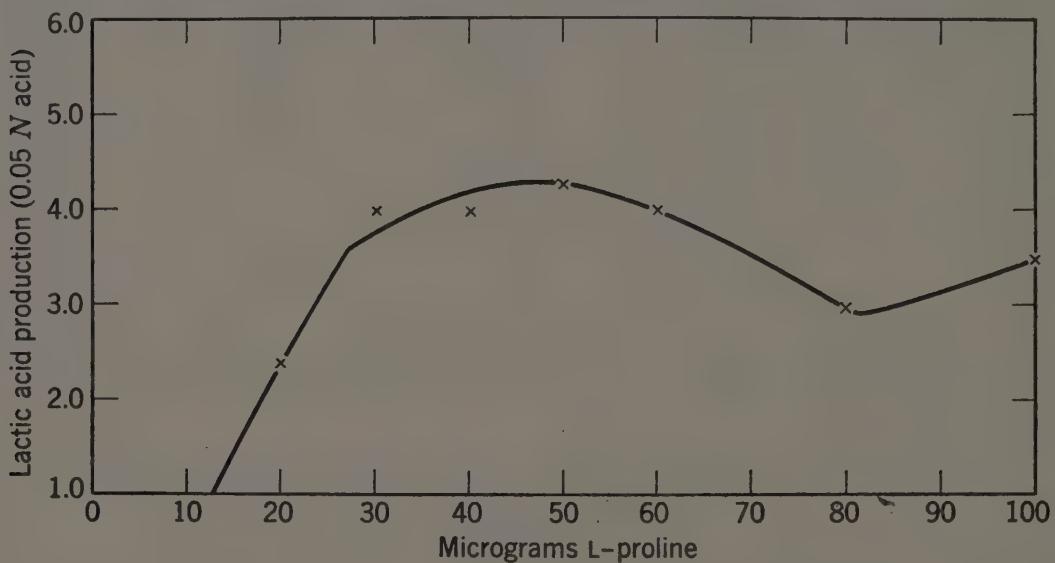


Fig. 6-4. Unusable growth response to L-proline with *Lactobacillus brevis*. In this case, alteration of the medium led to a usable standard curve.

2. Determination of the Relationship between the Concentration of Material and the Assay Response. Is the intensity of color in the colorimetric method a function of the concentration in dilute solutions? Do the solutions obey the laws (Lambert, Bouguer, Beer) relating color intensity and concentration? If they do not, can a regular relationship be found which permits the plotting of a standard curve as a reference in each determination? The last question applies also to microbiological assay. Many organisms will show a regular growth response to added amounts of an amino acid being determined, as in Fig. 6-3. Others, however, may give irregular responses, as in Fig. 6-4. If the answers to the above questions are negative, the method

is without value for the material for which it is being studied. If the relationship is regular, the concentrations as calculated from readings at different test levels should agree. In experiments with pure materials no assay should be applied to any material until it is first shown to meet the criteria set up in this paragraph.

3. Recovery Experiments. This criterion is a very practical and critical one. Even though the assay response may be regular with pure materials, the particular biological material being studied may contain within it some substance leading to irregularities by reaction with the reagents or with the substance under investigation. In order to carry out recovery experiments a known amount of the substance assayed is added to the biological material, and (at various stages of the assay procedure in different determinations) the amount added is determined by assay and this figure is compared with the known value. Ideally, recovery is 100 per cent plus or minus the probable error of the method (typically under ± 5 per cent for microbiological or Stein-Moore assay); when the recovery is not 100 per cent, interfering or contributing substances may be suspected. Here again, satisfactory recovery alone does not conclusively establish the validity of the assay, but it does represent a criterion that should be met before the results of an assay are taken seriously.

4. Precision and Accuracy. *Accuracy* measures the trueness of a value; *precision* is a measure of the degree of reproducibility in replicates. If the results are quite precise in separate runs, one may feel more certain of his method; even so, the method may be highly inaccurate.

Aside from isotope dilution or other methods employing actual isolation of the studied compound, accuracy, strictly speaking, cannot be quantitatively evaluated. The closest known approach to accuracy of an amino acid assay results from the employment, in a control, of a mixture of amino acids which simulates the composition of the protein or proteinaceous foodstuff being analyzed. This evaluation of accuracy is imperfect for a number of reasons, one of which is that unknown amino acids may be present in the protein, and another is that losses due to hydrolytic treatment may occur.

5. Statistical Evaluation. An important factor in the employment of assays is the skill of the technician. In a set of samples, statistical measures such as mean deviation from the mean, and standard deviation, permit evaluation of the improvement in the worker's ability. If other workers have reported statistical evaluation of their determinations, worthwhile comparisons can be made. One can especially

determine, in a relatively objective fashion, whether the particular method is as suited to one biological material as it is to others.

Criteria 2, 3 and 4 are special types of statistical evaluation; the wise analyst, however, will pursue his own further statistical treatment or will seek the aid of a statistician. Developed judgment is of considerable importance in applying an assay project and deciding what to accept as determinations, but part of this judgment should be the decision to use statistical help wherever it is applicable.

6. Comparison of Two Chemical Methods or of Two Microbiological Methods. If two microbiological assay procedures employing different microorganisms, or two colorimetric methods employing different reagents, agree in the values they provide, belief in the reliability of each method is greatly enhanced. If the two methods employed are microbiological, it is desirable to carry them out with organisms as different as possible, a desideratum which is analogous to employing colorimetry with different reagents. The use, for instance, of a dairy bacillus and of *Escherichia coli* would fulfill such a requirement.

7. Comparison of a Chemical Method with a Microbiological Method. Agreement between a chemical method and a microbiological method is even more satisfying than fulfillment of requirement 6. There is, however, little justification for generalizing that the biological methods must confirm chemical methods, or vice versa. For some amino acids, chemical methods are widely preferred (tyrosine), for others the biological (valine, leucine, isoleucine). The degree of disagreement is typified in Table 6-2.

Perhaps the most effective combination of the ideal and the practical is a comparison of Stein-Moore and microbial analyses. The two types of method can be most effectively and simply used in concert by employing each of those amino acids for which experience has shown the technique to be most reliable. Histidine, lysine, and arginine may be determined more reliably by appropriate microbial assays than by Stein-Moore analysis. On the other hand, microbial assay is least dependable for those amino acids for which biosynthetic pathways are not easily lost in evolution, such as glycine, alanine, and aspartic acid. On these, Stein-Moore assay is often to be preferred.

Other considerations besides accuracy are often critical. In large-scale biological experiments the demand for rapidity is often so great that the use of a less dependable method may be justified if it provides satisfactory comparative figures. In such cases an estimate of the accuracy is especially to be sought. Testing with all of the above criteria requires a great deal of chemical knowledge and judgment, as well as much work, but for results of significance it is not too much

to ask that all of these criteria be applied when possible. Criteria 2 through 5 should be applied in the institution of any method, with 6 or 7 preferably added.

For microbiological assay, some of the questions raised in this section can be answered by proper use of a simple log-log plotting. The treatment of such plottings is described by Barton-Wright (1952).

TABLE 6-2. Selected Assays Which Permit Comparison of Results by Different Methods¹

Amino Acid	Protein	Method	Amino Acid Content, grams per 16 gm. N ²
Histidine	Horse hemoglobin	Chromatographic	8.2
		Electrolytic	8.5
		Enzymatic	7.4
		Colorimetric	8.0
Leucine	Ovalbumin	Microbiological A	8.8
		Microbiological B	9.4
		Gravimetric	9.4
		Microbiological C	8.1
		Microbiological D	9.8
Phenylalanine	Silk fibroin	Gravimetric	1.3
		Microbiological A	1.3
		Microbiological B	1.2
		Microbiological C	1.1
		Microbiological D	1.3
Tryptophan	Vitellin	Colorimetric A	2.0
		Colorimetric B	1.3
		Colorimetric C	2.6
		Colorimetric C	1.9

¹ From R. J. Block and D. Bolling, *The Amino Acid Composition of Proteins and Foods*, Charles C Thomas, Springfield, Illinois, 1951.

² The discrepancies noticeable may be due to differences in homogeneity of preparations of protein as well as to differences in methodology. Some amino acids are more reliably determined than others, tryptophan particularly giving various values. As emphasized by Block, if each analyst would apply at least two methods to his determinations, the results would be revealing.

PROTEIN COMPOSITION

The amino acid contents of a number of proteins are given in Table 6-3.

A few features are especially worthy of note. This table, compiled from several surveys, includes values obtained by a variety of methods. As shown, most proteins contain virtually all of the amino acids. It may be noted that even though the botulinus toxin is particularly well supplied with the nutritionally essential amino acids, it is a violent

TABLE 6-3. Amino Acid Composition of Some Proteins¹

Sal-mine	Insulin	Edestin	β -Lacto-globulin	Ovalbumin	Horse Hemoglobin	Botulinus Toxin	Silk Fibroin	Gelatin	Pepsin	Casein
Alanine	1.1	4.5	4.3	6.4	6.7	7.4	3.9	29.7	9.3	3.2
Arginine	85.2	3.1	1.76	2.9	5.7	3.7	4.6	1.1	8.6	3.8
Aspartic acid	..	6.8	12.0	11.4	9.3	10.6	20.1	2.8	5.6	7.2
Cysteine	0.50	1.1	1.4	0.56	0.53	..	0.50	..
Cystine	12.5	0.93	2.3	0.51	0.45	0.27
Glutamic acid	18.6	20.7	21.5	16.5	8.5	15.6	2.2	11.2
Glycine	2.9	..	4.3	..	1.5	3.1	5.6	1.4	43.6	26.9
Histidine	4.9	2.9	1.6	2.4	8.7	1.0	0.36	0.73
Isoleucine	2.8	7.5	5.9	7.0	..	11.9	1.1	1.3
Leucine	..	1.6	13.2	4.7	15.5	9.2	15.4	10.3	0.91	10.8
Lysine	2.5	2.4	11.3	6.3	8.5	7.7	0.68	10.4
Methionine	2.4	3.2	5.2	1.0	1.1	..	1.3
Phenylalanine	5.5	4.0	7.7	7.7	1.2	3.4	1.7
Proline	..	5.8	2.5	4.3	5.4	3.6	3.9	2.6	0.74	6.4
Serine	..	9.1	5.2	6.3	4.1	8.2	5.8	4.4	16.2	0.9
Threonine	2.1	3.9	5.2	4.0	4.4	8.5	1.6	1.3
Tryptophan	1.5	1.9	1.2	1.7	1.9	1.2
Tyrosine	13.0	4.3	3.7	3.7	3.0	13.5	12.8	1.0
Valine	..	3.1	7.8	5.7	5.7	7.1	9.1	5.3	3.6	7.1
										6.9

¹ Units: grams amino acid per 100 gm. protein.

poison. This underlines the fact that amino acid residue arrangement or other structural features, rather than composition, account for many biological properties.

A mathematical treatment of the relationships in quantitative compositional values of protein is considered in Chap. 24.

SELECTED REFERENCES

Books and Reviews

"Analytical Chemistry of the Proteins," A. J. P. Martin and R. L. M. Synge, *Advances in Protein Chem.*, **2**, 1 (1945).

"The Microbiological Assay of Amino Acids," E. E. Snell, *Advances in Protein Chem.*, **2**, 85 (1945).

"The Principles of Chromatography," A. J. P. Martin, *Endeavour*, **6**, 21 (1947).

"Adsorption Analysis of Amino Acid Mixtures," A. Tiselius, *Advances in Protein Chem.*, **3**, 67 (1947).

"Amino Acid Composition of Purified Proteins," G. R. Tristram, *Advances in Protein Chem.*, **5**, 83 (1949).

"Methods for Microbiological and Chemical Determinations of Essential Amino Acids in Proteins and Foods," M. J. Horn, D. B. Jones, and A. E. Blum, *U. S. Dept. Agr. Misc. Publ.* 696, 1950.

The Amino Acid Composition of Proteins and Foods. Analytical Methods and Results, R. J. Block and D. Bolling, Charles C Thomas, Springfield, Illinois, 1951.

"Methods for the Determination of Amino Acids," H. S. Olcott in Greenberg's *Amino Acids and Proteins*, p. 56, Charles C Thomas, Springfield, Illinois, 1951.

The Microbiological Assay of the Vitamin B-Complex and Amino Acids, E. C. Barton-Wright, Sir Isaac Pitman & Sons, London, 1952.

"The Amino Acid Composition of Proteins," G. R. Tristram in Neurath and Bailey's *The Proteins*, Vol. IA, p. 181, Academic Press, New York, 1953.

Original Articles

"Über die Hydrolyse des Caseins durch Salzsäure," E. Fischer, *Hoppe-Seyler's Z. physiol. Chem.*, **33**, 151 (1901).

"The Origin of the Humin Formed by the Acid Hydrolysis of Proteins. II. Hydrolysis in the Presence of Carbohydrates and of Aldehydes," R. A. Gortner, *J. Biol. Chem.*, **26**, 177 (1916).

"The Photometric Determination of Tryptophane, Tyrosine, Diiodotyrosine, and Thyroxine," E. Brand and B. Kassell, *J. Biol. Chem.*, **131**, 489 (1939).

"A New Procedure for Quantitative Analysis by Isotope Dilution, with Application to the Determination of Amino Acids and Fatty Acids," D. Rittenberg and G. L. Foster, *J. Biol. Chem.*, **133**, 737 (1940).

"A Lysineless Mutant of Neurospora and Its Inhibition by Arginine," A. H. Doermann, *Arch. Biochem.*, **5**, 373 (1944).

"Paper Chromatography Applied to the Isotopic Derivative Method of Analysis," A. S. Keston, S. Udenfriend, and M. Levy, *J. Am. Chem. Soc.*, **69**, 3151 (1947).

"The Amino Acid Requirements of Twenty-three Lactic Acid Bacteria," M. S. Dunn, S. Shankman, M. N. Camien, and H. Block, *J. Biol. Chem.*, **168**, 1 (1947).

"Chromatography of Amino Acids on Starch Columns. Separation of Phenylalanine, Leucine, Isoleucine, Methionine, Tyrosine, and Valine," W. H. Stein and S. Moore, *J. Biol. Chem.*, **176**, 337 (1948).

"Buffered Filter Paper Chromatography of the Amino Acids," E. F. McFarren, *Anal. Chem.*, **23**, 168 (1951).

"Quantitative Determination of Amino Acids in Protein Hydrolyzates by Paper Chromatography," R. R. Redfield and E. S. G. Barron, *Arch. Biochem. and Biophys.*, **35**, 443 (1952).

"Influence of Cations on the Ninhydrin Reaction for the Determination of Amino-Acids," H. Meyer and E. Riklis, *Nature*, **172**, 543 (1953).

"Amino Acid Composition of Crystalline Carboxypeptidase," E. Smith and A. Stockell, *J. Biol. Chem.*, **207**, 501 (1954).

"Colorimetric Determination of Carbamylamino Acids and Related Compounds," S. B. Koritz and P. P. Cohen, *J. Biol. Chem.*, **209**, 145 (1954).

"Phenylthiohydantoins in Amino Acid Analysis," J. Sjöquist, *Biochim. et Biophys. Acta*, **16**, 283 (1955).

"Über quantitative Papierchromatographie von Aminosäuren in Proteinhydrolysaten," E. Kofranyi, *Hoppe-Seyler's Z. physiol. Chem.*, **299**, 129 (1955).

"The Dependence of R_f Values of Amino-Acids on the Shape and Size of the Filter Paper," H. C. Chakrabortty and D. P. Burma, *J. Indian Chem. Soc.*, **32**, 533 (1955).

"Sources of Error in Microbiological Determinations of Amino Acids on Acid Hydrolysates. II. Apparent Loss of Amino Acids on Storage," M. J. Horn, A. E. Blum, C. E. F. Gersdorff, and H. W. Warren, *Cereal Chem.*, **32**, 64 (1955).

"The Kinetics of Acid Hydrolysis of Dipeptides," J. I. Harris, R. D. Cole, and N. G. Pon, *Biochem. J.*, **62**, 154 (1956).

Metabolic and Nutritive Significance of the Amino Acids

Anabolism

Catabolism

Relationship of amino acids to other substances from nature

A few integrated pathways

Metabolism of some individual amino acids

Nutritional availability

The time factor

Cofactors

The essential amino acids

Effects of antipodes

As knowledge of metabolism and nutrition grows, it becomes increasingly apparent that the two fields have much in common. The differences between them are mainly differences in emphasis. From a comparative biological point of view, the amino acid requirements can be expressed by the equation

$$\begin{aligned} \text{Synthetic abilities} + \text{nutritional requirements} \\ = \text{metabolic requirements} \end{aligned}$$

The requirements for metabolism, however met, are very much alike from one organism to another. The roster of needed amino acids is relatively constant over the phylogenetic spectrum. Within the limits of this roster, however, considerable variation is found in synthetic abilities. Organisms differ markedly in their requirements of individual amino acids in the diet. Some can synthesize all of these

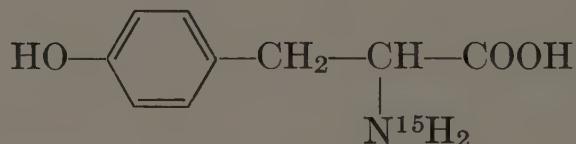
metabolic needs, others none, and various degrees of intermediate synthetic ability are found. Those amino acids that cannot be synthesized must be obtained in the food, a fact which emphasizes nutrition.

HISTORY

Among the early workers in the area of amino acid nutrition and metabolism, the protein chemist Emil Abderhalden carried out many studies with protein hydrolyzates. He helped to form the basis for the concept that some amino acids, such as tryptophan, are more critical components of the diet than others. Abderhalden's demonstration, early in the century, that he could maintain the tissue nitrogen of a twelve-year-old boy by parenteral feeding of a properly supplemented hydrolyzate after the boy's esophagus had been rendered useless by an accident, was dramatic proof of the possible applications of knowledge of this sort. In this country, studies of protein nutrition were initiated by Osborne and Mendel at Yale. The scientific lineage of much of the work that has been carried out since can be traced to the early Yale school.

A surer type of experimental approach to the problem of protein and amino acid nutrition was developed especially by William C. Rose at the University of Illinois. The use by earlier workers of hydrolyzates from which individual amino acids had been selectively removed, led frequently to ambiguous results in the state of knowledge that then existed. Rose and his students employed instead mixtures of pure amino acids. In 1935 the Rose group completed the list of amino acids which, with proper dietary supplements, must be present in order for the rat to attain approximately normal growth and development. These critical amino acids were designated as *essential amino acids*. The reasons why these and other amino acids are needed by the organism are rooted in the metabolic pathways involved either in the construction of protein or in its conversion to other biologically important substances.

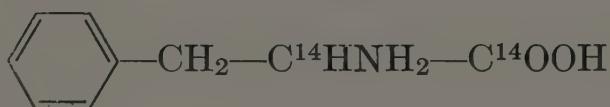
The modern era of our understanding of the metabolism of the amino acids began about 1930, when Rudolf Schoenheimer applied tracers to the studies in this field. For example, isotopic tyrosine



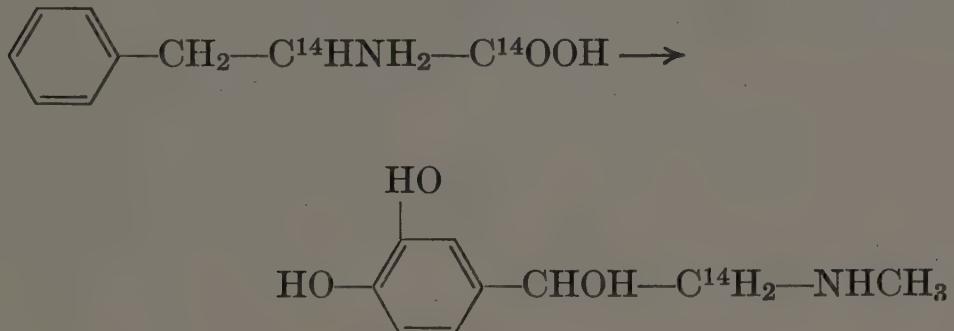
containing the heavy isotope N^{15} , was fed to rats. The carcass and

liver proteins were found to contain the isotope, not only in the form of tyrosine but also as other amino acids, such as arginine, histidine, glutamic acid, and aspartic acid. In yet other experiments it was learned that isotopic nitrogen incorporated into ammonium citrate could be recovered in the same amino acids. Thus was established a dynamic interconvertibility of some amino acids, a fact which helped to explain why some must be included in the diet and others need not be. From such studies grew a concept that invalidated previous assumptions; it was now clear that, in the main, protein metabolism was a dynamic type of process. The temporal factor in amino acid nutrition (p. 123) is also better understood.

As another illustration of the applicability of the tracer method, the work of Gurin and Delluva (1947) may be considered. By employing radioactive phenylalanine



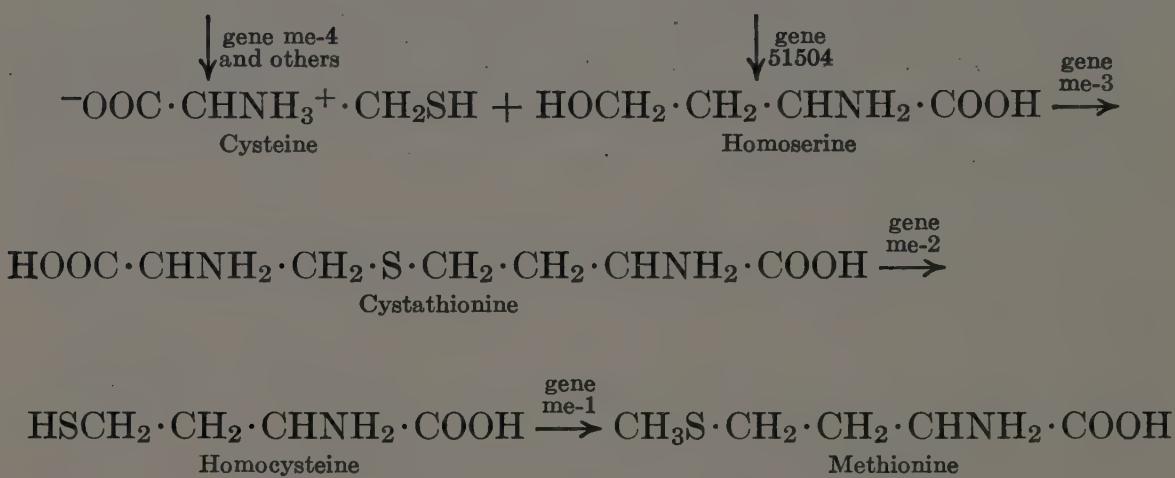
these workers were able to contribute to the elucidation of the biogenesis of adrenalin. From rats fed this radioactive amino acid, it was possible to isolate radioactive adrenalin in which only the β -carbon atom was radioactive:



The use of tracers is by no means the only technique that has contributed to the modern knowledge of metabolism. The use of biochemical mutants, for example, has proved of value in elucidating pathways of synthesis in microorganisms.

The application of mutant microorganisms has been especially fruitful in studies of biosynthesis (Beadle, 1945). Such mutants are designated by the metabolite for which synthesis has been lost, followed by the suffix “-less.” A mutant in which the capacity for synthesis of leucine has been lost is known as a *leucineless* mutant. There may be several *leucineless* mutants, corresponding to the several

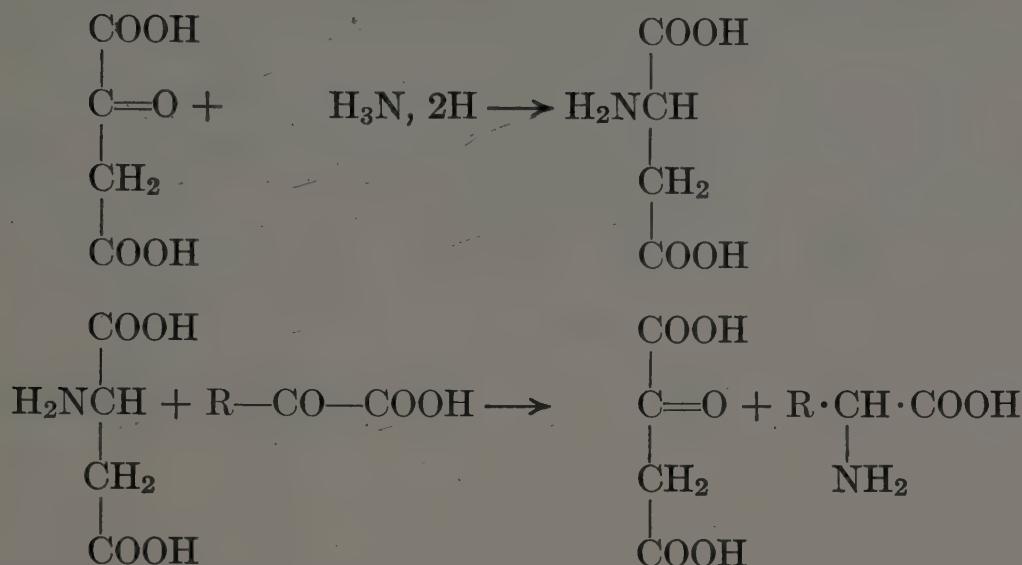
steps in synthesis of leucine by the organism. The accumulated biochemical and genetic data lead to the conclusion that each chemical reaction in the organism is controlled indirectly by a single gene. The substance for which the subsequent reaction has been impaired may, in a minor proportion of the cases, be recognized by its accumulation in the medium. In other instances, a variety of postulated intermediates is tested until the correct molecular link in the synthetic chain is discovered. A typical flow sheet of gene-controlled reactions as worked out for the biosynthesis of methionine is (Teas *et al.*, 1948):



A biochemical reaction may be specifically blocked by a chemical antimetabolite, as well as by impairment of the controlling gene. The metabolite is then similarly investigated through its ability, upon addition to the altered system, to overcome the antimetabolite. In the synthesis of tryptophan (p. 122) information has been obtained both through biochemical mutants that have lost the ability to synthesize the amino acid, and through the use of various methyl-indoles that block the conversion of indole to tryptophan. The technique of "chemical block" has not been as fully cultivated as the others given here, but with increased understanding of competitive inhibitions (Potter, 1951) this mode of investigation offers some uniquely attractive features.

ANABOLISM

General Anabolism in Plants. The primary source of amino acids and proteins is the plant kingdom. Plants convert inorganic nitrogen to ammonia, which can in turn react with keto acids such as oxaloacetic acid and be simultaneously reduced to aspartic acid. Aspartic acid is then capable of transferring amino groups to α -keto acids arising in the carbohydrate phase of photosynthesis. One scheme follows.



Microorganisms. The synthetic ability of microorganisms, as compared to that of man, is quite extensive. A small proportion of the total number of microorganisms which have been studied is capable of synthesizing its amino acids and vitamins from inorganic nitrogen. Many microorganisms can however be "trained" to synthesize various amino acids, such as tryptophan. The acquisition and loss of such abilities appear to be rather general phenomena. Retention of abilities to carry out some but not all steps in a synthesis is considered to be analogous to morphological evolutionary vestiges, like the tonsils. Stepwise recovery of synthetic steps, in fact, provides another method for charting biosynthetic pathways.

Animals. The capabilities of animals for synthesis are much more limited than those of plants and unicellular creatures in general. Mammals vary in their nutritional requirements; this is in part due to the fact that their digestive flora consist of microorganisms which can in turn carry out synthesis of nitrogen compounds. This is particularly true of the cow, whose ruminant metabolism relies heavily upon the synthetic abilities of a varied flora. Examples of synthesis by animals have been mentioned in Chap. 2 (thyroxine, epinephrin).

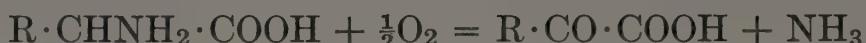
CATABOLISM

Two main stages in the breakdown of protein may be considered:

- (1) Protein $\xrightarrow{\text{H}_2\text{O}}$ peptides and amino acids
- (2) Amino acids \longrightarrow simpler nitrogen compounds

The first stage is described on p. 399. Only the catabolism of amino acids will be taken up in this chapter.

Other Organisms. The reactions involved in the breakdown of amino acids in animals concern the conversion of the grouping at the α -carbon. The acid may be decarboxylated, as indicated in the formation of histamine and of tyramine. It may be deaminated, the nitrogen being converted principally to urea and subsequently excreted. The remainder of the amino acid molecule, the side chain, undergoes more specific reactions, treated later in this chapter. The reaction of the alpha structure that has received the most study in animals is deamination, which is represented by the over-all equation



The keto acid product is important as an intermediate in the synthesis of non-nitrogenous substances. The ammonia is eliminated, to at least a considerable extent, as urea through the ornithine cycle (see urea, p. 111).

RELATIONSHIP OF AMINO ACIDS TO OTHER SUBSTANCES FROM NATURE

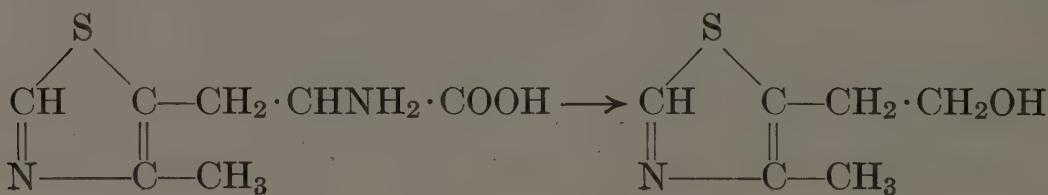
Carbohydrates. Proteins, or amino acids generally, are capable of conversion to carbohydrate. This is illustrated by Table 2-1, in which a number of amino acids are listed as being glycogenic. Glycogen represents the storage polysaccharide of the glucose obtained from the amino acids. This conversion to glycogen probably occurs largely through the keto acids that result from deamination of the amino acids.

Fat. The production of fat from protein may occur through the path of protein \rightarrow amino acids \rightarrow glucose \rightarrow fat, or from keto acids.

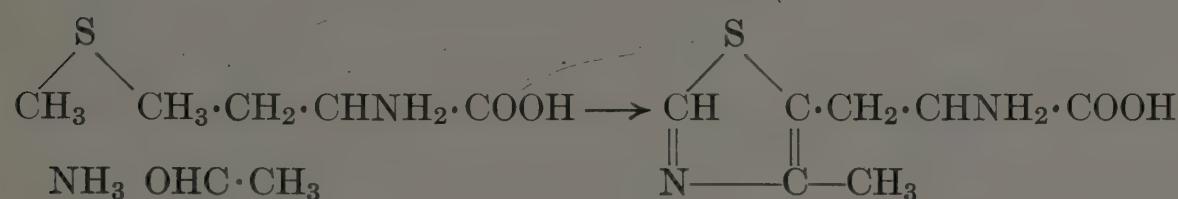
It has long been recognized that a 2-carbon intermediate must be involved in the synthesis of the fatty acids. It is now well established that this intermediate is the acetyl group of acetyl-coenzyme A.

Vitamins. The conversion of amino acids to vitamins has received some attention. The widespread occurrence of nitrogen in the water-soluble vitamins has argued for such relationships.

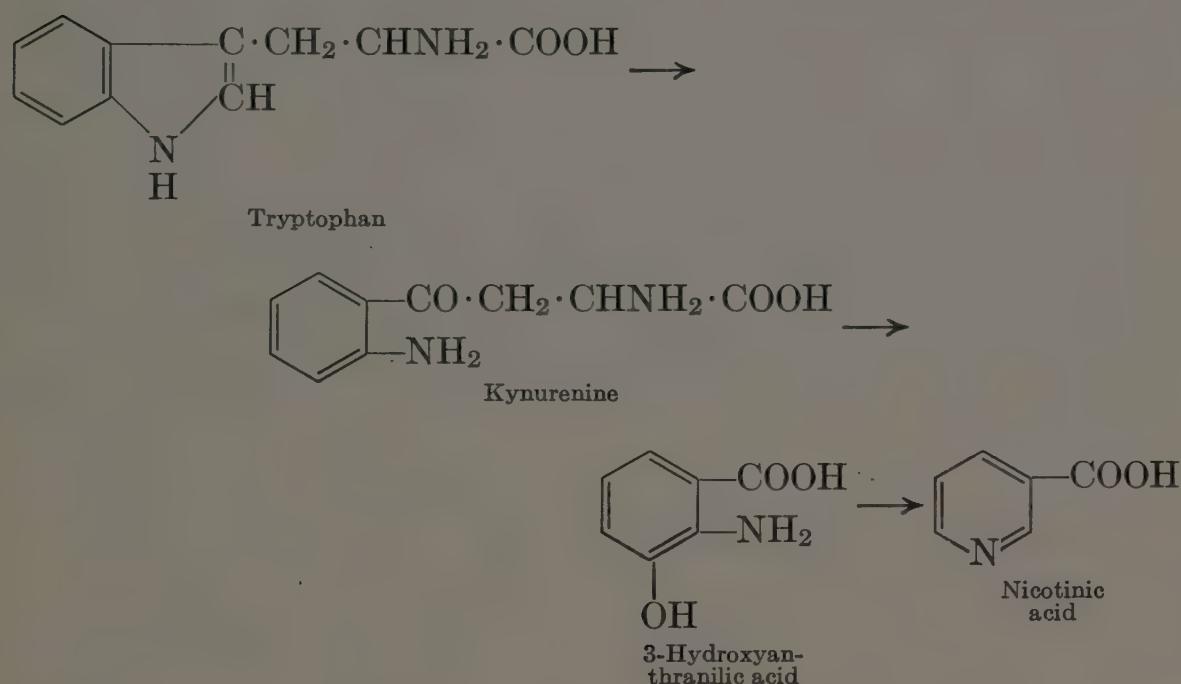
Yeast has been found to be capable of converting methylthiazolyl alanine into the thiazole group of thiamine:



The heterocyclic amino acid could theoretically be built from methionine, ammonia, and acetaldehyde.



Nicotinic acid can be partially replaced in the diet by tryptophan; tryptophan thus spares nicotinic acid. Part of the chemical pathway involved is as follows:

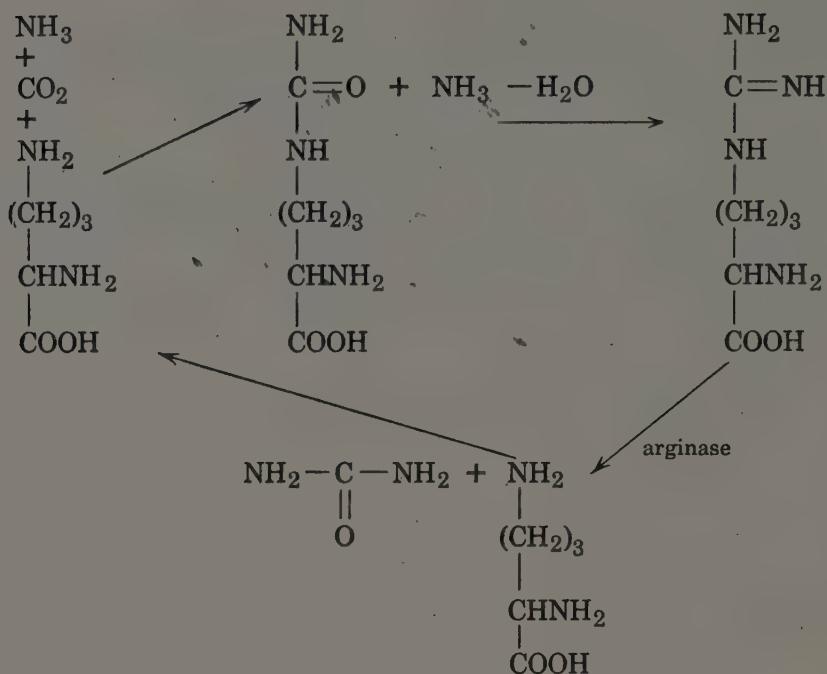


The β -alanine moiety of pantothenic acid arises in some cases by decarboxylation of aspartic acid:

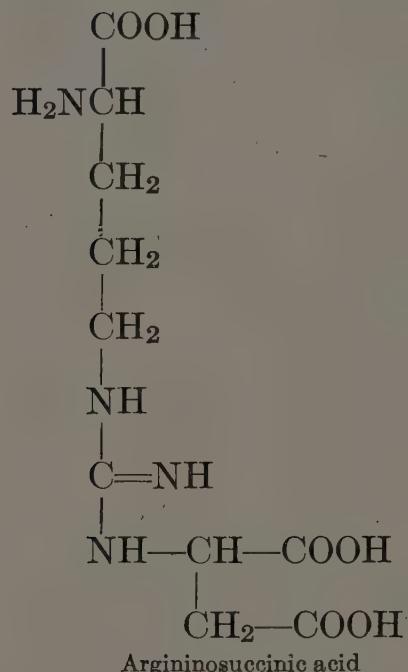


The folic acids are derivatives of glutamic acid and polyglutamic acid peptides.

Urea. The formation of urea in the breakdown of protein is quantitatively an important process, as evidenced by the fact that urea is the main solute found in urine. A principal mode of formation of urea in mammals (also in such phylogenetically distant forms as *Neurospora crassa*) is through the amino acids ornithine, arginine, and citrulline. This cycle is known as the urea cycle, or sometimes as the ornithine cycle. The following are the main reactions involved.

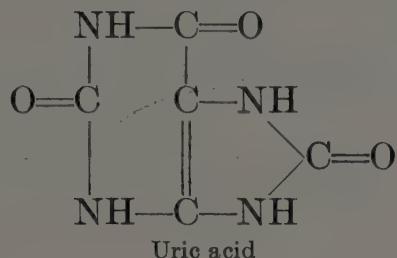


The ornithine becomes available for combination with more molecules of ammonia. These latter result from breakdown of nitrogen compounds. If they led only to ammonia, this would soon be present in toxic concentration. The ammonia is transferred not as such to citrulline, but through argininosuccinic acid, in a reaction which requires ATP.



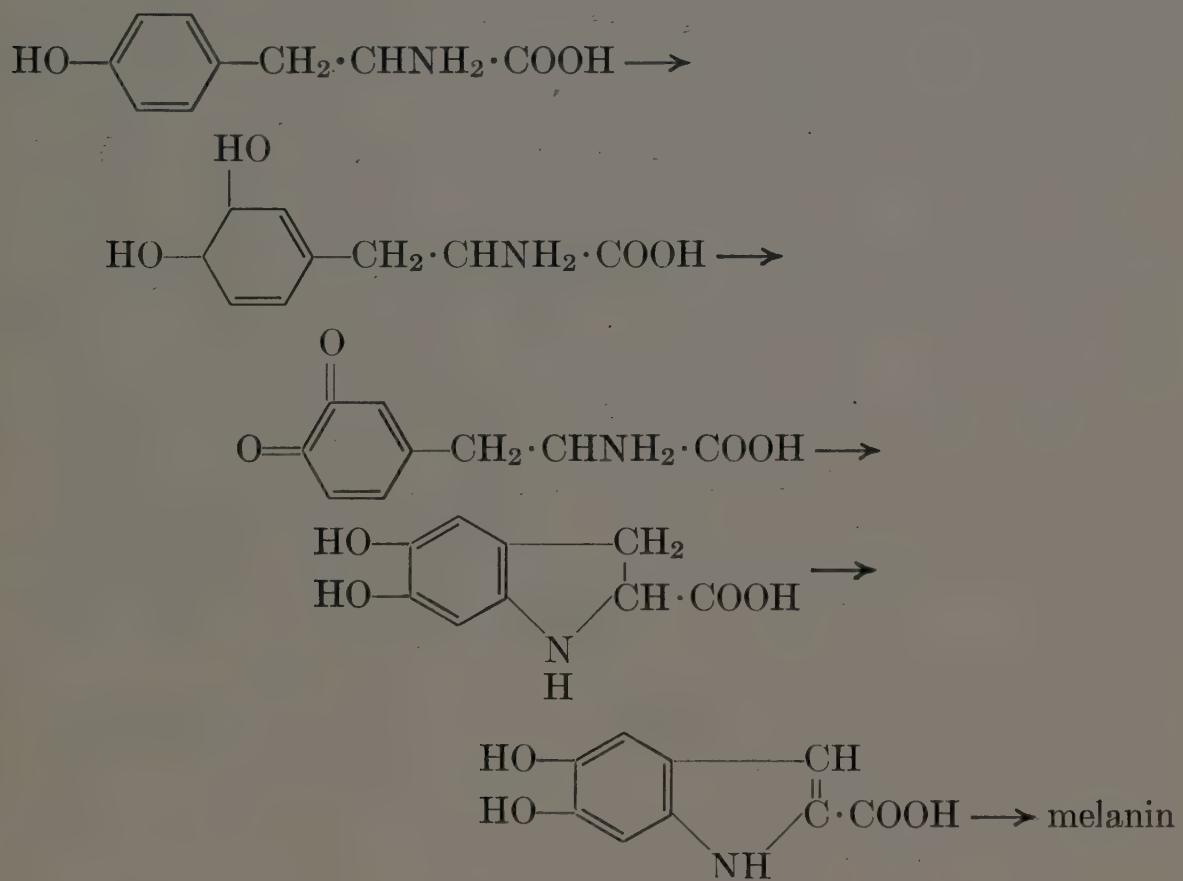
This entire process may be looked upon as a physiological means for combining ammonia and carbon dioxide to form urea. The synthesis is found to occur in most mammals, but birds, snakes, and lizards excrete uric acid, which may be looked upon as a derivative

of urea. The former type of metabolism is *ureotelic*, the latter *uricotelic*.



Uric acid is seen to be a cyclic derivative of urea.

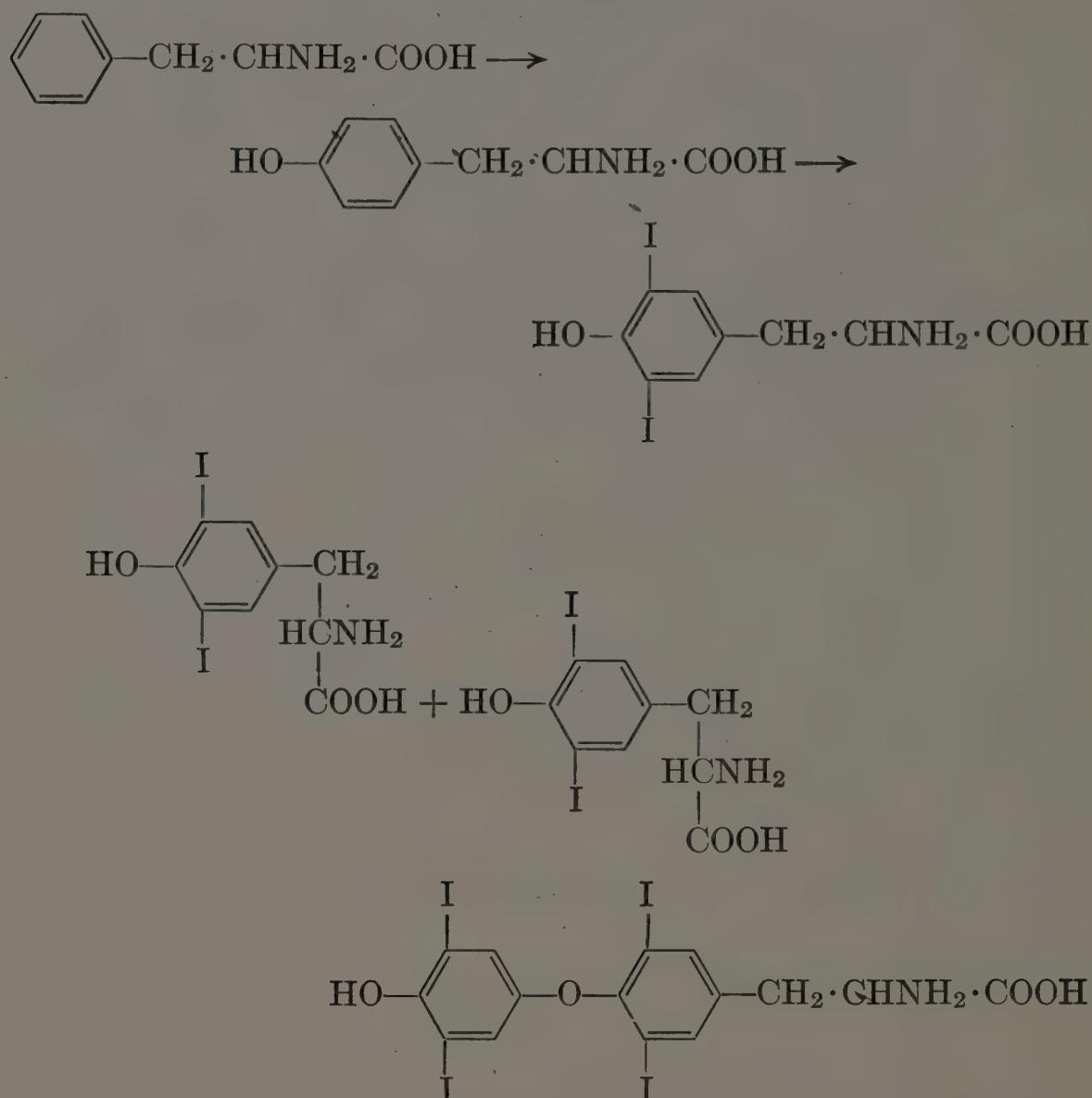
Melanin. The pigment of skin and hair, the coloring of slices of potato and apple, and related compounds, are formed by oxidation of the amino acid tyrosine. Reactions which have been suggested to explain this conversion are



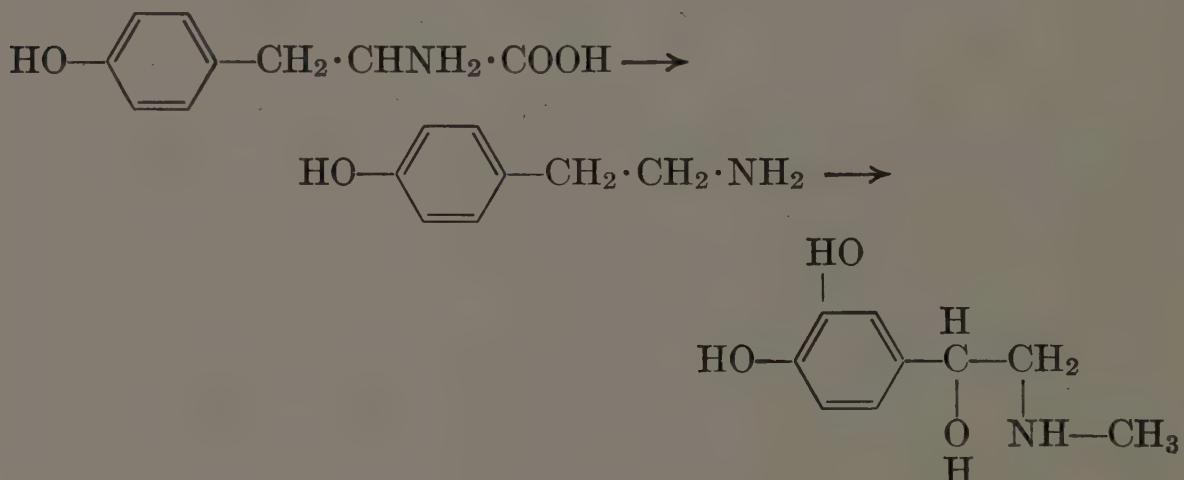
The constitution of melanin itself and of the related melanoidins is unknown. It is likely that they constitute a series of substances of somewhat diverse structure.

Hormones. Tyrosine is also of importance as an intermediate in the formation of thyroxine. As noted in Chap. 2, thyroxine is not designated as an essential amino acid even though it is essential to mammalian life, since it need not be included in the diet. This is a consequence of the presence in food of the essential amino acid phenylalanine, or in some cases tyrosine, either of which is transformed

to thyroxine. The reaction sequence representing the conversion of phenylalanine to the hormonal thyroxine is

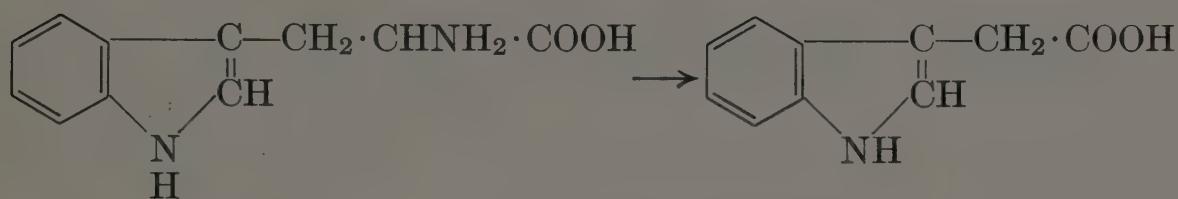


Tyrosine is also an intermediate in the mammalian synthesis of the hormone epinephrine (adrenalin, p. 107):



The order of the oxidations and methylation of tyramine is not known; however, the methyl group is at least partially derived from methionine (p. 121).

The plant hormone indoleacetic acid (heteroauxin) is derived from the indolyl amino acid tryptophan:



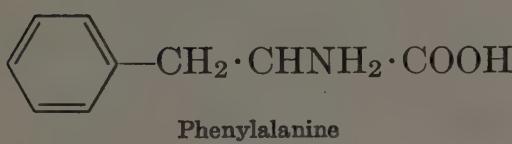
Indoleacetic acid was first studied as a growth hormone in corn. It has been found that as the corn matures, indoleacetic acid content increases at the expense of tryptophan.

Antibiotics. Many antibiotics have either a protein or a modified peptide structure. These are discussed in detail in Chap. 8, but it is appropriate here to point out that many antibiotics are derivatives of amino acids and in many cases are found to contain the unusual D-amino acid residue.

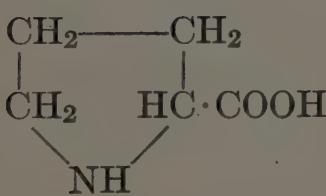
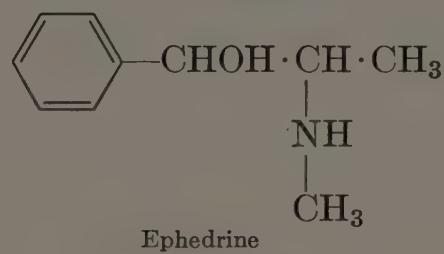
Alkaloids. The pharmacologically active nitrogen compounds found in some species of plants are believed to arise in the main from amino acids. One reason for this belief is that many of the heterocyclic rings are common to both classes of compounds. It is a relatively simple exercise to devise schemes by which the alkaloids might arise from the corresponding amino acids.

It has also been possible to produce, under physiological conditions, alkaloids from compounds obtainable from amino acids by catabolism. Examples of the structural relationship are given here.

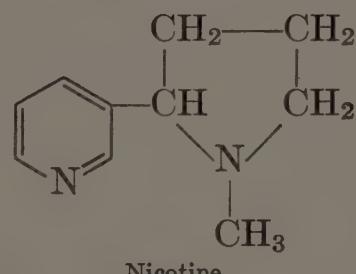
Amino Acid



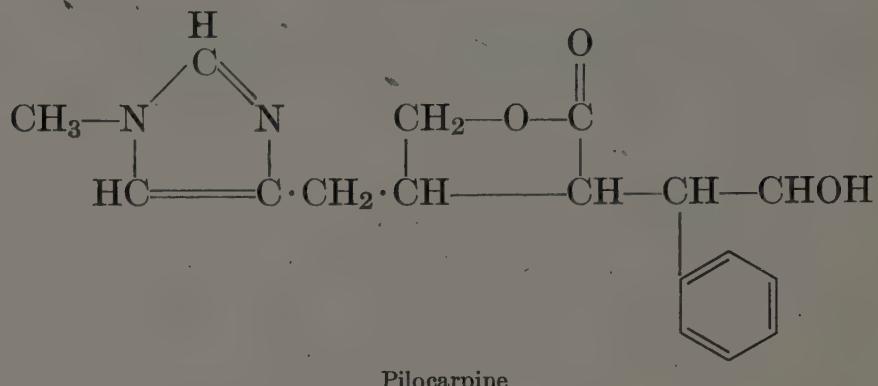
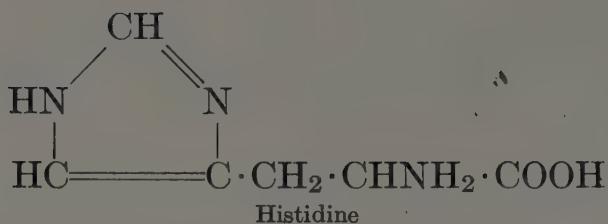
Alkaloid



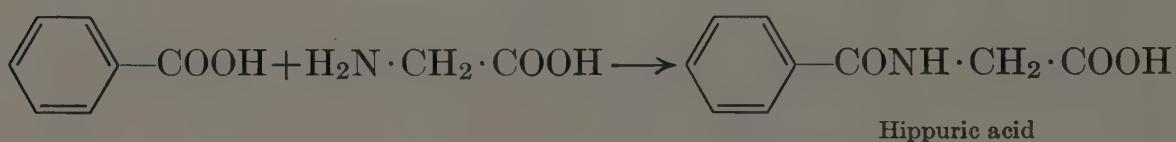
Proline



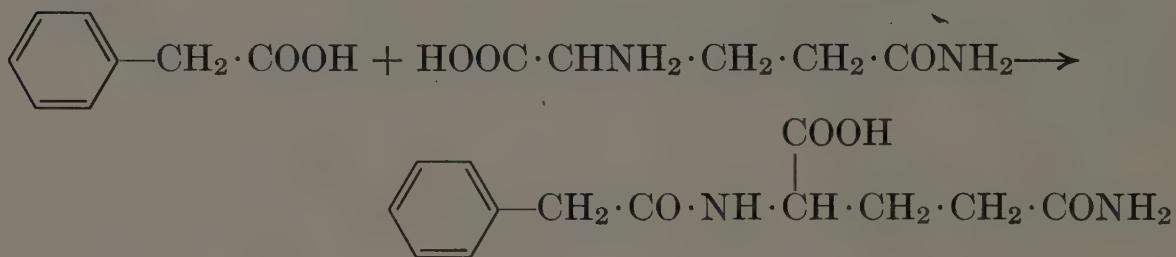
Nicotine



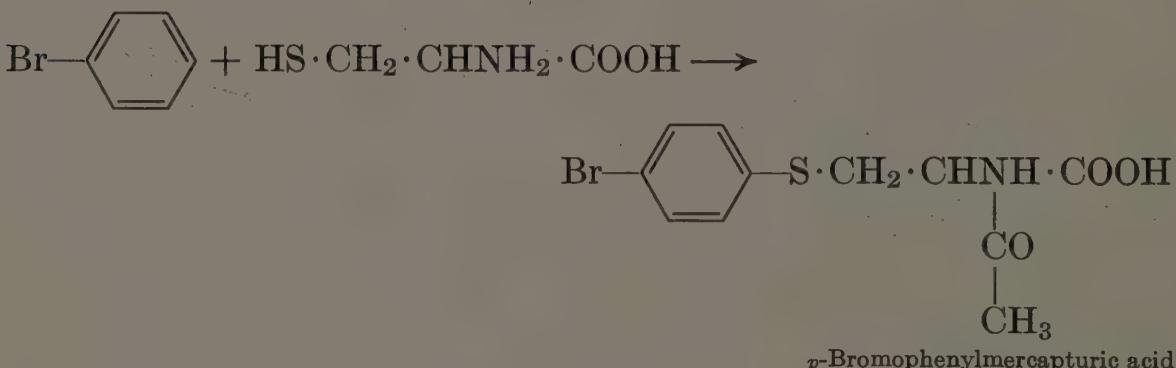
Detoxication. The amino acids of recognized significance in detoxication mechanisms are glycine, glutamic acid, and cysteine. Glycine detoxifies organic acids such as benzoic acid:



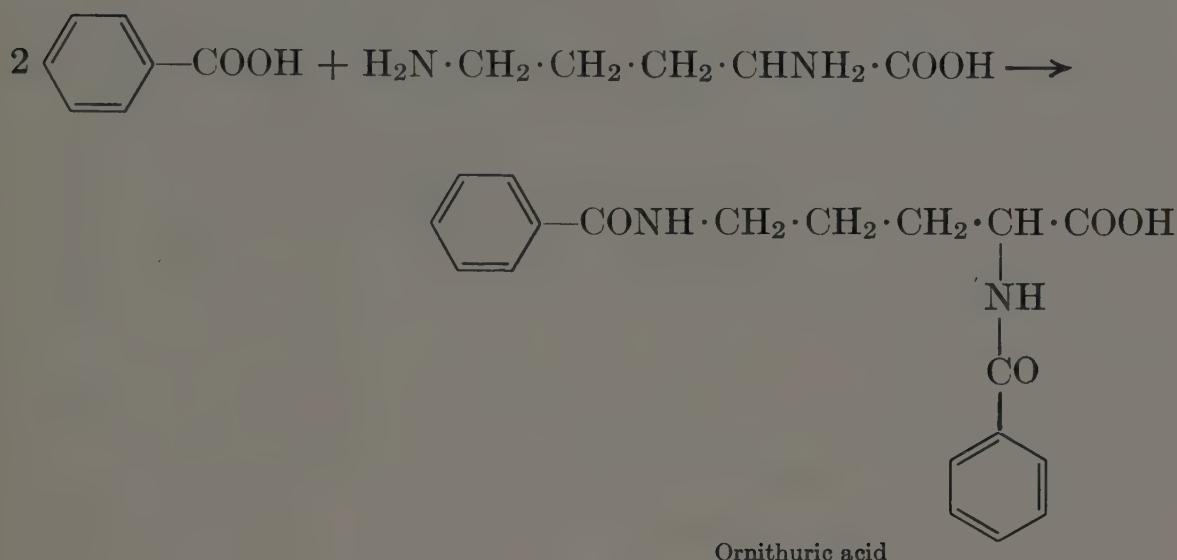
The amide of glutamic acid, glutamine, reacts, for example, with phenylacetic acid:



Cysteine can detoxify aromatic compounds in the following way:



Birds utilize ornithine for detoxication of organic acids:



A FEW INTEGRATED PATHWAYS

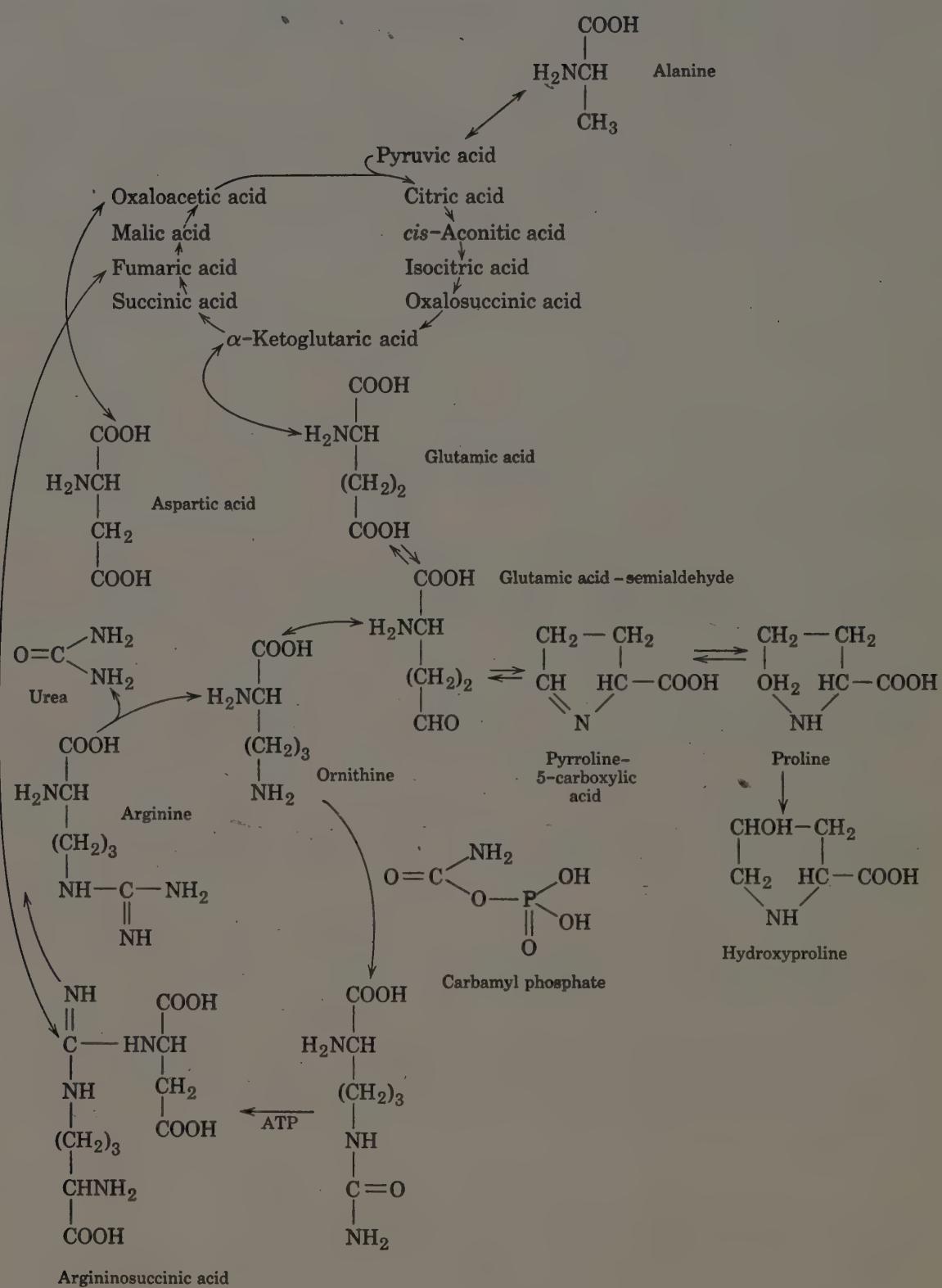
The complete exposition of reasonably well established pathways of amino acid metabolism is infeasible, but two will be given. These are selected in part for their involvement in other types of metabolism. Other metabolic pathways are partly illustrated in the section on metabolism of individual amino acids and in the preceding discussion.

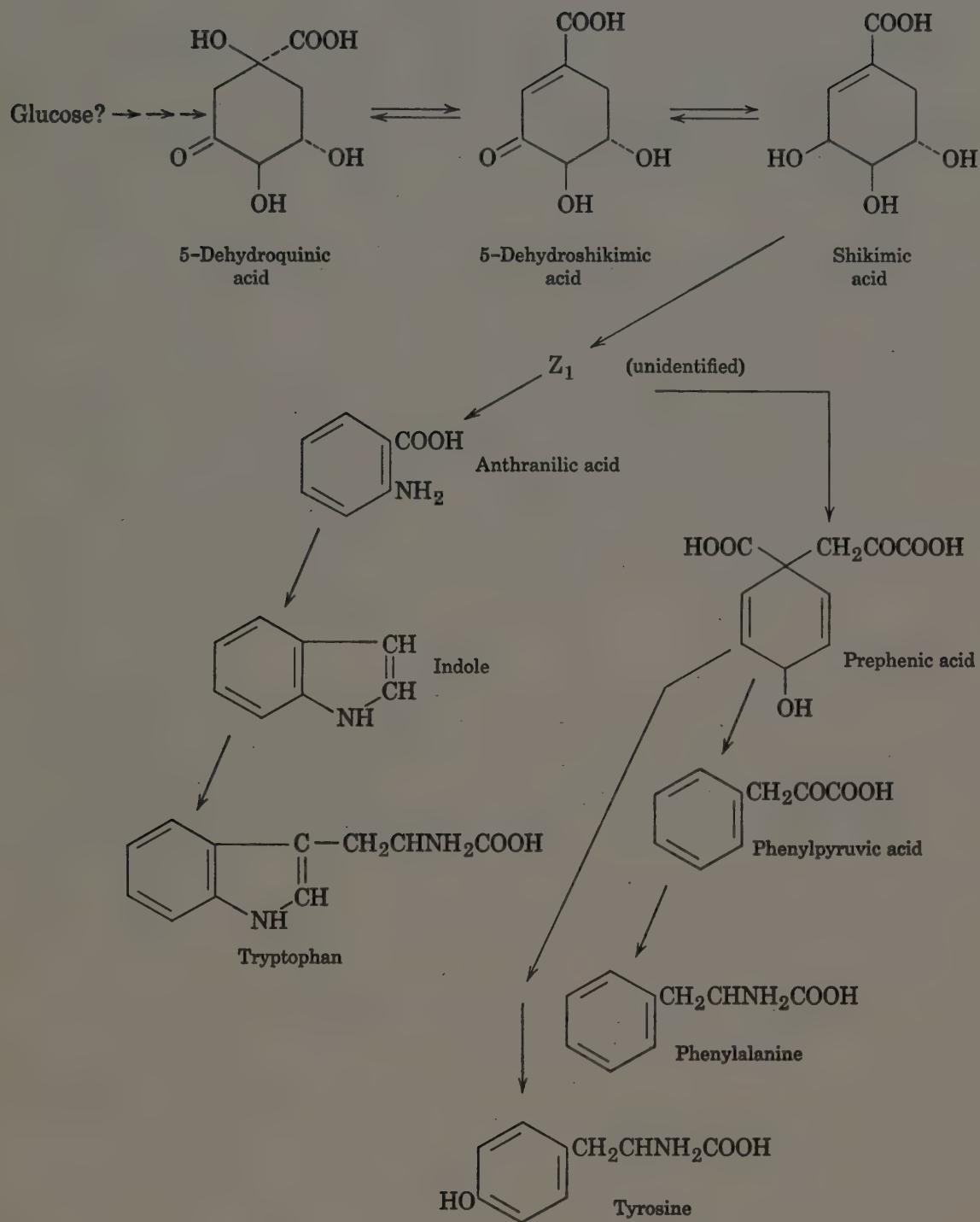
In the first scheme (p. 118) the direct linkage of the synthesis of aspartic acid, alanine, and glutamic acid to the Krebs cycle is shown. By virtue of transamination, many other amino acids also arise from the Krebs cycle, especially through glutamic acid and aspartic acid.

The second scheme (p. 119) presents some of the knowledge of pathways of aromatic biosynthesis. Available evidence suggests that the aromatic intermediates arise from glucose. The aromatic amino acids (not including histidine) thus originate at a later stage of anabolism than do aspartic acid, alanine, and glutamic acid. Phosphorylated monosaccharides such as hexosediphosphate and sedoheptulose-7-phosphate function in experiments as intermediates, whereas glucose does not.

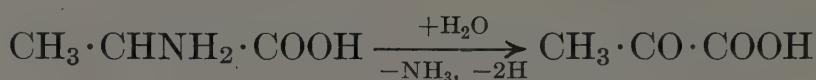
METABOLISM OF SOME INDIVIDUAL AMINO ACIDS

Aside from the general reactions which have been considered for breakdown of amino acids, some special pathways due principally to chemical vulnerability of side chains have been elucidated. Those given here will apply to the L-forms, which have received nearly all of the attention in this field.



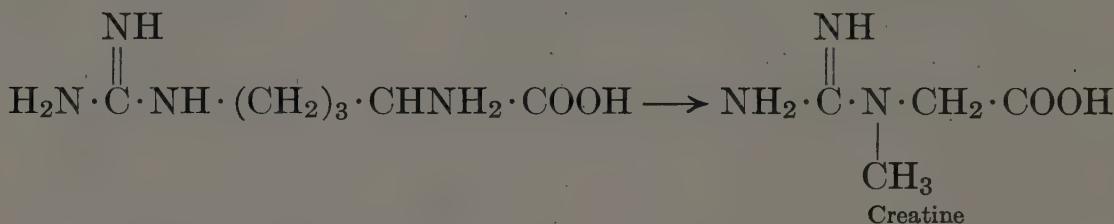


Alanine. Alanine, a glycogenic amino acid, is in part deaminated to pyruvic acid:



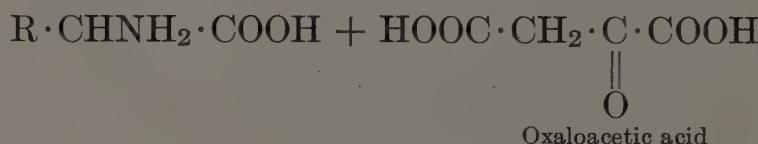
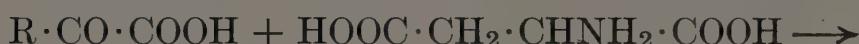
The resultant pyruvic acid is convertible to glucose, glycogen, and acetyl-coenzyme A.

Arginine. The guanidino group of arginine is transferrable to glycine, yielding creatine, a metabolic product of some interest. Creatine contains a methyl group, derivable from methionine:



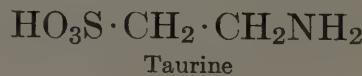
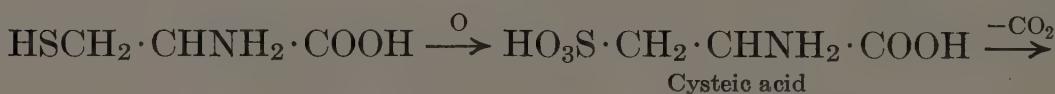
The importance of arginine in the urea cycle has also been indicated.

Aspartic Acid. Aspartic acid is metabolically important as a donor of amino groups to keto acids, in the transamination reactions

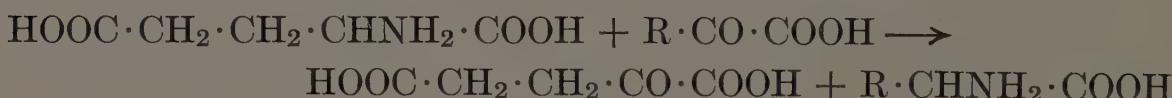


Through oxaloacetic acid, aspartic acid is closely connected to important respiratory pathways such as the Krebs citric acid cycle. This means it figures significantly both in carbohydrate catabolism and in photosynthesis, and is in a sense a nitrogen backlog in the fire of metabolism.

Cysteine and Cystine. A bile constituent, taurine, is believed to be formed from cysteine.



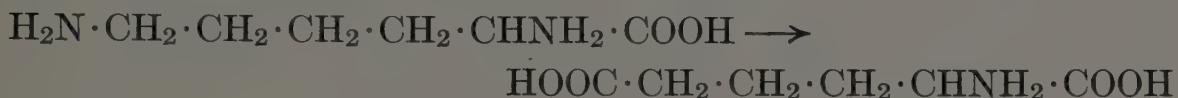
Glutamic Acid. Glutamic acid is especially important in the transamination reaction in animals.



Glycine. Glycine has been shown to be a probable precursor of heme, the pigment of hemoglobin. This conclusion is based upon

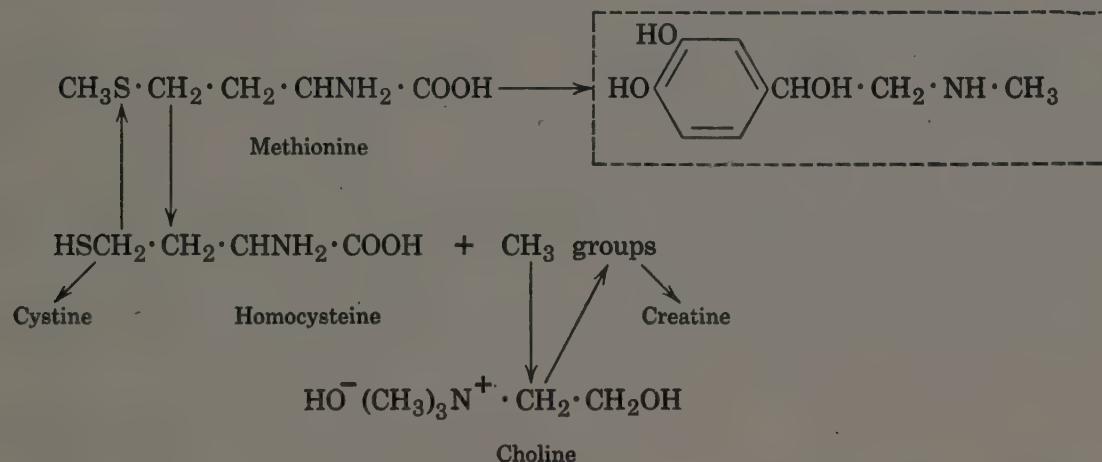
studies of the fate of ingested isotopic glycine. The available evidence indicates also that glycine is involved in a reversible conversion to serine.

Lysine. Borsook has shown that lysine is metabolized to α -amino-adipic acid in liver homogenate:



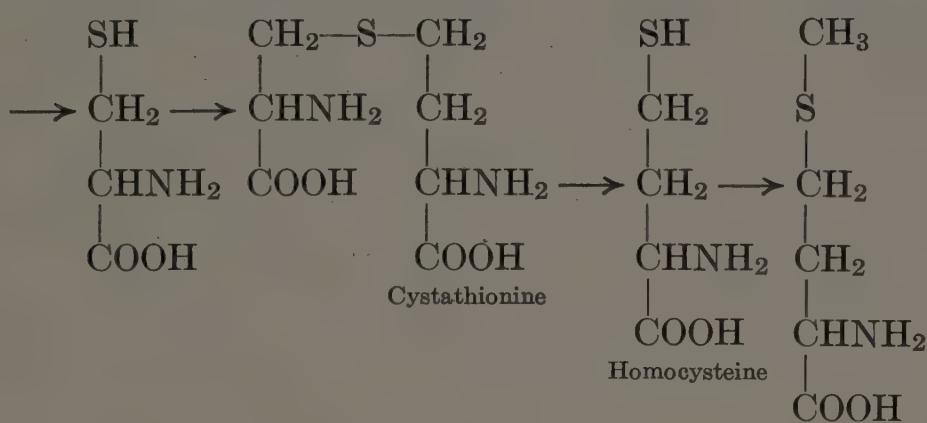
The reverse process can occur in a lysineless mutant of *Neurospora*.

Methionine. A summary chart of the metabolic relationships of methionine (H. B. Lewis) follows.



This chart illustrates especially the manner in which methionine participates in transmethylation.

As elaborated later in this chapter, methionine is convertible to cystine by mammals, but the reverse is not true. The synthesis of methionine in *Neurospora* has been shown by Horowitz to proceed through cysteine.

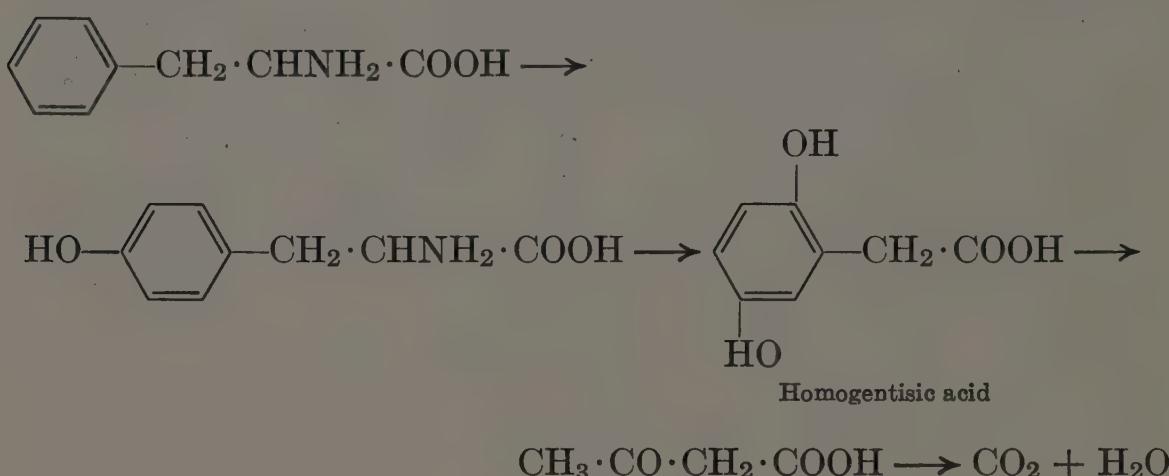


This is in essence the reverse of the pathway of mammalian catabolism. This reverse type of relationship between catabolism in mammals and anabolism in biosynthetic organisms is being increas-

ingly illustrated by experimental findings (compare, for example, lysine) and appears to be of fundamental significance.

Phenylalanine and Tyrosine. The essential nature of phenylalanine in the diet, as contrasted to tyrosine, is a situation comparable to that of methionine and cystine. Tyrosine is producible from phenylalanine by mammals. This conversion has been found to operate also in one strain of microorganism, *Escherichia coli*, but is recorded as not occurring in another, *Lactobacillus arabinosus*.

Aside from the previously detailed conversions of phenylalanine to such other natural products as epinephrin, thyroxine, and melanin, evidence exists for its catabolism through the following pathway.



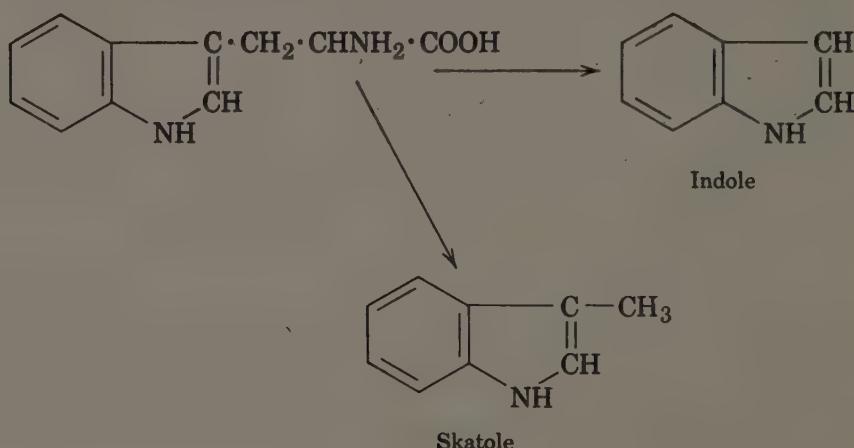
Serine. Serine is convertible to glycine in mammals, a capability that explains also its convertibility to heme.

It is probable that serine is a precursor of colamine (ethanolamine):

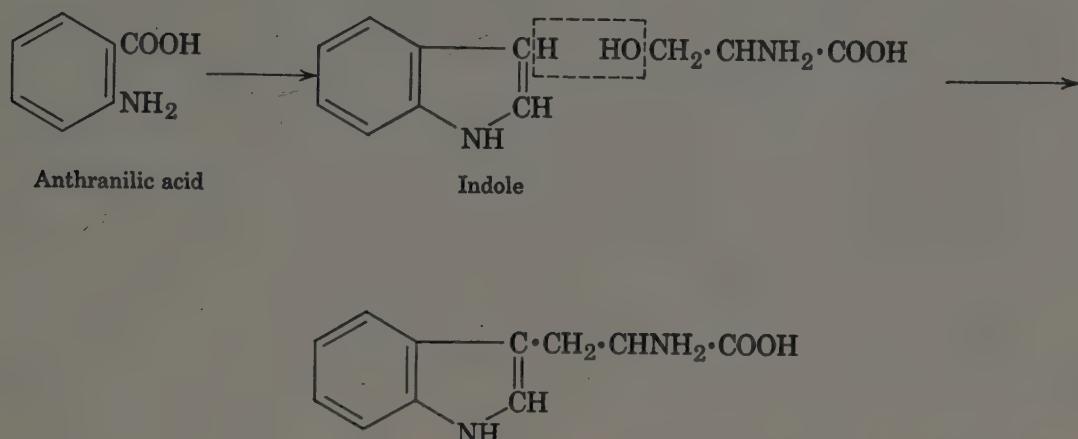


in a relationship not unlike that between cysteine and taurine.

Tryptophan. Putrefactive metabolism of tryptophan results in production of indole and skatole:



The synthesis of tryptophan by a process operating in the reverse direction has been found in a number of microorganisms:



Most of the early work on this synthesis was done with *Escherichia coli*.

In a classical experiment with *Neurospora*, a tryptophanless mutant failed to carry out the last step, and the failure at this specific point was recognized by a separation from the culture medium of the relatively insoluble indole. This discovery was a turning point in the field of chemical genetics.

NUTRITIONAL AVAILABILITY

The literature provides indications that the nutritional value of a peptide or protein may reside not only in the amino acid residue composition but also in the particular sequential arrangement. There are a few instances in which certain peptides are more readily available to a given microorganism than are others; on the other hand, there are known instances in which the availability of the fully hydrolyzed material is clearly greater than that of the unhydrolyzed.

The prime example of masked availability has been soybean meal, from which a nutritionally superior material could be obtained by heating. One school of thought holds that the heating breaks up certain refractory combinations of amino acid residues. It has also been suggested, however, that heating destroys a trypsin inhibitor which is known to occur in soybean. If the latter interpretation is correct, the impaired availability is due not to a special linkage but rather to the inhibition of action of enzymes concerned with utilization of the amino acids that are potentially available.

THE TIME FACTOR

Although suggested earlier, it became clear at a relatively late period in nutritional investigation that optimal utilization of the amino

acids required the simultaneous availability of adequate amounts of all of them. Omission of any one essential amino acid from the diet of a mammal, for instance, for a period of only three hours, could result in complete failure to use all of the other essential amino acids. The so-called nonessential amino acids can be synthesized from other components in the diet, but the rates of such syntheses appear to be limiting, and the time factor has been shown to be important in the case of the nonessential components as well as for the essential ones.

Work on the quantitative evaluation of the need for amino acids has suggested that their requirement conforms to stoichiometric proportions. This fact has tempted some workers to speculate that these amino acids are used in some grand molecular combination in the synthesis of protein.

COFACTORS

Following the discovery of threonine, many investigators held that the nutritional quality of a protein was totally expressible in the amino acids obtainable by hydrolysis. Others believed that selected whole proteins were nutritionally superior to any combination of amino acids. The consensus has gradually favored the latter point of view. This conclusion is supported not only by the knowledge of masked availability but also by the awareness that cofactors could enhance growth obtained with amino acid mixtures.

In research begun before 1942, Merrifield and Woolley (1955) described a cofactor important for optimal utilization of amino acid mixtures by certain microorganisms. This material, which has the properties of a peptide and can be isolated from whole proteins after gentle treatment, was named *strepogenin*. Evidence for the existence of such a factor in rat nutrition has also been brought forward.

Other growth stimulants which influence utilization of amino acids include *animal protein factor* (APF) and the vitamin B₁₂ which it contains. Various vitamins have been shown to influence both quantitatively and qualitatively the requirements for individual amino acids by microorganisms.

Another type of manifestation similar to that of the cofactor is illustrated by the supplementary action of tryptophan and nicotinic acid.

THE ESSENTIAL AMINO ACIDS

The *essential* amino acids or *indispensable* amino acids have been defined as those amino acids required in the diet for growth of the young and maintenance of the adult. Besides growth and maintenance,

modern investigations recognize the importance of other nutritional functions and conditions, such as repletion of the depleted animal and the requirements for lactation in the female. The latter are particularly stringent. Many other special needs are known, notably those for antibody production.

Among the necessary qualifications of the word essential, in this connection, is correction of the frequent inference that the amino acids not so labeled are not essential. Many, almost certainly all of them, are essential, but they need not be included in the diet because the body is capable of synthesizing them (for example, thyroxine).

Moreover, what is essential for one species may not be essential for another. It is therefore necessary to designate the species in listing indispensable amino acids. Among varieties of microorganisms, for example, many different sets of amino acid requirements are to be found, including some which closely parallel those of man.

It is also important to recognize that there are degrees of essentiality. The quantity of any one amino acid that must be included in the diet for a given species is often different from the required quantity of another amino acid. Furthermore, the ability to synthesize the other amino acids will vary. Rats, for instance, were early recognized as being unable to synthesize sufficient quantities of arginine, although they are capable of synthesizing part of the requirement for that amino acid. In microbiology, amino acids that occupy such an intermediate position are known as *accessory*.

Another qualification concerns the fact that the composition of the rest of the diet, both quantitatively and qualitatively, also governs which amino acids are essential and to what extent. By "the rest of the diet" is meant not only dispensable amino acids but also carbohydrate, particular vitamins, and other factors. This sort of information has especially been obtained in studies of the nutritional requirements of microorganisms, in which large numbers of experiments can be conducted within a feasible length of time.

Species Requirements. The requirements of rats for amino acids are given in Table 7-1. One should not overlook the fact that not only are there species differences, but there is evidence for difference in requirements of individuals. This is most strikingly illustrated in the development of bacterial colonies from individual cells that have been selected from cultures "trained" either to synthesize or to require an amino acid which was not needed by the parent.

In both mice and dogs, arginine is not required, all other needs being the same as in rats. From the viewpoint of comparative biochemistry, the fact that an amino acid only partially required by one

species is not needed at all in the diet of two other mammals, may be especially significant. The requirements of chicks, on the other hand, include the same ten amino acids as those indispensable to rats,

TABLE 7-1. Amino Acid Requirements of Rats

Indispensable	Per Cent ¹	Dispensable
Lysine	1.0	Glycine
Tryptophan	0.2	Alanine
Histidine	0.4	Serine
Phenylalanine	0.7	Aspartic acid
Leucine	0.8	Glutamic acid
Isoleucine	0.5	Hydroxyglutamic acid
Threonine	0.5	Proline
Methionine	0.6	Hydroxyproline
Valine	0.7	Citrulline
X Arginine (required in part)	0.2	Tyrosine

¹ The values given represent the minimum proportion of each essential amino acid for growth, when the dispensable amino acids are also included. The figures may be considered to be first approximations.

plus one other, glycine. For obvious reasons, the requirement of amino acids for growth in the human has not been evaluated. It has, however, been assessed for maintenance of nitrogen in the adult. Here, quite surprisingly, both arginine and histidine were found to be dispensable, although the other eight amino acids are needed.

EFFECTS OF ANTIPODES

Under certain conditions some of the D-amino acids at relatively high levels in the diet have been found to be inhibitory to growth. Some D-amino acids are particularly inhibitory; this is especially true

TABLE 7-2. Availability of Optical Forms of Amino Acids to the Rat

D and L (Both)	L (Only)
Tryptophan	Lysine
Histidine	Valine
Methionine	Leucine
Phenylalanine	Isoleucine
	Threonine

of serine, but this effect has been reported also for other amino acids, such as leucine and phenylalanine. Although contrary findings have been reported, retardation of rat growth has been observed in so many laboratories as to suggest that the discrepancy is due to variations in

experimental conditions, differences in the general D-amino acid level in particular.

Most of the studies on the effects of the antipodes have been more concerned with the availability of each of the two forms. It has been found that some of the amino acids, including all of the heterocyclic types that are essential, can be utilized in either form. The availability of the two forms to the rat is given in Table 7-2. Here again, scattered data indicate species specificity.

SELECTED REFERENCES

Books and Reviews

- "Synthesis in Biochemistry," R. Robinson, *J. Chem. Soc.*, **1936**, 1079.
"The Study of Intermediary Metabolism of Animals with the Aid of Isotopes," R. Schoenheimer and D. Rittenberg, *Physiol. Revs.*, **20**, 218 (1940).
"The Significance of the Sulfur-Containing Amino Acids in Metabolism," H. B. Lewis, *Harvey Lectures Ser.*, **36**, 159 (1940-41).
"Biochemical Genetics," G. W. Beadle, *Chem. Revs.*, **37**, 15 (1945).
"Metabolite Antagonists," R. O. Roblin, Jr., *Chem. Revs.*, **38**, 255 (1946).
"The Problem of Tissue Protein Synthesis," P. R. Cannon, *Federation Proc.*, **7**, 391 (1948).
"Amino Acid Mixtures as Parenteral Protein Food," R. Elman, *Am. J. Med.*, **5**, 760 (1948).
"Amino Acid Requirements of Man," W. C. Rose, *Federation Proc.*, **8**, 546 (1948).
"The Path of Carbon in Photosynthesis," M. Calvin, *J. Chem. Educ.*, **26**, 639 (1949).
Bacterial Metabolism, M. Stephenson, Longmans Green, London, 1949.
Protein and Amino Acid Requirements, A. A. Albanese, Academic Press, New York, 1950.
"Synthesis and Metabolism of Amino Acids," J. Bonner, *Plant Biochemistry*, p. 219, Academic Press, New York, 1950.
"Sequential Blocking of Metabolic Pathways *in vivo*," V. R. Potter, *Proc. Soc. Exptl. Biol. Med.*, **76**, 41 (1951).
"Nutritional Applications of the Amino Acids," H. J. Almquist in Greenberg's *Amino Acids and Protein*, Charles C Thomas, Springfield, Illinois, 586, 1951.
"The Comparative Biochemistry of Nitrogen Fixation," P. W. Wilson, *Advances in Enzymol.*, **13**, 345 (1952).
"Amino Acids and Proteins. Their Place in Human Nutrition Problems," N. W. Flodin, *J. Agr. Food Chem.*, **1**, 222 (1953).
"Diseases of the Nervous System. Metabolic Aspects of Some Neurological and Muscular Disorders," F. H. Tyler and M. D. Armstrong, *Ann. Rev. Med.*, **4**, 207 (1954).
"Neue Amino- und Ketosäuren in grünen Pflanzen und die Biosynthese der Aminosäuren," A. I. Virtanen, *Angew. Chem.*, **67**, 381 (1955).
Genetics and Metabolism, R. P. Wagner and H. K. Mitchell, John Wiley & Sons, New York, 1955.

Amino Acid Metabolism, W. D. McElroy and B. Glass, Johns Hopkins Press, Baltimore, 1955.

"Symposium on Amino Acid Metabolism," Am. Soc. Biol. Chemists, *Federation Proc.*, **14**, 683 (1955).

Original Articles

"Über die Verwertung von tief abgebautem Eiweiss im menschlichen Organismus," E. Abderhalden, F. Frank, and A. Schittenhelm, *Hoppe-Seyler's Z. physiol. Chem.*, **63**, 215 (1909).

"The Relative Value of Certain Proteins and Protein Concentrates as Supplements to Corn Gluten," T. B. Osborne and L. B. Mendel with E. L. Ferry and A. J. Wakeman, *J. Biol. Chem.*, **29**, 69 (1917).

"The Use of Soybean as Food," T. B. Osborne and L. B. Mendel, *J. Biol. Chem.*, **32**, 369 (1917).

"The Antitrypsin of Egg White," A. K. Balls and T. L. Swenson, *J. Biol. Chem.*, **106**, 409 (1934).

"Feeding Experiments with Mixtures of Highly Purified Amino Acids. VIII. Isolation and Identification of a New Essential Amino Acid," R. H. McCoy, C. E. Meyer, and W. C. Rose, *J. Biol. Chem.*, **112**, 283 (1935).

"Factors Influencing Bacterial Deamination. I. The Deamination of Glycine, *dl*-Alanine, and *l*-Glutamic Acid by *Bacterium coli*," M. Stephenson and E. F. Gale, *Biochem. J.*, **31**, 1316 (1937).

"Amino Acid Metabolism. V. The Metabolism of *l*-Cystine and *dl*-Serine in the Normal Animal," J. S. Butts, H. Blunden, and M. S. Dunn, *J. Biol. Chem.*, **124**, 709 (1938).

"Protein Metabolism. VIII. The Activity of the α -Amino Group of Histidine in Animals," R. Schoenheimer, D. Rittenberg, and A. S. Keston, *J. Biol. Chem.*, **127**, 385 (1939).

"*p*-Bromophenol as the Intermediate in Synthesis of *p*-Bromophenylmercapturic Acid from Bromobenzene in the Rat," J. A. Stekol and H. H. Dash, *Proc. Soc. Exptl. Biol. Med.*, **40**, 261 (1939).

"The Production of Amines by Bacteria. I. The Decarboxylation of Amino Acids by Strains of *Bacterium coli*," E. F. Gale, *Biochem. J.*, **34**, 392 (1940).

"Soybean Protein as a Source of Amino Acids for the Chick," H. J. Almquist, E. Mecchi, F. H. Kratzer, and C. R. Grau, *J. Nutrition*, **24**, 385 (1942).

"The Genetic Control of Biochemical Reactions in *Neurospora*: A Mutant Strain Requiring Isoleucine and Valine," D. Bonner, E. L. Tatum, and G. W. Beadle, *Arch. Biochem.*, **3**, 71 (1943).

"The Relative Toxicity of *l*- and *dl*-Serine in Rats," C. Artom, W. H. Fishman, and R. P. Morehead, *Proc. Soc. Exptl. Biol. Med.*, **60**, 284 (1945).

"Some Correlations of Growth-Promoting Powers of Proteins with Their Strepogenin Content," D. W. Woolley, *J. Biol. Chem.*, **162**, 383 (1946).

"Biochemical Transformations as Determined by Competitive Analog Metabolite Growth Inhibitions. I. Some Transformations Involving Aspartic Acid," W. Shive and J. Macow, *J. Biol. Chem.*, **162**, 451 (1946).

"The Effects of Baking and Toasting on the Nutritional Value of Proteins," R. J. Block, R. P. Cannon, R. W. Wissler, C. H. Steffee, Jr., R. L. Straube, L. E. Frazier, and R. L. Woolridge, *Arch. Biochem.*, **10**, 295 (1947).

"The Biological Synthesis of Radioactive Adrenaline from Phenylalanine," S. Gurin and A. M. Delluva, *J. Biol. Chem.*, **170**, 545 (1947).

- "Experiments with Delayed Supplementation of Incomplete Amino Acid Mixtures," E. Geiger, *J. Nutrition*, **34**, 97 (1947).
- "Interrelations of Phenylalanine and Tyrosine in the Chick," C. R. Grau, *J. Biol. Chem.*, **170**, 661 (1947).
- "Biological Precursors of Uric Acid," J. H. Buchanan, J. C. Soune, and A. M. Delluva, *J. Biol. Chem.*, **173**, 81 (1948).
- "Intermediate Role of Carbamyl-L-glutamic Acid in Citrulline Synthesis," P. P. Cohen and S. Grisola, *J. Biol. Chem.*, **174**, 389 (1948).
- "Degradation of L-Lysine in Guinea-Pig Liver Homogenate: Formation of α -Aminoadipic Acid," H. Borsook, C. L. Deasy, A. J. Haagen-Smit, G. Keighley, and P. H. Lowy, *J. Biol. Chem.*, **176**, 1383 (1948).
- "The Chemical Nature and Formation of Auxin in the *Avena* Coleoptile," S. G. Wildman and J. Bonner, *Am. J. Botany*, **35**, 740 (1948).
- "Melanin and Its Precursors. I. Synthesis of 5, 5', 6, 6'-Tetrahydroxyindigo," J. Harley-Mason, *J. Chem. Soc.*, **1948**, 1244.
- "Homoserine as a Precursor of Threonine and Methionine in *Neurospora*," H. J. Teas, N. H. Horowitz, and M. Fling, *J. Biol. Chem.*, **172**, 651 (1948).
- "The Influence of Heat-Processing on the Functional and Nutritive Properties of Protein," D. Melnick and B. L. Oser, *Food Technol.*, **3**, 57 (1949).
- "Effect of Enzymic Hydrolysis on the Nutritive Value of Casein. III. Intravenous Feeding in Dogs," J. B. Allison, B. F. Chow, J. I. White, and J. S. Roth, *J. Biol. Chem.*, **180**, 383 (1949).
- "The Comparative Availabilities of Mixtures of the L and DL Modifications of the Essential Amino Acids for Growth in the Rat," J. F. Van Pilsum and C. P. Berg, *J. Biol. Chem.*, **183**, 279 (1950).
- "Use of Isotopic Nitrogen in a Study of the Conversion of 3-Hydroxyanthranilic Acid to Nicotinic Acid in *Neurospora*," E. Leifer, W. H. Laugham, J. F. Nyc, and H. K. Mitchell, *J. Biol. Chem.*, **184**, 589 (1950).
- "Studies of Nutritionally Deficient Bacterial Mutants Isolated by Means of Penicillin," Bernard D. Davis, *Experientia*, **6**, 41 (1950).
- "Studies on Natural and Racemic Amino Acids with Rats," C. E. Graham, S. W. Hier, H. K. Waitkoff, S. M. Saper, W. G. Bibler, E. I. Pentz, *J. Biol. Chem.*, **185**, 97 (1950).
- "Synthesis of Phenylalanine and Tyrosine in Yeast," C. Gilvarg and K. Block, *J. Am. Chem. Soc.*, **72**, 5791 (1950).
- "Synthesis of Aromatic Amino Acids by Mutant Strains of *Lactobacillus arabinosus*," D. E. Atkinson and S. W. Fox, *Arch. Biochem. and Biophys.*, **31**, 212 (1951).
- "Metabolism of Labeled 2-Carbon Acids in the Intact Rat," S. Weinhouse and B. Friedmann, *J. Biol. Chem.*, **191**, 707 (1951).
- "The Bacterial Oxidation of Tryptophan. III. Enzymatic Activities of Cell-Free Extracts from Bacteria Employing the Aromatic Pathway," O. Hayaishi and R. Y. Stanier, *J. Bacteriol.*, **62**, 691 (1951).
- "The Effects of D-Amino Acids on the Stereonaturalization of D-Methionine," K. A. J. Wretlind, *Acta physiol. Scand.*, **25**, 267 (1952).
- "Aromatic Biosynthesis. V. Antagonism Between Shikimic Acid and Its Precursor, 5-Dehydroshikimic Acid," B. D. Davis, *J. Bacteriol.*, **64**, 749 (1952).
- "Biosynthesis of Urea. VI. Enzymatic Cleavage of Argininosuccinic Acid to Arginine and Fumaric Acid," S. Ratner, W. P. Anslow, Jr., and B. Petrack, *J. Biol. Chem.*, **204**, 115 (1953).

"Conversion of Citrulline to Ornithine by Cell-Free Extracts of *Streptococcus lactis*," M. Korzenovsky and C. H. Werkman, *J. Biol. Chem.*, **46**, 174 (1953).

"The Conversion of Lysine to Pipecolic Acid in the Rat," M. Bothstein and L. L. Miller, *J. Biol. Chem.*, **211**, 851 (1954).

"A General Mechanism for Vitamin B₆-Catalyzed Reactions," D. E. Metzler, M. Ikawa, and E. E. Snell, *J. Am. Chem. Soc.*, **76**, 648 (1954).

"The Precursors of Sphingosine in Brain Tissue," D. B. Sprinson and A. Coulon, *J. Biol. Chem.*, **207**, 585 (1954).

"Carbamyl Phosphate, the Carbamyl Donor in Enzymatic Citrulline Synthesis," M. E. Jones, L. Spector, and F. Lipmann, *J. Am. Chem. Soc.*, **77**, 819 (1955).

"Quantitative Determination of Terminal Methionine, Leucine, and Lysine in Raw and Toasted Soybean Oil Meal," S. W. Fox, C. Warner, and T. L. Hurst, *J. Agr. Food Chem.*, **3**, 704 (1955).

"The Effect of Methionine Supplementation upon the Tumor-Host Relationship in the Rat," J. B. Allison, R. W. Wannemacher, Jr., R. Helf, and C. A. Hetzel, *J. Nutrition*, **59**, 27 (1956).

"Tyrosinase-Produced Quinones and the Disappearance of Kynurenone in Larval Extracts of *Drosophila melanogaster*," E. Glassman, *Arch. Biochem. and Biophys.*, **67**, 74 (1957).

Naturally Occurring Peptides and Their Fractionation

Fractionation of peptides

Anserine and carnosine

Folic acids

Glutathione

Hypertensin (angiotonin)

Peptide antibiotics

Secretin

Other peptides

Occurrence of peptides in families

A number of biologically active peptides have been found in nature, as well as some substances closely related in structure to peptides. The number of such peptides is not remarkably large, but their significance seems to be considerable.

The first problem in the study of peptides is their isolation and fractionation to yield samples of essentially homogeneous molecules. The methods available for fractionation of peptides are, to a considerable degree, the methods available for fractionation of proteins. In this chapter are considered those methods which are primarily useful for fractionation of peptides, others being reserved for discussion in Chap. 15. The first group consists of the following techniques: countercurrent distribution, analysis on ion-exchange columns, paper chromatography, and selective precipitation.

It is first desirable to have a clear idea of the distinction between a peptide and a protein. It is of course easy to differentiate an amino

acid and a peptide, but the dividing line between a peptide and protein is highly arbitrary. Unfortunately, the molecular weight and size that permit biological activity are not adequate criteria for such distinctions. The most used arbitrary specification has perhaps been a molecular weight of 10,000. Another property which has been employed is diffusibility through cellophane. This criterion has provided a lower limit in size of about 10,000. Similar standards are precipitability by trichloroacetic acid and ammonium sulfate.

The smallest peptide is a dipeptide, which contains but a single peptide link. The prefix designating the size of the peptide is selected as equivalent to the number of amino acid residues, which is always one more than the number of peptide links. For example, $\text{H}_2\text{N}\cdot\text{CHR}\cdot\text{CO}\cdot\text{NH}\cdot\text{CHR}'\cdot\text{CO}\cdot\text{NH}\cdot\text{CHR}''\cdot\text{COOH}$ is a tripeptide, containing two peptide linkages but three amino acid residues. Occasionally peptides of a small number of residues are referred to as oligopeptides.

Aside from the discrete peptides found in nature, which will be discussed later, other peptides occur as intermediates in various processes. A number of these peptides have been isolated from such sources as intestinal contents, but this type of work has not furnished much critical information. Much more has been learned in this field by Fischer, Bergmann, Abderhalden, Neurath, Fruton, Smith, and others, by subjecting a variety of peptides to the action of proteolytic enzymes. The greatest variety of peptides was made available by the Bergmann-Zervas synthesis. The implications of the conclusions from study of the action of digestive enzymes on synthetic substrates lead into broader fields involving hydrolysis, synthesis, and rearrangement of peptide bonds.

Strictly speaking, peptides are composed of amino acid residues only. A number of naturally occurring compounds contain groups besides amino acid residues, but also contain one or more peptide linkages ("pseudopeptides"). Residues bearing α -amino groups are N-termini; those at the carboxyl end are C-termini.

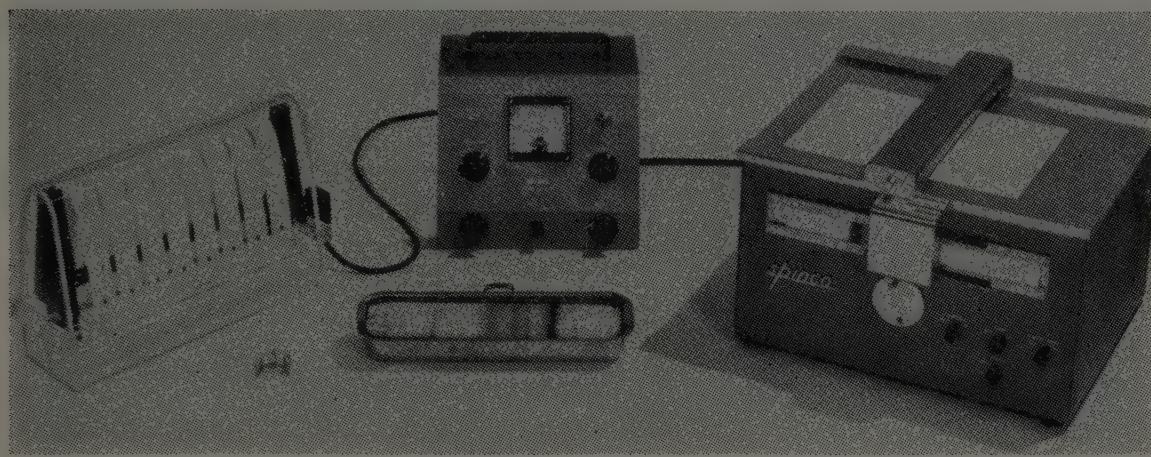
FRACTIONATION OF PEPTIDES

Selective Precipitation. Some of the reagents that are useful for precipitation of amino acids have been employed with peptides composed of those amino acids. This kind of technique received more attention in early work, but has been rather thoroughly supplanted by newer methods.

Abderhalden and Bahn showed that glycyltyrosine was precipitable by mercuric salts under conditions in which tyrosylglycine was not so

precipitated. This degree of selectivity is great enough that studies of related effects would be of much interest. The prognosis for such research projects seems good in view of the number of known selective precipitants for amino acids.

In the course of elucidating the structure of clupein, Felix and co-workers made much use of the selective precipitability of arginine-containing peptides obtained in partial hydrolysis of clupein. For this purpose flavianic acid was used. Flavianic acid has previously been described as a selective precipitant for the amino acid arginine.



Electrophoresis cell

Power supply

Recording scanner

Fig. 8-1. Apparatus for microfractionation of peptides on paper.

Paper Chromatography. Paper chromatography has been successfully used for the separation of micro quantities of peptides. This use has been largely restricted to fractionation of hydrolyzates. A fair degree of skill is required in this application (Fig. 8-1). This method has probably not received as much attention as it appears to merit.

Analysis by Ion Exchange. The separation of peptides on ion-exchange columns is, in principle, not different from the corresponding analysis of amino acids (p. 91). This technique has been successfully employed in fractionation of corticotropins and other peptides.

Countercurrent Distribution. A highly useful technique is that of countercurrent distribution, as developed by Craig (apparatus in Fig. 8-2). To explain this process a simplification of the actual situation may be imagined, along the lines of Fig. 8-3.

Consider that two solutes are extracted in a two phase liquid-liquid system, that each solute behaves as if the other were not present (often true), and that there are constant partition ratios over the

concentrations employed. Assume also that there are equal volumes of the two solvents, and that the partition ratio of the solute first to be considered is 1.00. One gram (1.00 gm.) of solute is dissolved in L_0 , U_0 is moved over it, and the system is equilibrated by shaking the apparatus. The phases separate, and the upper phase is shifted so that U_0 is over L_1 and U_1 over L_0 . This is the first transfer. The distribution is then as given in the row labeled Transfer No. 1. When the entire process is repeated, it may now be seen that the solute is

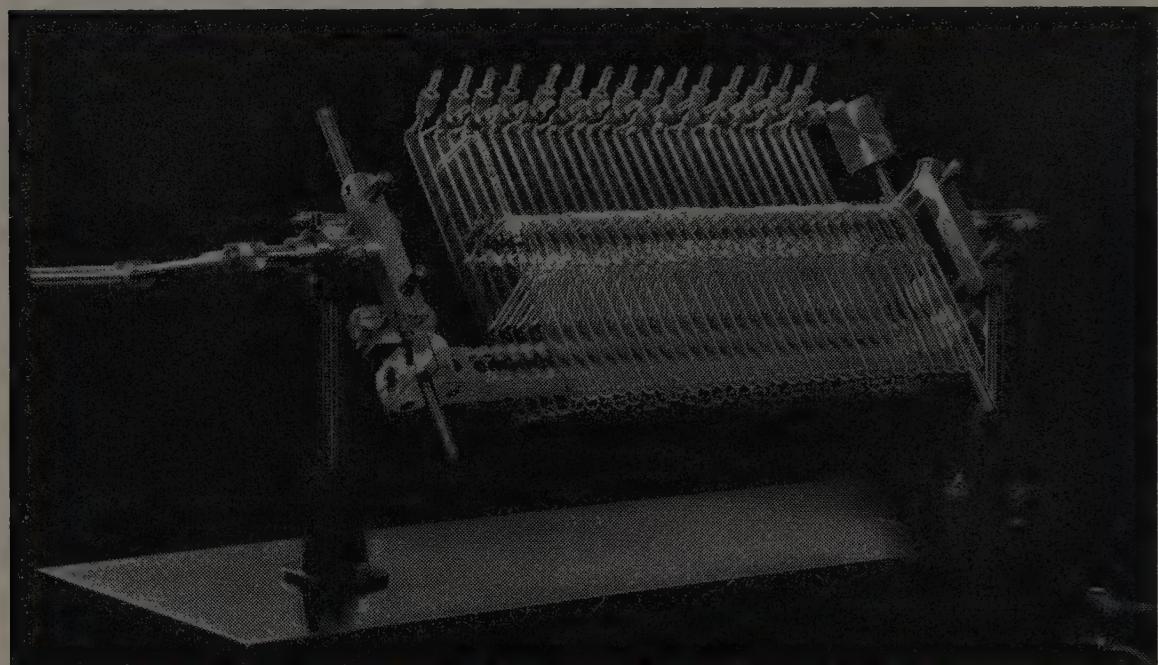


Fig. 8-2. Countercurrent distribution apparatus of Craig. Courtesy of Dr. L. C. Craig.

present in highest concentration in tube No. 1. After four transfers, the peak of concentration is in tube No. 2, as depicted in Fig. 8-3. Theoretical and experimental distribution patterns obtained by this technique for insulin are presented in Fig. 8-4.

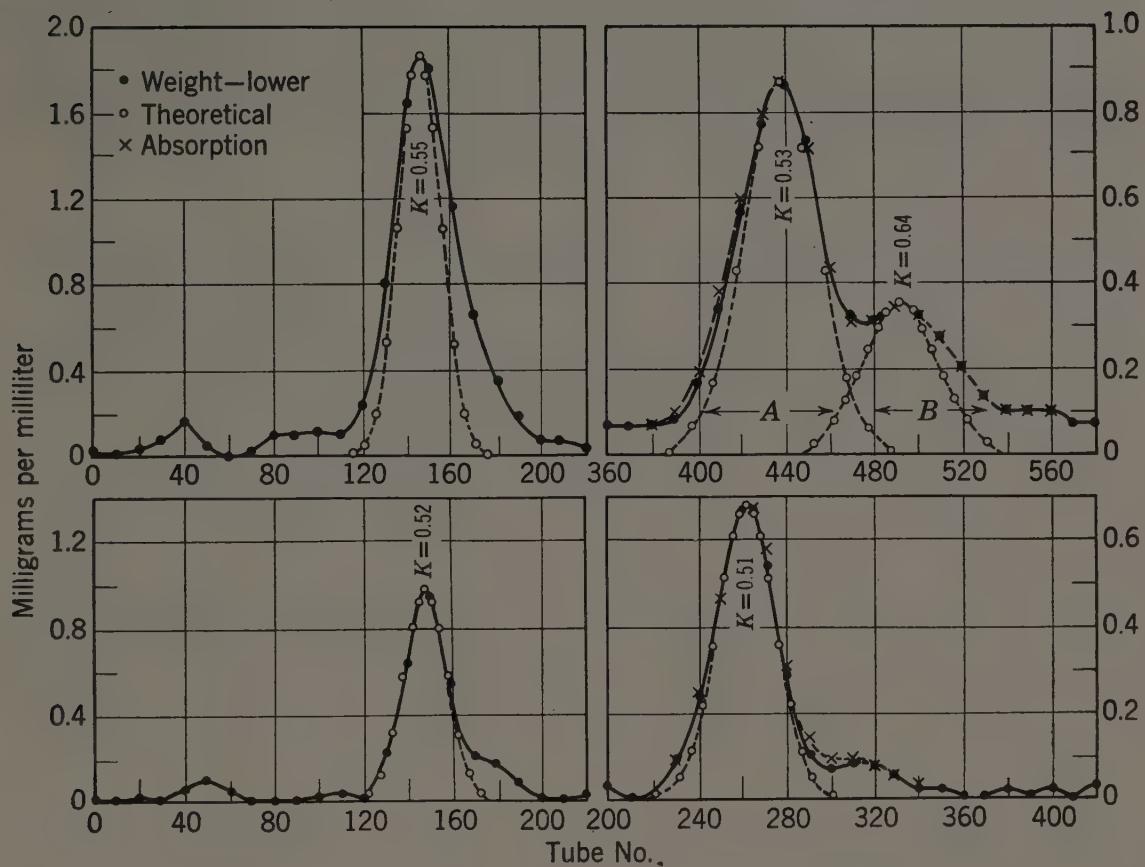
In practice a second solute or later solutes will have different partition ratios, and the tubes with the highest concentration will be other tubes in the series than those for the first solute. The process can be brought closer to perfect separation by rerunning critical tubes, by increasing the number of tubes, and by shifting to other solvent systems than the one employed in the initial distribution.

ANSERINE AND CARNOSINE

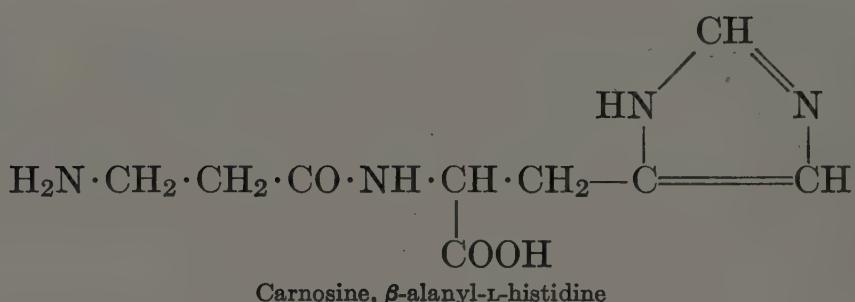
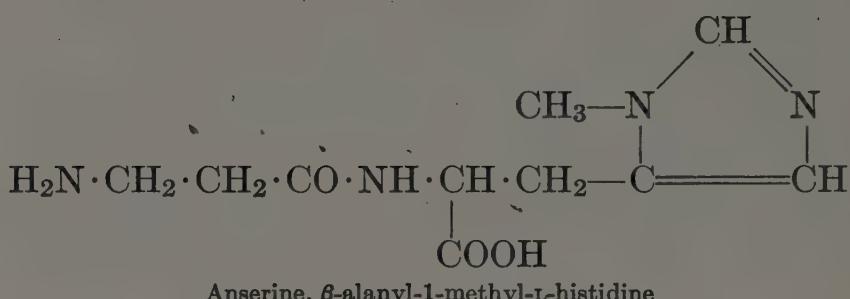
These two substances are extractable from muscle tissue of animals, birds, and fish. Their biochemical origin and function are not clearly

U_r	---				U_3	U_2	U_1	U_0	---				L_0	L_1	L_2	L_3	L_r	
Transfer No.	Tube No.	0	1	2	3	---	r											
0	0	1.0																
1	0.50		0.50															
2	0.25		0.50	0.25														
3	0.125		0.375	0.375	0.125													
4	0.0625		0.25	0.375	0.25	0.0625												
n																		

Fig. 8-3. Scheme for countercurrent distribution.

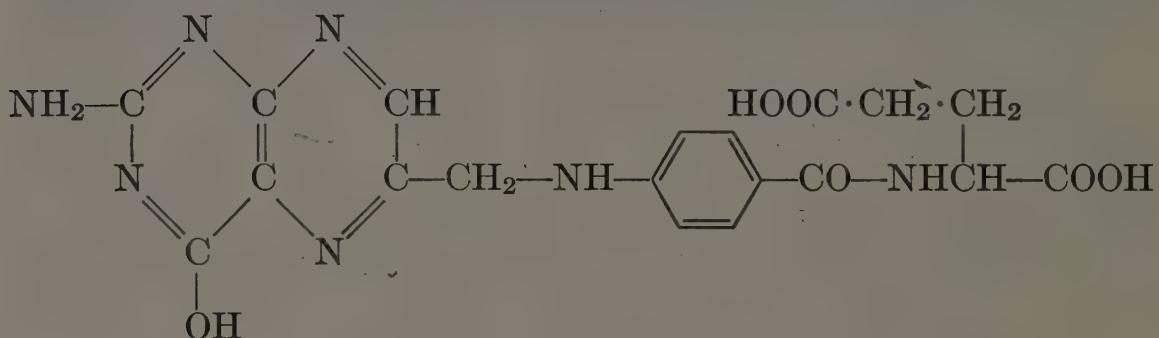
Fig. 8-4. Theoretical and observed distribution patterns of Lilly insulin. From E. J. Harfenist and L. C. Craig, *J. Am. Chem. Soc.*, 74, 3083 (1952).

understood. Both are derivatives of histidine.



FOLIC ACIDS

Members of the folic acid family of vitamins are derivatives of L-glutamic acid. The folic acids function as growth factors, and to a limited extent as antianemia factors. The progenitor of the family is pteroylglutamic acid, the *L. casei* factor:

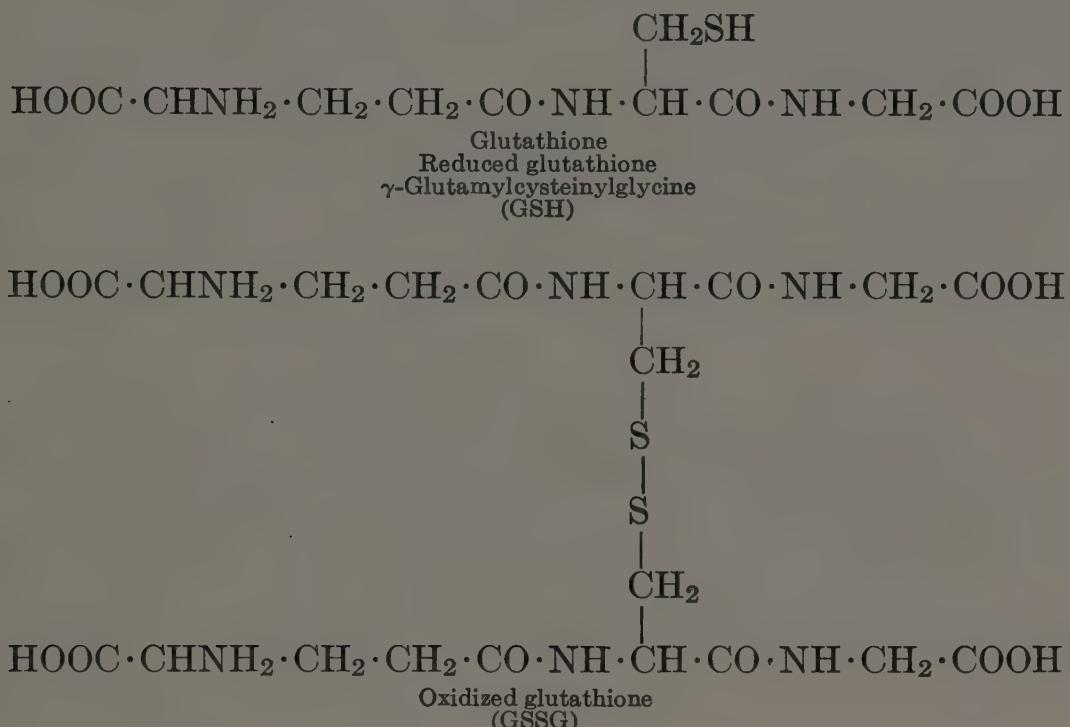


Another member of this family is pteroyldiglutamylglutamic acid, also known as *teropterin* and as fermentation *L. casei* factor. It contains two glutamic acid residues coupled at the γ - rather than the α -positions.

GLUTATHIONE

Glutathione, like folic acid and other vitamins, is found in quantity in yeast. It occurs in mammalian blood. In reduced form glutathione is found especially in tissue in which cells are dividing with relative

rapidity. The importance to the mammal is believed to be related to its functions in oxidative metabolism and in detoxication. The latter is suggested by the fact that glutathione is a tripeptide that yields on hydrolysis three amino acids which are themselves effective detoxifying agents of diverse types. Evidence has been offered that glutathione functions to mediate protein synthesis.



The linkage between the glutamic acid and cysteine residues is atypical for peptides of the protein type, since the amino group of the glutamic acid residue is not coupled. Glutathione was the first compound in which this abnormal peptide linkage was found.

HYPERTENSIN (ANGIOTONIN)

A substance which has properties that appear to be the cause of hypertension is the peptide *hypertensin* or *angiotonin*. Various peptide compositional types have been proposed for hypertensin, and preliminary evidence indicates that this material may also prove to be a family of peptides, although probably in a manner different from that of the folic acid group.

PEPTIDE ANTIBIOTICS

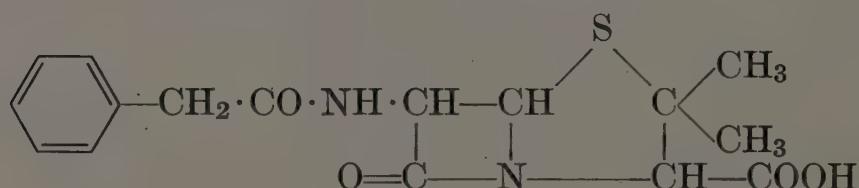
A number of antibiotics are peptides; others are peptide derivatives. *Gramicidin* and *tyrocidine* are polypeptides and represent two of the earliest crystallized antibiotics. Each of these seems to be a cyclo-

peptide, a type of structure which appears to be common to other peptide antibiotics.

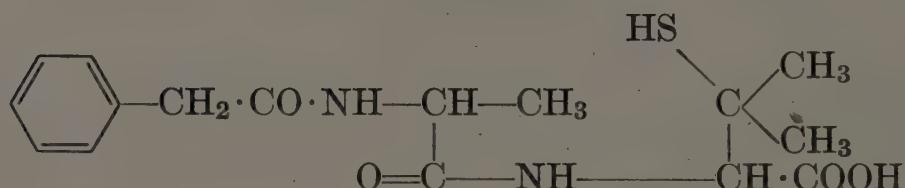
Gramicidin contains 22 amino acid residues in combination with two molecules of ethanolamine. Almost half of the amino acid residues are of the D-configuration. These are all D-valine and D-leucine. Gramicidin has been shown by Craig to exist in families, gramicidin A, gramicidin B, and so on. These differ in activity and also, to a minor extent, in composition. The amino acid residue sequences are necessarily also different.

Tyrocidine has also two unusual components, L-ornithine and D-phenylalanine. Tyrothricin is a mixture of tyrocidine and gramicidin. The structures of tyrocidine A and B have been elucidated.

Penicillin is one of the antibiotic peptide derivatives. Benzylpenicillin, the best known of the penicillin family, has the structure



This can be seen to be a closed-ring derivative of



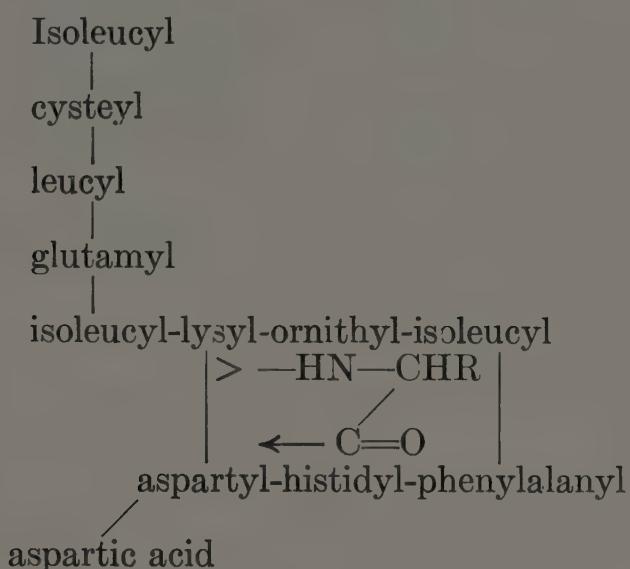
or phenacetyl-L-alanyl-D-dimethylcysteine. Penicillin is thus also a D-amino acid derivative. In other members of the family of naturally occurring penicillins, other groups are found in place of the benzyl residue.

Gramicidin S represents not only an antibiotic but the first peptide of its size for which the structure was completely elucidated. In common with tyrocidine this antibiotic contains two exotic components, L-ornithine and D-phenylalanine. Like penicillin, gramicidin S is cyclic, being a cyclodecapeptide.

Bacitracin, an antibiotic isolated from a *Bacillus subtilis* culture from the healing leg wound of a girl in New York City named Mary Tracy (from which fact the antibiotic gets its name), is a peptide containing D-amino acids. The molecular weight is less than 2000.

Bacitracin has been shown by Craig to be resolvable into a family of

molecules. He and his co-workers have proposed the following structure for bacitracin A:



The phenylalanine is known to be of the D-configuration.

Polymixins. Another family of peptide antibiotics is represented by the polymixins, of which a considerable number have been investigated. Among the amino acids isolated after hydrolysis are D-leucine, L-phenylalanine, L-serine, L-threonine, and the unique amino acid α , γ -diaminobutyric acid, $H_2N\cdot CH_2\cdot CH_2\cdot CHNH_2\cdot COOH$. It has been reported that this diaminobutyric acid occurs in the D-form as well as in the L-isomer, in at least one of the polymixins.

A large proportion of the medically unusable antibiotics consist of peptides or peptide derivatives. Only some of the more intensively studied types have been described. Others for which evidence of peptide constitution exists are: subtilins, elemycin, bacillin, and lichenformin. Aspergillic acid can be considered to be a diketopiperazine derivative.

SECRETIN

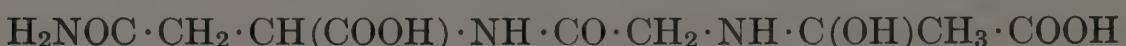
Secretin is a pancreatic hormone of molecular weight about 5000. Secretin stimulates the secretion of the pancreatic fluid. It has been found to contain glycine, alanine, tyrosine, histidine, serine, methionine, cystine, lysine, arginine, aspartic acid, glutamic acid, proline, valine, phenylalanine, and leucine or isoleucine. Workers in this field have referred to secretin as a protein.*

* When one notes that both secretin and protamines of approximately the same molecular weight are referred to as proteins and also as peptides, it is clear that no universally accepted size criterion exists for protein molecules, as pointed out at the beginning of this chapter.

OTHER PEPTIDES

The causative agents of allergies from such sources as ragweed and timothy have been shown to be polypeptides of molecular weight about 5000. The relatively small size of the molecules has been invoked as explaining the ease of entrance of these irritants through animal membranes.

Lycomarasmin is a peptide which produces wilt in tomatoes. Hydrolysis yields glycine, aspartic acid, and pyruvic acid. The structure is believed to be



A peptide of not entirely definite constitution but marked activity is strepogenin. Strepogenin aids the utilization of amino acids in mammalian and bacterial protein synthesis. Strepogenin has the properties of a peptide and is extractable by alcohol from some proteins, following brief tryptic digestion. Crystalline insulin and trypsinogen are among the richest sources. The molecular weight is probably in the range of 300 to 500. Its structure is probably serylhistidylleucylvalylglutamic acid (Merrifield and Woolley, 1955). Strepogenin has the interesting faculty of entering into competitive inhibition with lycomarasmin.

Peptides of indefinite constitution are believed to be responsible for the physiological response of inflammation. In this they may not be unrelated to the allergenic peptides.

Associated with pepsin and trypsin are found specific inhibitors (p. 391). The molecular weights of these peptides have been reported as being in the 5000 range. Presumably these inhibitors aid in the natural regulation of digestion. This group of peptides has been masterfully characterized by Northrop, Kunitz, and Herriott (1948).

Finally, a variety of peptides have been isolated from digestive media. A larger variety have been synthesized chemically for enzymic studies. It seems certain that many peptides must occur as intermediates in hydrolysis of proteins, and perhaps also in protein synthesis.

OCCURRENCE OF PEPTIDES IN FAMILIES

A large amount of evidence has been accumulated to indicate that natural peptides tend to occur in families of closely related molecules. This is believed by some workers to be true also for protein molecules, but the evidence is more conclusive for peptides. In the case of pro-

TABLE 8-1. Common Polypeptide Antibiotics Which Occur in Families

Polymixins	A,B,C,D,E
Licheniformins	A,B,C
Actinomycins	A,B,C
Tyrocidines	A,B
Gramicidins	A,B,C,D?
Penicillins ¹	Numerous

¹ Penicillin is a cyclic derivative of acyl-L-alanyl-D-penicillamine. The variation in this case is only in the nature of the acyl group.

TABLE 8-2. Composition of Each of Two Gramicidins

Residue	Gramicidin A	Gramicidin B
Alanine	9	10
Glycine	5	5
Leucine	20	21
Phenylalanine	0	5
Tryptophan	20	18
Valine	17	17
Ethanolamine	4	4

teins, it is difficult to rule out the possibility that the method of purification has itself led to some diversification in molecular type. Such an explanation is unacceptable for peptides in which the differences can be traced to structural variations that cannot be explained

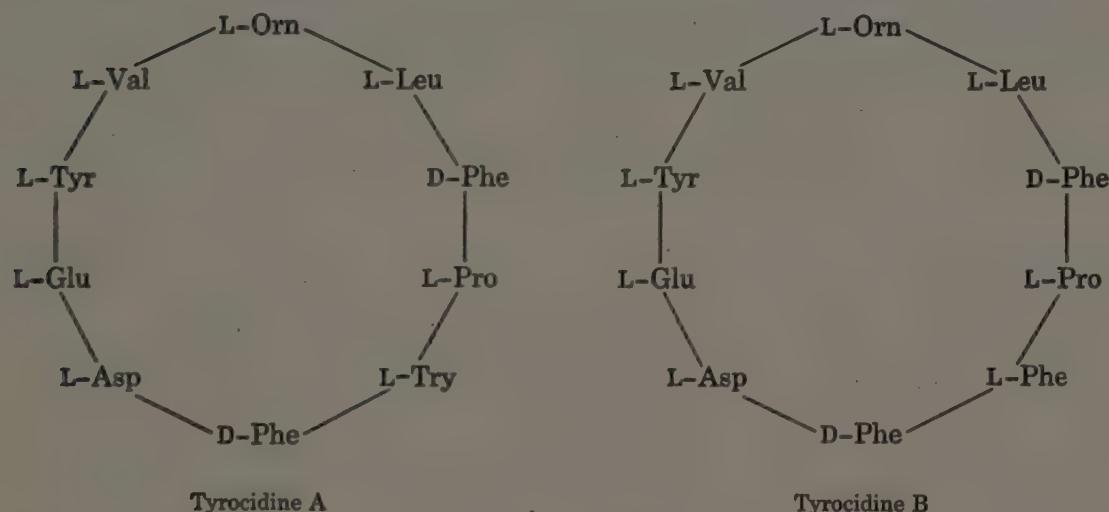


Fig. 8-5. Structures of tyrocidines A and B, revealing a difference in one residue only (L-tryptophan versus L-phenylalanine).

on these grounds (see especially oxytocin and vasopressin). The polypeptide antibiotics that have received most intensive study are predominantly recognized as occurring in family relationships. This phenomenon can be explained as an accumulation of biosynthetic

mistakes, or more fundamentally as an essential step in Darwinian evolution at the molecular level (Chap. 24).

The existence of families of peptide molecules is, in any event, now a widely recognized phenomenon. Common polypeptide antibiotics that occur in families are listed in Table 8-1. Differences in composition between two of the sibling gramicidin molecules is illustrated in Table 8-2. Tyrocidines A and B display differences in structure, as depicted in Fig. 8-5.

SELECTED REFERENCES

Reviews

"Carnosine and Anserine," V. du Vigneaud and O. K. Behrens, *Ergeb. Physiol. biol. Chem. u. exptl. Pharmakol.*, **41**, 917 (1939).

"Pepsin Inhibitor" (Chap. IV) and "Trypsinogen, Trypsin, and Trypsin Inhibitor" (Chap. VI), J. H. Northrop, M. Kunitz, and R. M. Herriott, *Crystalline Enzymes*, Columbia University Press, New York, 1948.

"Antibiotics Derived from *Bacillus polymyxa*," P. H. Long *et al.*, *Ann. N. Y. Acad. Sci.*, **51**, 853 (1949).

"Extraction and Distribution," L. C. Craig and D. Craig in Weissberger's *Technique of Organic Chemistry*, Vol. III, p. 171, Interscience, New York, 1950.

"Naturally Occurring Peptides," E. Bricas and C. Fromageot, *Advances in Protein Chem.*, **8**, 1 (1953).

"Application of Ion-exchange Resins to Organic Analysis," B. S. Miller and J. A. Johnson, *Trans. Am. Assoc. Cereal Chemists*, **12**, 29 (1954).

"Column Chromatography of Peptides and Proteins," S. Moore and W. H. Stein, *Advances in Protein Chem.*, **11**, 191 (1956).

"Countercurrent Distribution in Protein Chemistry," P. Von Tavel and R. Signer, *Advances in Protein Chem.*, **11**, 237 (1956).

Original Articles

"A Preliminary Note on the Nitrogen Metabolism of Seaweeds. Glutamic Acid Peptide," P. Haas and T. G. Hill, *Biochem. J.*, **25**, 1472 (1931).

"The Ergot Alkaloids. V. The Hydrolysis of Ergotinine," W. A. Jacobs and L. C. Craig, *J. Biol. Chem.*, **110**, 521 (1935).

"Über die Giftstoffe des Knollenblätterpilzes. V. Zur Konstitution des Phallobiodins," H. Wieland and B. Witkop, *Ann. Chem. Liebigs*, **543**, 171 (1940).

"Studies on the Biochemistry of Micro-organisms. 71. Fumaryl-*dl*-alanine (Fumaromono-*dl*-alanide), a Metabolic Product of *Penicillium resticulosum* sp. nov.," J. H. Birkinshaw, H. Raistrick, and G. Smith, *Biochem. J.*, **36**, 829 (1942).

"Analysis of Hay Fever-Producing Component of Ragweed Pollen Extract," H. A. Abramson, D. H. Moore, and H. H. Gettner, *J. Phys. Chem.*, **46**, 192 (1942).

"The Synthesis and the Enzymatic Degradation of L-Tyrosyl-L-Lysyl-L-Glutamyl-L-Tyrosine," A. A. Plentl, I. H. Page, and F. R. van Abeele, *J. Biol. Chem.*, **163**, 49 (1946).

"Ionophoresis in Silica Jelly. A Method for the Separation of Amino Acids

and Peptides," R. Consden, A. H. Gordon, and A. J. P. Martin, *Biochem. J.*, **40**, 33 (1946).

"The Capsular Substance of *Bacillus anthracis*," W. E. Hanby and H. N. Rydon, *Biochem. J.*, **40**, 297 (1946).

"Identification of Lower Peptides in Complex Mixtures," R. Consden, A. H. Gordon, and A. J. P. Martin, *Biochem. J.*, **41**, 590 (1947).

"Pteroic Acid Derivatives. IV. Pteroyl- α - γ -glutamylglutamic Acid," J. H. Mowat, A. L. Gazzola, B. L. Hutchings, J. H. Boothe, C. W. Waller, R. B. Angier, J. Semb, and Y. Subba Row, *J. Am. Chem. Soc.*, **71**, 2308 (1949).

"Purity Studies on Polypeptide Antibiotics," L. C. Craig, J. D. Gregory, and G. T. Barry, *J. Clin. Invest.*, **28**, 1014 (1949).

"A Microelectrophoretic and Microonophoretic Technique," E. L. Durrum, *J. Am. Chem. Soc.*, **72**, 2943 (1950).

"Calculations on Paper Chromatography of Peptides," A. B. Pardee, *J. Biol. Chem.*, **190**, 757 (1951).

"The Amino Acid Composition of Subtilin," J. C. Lewis and N. S. Snell, *J. Am. Chem. Soc.*, **73**, 4812 (1951).

"Uridine-5'-pyrophosphate Derivatives. III. Amino Acid-containing Derivatives," J. T. Park, *J. Biol. Chem.*, **194**, 897 (1952).

"Isolation and Determination of Structure of a Peptide with Streptokinase Activity," R. B. Merrifield and D. W. Woolley, *Arch. Biochem. and Biophys.*, **56**, 265 (1955).

"The Chemistry of Tyrocidine. V. The Amino Acid Sequence of Tyrocidine B," T. P. King and L. C. Craig, *J. Am. Chem. Soc.*, **77**, 6627 (1955).

"The Identical Behaviour of Some Peptides and Amino-acids in Paper Chromatography," K. R. Deane and E. V. Truter, *Biochim. et Biophys. Acta*, **18**, 435 (1955).

"Isolation of Cephalosporin C, a Penicillin-like Antibiotic Containing D- α -Aminoadipic Acid," G. G. F. Newton and E. P. Abraham, *Biochem. J.*, **62**, 651 (1956).

9

Determination of Structure of Peptides

History

Enzymic methods

Physicochemical methods

Amino condensations

Carboxyl condensations

Stepwise degradations

Other procedures

Examples of sequence determination

Application to large peptides and proteins

Quantitative analysis of residue sequence

Conclusions

HISTORY

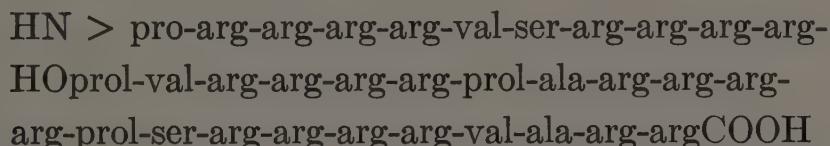
For determination of the structure of peptides, many methods have been available for decades. A widespread feeling that attempts to elucidate the complete structure of a protein were futile was replaced by optimism in the late forties, even though the possibilities had long been discernible in certain special work. This shift in assessment of probabilities was due to an accelerated production of methodology and applications thereof, and especially to proposals by Sanger and associates for complete sequences in insulin. One review of the subject (Fox, 1945) listed approximately two dozen chemically discrete methods for assignment of residue sequence.

The essential principle of a working method which has been most popular in recent years is that of Barger. This relied on the propensity

of trinitrotoluene to condense with the amino end of peptides, with the elimination of one of the nitro groups. Barger and Tutin used this method successfully in 1918 to elucidate the order of the β -alanine and histidine components of a peptide found in muscle extracts, carnosine. The same approach was later employed to confirm the structure of glutathione, as previously correctly suggested by Pirie and Pinhey in 1929 from titration data.

Abderhalden, who contributed tremendously to much of the information on peptides, proposed the arrangement of nine residues in silk fibroin in 1933. His approach, which involved reconstruction from fragments, pointed the way for so much later work that the details of his assignment of sequence in one tetrapeptide should be considered here. Individual amino acids were frequently identified by Abderhalden by N analysis and by melting points of the benzoyl derivatives. With an associate, Bahn, a tetrapeptide from silk was found to contain 2 tyrosine, 1 serine, and 1 proline. When this peptide was benzoylated and the benzoylpeptide was treated with the enzyme trypsin, it was possible to recover tyrosine in characteristic crystals plus the benzoylated tyrosylserylproline. On alkaline hydrolysis this latter yielded benzoyltyrosine (recognizable by melting point and analysis) and also a dipeptide composed of serine and proline. The order in this last dipeptide was assigned by the $\text{NH}_2 : \text{N}$ ratio, inasmuch as prolylserine would have yielded no N_2 with nitrous acid. The principles employed by Abderhalden have, with modification, been widely used since.

In 1937, Felix offered a complete structure for the protamine clupein, found in the spermatic tissue of herring. This structure



was based on much detailed work on conditions of partial hydrolysis of the purified substance and isolation of small amounts of the fragments obtained. A triarginylarginine unit was found. Some of the individual sequences in which arginine did not occur were worked out by Felix or others. Recent work places this structural proposal in greater doubt, although some of the sequences are undoubtedly correct.

The structure of a much smaller molecule, gramicidin S, was announced by Syngle in 1947 (Consden *et al.*, 1947). Despite the small size of this molecule, the work stands as a model for its thoroughness and the clarity of technique (p. 154).

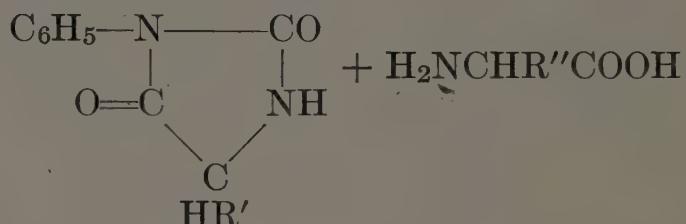
In a highly fruitful modification of the fragmentation technique, Sanger (1952), for example, treated DNP-phe-val-asp-glu (DNP is

dinitrophenyl) so that he was able to isolate the following:

DNP-phe-val-asp-glu
DNP-phe-val-asp
DNP-phe-val
DNP-phe

From these products the original sequence of the tetrapeptide could be reconstructed. This method has the advantage that the derivatized peptide in each case has properties quite different from those of the underivatized fragments, and isolation is thus aided. Among the disadvantages are the possibility of too rapid hydrolysis of some bonds (Table 5-1), rearrangement of structure, the known fact that such a group as dinitrophenyl is not completely stable during hydrolysis, and the need for fractionation. It has been possible, however, in a qualitative way to deduce long sequences from orders in overlapping fragments, and this method has been most used.

A method which permits stepwise removal of terminal residues offers some advantages. Of the few such methods that have been devised, one stands out in historical practice. In 1927, Bergmann and co-workers learned that phenylcarbamyl peptides released phenylhydantoin readily:



Abderhalden and Brockmann later (1930) showed that terminal amino groups could be successively removed by this reagent by refluxing in methanolic hydrochloric acid. The splitting of nonterminal bonds was small but appreciable. Edman further improved the method by replacing phenylisocyanate by phenylisothiocyanate, and by using nitromethane as a solvent for the fission. Later, dioxane was found to be more widely useful. Dry hydrogen chloride could also be used. It was learned that this method could be applied quantitatively as well as in a stepwise manner, and that fractionation could be minimized or eliminated.

Since methods such as the last can be employed to a considerable degree without need for fractionation of fragments, the fractionation of peptides as an adjunct to sequence assignment in larger molecules

is less urgent than formerly. The topic of fractionation has been treated in part in the preceding chapter, and is discussed more fully in Chap. 15.

Determination of terminal amino acid residues is also useful for other purposes, such as characterization of proteins and discovery of particular structural features in protein molecules. From investigation of this kind has come the knowledge that terminal amino acid residues bearing the free α -amino groups in proteins may be acidic, neutral, or basic. Terminal residue analyses and sequential analyses are furnishing new understanding of the relationships between proteins. When complete, this knowledge may supplant partially or entirely the present unsatisfactory system of classification of protein.

ENZYMIC METHODS

The peptidases of known specificity, such as carboxypeptidase, are of some value in this field of work and have been applied to the partial elucidation of the order in protamines and in ACTH. Theoretically, peptidases which attack only terminal residues have the necessary specificity, but the rate of scission is a function of the molecular environment, a fact which complicates the use of enzymes. Other objections, such as uncertain knowledge of enzyme specificity, have hampered use of this technique. A further limitation is posed in the increasingly recognized ability of proteolytic enzymes to catalyze transfer of peptide fragments. Carboxypeptidase, however, has been used with success in the controlled identification of C-terminal residues.

PHYSICOCHEMICAL METHODS

Titration. It is possible in some cases to determine by titration (p. 36) the sequence in small peptides of amino acid residues containing ionizable groups. Neighboring structures in the peptide molecule affect the ionization of individual groups in recognizable ways. This method is not so readily applicable to the elucidation of structure of many peptides of the monoaminomonocarboxylic acids, but is of value where the side chains contain dissociable groups.

Racemization. The amino acid residues within a peptide chain are found to be racemized more readily than those at the ends. Hydrolysis of a peptide treated in this way would be expected to yield optically active amino acids corresponding to the terminal residues; the method has been so used on synthetic peptides and on proteins.

Equilibrium Constant of Amino Group Reaction. The equilibrium constant of the reaction of such aldehydes as glucose with the free

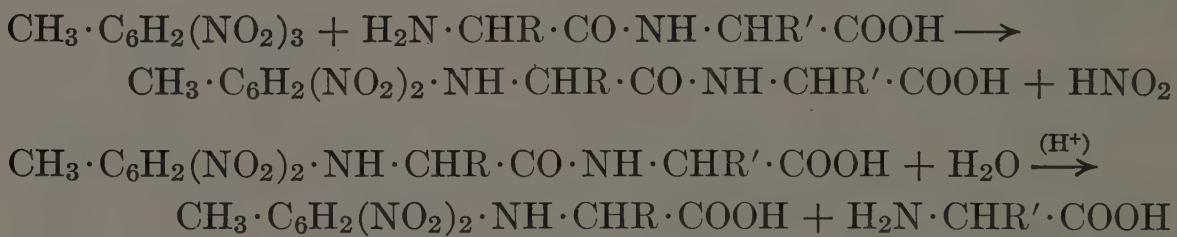
amino group of a peptide is in some cases characteristic of the type of amino acid residue at the free amino end. This property has received limited study. Difficulties arise from the propensity of glucose to initiate further decomposition.

Chromatographic Mobility (R_f value). Another characteristic physical property of utility is the R_f value of the peptide in a chromatogram. This can be determined by comparing the position or rate of travel of the unknown with those specifications of authentic synthetic peptides. A main difficulty here is the frequent unavailability of the reference peptide. If standard peptides are available, other physical properties, such as melting point, mixed melting point, titration behavior, and solubility behavior, may also be employed to establish the identity of the substance.

AMINO CONDENSATIONS

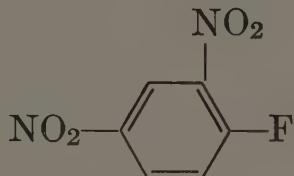
Reactions of the amino or carboxyl ends of peptides may be utilized to establish the amino acid residues at the ends. A number of appropriate reagents, especially for the amino terminus, have been used.

Trinitrotoluene. One of these reagents is trinitrotoluene, which on heating with the peptide forms a relatively nonhydrolyzable condensation product. The peptide bonds can then be broken and the resultant yellow dinitrotolyl peptide separated readily, because its solubility properties differ from those of the unsubstituted amino acids obtained by hydrolysis:



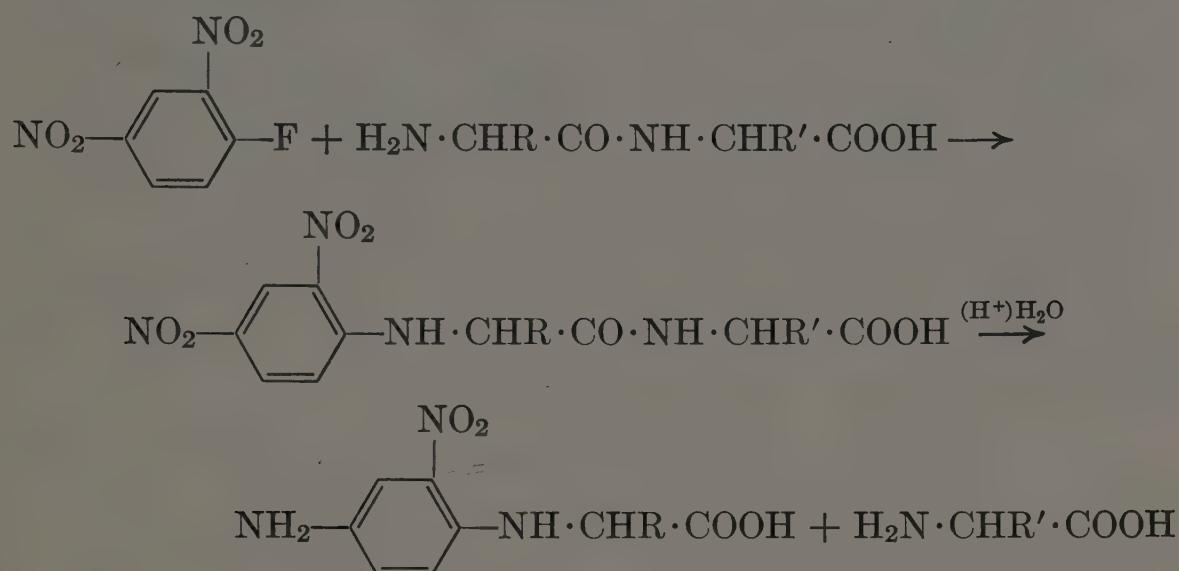
The trinitrotoluene can be any of the isomers that contain two nitro groups *ortho* to each other. This procedure was successfully used to assign the structures of glutathione, anserine, and carnosine.

Dinitrofluorobenzene. Dinitrofluorobenzene (DNFB) is a somewhat similar reagent,



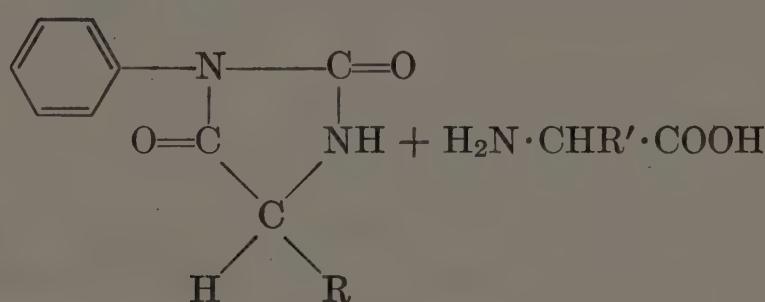
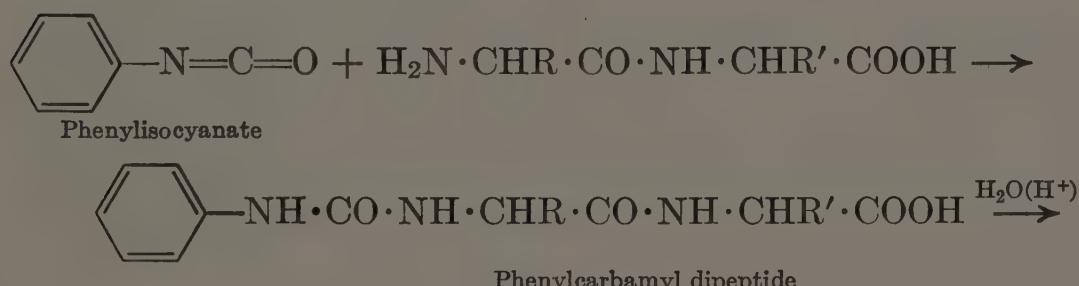
which also forms a dinitrophenyl derivative. Owing to the lability

of the fluorine atom, coupling occurs readily; the substituent is only slowly removed by agents which hydrolyze the peptide bond. This rate varies with the residue, however, being extremely rapid with DNP-proline and quite rapid with residues such as DNP-glycine. Dinitrochlorobenzene, introduced earlier by Abderhalden, does not react as rapidly as the fluoro analog.



With this reagent and successive fragmentation of reacted peptides, Sanger has provided much knowledge of the structure of insulins from various species.

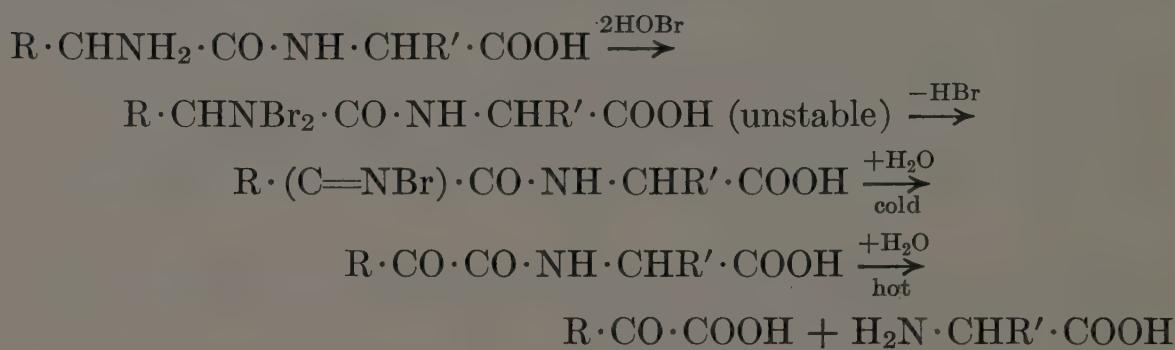
Aryl Isocyanate. Phenylisocyanate or naphthylisocyanate may be employed to give derivatives, at the amino group, which are subsequently converted to hydantoins in high yield by hydrolytic treatment with hydrochloric acid. The hydantoins are relatively stable to hot acid, and can be subsequently separated from the hydrolysate and identified by melting point.



This method has been particularly applied to identification of terminal amino acid residues in proteins.

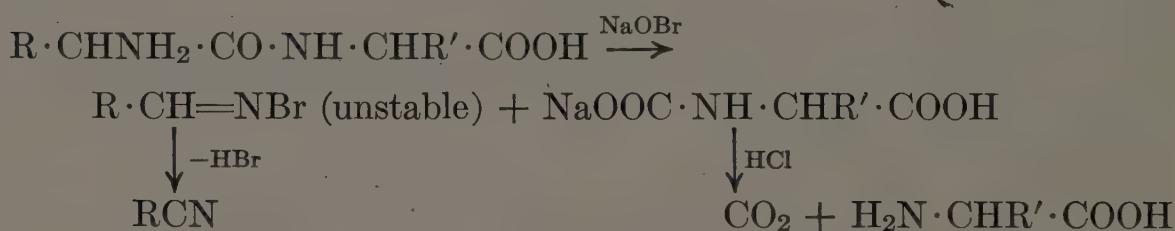
Acyl Halides. Those acyl halides traditionally used for identification of amino acids (benzoyl chloride, benzenesulfonyl chloride, and β -naphthalenesulfonyl chloride) have been employed. The use of these relied upon the more rapid hydrolyzability of peptide bonds than of the introduced amide linkage, plus the different solubility properties of the residue with masked amino group. One of the limitations of this type of procedure, appreciable hydrolyzability, applies particularly to acyl substituents. This factor is more important with some peptides than others, since particular linkages are known to be hydrolyzed relatively much more rapidly than others (Table 5-1).

Hypobromite. Hypobromite, both acid and alkaline, has been used. In acid solution the reactions are



The keto acid may be identified as the phenylhydrazone.

Alkaline hypobromite releases a nitrile with one carbon less than the terminal residue bearing the free amino group:

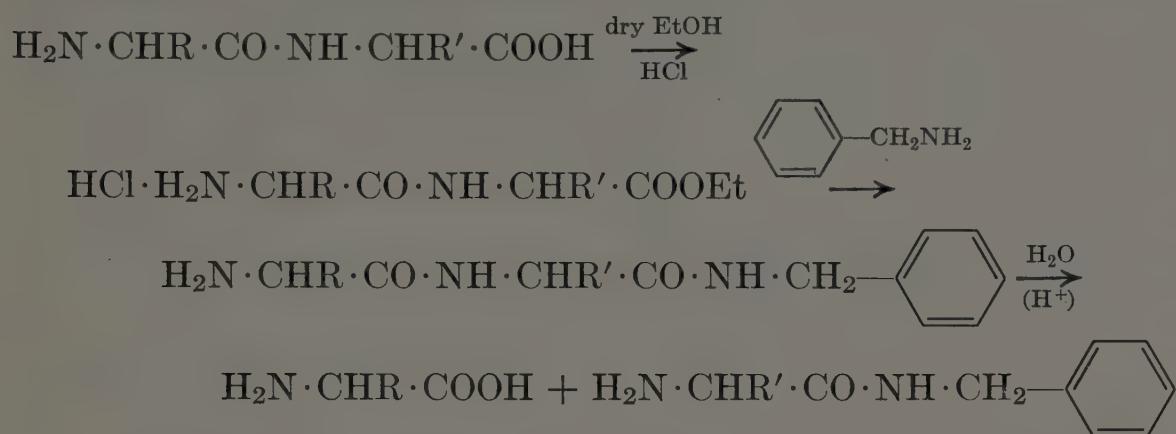


Many synthetic peptides have responded to this procedure, including: alanylvalylglycine, alanylleucylglycine, alanylalanylleucine, alanylalanylglycine, alanylalanylalanylglycine, valylalanylglycine, and leucylalanylvalylglycine. With tripeptides and tetrapeptides, dehydrodantoins form from the peptide residue.

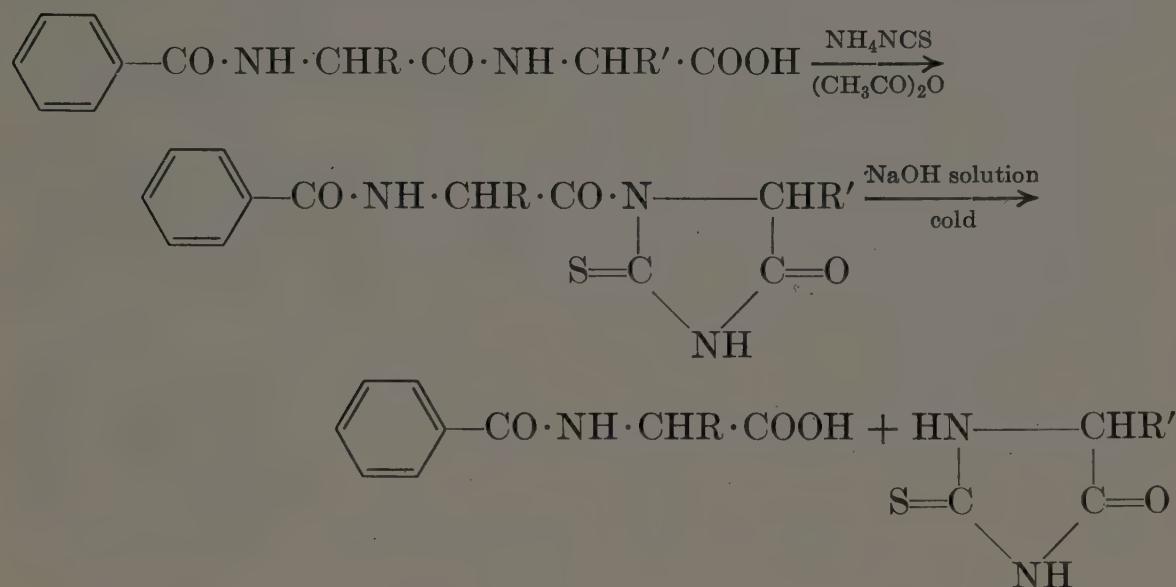
CARBOXYL CONDENSATIONS

Reaction with Benzylamine. Benzylamine has been employed on the carboxyl terminus in a manner analogous to that for some of the acyl blocking agents for the amino group. The peptide is converted to an ester and heated in suspension in benzylamine, to yield a benzyl-

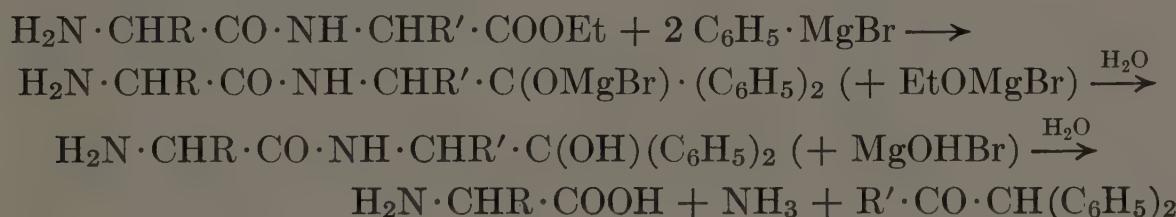
amide the peptide linkage of which hydrolyzes less rapidly than those between two amino acid residues.



Thiohydantoin Formation. The residue with a free carboxyl is converted to a thiohydantoin, which is liberated by cold alkali and identified. The free amino group must first, however, be protected.



The Grignard Reaction. This classical reaction has been applied to synthetic peptides and to the location of glycine in glutathione.

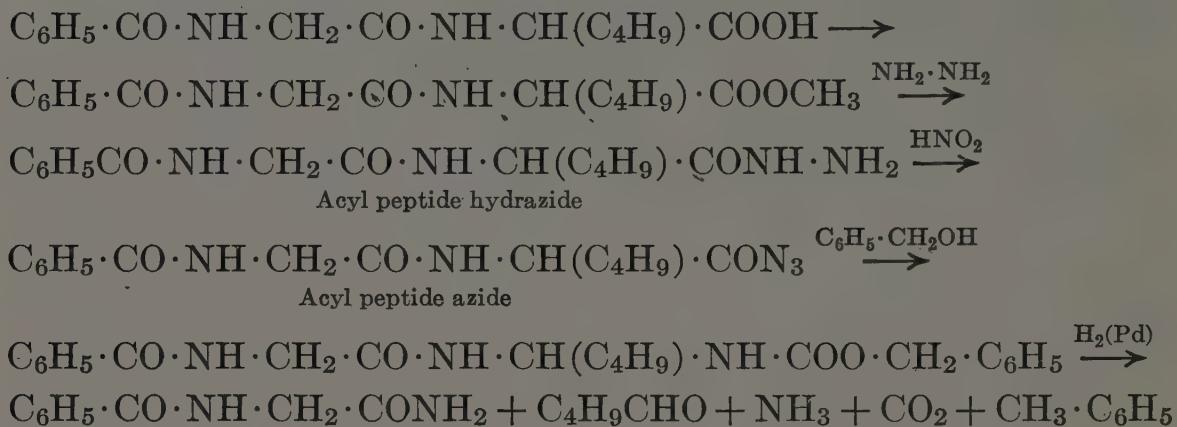


The ketone obtained corresponds to the terminal amino acid.

STEPWISE DEGRADATIONS

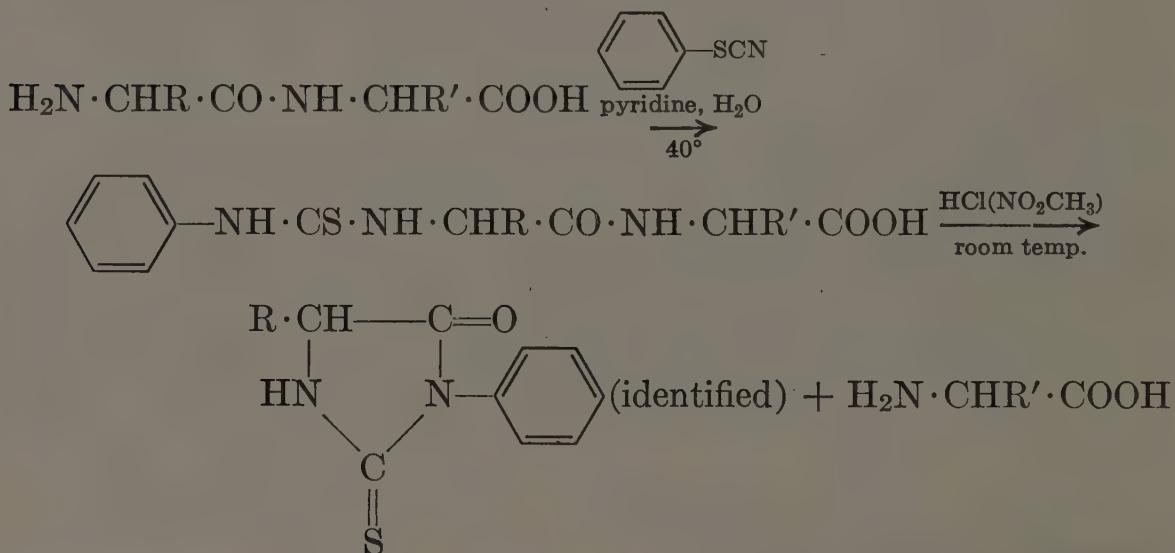
The Bergmann-Zervas-Schneider Degradation. This stepwise degradation relies upon identification of an aldehyde with one carbon

atom less than the terminal amino acid bearing the free carboxyl group. The method is illustrated by the following reaction sequence.



In this case the terminal leucine is recognized through the derived isovaleraldehyde, which is separated as a characteristic derivative such as the nitrophenylhydrazone. The peptide amide left is convertible to a hydrazide and subject to further degradation, in the same manner as an ester. Since many reactions are involved, the over-all yield is low, about 10 per cent in such cases as that given. On this basis a dipeptide would give an over-all yield of the second aldehyde of about 1 per cent. Nevertheless, the method has been successfully applied to the determination of sequence in a synthetic tetrapeptide, glycylalanyleucylglutamic acid.

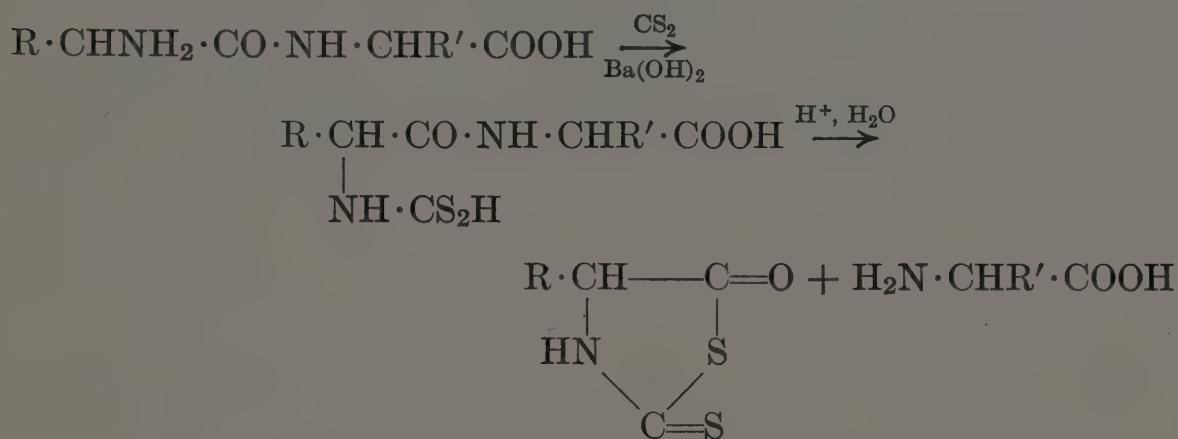
The Edman Method. A modification (Edman, 1950) of Abderhalden's phenylisocyanate procedure, except that zero peptide bond hydrolysis is claimed, is carried out under the following conditions with phenylisothiocyanate (PTC).



This method has been applied to alanylglycine, leucylglycine, leucyltyrosine, leucylglycylglycine, and alanylleucylglycine. Inasmuch as

dioxane-HCl gives almost the same result as nitromethane-HCl, the selective fission of the altered terminal residue is probably more a function of the ring structure than of the solvent.

The Thiothiazolidone Method. The thiothiazolidone procedure requires no elevated temperatures.

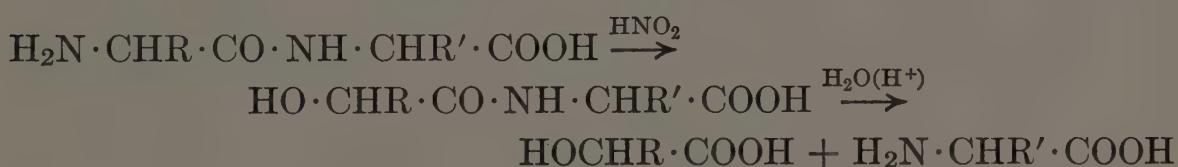


Since these reactions are conducted at room temperature, stepwise degradation is possible.

OTHER PROCEDURES

Comparison with Synthetic Peptides. Isolation and comparison with synthetic peptides has been employed. This is limited practically by the general unavailability of the necessary peptides. When a particular peptide is suspected, however, it may prove feasible to synthesize one or a few for comparison. This was done for peptides obtained by partial hydrolysis of gramicidin S. The comparison was carried out by study of the R_f values on papergrams.

Nitrous Acid. The Van Slyke reagent has been used in a number of ways. It has served to establish sequence in dipeptides composed of proline, nitrogen being liberated only when the proline carboxyl is free. It has also been used to yield hydroxy acid, which is recoverable after hydrolysis of a peptide,



and is separable from the amino acids resulting from hydrolysis.

Perhaps the most interesting application is in the identification of dipeptides on paper strips in a subtractive manner. The dipeptide is first hydrolyzed and chromatographed on paper to indicate the component amino acids. Another aliquot is treated with nitrous acid fumes until only one spot is obtained. The spot which has disappeared

corresponded to the amino acid that bore the amino group. This micro method was used for pilot studies on gramicidin S, and was followed for confirmation by comparison of the hydrolytic fragments with synthetic peptides.

EXAMPLES OF SEQUENCE DETERMINATION

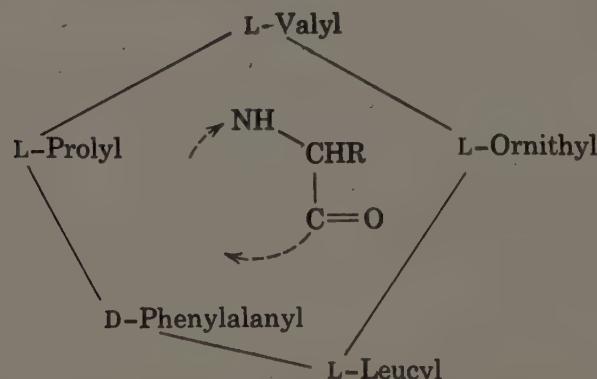
Glutathione. After the three amino acids of glutathione had been recognized, their sequence was first established by titration studies; the conclusions correctly pointed to the atypical glutamic acid linkage. Titration data would probably not have been so valuable in a tripeptide less well provided with functional groups. The structure of each of the dipeptides from glutathione was also determined, by the trinitrotoluene method.

Many of the methods previously discussed in this chapter, including nitrous acid treatment, the Grignard reaction, the thiohydantoin formation, and carboxypeptidase behavior, were used to confirm phases of the evidence.

Gramicidin S. Consden, Gordon, Martin, and Synge, who are largely responsible for the practical applications of paper chromatography to amino acids and peptides, isolated the following dipeptides from hydrolysis of the antibiotic: α -valylornithine, ornithylleucine, leucylphenylalanine, and phenylalanylproline. They had evidence also for the tripeptides valylornithylleucine and phenylalanylprolylvaline.

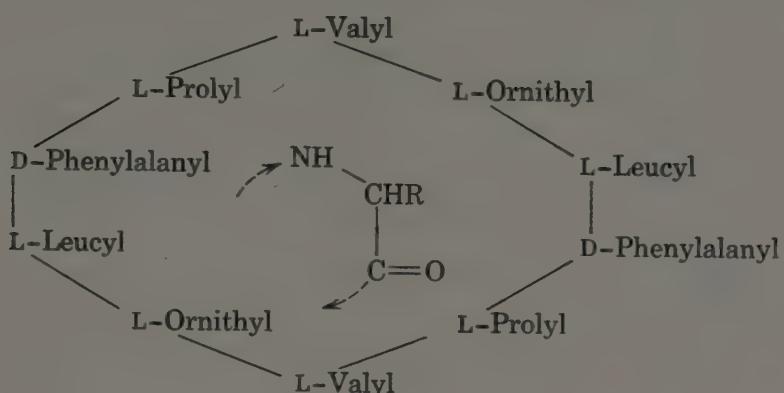
They identified these principally by the subtractive nitrous acid method, and confirmed their findings by comparison on paper with the synthetic peptides α -L-valyl-L-ornithine, L-ornithyl-L-leucine, L-leucyl-D-phenylalanine, D-phenylalanyl-L-proline, and L-prolyl-L-valine.

Gramicidin S was found to have no free α -amino nor α -carboxyl group, and the structure was accordingly first considered to be

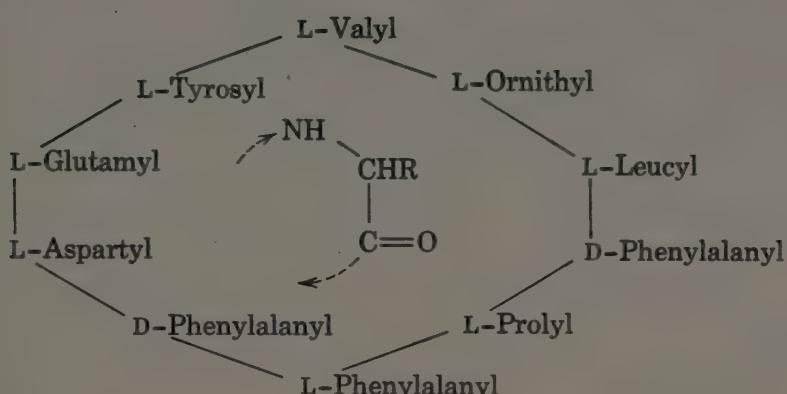


Battersby and Craig, however, showed that the pentapeptide unit

was once repeated and the structure of gramicidin S is therefore



Tyrocidine A. By use of countercurrent extraction of peptide fragments, by interpretation similar to that of Consden, Gordon, Martin, and Syngle, and by some auxiliary use of the DNP method, Paladini and Craig revealed the structure of tyrocidine A to be



The right hand half of this cyclodecapeptide is identical to that of gramicidin S.

Oxytocin-Vasopressin. Determination of the structure of oxytocin and vasopressin relied upon the judicious employment of several methods of peptide analysis. This beautiful work is outlined in Chap. 20. It deserves the reader's perusal of the original articles for a number of reasons, one of which is the masterful coordination of several methods.

APPLICATION TO LARGE PEPTIDES AND PROTEINS

In principle, most of the methods described in this chapter are applicable to peptide chains of any size if one obtains fragments which overlap and which are small enough to permit convenient assignment of sequence. One is not limited to the smaller peptides of the preceding section if he can obtain from larger molecules peptides which overlap in their sequences. Such a process is more feasible if relatively

specific modes of fragmentation can be applied prior to separation of the hydrolytic products.

Sequence assignment in molecules such as insulin, corticotropin, and ribonuclease have relied largely upon fragmentation with proteolytic enzymes, as well as with mineral acids. Use of both types of hydrolysis including several proteases has provided fragments which were found to overlap. The proteolytic enzymes have been useful for attack at relatively predictable loci in the chains. Trypsin, in particular, has been found to split large peptides at the carbonyl group of basic amino acids and to leave other linkages relatively untouched.

QUANTITATIVE ANALYSIS OF RESIDUE SEQUENCE

When an analytical method can be employed in a quantitative manner, the resulting data are frequently more meaningful than those that may be obtained from a purely qualitative application. Peptide analysis has afforded no exception to this rule. The advantages of such an approach include greater certainty in structural determina-

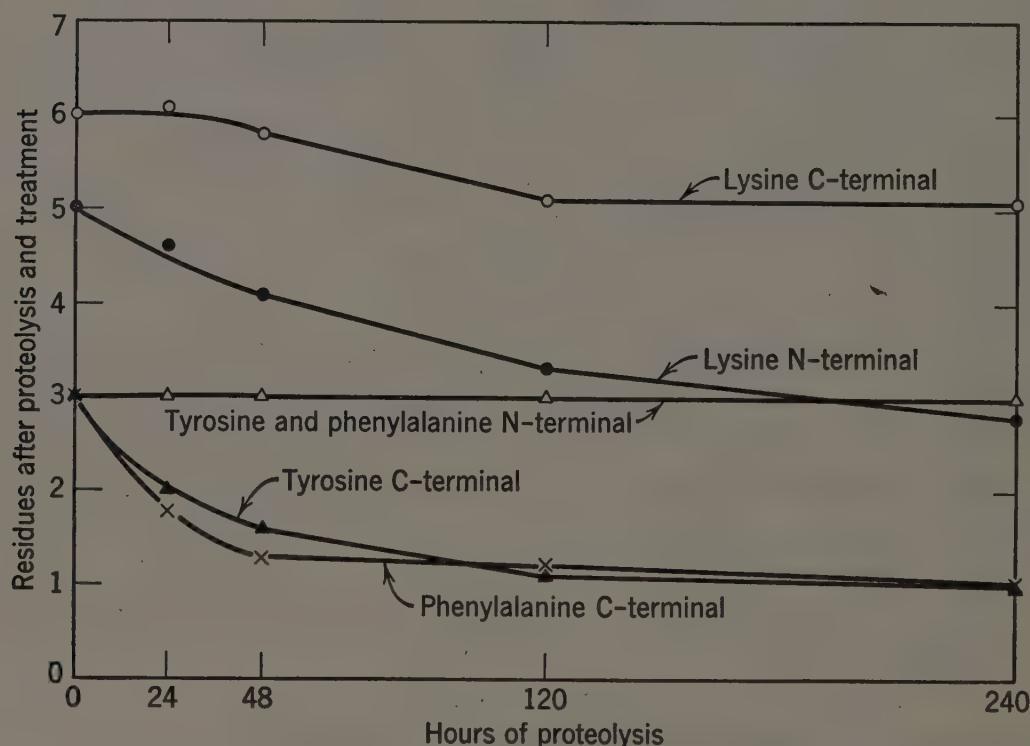


Fig. 9-1. Changes in exposure of individual amino acids during chymotryptic proteolysis of lysozyme ($\text{pH} 8.5, 37^\circ \text{C}$). Data of T. L. Hurst and S. W. Fox.

tions, the basing of negative conclusions on positive data, and the opportunity to learn of peptide changes in dynamic systems. Problems like the nature of protein hydrolysis, changes in biologically active materials due to radiation, and metamorphoses such as that from the

inactive precursor of an enzyme to the active enzyme, may be studied more explicitly. In Fig. 9-1 are presented some of the special changes occurring during proteolysis, as studied by this method. Such information is taken up in greater detail in Chap. 22.

CONCLUSIONS

By way of summary, some comparative advantages and disadvantages of a few of the most used methods may be considered.

The DNP method provides a DNP amino acid which has properties that permit ready separation. The separation of the yellow DNP amino acids on columns or paper strips can be followed easily. This procedure offers relatively definite indication of small amounts of free amino groups.

The quantitative application of this method is impaired by the fact that the DNP substituent is not entirely nonhydrolyzable. Correction factors based on incompletely controllable experiments can, however, be provided. Certain amino acids form DNP derivatives that are particularly unstable to the conditions producing hydrolysis of peptide bonds. Notable among these are proline, glycine, and cystine. The behavior of histidine is not clear. Artifacts are known to be formed from DNP derivatives of methionine, tyrosine, cystine, and tryptophan. The method fails to reveal its own defects, inasmuch as products of aberrant reactions may be lost on the columns. It has not been developed by tests on many peptides of known constitution.

The stepwise procedure employing phenylisothiocyanate and the resultant phenylthiohydantoin is of value in qualitative studies. None of the variations of the method relying on extraction of the phenylthiohydantoin are truly quantitative. When applied subtractively, that is, by determination of all amino acids and estimation (columnar or microbiological) of unblocked amino acids, one can construct a quantitative balance sheet at each step in the series of treatments. This is tedious, but it avoids the losses attendant upon extraction of the phenylthiohydantoin.

The reagent most used for determination of C-termini has been carboxypeptidase. When employed in conjunction with paper chromatography, carboxypeptidase yields results conveniently. There is reason to believe that this reagent responds differently, and perhaps in some cases not at all, to peptide linkages composed of different amino acid residues.

This is by no means a complete cataloguing of advantages and disadvantages of each of the methods mentioned. The choice of method in each case must depend upon the specific experience of the

worker, his particular problem, and the consequently variable weight of each factor listed here, as well as other factors. The summary is intended principally to underline the point that many emphases may enter into a choice. It is also clear that there is much value in having available methods which may supplement and confirm or deny the results of other methods.

Although methods of terminal residue assignment and peptide sequential analysis are the proper concern of this chapter in particular, their use has spread to many areas of protein chemistry.

SELECTED REFERENCES

Reviews

"Partial Hydrolysis Products Derived from Proteins and their Significance for Protein Structure," R. L. M. Synge, *Chem. Revs.*, **32**, 135 (1943).

"Terminal Amino Acids in Peptides and Proteins," S. W. Fox, *Advances in Protein Chem.*, **2**, 155 (1945).

"The Arrangement of Amino Acids in Proteins," F. Sanger, *Advances in Protein Chem.*, **7**, 1 (1952).

"Peptides: Methods of Synthesis and Terminal-Residue Studies," H. D. Springall and H. D. Law, *Quart. Revs. London*, **10**, 230 (1956).

Original Articles

"Glutathione: A Reinvestigation," F. G. Hopkins, *J. Biol. Chem.*, **84**, 269 (1929).

"Beitrag zur Konstitutionsermittlung von Proteinen bzw. Polypeptiden," E. Abderhalden and H. Brockmann, *Biochem. Z.*, **225**, 386 (1930).

"Method for the Stepwise Degradation of Polypeptides," M. Bergmann, L. Zervas, and F. Schneider, *J. Biol. Chem.*, **113**, 341 (1936).

"Clupein. VIII," K. Felix and A. Mager, *Hoppe-Seyler's Z. physiol. Chem.*, **249**, 111 (1937). (See also papers cited in bibliography.)

"Chemical Properties of Bactericidal Substances Isolated from Cultures of a Soil Bacillus," R. D. Hotchkiss and R. Dubos, *J. Biol. Chem.*, **132**, 793 (1940).

"The Chemical Nature of Gramicidin and Tyrocidine," R. D. Hotchkiss, *J. Biol. Chem.*, **141**, 171 (1941).

"The Identification of Lower Peptides in Complex Mixtures," R. Consden, A. H. Gordon, and A. J. P. Martin, *Biochem. J.*, **41**, 590 (1947).

"Gramicidin S. The Sequence of the Amino Acid Residues," R. Consden, A. H. Gordon, A. J. P. Martin, and R. L. M. Synge, *Biochem. J.*, **41**, 596 (1947).

"The Nature of Bacitracin," G. T. Barry, J. D. Gregory, and L. C. Craig, *J. Biol. Chem.*, **175**, 485 (1948).

"Method for Determination of Amino Acid Sequence in Peptides," P. Edman, *Acta Chem. Scand.*, **4**, 283 (1950).

"A Microbiological Method for the Determination of Sequences of Amino Acid Residues," S. W. Fox, T. L. Hurst, and K. F. Itschner, *J. Am. Chem. Soc.*, **73**, 3573 (1951).

"The Amino End-Group and the Amino Acid Composition of Salmine," S. P. Velick and S. Udenfriend, *J. Biol. Chem.*, **191**, 233 (1951).

"The Stepwise Degradation of Peptides," S. G. Waley and J. Watson, *J. Chem. Soc.*, **1951**, 2394.

"Sequence of Four Amino Acids at the Amino End of the Single Polypeptide Chain of Lysozyme," W. A. Schroeder, *J. Am. Chem. Soc.*, **74**, 5118 (1952).

"Determination of the Terminal Carboxyl Residues of Peptides and of Proteins," V. H. Baptist and H. B. Bull, *J. Am. Chem. Soc.*, **75**, 1727 (1953).

"The Sequence of Amino Acids in Oxytocin, with a Proposal for the Structure of Oxytocin," V. du Vigneaud, C. Ressler, and S. Trippett, *J. Biol. Chem.*, **205**, 949 (1953).

"The Chemistry of Tyrocidine. III. The Structure of Tyrocidine A," A. Paladini and L. C. Craig, *J. Am. Chem. Soc.*, **76**, 688 (1954).

"A New Method for Identifying C-Terminal Residues in Peptides," R. A. Turner and G. Schmerzler, *J. Am. Chem. Soc.*, **76**, 949 (1954).

"Studies on the Determination of the Sequence of Amino Acids in Peptides and Proteins. 3. The Synthesis of 3-(4'-Dimethylamino-3':5'-Dinitrophenyl) Hydantoin Derivatives of Various Amino Acids, and Their Use for the Determination of N-Terminal Amino Acids," G. G. Evans and W. S. Reith, *Biochem. J.*, **56**, 111 (1954).

"Structure of β -Corticotropin: Final Sequence Studies," K. S. Howard, R. G. Shepherd, E. A. Eigner, D. S. Davies, and P. H. Bell, *J. Am. Chem. Soc.*, **77**, 3419 (1955).

"Studies on Adrenocorticotropin. XII. Action of Aminopeptidase on Corticotropin-A; Effect on Biological Activity," W. F. White, *J. Am. Chem. Soc.*, **77**, 4691 (1955).

"Dégradation systématique des dinitrophenylpeptides," M. Jutisz and W. Ritschard, *Biochim. et Biophys. Acta*, **17**, 548 (1955).

"Products of Hydrolysis of Dinitrophenylproline and N-Terminal Proline-peptides with Hydrochloric Acid," F. S. Scanes and B. T. Tozer, *Biochem. J.*, **63**, 282 (1956).

"The N-Terminal Amino Acid Residues of Normal Adult Human Hemoglobin: A Quantitative Study of Certain Aspects of Sanger's DNP Method," H. S. Rhinesmith, W. A. Schroeder, and L. Pauling, *J. Am. Chem. Soc.*, **79**, 609 (1957).

10

Synthesis of Peptides

Aminoacyl halide reaction

Diketopiperazine hydrolysis

Haloacyl halide method

Large synthetic peptides

The Bergmann-Zervas carbobenzoxy synthesis

The N,N'-dicyclohexylcarbodiimide method

The phosphite amide synthesis

N-carboxy amino acid anhydrides and polyamino acids

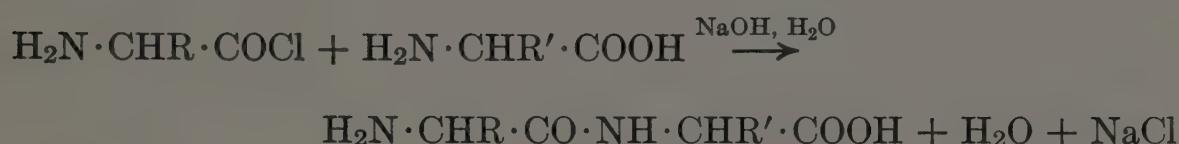
Other methods

Three principal objectives have been discernible in the voluminous reports of work on synthesis of peptides. One of these has been the characteristic goal of the organic chemist in seeking to prepare artificially the biologically active products of nature. A most notable instance of success of this sort was the first synthesis of a peptide hormone, oxytocin, by du Vigneaud and co-workers in 1953. Another objective has been the preparation, in a controlled manner, of a wide variety of peptides of definite constitution (in contrast to some of the mixtures obtained by isolation) for experimental studies, as in protease-substrate interaction. A third objective has been the fundamental one of attempting to imitate the natural synthesis of protein.

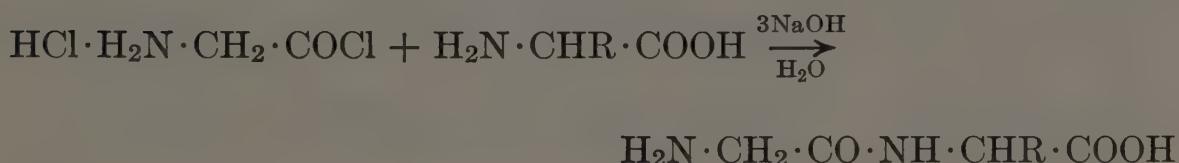
AMINOACYL HALIDE REACTION

The successful accomplishment of peptide synthesis might be considered basically to involve coupling of two or more amino acid

molecules. The direct combination of two such amino acid molecules has proved to be feasible in only a few special instances. Direct reaction of two molecules to form a peptide linkage by loss of water can be effected, but is complicated by the necessity for high temperatures and consequent side reactions. In this case the other amino and carboxyl groups also react. The reaction can be conducted at lower temperatures, and without appreciable side reaction, if the carboxyl group is converted to an acid chloride, thus rendering it an acylating agent.



The classical reagent for conversion of carboxyl groups to carboxyl chlorides, phosphorus pentachloride, was believed to bring about deep-seated changes in the amino acid molecule. Emil Fischer was able to overcome this difficulty for glycine by reacting the pentachloride with glycine in acetyl chloride as a solvent. This yielded glycyl chloride hydrochloride, which, when carefully handled, could be coupled by the Schotten-Baumann reaction.

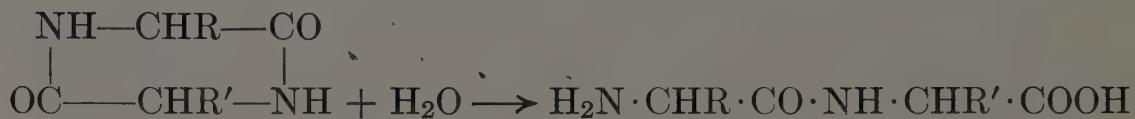


The Schotten-Baumann reaction, fundamental to much of peptide synthesis in general, employs alkali to neutralize the organic acid present and the HCl or HX which is split out (the reaction is greatly slowed down in acid solution). In this instance 3 moles of alkali are essential, since the hydrochloride must also be neutralized. Although Fischer used glycyl chloride hydrochloride in the case just described, he employed other methods for almost all of the numerous peptides that he prepared. The reason for this was undoubtedly the difficulty of preserving the acid chlorides. When this work was continued many decades later, it was found that the aminoacyl chlorides tended to form anhydropolymers of amino acids.

This type of reaction has been employed, however, in the synthesis of peptide analogs of penicillin.

DIKETOPIPERAZINE HYDROLYSIS

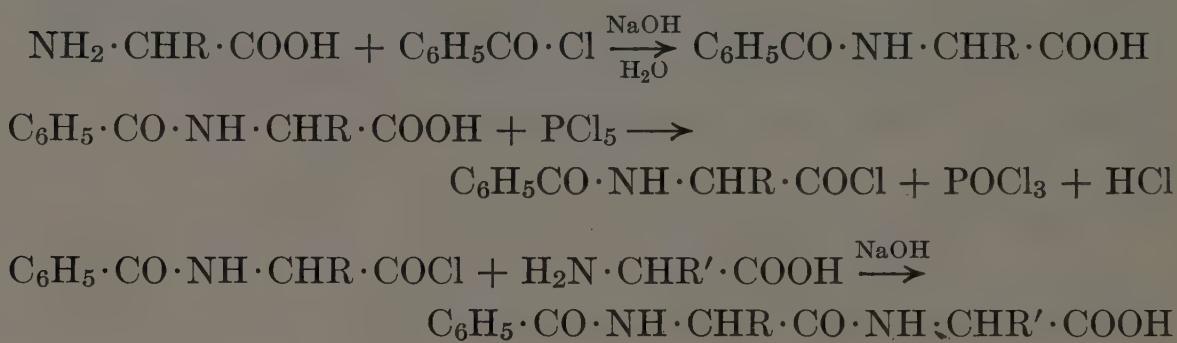
Another method employed hydrolysis of diketopiperazines:



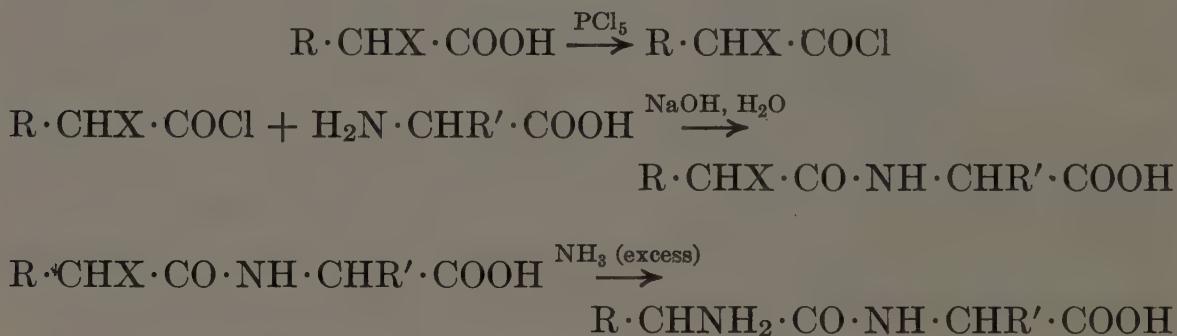
It is possible to hydrolyze one of the peptide linkages more rapidly than the other. One minute of boiling with concentrated hydrochloric acid solution, for instance, hydrolyzes glycine anhydride to glycyl-glycine without excessive hydrolysis of the second peptide linkage. This method has limited applicability, being of practical value only for dipeptides from a single amino acid.

HALOACYL HALIDE METHOD

In obviating the lability of amino groups to phosphorus pentachloride, Curtius protected the amino group with a benzoyl substituent:



Regeneration of the peptide was not accomplished. Hydrolysis of the benzoyl group also leads to hydrolysis of the peptide linkage. Fischer overcame this difficulty by employing haloacids that could be aminated after the peptide linkage was formed.



This method was the principal mode of synthesis of peptides until 1932. Its main disadvantages are (1) that it is apparently limited to amino acids with hydrocarbon side chains and (2) that use of opti-

cally active intermediates is difficult or often impractical. Many peptides have been made by this process however.

LARGE SYNTHETIC PEPTIDES

The largest synthetic peptides of known constitution have been fabricated by repeated coupling of haloacyl halides with peptides and subsequent combination of polypeptide units.

By this procedure, Fischer synthesized L-leucylglycylglycylglycyl-L-leucylglycylglycylglycyl-L-leucylglycylglycylglycylglycylglycylglycylglycylglycylglycine. Abderhalden, one of Fischer's students, with Fodor outdid the master by one amino acid residue in preparing L-leucylglycylglycylglycyl-L-leucylglycylglycylglycyl-L-leucylglycylglycylglycyl-L-leucylglycylglycylglycylglycylglycine. Fischer thus prepared an octadecapeptide, and Abderhalden a nonodecapeptide. Abderhalden also tabulated some properties for some of his intermediates. These indicate a progression toward properties of the kinds that are associated with proteins (Table 10-1).

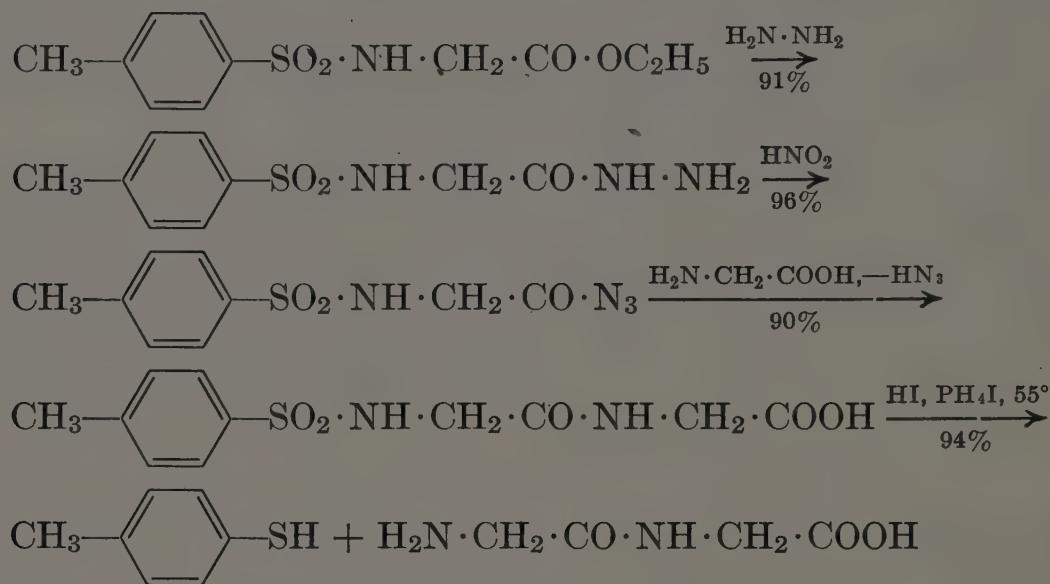
TABLE 10-1. Properties of Abderhalden's Peptides

Number of Amino Acid Residues	Molecular Weight	Solubility in Cold Water	Behavior on Saturation of Aq. Solution with $(\text{NH}_4)_2\text{SO}_4$
3	295	Very easily soluble	Not salted out
5	359	Easily soluble	Not salted out
7	473	Fairly easily soluble	Salted out
11	757	Difficultly soluble	Salted out
15	1041	Difficultly soluble	Salted out
19	1325	Difficultly soluble	Salted out

With the exception of the special preparation of anhydropolymers of amino acids (p. 170), no method of synthesis has been generally applied to the fabrication of such large peptides. The haloacyl halide method is readily applicable only to amino acids in the monoamino-monocarboxylic acid series. Even in this series, the preparation of peptides containing optically active units is tedious, whereas optically active peptides are the type principally desired for enzymic and related studies. The chief limitation, however, is that more complex amino acids such as histidine and tryptophan are not smoothly brought into combination. For this purpose the first method shown to be of comprehensive value was that of Bergmann and Zervas (1932).

Schönheimer (1926), however, had reported that peptides could be synthesized through tosyl derivatives from which the blocking group

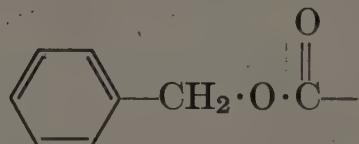
was finally removed by reduction in aqueous solution with hydriodic acid and phosphonium iodide. Schönheimer established that the conditions employed resulted in almost no hydrolysis of peptide bonds.



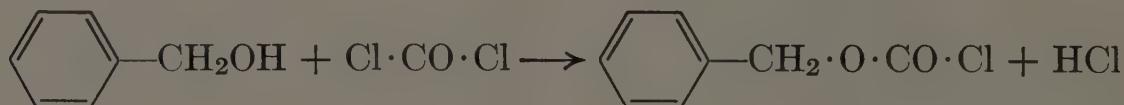
Yields of glycyl-DL-alanine, glycyl-DL-leucine, DL-alanylglycine, DL-leucylglycine, and L-alanyl-L-leucine in the last step were respectively 90, 87, 91, 91, and 84 per cent.

THE BERGMANN-ZERVAS CARBOBENZOXY SYNTHESIS

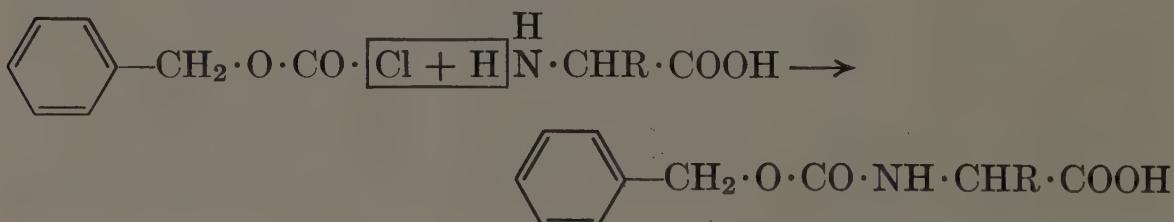
The Bergmann-Zervas method succeeded in combining optically active amino acids by the use of a blocking group which could be removed from the penultimate product by nonhydrolytic means. This blocking group is carbobenzoxy (benzylcarbonato, benzylloxycarbonyl),



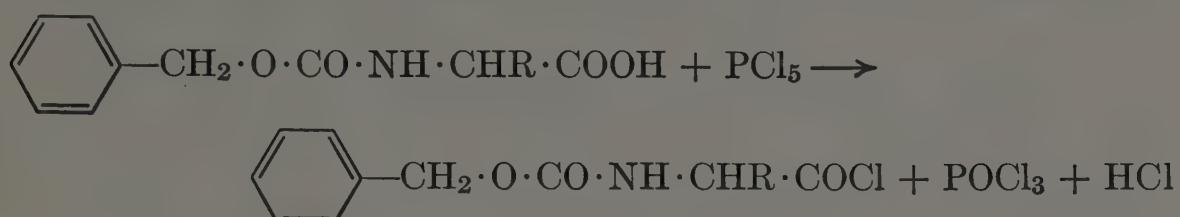
The carbobenzoxy chloride is prepared in toluene solution from benzyl alcohol and phosgene:



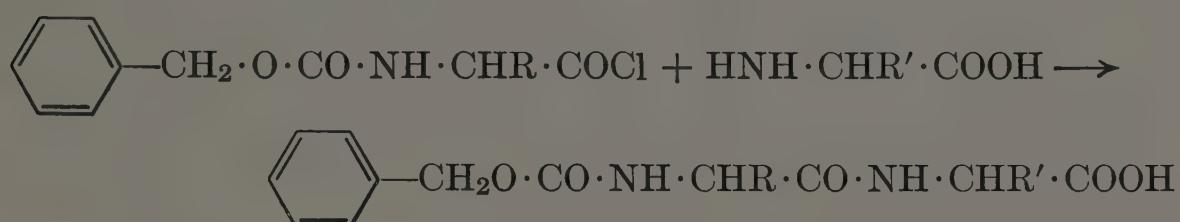
The resultant carbobenzoxy chloride may then be condensed with the amino acid by the Schotten-Baumann reaction:



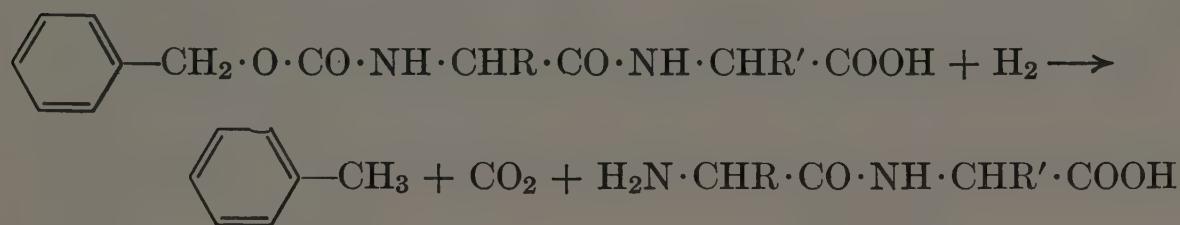
This product can be converted to the acid chloride with PCl_5 :



By a second Schotten-Baumann reaction this chloride can be condensed with a second amino acid:



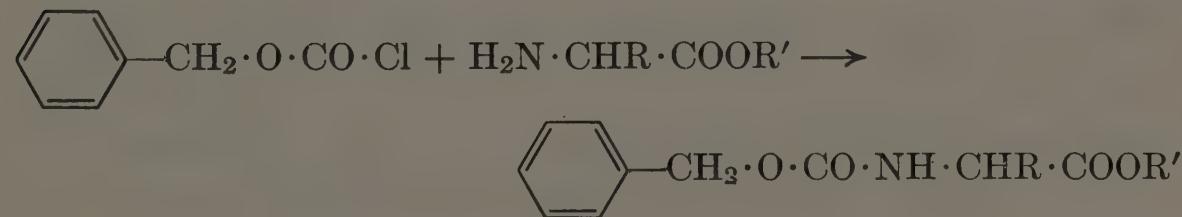
This process may be repeated. At any stage the carbobenzoxy residue may be removed by catalytic hydrogenolysis, which yields the peptide and volatile by-products:



The optically active dipeptide results, and the method has been found to be applicable (with occasional modification) to virtually all amino acids. All of the reactions may be carried out at room temperature.

Difficulties are often encountered in conversion of a carbobenzoxy-amino acid to the corresponding acyl chloride with phosphorus pentachloride or with thionyl chloride. This is true not only for peptides that contain obviously reactive side chains, such as in serine and glutamic acid, but also for such starting materials as carbobenzoxy-L-leucine. For this reason, the interpolation of Curtius' azide coupling has become popular in carrying out peptide syntheses. The reactions resemble in large part those of the Bergmann-Zervas-Schneider degradation, described in Chap. 9.

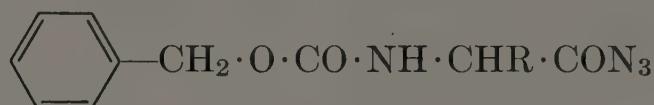
The initial point of departure from the acyl halide avenue is the formation of a carbobenzoxyamino acid ester (R' is usually CH_3):



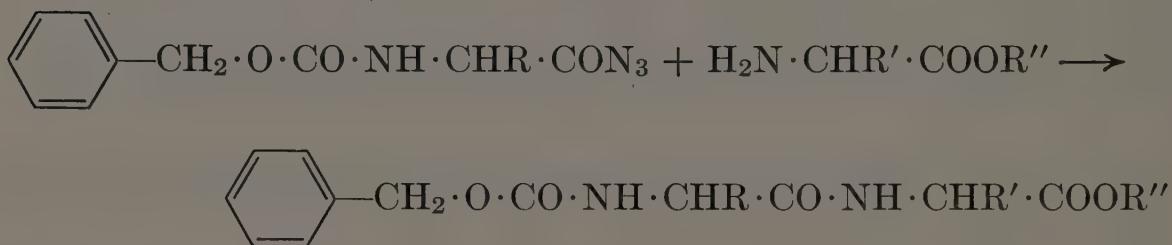
The ester is treated with hydrazine to form a hydrazide



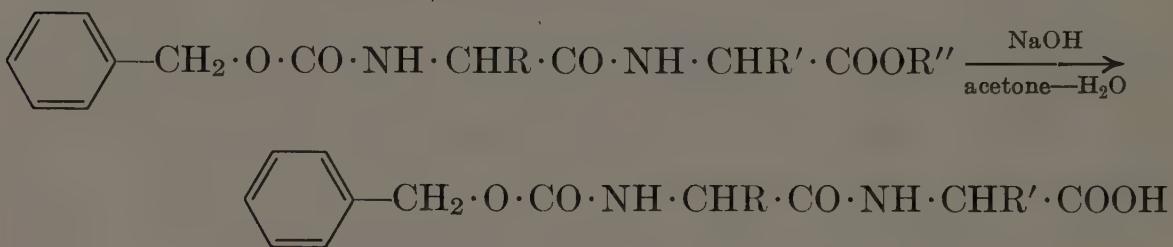
On reaction with nitrous acid, an acid azide results:



To construct a peptide it is now necessary to bring the azide into prolonged contact (room temperature overnight, typically) with not an amino acid but an amino acid ester. The acid azide group has reaction properties similar to that of an acyl halide; in this case, however, the hydrazoic acid eliminated is liberated as a gas:



The ester may be converted to the substituted dipeptide by mild alkaline hydrolysis,

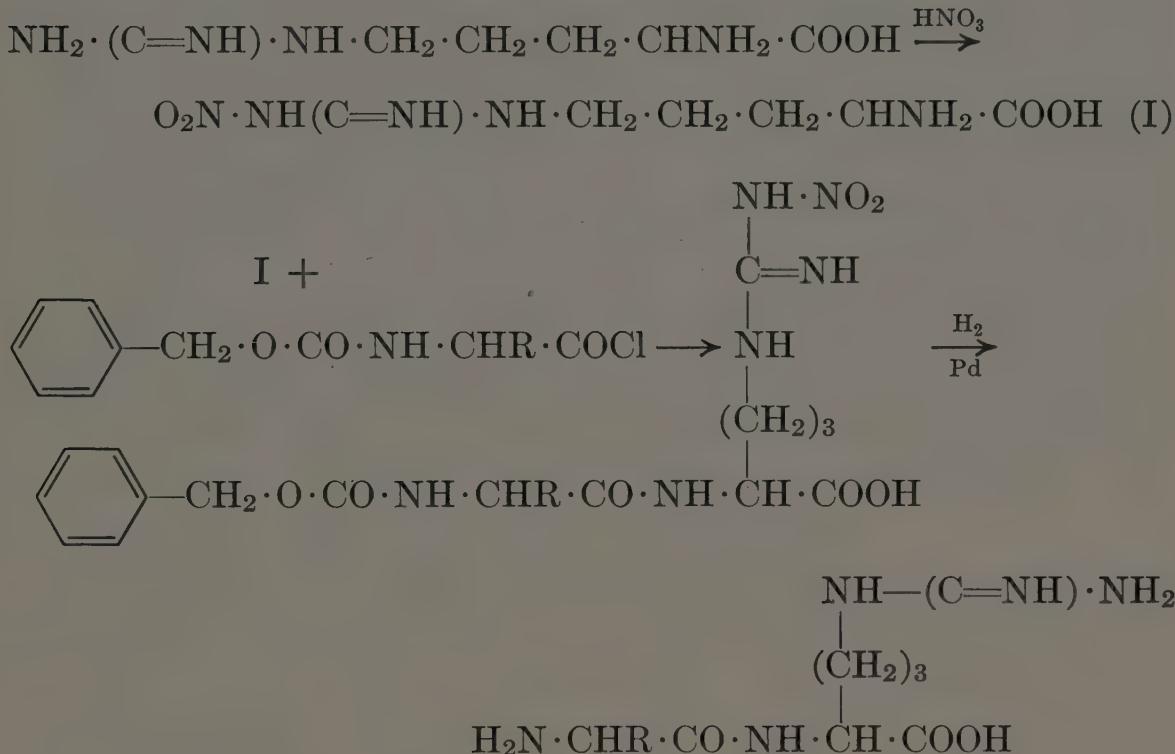


and then hydrogenated to the dipeptide, or the ester may be carried through another azide formation and reaction.

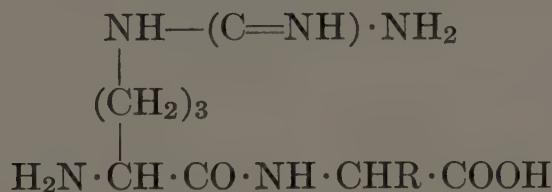
Cystine is among the amino acid residues which require special treatment. Appropriate procedures have been developed for this amino acid, especially because of its occurrence in glutathione. The sulfur compounds poison the palladium or platinum catalysts used in the hydrogenation, but other reducing agents, such as phosphonium iodide and sodium in liquid ammonia, have been successfully employed. When the cysteine residue is involved, the $-\text{SH}$ group may

be protected by a benzyl residue, which is removed on reduction of the entire compound with sodium in liquid ammonia.

Another troublesome amino acid residue in this synthesis was arginine. Bergmann and co-workers (1934) found that the difficulties introduced by the guanidino group of arginine could be eliminated by initial nitration of the basic group, followed by its final removal at the reduction stage.

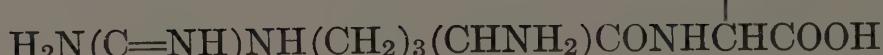
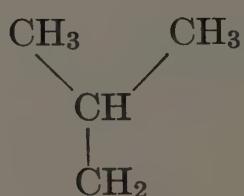
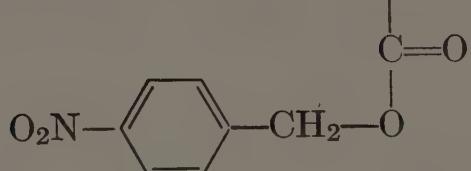
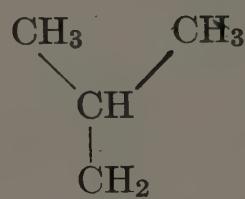
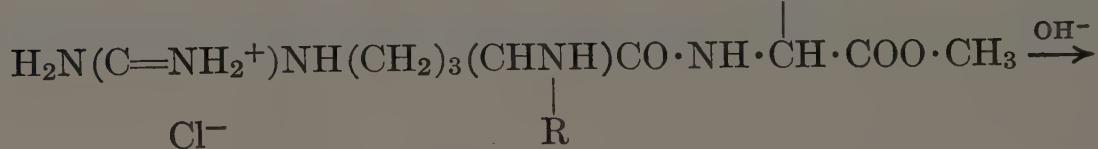
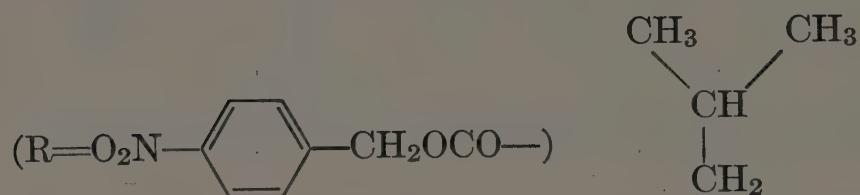
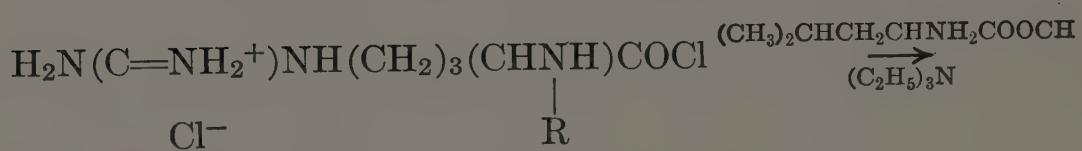
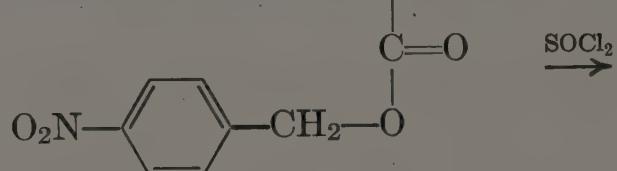
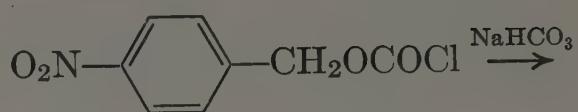


This synthesis, however, yielded only arginine peptides, not arginyl peptides; that is, the free terminus of the arginine residue in the peptide was the carboxyl group. For syntheses producing arginyl peptides, such as



the problem was suddenly solved by work from three laboratories, all described in a single issue of the *Journal of the American Chemical Society* (Gish and Carpenter, 1953; Anderson, 1953; Hofmann *et al.*, 1953).

One of the syntheses (Gish and Carpenter) proceeded through an N-p-nitrobenzyloxycarbonylarginine as follows.

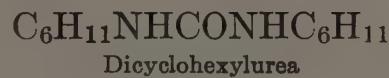
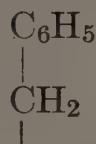
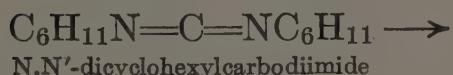
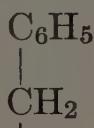


For other amino acids, glutamic acid and lysine for example, other special modifications have been introduced. The examples given will serve to indicate that the difficulties are not insurmountable. The richest profusion of varied synthetic peptides has resulted from the carbobenzoxy method. Aside from the advantages of nonhydrolytic removal of the protective group, and the applicability of the method to almost any amino acid in its optically active form, the various reactions may be accomplished at sufficiently low temperatures that deleterious effects from such processes as coupling amino acids to proteins can be kept to a minimum. Harington, for example, has found it possible to "load" protein molecules peripherally with thyroxine by this method, in order to study the immunological effects of the thyroxine residue.

Ben-Ishai and Berger (1952) showed that the long-held assumption that decarbobenzoylation required a hydrogenolysis process was incorrect. The reaction could be acid-catalyzed in anhydrous media. This reinterpretation of obligatory mechanism made possible simple decarbobenzoylation with hydrogen bromide, and also debenzylation of intermediate benzyl esters.

THE N,N'-DICYCLOHEXYLCARBODIIMIDE METHOD

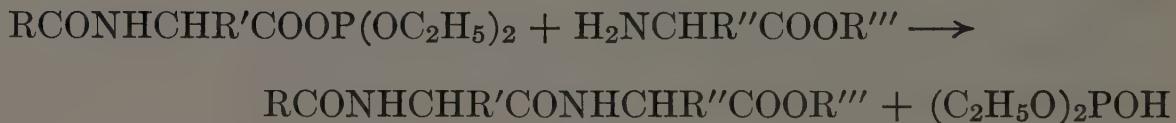
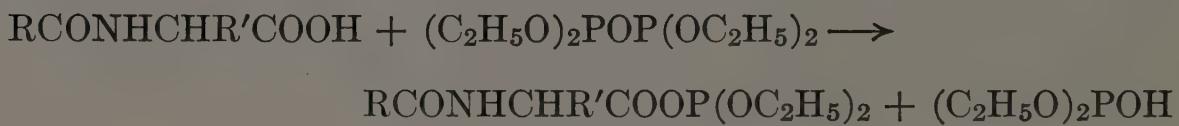
Sheehan and Hess (1955) have introduced a method which provides the unique utility of combining a carboxyl group with an amino group to form a peptide. The nonparticipating carboxyl and amino groups of the amino acids involved must be substituted. The reagent which makes this reaction possible is N,N'-dicyclohexylcarbodiimide. An exemplary synthesis takes the following path:



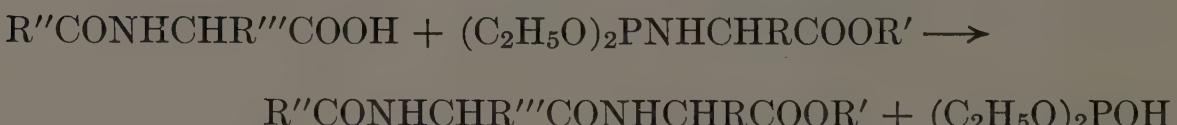
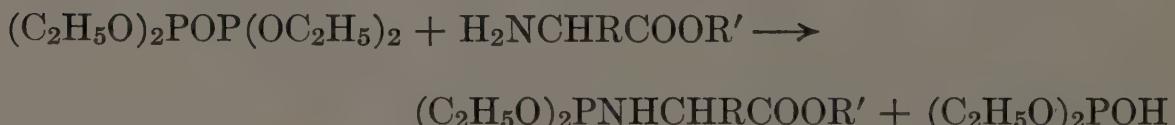
The reaction proceeds rapidly in high yield in various solvents. The reaction is also insensitive to moisture, and can in fact be conducted in aqueous solutions. The by-product dicyclohexylurea has a low solubility in most of the solvents that can be used, and is therefore easily separated. A salient feature of this mode of synthesis is that a single reaction serves to convert both the reacting carboxyl and reacting amine to a peptide bond.

THE PHOSPHITE AMIDE SYNTHESIS

Anderson and co-workers (Young *et al.*, 1956) have developed a number of ways of employing phosphorus derivatives in the synthesis of peptides. One of these, involving tetraethyl pyrophosphite, has been used in the synthesis of oxytocin and is for this and other reasons of much practical interest. The essential reactions involved are:



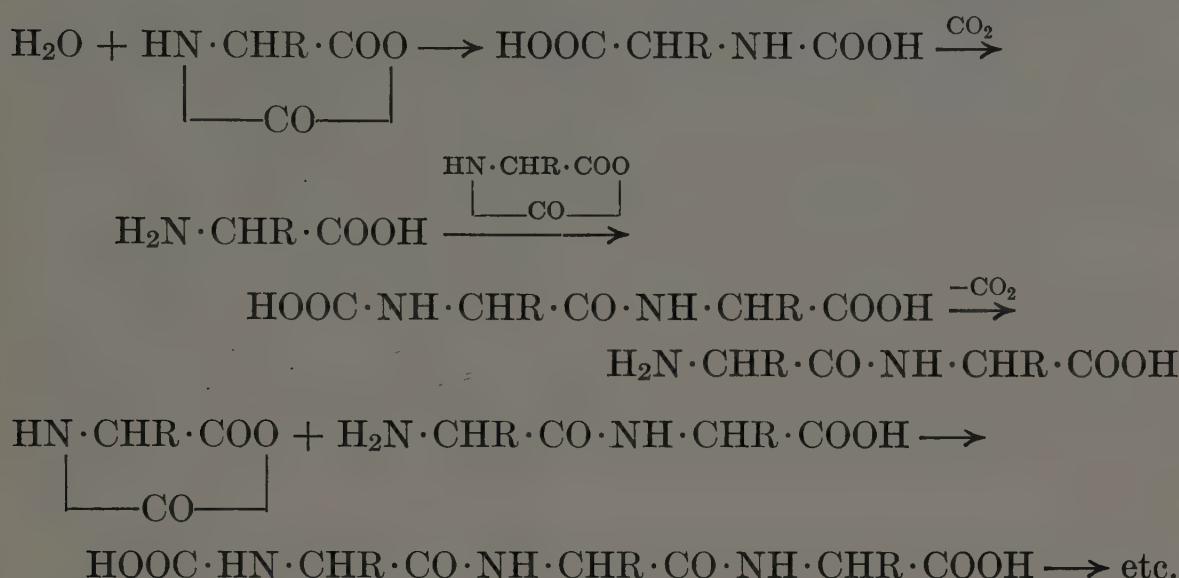
These reactions represent the "anhydride" modification of the synthesis. The "amide" modification, which was of use particularly in the du Vigneaud synthesis of oxytocin (du Vigneaud *et al.*, 1954), followed a course of this description:



N-CARBOXYAMINO ACID ANHYDRIDES AND POLYAMINO ACIDS

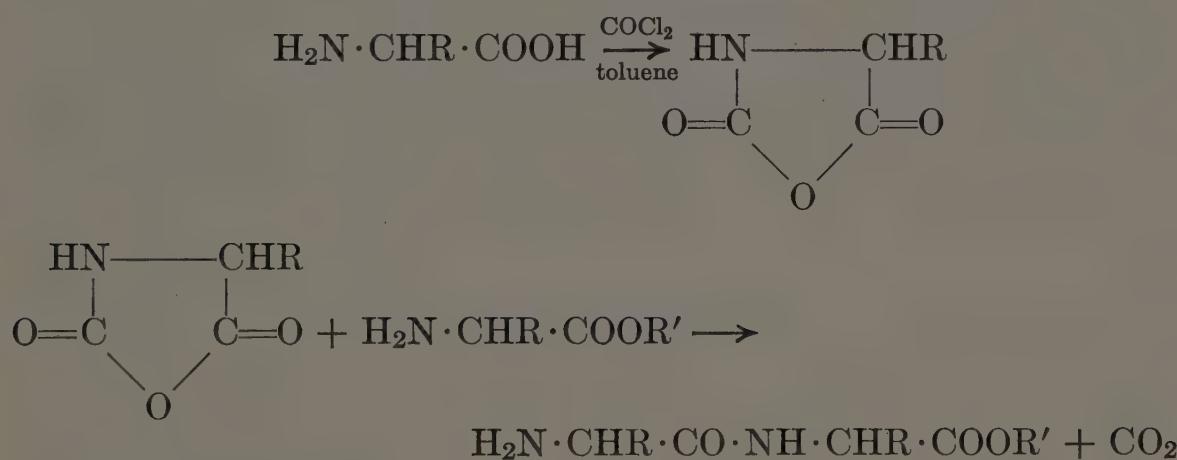
One type of peptide synthesis involves the polymerization of amino acids to large molecular types in which one or two residues are

included. This synthesis was described by Leuchs and Geiger (1908). The reactive intermediate is an N-carboxyamino acid anhydride.



Such synthetic peptides, *polyamino acids*, are being found useful for studies of configuration by X-ray diffraction and other physical methods.

A promising method for small peptides is that involving the use of N-carboxyamino acid anhydrides. These may in turn be prepared from the amino acid and phosgene or other reagents:



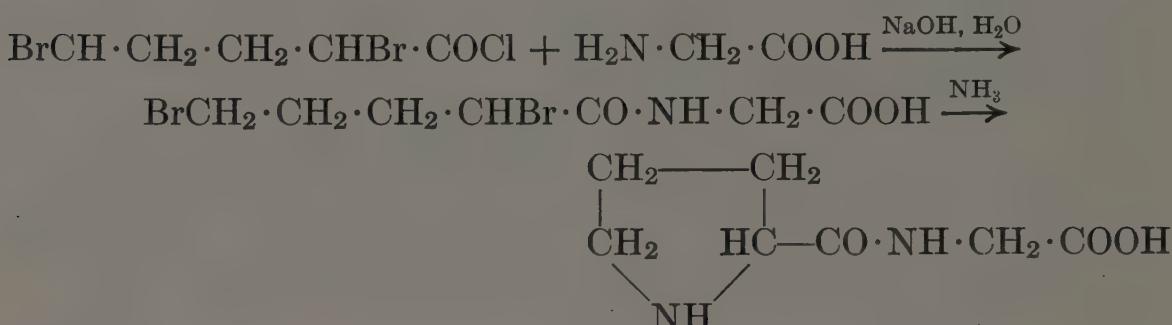
By the use of an excess of ester as condensing agent, and by control of other details, it has been possible to prepare a number of peptides of diverse structure. In principle, this method would appear to be difficult to improve upon.

OTHER METHODS

When peptides composed of simple DL-amino acid residues can be employed in investigative work, the carbobenzoxy and other methods are often not chosen. Special methods, other than those described, are

also employed for particular amino acids, and these may differ further, depending upon the position of the residue in the peptide.

As one example, glycylproline has been synthesized through the carbobenzoxy method, whereas prolylglycine may conveniently be obtained through a dibromo precursor:



Other general methods of synthesis involving azlactones, ester condensations, phthaloyl amino acids, or enzymic synthesis have all received attention. It is entirely conceivable that proper control of any of these, or yet some other method, will provide easier access to many peptides than is now possible. The use of the toluenesulfonyl group as a blocking group, subsequently removable by reduction, has been successfully applied to peptide synthesis in a number of laboratories. Reasons for its not having been used more appear to be missing from the literature. The trifluoroacetyl group has been used, and is found to be removable from peptides at pH 12 at room temperature.

It may be anticipated, with the growing interest in peptides, that new easily removable amine substituents and new modes of coupling carboxyl and amino groups will appear.

SELECTED REFERENCES

Reviews

"Peptides, Peptidases, and Diketopiperazines," J. P. Greenstein in Schmidt's *Chemistry of the Amino Acids and Proteins*, pp. 252, 1080, Charles C Thomas, Springfield, Illinois, 1943.

"The Synthesis of Peptides," J. S. Fruton, *Advances in Protein Chem.*, **5**, 1 (1950).

"Poly- α -Amino Acids," E. Katchalski, *Advances in Protein Chem.*, **6**, 123 (1951).

"Peptidsynthesen. II. 9. Mitteilung," T. Wieland, *Angew. Chem.*, **66**, 507 (1954).

Original Articles

"Synthese von Polypeptiden. XV," E. Fischer, *Ber.*, **39**, 2893 (1906).

"Über die Anhydride von α -Amino-N-Carbonsäuren und die von α -Aminosäuren," H. Leuchs and W. Geiger, *Ber.*, **41**, 1721 (1908).

"Ein Beitrag zur Bereitung von Peptiden," R. Schönheimer, *Hoppe-Seyler's Z. physiol. Chem.*, **154**, 203 (1926).

"Über ein allgemeines Verfahren der Peptid-Synthese," M. Bergmann and L. Zervas, *Ber.*, **65**, 1192 (1932).

"Synthese von Peptiden des D-lysins: D-lysyl-D-glutaminsäure und D-lysyl-L-histidin," M. Bergmann, L. Zervas, and J. P. Greenstein, *Ber.*, **65**, 1692 (1932).

"Neues Verfahren zur Synthese von Peptiden des Arginins," M. Bergmann, L. Zervas, and H. Rinke, *Hoppe-Seyler's Z. physiol. Chem.*, **224**, 40 (1934).

"Synthesis of Glutathione," C. R. Harington and T. H. Mead, *Biochem. J.*, **29**, 1602 (1935).

"A Synthesis of Glutathione," V. du Vigneaud and G. L. Miller, *J. Biol. Chem.*, **116**, 469 (1936).

"Studies in Synthetic Immunochemistry. III. Preparation and Antigenic Properties of Thyoxy Derivatives of Proteins, and Physiological Effects of their Antisera," R. F. Clutton, C. R. Harington, and M. E. Yuill, *Biochem. J.*, **32**, 1119 (1938).

"Synthese Penicillin-ähnlicher, äcylierter Dipeptide. I" (use of glycyl chloride hydrochloride), R. Neher, A. Wettstein, and K. Miescher, *Helv. Chim. Acta*, **29**, 1815 (1946).

"Poly-condensation of α -Amino Acid Derivatives. III. Poly-lysine," E. Katchalski, I. Grossfeld, and M. Frankel, *J. Am. Chem. Soc.*, **70**, 2094 (1948).

"Strepogenin Activity of Derivatives of Glutamic Acid," D. W. Woolley, *J. Biol. Chem.*, **172**, 71 (1948).

"A New Synthetic Route to Peptides," J. C. Sheehan and V. S. Frank, *J. Am. Chem. Soc.*, **71**, 1856 (1949).

"Synthesis of Peptides of Methionine and Their Cleavage by Proteolytic Enzymes," C. A. Dekker, S. P. Taylor, and J. S. Fruton, *J. Biol. Chem.*, **180**, 155 (1949).

"Studies in the Azole Series. Part XXVIII. A New Method of Peptide Synthesis: Alanyl-peptides," A. H. Cook and A. L. Levy, *J. Chem. Soc.*, **1950**, 651.

"The Synthesis of Simple Peptides from Anhydro-N-Carboxyamino-Acids," J. L. Bailey, *J. Chem. Soc.*, **1950**, 3461.

"Une nouvelle méthode de synthèse peptidique," R. A. Boissonnas, *Helv. Chim. Acta*, **34**, 874 (1951).

"Über eine neue Peptidsynthese," O. Sus, *Ann. Chem. Liebigs*, **572**, 96 (1951).

"Cleavage of N-Carbobenzoxy Groups by Dry Hydrogen Bromide and Dry Hydrogen Chloride," D. Ben-Ishai and A. Berger, *J. Org. Chem.*, **17**, 1564 (1952).

"The Use of Phosphite Amides in Peptide Synthesis," G. W. Anderson, J. Blodinger, R. W. Young, and A. D. Welcher, *J. Am. Chem. Soc.*, **74**, 5304 (1952).

"Tetraethyl Pyrophosphate as a Reagent for Peptide Synthesis," G. W. Anderson, J. Blodinger, and A. D. Welcher, *J. Am. Chem. Soc.*, **74**, 5309 (1952).

"Preparation of Arginyl Peptides," D. T. Gish and F. H. Carpenter, *J. Am. Chem. Soc.*, **75**, 5872 (1953).

"The Synthesis of an Arginyl Peptide," G. W. Anderson, *J. Am. Chem. Soc.*, **75**, 6081 (1953).

"Studies on Polypeptides. V. The Synthesis of Arginine Peptides," K. Hofmann, A. Rheiner, and W. D. Peckham, *J. Am. Chem. Soc.*, **75**, 6093 (1953).

"Acylchlorides of Amino Acids and Peptides as Monomers for the Preparation of Polymeric Polypeptides. I," M. Frankel, T. Liwschitz, and A. Zillha, *Experientia*, **9**, 179 (1953).

"The Synthesis of Oxytocin," V. du Vigneaud, C. Ressler, J. M. Swan, C. W. Roberts, and P. G. Katsoyannis, *J. Am. Chem. Soc.*, **76**, 3115 (1954).

"A New Method of Forming Peptide Bonds," J. C. Sheehan and G. P. Hess, *J. Am. Chem. Soc.*, **77**, 1067 (1955).

"Enzymatic Removal of the Protecting Group in Peptide Synthesis," R. W. Holley, *J. Am. Chem. Soc.*, **77**, 2552 (1955).

"Ethyl Thioltrifluoroacetate as an Acetylating Agent with Particular Reference to Peptide Synthesis," E. E. Schallenberg and M. Calvin, *J. Am. Chem. Soc.*, **77**, 2779 (1955).

"Peptide Derivatives Containing Hydroxyamino Acids," J. C. Sheehan, M. Goodman, and G. P. Hess, *J. Am. Chem. Soc.*, **78**, 1367 (1956).

"The Use of Phosphorous Acid Chlorides in Peptide Synthesis," R. W. Young, K. H. Wood, R. J. Joyce, and G. W. Anderson, *J. Am. Chem. Soc.*, **78**, 2126 (1956).

"Studies on Polypeptides. VIII. Synthesis of Peptides Related to Corticotropin," K. Hofmann, A. Jöhl, A. E. Furlenmeier, and H. Kappeler, *J. Am. Chem. Soc.*, **79**, 1636 (1957).

Amphoteric Properties of Proteins

The groups involved

Protein titration curves

General interpretation

Interaction effects

Configurational changes associated with protein titration

Masked acidic and basic groups

The earliest observation that proteins are capable of combining with (neutralizing) acids and bases dates back to the late nineteenth century. During the first years of the present century attempts were made to interpret these phenomena on the basis of surface adsorption through the then popular and ill-defined "colloid forces." But as early as 1904, Jacques Loeb suggested that such binding, and the consequent electrical charge on proteins, could be explained on a sound chemical basis, in terms of the known acidic and basic functional groups occurring in proteins. Through his classic studies, as well as those of such figures as Sørensen and Cohn, this concept has become firmly established and generally accepted. These studies, together with the sedimentation experiments of Svedberg, removed the proteins from the obscure class of colloids to the status of large molecules amenable to treatment from a sound chemical point of view.

THE GROUPS INVOLVED

Most of the amino acids in a protein molecule are linked through the α -peptide linkage, hence their α -amino and α -carboxyl groups are nonfunctional. In other words, they cannot take part in acid-base

equilibria. Nevertheless, all proteins contain acidic and basic groups, which may consist of:

- (1) A few terminal α -amino and α -carboxyl groups.
- (2) Possibly some α -amino and α -carboxyl groups of amino acids held by linkages other than the typical α -peptide bond.
- (3) Functional groups residing in the side chains of amino acid residues.

Of the three possibilities, the third is without doubt the most important one in most, if not all, proteins. In most cases, some terminal groups are probably also present, but usually in such small proportions that their titration behavior is masked by groups of class 3. As to groups of class 2, little evidence for their presence in proteins can be cited, but the possibility of there being such groups should be kept in mind.

TABLE 11-1. Reasonable pK Values for Titratable Groups in Proteins

Group	pK
Carboxyl (α)	3.0– 3.6
Carboxyl (aspartyl and glutamyl)	3.0– 4.7
Phenolic hydroxyl (tyrosine)	9.8–10.4
Phenolic hydroxyl (diiodotyrosine)	6.5
Sulphydryl	9.1–10.8
Imidazolium (histidine)	5.6– 7.0
Ammonium (α)	7.5– 8.4
Ammonium (ϵ , lysine)	9.4–10.6
Guanidinium (arginine)	12.0–13.0

The major groups contributing to the amphoteric behavior of proteins are listed in Table 11-1, together with reasonable pK values corresponding to each. These pK values have been inferred from actual analysis of protein titration curves and are subject to some uncertainty. In general, they are very close to the corresponding values observed in amino acids and peptides (Table 3-2).

PROTEIN TITRATION CURVES

The method of determination of protein titration curves is in essence identical with that used for simpler ampholytes such as amino acids and peptides, or, for that matter, for simple acids and bases. It is desirable to maintain a constant concentration of protein. Further, as will be seen later, the titration behavior of proteins is extremely sensitive to the presence of other ions. It is therefore important to maintain, as nearly as possible, constant ionic strength. The usual procedure is to add a definite amount of protein solution

to each of a series of tubes, to which are also added varying but carefully determined amounts of acid or alkali. The volume and ionic strength of each solution are then brought to a common value by appropriate additions of dilute salt solution (usually KCl), and the pH of each of these solutions is determined after careful mixing.

The determination of pH must be carried out with the greatest possible precision, not so much for its own sake as because the hydrogen ions bound are calculated by the difference between the amounts added and that free (as calculated from the pH). This may involve the taking of small differences between relatively large quantities. Formerly, careful titration curves were always based on measurements made with the hydrogen electrode. This method suffers certain experimental disadvantages, notably the fact that equilibration is rather slow, and in the case of proteins serious alterations in properties may take place during this equilibration period. The past decade has seen the development of glass electrodes of such precision that they are now normally employed instead of hydrogen electrodes.

To complete the electrochemical cell, a calomel electrode and a salt bridge are customarily employed. Such cells possess, in general, a contribution to the potential, owing to "transference" at the liquid-liquid junctions. It is customary to assume that this potential is not influenced by the presence of protein and is taken care of by the blank on water. This problem has been considered in a review by Tanford (1955).

One of the major uncertainties lies in the fact that the electrode measurement yields the "activity" of hydrogen ions, not the concentration. It is necessary, however, to infer concentrations in order to calculate the bound or liberated H⁺. The customary procedure is to carry out a blank titration on the system, in the absence of protein but *at the same ionic strength* as used in the protein solutions. The assumption is then made that the activity coefficient of H⁺ is the same in the protein solutions, at a given pH, as in the blank solutions. The implicit assumption is that the protein does not affect the activity coefficient of H⁺, an assumption that is almost certainly wrong. By working at relatively low protein concentration and at relatively high ionic strength, one feels some confidence that the small ions are substantially the sole determiners of activity coefficient. On the other hand, the lower the protein concentration, the less the binding and hence the smaller the difference between the protein solution and the blank which is to be used in calculation of the H⁺ bound. This increases the uncertainty in the result. One is forced to compromise on some intermediate protein concentration, and prefer-

ably should carry out titrations at several concentrations to ascertain the extent of agreement.

Protein titration curves are normally plotted as equivalents of H^+ bound (or liberated) per mole of protein, providing the molecular weight of the protein is known. In case the molecular weight is

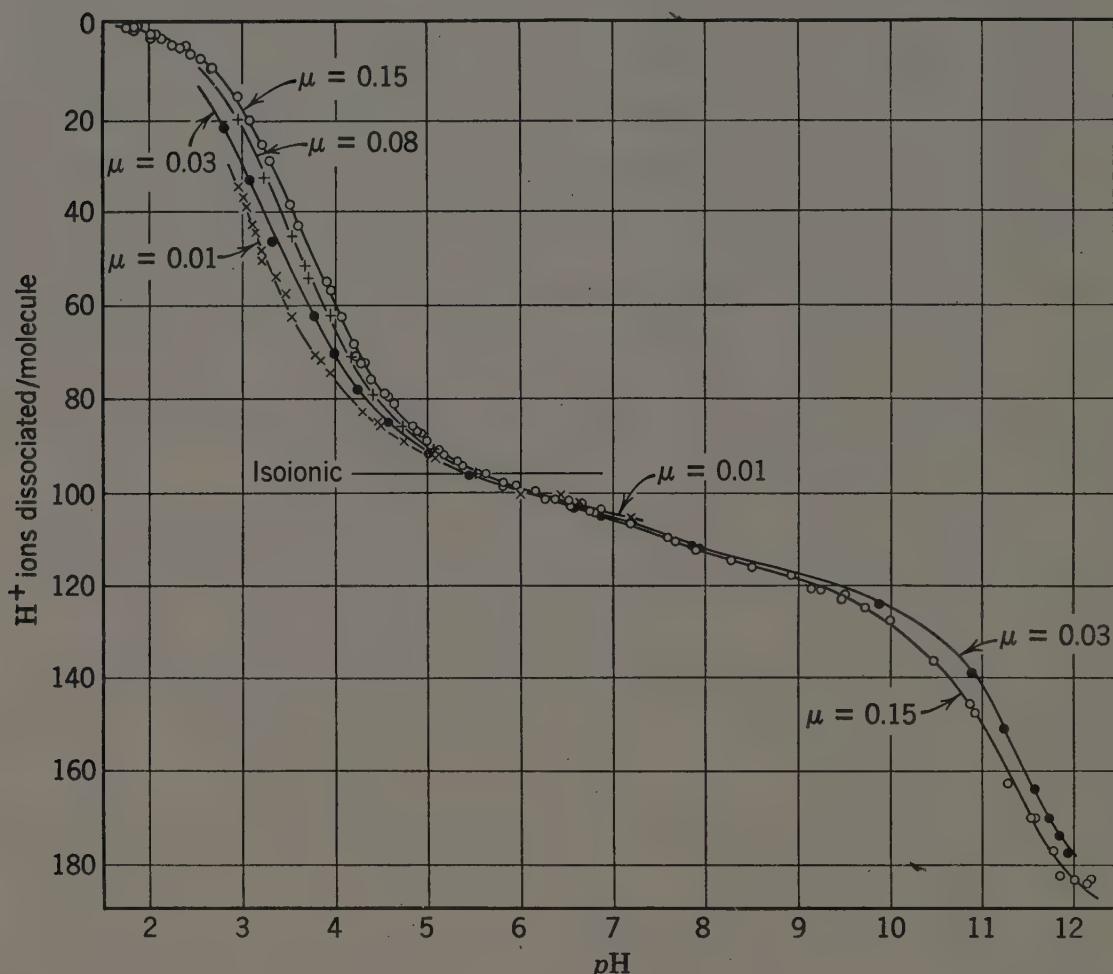


Fig. 11-1. Titration curves for bovine plasma albumin at four ionic strengths. From C. Tanford et al., J. Am. Chem. Soc., 77, 6414 (1955).

uncertain, it is common to plot equivalents bound (or liberated) per 100,000 g. of protein. The scale may be given as H^+ bound, relative to a zero which is taken as the protein in strong alkali (pH about 14), where binding is nil. More commonly it is given as H^+ liberated, starting from a zero value which corresponds to the protein in strong acid (pH about 1), where binding is optimal. (The latter limit can be determined with considerably more precision because of the great difficulty of pH measurement above pH 13, plus the lability of most proteins at such extreme alkalinity.) Even better is a plot of H^+ bound in solutions acid to the isoionic point, and of H^+ liberated alkaline to this point, the zero value corresponding to the isoionic

protein. This, however, necessitates a knowledge of the isoionic pH , a quantity which is not directly determinable from titration data alone but must be ascertained by separate experiments such as exhaustive dialysis or exchange deionization.

Figures 11-1 and 11-2 show some of the most reliable titration results yet published on two well-defined proteins, bovine plasma

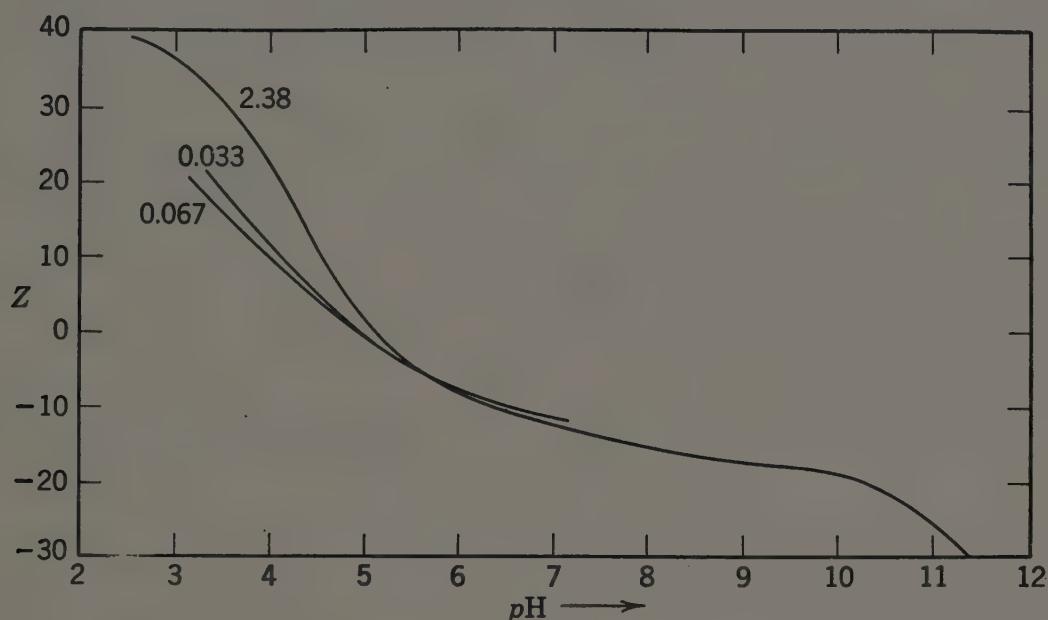


Fig. 11-2. Titration curves for ovalbumin at various ionic strengths. From R. K. Cannan et al., *Ann. N. Y. Acad. Sci.*, **41**, 243 (1941).

albumin and ovalbumin. It is clear that they have characteristics in common, notably a strong binding region in the pH range 3 to 5, a region of low buffering power from about 5 to 9, and another strongly buffering region above 9.

GENERAL INTERPRETATION

The qualitative and even the semiquantitative interpretation of protein titration curves is reasonably straightforward. The strongly buffering region between pH 3 and 5 can be attributed to the titration of carboxyl groups, largely those of aspartic acid and glutamic acid side chains. Most proteins are rich in such groups. The other strongly buffering region, above pH 9, corresponds to the titration of ammonium groups primarily, to a lesser extent sulphhydryl and phenolic groups (since most proteins contain relatively few such groups). The titration of guanidinium groups must also begin in this region, but doubtless is never completed at the pH at which most protein titrations are discontinued. (An appreciable proportion of the guanidinium groups must retain their positive charge at pH values as high as 13;

hence the great difficulty of ascertaining the limit of proton dissociation, previously mentioned.)

That these qualitative assignments are correct may be substantiated in a number of ways. Most convincing, perhaps, is the study of the effect of temperature on the titration curves. The dissociation of simple carboxylic acids is known to be relatively temperature independent; thermodynamically, the enthalpy or heat of ionization is close to zero. In agreement with this, the pH 3 to 5 region of protein titration curves is little affected by temperature. The alkaline portion, on the other hand, is markedly affected, as is to be expected from the large positive heat changes normally associated with the dissociation of protons from ammonium and guanidinium groups. Furthermore, it is possible to mask certain groups preferentially by chemical modification; for example, amino groups can be acylated. Such alterations have, qualitatively at least, the expected effect on the titration curve. Finally the effect of formaldehyde on protein titration curves is seen in the pH 8 to 10 region, as is to be expected on the basis of considerations presented in Chap. 3 (formol titration).

The complexity of a protein titration, there being involved typically 50 to 100 or more titratable groups per molecule, renders complete interpretation and analysis out of the question. The earliest attempts at quantitative correlations, which helped to put the chemical interpretation of protein amphotericity on a firm basis, were concerned with the magnitude of the binding limits. These limits were termed the maximum acid-binding and base-binding capacities.

At very low pH, 1.0 or below, the binding curve approaches a maximum and levels off. This maximum, if measured from a zero reference point at the isoionic pH, corresponds to the maximum number of excess protons with which the protein can combine over and above those combined at the isoionic point. In absence of any combination with ions other than protons, it represents the maximum positive charge which the protein can attain. The value should thus correspond to the number of cationic groups, or the number of groups which are uncharged in the basic form; in other words, should correspond to the number of so-called basic groups or basic amino acids.

Table 11-2 summarizes some comparative results on maximum acid-binding capacity and basic amino acid content for several proteins. The correlation is seen to be reasonably satisfactory in most cases. At the other extreme, the maximum "base-binding" capacity or maximum negative charge should correspond to the number of uncharged acid groups in the protein ($-\text{COOH}$, $-\text{SH}$, $-\text{OH}$). Correlations in

this case have never been very satisfactory, partially because of the aforementioned difficulty of carrying out protein titration curves to high enough alkalinity, partially because of uncertainty as to the extent to which carboxyl groups in a protein are free and to what extent present in the amide form.

TABLE 11-2. Comparison of Maximum Acid-Binding Capacity of Proteins with Content of Basic Amino Acids¹

Protein	Equivalents of Basic Amino Acids per 100,000 gm. (analysis)	Equivalents of Protons Bound per 100,000 gm.
Ovalbumin	76	80-87
Insulin	95	101
Zein	15	18-21
Casein	80	76-90
Edestin	128	134
Hemoglobin	125	134-148
Serum albumin (human)	144	145
Lysozyme	125	133

¹ Taken in part from a compilation by E. J. Cohn in Cohn and Edsall's *Proteins, Amino Acids and Peptides*, p. 354, Reinhold, New York, 1943.

A rough analysis of the total titration curve may be attempted by breaking down the titratable groups into a few, perhaps four or five, classes and assigning definite assumed *pK* values to the groups of a given class (such as the *pK* values given in Table 11-1). If one is willing to assume complete independence of behavior on the part of each group (an assumption which is by no means safe in the case of protein, as will be seen), the titration of a large number, *n*, of identical groups in a given protein molecule can be handled by the same methods discussed in Chap. 3 for simple acids and bases. At a given *pH*, a fraction α of the carboxyl groups, for example, will be in the ionized (COO^-) state, and a fraction $1 - \alpha$ in the undissociated (COOH) form. The simple relation

$$\text{pH} = \text{p}K + \log \alpha / (1 - \alpha) \quad (1)$$

analogous to equation 14 of Chap. 3 is then applicable. At any given *pH*, then, it is possible to compute that fraction of each class of groups which is in each form, and sketch a theoretical titration curve. By a method of successive approximation, varying the assumed numbers of groups in each class and also varying the assigned *pK*'s somewhat, it is possible to fit the experimental curve. Some results

of this type are given in Table 11-3. Agreement with analytical data is seen to be quite satisfactory.

This treatment is equivalent to disregarding the fact that n ionizing groups are attached to the same molecule, and considering them as n entirely independent small molecules or ions. An alternative but entirely equivalent approach is to consider that the large molecule contains n equivalent sites, having the same *intrinsic* dissociation (or association) constant. In this case the ionization of successive equivalent groups will have to be corrected for statistical factors.

TABLE 11-3. Agreement of Protein Titration Data with Analytical Data on Amino Acid Composition¹

Group	Serum Albumin ² (Human)		Oval- bumin ³		Insulin ⁴ (Zinc-Free)		Lysozyme ⁵	
	A	T	A	T	A	T	A	T
α -Carboxyl	106	106	46	51	4	12.5	1	10.5
β - plus γ -carboxyl					8-8.5		1.3-6.5	
Imidazole	16	16	4	5	4	4	1	1
α -NH ₂	9	4	15	22	4	4	1	1
ϵ -NH ₂	58	56			2		5.5-5.8	5-6
Phenolic	18	18	8	10	2.8-2.9	3
Guanidinium	24	24	14	14	2	(2)	10-12	11-12

¹ Data given in moles of amino acid per mole protein. In each case the analytical figure is given under column A, the value deduced from the titration curve under column T.

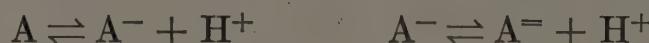
² C. Tanford, *J. Am. Chem. Soc.*, **72**, 441 (1950).

³ R. K. Cannan, A. Kibrick, and A. H. Palmer, *Ann. N. Y. Acad. Sci.*, **41**, 243 (1941).

⁴ C. Tanford and J. Epstein, *J. Am. Chem. Soc.*, **76**, 2163 (1954).

⁵ C. Tanford and M. L. Wagner, *J. Am. Chem. Soc.*, **76**, 3331 (1954).

Consider a dibasic acid in which the two acidic functions are entirely separate and independent (a situation which is approached in the case of very long α - ω -dicarboxylic acids). In this case the two K 's corresponding to the two ionizations



are found to be related by the factor 4, or

$$K_1 = 4 \times K_2$$

The reason for this is that the ways in which the first dissociation can take place are two as compared to only one for the second dissociation, introducing a factor 2 into the ratio. Further, there are two ways

that protons can recombine with A^- to give A^- , but only one way in which combination can take place with A^- , introducing another factor of 2. In the case of a large number of equivalent sites, for example 100, there are 100 ways in which the first ionization can take place, as compared to 99 for the second. Moreover, there are two recombination pathways for the second as compared to one for the first. Thus

$$K_1 = \frac{100 \times 2}{99} K_2$$

or in terms of an arbitrary number of sites n ,

$$K_1 = (2n/n - 1)K_2 \quad (2)$$

In terms of an intrinsic constant K_0 ,

$$K_1 = nK_0$$

$$K_2 = [(n - 1)/2]K_0$$

$$K_3 = [(n - 2)/3]K_0$$

$$K_n = K_0/n$$

or in general

$$K_i = [(n - i + 1)/i]K_0 \quad (3)$$

This approach has not been much used in the interpretation of protein titration curves, but it has been very popular and useful in the case of interpretation of binding of other ions by proteins, as will be seen in Chap. 14.

As a consequence of the large number of binding equilibria, a number of charge species always exist in a protein solution under any given set of conditions. This is graphically illustrated in Fig. 11-3, where the charge distribution for a hypothetical protein containing 100 carboxyl groups has been plotted at various acid pH values on the basis of certain assumed electrostatic interactions (see p. 185). For simplicity it has been assumed here that all groups other than carboxyl are in their acid form at the isoionic pH , and that none of the carboxyls is in its acid form at this pH . At any pH a rather broad distribution of forms exists, particularly toward the mid-point of the titration range. This by no means expresses the full heterogeneity, since there are many isomers for each particular charge form. It should be emphasized that such considerations apply also at the isoionic point. By no means are all of the protein molecules isoionic at all times at this pH ; it is only the average charge which is zero. These considera-

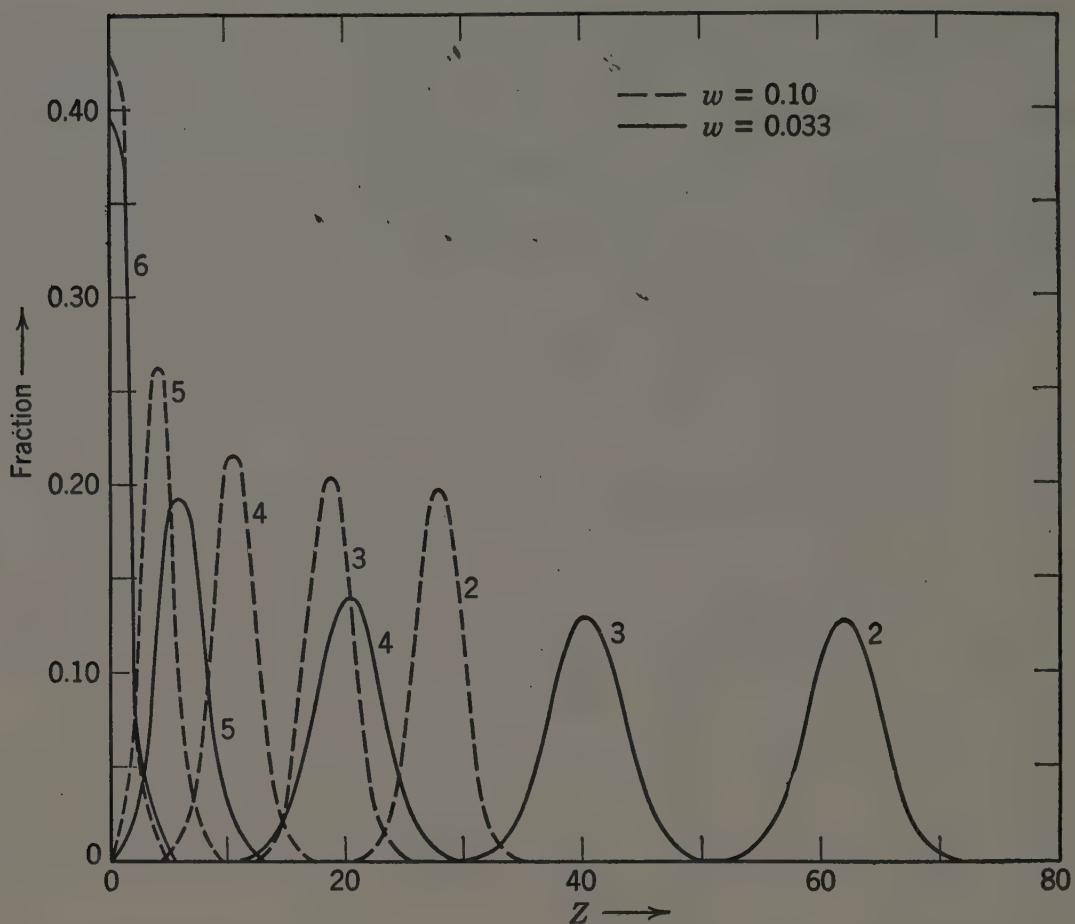


Fig. 11-3. Theoretical distribution of charge species for an assumed model approximating the composition of bovine plasma albumin, in the acid titration region. Distributions are given for two values of the electrostatic interaction parameter w (see text). Numbers on curves denote pH values.

tions have been discussed in a particularly lucid fashion by Edsall (1943), who calculates, for example, that in the case of hemoglobin some 350 million species exist in reasonable amount at the isoionic point.

INTERACTION EFFECTS

As was previously mentioned, assumption that the various titratable groups of a given class in a protein molecule are free and independent is a gross oversimplification. Interaction between groups can take place in at least two general ways: (1) through inductive interactions, and (2) through electrostatic interactions.

Inductive effects arise through subtle shifts in electronic configuration in molecules, which take place upon alterations in chemical substitution or state of ionization. They are of great importance in aromatic compounds but generally of much less importance in saturated aliphatics. The presence of several single-bonded carbon atoms in a molecule between two functional groups is generally considered to

rather thoroughly insulate them from one another so far as inductive effects are concerned. To date there is little evidence to indicate the extent to which such interactions might be of importance in protein molecules. To be sure, proteins are essentially aliphatic in nature, but there is good evidence of an appreciable degree of double-bond character in the C-N linkage of the peptide chain, as will be mentioned in Chap. 16.

Electrostatic effects, on the other hand, are of great importance in modifying the acid-base equilibria of proteins. That this must be so is apparent when one considers that at extreme *pH* there may be of the order of 100 or so excess positive or negative charges packed in the volume of the protein molecule, a volume of only a few thousand cubic angstroms.

Detailed consideration of these interactions would necessitate a full knowledge of the size and shape of the protein molecule and the precise location of each titratable group therein. Such information is, unfortunately, not yet attainable. Nevertheless, some most interesting calculations with respect to such interactions have been carried out,* leading to important deductions as to the behavior of proteins at extreme *pH*. The treatment is based on the notion that the protein can be regarded as a sphere with the titratable groups distributed completely at random over its surface.

By fundamental thermodynamic theory, the standard free energy of ionization, ΔF° , is related to the equilibrium constant for that ionization through the relation

$$\Delta F^\circ = -RT \ln K = 2.303RTpK \quad (4)$$

The free energy change may be considered to be composed of two parts, an intrinsic component and an electrostatic one. The latter may be considered as arising from the electrostatic attraction or repulsion (depending on the sign of the charge) of the protein of charge *Z* for the proton. Thus

$$\Delta F^\circ = \Delta F_{\text{int}}^\circ + \Delta F_{\text{elec}}^\circ$$

The magnitude of the electrostatic term has been derived, on the basis of the Debye-Hückel theory of ionic solutions, and on the basis of the preceding assumptions concerning the protein structure, as proportional to the square of the charge through the relation

$$\Delta F_{\text{elec}}^\circ = (Z)^2 RT w \quad (5)$$

* For further details the reader is referred to the excellent publications of Tanford (1954, 1955).

where

$$w = \frac{N\epsilon^2}{2DRT} \left(\frac{1/b - \kappa}{1 + \kappa a} \right)$$

In these equations, N and ϵ are the Avogadro number and the charge on the electron (or proton), hence are universal constants; a and b are properties of the protein, b being the radius of the (spherical) molecule itself and a its "radius of exclusion" (a value slightly larger than b but, in the case of a protein, not significantly different); and D and κ are properties of the solution, D being the dielectric constant and κ a parameter of great importance in the Debye-Hückel theory, governed by the ionic strength for all practical purposes.

In a protonic equilibrium in which a protein of negative charge Z loses a proton to give a charge $Z + 1$, the $\Delta F_{\text{elec}}^\circ$ term will be the difference between that appropriate to the two forms, that is

$$-\Delta F_{\text{elec}}^\circ = [(Z + 1)^2 - Z^2]RTw = (2Z + 1)RTw$$

Or if the corresponding equilibrium constant is termed K_z ,

$$pK_z = -(2Z + 1)/2.303w + \Delta F_{\text{int}}^\circ/2.303RT$$

By the same symbolic notation, pK_0 would correspond to the protonic dissociation in which the product has a charge $Z = 1$ and the reactant $Z = 0$, so that by the same method

$$pK_0 = -1/2.303w + \Delta F_{\text{int}}^\circ/2.303RT$$

Therefore,

$$pK_z = pK_0 - 2Zw/2.303$$

This pK_z may now be utilized in equation 1 to give

$$\text{pH} = pK_z + \log \alpha/(1 - \alpha)$$

or

$$\text{pH} = pK_0 + \log \alpha/(1 - \alpha) - 2Zw/2.303 \quad (6)$$

By means of this equation it is possible to fit protein titration curves in a more precise manner than with equation 1. In a given titration range where substantially only one type of dissociable group is functional, in addition to the two previously mentioned variables of pK (now pK_0) and number of groups n , one has the variable w with which to work. It has been found that the w values so obtained are entirely reasonable, that is, they correspond to reasonable values of the molecular radius b . Since w varies with ionic strength through the κ term, it is possible to account very nicely for the observed effect of ionic strength on titration curves of some proteins.

Figure 11-2 gives examples of the ionic-strength effect in the case of ovalbumin, for which this simple electrostatic theory appears to be adequate. In other cases there are added complications, due to binding of ions other than protons. Anions, even Cl^- , are bound in appreciable quantities by some proteins, notably serum albumin. In such cases the charge Z can clearly not be inferred directly from the proton binding. In particular, on the acid side the charge increases much less rapidly than in proportion to the proton binding, so that the electrostatic correction term is less than would be inferred from theory. The matter of the binding of such ions is considered further in Chap. 14.

CONFIGURATIONAL CHANGES ASSOCIATED WITH PROTEIN TITRATION

Another important and interesting deduction has come out of such considerations of electrostatic effects in protein titration behavior. In some cases, notably with human and bovine serum albumin, it is found that a constant value for the w term is not adequate to fit the shape of the acid titration limb of the curve. The deviation is in such a direction as to suggest that the electrostatic correction term does not build up in proportion to Z as expected. A part of this effect is due to Cl^- binding, but reasonable corrections for this effect by no means suffice, the calculated values of w decreasing strongly with decreasing $p\text{H}$ even after this correction.

One plausible interpretation which has been offered is that the protein molecule expands with decreasing $p\text{H}$ (increasing charge), with the result that radius b increases, decreasing the w term. Intrinsic viscosity studies now offer excellent independent evidence for such an expansion of serum albumin. These studies suggest an expansion in volume of the protein molecule of some 20-fold between the isoionic point and $p\text{H } 2$. Expansion is favored by decreased dielectric constant of the medium and repressed by increasing ionic strength, suggesting electrostatic repulsion to be the motivating force. Such changes, which in the case of serum albumin appear to be rapidly and completely reversible, introduce additional complications into the interpretation of titration data.

MASKED ACIDIC AND BASIC GROUPS

We have seen that, in general, titration curves can be interpreted, at least in a semiquantitative manner, in terms of the known amino acid composition of the proteins in question. This implies that all of the acidic and basic groups present in the amino acid side chains are

exposed and in equilibrium with the hydrogen ions in the aqueous environment. Strictly speaking, application of the electrostatic equations given requires that such groups be actually projecting out of the surface into the aqueous environment in order that the dielectric constant D may be taken as that of water.

While this situation may in general be approximately true, it is by no means always so. It is known, for example, that the phenolic groups of ovalbumin cannot be titrated except at pH much higher than that expected, and also that this ionization is accompanied by irreversible denaturation (structural alteration) of the protein. This suggests that these groups are buried in the molecule or at least strongly bonded in some way (see Chap. 16).

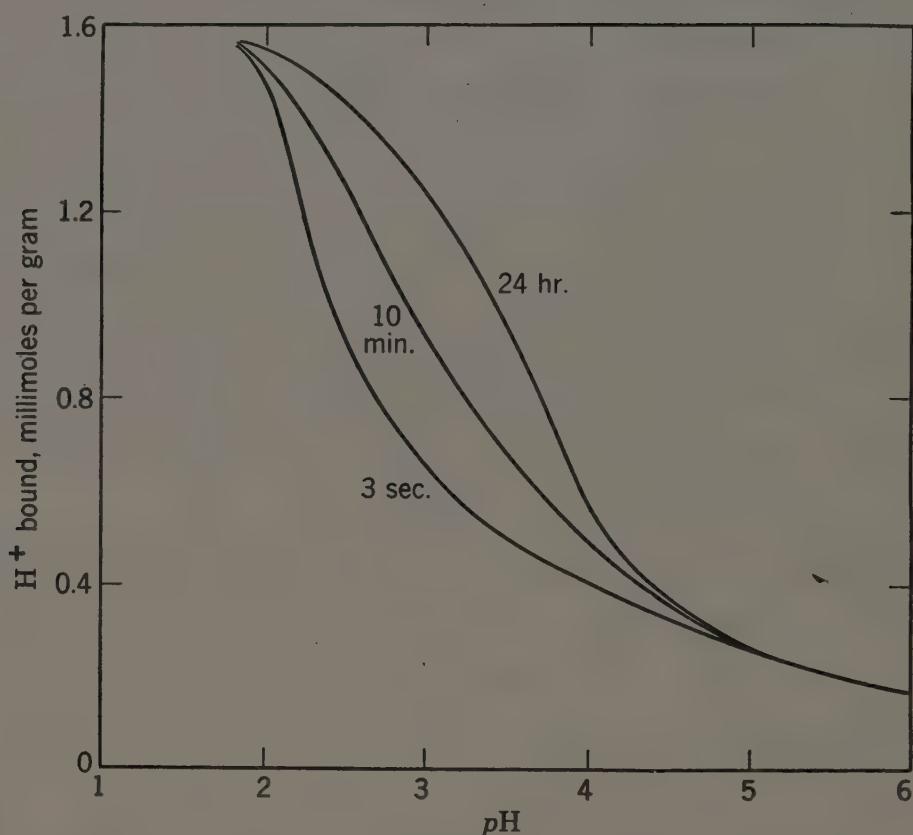


Fig. 11-4. Titration curves on horse carbonylhemoglobin as a function of time, demonstrating slow unmasking phenomenon. Redrawn from Steinhardt and Zaiser, *J. Biol. Chem.*, 190, 197 (1951)

Another interesting case is that of zinc insulin, where titration is only sluggishly reversible. In this case titration in the upward pH direction yields a different curve than that obtained on back titration when the protein is first brought to high pH and then titrated with acid. The explanation doubtless lies in the blocking of certain groups, probably imidazole, by the bound zinc. The fact that insulin is insoluble in water over the pH range 4 to 7 is also involved; however, zinc-free insulin yields a normal reversible titration curve even through

the region of insolubility, indicating that the precipitated phase is loose and permeable to ions.

Particularly interesting in this connection are certain derivatives of horse hemoglobin, carbonylhemoglobin, and ferrihemoglobin, elucidated so beautifully through the efforts of Steinhardt and Zaiser. The key to this elucidation is the application of a rapid-flow technique which permits measurement of *pH* within seconds after mixing protein and acid. In the first place, it was observed that the acid titration limb of carbonylhemoglobin, determined in the conventional manner, is much steeper than that of other typical proteins but similar to the curves given by plasma albumins. There is also a characteristic shift in the absorption spectrum of the colored protein in acid, and the titration curve is not fully reversible, the back titration curve not coinciding with that obtained by direct titration with acid.

Figure 11-4 shows the titration curves obtained at various time intervals after mixing. It is seen that on adding acid there is a progressive increase in the binding of protons, which takes place at a readily measurable rate. This increase corresponds to the liberation of 36 additional basic groups, which are evidently masked at the isoionic point, half of the total available for combination with protons in the acid range. These groups are unmasked through a slow structural rearrangement.

Additional aspects of this important work are discussed later under the subject of protein denaturation. It need only be pointed out here that these findings open up important new vistas in the area of protein amphoteric behavior. There is no reason to believe that such effects are restricted to the hemoglobin derivatives. Indeed, it seems probable that a part of the anomaly in the low *pH* titration of plasma albumin is due to a similar structural alteration.

SELECTED REFERENCES

Books and Reviews

"Proteins as Acids and Bases," J. T. Edsall, Chap. 20 in Cohn and Edsall's *Proteins, Amino Acids and Peptides*, Reinhold, New York, 1943.

Electrometric pH Determination, R. G. Bates, John Wiley & Sons, New York, 1954.

"Hydrogen Ion Titration Curves of Proteins," C. Tanford in *Electrochemistry in Biology and Medicine* (T. Shedlovsky, ed.), John Wiley & Sons, New York, 1955.

"Hydrogen Ion Equilibria in Native and Denatured Proteins," J. Steinhardt and E. Zaiser, *Advances in Protein Chem.*, **10**, 152 (1955).

Original Articles

- "The Relationship between the Composition of Zein and Its Acid and Basic Properties," E. J. Cohn, R. Berggren, and J. Hendry, *J. Gen. Physiol.*, **7**, 81 (1924).
- "The Amphoteric Properties of Egg Albumin," R. K. Cannan, A. Kibrick, and A. H. Palmer, *Ann. N. Y. Acad. Sci.*, **41**, 243 (1941).
- "Hydrogen Ion Equilibria in Native and Modified Human Serum Albumins," C. Tanford, *J. Am. Chem. Soc.*, **72**, 441 (1950).
- "Masking of Acid-Binding Groups in Native Horse Carbonylhemoglobin," J. Steinhardt and E. Zaiser, *J. Biol. Chem.*, **190**, 197 (1951).
- "Thermodynamic Consideration of Protein Reactions. I. Modified Reactivity of Polar Groups," M. Laskowski, Jr., and H. Scheraga, *J. Am. Chem. Soc.*, **76**, 6305 (1954).
- "Hydrogen Ion Titration Curve of Crystalline Zinc Insulin. The Nature of Its Combination with Zinc," C. Tanford and J. Epstein, *J. Am. Chem. Soc.*, **76**, 2170 (1954).
- "Phenolic Hydroxyl Ionization in Proteins. II. Ribonuclease," C. Tanford, J. Hauenstein, and D. Rands, *J. Am. Chem. Soc.*, **77**, 6409 (1955).

Electrophoresis of Proteins

Simple theory

Direct observation

Moving-boundary method

Zone electrophoresis

Reversible boundary spreading

Of the many physicochemical techniques which have contributed to our present state of knowledge of proteins, electrophoresis occupies a position of primary importance. This technique provides, within certain limits, a method for analyzing complex protein mixtures both qualitatively and quantitatively, a fact which has made possible the fractionation of such systems (Chap. 15). In addition, electrophoresis enables the separation of closely related proteins which are not readily separable by the more conventional fractional precipitation techniques. A third general application is receiving more and more attention, namely, the use of electrophoresis for studying the interaction of proteins with one another and with small ions.

SIMPLE THEORY

Fundamentally the method is based on the fact that proteins in solution are usually charged and hence migrate in an electric field. It is a well known fact that different ions, for example hydroxyl and chloride, migrate at different velocities in an electric field of given strength. Such differences of *electrophoretic mobility* exist also, and to an even greater extent, in large ions such as proteins. This mobility depends directly on the magnitude of the net charge on the protein ion,

and inversely upon the frictional resistance exerted on the protein by the solvent.

The latter effect depends in a complex manner on the size and shape of the protein molecule and the extent to which it is solvated; the greater its size and the more asymmetric its shape, the greater will be the frictional resistance and hence the lower the mobility, other factors being equal. The problem of the frictional resistance is considered in somewhat more detail in Chap. 13, in connection with sedimentation, diffusion, and viscosity of protein solutions, where it will be seen that the problem is a very complex one.

The magnitude of charge on the molecule depends, above all else, on the *pH* of the solution, as would be deduced from the discussion of the previous chapter. A general correlation between electrophoretic mobility and the titration curve would therefore be expected. For any protein there exists a *pH* of zero mobility, the isoelectric point (p. 30). At lower *pH* values the protein migrates positively, or as a cation, the mobility increasing with decreasing *pH*. Conversely, above the isoelectric point the protein migrates negatively, or as an anion, the negative mobility increasing with increasing *pH*.

The net charge on the protein can be inferred from the titration curve only to a first approximation. Of considerable importance in this connection is the extent to which the protein molecule binds other ions in the solution. This problem was mentioned briefly in the preceding chapter and is discussed in Chap. 14. Since anions are in general bound much more tenaciously by protein molecules than are cations, at least monovalent cations, it is to be expected that the actual net charge on the acid side of the isoelectric point will usually be less than that inferred from titration data.

Given a detailed knowledge of the net protein charge as a function of *pH*, it might appear possible to infer the magnitude of the frictional resistance from mobility versus *pH* data. Alternatively, if the frictional resistance were also known from independent hydrodynamic studies, it should be possible to calculate mobilities for comparison with experimental results. Unfortunately, no great success has as yet been attained in this direction because of the complexity of the actual relationship between charge and mobility. A number of theories exist for various assumed situations, probably no one of which ever applies in detail for a given protein.

On the basis of fundamental electrostatic principles, the force exerted on an ion in an electric field depends not on the absolute value of the potential but on the potential gradient, or the rate of change of potential with respect to position. This is given generally

by the derivative dE/dX , where X is the position or distance, but in a homogeneous solution the derivative can be replaced simply by E/X . The actual force exerted is equal to the product of this gradient and the charge on the ion; that is,

$$F = QE/X \quad (1)$$

The frictional resistance is proportional to the velocity V through the proportionality constant f , which is a function of the various factors just enumerated. The velocity V attained will be such that the two forces, that of the electrostatic field and that of resistance, are equal so that

$$fV = QE/X$$

or

$$V = (Q/f)(E/X) \quad (2)$$

The electrophoretic mobility μ is defined as the velocity attained in a unit potential gradient, that is, V divided by the gradient. Thus

$$\mu = V/(E/X) = Q/f \quad (3)$$

For an ideal sphere, that is, one which is rigid, compact, and unsolvated, f is given by the Stokes-Einstein relation as

$$f = 6\pi\eta r.$$

In general (see Chap. 13) f may be related to the diffusion coefficient D of the protein, through the relation

$$f = kT/D$$

This simple theory is based on the assumption that the protein can be considered as a simple isolated ion acting completely independently of other ions in the system. This requires that it be far removed from all other ions, a situation which exists only at vanishingly small concentrations and at substantially zero ionic strength. In an actual situation, the protein ion in question is surrounded by other ions, including other protein ions and particularly small ions such as H^+ or OH^- , and usually various buffer ions. These too will be subject to the field, and will tend to carry small volumes of solvent with them because they are solvated. On the average the protein molecule will find more ions of opposite charge in its environment than ions of like charge. In effect, it is immersed in a cloud of opposite charge, moving slightly in the opposite direction and tending to retard its motion. Various equations exist for attempting to take this factor into account,

but they will not be detailed here.* It is sufficient to point out that the idealized mobility will usually be attained only at zero ionic strength, and the actual mobility will decrease with increasing ionic strength.

The many experimental devices which have been utilized in studying the electrophoretic behavior of proteins can be grouped into three categories: (1) direct observation, (2) the moving-boundary method, and (3) zone electrophoresis. These are discussed in order of their historical development.

DIRECT OBSERVATION

This method is applicable to particles which can be seen with the aid of a microscope, that is, to gross suspensions. Migration is carried out in a suitable cell mounted on the stage of a microscope, and the velocity of the particles is determined directly by timing them. This method is not particularly useful for either analysis or separation. It has proved of most value in determining the manner in which the mobility of the particles (hence the charge if the size and shape remain constant) varies with pH. Microorganisms and other living cells can be studied nicely by this means. It is possible to study proteins, themselves invisible, by adsorbing them on the surface of microscopic particles, in which case the charge on the particles becomes primarily a function of that on the proteins. This method has been largely supplanted in protein studies, however, by the other methods to be discussed.

MOVING-BOUNDARY METHOD

General Principles of the Moving-Boundary Method. It is much more convenient and meaningful to study proteins in solution. If a sharp boundary is formed between the protein solution (usually containing buffer salts in addition) and a buffer solution, this boundary will in general migrate upon application of a potential gradient. Since it is important that the ionic composition of both solutions be as nearly identical as possible, the protein solution is usually first equilibrated with the buffer through a semipermeable membrane (dialysis). The direction of migration will depend on the sign of the charge on the protein molecules, which in turn depends primarily on the pH. If the protein solution contains more than one electrophoretic component (that is, components having differing electrophoretic mobilities), upon migration the originally homogeneous boundary will

* For an excellent discussion of the problem of electrophoretic mobility in general, the book by Abramson, Moyer, and Gorin (1942) is recommended.

separate into two or more boundaries, depending on the number of components.

Such a separation is indicated schematically in A and B of Fig. 12-1, where three components are assumed, *a*, *b*, and *c*, whose mobilities decrease in that order. Behind the advancing *a* boundary is a region

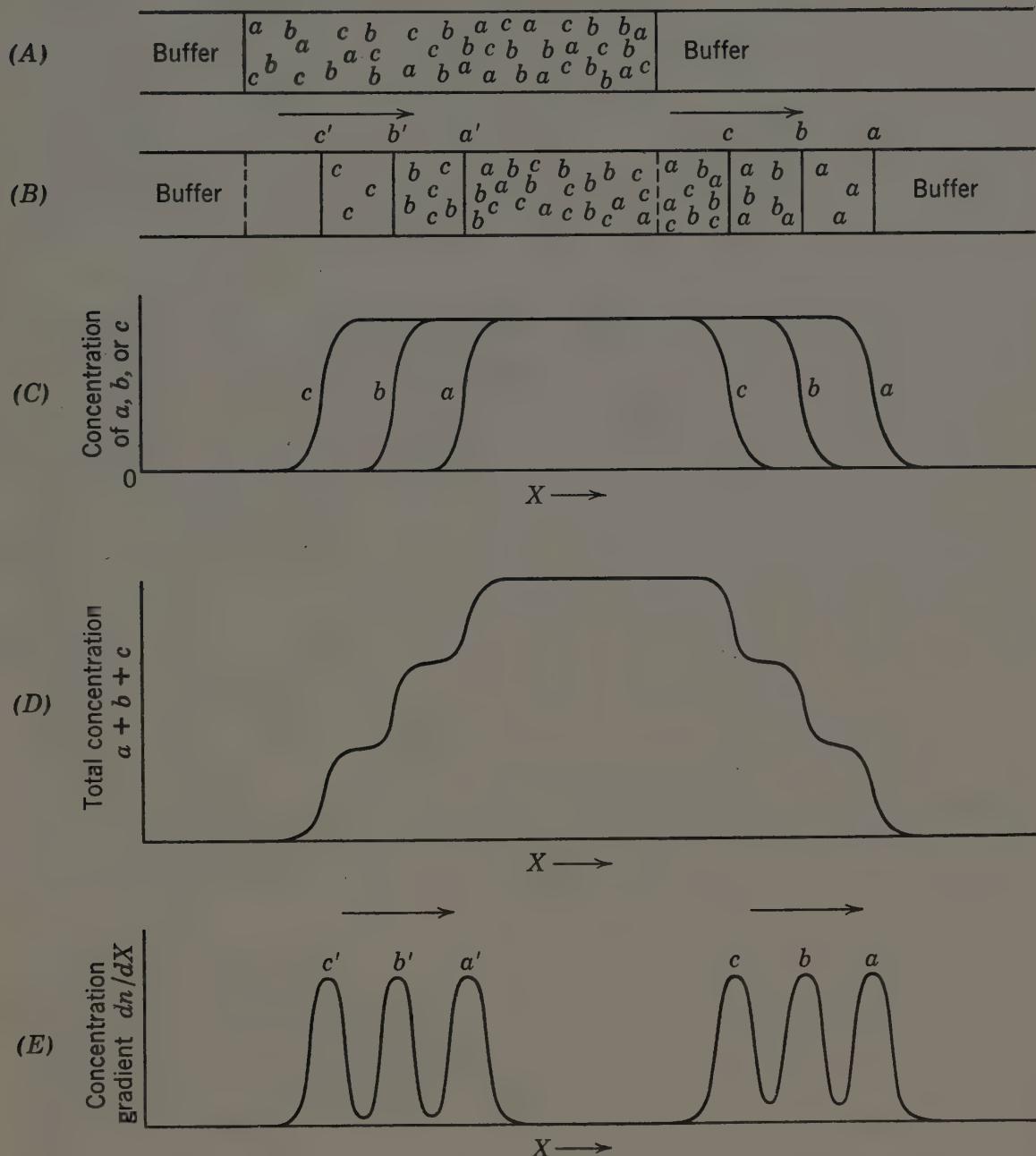


Fig. 12-1. Schematic representation of the electrophoretic separation of three protein components, *a*, *b*, and *c* (mobility decreasing in that order). Components are assumed to be present in roughly equal weight quantities.

containing pure *a*; behind the *b* boundary there is a mixture of *a* and *b*, and so on. Inspection of the figure will show that the fastest component, *a*, could be isolated in pure form behind the advancing *a* boundary; the slowest component, *c*, could be isolated behind the

receding b boundary, noted as b' . Components of intermediate mobility could not be isolated in a pure state.

In C of Fig. 12-1 the concentration of each of the three components is plotted as a function of X , the position along the cell. In D the summation of these concentrations is plotted, giving total protein concentration as a function of X . Finally, E shows the first derivative (without regard to sign) of the protein concentration with respect to X ,* in other words the concentration gradient, plotted as a function of X . Perfectly sharp boundaries can never be attained, and even if formed they would rapidly go over to the type of curves illustrated in D , owing to diffusion. The first derivative of such curves is simply the standard curve of error, or Gaussian distribution curve (see Chap. 13).

It is not practicable to carry out electrophoretic experiments in a horizontal cell such as was illustrated, for convenience, in Fig. 12-1. Because of the difference in density that usually exists between the two solutions forming the boundary, convection would result, with consequent mixing of the two solutions and destruction of the boundaries. The boundaries must therefore be horizontal, with the less dense solution always overlying the solution of greater density. Since the protein solution is invariably more dense than the buffer, the latter must lie above the boundary. This requires the use of a U-shaped cell. The advancing boundary now becomes an "ascending" boundary, the receding boundary "descending." These are the terms used to designate the two electrophoretic boundaries.

Observation of Boundaries. There remains the problem of observing the position and estimating the magnitude (or the change in protein concentration) of the various boundaries. If the protein components were colored, it would be a simple matter to examine the light absorption at various heights in the cell and calculate the protein concentration curve. Such a method has actually been used, for example with hemoglobin. Most proteins are not colored, however, and to make the procedure really useful a more general method of observation is needed. Such a method was introduced into electrophoretic analysis by Tiselius in 1937. This method, based on the Toepler *schlieren*† principle, had been adapted previously to the study of boundaries in diffusion

* The expression "first derivative with respect to X ," for those readers without a knowledge of calculus, means the rate at which the concentration changes with change in position X . Thus the value of the ordinate on the concentration gradient curve in E of Fig. 12-1 is just the *slope* of the curve in D at the corresponding position X .

† After the German for "streaks," for reasons which will be obvious from the discussion.

measurements by Lamm, and of sedimentation boundaries by Svedberg (Chap. 13).

Before considering the true schlieren techniques it might be well to consider briefly a somewhat simpler technique, namely scale-line displacement. In this technique a transparent scale, placed just ahead of the cell to be examined, is photographed through the cell. If the solution in the cell is homogeneous and the cell is optically perfect, an undistorted image of the scale results. A refractive gradient in the cell leads to a distortion of the image, the scale lines in the vicinity of the boundary being shifted in position. The magnitude of this displacement can be shown to be directly proportional to the gradient of the refractive index, which in turn is very nearly proportional to the gradient in concentration, so that a plot of scale-line displacement versus undisplaced position gives directly a diagram of the type shown in *E* of Fig. 12-1. This method, potentially very accurate, has been largely supplanted. Unfortunately, it requires much tedious work in measuring the line positions, hence is not desirable for routine electrokinetic analyses.

The schlieren techniques can be understood best, perhaps, by an examination of Fig. 12-2. Light from a horizontal slit *S* is focused

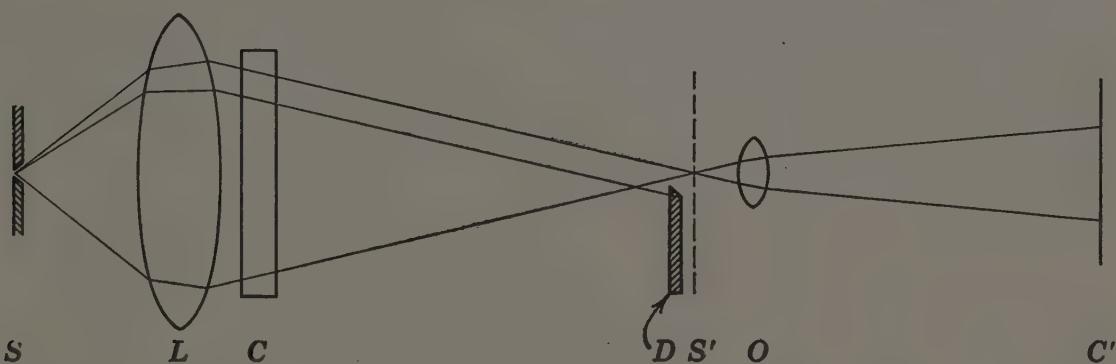


Fig. 12-2. Schematic representation of the optical arrangement for formation of the schlieren pattern. *S*, horizontal slit; *L*, schlieren lens; *C*, cell; *D*, diaphragm; *S'*, focal plane for slit *S*; *O*, objective lens which focuses cell on the film or ground glass at *C'*. For the formation of the scanning pattern, the diaphragm *D* is moved vertically while the film is moved laterally (in or out of the plane of the paper).

at *S'* by means of a lens *L* of large aperture. The cell *C* is placed just following the lens. If the solution in the cell is homogeneous, it does not affect the light, and a sharp image of *S* results in the plane *S'*. If, on the other hand, a refractive gradient exists, light coming through this region will not focus at *S'* but will be deflected (in general downward, since the solution which is more dense gravitationally is usually also more dense optically). The undeflected light passes

through the objective lens O of the camera, which forms an inverted image of the cell on the camera film C' . Obviously, if the deflected light is intercepted, there will be a resulting dark region (streak) in the image C' at a position corresponding to the position of the refractive gradient in the cell (B of Fig. 12-3).

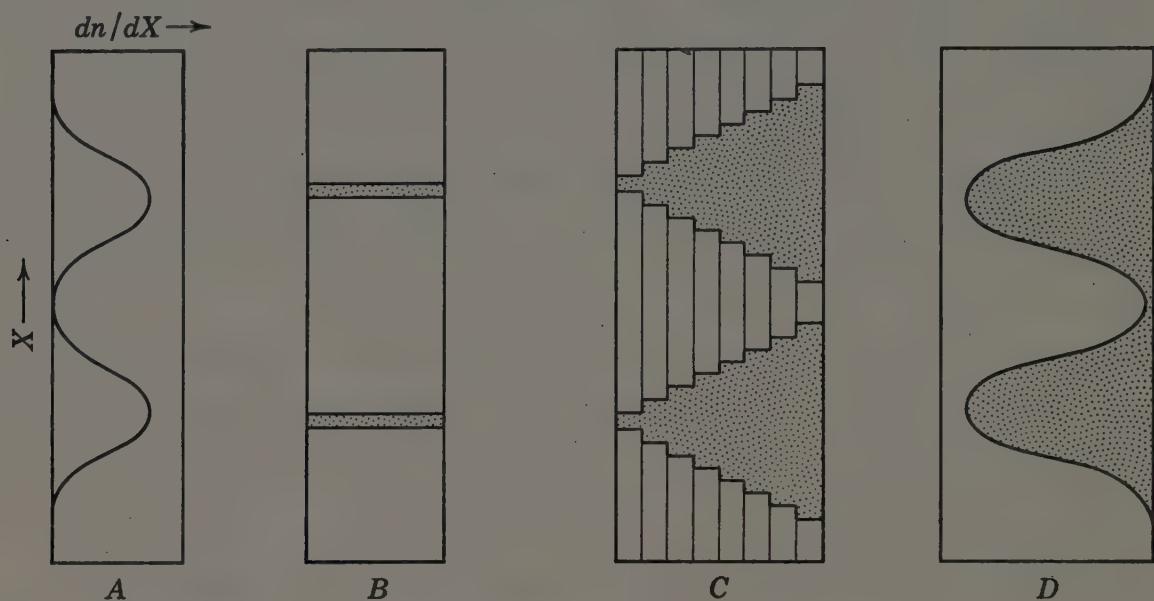


Fig. 12-3. The formation of the schlieren pattern. *A*, the electrophoretic cell, with the refractive gradient plotted as a function of height in the cell; *B*, the simple schlieren pattern obtained with the diaphragm raised to a point where it just intercepts the light which is deflected maximally, yielding two dark bands corresponding to the positions of maximum gradient; *C*, an idealized pattern obtained by raising the diaphragm by regular increments and at the same time moving the film laterally before exposure; *D*, the scanning pattern obtained by raising the diaphragm slowly and continuously in synchronization with a lateral motion of the film.

Such a simple arrangement could be used to define the positions of the boundaries. It is important, however, to have some idea as to the form of the boundaries, and particularly their magnitudes, since this gives information as to the concentration of the components. This could be done by taking, instead of a single picture, a series of pictures while progressively more and more of the deflected light is intercepted by raising the knife-edge diaphragm (*D* of Fig. 12-2) in a stepwise fashion, as illustrated in *C* of Fig. 12-3. Finally, one might move the film laterally past a vertical slit in synchronization with the vertical movement of the knife-edge, to build up a continuous pattern as in *D* of Fig. 12-3. This is the basis of the so-called scanning method of Longsworth, the method most used in this country.

An even more ingenious system is the cylindrical-lens method (Fig. 12-4), called the Philpot-Svensson system after the men who developed and refined it. If, in place of the knife-edge diaphragm previously used, one uses a slit which is inclined to the horizontal at some angle,

for example 45 degrees, a single spot of light will pass at the point where it is illuminated by the horizontal slit image. A downward deflection of the slit image is then resolved into a lateral displacement by the diagonal slit. This horizontal deflection can be recorded on the film without disturbing the focus of the cell, in the vertical direction, by means of a properly oriented and positioned cylindrical lens. There results a pencil of light, composed of rays the vertical position

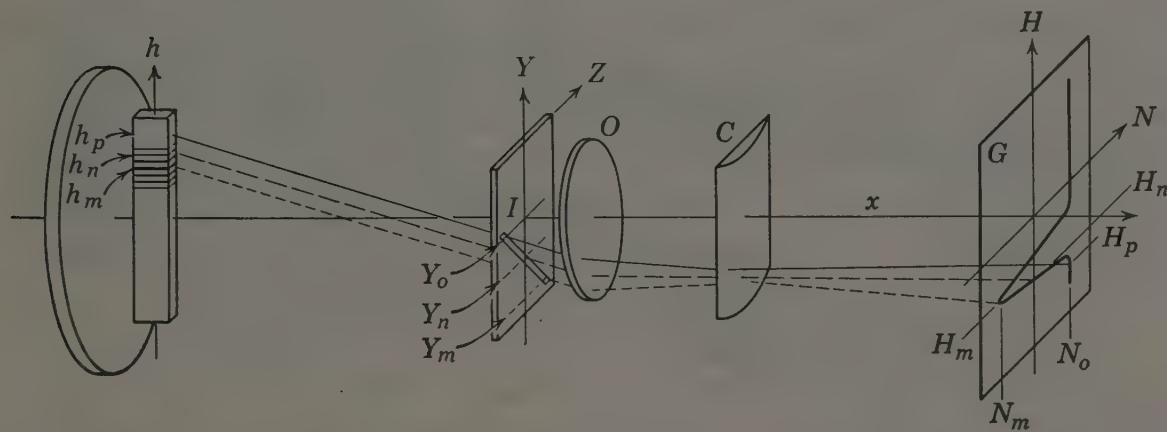


Fig. 12-4. The resolution of a refractive gradient by the Philpot-Svensson cylindrical-lens optical system. Undeflected light, for example from h_p , strikes the diagonal slit at Y_o and the film along the base line N_o . Light passing through the region of a gradient, for example h_n or h_m (the position of maximum gradient), passes the diagonal slit at a lower position, such as Y_n or Y_m , and is focused by the cylindrical lens C at a position off of the base line, in this case at H_n or H_m . Reproduced through the courtesy of Dr. L. G. Longsworth, from his paper "Optical Methods in Electrophoresis," *Ind. Eng. Chem., Anal. Ed.*, **18**, 219 (1946).

of which is associated with position X along the cell, their lateral position being proportional to the magnitude of the refractive gradient dn/dX at the given point in the cell. Since refractive gradient is proportional to concentration gradient, this pencil of light traces out the type of curve shown in E of Fig. 12-1. The method has the advantage that the entire pattern is available for visual observation at any moment, without the necessity of going through the scanning procedure and developing the picture. Further, there are no moving parts to get out of adjustment.

The proportionality constant between concentration gradient and refractive gradient, the so-called refractive increment, is defined through the equation

$$dn/dX = K(dc/dX)$$

and is approximately equal for most proteins. The value ranges from about 0.0018 to 0.0020 for the common proteins if concentration is expressed in per cent. On this basis the integral of dn/dX across the

region of the boundary (total area under the gradient curve) is proportional to the total concentration change at that boundary. In other words, the relative concentrations of protein species are directly proportional to the relative areas of their respective gradient curves. For very precise work it would be necessary to determine the refractive increments of the various species individually.

Limitations. The limiting factor on the amount of current that can be used, hence on the rate at which a separation can be made, is the generation of heat in the cell. Excessive heating is serious in that it leads to thermal gradients resulting in destruction of the sharpness of the boundaries. To minimize this effect, Tiselius made two innovations. (1) He thermostated the cell at approximately 4° C. At this temperature water has its maximum density, and the change in density with temperature is at a minimum. For a given temperature gradient, the density gradient is thereby reduced. (2) He introduced a flattened rectangular cell in place of the previously used design of circular cross section. This facilitates heat interchange between cell and thermostat, thus reducing the temperature gradient. Through these contributions, together with his introduction of the schlieren optical system, Tiselius converted electrophoresis from a laboratory curiosity to a really useful tool in protein chemistry (Fig. 12-5).

In electrophoresis experiments gradients usually occur which cannot be associated with individual ionic species. These are often termed anomalies, although their source is now fairly well understood. Most prominent are the standing anomalies which, as the name implies, occur in the vicinity of the initial boundary. The one on the ascending side, which is most prominent, results from an adjustment that takes place in the concentration of all protein components, and is termed the δ -boundary. (It was first thought to be a fourth member of the plasma globulins, which are denoted by the letters of the Greek alphabet.) The standing anomaly on the descending side results from an adjustment in buffer ion concentrations only, and in many cases is scarcely visible. It is termed the ϵ -boundary.

In practice the smaller the δ - and ϵ -anomalies, other things being equal, the better the experiment. These anomalies in themselves are not particularly objectionable except that, if large, they may obscure boundaries due to slow-moving protein components. The real objection arises from the fact that if large standing anomalies exist, appreciable ionic concentration gradients across the protein boundaries can also be expected. Under such conditions these ionic gradients may make an appreciable contribution to the areas under the refractive gradient curves in the resultant pattern, and the calculated protein

composition will be in error. Furthermore, these ionic gradients lead to uncertainties as to the conductivity in the various regions of the cell, with consequent errors in the calculated protein mobilities. There may even be significant differences in pH in the different regions,

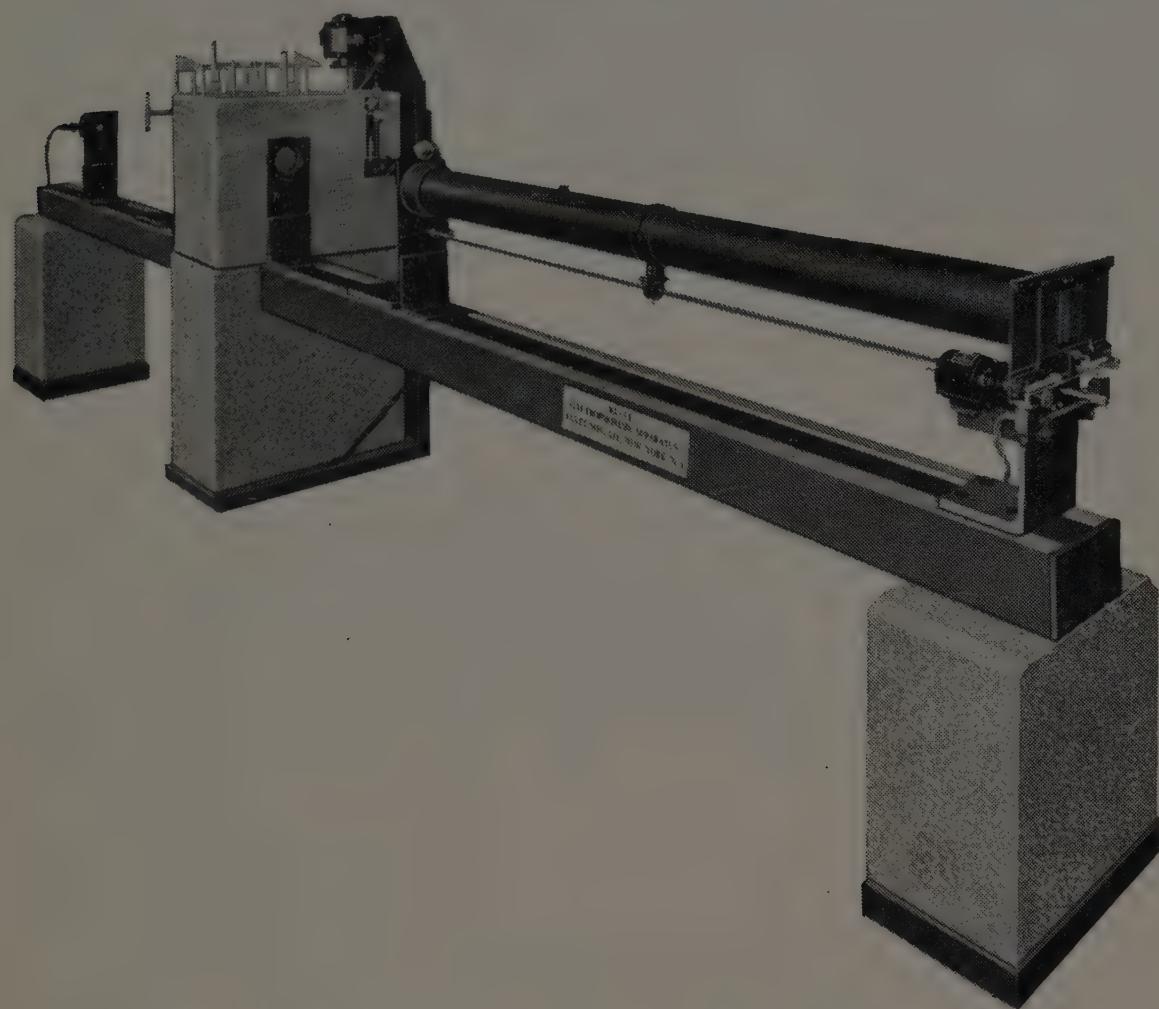


Fig. 12-5. Complete electrophoresis apparatus, showing light source (at the extreme end of the optical bench), insulated thermostat, and optical system. The pattern is recorded by means of the Longsworth scanning camera, on a photographic plate placed in the plate holder at the right. Reproduced through the courtesy of the Klett Manufacturing Co., New York.

another factor which can influence the pattern in the case of ampholytes such as proteins. Since these factors enter differently in the ascending and descending limbs of the cell, the two patterns are affected differently. As a consequence the two patterns, which should ideally be mirror images, may appear quite different. The effects are usually, though not invariably, such that the ascending boundaries are artificially sharpened and the descending pattern broadened.

A well-conducted electrophoretic experiment is indicated, then, if the standing anomalies are not excessively large and if the patterns resulting in the two sides of the cell are approximately enantiographic,

that is, mirror images. If these requirements are not met, the implication is that an insufficient buffer concentration was used, and that the patterns can be interpreted with reservations.

Applications. Undoubtedly the most important use of the Tiselius method has been in the analysis of blood serum and plasma: the establishment of normal patterns for several species, study of the changes

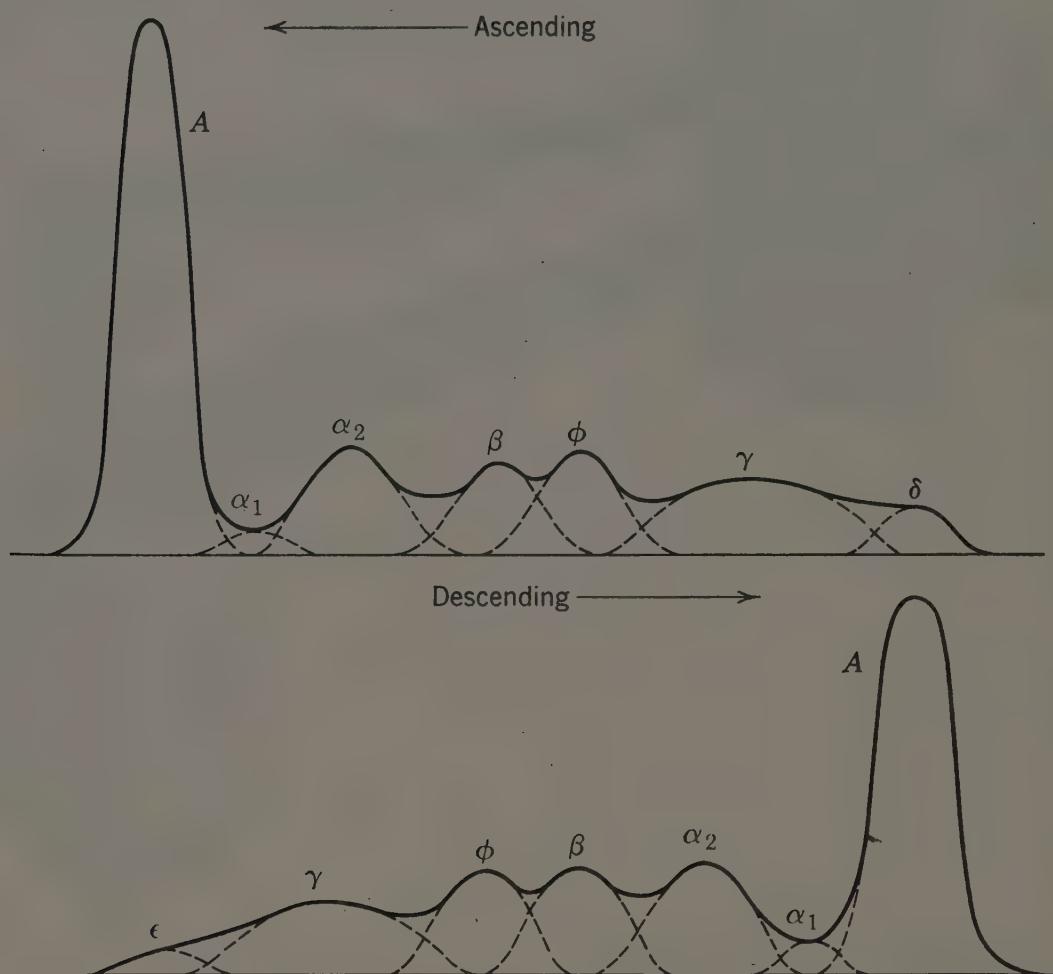


Fig. 12-6. Tracing of a typical electrophoretic pattern for swine plasma. Broken curves outline idealized boundaries, sketched in to sum to the total refractive gradient curve. The relative amounts of the various protein components, designated by A and by the Greek letters, are presumed proportional to the relative areas under the respective broken curves. Conditions: pH 7.5, phosphate buffer, ionic strength 0.20.

in pattern resulting from disease and nutritional deficiency, and use in following the fractionation of the proteins in plasma. Many other protein systems have been studied, and a voluminous literature has developed on the subject.

A typical electrophoretic pattern obtained with blood plasma is reproduced in Fig. 12-6. The various peaks correspond to the principal protein fractions, which are designated in order of increasing mobility: γ -globulin, fibrinogen, β -globulin, α_2 - and α_1 -globulin, and

albumin (A). These plasma components are discussed in more detail in Chap. 18. The various boundaries are seen to be rather poorly resolved. In order to estimate the relative amounts of the various components, it is necessary to sketch in normal boundary curves (Gaussian curves) in such a manner that their ordinates sum at all positions to give the observed outline curve. This procedure, indicated in the figure, is somewhat arbitrary at best, particularly in so complex a system as blood plasma. Analysis is sometimes made simply by dropping perpendiculars from the minima between the peaks, a somewhat less justifiable method.

This pattern is probably more enantiographic than most published patterns on plasma, a consequence of the fact that a rather high ionic strength (0.2) was used in proportion to the protein concentration (1 per cent). Nevertheless, the ascending boundaries are clearly sharper than the descending. The δ -anomaly is not excessively large, but is larger than the ϵ -anomaly. Also the δ -anomaly is seen to be very poorly resolved from the γ -globulin. (This situation is much improved by using veronal buffer systems.) The very broad character of the γ -globulin boundary is a reflection of the heterogeneity of this protein fraction. The curve sketched in for this component is scarcely Gaussian; to fit the gradient curve would require several Gaussian curves, but there is little justification for trying to break the boundary up in this manner.

ZONE ELECTROPHORESIS

In zone electrophoresis, in contrast to moving-boundary electrophoresis, the solution to be separated is originally placed in a restricted zone within the buffer system, and each resolved component subsequently migrates as a discrete zone. Such an experiment is clearly impossible in free solution because of convective mixing. But convection can be eliminated through the use of a suitable supporting medium. Media employed successfully include filter paper, various gels such as agar, and columns packed with such materials as starch or silica. Of these, filter paper has been particularly popular, and commercially available equipment now exists for filter-paper electrophoresis.

In such experiments the support is first saturated with the buffer solution. The solution to be separated, which should be similar to the buffer in all respects except for the solute contained, is then placed near one end of the column or paper strip. Reservoirs of buffer solution are provided at each end of the system, and current is permitted to flow by means of suitable electrodes. Ideally, one then obtains,

after electrophoresis, a series of localized zones, either spots on the paper or layers in the gel or column. For localization of the spots, suitable stains may be applied by a technique analogous to that followed in chromatography.

Zone electrophoresis is still in its infancy but possesses many potentialities. Among its advantages are the following:

(1) The equipment required is simple, particularly in the case of paper electrophoresis.

(2) Small quantities of material are required. On paper, a small drop of solution may suffice.

(3) There is less interference from boundary anomalies. The various zones occupy such short distances that diffusion tends to level out any inequalities in buffer composition.

(4) It is possible to obtain complete separation of all components, in contrast to the situation in moving-boundary electrophoresis where only the fastest and slowest components are obtained in pure form and then in only relatively poor yields.

On the other hand, the method possesses certain inherent limitations. Its utility as a preparative method is limited by the small quantities of solute which can be handled. Absolute mobilities cannot be obtained, in general, because of (1) the severity of electroosmosis and (2) adsorption of the solute on the supporting medium. The first of these effects is due to the fact that the supporting medium usually possesses a charge relative to the liquid, leading to a counterstreaming of solution (electroosmosis). Correction for this effect has been made by adding uncharged reference solutes, such as dextran. The second, perhaps more serious effect limits use of the method to proteins possessing the same sign of charge as the supporting substance, usually negative; in other words, studies are normally limited to *pH* values above the isoelectric point. Adsorption of proteins is a troublesome possibility, however, even if the sign of charge is the same as that of the matrix.

At present it would appear that these advantages and limitations render the method most attractive for study of small ions like those of amino acids and peptides, although good agreement has been obtained between paper electrophoresis and moving-boundary electrophoresis in a few protein systems such as blood serum.

Many variations are possible; for example, two-dimensional electrophoresis on paper (electrophoresis in one buffer system in one direction, followed by electrophoresis perpendicular thereto in another buffer system), and combined electrophoresis and chromatography. In the latter case one takes advantage of not only differences in electro-

phoretic mobility, but also differences in adsorption affinity for the substrate.

REVERSIBLE BOUNDARY SPREADING

As was shown previously, electrophoresis experiments sometimes yield single broad boundaries that do not resolve into components. The question then arises as to how much of the broadening is due to electrophoretic inhomogeneity and how much to the normal spreading due to diffusion. It is possible to distinguish between the two factors by carrying out electrophoresis for a given time, then reversing the direction of the current for precisely the same period of time and at the same current density. To the extent to which the spreading of the boundary in the initial electrophoresis is reversible, it may be assumed to be due to inhomogeneity in the protein comprising the boundary, although there are small secondary effects which are also reversible. The spreading due to diffusion is entirely irreversible and will indeed proceed at the same rate, independent of the direction of the current flow.

The study of *reversible boundary spreading* is usually carried out at or near the isoelectric point (or average isoelectric point) of the protein forming the boundary, so that there is little actual migration of the center of the boundary. This technique promises to afford a powerful criterion of the homogeneity of protein preparations. Indeed, to date no protein has been found which passes this test of purity.

SELECTED REFERENCES

Books and Reviews

- "Electrophoresis" (symposium), *Ann. N. Y. Acad. Sci.*, **39**, 105 (1939).
- "Electrophoresis," L. G. Longsworth and D. A. MacInnes, *Chem. Revs.*, **24**, 271 (1939).
- "Recent Advances in the Study of Proteins by Electrophoresis," L. G. Longsworth, *Chem. Revs.*, **30**, 323 (1942).
- Electrophoresis of Proteins*, H. A. Abramson, L. S. Moyer, and M. H. Gorin, Reinhold, New York, 1942.
- "Electrophoresis by the Moving Boundary Method" (thesis), H. Svensson, *Arkiv. Kemi. Mineral. Geol.*, **22**, No. 10 (1946).
- "Biological and Medical Applications of Electrophoresis," J. A. Leutscher, *Physiol. Revs.*, **27**, 621 (1947).
- "Methods and Applications of Electrophoresis," E. Wiedemann, *Chimia Switz.*, **2**, 25 (1948).
- "Zone Electrophoresis," A. Tiselius and P. Flodin, *Advances in Protein Chem.*, **8**, 461 (1953).

Original Articles

- "A New Apparatus for Electrophoretic Analysis of Colloidal Mixtures," A Tiselius, *Trans. Faraday Soc.*, **33**, 524 (1937).
- "Electrophoretic Study of Mixtures of Ovalbumin and Yeast Nucleic Acid," L. G. Longsworth and D. A. MacInnes, *J. Gen. Physiol.*, **25**, 507 (1942).
- "Progressive Boundary Spread in Electrophoresis of Proteins in Solution," D. G. Sharp, M. H. Hebb, A. R. Taylor, and J. W. Beard, *J. Biol. Chem.*, **142**, 217 (1942).
- "The Quantitative Interpretation of the Electrophoretic Patterns of Proteins," L. G. Longsworth, *J. Phys. & Colloid Chem.*, **51**, 171 (1947).
- "Quantitative Interpretation of Electrophoretic Schlieren Diagrams of Normal Human Plasma Proteins," S. H. Armstrong, Jr., M. J. E. Budka, and K. C Morrison, *J. Am. Chem. Soc.*, **69**, 416 (1947).
- "A Quantitative Study of Reversible Boundary Spreading in the Electrophoresis of Proteins," R. A. Alberty, *J. Am. Chem. Soc.*, **70**, 1675 (1948).
- "Chemical Reactions in Moving-Boundary Systems of Weak Electrolytes," R. A. Alberty and J. C. Nichol, *J. Am. Chem. Soc.*, **70**, 2297 (1948).
- "Homogeneity and the Electrophoretic Behavior of Some Proteins," R. A. Alberty, E. A. Anderson, and J. W. Williams, *J. Phys. & Colloid Chem.*, **52**, 217 (1948).
- "Electrophoretic Study of Interaction of Serum Albumin and Thymus Nucleic Acid," E. Goldwasser and F. Putnam, *J. Phys. & Colloid Chem.*, **54**, 79 (1950).
- "Theory of Reversible Boundary Spreading," R. Brown and J. R. Cann, *J. Phys. & Colloid Chem.*, **54**, 364 (1950).
- "New Type of Preparative Electrophoresis Cell," R. Kekwick *et al.*, *Biochem. J.*, **49**, 253 (1951).
- "Weak Electrolyte Moving Boundary Systems Analogous to the Electrophoresis of a Single Protein," E. Dismukes and R. Alberty, *J. Am. Chem. Soc.*, **76**, 191 (1954).
- "The Characterization of Proteins of Animal Sera Separated by Zone Electrophoresis on Starch," J. Schultz, G. Grannis, H. Kimmel, and H. Shay, *Arch. Biochem. and Biophys.*, **55**, 169 (1955).
- "An Easily Assembled Continuous Electrophoresis Apparatus," H. Saroff, *Nature*, **175**, 896 (1955).

Physical Methods for Investigation of Proteins

Osmotic pressure

Light-scattering

The ultracentrifuge

Diffusion

Solution viscosity

Rotary Brownian motion

The electron microscope

Methods for obtaining further structural details

As early as the beginning of the present century it became apparent to some workers in protein chemistry that certain features of protein behavior are not susceptible to attack by the classical methods of the organic chemist. In particular, it was concluded by some that protein solutions possess the properties of colloidal "sols" rather than "true solutions," suggesting that the dispersed units were of the nature of indiscriminate aggregates of some smaller structural units, for example peptides. The attack on the question of the colloidal nature of protein solutions demanded physical techniques, some already available, others to be developed specifically for this problem.

The outstanding milestone, in the third decade of this century, was the development by The Svedberg and collaborators of the ultracentrifuge, and the demonstration therewith that purified proteins are by no means indiscriminate colloidal aggregates, but definite entities which can rightly be called molecules. This development, far from ending the matter, only served to catalyze an even more intensive

study of the physicochemical properties of the solutions of proteins and of other macromolecules, both synthetic and naturally occurring. Such solutions have been shown to obey the same fundamental laws as ordinary solutions, but with modifications that are a natural consequence of the enormous size of the solute molecules. Today we can speak with some confidence regarding the "molecular weight" of many proteins; unfortunately, we still cannot speak with comparable authority about the equally important properties "molecular shape" and "solvation."

In succeeding chapters it will be seen that many of our present concepts of protein structure and behavior are based on the physicochemical investigation of protein solutions. Many of these techniques are quite foreign to even the trained chemist unless he has had previous experience with proteins or with other macromolecules or polymers. It therefore seems desirable, before continuing with the discussion of protein structure and properties, to outline briefly the more important of these methods, together with some of the results they have contributed or are capable of contributing. No attempt will be made to derive the theoretical equations involved, nor to go into the experimental details. All of the methods are complex, but all rest basically on relatively simple physical principles.

OSMOTIC PRESSURE

The behavior of "ideal" solutions is governed by the fundamental law of Raoult, which states that the partial vapor pressure of any volatile constituent of a solution (the solvent, normally water, in the case of protein solutions) is equal to the vapor pressure of the pure constituent multiplied by the mole fraction of that constituent in solution.*

With little loss of generality, we can define the thermodynamic activity of the solvent component as the ratio of the actual vapor pressure to the vapor pressure of the pure constituent; that is,

$$a_1 = p_1 / p_1^{\circ} \quad (1)$$

(The subscript 1 is by convention employed to refer to the solvent component.) Thus, any measurement which is capable of yielding the activity a_1 is in principle capable of measuring the mole fraction. This in turn tells us the concentration of "kinetic units," the units of solute which are free to move independently of one another. In

* In the simplest possible case, a solution containing only protein and water, the mole fraction of water would be simply moles of water divided by total moles, that is, divided by moles of water plus moles of protein.

covalently bonded molecules, these units would of course be the molecules per se. In the case of salts, they are the ions of which the salt is composed. Knowing the weight concentration of solute, one can immediately calculate the weight of such kinetic units.

Traditionally, four types of measurement have been employed for such purposes, the so-called colligative properties: (1) vapor pressure lowering, (2) boiling point elevation, (3) freezing point depression, and (4) osmotic pressure. Of these, the osmotic pressure is most difficult to measure; however, it constitutes the largest of the effects, and for protein solutions with very large kinetic units is the only one of the four which is useful.

It should be emphasized that Raoult's law, and consequently all theoretical equations based upon it, is strictly applicable only to "ideal" solutions. Fortunately, it can be shown rather rigorously that any solution must approach this ideal behavior as the solute concentration approaches zero. The most important development with respect to the physical chemistry of solutions of proteins, and of macromolecules in general, is the realization and demonstration that even these solutions are also subject to Raoult's law in the limit of zero concentration. It is therefore possible to make measurements of a given property, for example osmotic pressure, at various concentrations, and extrapolate a suitable function to zero concentration, where the ideal equations hold.

If pure solvent is separated from a solution by means of a semi-permeable membrane (one which is permeable to solvent but not to solute), a difference in activity of solvent exists on the two sides of the membrane. This difference in activity manifests itself in a tendency of the solvent to flow from pure solvent to solution, a tendency which can be measured quantitatively in terms of the excess pressure that must be applied to the solution to balance this effect and yield equilibrium.* The equilibrium pressure π is known as the osmotic pressure, and is related to the solute concentration through the van't Hoff equation

$$\pi = (g/M)RT/V = (g/V)RT/M \quad (2)$$

where π is the osmotic pressure, R the universal gas constant, T the

* According to a fundamental thermodynamic relationship, the free energy of solvent \bar{F}_1 , proportional to the logarithm of its activity, varies with pressure according to the relation

$$(\partial\bar{F}_1/\partial P)_T = \bar{V}_1$$

where \bar{V}_1 is the partial molal volume of solvent. The pressure π is that pressure which makes a_1 equal to the activity of solvent in pure form.

absolute temperature, M the molecular weight of the solute and g the number of grams of solute in a volume V of solution.

In order to obtain a reliable measurement of molecular weight, it is important that the data be extrapolated to infinite dilution. This is usually best done by measuring π at several concentrations ($C = g/V$) and plotting π/C versus C . Usually a straight line is obtained, so that an accurate extrapolation can be made.

Small scale instruments have been developed for carrying out osmotic pressure measurements on samples of solution as small as 1 cc. or less. The time of equilibration, which was many hours or even days with the older forms of apparatus, is thereby cut down to as low as one-half hour. The method is at its best for molecules in the size range of most proteins, 10,000 to 100,000 molecular weight. But proteins present a difficulty as a result of their ionic nature, namely, the unequal distribution of small (diffusible) ions across the membrane because of the presence of nondiffusible ions (the Donnan equilibrium).* For this reason measurements on proteins must be carried out and interpreted with care. In particular, it is desirable to carry out the measurements at a pH as close as possible to the isoelectric point of the protein in order to minimize the net charge. In addition, a reasonably high concentration of diffusible ions should be present.

LIGHT-SCATTERING

One of the earliest observed properties of colloidal solutions was their ability to scatter light (Tyndall effect). The ultramicroscope made it possible to observe and count very large colloidal particles, particularly those in the class of so-called lyophobic systems. The technique was of little use in protein solutions. Only very recently has the phenomenon of light-scattering been placed on a quantitative basis for such systems, but already it has become one of the most convenient and reliable methods for measuring the size of protein molecules as well as other macromolecules. The basis of the method lies in the demonstration by Einstein that there is a simple relationship between the ability of a solution to scatter light and the colligative properties of the solution (for example osmotic pressure). The

* The Donnan equilibrium refers to the imbalance of the concentrations (more properly activities) of ions in two regions of a solution, resulting from the presence of one ionic constituent which is restricted in its movement to one of the regions. Most commonly the restriction is furnished by a semipermeable membrane, as in the case of osmosis. The effect is such as to cause abnormally high ionic concentrations in the protein-containing solution, leading to osmotic pressures which are too high.

limiting relation is

$$HC/\tau = 1/M \quad (3)$$

where C is the concentration, τ the turbidity, M the molecular weight, and H a constant involving the wave length of the light used, the refractive index of the solvent, and the refractive increment of the solute (dn/dC). The simple theory considers the solute molecules

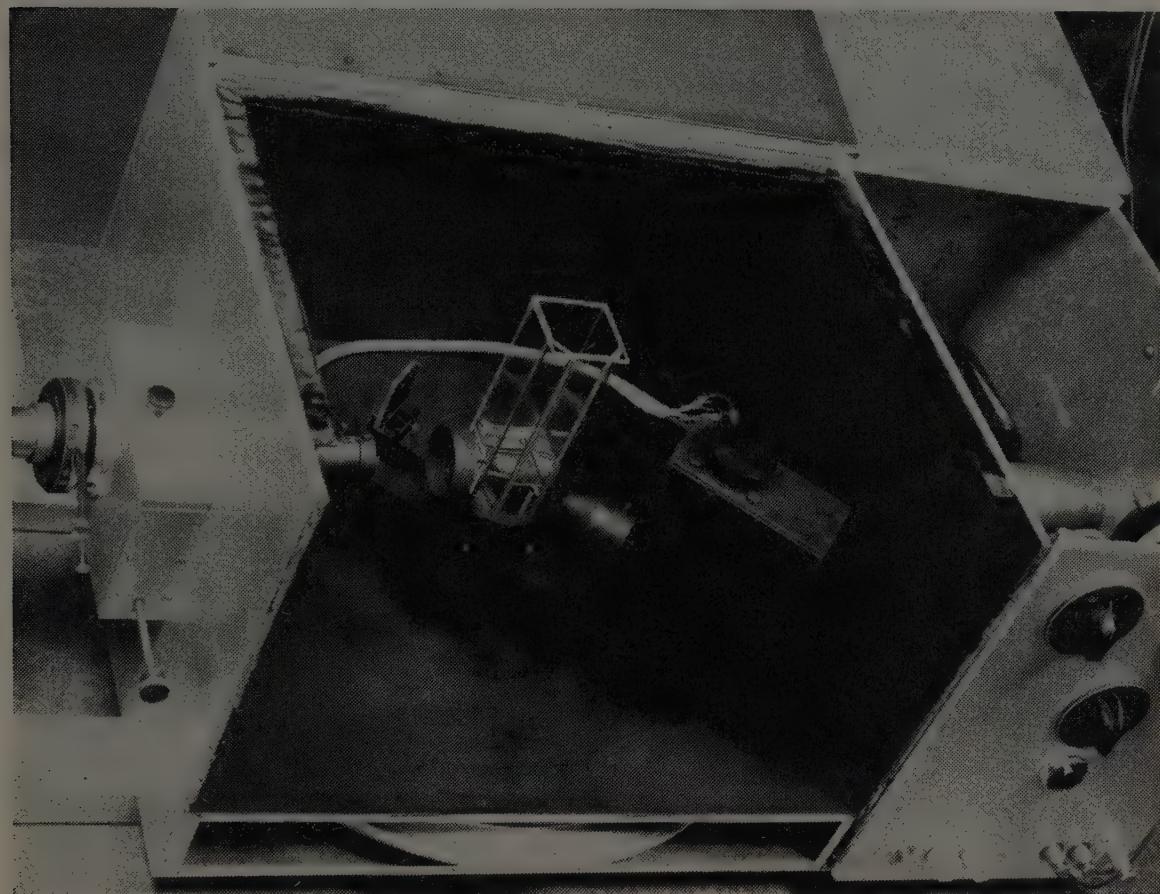


Fig. 13-1. An instrument for the measurement of light-scattering by solutions. The light originates at the left side, just outside the field of the photograph, passes through a suitable collimating system, and through the six-sided cell containing the solution. The intensity of the light scattered in various directions is measured photoelectrically by means of the photomultiplier tube contained in the black box. From B. A. Brice, M. Halwer, and R. Speiser, *J. Opt. Soc. Amer.*, **40**, 768 (1950).

as point sources of scattered light, and is satisfactory provided no dimension of the molecule exceeds about $1/20$ the wave length of the light used, that is, about 200 Å. Many proteins fulfill this requirement. For such systems it is only necessary to measure the scattering at any one angle, usually 90 degrees to the direction of the incident light, as a function of concentration of solute and to carry out a suitable extrapolation to zero concentration. Usually the ratio of concentration to turbidity, C/τ , is plotted as a function of C . For

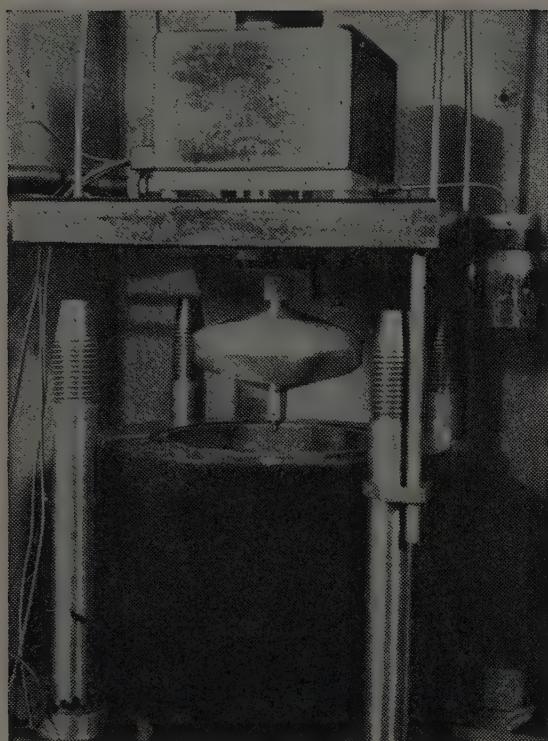
larger or more elongated molecules it is necessary, in measuring τ , to make a correction for the dissymmetry of scattering. This results from the fact that the light scattered in the forward direction is somewhat greater in intensity than that scattered in the backward direction as a consequence of interference. Since most globular proteins do not possess any dimension greater than $\frac{1}{2}_0$ the wave length of visible light (normally the green emission line of the mercury arc of wave length approximately 5400 Å. is employed), such dissymmetry is not usually observed. In a sense this is unfortunate, for careful study of the interference effect provides additional information on molecular shape. Limited studies of the angular dependence of scattering of X-rays (of very short wave length) have provided important information on the shape of globular proteins.

Application of light-scattering in protein chemistry can be expected to expand rapidly. The method should be especially useful for studying changes in proteins, such as aggregation, dissociation, and unfolding, and as another test for purity. Light-scattering, in the case of inhomogeneous solutions, yields a weight-average molecular weight. In other words, the results are weighted relatively heavily by the larger molecules present. Osmotic pressure, on the other hand, yields a simple number-average value, all molecules contributing equally. In the case of several proteins, notably serum albumin, ovalbumin, β -lactoglobulin, lysozyme, and fibrinogen, good agreement has been obtained between the two methods, indicating the high degree of homogeneity of some proteins insofar as molecular size is concerned.

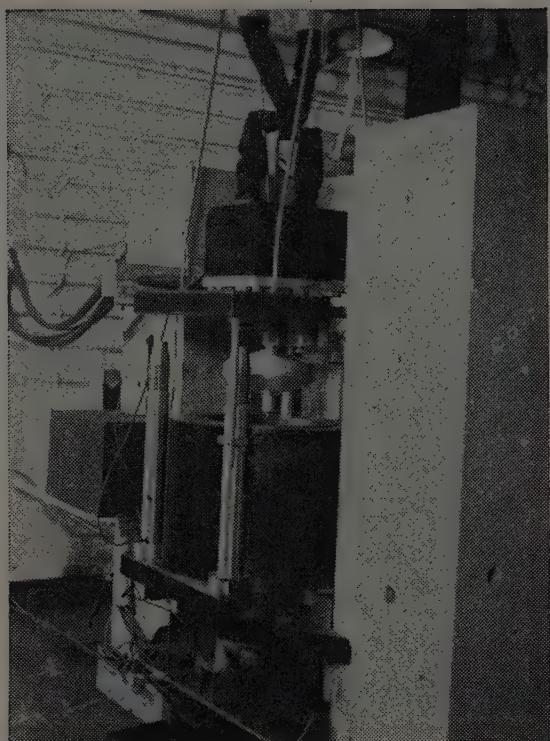
THE ULTRACENTRIFUGE

The physical method which has probably contributed more than any other to our present state of knowledge of the protein molecule, and perhaps of colloidal systems in general, is sedimentation in the ultracentrifuge. The development of this instrument is largely the product of The Svedberg and his students at Uppsala, Sweden, who also conducted the early measurements on proteins. For this work Svedberg received the Nobel prize in 1926.

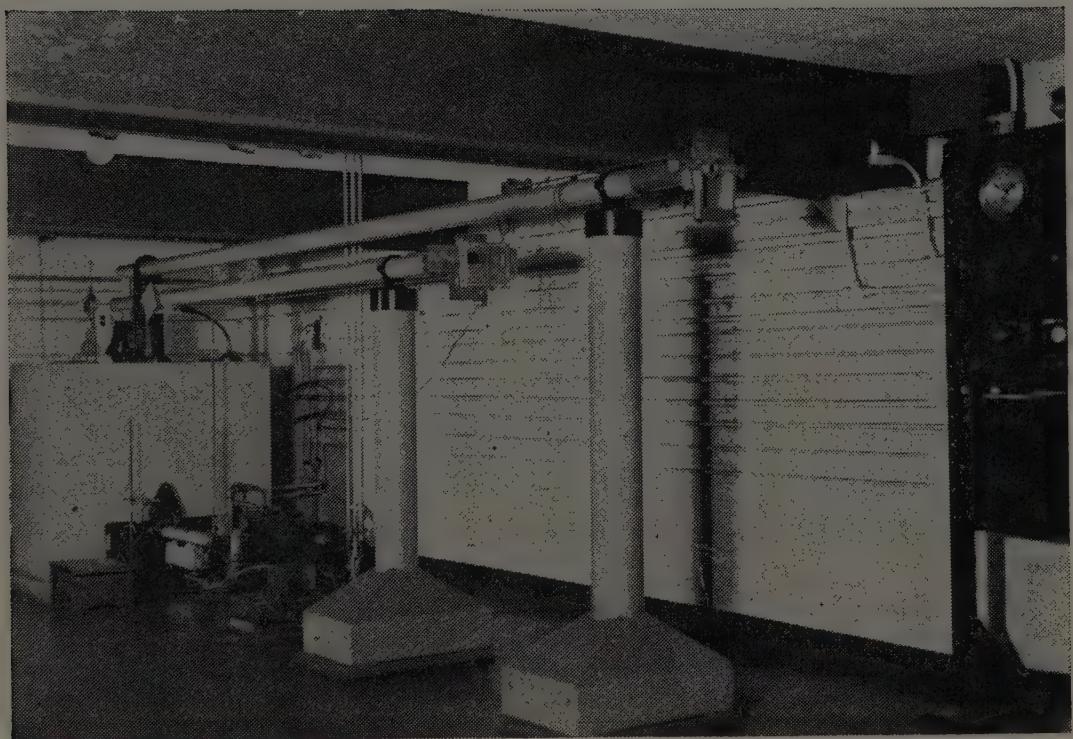
Svedberg realized that colloidal particles such as protein molecules are of such a size that it should be possible to cause them to sediment out of solution under the influence of gravitational fields that are practically attainable. Further, he deduced that it should be possible to gather information regarding the size and homogeneity of solute particles by such means. Accordingly, in the early 1920's he constructed the first high-speed centrifuge, and supplanted this by more and more refined instruments capable of rotating at 50,000 to 100,000



A



B



C

Fig. 13-2. The air-driven ultracentrifuge of the University Laboratory of Physical Chemistry, Harvard. A, the rotor. The drive mechanism is enclosed in the box at the top. When in operation, the whole upper assembly is lowered so that the rotor is contained within the steel chamber. B, a portion of the protective concrete barricade. C, the optical systems and control panel. The centrifuge proper is behind the concrete barricade at the extreme left of the picture. Courtesy of Dr. J. L. Oncley.

r.p.m. and of producing gravitational fields as high as 500,000 times that due to the earth.

The modern ultracentrifuge (Fig. 13-2) consists essentially of a rotor 6 to 10 in. in diameter, usually resembling a large top, constructed of a light but strong alloy such as Duralumin. This rotates in a high-vacuum chamber to reduce air friction, and is driven through a flexible shaft, such as a piano wire, by a turbine. The turbine may be driven by either oil or air under pressure.*

The solution is contained in a disk-shaped metal cell of about 1-in. diameter, which is channeled out in the form of a sector about the center of the rotor. This cell, covered above and below by glass or quartz windows to provide for optical examination during the run, fits into a hole near the periphery of the rotor. Usually there are two such cells, the second being filled with solvent and placed opposite in the rotor to serve as a counterbalance.

The concentration distribution of the dissolved proteins in the cell can be studied by any of the previously mentioned optical systems. This requires windows in the top and bottom of the vacuum chamber, in such positions that the cell will pass between them once each revolution.

There are two different ways in which the ultracentrifuge can be used, and the construction of the instrument varies considerably depending on which purpose it is to fulfill.

The Equilibrium Method. In this method the gravitational field is increased only to a level where the normally random distribution of the solute molecules is modified to a moderate extent. Centrifugation is continued at a constant speed until equilibrium between the gravitational force and thermal (diffusion) force is established. The molecular weight of the solute, provided it is homogeneous, can then be determined from the concentration distribution, using optical methods similar to those used in electrophoresis, by means of the expression

$$M = \frac{2RT \ln C_2/C_1}{(1 - \bar{V}\rho)\omega^2(X_2^2 - X_1^2)} \quad (4)$$

where C_1 and C_2 are concentrations at any two positions in the cell, X_1 and X_2 ; ω is the velocity of rotation in radians per second; \bar{V} is the

* An electrically driven ultracentrifuge capable of producing gravitational fields of 250,000 times gravity is now in commercial production in the United States (Spinco ultracentrifuge, manufactured by the Specialized Instruments Corp., Belmont, California). This machine has increased several fold the number of ultracentrifuges in use in this country.

partial specific volume of solute; ρ is the density of the solvent; and R and T have their usual significance. If the solute is nonhomogeneous, the calculated values of M will be found to drift with X . Interpretation in such cases involves knowledge as to the nature of the polydispersity, that is, whether the system consists of a random distribution or of discrete components.

The equilibrium method may be regarded as an exact thermodynamic method in the same sense as those based on osmotic pressure and the other colligative properties, and involves no assumptions as to the shape of the molecules. As in the other cases, it is important to carry out measurements at various concentrations and to extrapolate the results to zero concentration in order to obtain correct values of molecular weight.

This method does not give information as to the number of components in a polydisperse system, and has the further disadvantage in protein studies that it often requires several days to attain equilibrium. During this time changes in the properties of the protein may ensue. The centrifuge, although it rotates at only relatively low speeds (20,000 r.p.m. or less) as compared to the velocity ultracentrifuge described later, must be carefully insulated against mechanical shock and thermal fluctuations. This imposes a considerable mechanical problem.

The Velocity Method. In this method the gravitational field is increased to a point where the solute sediments out of solution at a finite rate. The sedimentation coefficient is the velocity attained in a unit centrifugal field and has the dimensions of time, expressed in the c.g.s. system in seconds. For most proteins it is of the order 10^{-13} sec. Hence a new unit has been defined, the Svedberg unit S , which is a sedimentation constant of 10^{-13} sec. Since the velocity depends on the viscosity of the solvent and on its density, results are usually corrected, for uniformity, to a basis of sedimentation in water at 20° ($S_{20,w}$).

With charged particles, as is normally the case when dealing with proteins, the sedimentation rate is retarded by the drag of the surrounding oppositely charged ions (electroviscous effect). To reduce this effect it is desirable to carry out sedimentation experiments in solutions containing moderately high concentrations of ions. The corrected sedimentation constant $S_{20,w}$ is a characteristic property of the solute, depending directly on the mass and inversely on the so-called frictional coefficient f of the solute particles; that is,

$$S = M(1 - \bar{V}\rho)/f \quad (5)$$

The value of f depends on the shape of the molecule and also on the extent to which it binds solvent, that is, its solvation. Interpretation of the data in terms of molecular weight therefore requires further knowledge of these properties. In Svedberg's work this information was ordinarily obtained from diffusion measurements, to be discussed. The frictional coefficient is uniquely defined in terms of the diffusion constant D , through the relation $f = RT/D$. Viscosity studies can also be used, although the relation between viscosity and the frictional coefficient is not simple and involves certain assumptions. Alternatively, independent molecular weight measurements, as from osmotic pressure or sedimentation equilibrium, could be used, and the $S_{20,w}$ value used to compute the frictional coefficient. If diffusion measurements are used, the fundamental relation is

$$M = \frac{RTS}{D(1 - \bar{V}\rho)} \quad (6)$$

where S is the sedimentation constant, D the diffusion constant (measured under the same conditions of solvent viscosity or corrected to the same conditions), and \bar{V} and ρ have the same significance as previously given.

In addition to providing data of value in determining the size and shape of large molecules, the velocity centrifuge gives information regarding the homogeneity of the system. Such information is derived from an examination of the shape of the sedimenting boundary or boundaries. Heterogeneous protein preparations containing two or more species which differ appreciably in S give rise to multiple sedimenting boundaries, analogous to multiple boundaries in electrophoresis. The Svedberg school has used sedimentation as a control method in fractionation studies of protein systems such as blood plasma, supplementing the more commonly used electrophoretic technique.

DIFFUSION

The experimental approach most used in conjunction with the velocity ultracentrifuge is the determination of the diffusion constant D . This coefficient, like the sedimentation constant, depends both on the mass and on the frictional coefficient of the solute molecules, and hence gives useful information only when used in conjunction with some other method that depends on one or both of these properties. The diffusion constant is a measure of the rate at which the solute molecules assume a random distribution in a solvent.

Mathematically it is defined through the Fick first-law equation

$$ds/dt = -D \cdot A \frac{\partial C}{\partial X} \quad (7)$$

where ds/dt is the time rate of diffusion across a boundary of area A under a concentration gradient $\partial C/\partial X$.*

The measurement of D can be made in several ways, perhaps the best of which is to form a sharp boundary between the protein solution and the solvent (as in electrophoresis) and to observe the rate at which this boundary broadens. In the ideal case of a homogeneous solute whose rate of diffusion does not depend appreciably on concentration, the concentration (or refractive index) gradient curve will be of the Gaussian† type. An asymmetric curve indicates nonideal diffusion (dependence on concentration), whereas a symmetrical non-Gaussian curve indicates nonhomogeneity, in which case the curve is actually a summation of two or more Gaussian curves. Interferometric optical techniques are best for these measurements, although the Philpot-Svensson type of optical system is capable of giving fairly reliable values. Hence, measurements can be made in the ordinary electrophoretic apparatus.

Another important method of measurement of the diffusion constant is based on the rate of diffusion through a sintered glass plate. This method is not an absolute one, in that the plate must be calibrated by using solutes of known diffusion constant. The method has been used some in the protein field, particularly in the case of enzymes, where the activity can be employed as a sensitive measure of the amount of protein diffusing through the sintered glass in a given time.

Table 13-1 summarizes the results of sedimentation and diffusion analyses on several important proteins. In this table, V_{20} is the partial specific volume, S_{20} the sedimentation velocity in Svedberg units, and D_{20} the diffusion constant in square centimeters per second, all corrected to water at 20° C. The molecular weight derived from

* This relation assumes that the concentration varies along only one direction, namely the X direction. From the Fick first law, the second law may be derived in the form

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial X^2}$$

This equation relates the time rate of change of concentration in a small element of volume to the second derivative of concentration with respect to position.

† The symmetrical bell-shaped distribution frequently obtained in a statistical measurement.

INTRODUCTION TO PROTEIN CHEMISTRY

TABLE 13-1. Sedimentation and Diffusion Results on Some Important Proteins, and Derived Values of Molecular Weight and Shape

Protein	\bar{V}_{20}	S_{20}	$D_{20} \times 10^7$	M_s	M_e	f/f_0	a/b^1	Oblate	
								10	0.08
Lysozyme	0.75	1.8-	8.6-	14,000-	...	1.58			
Lactalbumin	0.751	1.9	11.2	17,000	19,000	1.16	4.0	0.25	
Cytochrome C	0.707	1.9	10.6	17,400	...	1.29	5.8	0.17	
Myoglobin	0.751	2.0	10.1	15,600	...	1.11	3.2	0.32	
Ribonuclease	0.709	1.85	11.3	16,900	17,500	1.04	2.2	0.45	
Gliadin	0.722	2.1	13.6	12,700	13,000	1.04	1.1	0.07	
Hordein	0.729	2.0	6.72	27,500	27,000	1.60	1.2	0.06	
Zein	0.771	1.9	6.5	27,500	...	1.64	2.5	0.01	
β -Lactoglobulin	0.751	3.12	7.3	41,500	38,000	1.26	6.2	0.18	
Pepsin	0.750	3.3	9.0	35,500	39,000	1.08	2.8	0.35	
Insulin	0.749	3.5	8.2	41,000 ²	35,000	0.96	1.0	1.1	
Ovalbumin	0.749	3.55	7.8	44,000	40,500	1.16	4.0	0.25	
Hemoglobin, horse	0.749	4.41	6.3	68,000	68,000	1.24	5.0	0.18	
Hemoglobin, man	0.749	4.48	6.9	63,000	...	1.16	4.0	0.25	
Serum albumin, horse	0.748	4.46	6.1	70,000	68,000	1.27	5.4	0.17	
Serum albumin, man	0.733	4.6	5.9-	69,000-	...	1.28	5.7	0.17	
Serum α_1 -globulin, man			6.1	72,000	...				
Serum α_1 -globulin, man	0.841	5.0	...	200,000	...	1.38	7.5	0.12	
Serum β_1 -globulin, man	0.693	9	...	(300,000)	...	1.58	10	0.08	
Serum β_1 -globulin, man (lipid rich)	0.725	5.5	...	90,000	...	1.37	7.2	0.13	
Serum γ -globulin, man (two components)	0.950	2.9	...	1,300,000	...	1.7	13	0.06	
Fibrinogen, man			7.2	156,000	...	1.38	7.5	0.12	
		10	...	(300,000)	...				
		9	...	400,000	...	1.98	20	0.02	

¹ Axial ratio calculated on basis of no hydration; actual value would be less than this.

² Dissociates under other conditions into subunits of about 12,000 molecular weight.

³ In these calculations viscosity data were used rather than diffusion data.

these data is given as M_s ; M_e is molecular weight from sedimentation equilibrium.

From the measured molecular weight it is possible to compute f_0 , the frictional coefficient of an idealized compact spherical molecule of the same molecular volume. The ratio of the observed frictional value to this idealized value, f/f_0 , is the frictional coefficient. It will obviously be 1.0 for an unsolvated spherical molecule, greater than unity

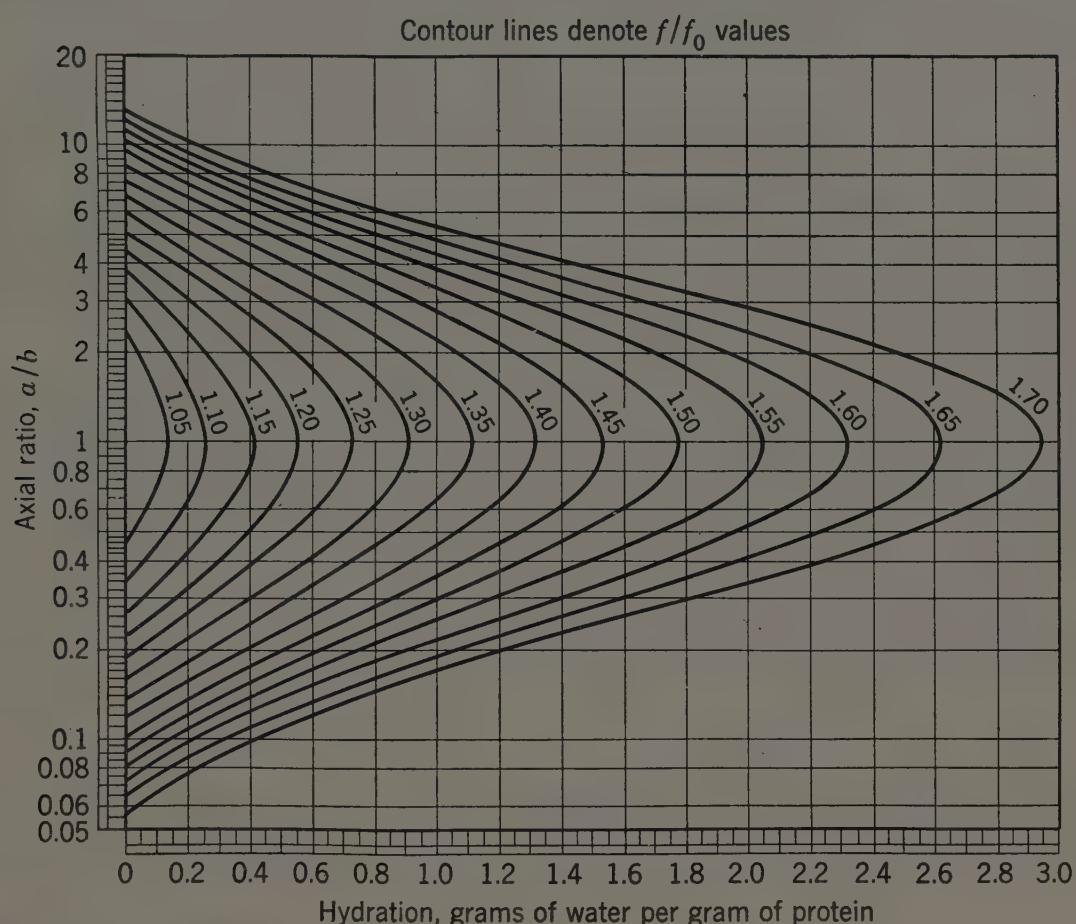


Fig. 13-3. Contour chart showing values of axial ratio and degree of hydration which are in accord with various values of the frictional ratio f/f_0 . From J. L. Oncley, Ann. N. Y. Acad. Sci., 41, 121 (1941).

for an actual molecule which is either asymmetric or solvated. Theoretical relationships between f/f_0 and axial ratio for various degrees of hydration are shown graphically in Fig. 13-3. The axial ratios in Table 13-1 are obtained from such relationships on the assumption of no hydration and hence represent an upper limit. There is unfortunately no known way of separating the effects of molecular asymmetry and solvation.*

* This problem has been most clearly emphasized by Scheraga and Mandelkern (1953).

SOLUTION VISCOSITY

Measuring the viscosity of polymer solutions has long been a popular means of estimating the size and/or shape of the solute molecules. The reason for this lies in the relative simplicity of the measurements as compared to the other techniques available. From the theoretical standpoint, interpretation of the data is difficult and depends on assumptions concerning the nature and properties of the solute. The intrinsic viscosity η_i of a solution is defined by the relation

$$\eta_i \equiv \lim_{C \rightarrow 0} \frac{\eta_{sp}}{C} \equiv \lim_{C \rightarrow 0} \frac{\ln \eta_R}{C} \quad (8)$$

where η_R (relative viscosity) is the ratio of the viscosity of the solution to that of the solvent; and η_{sp} (specific viscosity) is the difference between the viscosity of the solution η and that of the solvent η_0 , divided by the latter: $[(\eta - \eta_0)/\eta_0$ or $(\eta_R - 1)]$.

Essentially the intrinsic viscosity of a solution is a measure of the effective "hydrodynamic" volume of the solute (contributions due to interaction between solute molecules are eliminated by the extrapolation to zero concentration). This effective volume is at a minimum for compact, rigid, unsolvated spheres; in which case, according to Einstein, η_i should be equal to 2.5 if C is expressed as volume fraction. Asymmetry increases the effective volume, since the molecule rotates in a more or less random fashion and can be considered as occupying a volume roughly equivalent to a sphere of diameter equal to its longest dimension. If the molecule is strongly solvated, it can be considered as carrying an envelope of solvent with it, thus further increasing its effective volume and viscosity. If the molecule is a chain, the effective volume depends upon the degree to which this chain is coiled, the rigidity of the coiled structure, the ease with which solvent can flow through the interstices of the coils, and so on.

In the case of native corpuscular proteins, viscosity data can probably best be interpreted on the basis of rigid molecules, more or less solvated and deviating somewhat from spherical form. Deviations from sphericity have so far been treated mathematically on the basis of ellipsoidal shape, either prolate (cigar-shaped) or oblate (disk-shaped), improbable as such structures are. The relationships for proteins are expressed reasonably well in the graph shown in Fig. 13-4.

In spite of the limitations, these relations have been found to yield values for the axial ratio a/b of some proteins which are in close agreement with values obtained by other physical measurements, particularly sedimentation and diffusion. However, it should be empha-

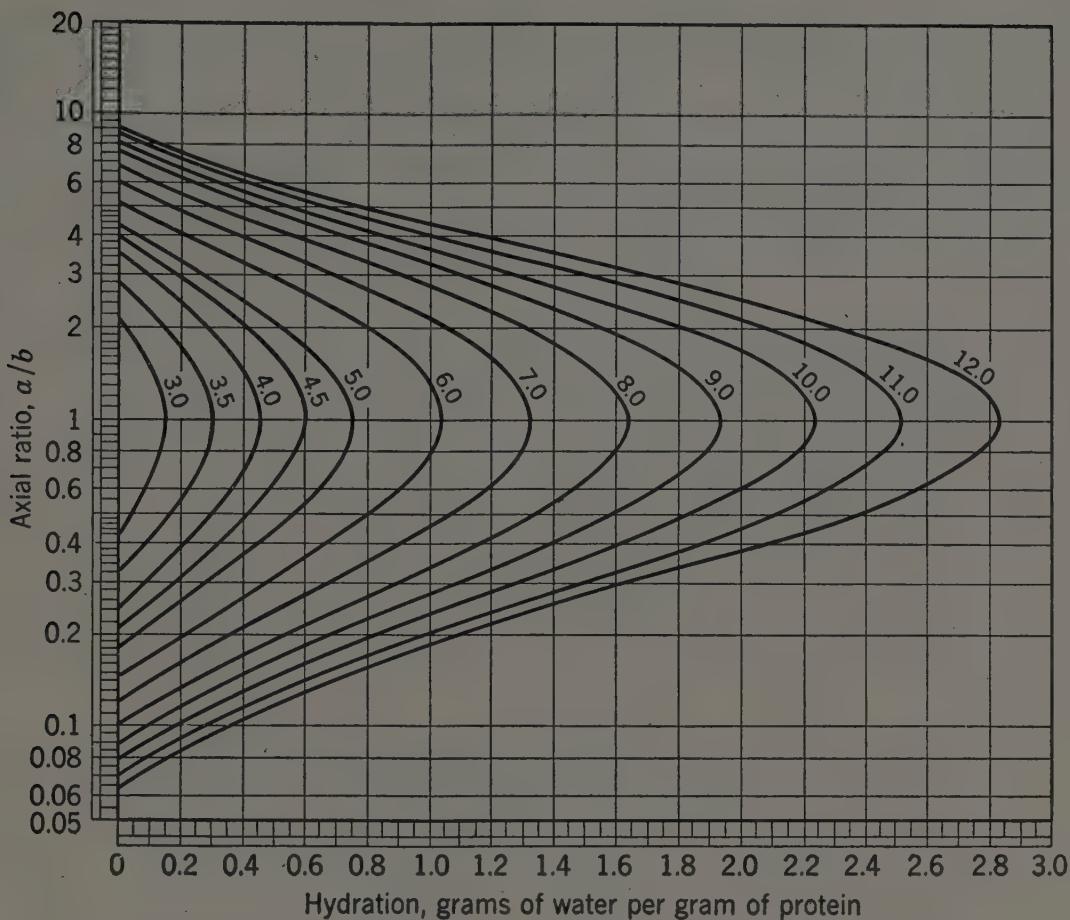


Fig. 13-4. Contour chart showing values of axial ratio and degree of hydration which are in accord with various values of the viscosity coefficient. Contour lines denote viscosity coefficient η_i , concentration being expressed in units of volume fraction. From J. L. Oncley, *Ann. N. Y. Acad. Sci.*, 41, 121 (1941).

sized again that such calculations necessitate some assumption about solvation and the results must not be taken too seriously.

ROTARY BROWNIAN MOTION

Analogous to the translational diffusion of molecules in solution, there is also rotational motion about their axes. A *rotary diffusion constant* Θ can be defined, and will depend on the size and shape of the molecule in question. An ellipsoidal molecule possesses two such constants, each characteristic of rotation about one of its two principal axes. Two methods are of primary importance in evaluating these constants, dielectric dispersion and streaming birefringence.

Dielectric Dispersion. At the isoelectric point, protein molecules carry equal numbers of positively and negatively charged groups. It seems highly improbable a priori that these charges would be symmetrically placed; hence it is to be expected that protein molecules would have an appreciable dipole moment (a separation between the centers of positive and negative charge). If a solution of such

molecules is placed between electrodes and subjected to an alternating current (to form a condenser), the dipolar molecules tend to rotate or oscillate in phase with the current, making a large contribution to the dielectric constant of the system. Resisting this oscillation are the frictional forces between solute molecules and solvent, which forces are a function of the size and shape of the solute molecules and of the viscosity of the solvent. As the frequency is increased, the molecules have more and more difficulty in keeping in phase, and as a consequence the dielectric constant is found to drop, more or less in a stepwise manner. These steps yield one or more characteristic frequencies, which are related to the time required for rotation about various axes of the molecule (rotary diffusion constants). For a cigar-shaped molecule there would be two such frequencies, the lower corresponding to orientation of the long axis (rotation about the short axis), and the higher corresponding to orientation of either of the short axes.

Analysis of the data contributes information as to the axial ratio of the molecule, the results again depending on the molecular weight, the degree of solvation, and the assumptions as to shape and rigidity. Though the method is most readily applied to proteins soluble in media of low polarity, such as the prolamins, it has been extended to the study of albumins and globulins. The total dipole moment of the molecule can be estimated, and also the orientation of this moment with respect to the geometric axes of the molecule. Such measurements on variously modified proteins could conceivably supply some information as to the location of polar groups. The dipole moment of proteins is usually found to be quite small, considering the size of the molecules involved, indicating a high degree of symmetry in the distribution of the charges.

It has been pointed out by Kirkwood and Shumaker (1952) that the dielectric dispersion exhibited by protein solutions may also be explainable on the basis of migration of hydrogen ions on the protein surface. If this is true, dielectric dispersion would clearly be of no value in elucidating questions of molecular size and shape. Unfortunately, no good experimental means of distinguishing the two contributions to dielectric dispersion has yet been conceived.

Streaming Birefringence. If a solution of asymmetric molecules is subjected to a strong velocity gradient, for example, by being placed between concentric cylinders one of which is rotating, there is a tendency for the molecules to orient with their long axes parallel to the direction of flow. Opposed to this is the tendency of the molecules to assume a completely random orientation owing to their thermal

energy. The resulting equilibrium state is a compromise between these two tendencies. The partial orientation results in the normally isotropic solution becoming doubly refractive (birefringent), and the magnitude of the birefringence and the preferred direction of orientation (optic axis) can be measured. The angle χ , between this direction and the direction of flow, is related through an elaborate theory to Θ ,

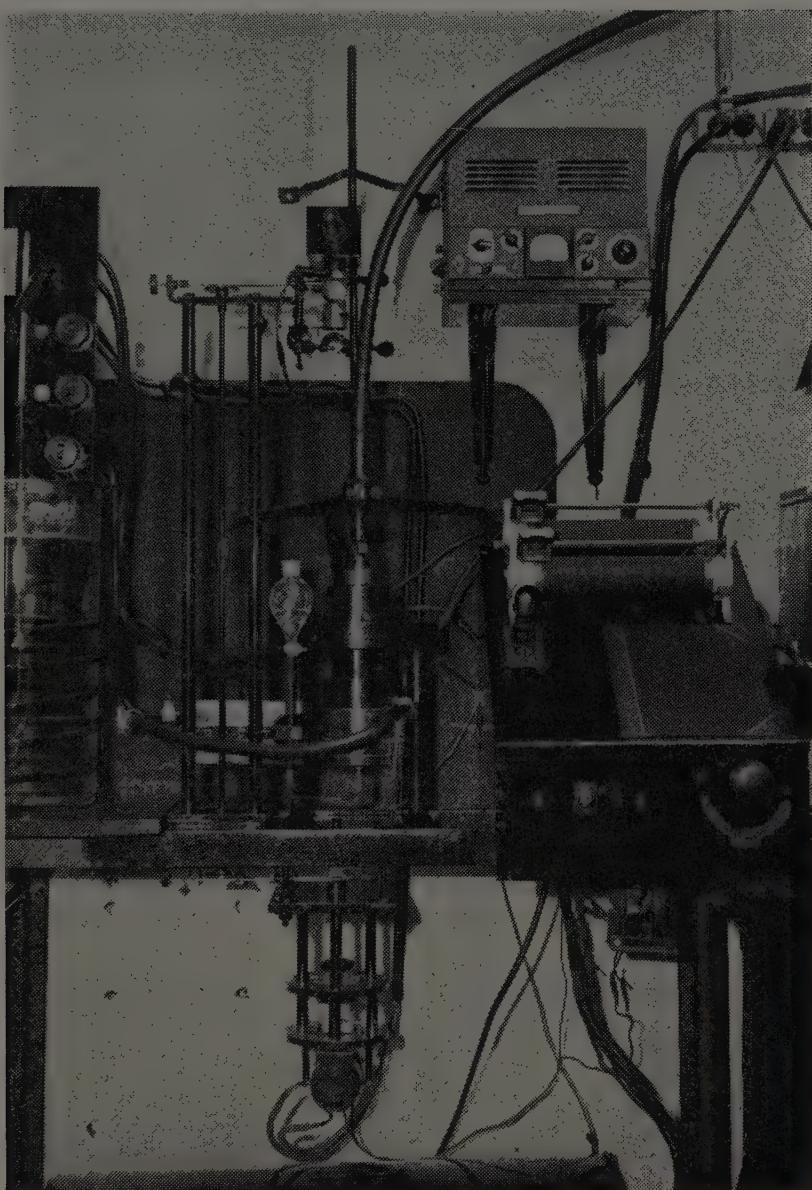


Fig. 13-5. Apparatus for the measurement of streaming birefringence.

the rotary diffusion constant of the molecule (about its short axis). This rotary diffusion constant depends in part on the axial ratio of the molecule, but principally on the actual magnitude of the major axis; the method can therefore be considered suitable for measuring the length of large molecules. Results on several corpuscular proteins have yielded lengths in close agreement with values calculated from

other physical methods (Table 13-2). For nonrigid molecules the theory is not yet well clarified.

It should be noted that the proteins included in Table 13-2 are relatively large and/or relatively asymmetric. The flow birefringence method cannot, unfortunately, be employed with most of the globular proteins of interest; molecules possessing no dimension greater than

TABLE 13-2. Length of Protein Molecules As Determined by Streaming Birefringence

Protein	Length from Sedimentation and Diffusion Analysis, A.	Length from Flow Birefringence, A.
Myosin ¹	...	
Rabbit	...	11,600
Snail	...	ca. 28,000
Octopus	...	18,000
Tobacco mosaic virus		
pH 6.8	3000 ²	7,200
pH 4.5		24,000
Casein, alkaline solution	...	2,200
Helix Hemocyanin		
Mol. wt. 8.9×10^6	1130	890
Mol. wt. 4.3×10^6	820	890
Mol. wt. 1.0×10^6	820	960
Mol. wt. 0.91×10^6	960	1,280
Fibrinogen, human	700-750	700
Zein	425 ²	345-425
γ -Globulin, human	235 and 470 ²	230
Serum albumin, bovine	150	190-200

¹ Actually actomyosin in the new terminology (see p. 360).

² Known to be inhomogeneous.

approximately 300 Å. rotate too rapidly to be oriented readily in the streaming gradient. The results reported on γ -globulin and serum albumin were obtained in media of very high glycerol content to reduce the speed of Brownian rotation and the results are probably of only qualitative significance. Some of these globular protein molecules become more elongated upon denaturation (Chap. 17), and the method of streaming birefringence is of importance for the study of such changes.

Polarization of Fluorescence. Another method for studying rotatory diffusion should be mentioned although it has not as yet been widely employed. This is the study of the polarization of fluorescent light emitted by proteins which have been tagged with suitable fluorescent groupings. In fluorescence a certain finite time elapses between excitation and the emission of radiation. This can be characterized through

the fluorescence half life. If in this period of time the fluorescing units cannot undergo appreciable rotation, the light emitted at 90 degrees to the direction of the incident (exciting) radiation should be completely polarized. Significant rotation between absorption and emission will lead to partial depolarization. Study of the degree of depolarization by use of fluorescing groups of known fluorescence half life can thus permit calculation of a mean rotary diffusion coefficient.

It is important that the fluorescent tag not be able to rotate independently of the protein molecule as a whole. The problem of firmly affixing a group of the right properties without significantly altering the macromolecule itself is a formidable one in the case of proteins. Nevertheless, it can be anticipated that this method will find increased application in the future because of the simplicity of the equipment required, and because it should be applicable in situations where streaming birefringence is out of the question.

THE ELECTRON MICROSCOPE

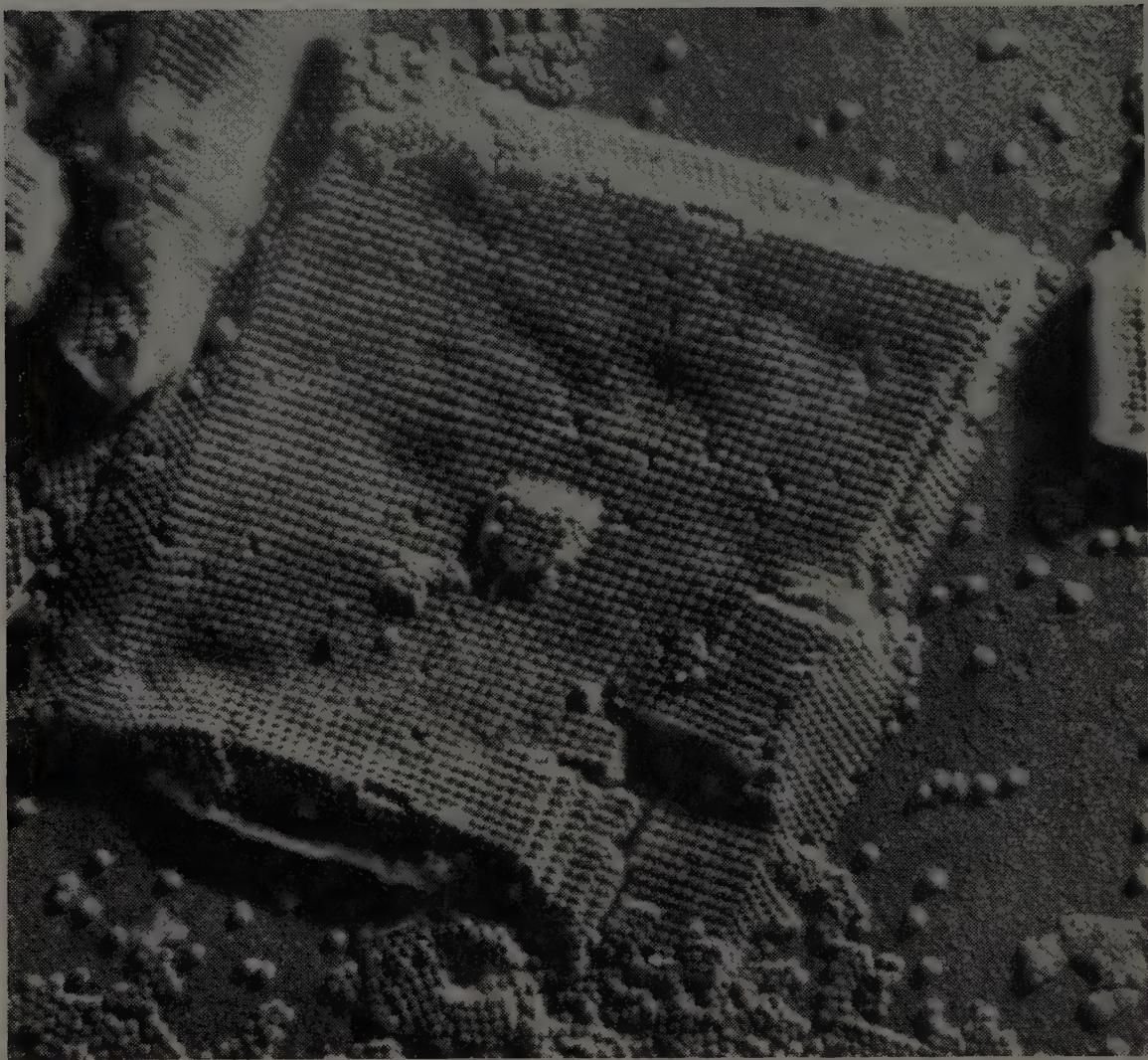
The principal factor limiting the resolving power of a microscope is the wave length of the light used. In recent years it has been found possible to focus electron beams by means of magnetic or electrostatic fields, making available a new radiation in microscopy. Such a beam can be considered by the De Broglie concept to have a wave length of the order 0.01 Å. It should therefore be possible, theoretically, to photograph atoms which have diameters of the order 1 Å. Practical limitations have so far restricted the resolution to a level not much below 50 Å., but it should still be possible to study most proteins by this means.

TABLE 13-3. The Dimensions of Tobacco Mosaic Virus Particles¹

Methods	Diameter, Å.	Length, Å.	Molecular Weight
Sedimentation and viscosity	136	2760	33.2×10^6
Sedimentation and diffusion	138	2560	31.6×10^6
Viscosity and diffusion	140	2830	36.0×10^6
Electron microscope	152	2700	40×10^6

¹ From M. A. Lauffer, *J. Am. Chem. Soc.*, **66**, 1188 (1944).

Beautiful electron micrographs have been obtained of very large proteins, the viruses. The "molecules" of tobacco mosaic virus are found to be rodlets, primarily 2700 by 150 Å. though there is some distribution in size, which may arise in large part from the fragmentation of the particles in preparation of the specimens. Some examples



A

Fig. 13-6. Electron micrographs of protein crystals. A, crystal of one of the tobacco necrosis virus proteins, showing the way its molecules are arranged. Each of these molecules has a diameter of about 250 Å. Note dissolution which has occurred at the lower right-hand corner. B, two crystals of a tobacco protein, showing their molecular arrangement. These molecules have diameters of about 130 Å. Reproduced through the courtesy of Dr. R. W. G. Wyckoff.

of electron micrographs are shown in Fig. 13-6. Because of the low electron-scattering power of organic materials, such pictures are frequently obtained by "shadowing" with metal. This is done by evaporating metal onto the specimen (in vacuum) at an angle, so that in effect the specimen casts its shadow in the deposited metal. The metal deposit is then photographed in the electron beam, reproducing the protein indirectly.

Table 13-3, taken from a paper by Lauffer, summarizes information on the size and shape of tobacco mosaic virus as determined by various physical methods. The confirmation, by direct observation with the electron microscope, of the values calculated by indirect physical methods is most satisfying and from one point of view somewhat



surprising. One of the limitations of the electron microscope as applied to biological materials is the need of operating under high vacuum, hence of using only thoroughly desiccated samples. The electron microscope cannot in general be expected to yield a picture of the solvated protein molecule which exists in solution.

Other viruses have been studied in the electron microscope. As yet it has not been possible to examine by this means smaller proteins, especially enzymes and antibodies. One problem involved is that of finding surfaces of a high enough degree of smoothness on which to prepare the specimens for photography. The best polished metal or diamond surface becomes, on this scale, a rugged mountain range. Successful pictures of hemoglobin have been obtained, but some doubt in interpretation remains, owing to the possibility of denaturation during preparation of the specimen.

METHODS FOR OBTAINING FURTHER STRUCTURAL DETAILS

Monolayer Studies. In many solutions the solute has a profound tendency to concentrate in the interface, in which case the solute is said to be surface-active. Examination of such systems by Langmuir and others has shown that this results from surface adsorption of the solute in the form of a monomolecular layer. If such a monolayer is confined in a shallow trough between two movable surface barriers, one of which is attached to a sensitive balance, it is possible to find the force on this latter barrier as a function of the area on which the monolayer is confined. Force-area curves can be plotted and are analogous to the pressure-volume curves given, for example, by gases in conventional three-dimensional space. Gaseous, liquid, and solid regions can be identified as the area is decreased by increasing the lateral force on the surface film. In the gaseous region, application of the gas laws is sometimes made to determine the molecular weight of the spread substance. Study of the solid region, by measuring the area occupied by a given weight of solute, yields the film thickness, a measure of one of the dimensions of the spread molecule. For proteins the latter value is usually found to be about 8 to 10 Å., much less than any of the dimensions of the molecule in solution if reliance can be placed in the various physical measurements. This discrepancy has led to some interesting ideas with regard to protein structure, as will be pointed out in Chap. 16.

It is further possible to build up polymolecular layers by raising and lowering a plate through a spread monolayer. By such means, films of as many as 1700 molecular layers have been formed. Such films can be studied by X-ray diffraction techniques, and the thickness can be measured by ordinary interferometric methods. The thickness was found by this means to correspond to about 10 Å. per monolayer, in agreement with the surface-balance method.

X-Ray Diffraction. X-rays result when metals are bombarded with high-voltage electrons, and can be considered as electromagnetic radiations of wave length in the range of a few hundredths to a few hundred angstrom units. Hence they are potentially capable of yielding detail on an atomic scale. It has not yet been feasible to devise X-ray lenses to utilize the potential resolving power, although suggestions as to possible means of doing this have been made. Nevertheless, much can be learned about the location of atoms in a crystalline material by observing the characteristic reflections (scattering) of X-rays from the atomic planes. In general the scattered rays are eliminated by destructive interference, but at certain charac-

teristic angles of incidence (and reflection) constructive interference results, the waves reflected from various planes being retarded by whole numbers of wave lengths. This is summarized in the Bragg law, which forms the basis for X-ray crystallography:

$$n\lambda = 2d \sin \theta \quad (9)$$

where n is a whole number, the "order" of the reflection; λ the wave length of the radiation; d the characteristic spacing of the atomic planes in question; and θ the angle of incidence.

The essence of the method lies in irradiating a crystal* with X-rays and photographing the reflected rays in a suitable manner, measuring the $\sin \theta$ values of the various resulting spots, arcs, or circles (depending on the technique used), and identifying each of these with the various orders of the various spacings in the crystal. One can then compute the size and shape of the *unit cell* (the unit which, when repeated along the various axes, can reproduce the crystal), the number of molecules in the unit cell, something of the symmetry of the molecule in some cases, and eventually the position of each atom involved. The last step involves measurement of the relative intensities of various scatterings and a synthesis, by mathematical means, of a three-dimensional plot of the electron density (the scattering is primarily due to electrons). The most complex structure which has been carried through to this ultimate state is that of benzylpenicillin. This molecule contains 24 atoms, exclusive of hydrogen, but does not begin to approach in complexity the proteins which may contain several thousand atoms.

Astbury studied proteins of the fibrous type, such as silk fibroin and keratin. He distinguished between the so-called α -keratin and β -keratin fibrillar structures, the X-ray patterns of which are entirely distinct. In the α -keratin structure there is a characteristic repeat period of 5.1 Å. along the fiber axis, in β -keratin, a repeat period of 3.4 Å. Some natural fibers, such as silk fibroin, occur in an elongated configuration resembling the β -keratin configuration. Keratin occurs

* If a single crystal is used with monochromatic X-rays, it is necessary to rotate or oscillate the crystal so that the various possible reflections will be brought out. Rotation about various axes will lead to different sets of reflections, and this greatly facilitates the indexing of the pattern. A mass of small crystals is sometimes used to yield a "powder" pattern. No rotation is then necessary, since statistically there will be crystals oriented in all possible ways and all the various reflections will occur. However, one does not then have the advantage of knowing which reflections are associated with which axes, and indexing is much more difficult, in many cases impossible.

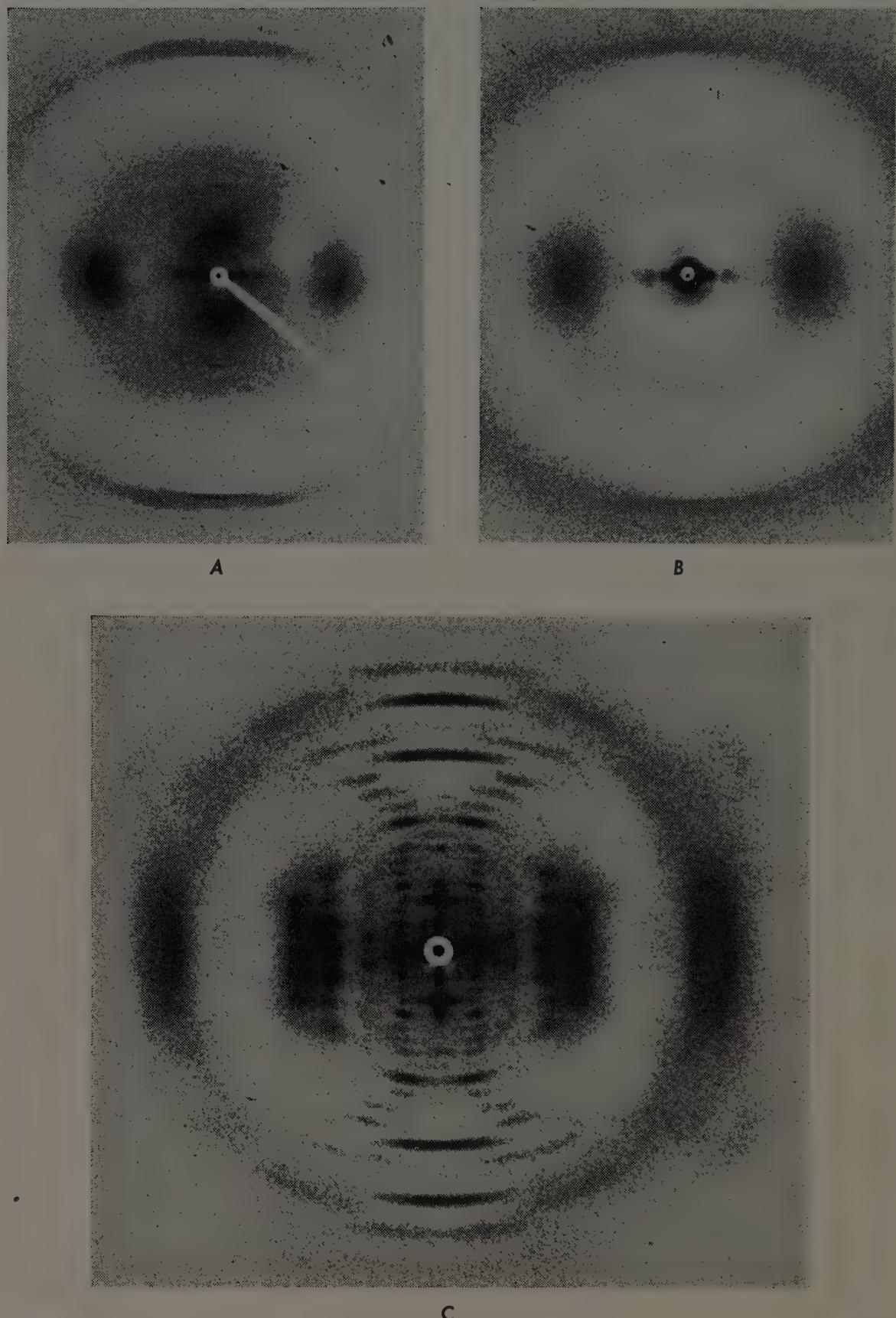


Fig. 13-7. X-ray diffraction patterns: A, porcupine quill; B, unstretched human hair; C, sea gull feather rachis. Patterns A and B are typical of the α -keratin configuration, pattern C of β -keratin. In all cases the fiber axis is vertical. Reproduced from R. S. Bear and H. J. Rugo, Ann. N. Y. Acad. Sci., 53, 627 (1951), through the courtesy of Professor Bear.

naturally in the α -configuration, but goes over to the β -configuration upon stretching.

Astbury found two other characteristic spacings in the β -configuration. One of these, about 10 Å., he associated with the "thickness" of the peptide chain in the direction of the amino acid side chains. As is to be expected, this spacing varies depending on the amino composition of the protein involved. The other, about 4.5 Å., is more constant and is presumably the lateral or "backbone" width of the peptide chain. The significance of these spacings is discussed further in Chap. 16 and may be clarified by inspection of Fig. 16-1. Recent work, principally by Bear, indicates the presence in many fibrillar proteins of long spacings, ranging from about 100 to 700 Å.

The fibrillar proteins do not give true crystals; instead they contain both ordered and amorphous regions. A single fibrillar molecule may run through several alternate crystalline and amorphous regions (Fig. 16-2). Crystallographic studies of these proteins can obviously never yield a picture of the molecule as a whole, though they can yield valuable information with regard to the configuration of peptide units. On the other hand, some globular proteins, such as insulin, hemoglobin, and β -lactoglobulin, give true crystals, entire protein molecules being arranged in a definite space lattice. X-ray diffraction has yielded information as to the molecular weight of such proteins, and in a few cases some information as to the symmetry of the molecule.

The most complete crystallographic studies on proteins to date are those of Perutz and co-workers on hemoglobin. These results will be discussed in Chap. 18. Even this work, however, encompasses only the gross features of the molecule. Actual location of the atoms or even the amino acid residues in a protein appears at the moment to be an almost overwhelming task. Nevertheless, the large amount of effort being expended in this direction can be expected to yield valuable additional information on the internal structure of the protein molecule.

SELECTED REFERENCES

Books and Reviews

The Ultracentrifuge, The Svedberg and K. O. Pedersen, Clarendon Press, Oxford, 1940.

"The Investigation of Proteins by Diffusion Measurements," H. Neurath, *Chem. Revs.*, **30**, 357 (1942).

"The Investigation of Proteins by Dielectric Measurements," J. L. Oncley, *Chem. Revs.*, **30**, 433 (1942).

"Rotary Brownian Movement: The Shape of Protein Molecules as Determined from Viscosity and Double Refraction of Flow," J. T. Edsall, Chap. 21 in Cohn and Edsall's *Proteins, Amino Acids and Peptides*, Reinhold, New York, 1943.

"X-Ray Diffraction and Protein Structure," J. T. Edsall, Chap. 14 in Cohn and Edsall's *Proteins, Amino Acids and Peptides*, Reinhold, New York, 1943.

"Spread Monolayers of Proteins," H. B. Bull, *Advances in Protein Chem.*, **3**, 95 (1947).

"X-Ray Studies on Amino Acids and Peptides," R. B. Corey, *Advances in Protein Chem.*, **4**, 385 (1948).

"The Scattering of Light and Its Applications to Chemistry," G. Oster, *Chem. Revs.*, **43**, 319 (1948).

"The Electron Microscopy of Macromolecules," R. W. G. Wyckoff, *Advances in Protein Chem.*, **6**, 1 (1951).

"Light Scattering in Protein Solutions," P. Doty and J. T. Edsall, *Advances in Protein Chem.*, **6**, 35 (1951).

"The Size, Shape and Hydration of Protein Molecules," J. T. Edsall in Neurath and Bailey's *The Proteins*, Vol. IB, Academic Press, New York, 1953.

"Rotational Brownian Motion and Polarization of the Fluorescence of Solutions," G. Weber, *Advances in Protein Chem.*, **8**, 416 (1953).

Light-Scattering in Physical Chemistry, K. Stacey, Academic Press, New York, 1956.

"Measurement and Interpretation of Diffusion Coefficients of Proteins," L. J. Gosting, *Advances in Protein Chem.*, **11**, 429 (1956).

Original Articles

"The Apparent Shape of Protein Molecules," H. Neurath, *J. Am. Chem. Soc.*, **61**, 1841 (1939)

"Studies in the Physical Chemistry of Insulin," E. J. Cohn, J. D. Ferry, J. J. Livingood, and M. H. Blanchard, *Science*, **90**, 183 (1939).

"Evidence from Physical Chemistry Regarding the Size and Shape of Protein Molecules from Ultra-centrifugation, Diffusion, Viscosity, Dielectric dispersion and Double refraction of flow," J. L. Oncley, *Ann. N. Y. Acad. Sci.*, **41**, 121 (1941).

"Evidence from X-rays Regarding the Structure of Protein Molecules," I. Fankuchen, *Ann. N. Y. Acad. Sci.*, **41**, 157 (1941).

"The Size and Shape of Tobacco Mosaic Virus Particles," M. A. Lauffer, *J. Am. Chem. Soc.*, **66**, 1188 (1944).

"Studies on Double Refraction of Flow. II. The Molecular Dimensions of Zein," J. F. Foster and J. T. Edsall, *J. Am. Chem. Soc.*, **67**, 617 (1945).

"Osmotic Pressure of β -Lactoglobulin Solutions," H. B. Bull and B. T. Currie, *J. Am. Chem. Soc.*, **68**, 742 (1946).

"Physical-Chemical Characteristics of Certain of the Proteins of Normal Human Plasma," J. L. Oncley, G. Scatchard, and A. Brown, *J. Phys. & Colloid Chem.*, **51**, 184 (1947).

"Light Scattering in Solutions of Serum Albumin: Effects of Charge and Ionic Strength," J. T. Edsall, H. Edelhoch, R. Lontie, and F. R. Morrison, *J. Am. Chem. Soc.*, **72**, 4641 (1950).

"An X-Ray Investigation of the Shapes and Hydrations of Several Protein

Molecules in Solution," H. N. Ritland, P. Kaesberg, and W. W. Beeman, *J. Chem. Phys.*, **18**, 1237 (1950).

"Electron Microscopy of Crystalline Edestin," C. E. Hall, *J. Biol. Chem.*, **185**, 45 (1950).

"Electron Microscopic Observations on the Unit of Length of the Particles of Tobacco Mosaic Virus," R. C. Williams and R. L. Steere, *J. Am. Chem. Soc.*, **73**, 2057 (1951).

"Molecular Weight of Lactoglobulin, Ovalbumin, Lysozyme and Serum Albumin by Light Scattering," M. Halwer, G. C. Nutting, and B. A. Brice, *J. Am. Chem. Soc.*, **73**, 2786 (1951).

"The Influence of Dipole Moment Fluctuations on the Dielectric Increment of Proteins in Solutions," J. Kirkwood and J. Shumaker, *Proc. Natl. Acad. Sci. U. S.*, **38**, 855 (1952).

"Small Angle X-ray Scattering from Serum Albumin Solutions," J. Anderegg and W. Beeman, *Physiol. Revs.*, **32**, 186 (1952).

"Consideration of the Hydrodynamic Properties of Proteins," H. Scheraga and L. Mandelkern, *J. Am. Chem. Soc.*, **75**, 179 (1953).

"The Size and Shape of Bovine Fibrinogen. Studies of Sedimentation, Diffusion, and Viscosity," S. Shulman, *J. Am. Chem. Soc.*, **75**, 5846 (1953).

"Determination of the Form and Dimensions of the Serum Albumin Molecule in Solution," M. Champagne, *Compt. rend.*, **237**, 521 (1953).

"Molecular-Kinetic Properties of Crystalline Trypsinogen," F. Tietze, *J. Biol. Chem.*, **204**, 1 (1953).

"Certain Physical Properties of Bovine Prothrombin," F. Lamy and D. Waugh, *J. Biol. Chem.*, **203**, 489 (1953).

"Ultracentrifuge Studies with a Synthetic Boundary Cell," H. K. Schachman and W. F. Harrington, *J. Polymer Sci.*, **12**, 379 (1954).

"Viscometric Investigation into the Influence of pH and Ionic Strength on the Molecular Shape of a Structural Muscle Protein," A. Wassermann, *Rec. trav. chim.*, **74**, 978 (1955).

"Determination of the Shape of γ -Globulin by Means of the X-ray Low Angle Method," O. Kratky, G. Porod, A. Sekora, and B. Paletta, *J. Polymer Sci.*, **16**, 163 (1955).

"Simultaneous Determination of Molecular Weights and Sedimentation Constants," S. M. Klainer and G. Kegeles, *J. Phys. Chem.*, **59**, 952 (1955).

The Solubility Behavior of Proteins

The forces involved

Factors governing the solubility of proteins

Binding of ions by proteins

Solubility as a criterion of purity

Protein-protein interaction

Gelation of proteins

The proteins differ widely in solubility, ranging from the very insoluble keratin through the prolamines, soluble in aqueous alcohol or in water at high *pH*, to the albumins, which are soluble to the extent of 25 per cent or more in pure water. The solubility behavior of proteins is of extreme importance from several points of view. From the practical standpoint, intelligent handling of proteins in the processes of purification and fractionation demands considerable knowledge of the solubility properties of the proteins in question. In addition, however, some light can be cast on the structure of proteins from a consideration of their solubility properties.

It is of interest that the historical classification scheme for the simple (unconjugated) proteins was based largely on solubility behavior. Those proteins soluble in water at the isoelectric point were termed *albumins*, those insoluble in water but solubilized by low concentrations of salt were classified as *globulins*. It has been observed that globulins are more readily "salted out" of solution than are albumins. *Glutelins* are proteins insoluble in all neutral solvents but readily soluble in dilute acid or alkali. The *prolamines* have the peculiar property of being soluble in mixtures of ethanol and water

but insoluble in either of the pure solvents. The *scleroproteins*, which are the proteins of structural tissue, such as collagen, elastin, and keratin, are insoluble in all of the solvents mentioned.

It is not the purpose of this chapter to survey the experimental data which are available on the solubility of proteins. Rather, an attempt is made to present the fundamental principles underlying the solubility of proteins, and certain closely related problems such as gelation and ion binding are discussed.

THE FORCES INVOLVED

Four types of forces significantly affect the solubility behavior of proteins: ionic attraction, dipolar forces, hydrogen bonds, and van der Waals forces.

Ionic Attraction. The fundamental law governing the force of attraction between oppositely charged ions is known as Coulomb's law, which may be written as

$$f = \frac{q_1 \cdot q_2}{D \cdot r^2} \quad (1)$$

Here f is the force of attraction between two charges q_1 and q_2 , separated by a distance r , the intervening space having a dielectric constant D . The force falls off quite rapidly with increasing distance of separation, since r enters in the denominator as the square. Other forces drop off even more rapidly with increasing distance, so that at long range (greater than say 5 A. or so) the important forces are in all probability coulombic.

It is important to emphasize the dielectric constant D . In an ionic crystal the space between ions is merely a vacuum, having a dielectric constant of unity. The force of attraction in ionic crystals is enormous and may easily be equal to that of an ordinary covalent bond. Yet when sodium chloride is placed in water, it passes into solution very readily. This is chiefly a result of the fact that water, having a dielectric constant 80 times that of a vacuum, reduces the force of attraction between the ions by this factor. Moreover, the ions have a tendency to hydrate, the solvation layer making it impossible for the ions to approach one another as closely as in the crystal, thus further lowering the force of attraction. Indeed, the force of attraction between sodium ions and chloride ions in aqueous solution is so low that for all practical purposes they can best be considered as being completely dissociated.

Coulomb's law is valid for the case where the charge can be considered as a point. In proteins we have a much more complex

situation, both positive and negative charges being distributed over the surface of a giant molecule, as in Fig. 14-1. It might still be expected that the simple Coulomb law would be applicable in the case of the force between an ion and a singly charged group of the protein if the ions were close enough to one charged group and far enough from all others that the effect of the latter could be ignored.

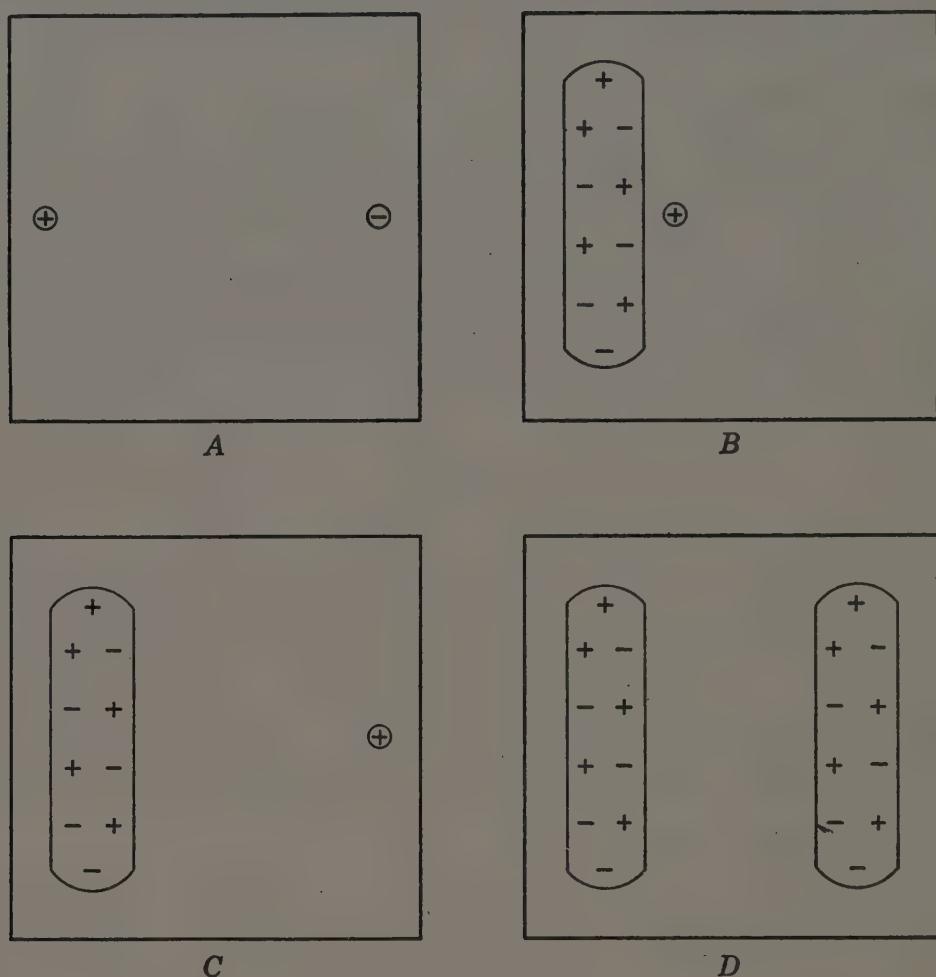
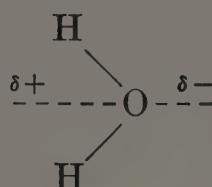


Fig. 14-1. Schematic illustration of various important cases of interaction between charged particles. In A the simple Coulomb law would apply; in B, C, and D the situation would be more complex.

If it were farther from the protein, however, so that all charged groups would have to be considered, the interaction would be more complex. Even more complicated is the force of attraction between two protein molecules, each carrying many charged groups of both signs. This problem is so complex that it has not yet been possible to treat it adequately.

Dipolar Forces and Hydrogen Bonds. The high dielectric constant of water has been mentioned. It is because of this property that

water is the "universal solvent" for ionic materials. All substances have a dielectric constant greater than unity, that is, greater than that of a vacuum. Substances of high dielectric constant are usually found to possess permanent dipoles; the centers of their positive and negative charges do not coincide. Water, for example, as a result of its angular nature,



possesses a permanent dipole along the bisecting axis. The covalent hydrogen-oxygen bonds are "polarized" to a certain extent, since the oxygen atom has a stronger affinity for the binding electrons than has the hydrogen atom; hence the tendency for water to dissociate into hydrogen and hydroxyl ions. As a consequence the electron density is heavier toward the oxygen end of the molecule.

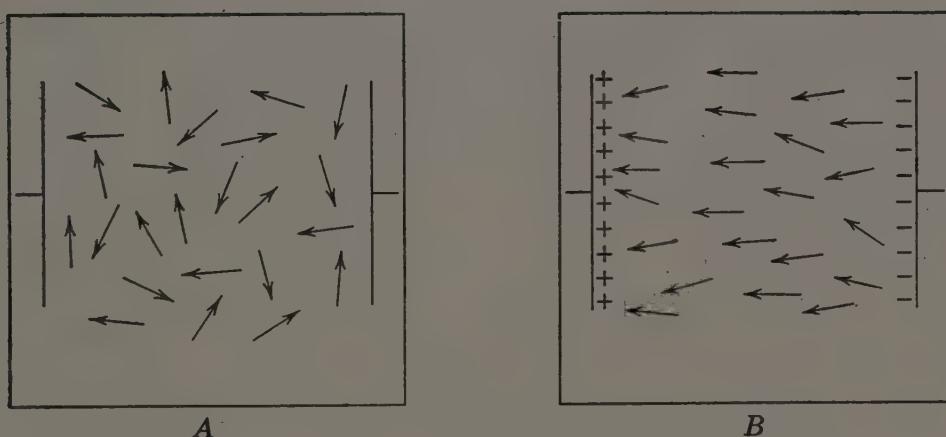


Fig. 14-2. Orientation of dipoles by an electric field: A, random arrangement prior to application of the field; B, orientation under influence of the field.

In the absence of an electric field, dipolar molecules in the bulk will have a random arrangement (A of Fig. 14-2). When placed between the charged plates of a condenser or between positive and negative ions, they tend to orient with their negative ends toward the positive pole and vice versa (B of Fig. 14-2). This results in a diminution of the force of attraction between the poles.

Even in the absence of an electric field one would find that the random distribution did not prevail if he could examine sufficiently small volumes of the medium. In water there are small clusters of

water molecules in which the individual molecules are arranged in a more or less definite crystalline arrangement. The hydrogen atoms of one water molecule tend to be aligned toward the oxygens of other water molecules. In this way each oxygen atom tends to surround itself, in a tetrahedral configuration, with four hydrogen atoms of which two can be considered as being held by covalent bonds and two held by electrostatic attraction of the positive protons for negative oxygen. A given water molecule moves about from one cluster to another, and its hydrogen atoms undergo frequent exchange with the hydrogen atoms of other molecules through ionization.

Such dipole-dipole attractions must always be considered when dealing with polar molecules, but are especially important in the case of water and in certain other hydrogen-containing molecules. This results from the fact that the hydrogen atom, having only a single electron, becomes essentially a bare nucleus when involved in a partially polarized covalent bond, and hence has an unusual affinity for the negative moiety of other molecules. Such attractive forces, though relatively weak as compared to true covalent linkages (energy of the order of 5000 or so calories per mole as compared to 50,000 to 100,000 for covalent bonds), are of such importance that they have been graced with a name, the *hydrogen bond* (or *hydrogen bridge*).

Hydrogen bonds are strongest in H_2F_2 , somewhat less strong in compounds containing —OH groups such as water and alcohols, weaker yet but still important in —NH— compounds such as amines and amides. The peptide linkages in proteins provide excellent opportunities for the formation of hydrogen bonds (Chap. 16).

As a consequence of the force of attraction of ions for dipoles there usually results, when ions are dissolved in water, a partial breakdown of the water clusters and a reorientation of the water molecules around the ions. This process is called *hydration*. Since the tetrahedral clusters are loosely packed (contain considerable waste space), the hydration process usually involves a decrease in total volume of water. The volume of a solution containing an ionic solute may actually be less than that of the water alone. This shrinkage in volume, termed *electrostriction*, may be regarded as a manifestation of the strong forces of interaction between ions and dipoles.

Mention should also be made of other dipolar forces. Molecules of themselves symmetrical, and hence carrying no permanent dipole moment, may be polarized in the presence of an electric field. This is due primarily to a shift of the electrons from their normal average position toward the positive pole, or electronic polarization. It is for this reason that all matter has a dielectric constant greater than

unity. Such an induced dipole will be attracted by ions in a manner similar to that in which permanent dipoles are attracted.

Van der Waals Forces. Finally, we must mention the forces of attraction which always exist between atoms and molecules, even though they are completely uncharged and carry no permanent dipole. These are the *van der Waals* or *London dispersion* forces. They result from the fact that electronic symmetry in molecules is a time-average situation; at a given instant it is most probable that there will be some dipole moment extant. This moment can influence the distribution of electrons in a neighboring molecule. Two closely neighboring molecules can thus mutually induce dipoles in one another, the direction of the dipoles oscillating in phase at high frequency. This can cause a considerable force of attraction, amounting to as much as 3000 cal. per mole. The strength of van der Waals forces drops off very rapidly with distance, becoming negligible at distances of even 5–10 Å. Closeness of fit is therefore an important factor in determining the strength of such forces.

FACTORS GOVERNING THE SOLUBILITY OF PROTEINS

The solubility of a solute *A* in a solvent *B* is a consequence of several factors. Obviously of significance is the force of attraction between the atoms or molecules that comprise the solute. The greater such forces, the lower the solubility should be, other factors being the same, since such forces must be overcome in the solution process. (The same forces must be overcome in fusion.) The solution process also involves the separation of solvent molecules to form "holes" for the solute. Both of these processes require an input or expenditure of energy, and solution would not ordinarily be expected to take place under such conditions. A counterbalancing source of energy is needed; this comes from the formation of new forces of attraction between solvent and solute, or *solvation*. Thus, ionic crystals dissolve in water in spite of the high ion-ion forces and water-water forces because of the compensating effect of high water-ion forces. Nonpolar organic compounds dissolve in hydrocarbons because all forces involved are small (van der Waals). They do not dissolve in water, because the necessity of breaking strong water-water forces is not compensated by correspondingly high forces of solvation. It is such considerations that provide a theoretical basis for the rule "like dissolves like."

In this discussion, mention of the entropy changes involved in the solution process has been tacitly avoided. The actual solubility will depend on the free-energy change involved. This is the difference between the total energy change and the entropy change multiplied

by the absolute temperature:

$$\Delta F = \Delta H - T\Delta S \quad (2)$$

where ΔS is the entropy change and T the absolute temperature.

The more negative ΔF the greater the solubility, so that a large increase in entropy favors solution. The entropy of a crystalline solute will increase upon dissolving, since entropy is essentially a measure of randomness and the solute in solution represents a much more random state than does the crystal. In many cases, however, this is compensated or more than compensated by a negative entropy change due to the solvation process (increased order of the solvent). The relations become very complex. For an intensive treatment of the many considerations involved, the reader is referred to an excellent chapter in the book by O. K. Rice (1940) cited in the reference list.

Solubility and Composition. The proteins which command the most biological interest are soluble in water or in dilute salt solution, that is, they are either albumins or globulins, but proteins with very different solubility properties are known. The factors influencing solubility of such complex substances as proteins are multifarious, and it is desirable to study as many as possible of these effects separately.

To do this, it is of interest to examine first the effect of the amino acid make-up of the protein. If consideration is limited to the solubility of the isoelectric protein in water, in which case the solubilizing influence of the charged groups is at a minimum, a crude correlation between composition and solubility is observed. For convenience, the amino acids can be divided into three classes, depending on the nature of their side chains: (1) *nonpolar*, including alanine, valine, leucine and isoleucine, phenylalanine, cystine, methionine, and proline, which will tend to reduce the solubility of the protein in water and increase its solubility in solvents of low dielectric constant (such as alcohol); (2) *ionic*, those side chains which are capable of contributing charged groups at the isoelectric point of the protein, including lysine, arginine, and histidine (cationic), and aspartic and glutamic acids (anionic); (3) *polar*, including serine, threonine, cysteine, tyrosine, hydroxyproline, and any dicarboxylic acids whose side chains exist in the form of amides. Both ionic and polar groups can be expected to contribute to the water solubility, though ionic groups are doubtless much more important in this respect.

Analysis of the amino acid composition of several well characterized proteins indicates that the water-soluble or salt-soluble proteins have roughly equal amounts of residues of the amino acids classified above as polar or ionic and nonpolar. The ratio of polar plus ionic to

nonpolar residues is for ovalbumin about 0.9 (based on relatively incomplete analysis); human serum albumin, 1.4; γ -globulin, 1.0; β -globulin, 1.2; fibrinogen, 1.6; gelatin, 0.65; and insulin, 1.0. The α -globulin component of human blood serum shows an unusually high ratio of 2.8. The prolamines, which are soluble in media of low dielectric constant (alcohol-water), have a low ratio of polar plus ionic to nonpolar groups, approximately 0.5 in the case of zein. Furthermore, although these proteins have a fairly high content of the dibasic amino acids, most of the side-chain carboxyl groups are present as the amides. The content of ionic groups is extremely small indeed.

The highly insoluble skeletal proteins, the keratins, are insoluble primarily because their molecules, if one can properly speak of keratin molecules, are laced together into large aggregates. This cross linking is thought to be due primarily to disulfide linkages (note that keratins contain up to nearly 20 per cent of cystine), and keratin is usually dissolved by treating with a reducing agent such as Na_2S , which serves to reduce the disulfide groups to —SH groups. Such a process might more properly be considered a degradation rather than a simple solution process.

The Effect of pH. The solubility of proteins in water is characteristically minimal at the isoelectric point; indeed, solubility measurement provides one of the commonly used methods of determining this property (Fig. 14-3). This situation is commonly described by saying that proteins are more soluble when combined with bases or acids than in the neutral state. As a first approximation it is better to consider the proteins as combining with or releasing only hydrogen ions, as discussed in Chap. 11. (Some deviations from this ideal behavior will be considered later in this chapter.) The effect of adding acid is to increase the positive charge of the protein through the binding of protons; alkali increases the net negative charge. There is a close enough correlation between water solubility and net charge (as estimated from electrophoretic mobility) to indicate that this charge is the important factor governing the effect of pH on solubility.

It is not possible to give a quantitative analysis of so complex a situation, but qualitatively one can see that forces of repulsion between the protein molecules would be much less in the isoelectric state than under conditions in which the protein possesses a strong net charge, either positive or negative.

The Effect of Salt Concentration. The water solubility of most proteins at their isoelectric points is increased by the addition of small amounts of salt. This effect is illustrated in Fig. 14-4.

The salt concentration is expressed along the abscissa, in this figure,

in terms of the "ionic strength" $\Gamma/2$ or μ , defined as

$$\mu = \frac{\Gamma}{2} = \sum_i \frac{C_i Z_i^2}{2} \quad (3)$$

where C_i is the molar concentration of the i th species of ion, and Z_i is the net charge of the corresponding ion.

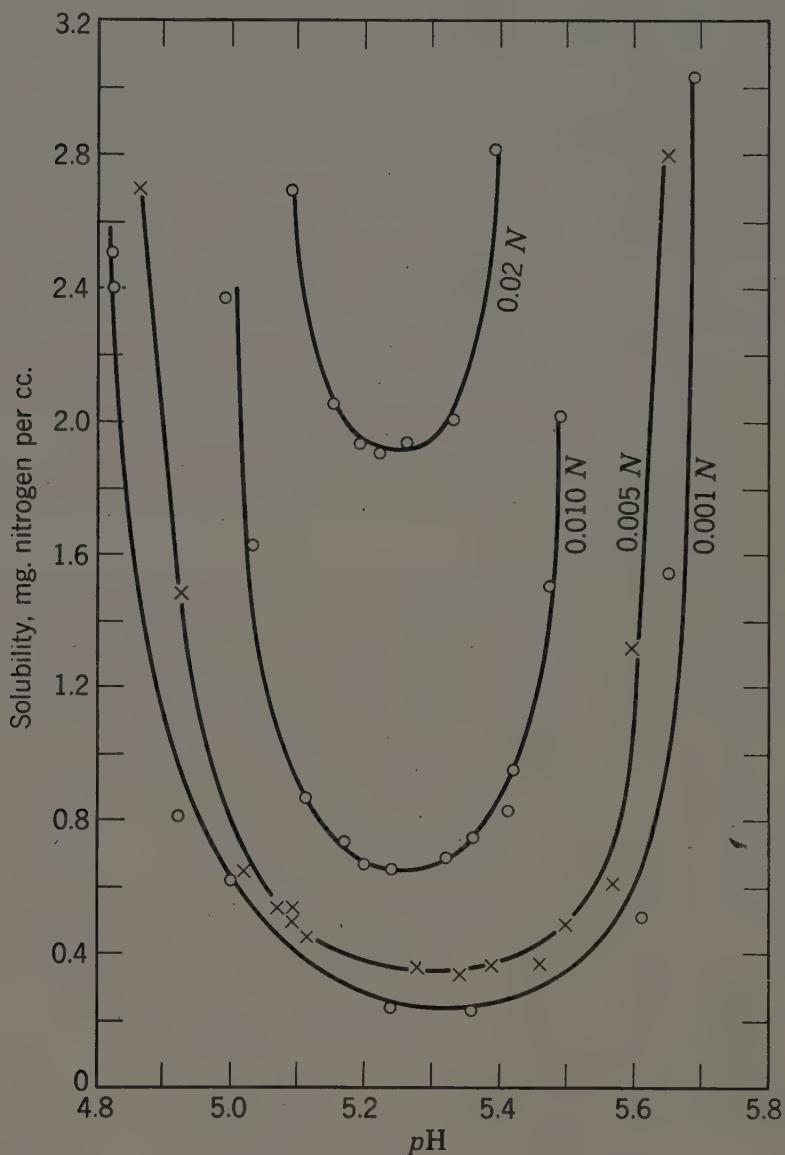


Fig. 14-3. Solubility of β -lactoglobulin as a function of pH at four different concentrations of sodium chloride. From the data of A. Gronwall, Compt. rend. trav. lab. Carlsberg, Sér. chim., 24, No. 8 (1942).

The solubility is usually expressed as $\log S/S'$, the logarithm of the ratio of the solubility S at a given ionic strength, to the solubility S' at zero ionic strength. (The reason for using these functions will later be apparent.)

The curves for various proteins are similar in character but differ in detail. The solubility is found to increase with increasing ionic

strength, to pass through a maximum, and then to decrease (Fig. 14-5). Thus, at low concentrations ions have a *salting-in* effect, at high concentrations a *salting-out* effect. This behavior is not limited to proteins but is observed also with the amino acids and most ionic substances. With globulins, the solubility at zero ionic strength is negligible and the salting-in effect is pronounced. Albumins, on the other hand, since they are readily soluble in the absence of salt, may

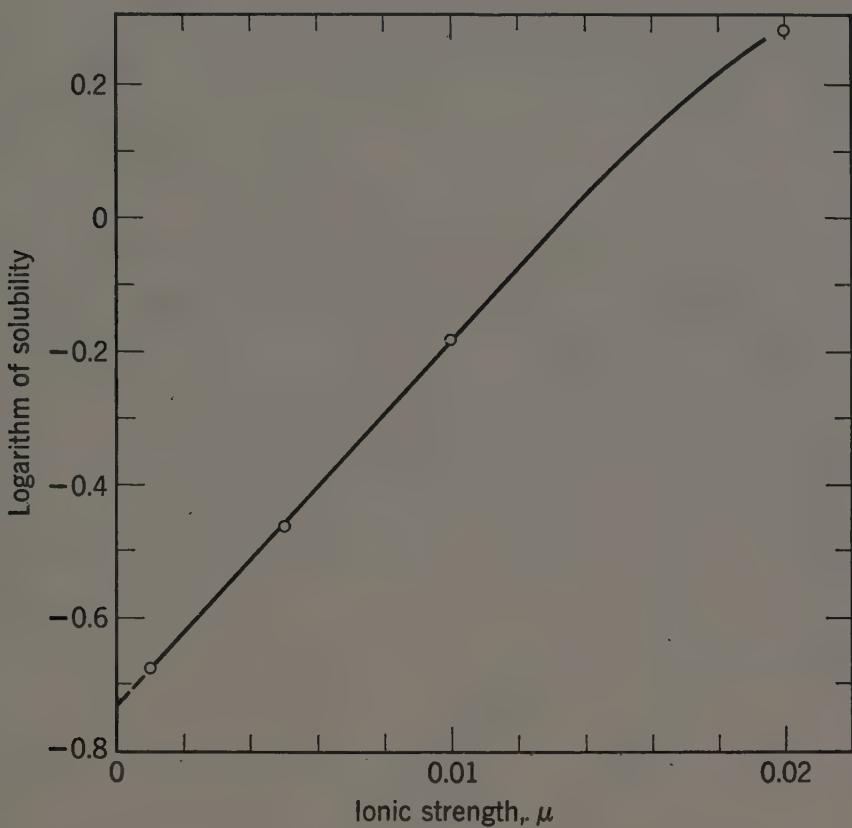


Fig. 14-4. Solubility of β -lactoglobulin as a function of ionic strength. From the data of A Gronwall, *Compt. rend. trav. lab. Carlsberg, Sér. chim.*, **24**, No. 8 (1942).

not exhibit much of a salting-in effect. Addition of alcohol to such systems reduces the solubility, producing a typical salting-in region.

It was long ago observed that polyvalent ions are more effective, on an equivalent basis, than monovalent ions in exerting both salting-in and salting-out effects. Qualitative theoretical justification for this may be seen in the following way. Considering the equilibrium between solution and undissolved solute, it is apparent on the basis of thermodynamic principles that the state of saturation is that in which the activity of the solute is the same in both solution and solute, since this is a necessary requirement for equilibrium. It is customary to take the activity of a solute as equal to the product of two factors, the concentration and the activity coefficient. Assuming the activity of the solute in the solid state to be a constant, independent of the

salt concentration,

$$S\gamma = k \quad \text{or} \quad S = \frac{k}{\gamma} \quad (4)$$

where S is the solubility and γ the activity coefficient of the solute in solution. This coefficient is defined as unity in the infinitely dilute solution, where the solute is supposed to behave ideally, but may in the general case be either greater or less than unity.

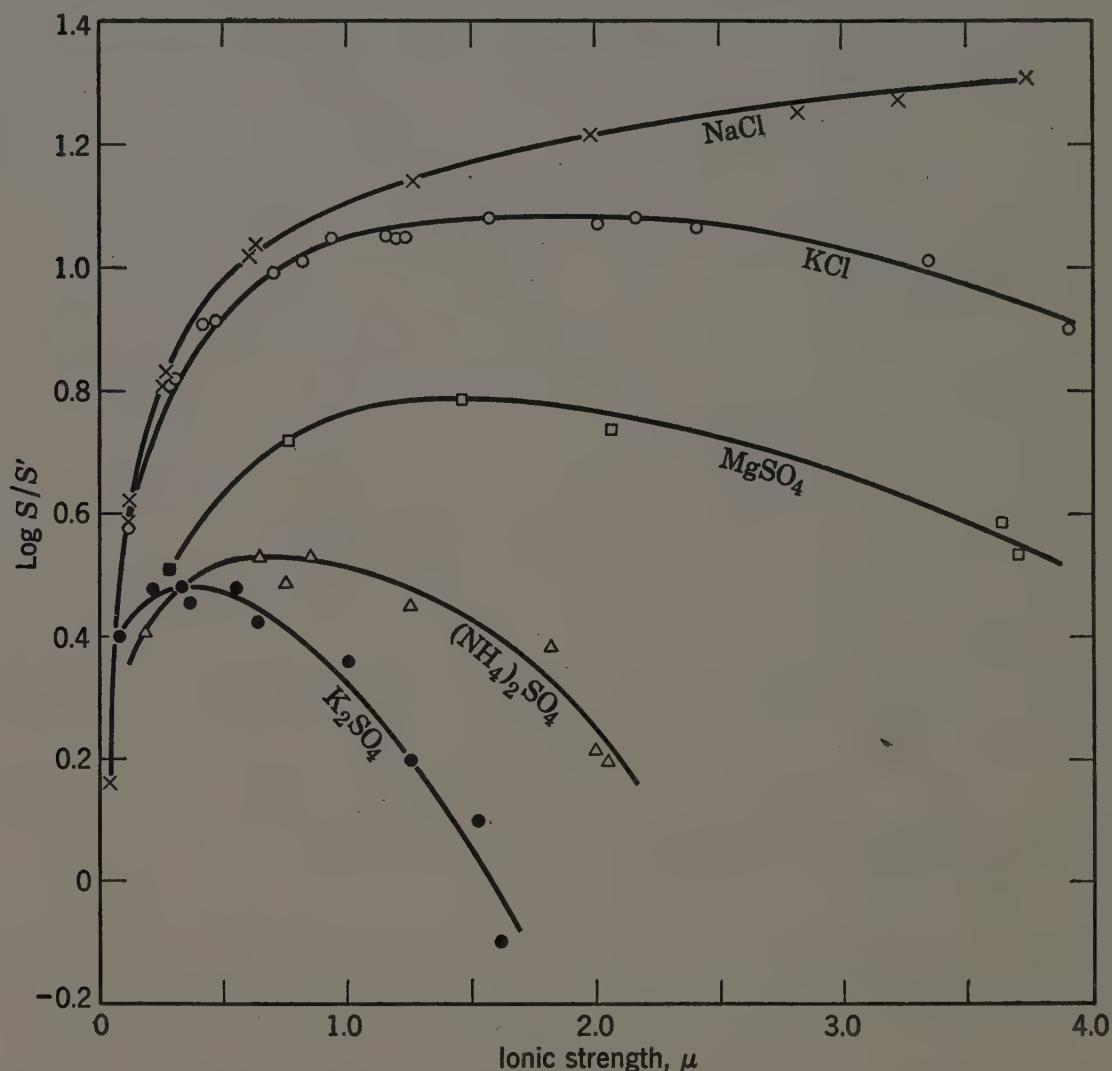


Fig. 14-5. Solubility of isoelectric carboxy hemoglobin as a function of ionic strength and ion type. From data of A. A. Green, *J. Biol. Chem.*, 95, 47 (1932).

Debye and Hückel have derived an expression relating the activity coefficient of a given ion to the ionic strength of the form

$$-\log \gamma = \frac{1.81 \times 10^6}{D^{3/2} T^{3/2}} Z_1 Z_2 \sqrt{\Gamma/2} \quad (5)$$

where D is the dielectric constant of the medium, T the absolute

temperature, and Z_1 and Z_2 the valences of the ions of the salt. Combining equations 4 and 5, it is seen that

$$\log S = \log k - \log \gamma = \log k + \frac{1.81 \times 10^6}{D^{3/2} T^{3/2}} Z_1 Z_2 \sqrt{\Gamma/2} \quad (6)$$

This equation predicts that the logarithm of the solubility should be proportional to the square root of the ionic strength, a result which is borne out by experiment with simple electrolytes. The proportionality constant depends on Z_1 and Z_2 , the ionic valences, thus explaining the fact that polyvalent electrolytes are salted-in more readily than univalent ones. Further, polyvalent electrolytes are more effective in increasing the ionic strength (equation 3) and are hence more efficient in producing the salting-in effect.

The Debye-Hückel relation can be considered as only a very rough approximation for such complex ions as proteins or even amino acids. A more involved treatment for large spherical particles with a dipole moment, and a much better model for a protein molecule than the point charge assumed in the Debye-Hückel treatment, predicts a proportionality between the logarithm of the solubility and the first power of the ionic strength. This theory is obeyed rather well by β -lactoglobulin as shown in Fig. 14-4.

The effect of small dipolar ions such as isoelectric amino acids and peptides on the solubility of proteins can be attributed almost quantitatively to the effect of the dipolar ion in enhancing the dielectric constant of the medium. Figure 14-6 shows that the logarithm of the solubility of β -lactoglobulin is proportional to the reciprocal of the dielectric constant.

At higher values of $\Gamma/2$, the $\log S/S'$ curve (Fig. 14-5) is found to reach a maximum and then drop. Again, this salting-out effect is common not only to all proteins but also to amino acids and most other organic compounds. The effect can be considered most simply as a reduction in the activity of the water owing to hydration of the added salt ions. With less solvent available, then, less solute will be dissolved. The classical protein fractionation techniques depend largely on this salting-out phenomenon. Globulins salt-out much more readily than albumins, and can be separated by properly adjusting the ionic strength. Since this effect is again roughly proportional to the ionic strength, divalent ions are much more effective than monovalent. Ammonium sulfate has been a favored salting-out reagent because of its high solubility, although other sulfates have frequently been used.

Effect of Organic Solvents. Addition of organic solvents to aqueous solutions of typical proteins ordinarily leads to a reduction in solubility, and usually to precipitation. This effect can be understood, at least qualitatively, from the previous consideration of the properties

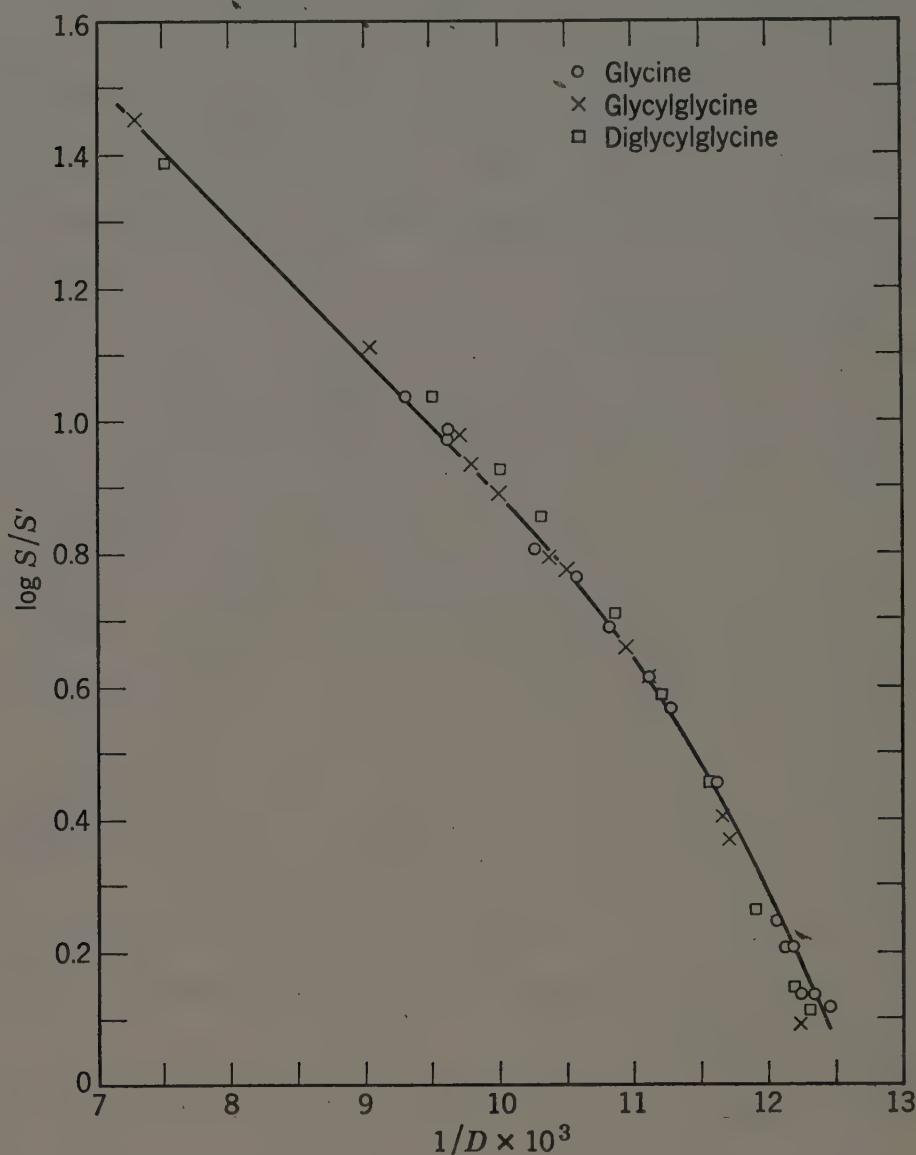


Fig. 14-6. Solubility of β -lactoglobulin in aqueous solutions of glycine, glycylglycine, and diglycylglycine at various concentrations. The abscissa is the reciprocal of the dielectric constant of the solvent. From data of A. Gronwall, Compt. rend. trav. lab. Carlsberg, Sér. chim., 24, No. 9 (1942).

of coulombic forces. Reduction in the dielectric constant leads to strongly increased ionic attraction, both protein-protein (if near the isoelectric point) and protein-ion. This favors insolubility.

As has been noted, decrease in dielectric constant, as by the addition of ethanol to an aqueous solution, should and does lead to a decrease in the solubility of ionic substances in general and proteins in particular. On the other hand, inspection of the Debye-Hückel equation,

equation 5, shows that the effect of ionic strength on the activity coefficient (salting-in effect) should be increased, since D appears in the denominator. In other words, a protein should be more "salt-sensitive" in such systems. A magnification of differences in the salting-in behavior of proteins would be expected. Furthermore, since the over-all solubility of the protein is reduced, it might be expected that even albumins would show a pronounced salting-in effect similar to globulins. This is one of the important advantages of the low-temperature, low dielectric constant method of protein fractionation introduced by Cohn and co-workers (Chap. 15).

BINDING OF IONS BY PROTEINS

In the treatment of the titration behavior of amino acids and proteins in Chaps. 3 and 11, it was tacitly assumed that the various equilibria involved only H^+ ions and the various dissociable groups of the amino acid; in other words, that no ions other than protons were bound. Such a treatment appears to be very satisfactory for ordinary acids and bases of low molecular weight, and even for amino acids.* However, there is growing, and now abundant, evidence that such simple considerations are not entirely adequate for proteins, for which the binding of other ions must also be taken into account.

At first this might seem like a backward step, since well up into the present century the proteins, in common with other colloids, were thought to derive their charge from adsorption of ions on their surface by "colloidal forces." The proposal by Loeb and others that the charge could be explained on the basis of the acidic and basic groups of the protein and the fundamental laws of homogeneous equilibrium, represented a considerable change in point of view. It seems clear that Loeb's point of view is correct to a reasonable degree of approximation in most cases. This section treats of the deviations from this "ideal behavior." We are now in the advantageous position of being able to regard the binding of ions other than protons from a more exact chemical point of view than was possible prior to Loeb's work.

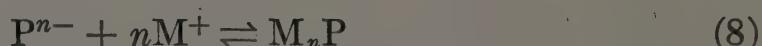
Evidence for Ion Binding. Perhaps the earliest indication that proteins bind ions was the observation that protein fibers absorb dyes. Most dyes consist essentially of complex organic cations or anions. The precipitation of proteins from aqueous solution by certain complex acids, such as trichloroacetic acid and phosphotungstic acid, can be regarded as the formation of an un-ionized salt, the protein acting

* There is, nevertheless, evidence that polyvalent cations are bound to a significant extent by small anions.

as the cation and the acid providing the anion according to the reaction

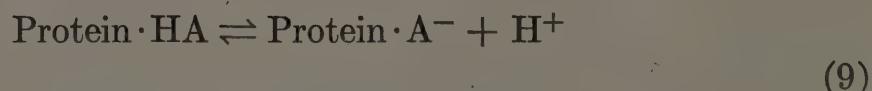


Other examples of anions which form protein precipitates (and in some cases precipitates with certain amino acids) are picric acid, flavianic acid, picrolonic acid, and the anion $[(NH_3)_2Cr(CNS)_4]^-$ or Reinecke salt. Many proteins also form insoluble precipitates with certain heavy metal cations, such as Ag^+ , Zn^{++} , Hg^{++} , Cu^{++} , Pb^{++} , and Au^{+++} . In such cases the precipitation may be regarded as the formation of an insoluble un-ionized salt in which protein serves as the anion (metal proteinate):



The metal ions forming such precipitates are frequently polyvalent and are known to form many insoluble salts.

These examples are obvious cases of ion binding by proteins. More subtle cases have come into prominence more recently. In 1941–1942, Steinhardt and co-workers published the results of a series of investigations of the behavior of wool fibers and of ovalbumin solutions when titrated with various strong acids. On the basis of the simple theory concerned only with the binding of protons, the titration curves, which are plots of H^+ bound to protein versus pH , should be independent of the acid used. Such was not the case. The curves had the same general shape, but were shifted along the pH axis (Fig. 14–7). Since the protein groups being titrated in this pH range are essentially the carboxyl groups, the obvious conclusion is that the dissociation constant of the carboxyl groups depends on the nature of the acid added, or on the nature of the other anions in the solution. Steinhardt *et al.* presented an explanation of the results on the basis that anions are also bound by the protein so that it is necessary to consider the dissociation constants of new groups, the protein-anion complex groups:



Assuming the curve with HCl to be the normal one, the affinity of the various anions can be ordered according to the magnitude of the shift in titration curve. The results show that anionic dyes have strong affinities, as is to be expected. Typical synthetic anionic detergents such as dodecylbenzenesulfonic acid also have extremely

high affinities for protein. An analogous behavior was also observed with water-soluble proteins, although in this case the results were more limited because of precipitation of the complexes at low pH .

Perhaps a somewhat more rational explanation of the effect of bound anions on the titration curve can be given in terms of the electrostatic interaction term $-2Zw/2.303$, given in equation 6 of Chap. 11. In the low pH titration region the protein possesses a net

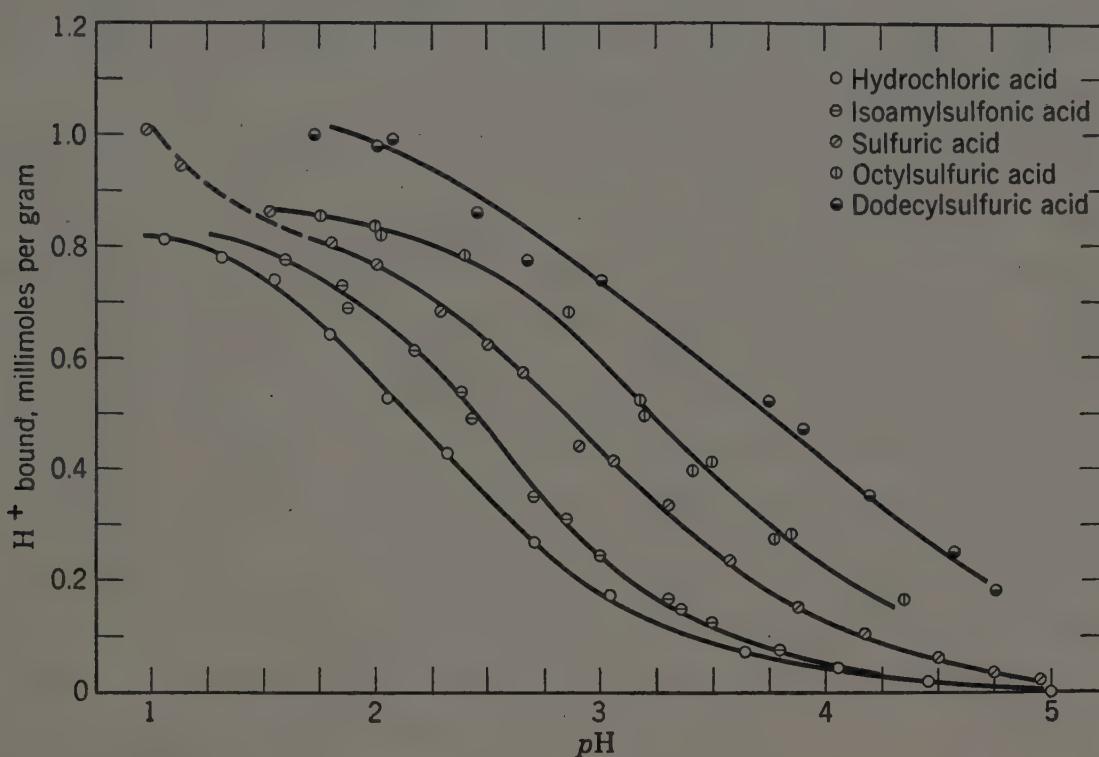


Fig. 14-7. Titration of wool protein with various strong acids, showing shift in equilibrium constants due to anion binding. From the data of J. Steinhardt, C. H. Fugitt, and M. Harris, *J. Research Natl. Bur. Standards*, **28**, 201 (1942).

positive charge which is diminished by anion binding. Thus, the greater the degree of binding of anions, the smaller Z and the less important the electrostatic term. The net effect will be to enhance hydrogen ion binding at a given pH , as is observed.

The question arises as to whether it can be assumed that the titration curve with HCl is indeed the ideal one, or whether Cl⁻ ions are also bound to an appreciable degree. It now appears that even the binding of Cl⁻ cannot be ignored in protein solutions.

Other Manifestations of Ion Binding. Another very interesting manifestation of the affinity of proteins for anions is the ability of certain anions to stabilize protein solutions toward denaturation (Chap. 17). Whereas serum albumin, in common with many proteins, is rendered completely insoluble by heating in water even briefly to about 60° C., addition of relatively small amounts of certain anions,

particularly those of the fatty acids containing four to eight carbons, renders such solutions stable at this temperature for periods up to many days or even months. Such effects certainly indicate adsorption of anions to the protein, particularly in view of the striking effectiveness of the ions at even low concentration. This view has been confirmed by electrophoretic studies, which show the mobility of the protein to be increased, presumably due to increase in negative charge, in the presence of such anions. This stabilization effect has also been shown to hold in the case of denaturation of serum albumin by urea and by guanidine hydrochloride.

Another useful method of studying anion binding is by examining the shift in absorption spectrum of anionic dyes in the presence of protein. The binding of noncolored anions can also be studied by examining the effect of such anions on the protein-dye absorption curves, since there is a competition for the available binding sites in the protein. Such measurements have confirmed the fact that proteins are generally capable of binding complex organic anions. On the other hand, certain proteins, notably serum γ -globulin and gelatin, have been found almost completely lacking in this dye-binding ability, for reasons not yet clear.

Protein-Detergent Interaction. An extreme case of the binding of ions by proteins is the binding of the highly asymmetric surface-active ions of soaps and the synthetic detergents. The binding of soaps to protein fibers was long ago observed to be an important aspect of laundering. It is now clear that proteins in solution have a strong affinity for soaplike ions, and that the resultant interaction can be exemplified by many diversified manifestations.

In the first place, proteins can be precipitated by detergents under appropriate conditions, but they usually redissolve in the presence of an excess of the reagent. Water-insoluble proteins such as zein may be dissolved by detergents, and the precipitation of soluble proteins by reagents such as trichloroacetic acid can be prevented. Detergents can lead to denaturation of proteins in remarkably low concentration (Chap. 17); indeed, from the standpoint of the low concentration required, they constitute the most potent denaturing agents. Conversely, there are claims of instances in which detergents, or at least closely related compounds, stabilize proteins against denaturation. Some proteins are disaggregated or dissociated in the presence of detergents, for example, insulin and tobacco mosaic virus. It is apparent, then, that detergent ions can lead variously to precipitation, solubilization, denaturation, disaggregation, and possibly to stabilization.

The precipitation of proteins by anionic detergents is confined to the acid side of the isoelectric point of the protein. This is entirely reasonable, for under such conditions the protein carries a net positive charge, and binding of detergent anions leads to a reduction in net charge. Binding of further detergent under such conditions can lead to an inversion of charge and eventually to re-solution. The precipitation with cationic detergents, such as alkylammonium types, is limited to the alkaline side of the isoelectric point, as would be expected. The amount of detergent required to give quantitative precipitation has been found invariably to be proportional to the amount of protein present, confirming the concept of combination of protein and detergent.

The most elegant demonstration of the combination between protein and detergents was given by Lundgren *et al.* (1943) through electrophoretic investigations. The detergent used, alkylbenzenesulfonate, migrates electrophoretically as a single component of relatively high mobility, approximately 19×10^{-5} cm.² sec.⁻¹ volt⁻¹. Detergents are known to exist in aqueous solution in the form of large aggregates or "micelles." This mobility then corresponds essentially to that of the micelles. Upon adding ovalbumin so that the ratio of detergent to protein, D : P, is less than 0.3, the detergent peak disappears and two slower peaks are observed. The slower of these corresponds in mobility to native ovalbumin; the faster, with mobility between that of ovalbumin and detergent, is presumed to be a complex. As the D:P ratio is increased, the proportion of the faster or complex peak increases but the mobilities of the two remain nearly constant, indicating constant composition. The interaction under such conditions is hence an "all-or-none" phenomenon. If a limited amount of detergent is present, it concentrates on part of the protein molecules to give a definite stoichiometric complex rather than distributing equally among all available protein molecules. Interestingly, if the protein is previously denatured, for example by heat, this all-or-none binding is not observed. This is unquestionably one of the most significant features of the interaction, and mention of this fact will be made later in connection with the discussion of protein structure and denaturation.

When the D:P ratio exceeds that value corresponding to complete formation of the stoichiometric complex, a second binding range occurs, in which a single electrophoretic peak exists. The mobility of this component increases with increasing D:P ratio, owing to the increase in net negative charge resulting from binding of the anions, until a limiting ratio is reached. After this, a peak corresponding to free detergent micelles appears for the first time. In this second binding

range the detergent distributes itself uniformly over all the available protein molecules. Lundgren has termed the detergent bound by this latter mechanism "extra bound detergent." The composition of the stoichiometric complex corresponds, for a number of proteins, to the binding of one detergent anion per cationic group. The saturation limit for total detergent, on the other hand, appears to be nearly the same, approximately 3.5 gm. detergent per gram of protein, for all proteins studied, ranging from ovalbumin with many cationic groups to zein, which has almost none. It has been suggested that this saturation limit is fixed by the surface area of the protein available for formation of a monomolecular detergent layer.

The "extra bound detergent" can be readily extracted by means of 60 per cent acetone, but the "stoichiometric detergent" is not extracted by such a solvent unless an electrolyte is present. This indicates that ionic forces are important in the tenacious binding in the region of stoichiometric binding. Van der Waals forces must also be of great importance, however.

Two stoichiometric detergent complexes have been observed with serum albumin. One corresponds closely to the binding of one anion per two cationic groups of the protein, the other to one-to-one binding. It has been suggested that approximately half of the cationic groups are more strongly basic than the remainder. It seems more likely, however, that the difference is in the availability of the cationic groups; perhaps roughly half are folded into the interior of the protein molecule in such a way that they are accessible to protons but not so readily accessible to the large anions.

Manifestations of Protein-Ion Interaction in the Solubility Behavior of Proteins. The nature of the solubility-pH curves for proteins is not independent of ionic strength. Hardy, as early as 1905, showed that the pH for minimum solubility depends on the concentration of salt, increase in ionic strength shifting the minimum in the acid direction. This would imply a shift in the isoelectric point toward the acid side, which has been verified by comparative electrophoretic studies in media of varying ionic strength (Fig. 14-8).

Sørensen long ago recognized that binding of ions other than H⁺ would affect the isoelectric point, and distinguished between this quantity and the isoionic point. In absence of salt (zero ionic strength) only hydrogen (and possibly hydroxyl) ions can be bound so that the isoionic and isoelectric points will coincide. The decrease in the isoelectric point with increasing salt concentration results from the fact that anions are more strongly bound than cations. This binding tends to increase the negative charge at a given pH, necessi-

tating a lower pH for total charge neutrality. On the other hand, at constant hydrogen ion binding (isoionic state) the protein takes on a progressively increasing negative charge with increasing salt concentration. This negative charge enhances hydrogen ion binding by virtue

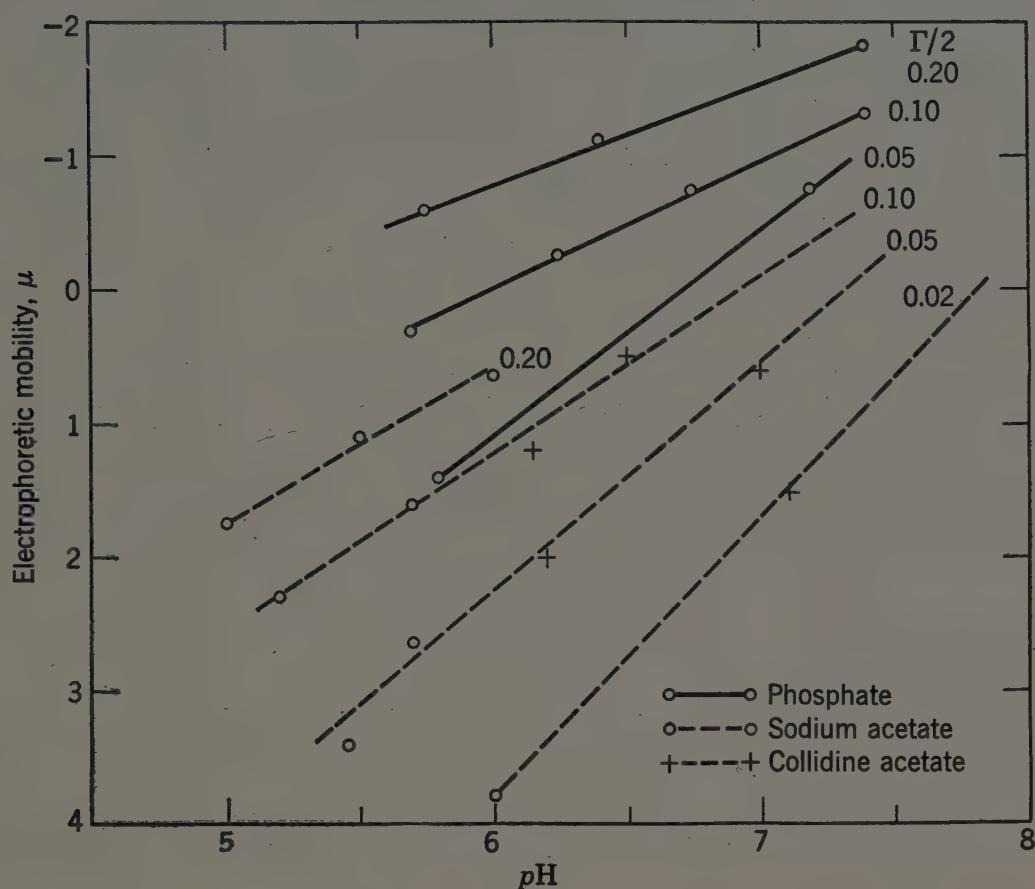


Fig. 14-8. The electrophoretic mobility of aldolase as a function of pH and ionic strength in phosphate, sodium acetate, and collidine acetate buffers. Note the shift in isoelectric point toward lower pH as the ionic strength is raised. From S. F. Velick, *J. Phys. & Colloid Chem.*, 53, 135 (1949).

of the electrostatic factor (equation 6 of Chap. 11) so that a higher pH is required to maintain the isoionic condition. To summarize, with increasing ionic strength the isoelectric point and isoionic point become more divergent, the one moving downward, the other upward, both effects being due to the excess binding of anions over cations.

SOLUBILITY AS A CRITERION OF PURITY

One of the most fundamental rules in physical chemistry is the Gibbs phase rule, which relates the number of degrees of freedom (number of variables) in a system to the number of components and the number of phases. If the temperature and pressure are fixed, this rule takes the form

$$F = C - P \quad (10)$$

where F is the number of degrees of freedom, C the minimum number of components needed to specify completely the composition of the system, and P the number of phases. The number of phases in an ordinary solubility study is 1 (the solution) if the solution is not saturated, and 2 if a single crystalline phase exists at saturation.

If one fixes the pH and the concentration and composition of salt in a system, it can be shown that the minimum number of components C is 2; these might be considered as water and protein. The number of degrees of freedom F is then $2 - 1$ or 1 as long as a

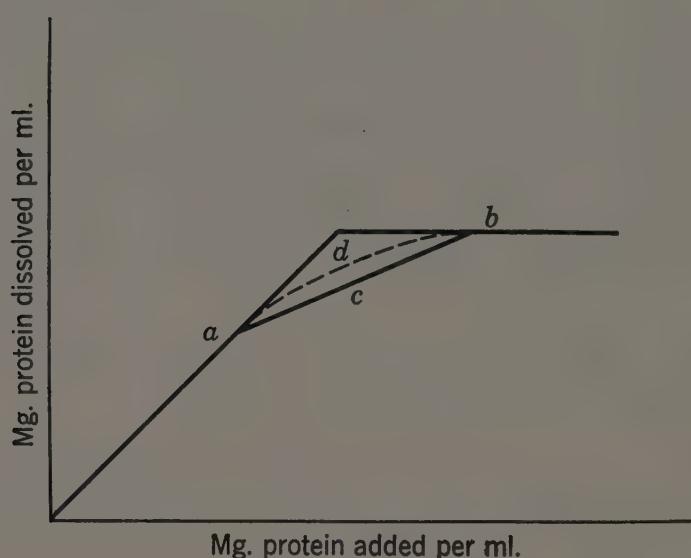


Fig. 14-9. Schematic illustration of the application of the phase rule to determination of the purity of a protein.

single (solution) phase is present. This is exemplified by a linear increase in protein concentration with addition of protein (a of Fig. 14-9). When saturation is reached, a second phase appears and F is then $2 - 2$ or 0. The protein concentration in solution is no longer variable but is constant, as shown in limb b of Fig. 14-9. The criterion of purity, then, is the manifestation of a sharp break in the curve. If two proteins are present, two breaks should be obtained, as shown by limb c . In many protein systems no sharp break is obtained, but only a curved region as shown by d .

This criterion of purity is so sensitive that few proteins have met it. It is possible that some homogeneous proteins fail to meet the theoretical criterion because of reversible association-dissociation phenomena. It is also possible for an inhomogeneous protein preparation to behave as a homogeneous one if the component proteins are present in the solid phase in the exact proportions of their ratios of solubility. For this reason solubility measurements should be carried out under different conditions of salt concentration and pH .

PROTEIN-PROTEIN INTERACTION

Many of the considerations involved in the interaction of proteins with molecules of low molecular weight and ions must hold also in the interaction of proteins with proteins. Such interaction must be especially complex and important, for example, in complex systems like blood plasma and egg white. In both of these cases certain anomalies in the electrophoretic patterns indicate interaction.

It is well known that in the special case of two proteins carrying opposite net charge, strong coulombic forces lead to aggregation and possibly to the formation of an insoluble precipitate. A classic example is the formation of the slightly soluble protamine-insulin complex, which is used therapeutically to provide a gradual release of insulin. The protamine, having a very high isoelectric point, carries a strong net positive charge at intermediate values of *pH*, where the insulin is negatively charged. Interactions between nucleic acids and proteins have been demonstrated electrophoretically at *pH* values acid to the isoelectric point of the proteins.

The possibility of protein-protein interactions is of prime significance in protein fractionation studies. It would appear undesirable, as a rule, to attempt separation at a *pH* intermediate between the isoelectric points of any of the protein species; however, in some cases one may actually take advantage of such interactions to attain greater selectivity in the fractionation of complex systems. Because coulombic interactions are diminished by increased ionic concentration, the presence of salt in fractionation experiments is usually to be desired. The problem of protein fractionation will be considered further in the next chapter.

In addition to this relatively simple coulombic interaction there is also the possibility of much more specific interaction, as in protein crystallization and in antibody-antigen and enzyme-substrate combination. Such interactions are presumably largely of van der Waals origin, but perhaps also involve hydrogen bonds and dipolar and coulombic interactions between favorably situated groups.

It has recently been suggested that protein-protein interactions may also arise as a result of fluctuations in surface charge density, owing to migration of protons between basic sites on the protein surface. In a sense this effect may be considered analogous to the London dispersion forces but involving fluctuations of protons rather than electrons. It is too early to assess the importance of such forces, though the possibility arises that they might be highly specific and might extend over relatively long range in solution. They would

probably be important in protein solutions only at very low ionic strength.

GELATION OF PROTEINS

An extreme case of the delicate balance between protein-solvent and protein-protein interactions is the formation of protein gels. The gel state may be regarded as a mesomorphic state, somewhere between the typical solid (crystalline) and liquid states. Technically a gel is distinguished from a liquid or an ordinary solution by its ability to resist a shearing stress, its *rigidity*. This property is distinct from viscosity. One may have a very viscous liquid, such as glycerol, which shows little or no rigidity to shear.

The classical picture of gel structure is that of a matrix of fibrillar solid phase running through the entire mass of the gel, with the solvent immobilized in the interstices. This picture still has much to recommend it.

Most gel-forming materials are fibrillar in nature. The best known gel-forming protein is gelatin, which is known to be a highly elongated molecule. It is obvious that for such a matrix to result at low solute concentration, the solute must be extremely asymmetric. Gelatin forms a gel in concentrations as low as 1 per cent or less. As is well known, the gelation of gelatin is entirely temperature reversible, the gel "melting" fairly sharply at a temperature which depends on such factors as gelatin concentration, *pH*, and salt concentration.

Fibrin (Chap. 18), in dilute solutions yields excellent gels (clots), which may be clear or opaque, depending on conditions. Gel formation may ensue in solutions containing only 0.004 per cent fibrinogen! Because fibrinogen is an elongated molecule, approximately 600 Å. long, network formation might result from the aggregation of such units end-to-end. The formation of the fibrin clot is discussed further in Chap. 18. Figure 18-3 shows in a striking manner the fibrillar matrix.

Many denatured proteins form gels. Such gelations require rather precise conditions of *pH*, and so forth. Apparently the proper balance between protein-protein and protein-solvent attraction is important. Gels of this kind are probably much more haphazard in nature than gelatin gels and especially fibrin clots. Denaturation probably involves the unfolding of the native protein to a more elongated structure (see Chap. 17). If so, the formation of a network matrix would be quite possible and gelation is not surprising. The forces of interaction are probably hydrogen bonds, coulombic attractions, and van der Waals forces.

A few puzzling instances are known in which nearly spherical proteins can form gels, and can be recovered without apparently having undergone denaturation. A case in point is β -lactoglobulin, which under appropriate conditions in 50 per cent ammonium sulfate forms a gel at about 5 per cent concentration. This process can be reversed readily by dialysing out the salt, and the protein can be crystallized. It has been calculated that network formation could result at this concentration, assuming an axial ratio of 5:1, if

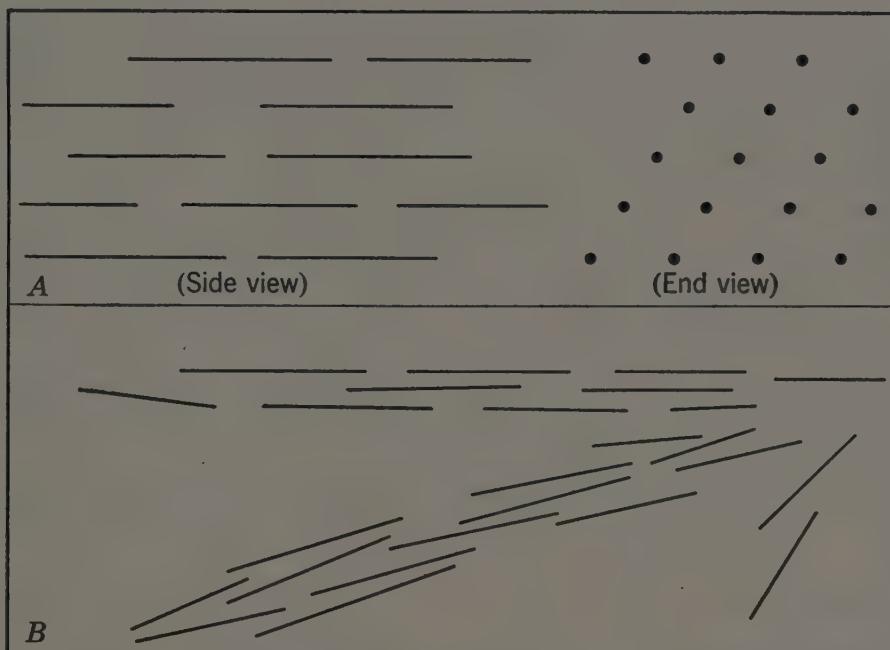


Fig. 14-10. Schematic depiction of the hexagonal array of tobacco mosaic virus particles: A, in gels in concentrated solutions; B, in tactoids. The spacings between rodlets (laterally) may be as great as several hundred angstroms.

association is entirely end-to-end. Another very interesting case is insulin, which forms gels at 2 per cent concentration when heated in acid solution. This is a consequence of the formation of long fibrils, apparently as a result of end-to-end aggregation and not by unfolding.

The possibility of gel formation in the absence of a network matrix still exists, is indeed probable in some systems. Especially interesting in this regard is tobacco mosaic virus. TMV consists essentially of rod-shaped "molecules" or particles about 3000 Å. long and 150 Å. in diameter. In the concentration range of 13 per cent and up, gels can be obtained. Study of such gels by X-ray diffraction has demonstrated that the TMV particles are not in contact but are arranged in a hexagonal array with their axes parallel, as shown in A of Fig. 14-10. The distance between the rods, depending on concentration, may be as great as 400 Å. The forces which are operative over such great distances must be coulombic in character. Presumably

this hexagonal arrangement represents the tendency of the system to minimize the coulombic repulsion between TMV particles. Under other conditions "tactoids" may result, which can lead to gelation by matrix formation, as in *B* of Fig. 14-10. As would be expected, such gels can form at much lower TMV concentrations.

SELECTED REFERENCES

Reviews

"The Structure of Water, Hydrates and Aqueous Solutions," Chap. 19 in *Electronic Structure and Chemical Binding*, O. K. Rice, McGraw-Hill, New York, 1940.

"Solubility of Amino Acids, Peptides and Related Substances in Water and Organic Solvents," J. T. Edsall and G. Scatchard in Cohn and Edsall's *Proteins, Amino Acids and Peptides*, Reinhold, New York, 1943.

"The Solubility of Proteins," E. J. Cohn in Cohn and Edsall's *Proteins, Amino Acids and Peptides*, Reinhold, New York, 1943.

"Surface Active Agents in Biology and Medicine," E. I. Valko, *Ann. N. Y. Acad. Sci.*, **46**, 451 (1946).

"Protein Gels," J. D. Ferry, *Advances in Protein Chem.*, **4**, 2 (1948).

"The Interactions of Proteins and Synthetic Detergents," F. W. Putnam, *Advances in Protein Chem.*, **4**, 80 (1948).

"Protein-protein Interactions," D. Waugh, *Advances in Protein Chem.*, **9**, 326 (1954).

Original Articles

"Theory of Solutions of Molecules Containing Widely Separated Charges with Special Applications to Zwitterions," J. G. Kirkwood, *J. Chem. Phys.*, **2**, 351 (1934).

"Fractionation of Pepsin. I. Isolation of Crystalline Pepsin of Constant Activity and Solubility from Pepsinogen or Commercial Pepsin Preparations," R. M. Herriott, V. Desreux, and J. H. Northrop, *J. Gen. Physiol.*, **24**, 213 (1940).

"Further Investigations of the Affinities of Anions of Strong Acids for Wool Protein," J. Steinhardt, C. H. Fugitt, and M. Harris, *J. Research Natl. Bur. Standards*, **28**, 201 (1942).

"Electrophoretic Study of the Action of Alkylbenzenesulfonate Detergents on Egg Albumin," H. P. Lundgren, D. W. Elam, and R. A. O'Connell, *J. Biol. Chem.*, **149**, 183 (1943).

"The Combination of Fatty Acids and Related Compounds with Serum Albumin. I. Stabilization Against Heat Denaturation," P. D. Boyer, F. G. Lum, G. A. Ballou, J. M. Luck, and R. G. Rice, *J. Biol. Chem.*, **162**, 181 (1946).

"Interaction of Homologous Alkyl Sulfates with Bovine Serum Albumin," F. Karush and M. Sonenberg, *J. Am. Chem. Soc.*, **71**, 1369 (1949).

"The Effect of Salts on the Isoionic and Isoelectric Points of Proteins," G. Scatchard and E. S. Black, *J. Phys. & Colloid Chem.*, **53**, 88 (1949).

"An Electrophoretic and Equilibrium Analysis of Aldolase in Phosphate and Acetate Buffers," S. F. Velick, *J. Phys. & Colloid Chem.*, **53**, 135 (1949).

"Participation of the Guanido Group of Arginine in the Gelling of Gelatin," P. Grabar and J. Morel, *Bull. soc. chim. biol.*, **32**, 643 (1950).

- "Interaction of Egg Albumin with Dodecylamine Hydrochloride," S. Timasheff and F. Nord, *Arch. Biochem. and Biophys.*, **31**, 309 (1951).
- "Precipitation and Inhibition of Lysozyme by Surface-Active Agents," H. Glassman and D. Molnar, *Arch. Biochem. and Biophys.*, **32**, 170 (1951).
- "Structure and Properties of Ice," N. Bjerrum, *Science*, **115**, 385 (1952).
- "Forces Between Protein Molecules in Solution Arising from Fluctuations in Proton Charge and Configuration," J. Kirkwood and J. Shumaker, *Proc. Natl. Acad. Sci. U. S.*, **38**, 863 (1952).
- "Statistical and All-or-None Binding of Alkylbenzenesulfonate by Albumins," J. Yang and J. F. Foster, *J. Am. Chem. Soc.*, **75**, 5560 (1953).
- "The Influence of Salt on the Size and Shape of a Protein-Detergent Complex," B. Harrap and J. Schulman, *Discussions Faraday Soc.*, **13**, 197 (1953).
- "A Study of the Interaction of Dodecyl Sulfate with Bovine Serum Albumin," M. Pallansch and D. Briggs, *J. Am. Chem. Soc.*, **76**, 1396 (1954).
- "Influence of Structure on the Association between Proteins and Organic Ions," E. Fredericq, *Bull. soc. chim. Belges*, **63**, 158 (1954).
- "On the Mode of Interaction of Surface Active Cations with Ovalbumin and Bovine Plasma Albumin," J. F. Foster and J. T. Yang, *J. Am. Chem. Soc.*, **76**, 1015 (1954).
- "The Binding of Uncharged Molecules to Proteins. II. Testosterone and Bovine Serum Albumin," J. Schellman, R. Lumry, and L. Samuels, *J. Am. Chem. Soc.*, **76**, 2808 (1954).
- "Effect of Temperature on the Interaction of Human Serum Mercaptalbumin with Zinc Ions," F. Gurd, *J. Phys. Chem.*, **58**, 788 (1954).
- "Potentiometric Studies of Binding of Alkali and Alkaline Earth Ions to Bovine Plasma Albumin," A. Kepes, *Compt. rend.*, **239**, 117 (1954).
- "Binding of Salt Ions by Bovine γ -Pseudoglobulin," J. Cann and R. Phelps, *J. Am. Chem. Soc.*, **77**, 4266 (1955).
- "Slow Intramolecular Changes in Copper Complexes of Serum Albumin. The Role of Neighboring Groups in Protein Reactions," I. Klotz, J. Urquhart, T. Klotz, and J. Ayers, *J. Am. Chem. Soc.*, **77**, 1919 (1955).
- "The Binding of Organic Ions by Proteins. Volume Changes," R. Rosenberg and I. Klotz, *J. Am. Chem. Soc.*, **77**, 2590 (1955).
- "An Electrophoretic Study of Bovine Serum Albumin-Methyl Orange Complex at 25°," M. Cohly, *J. Indian Inst. Sci.*, **37A**, 89 (1955).
- "Interactions of Homologs of Carcinogenic Azo Dyes and Bovine Serum Albumin," R. Burkhard and F. Moore, *J. Am. Chem. Soc.*, **77**, 6057 (1955).
- "Interactions of Neutral Amino Acids with Human Serum Albumin and γ -Globulin," M. Hunter and S. Commerford, *J. Am. Chem. Soc.*, **77**, 4857 (1955).
- "Theory of Protein Solutions. II," T. Hill, *J. Chem. Phys.*, **23**, 2270 (1955).
- "Association Energy of Bovine Plasma Albumin and Organic Ions," E. Fredericq, *Bull. soc. chim. Belges*, **64**, 639 (1955).
- "The Interaction of *n*-Octylbenzene-*p*-Sulfonate with β -Lactoglobulin," R. Hill and D. Briggs, *J. Am. Chem. Soc.*, **78**, 1590 (1956).
- "Solubility Transformation of α -Lactoglobulin," C. A. Zittle, *Arch. Biochem. and Biophys.*, **64**, 144 (1956).
- "The Binding of Sodium and Potassium Ions in Solutions of Various Proteins," C. W. Carr, *Arch. Biochem. and Biophys.*, **62**, 476 (1956).
- "The Binding of Some Small Anions to Serum Albumin," G. Scatchard, J. S. Coleman, and A. L. Shen, *J. Am. Chem. Soc.*, **79**, 12 (1957).

Preparation and Purification of Proteins

Criteria of purity

Quantitative determination of proteins

Separation from nonprotein material

Protein fractionation

Proteins exist which differ rather widely in amino acid composition and molecular weight. The theoretical number of possibilities is virtually without limit. Even if one considers only the possible number of isomeric structures with respect to amino acid sequence in a protein of definite molecular weight and composition, the number of possibilities is staggering. Fortunately, Nature has evidently elected to synthesize only a relatively small fraction of the total conceivable.

Protein fractions can be prepared which superficially appear homogeneous. Such separations, involving as they do enormously complex molecules which differ very little, relatively speaking, are accomplished in most cases only with great difficulty. Even so, the end products so obtained may prove to be only virtually pure. No one can yet be sure that the best available protein preparations are not actually mixtures of several closely related species, or perhaps even continuous spectra of molecular structures differing, for example, in the detailed folding of the peptide chain or in hydrogen bonding.* The first problem that will be considered, then, is the question of criteria available for judging the purity of protein preparations.

* The "microheterogeneity" of proteins has been discussed in an interesting review by Colvin *et al.* (1954).

Superposed on this problem of the separation of closely related chemical compounds is the fact that proteins are extremely labile bodies, subject to various and not yet clearly understood alterations lumped under the term *denaturation* (literally, change from the natural state), discussed in Chap. 17. This makes it necessary to handle proteins with "kid gloves," and makes unfeasible certain methods of operation commonly used in chemistry.

CRITERIA OF PURITY

Before embarking on a protein purification program, it is important to have some yardstick whereby the purity of a given preparation may be ascertained. A definition of the term *purity* imposes some interesting philosophical questions. In the first place, it is obvious that a definition of purity will depend on the use which is to be made of the material in question. The expression "pure water" will have a very different connotation to the public health official, the chemist, and the physicist. In addition, the implications of the term pure must be dependent on the criteria available, the methods of "measuring" purity. Even if the public health official, the chemist, and the physicist were to agree that pure water consisted of 100 per cent H_2O , they would not agree on the purity of a given sample because they would use different yardsticks for its measurement. Furthermore, as science advances, that which is considered the ultimate in purity today may be recognized as impure in the future as new methods of evaluation are developed. These considerations become of particular importance in the protein field.

Constancy of Chemical Composition. Chemical analysis can be used to a limited extent as a criterion for the purity of proteins. If a protein is subjected to successive purification steps and analyzed for certain amino acids or other characteristic groupings (prosthetic groups), it could be expected that the composition, with regard to the assayed group, should become constant at such time as all impurities are removed. As an example, some purified samples of zein have been found to contain no tryptophan. Relatively crude preparations contain an appreciable amount of this amino acid, presumably owing to contamination by other proteins. Analysis for tryptophan might therefore be used as one criterion of purity in this case. Obviously, though, other impurities not containing tryptophan might also be present and would not be detected at all by this method. In the purification of certain enzymes the constancy of composition with respect to the aromatic amino acids has been employed. This method is of very limited applicability in the protein field because the con-

taminating proteins usually differ only very slightly in chemical composition from the protein in question.

Electrophoresis. Perhaps the most widely used criterion of purity in the protein field is electrophoretic analysis (Chap. 12). For controlling the preliminary fractionation of a complex protein mixture such as blood plasma, this analytical tool is invaluable. Blood plasma consists of about a half-dozen principal electrophoretic components (Fig. 12-6), and these have been separated into substantially pure fractions as adjudged by this technique. It is important, however, to emphasize that electrophoretic homogeneity should never be accepted as final evidence for purity. Theoretically, two proteins may have exactly the same amino acid composition, size, and shape (and hence electrophoretic mobility) and still differ biologically and chemically because of differences in the arrangement of the amino acid residues. Electrophoresis would not be expected to distinguish between such isomers. Two proteins differing both in composition and in size and shape may yet have the same electrophoretic mobility, owing to a fortuitous compensation of these properties. It is much less likely that in such a case the mobility of the two would be the same over a wide pH range. Electrophoretic homogeneity over a range of conditions is therefore more meaningful than homogeneity under a single set of conditions.

An interesting example in this connection is serum γ -globulin which can be separated in a form that is homogeneous in the sense that no split of the electrophoretic boundary occurs. Still, it is known that this fraction contains antibodies to many different diseases. Careful examination of the electrophoretic pattern shows that while the boundary does not split into "components," it is much broader than would be expected for a homogeneous protein. Furthermore, application of the reversible boundary spreading criterion shows this fraction to be very heterogeneous. From this it must be inferred that there exists in the γ -globulin fraction a range of protein species differing only slightly in mobility. This result points up the power of the reversible boundary spreading technique as a criterion of purity. Apparently, no protein preparation has yet met this test fully.

A possible source of error in electrophoresis is the interaction which may occur in protein systems, particularly between proteins carrying opposite charge. Such a protein complex might migrate as a single component.

Ultracentrifugation. The sharpness of the sedimenting boundary obtained in the velocity ultracentrifuge is frequently used as a criterion of purity. Since sedimentation behavior depends, essentially, only on

size and shape, not on chemical composition, it can be considered an even less sensitive criterion of purity than electrophoresis.*

Combined study by the two methods is of course much better than the use of either alone. Nevertheless, it is possible for two biologically different proteins to have the same sedimentation constants and electrophoretic mobilities.

Crystallinity. Formerly it was common to suppose that a crystalline protein is a pure one. In the protein field, crystallization is difficult and usually involves so much preliminary fractionation and purification that crystallization is viewed, justifiably, with triumph. Evidence is now abundant, however that beautifully crystalline protein preparations may be impure as adjudged by other criteria. For example, β -lactoglobulin, long considered one of the best examples of a pure protein, is now known to be inhomogeneous electrophoretically. It even appears possible to vary the proportion of the electrophoretic components in the crystals, so that we are here dealing with isomorphous proteins (proteins which crystallize similarly so that they can replace one another freely in the crystalline lattice). Crystalline ovalbumin is another, though less striking, example. In another case a crystalline protein preparation from muscle has been found to be homogeneous electrophoretically although it contained at least two enzymatic components.

Solubility Studies. Perhaps the best physical method of testing for purity of protein preparations is a study of the solubility properties (Chap. 14). There are three general ways in which solubility behavior can be used for this purpose.

(1) The protein preparation can be fractionated or purified to a point where its solubility under a standard set of conditions is no longer changing. This criterion of *constancy of solubility* is somewhat analogous to constancy of chemical composition, previously mentioned.

(2) The phase rule theory can be applied. The solubility of the protein preparation is measured as a function of the amount of solid added, under carefully controlled conditions. The criterion of purity is a sharp break from a slope of unity to zero in the graph of protein in solution versus protein added. It is important to obtain a number of experimental points in the vicinity of the change of slope in order to show that this is really a sharp break. Few proteins have passed this very rigorous test when it has been carried out carefully under

* A system may be very heterogeneous with respect to molecular weight and still yield a single sharp boundary if the different species have a similar cross-sectional area, that is, if they differ in length only. Gelatin offers an interesting example of this behavior.

several sets of conditions, but it is doubtful that any protein which has successfully fulfilled the requirements of this method has ever been shown to be impure by other criteria.

(3) The third method involves an examination of the solubility of a protein preparation as a function of solvent composition, for example, salt concentration. An abrupt change in the slope of the solubility versus salt concentration curve indicates the precipitation of a component. A pure protein under such conditions should exhibit a smooth curve, linear if plotted as the logarithm of the solubility. This test can be amplified by following also the specific biological activity of the dissolved protein, if such exists.

Biological Criteria. Sensitive biological properties may be utilized to follow the extent and degree of purification of those proteins which possess such properties. Enzymes obviously do, and this criterion has been extensively employed in the recrystallization of enzymes. The enzymatic activity approaches a maximum value beyond which it does not rise with further recrystallization or further purification. This stage is then considered to represent full purification. For crystalline carboxypeptidase, two to five crystallizations typically suffice for the attainment of maximal activity.

Serological tests (precipitin test) have served very well to establish the absence of minute amounts of suspected impurities. Commercially produced serum albumin is routinely checked for absence of serum globulin by testing with antiserum to the latter.

QUANTITATIVE DETERMINATION OF PROTEINS

Methods for the quantitative determination of proteins* are of importance in separation and fractionation studies, as well as in protein studies in general. Classically, the method most often employed has been nitrogen analysis. Because most of the simple proteins contain approximately 16 per cent nitrogen, it has been fairly common to estimate the total amount of protein present in a sample by multi-

* Several qualitative color tests for the presence of proteins have been found useful. *Millon's reagent*, obtained by heating metallic mercury with nitric acid, yields a red coloration upon boiling with protein. The test depends upon the presence of tyrosine and is particularly sensitive. The *xanthoproteic reaction* results upon boiling protein with concentrated nitric acid, a yellow color appearing. This reaction involves the aromatic amino acids, and has been observed by any student of chemistry who has inadvertently spilled nitric acid on his fingers. The *glyoxylic acid reaction* results in a reddish-violet coloration and depends on the presence of tryptophan. The *biuret reaction* involves the formation of a pink-violet color when the protein is treated with cupric ion under alkaline conditions. This test is given also by any peptide containing two or more peptide linkages. The *ninhydrin reaction* is discussed on p. 52.

plying the nitrogen content by the factor 6.25 (100/16). It should be emphasized, however, that this method is not absolute, since the nitrogen content of even the simple proteins varies appreciably, depending on the amino acid composition, and the conjugated proteins may contain much less nitrogen (below 10 per cent in some cases). Nitrogen determination is usually made by the Kjeldahl method, which often yields low values unless great care is given to the problem of obtaining complete digestion. It should also be noted that nitrogen analysis is worthless unless all impurities containing nonprotein nitrogen have been removed.

For determination of protein in solution, ultraviolet absorption is convenient and rapid. Most proteins have an intense absorption in the wave-length region around $280m\mu$. This absorption is primarily, if not exclusively, due to the aromatic amino acids present. Again, the method is not absolute but depends on the composition of the protein.

Precipitation with trichloroacetic acid is often used, with the assumption that only protein is precipitated. The method is probably valid if all the contaminants are impurities of low molecular weight, such as amino acids, small peptides, and sugars, but it should be employed judiciously.

Fundamentally, then, for the quantitative determination of a given protein there is no substitute for isolation in "pure" form. Once a preparation has passed the various tests for purity, it is desirable to determine its nitrogen content and ultraviolet absorption spectrum, which properties then may be used conveniently for concentration determination on solutions of the "pure" protein.

SEPARATION FROM NONPROTEIN MATERIAL

Protein Extraction. Many important proteins occur in solution in various biological fluids, such as blood plasma, milk, egg white, and urine. Their preparation is simplified in that extraction is not necessary. Those proteins occurring as a part of a tissue, as well as intracellular proteins (even though in solution in the cell cytoplasm), must first be extracted from their natural environment.

For cytoplasmic proteins the task is essentially that of disrupting the cell structure. This may be accomplished in various ways, especially by mechanical rupture using various types of homogenizers, blenders, or grinding mills. High-speed agitation in the presence of small glass beads is particularly effective. Ultrasonic vibrations are sometimes employed. Simply suspending the cells in a medium of low ionic strength (hypotonic solution) may suffice to cause rupture by osmotic pressure.

Proteins occurring in structural tissues are generally most difficult to extract. Their insolubility may be due to an extremely high molecular weight of the protein or to strong interaction, possibly even covalent bonding with nonprotein material. As an example, the proteins of hair and wool (keratins) are insoluble because of polymerization through —S—S— linkages. Solubilization results only after rupture of such bonds, for example by reduction with sulfide or sulfite. The same is probably true of the endosperm proteins of corn and other cereals.

Removal of Other Colloidal Constituents. If an extraction is carefully carried out, one does not usually extract an appreciable amount of starch, cellulose, or other such colloidal material. Biological fluids do not contain such constituents, although certain mucinous secretions contain some large carbohydrate polymers, separation of which may prove extremely difficult. To remove such impurities one usually relies on exhaustive centrifugation, preferably with a high-speed supercentrifuge, or on a careful filtration. It is important to check the possibility of removal of protein along with the nonprotein material, particularly in the case of filtration. For this purpose careful nitrogen analyses are usually applied.

Removal of Impurities of Low Molecular Weight. The principal constituents which need to be considered in this class are salts, sugars, and perhaps certain lipids. If lipids are present, they will be attached more or less firmly to protein and their separation may impose a severe problem. Such procedures as ether extraction are potentially harmful to proteins and should be used only with care, preferably at low temperature. It would seem preferable to proceed without going to such steps, and to isolate the proteins as nearly in the native state as possible, even though this may involve an attached lipid, carbohydrate, or other nonprotein moiety.

Separation from salts, low-molecular carbohydrates, and any other bodies of low molecular weight which are not bonded to the protein, can usually be accomplished best by *dialysis*. In this technique the protein-containing solution is placed on one side of a *semipermeable membrane*, a membrane which is permeable to solvent and small molecules and ions but impermeable to the large protein molecules; the solvent, usually water, is placed on the other side. The impurities are then removed by a process of diffusion through the membrane. Perhaps the simplest material for carrying out this procedure in the laboratory is commercial cellophane tubing, sausage casing. The protein solution is placed in the tubing and immersed in a large volume of solvent. Some form of agitation should be maintained to speed the

process. It is also important to change the outer solution several times to accomplish complete removal of impurities. Where dialysis against water is being used, the water may sometimes be allowed to flow slowly and continuously through the system.

The process of removing ionic impurities can sometimes be speeded by means of *electrodialysis*, dialysis carried out in an electric field to promote the migration of ions out of the protein system.

The most elegant method for removal of ions from aqueous solution is ion-exchange, using one or a combination of the numerous exchange resins now available. Cation exchangers, when loaded with hydrogen ions (that is, when they are in the acid form), replace cations in solution with hydrogen ions. Similarly, anion exchangers in the hydroxyl form replace anions with hydroxyl ions. By combination of such resins it is possible to prepare deionized "conductivity water" comparable to the best obtainable by distillation. Two problems arise, however, in the application to protein solutions. First, globulins will tend to precipitate upon deionization and adsorb to the resin bed. The technique is therefore restricted to proteins that are soluble in pure water at the isoionic point. Second, extreme pH values that may exist during the exchange process may cause modification of the protein. A clever means for surmounting this difficulty is illustrated in Fig. 15-1. The technique used here is to first replace cations with ammonium ions and anions with acetate ions. These ions exert a strong buffering action on the solution, preventing extremes in pH. The ammonium and acetate ions are then replaced by hydrogen and hydroxyl ions. The entire process is conveniently and rapidly carried out by passing the protein solution through a single column packed as

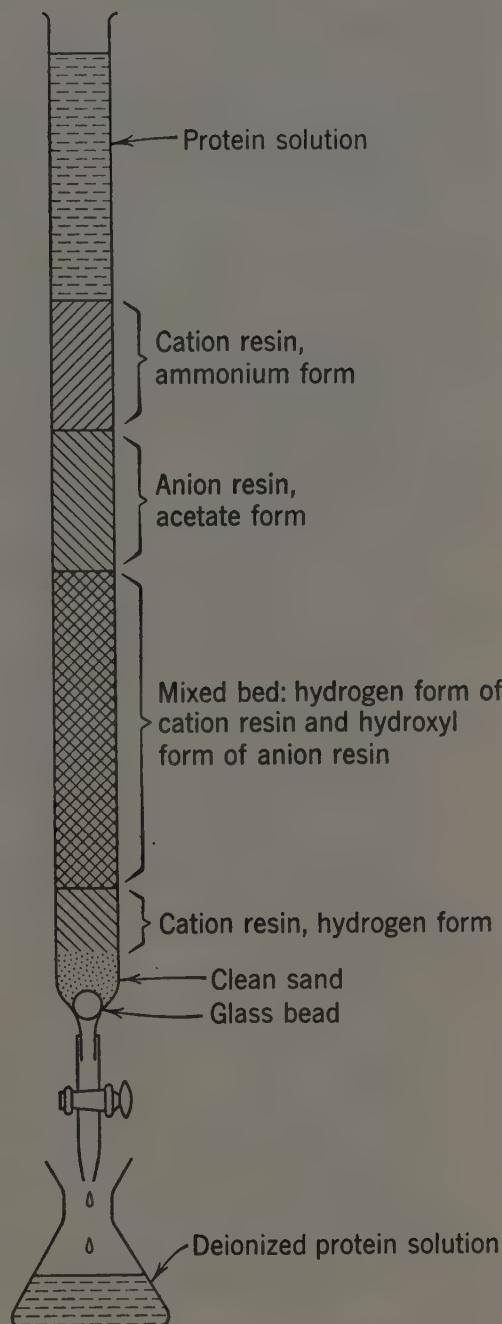


Fig. 15-1. Convenient column for deionization of protein solutions, developed by H. M. Dintzis, Ph.D. thesis, Harvard University, 1952.

These ions exert a strong buffering action on the solution, preventing extremes in pH. The ammonium and acetate ions are then replaced by hydrogen and hydroxyl ions. The entire process is conveniently and rapidly carried out by passing the protein solution through a single column packed as

indicated. This is now by far the most convenient method for preparing isoionic solutions of water-soluble proteins.

A technique closely related to dialysis is *ultrafiltration*. Here again a semipermeable membrane is used, but pressure is applied to the protein solution to force solvent and low-molecular impurities through the membrane, the proteins being retained. The relationship of this technique to filtration is obvious, hence the name.

It is also possible to effect a separation from low-molecular impurities by removing the proteins by precipitation. This calls for the use of one of the exhaustive protein precipitating agents, such as picric acid or trichloroacetic acid. Such precipitations are apt to damage proteins and should be employed only when interest is focused on the low molecular weight substances, not as a step in the purification of the protein. A few atypical proteins, particularly carbohydrate-rich glycoproteins, are not precipitated by these reagents. On the other hand, some substances containing nonprotein nitrogen are precipitated.

Removal of Solvent. In preparing proteins one finally encounters the problem of removing the solvent, usually water. The simple technique of drying in an oven, even a vacuum oven, is not suitable for most proteins, since changes in properties are apt to result. Simple air-drying, while somewhat better, is not usually satisfactory because of the possibility of denaturation and also because the product obtained is apt to be in a hard, horny state. A classical scheme for effecting dehydration is through the agency of a hydrophilic organic solvent. The aqueous protein solution is precipitated by pouring it slowly into a large volume of alcohol or acetone, usually at low temperature; the precipitate is further dehydrated by washing with more of the organic solvent; and the organic solvent is finally removed by air-drying or vacuum drying. The objective in such procedures is to remove the water as quickly as possible. Once the protein is anhydrous, denaturative changes usually take place only slowly. Furthermore, the product is obtained in a physical state in which it will redissolve readily.

Freeze-drying or *lyophilization* is now without question the method of choice for drying proteins (Fig. 15-2). In this method the protein solution is rapidly frozen in the form of a thin layer, and the solvent is removed under high vacuum by sublimation. Usually the freezing is carried out in a round-bottom flask by twirling in a Dry Ice and alcohol bath. The solution then remains frozen, owing to the high heat of sublimation, provided the condenser is maintained at a low temperature (for example, by immersion in a Dry Ice bath). The

product so obtained is in a very light, fluffy state and redissolves rapidly. The protein molecules are "caught" by quick-freezing in the solution state, and the solvent is removed without their having much chance to aggregate. The product might almost be regarded as a

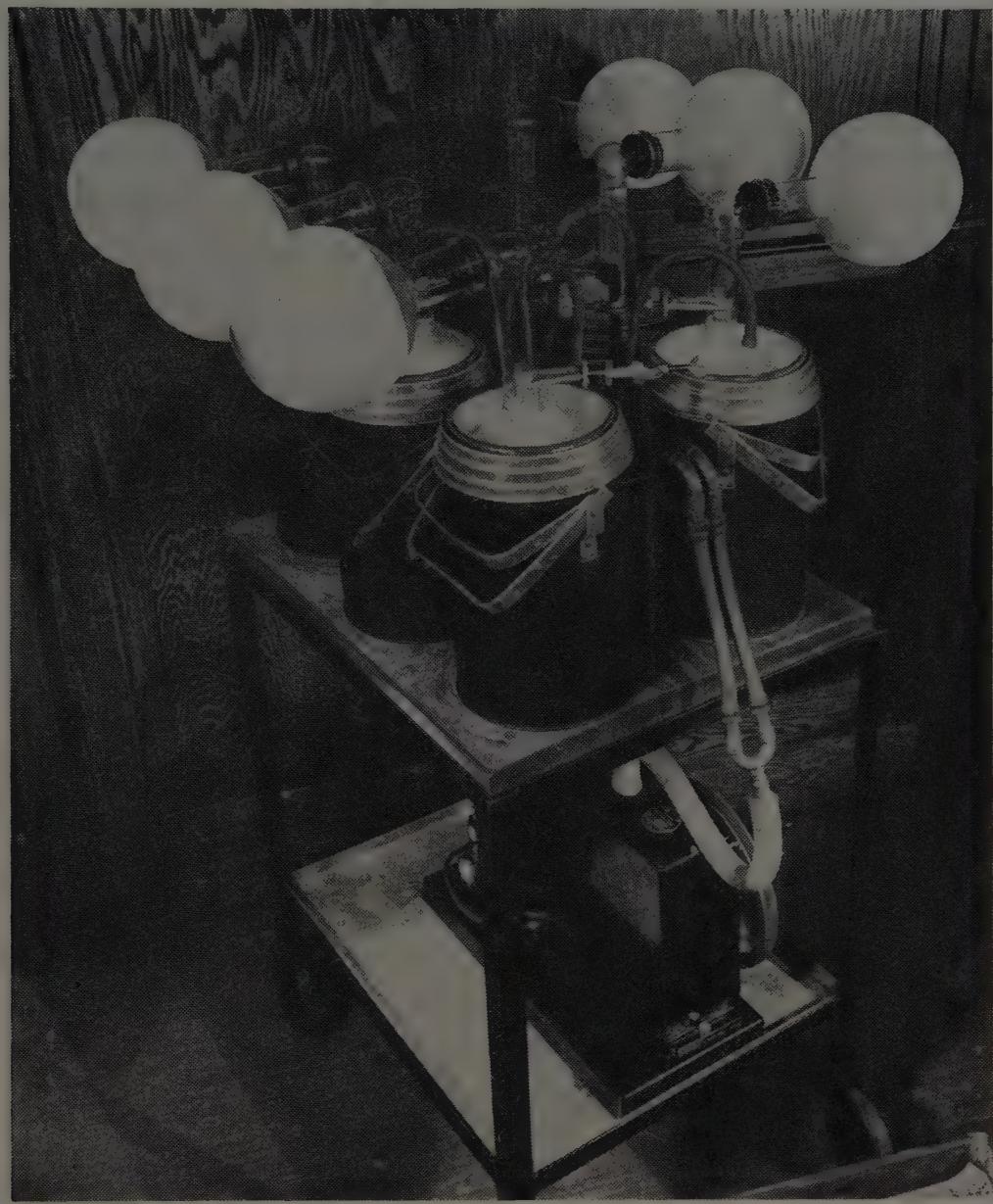


Fig. 15-2. Apparatus for the lyophilization of proteins. Courtesy of the University Laboratory of Physical Chemistry, Harvard Medical School.

"molecular solid." Freeze-drying represents a great step forward in the drying of protein, but it is not entirely safe. Some proteins, for example ovalbumin, are known to be damaged (denatured) in this process.

For concentrating protein solutions the technique of *pervaporation* is useful. Protein solution contained in a semipermeable bag or tube is suspended in a strong current of air, the water evaporating through

the membrane. This can be carried out in the cold room so that denaturation can be kept to a minimum.

PROTEIN FRACTIONATION

The really difficult part of protein preparation lies in the separation of the desired protein species from other proteins in the system. Such separations are based, in general, on differences in the solubility properties of the proteins present, which often are very minor. It is appropriate here to consider only the basic techniques, all of which are of themselves relatively simple. The difference between success and failure in a protein fractionation lies in the skill and ingenuity with which the various basic operations are combined.

Variation of pH. Since in general the solubility of proteins is very pH-dependent, variation of the acidity or alkalinity of the solution is useful in attaining separations. Proteins are least soluble at or near the isoelectric point. In a system consisting of several proteins, one of which is to be precipitated, it is therefore logical to adjust the pH of the medium to the isoelectric point of the protein in question. In some cases this may be all that is necessary; usually, however, it will be necessary to control carefully other factors, such as salt concentration, to effect removal of the desired protein without excessive precipitation of other proteins.

Variation of Temperature. Normally the solubility of proteins decreases with decreasing temperature.* Edestin, for example, has commonly been purified by cooling and diluting a warm salt solution of the protein. The temperature effect is rather nonspecific; all proteins in the system can be expected to have comparable temperature coefficients of solubility. Components with unusually high temperature coefficients are encountered occasionally, and control of temperature then becomes one of the important factors in making separations. Because of the marked thermal instability of most proteins it is desirable to operate at the lowest temperature practicable.

Variation of Salt Concentration. As was seen in Chap. 14, salt concentration (ionic strength) plays a vital role in determining the solubility of proteins, and control of this quantity must certainly be exercised in protein fractionation. The classical protein fractionation schemes and even the classification of proteins are based predominantly on the effect of salt concentration on solubility. Globulins are insoluble in pure water at the isoelectric point, but are dissolved by relatively small amounts of salt. Albumins, on the other hand, are

* Sometimes, especially in concentrated salt solution, the solubility of protein may decrease with increasing temperature.

soluble in the absence of salt. Thus, in a solution containing a mixture of an albumin and a globulin, the latter can often be precipitated by adjusting the solution to a pH near the isoelectric point and dialyzing out the salt.

Proteins differ also in their salting-out behavior, the ease with which they can be precipitated by adding large quantities of salt to the system. Many classical schemes call for the precipitation of globulins by making the system 30 to 50 per cent saturated with respect to salt, for example ammonium sulfate. Albumins usually require higher concentrations of salt for precipitation. Such salting-out schemes have proved highly effective and are still in wide use. Again, judicious combination with variation of pH and temperature is helpful.

Control of Dielectric Constant. The influence of the dielectric constant of the medium on protein-protein and protein-ion interactions was pointed out in Chap. 14. Attempts have long been made to utilize organic precipitants in protein fractionation, the principal effect of which is to diminish the dielectric constant. Until 1940 such methods met with limited success, primarily because in many cases reduction in dielectric constant is conducive to denaturation changes in the protein. Great advances have been made in the fractionation of blood plasma by E. J. Cohn and his collaborators by controlling dielectric constant through the addition of alcohol. In these methods care is taken to maintain the system at all times at a low temperature, usually just above the freezing point of the solvent mixture. This technique has several advantages. (1) One more variable is introduced, control of which increases the number of possible separations. (2) Separations are always made in the salting-in region of the solubility curve, never in the salting-out region. Proteins differ more in salting-in behavior; furthermore, this salting-in effect is accentuated by the reduction in dielectric constant. (3) There is the very practical advantage that alcohol can be removed simply by lyophilization of the aqueous-alcohol solution, whereas removal of the high concentrations of salt used in salting-out procedures usually requires prolonged dialysis. In addition to their application to the separation of the plasma proteins, the low dielectric constant procedures have been used with some success for egg white and liver proteins. These methods should be applicable to many other biological systems.

Selective Precipitation. In some cases a component can be precipitated preferentially by formation of a complex. By means of an anionic detergent it is possible to precipitate components that carry a net positive charge, leaving in solution the negatively charged com-

ponents. If the pH of the solution is adjusted to a value intermediate between the isoelectric points of two constituents, separation should result. Negatively charged colloidal substances such as the nucleic acids behave similarly. With a positively charged precipitant the

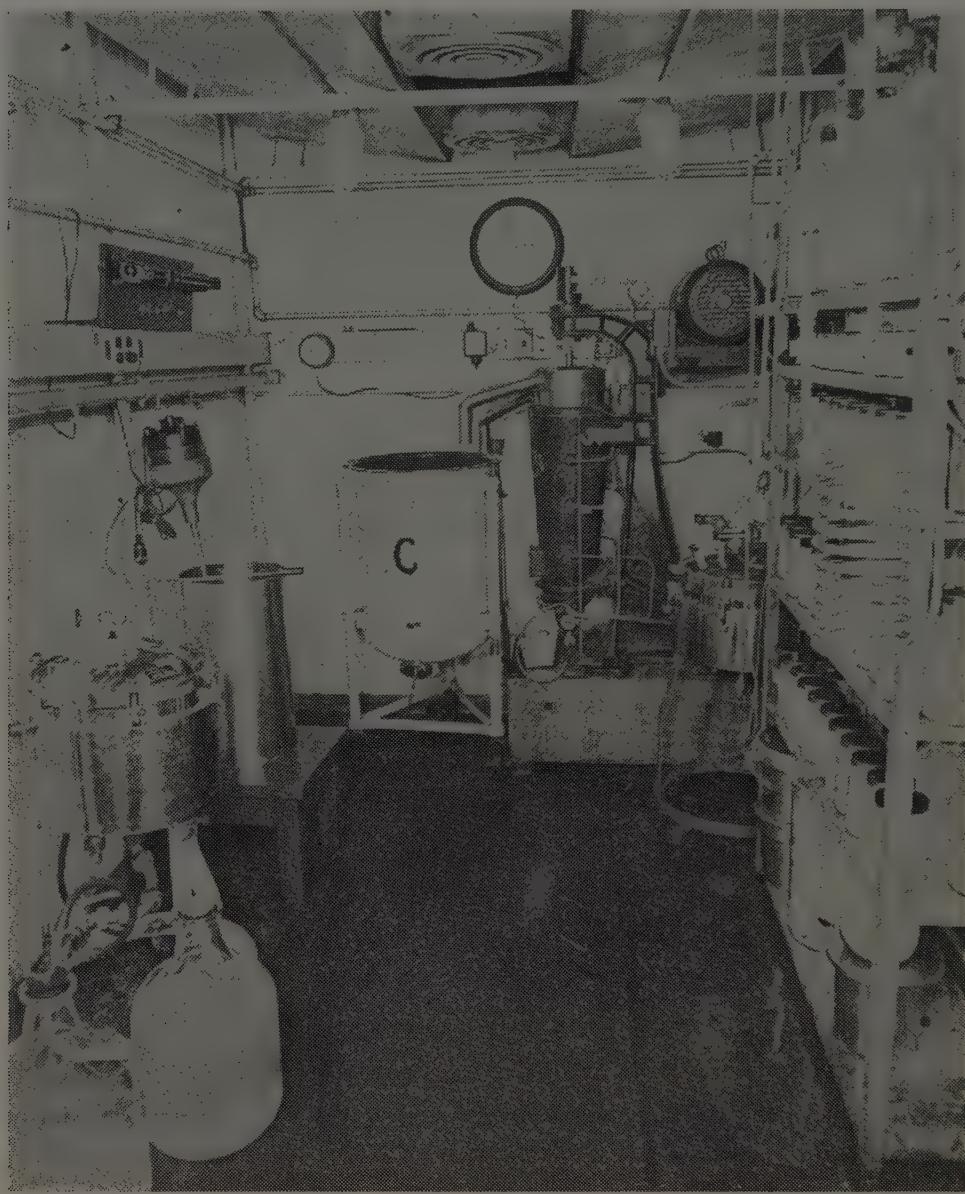


Fig. 15-3. Low-temperature laboratory for the fractionation of plasma proteins, on a pilot-plant scale, by low-dielectric procedures. Courtesy of the University Laboratory of Physical Chemistry, Harvard Medical School.

inverse situation holds, components with a negative charge being precipitated. These methods are of little significance from a preparative point of view because of possible alterations of the proteins, both those precipitated and those remaining in solution. They do provide a possible analytical technique.

The ability of proteins to form complexes with smaller organic ions, dipolar ions such as amino acids, and other organic compounds, as

well as with metal ions such as zinc, manganese, barium, calcium, mercury, iron, and copper, can be utilized in some cases to effect selective precipitations. Advantage of this property has been taken by Cohn and co-workers in improving the fractionation of the plasma protein system.

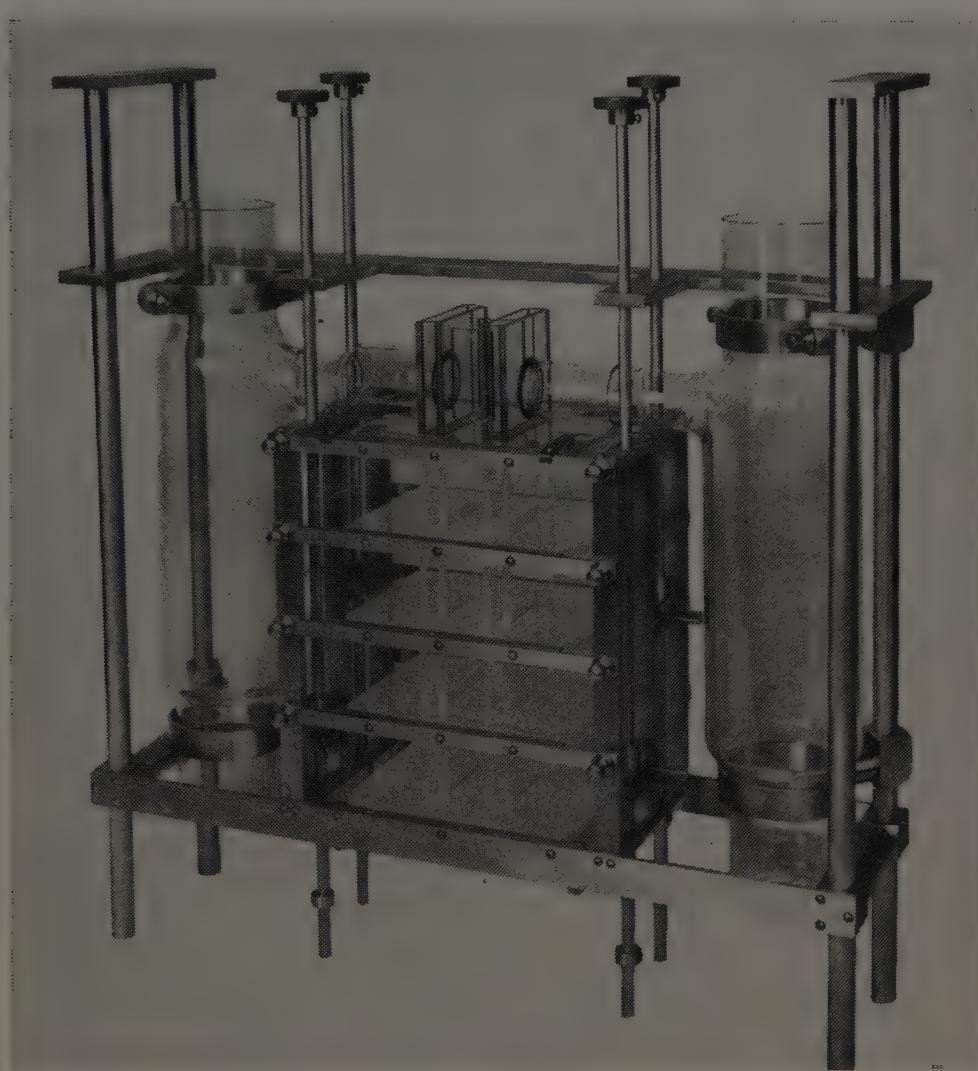


Fig. 15-4. Cell for preparative electrophoresis. Courtesy of the Klett Manufacturing Co., New York.

Preparative Electrophoresis. Electrophoresis has proved most useful as an analytical technique. It is possible, however, to isolate components in an electrophoresis cell. The fastest component can be isolated in essentially pure form behind the leading ascending boundary; the slowest component, ahead of the slowest descending boundary (Chap. 12). The ordinary analytical cell is of little use for this purpose except where the isolation of only a few milligrams of a protein is desirable. Large preparative cells have been developed and have proved quite useful (Fig. 15-4).

Electrophoresis-Convection. As the name implies, this method involves a combination of effects due to electrophoretic migration and convective flow. In a relatively primitive form the method was used as early as the 1920's in attempts to fractionate starch, latex rubber,

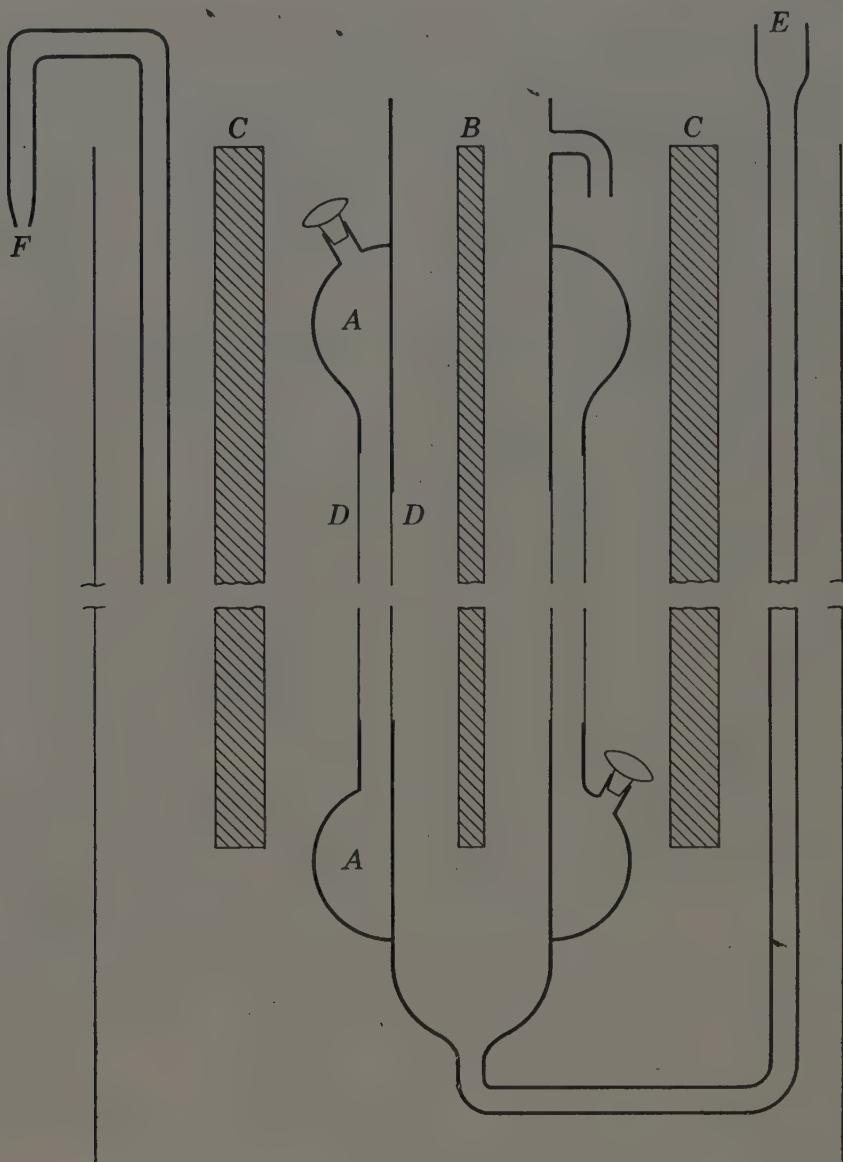


Fig. 15-5. Apparatus for the purification of proteins by electrophoresis-convection, as developed by L. Nielsen and J. Kirkwood. A, reservoirs; B, inner electrode; C, outer electrodes; D, cellulose membranes; E, inlet for buffer solution; F, outlet for buffer solution. From *J. Am. Chem. Soc.*, 68, 181 (1946).

and other colloidal substances. The solution to be separated is subjected to electrophoresis in a cell similar to the type used for electro-dialysis, in which the electrodes are separated from the solution by means of semipermeable membranes (Fig. 15-5). The colloidal constituent of higher mobility collects at one membrane, owing to its inability to pass through, thus yielding a layer of enhanced density. This layer tends to settle and take with it the more mobile constituent;

the less mobile constituent or constituents collect in the top. This method has been successfully applied to the separation of the various electrophoretic constituents of plasma. The γ -globulin fraction has been resolved into eight fractions, differing in mean isoelectric point from 5.74 to 7.31, but these fractions are by no means pure. It has been possible to effect a partial separation of the two principal electrophoretic components of ovalbumin, A_1 and A_2 (Chap. 19). A much expanded application of this technique can be anticipated.

Preferential Adsorption. Adsorption has been used to a limited extent in the purification of proteins, particularly in the case of enzymes. The adsorption of proteins is usually irreversible, and has been viewed with suspicion because of the possibility of denaturation taking place. Furthermore, extreme conditions such as strongly alkaline buffers are usually needed for elution.

By operating at high salt concentrations, somewhat below the level needed for salting-out, Tiselius and co-workers (1949) have been able to attain what they considered to be reversible adsorption. Under such conditions adsorbents which show little affinity for proteins at low salt concentration become effective. These workers have also been able to effect a satisfactory separation of serum albumin and globulin by chromatography on silica gel columns at low ionic strength. Enzymes have been successfully separated on paper piles.

Another chromatographic procedure is based on the use of a cation-exchange resin, the adsorption taking place primarily by virtue of coulombic attractions. Intensified research on chromatographic separation of proteins can be anticipated.

SELECTED REFERENCES

Books and Reviews

- "Criteria of Purity Used in the Study of Large Molecules of Biological Origin," N. W. Pirie, *Biol. Revs. Cambridge Phil. Soc.*, **15**, 377 (1940).
- "The Preparation of Pure Proteins," G. Schramm, *Angew. Chem.*, **54**, 7 (1941).
- "The Plasma Proteins and Their Fractionation," J. T. Edsall, *Advances in Protein Chem.*, **3**, 384 (1947).
- "Preparative Electrophoresis and Ionophoresis," H. Svensson, *Advances in Protein Chem.*, **4**, 251 (1948).
- Freeze-Drying; Drying by Sublimation*, E. W. Flosdorff, Reinhold, New York, 1949.
- "The Isolation of Proteins," J. F. Taylor, Chap. 1 in Neurath and Bailey's, *The Proteins*, Vol. 1A, Academic Press, 1953.
- "The Microheterogeneity of Proteins," J. Colvin, D. Smith, and W. Cook, *Chem. Revs.*, **54**, 687 (1954).

Original Articles

"The Fractionation of Proteins by Electrophoresis-Convection," L. E. Nielsen and J. G. Kirkwood, *J. Am. Chem. Soc.*, **68**, 181 (1946).

"A System for the Separation into Fractions of the Protein and Lipoprotein Components of Biological Tissues and Fluids," E. J. Cohn *et al.*, *J. Am. Chem. Soc.*, **68**, 459 (1946).

"Adsorption Separation by Salting-Out," A. Tiselius, *Arkiv. Kemi. Mineral. Geol.*, **26B**, No. 1 (1948).

"The Fractionation of Proteins by Electrophoresis-Convection," J. R. Cann and J. G. Kirkwood, *Cold Spring Harbor Symposia Quant. Biol.*, **14**, 9 (1949).

"Chromatography of Proteins. The Effect of Salt Concentration and pH on the Adsorption of Proteins on Silica Gel," C. C. Shepard and A. Tiselius, *Discussions Faraday Soc.*, **7**, 275 (1949).

"Chromatographic Analysis of a Mixture of Egg White Proteins," H. A. Sober, G. Kegeles, and F. J. Gutter, *Science*, **110**, 564 (1949).

"Separation of Enzymes on the Filter Paper Chromatopile," H. K. Mitchell, M. Gordon, and F. A. Haskins, *J. Biol. Chem.*, **180**, 1071 (1949).

"Quantitative Procedures for the Separation of the Protein Components of Human Plasma," E. J. Cohn *et al.*, *J. Am. Chem. Soc.*, **72**, 465 (1950).

"Chromatography of Proteins. Ribonuclease," C. H. W. Hirs, W. H. Stein, and S. Moore, *J. Am. Chem. Soc.*, **73**, 1893 (1951).

"The Application of Four Methods for Assessing Protein Homogeneity to Crystalline β -Lactoglobulin. An Anomaly in Phase Rule Solubility Tests," D. Smithies, *Biochem. J.*, **58**, 31 (1954).

"Separation of Neutral Proteins on Ion-Exchange Resins," N. Boardman and S. Partridge, *Biochem. J.*, **59**, 543 (1955).

"Continuous Zone Electrophoresis by Crossed Velocity Fields in a Supporting Medium," I. Brattsten, *Arkiv. Kemi.*, **8**, 347 (1955).

"Separation of Proteins by Gradient Solvent Extraction of a Protein Precipitate," C. Zittle and E. Dillamonica, *Arch. Biochem. and Biophys.*, **58**, 31 (1955).

"Chromatography of Proteins. I. Cellulose Ion Exchange Adsorbents," E. Peterson and H. Sober, *J. Am. Chem. Soc.*, **78**, 751 (1956).

16

Protein Structure

Chemical composition

Linkages

Arrangement of amino acid residues

Fibrillar proteins

Possible configurations for polypeptide chains

Globular proteins

The nature of folded structure

Determination of the complete detailed structure of the proteins has been a challenge to workers in the field for over fifty years. Much progress has been made, but admittedly a complete understanding of these most complex of chemical compounds has not yet been attained. The importance of such structural investigations cannot be over-emphasized.

It is the purpose of this chapter to review the salient features of protein structure. A few things can be said with assurance. Usually, however, ideas will be examined that are still quite nebulous, the evidence for which is not at all adequate. In a few cases, hypotheses presented are actually in conflict with the present evidence and hence no longer tenable. They are included here by way of broadening the perspective and rounding out the historical development of the subject.

The difficulty and importance of the attack on protein structure has been summarized most elegantly by Sir Lawrence Bragg:

One must be prepared to abandon without hesitation any proposed model, however attractive it may appear, if further evidence indicates that it is not valid. We are like a climber, striving to conquer a most difficult pitch, and

grateful for the slenderest cracks and projections which seem to offer a possible hold for hand or foot. The results obtained so far may seem meagre, but a glittering prize draws us on. There must be some very deep and fundamental reason why Nature has chosen the polypeptide chain as the building principle for all forms of living matter, a reason which should become clear if we can solve the mystery of the structure of the protein molecule.*

CHEMICAL COMPOSITION

The "Building Units." The proteins may be considered, chemically, as anhydrocopolymers of amino acids which are predominantly, if not entirely, of the L-configuration. The problems involved in the determination of the amino acids have been discussed (Chap. 6). In spite of the difficulties many data are at hand, and for some proteins the entire amino acid composition may be considered to be known with a reasonable degree of accuracy.

One of the earliest studies to emphasize the advance of knowledge in this field is the characterization of β -lactoglobulin, by Brand and co-workers (1945), as $\text{Ala}_{29}\text{Arg}_7\text{Asp}_{36}\text{CySH}_4(\text{CyS}-)_8\text{Glu}_{24}(\text{GluNH}_2)_{32}\text{Gly}_8\text{His}_4\text{Iso}_{27}\text{Tyr}_9\text{Try}_4\text{Val}_{21}$. (It is now common to indicate an amino acid by the first three letters of the name, glutamine by GluNH_2 , etc.)

When one compares the composition of β -lactoglobulin with that of any other protein for which a balance sheet is available, he will usually observe certain similarities. Leucine, for example, is ordinarily present in high proportion, 50 to 120 residues per 10^5 gm. Isoleucine is present almost always in smaller proportion than is leucine. Tryptophan is present in small proportion, and is absent in some cases, for example, insulin. Aspartic acid and glutamic acid together often constitute about 25 per cent of the total, and aspartic acid is present almost always in lesser proportion. Amino acids present in relatively variable proportion include valine, methionine, threonine, glycine, alanine, proline, and serine. It is possible to emphasize either the similarities or the dissimilarities.

The relationships have been subjected to statistical analysis. When this was performed on homologous proteins (proteins which occupy a similar biochemical place in different organisms), the compositions were found to be almost identical. When the same calculations were performed on two dozen *heterologous* (of varied function) proteins, the correlation in composition was found to be closer to complete identity than to entirely random composition. It follows that composition of one protein is not independent of that of another.

The theoretical number of possible isomers of a molecule of equine

* *Nature*, 164, 7 (1949).

hemoglobin, from which the isomers are obtained solely by rearrangement of amino acid residues, is 4×10^{619} . Some idea of the enormity of this figure can be gained from the fact that a cube one billion light years long per edge could hold no more than 10^{103} molecules of hemoglobin. From this sort of consideration and the statistically correlatable similarity in composition, it follows that the actualities are an infinitesimal fraction of the possibilities. Typical compositions of some common proteins are given in Table 6-3.

Prosthetic Groups. Although it should be recalled at this point that some proteins are composed of amino acid residues only, other (*prosthetic*) groups are present in many proteins. These are the so-called *conjugated* proteins. Most important of these are:

(1) The *nucleoproteins*, which are conjugates of protein and nucleic acid, either ribose nucleic acid or deoxyribose nucleic acid. These very important substances, found particularly in the cell nuclei (chromosomes) and in viruses, are discussed further in Chap. 23.

(2) The *chromoproteins*, which contain a prosthetic group consisting of a central metal atom surrounded by a tetrapyrrole structure (one of the porphyrin structures). Important in this class are various of the oxidative enzymes such as cytochrome and peroxidase, hemoglobin and chlorophyll. As the name implies, they are highly colored.

(3) The *glycoproteins*, which contain carbohydrates or carbohydrate derivatives and are found particularly in various mucus secretions.

(4) The *phosphoproteins*, which contain phosphate groups, are presumably esterified with the hydroxy residues of the hydroxyamino acids. Casein is the best known example of this type.

It is not always easy to decide whether such groups are truly part of the protein molecule or are tenaciously bound "impurities." There may be, in fact, no real difference. Some of the prosthetic groups have special significance chemically or biologically, for example, the heme moiety of hemoglobin.

LINKAGES

The Peptide Bond. After consideration of the composition of proteins, the question which naturally arises is that of the mode of linkage of the building units. It is now generally accepted that the predominant chemical linkage in proteins is the α -peptide linkage, that is, an amide linkage between α -amino and α -carboxyl groups of adjacent amino acid residues. The peptide concept of protein structure was proposed independently by Emil Fischer and F. Hofmeister in 1902. Among the more important pieces of evidence for the validity of the peptide concept may be mentioned the following:

(1) The biuret reaction, which characterizes structures containing multiple peptide bonds, is positive for proteins.

(2) Although native proteins contain little α -amino nitrogen and relatively few free carboxyl groups, these are numerous after hydrolysis.

(3) Hydrolysis of proteins liberates equivalent amounts of amino and carboxyl groups. Experiments which illustrate this have had to be corrected for amide nitrogen and for proline residues. When so corrected, the equivalence observed is quite satisfactory.

(4) Synthetic peptides containing appropriately placed residues are hydrolyzed by proteases which also hydrolyze proteins.

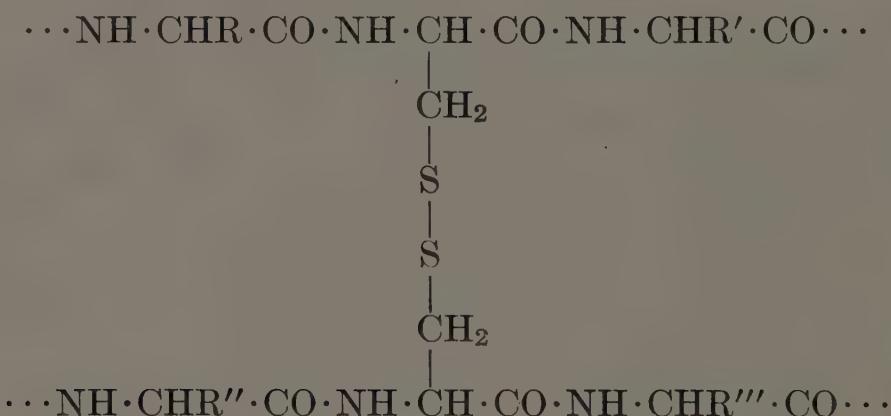
(5) Proteolytic enzymes are capable of catalyzing the synthesis of compounds which contain the peptide bond.

(6) The absorption spectra of proteins in the infrared show maxima attributable to the —NH— and C=O groups.

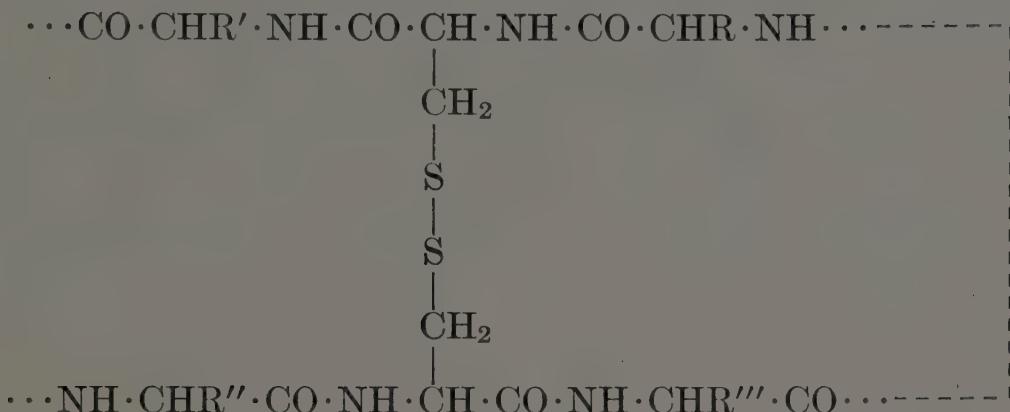
The peptide theory is now widely accepted, but acceptance came slowly, requiring several decades. Many workers still doubt that all bonds between all amino acid residues in proteins are the simple peptide linkage of the Fischer-Hofmeister theory. Some proteins yield color tests for diketopiperazine rings, and although the ring structure can hardly represent a large proportion of the total structure of a protein, it is not yet possible to rule out a mixed structure in which the diketopiperazine contribution is a minor one on a percentage basis.

It is also true that the peptide bond need not be an α -bond. Glutamic acid and aspartic acid can be linked through their δ - or γ -carboxyls respectively.

Linkages Other than Peptide Bond. (1) *The disulfide linkage* (—S—S—) is present in proteins containing cystine residues. This structure may cross-link two separate polypeptide chains, as indicated schematically by the structure:

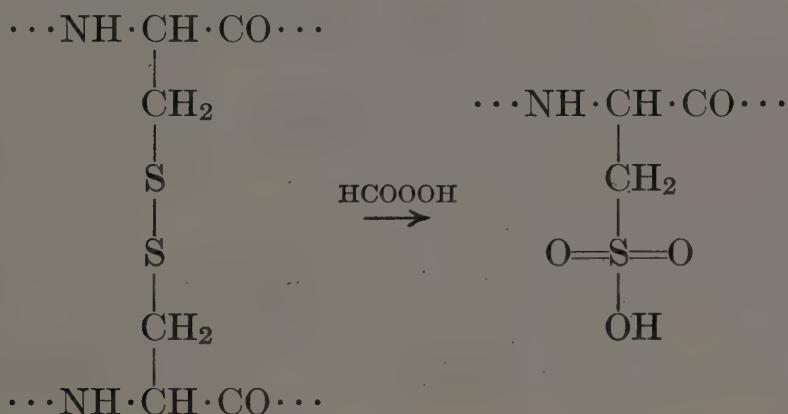


or may cross-link separate segments of a single chain:



In the latter case a closed loop results.

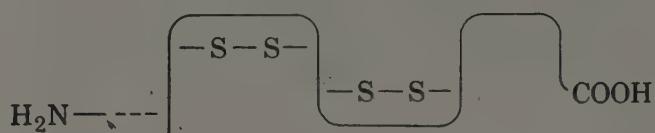
There is little doubt that the disulfide bridge contributes to the folded nature of many protein molecules. Following oxidation with performic acid, insulin can be separated into individual large peptides containing cysteic acid residues:



The resulting two chains, A and B, differ in amino acid composition and sequence.

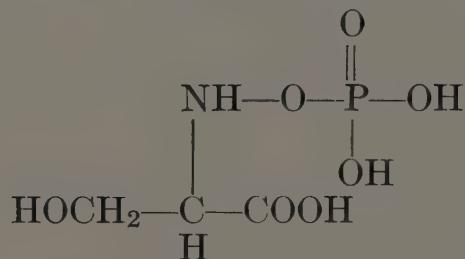
In other cases disruption of the disulfide bonds in proteins yields no reduction in molecular weight, demonstrating cross-linking of the second type. The molecular weight of performic acid-oxidized ribonuclease (about 14,000) is the same, within experimental error, as that of the native protein, although the properties of the protein are drastically altered. Another important example is lysozyme, which possesses one terminal amino, one terminal carboxyl, and several cystine residues per molecule. When lysozyme is reduced, all of the cystine is converted to cysteine residues, and these can be alkylated. It has been demonstrated that in such a process there is no measurable reduction in molecular weight. It has been suggested

that the likely interpretation of such studies is a coiled cross-linked structure,



The insolubility of the cystine-rich keratins has been attributed to a high degree of cross-linking. The fact that such proteins are solubilized by reducing agents lends credence to this view.

(2) *Ester linkages* have been proposed. These would necessarily be formed from the alcoholic groups of amino acid side chains, as in serine and threonine, plus available carboxyls. The evidence for such ester linkages is to a degree circumstantial. Serine appears to be unreactive in some proteins. Vitamin phosphates are known to be of significance in enzyme structures. Following enzymic hydrolysis, phosphorylserine



has been isolated from casein.

(3) *Coulombic attractions* (and repulsions) must be of great significance in proteins, owing to the presence of many positive and negative charges at or near the isoelectric point. Coulombic attractions may well be one of the important forces holding a peptide chain in the specific folded configuration in the globular proteins. In the extreme case, such interactions could take the form of "salt linkages" or ion-pair bonds of the type



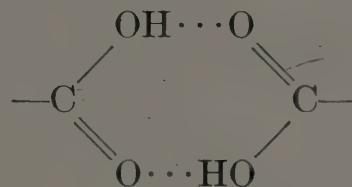
Jacobsen and Linderstrom-Lang (1950) have attempted to rule out such interactions on the basis that volume changes accompanying the combination of proteins with hydrogen ions indicate the ionic groups to be highly solvated (electrostriction). The volume changes are relatively small, however, and in view of the complexity of proteins the argument does not seem entirely convincing. In any event, it does not seem justifiable to rule out completely the possibility of such interactions in all proteins.

(4) *Hydrogen bonds* have been suggested repeatedly as important intramolecular bonding forces, and may indeed be the principal such

forces. There is ample opportunity for hydrogen bonding as a consequence of the carbonyl and amide groupings of the peptide chain. A possible schematic pattern for such bonding is indicated by the dotted lines in Fig. 16-1. Also, carbonyl-amide hydrogen bonds assume a key significance in the Pauling-Corey helix structures, to be discussed. In addition, certain of the side-chain groupings present the possibility of additional stabilization through hydrogen bonding. There is excellent evidence that the phenolic group of tyrosine participates in hydrogen bonds in at least some proteins, possibly in conjunction with carboxylate anions, schematically,



Bonds involving two COOH groups, known to exist in acetic acid, have also been suggested.



ARRANGEMENT OF AMINO ACID RESIDUES

Regularity in Order. A number of attempts to find regularity in the order of amino acid residues in proteins are recorded in the literature. One such proposal, which received serious consideration in the late 1930's and early 1940's, was Bergmann's periodicity hypothesis. A principal proposition of this hypothesis stated that each amino acid recurred in the peptide chain at a regular interval. This hypothesis is now generally considered to be untenable against the accumulated evidence, but its dominance for a while stimulated much interesting work. At the present time, there is evidence for a kind of regularity of order between protein molecules rather than within protein molecules. Another way of saying this is that proteins appear to differ through a continuous series of small changes, which are found both for molecules within the same organism and for molecules between organisms.

Chibnall's Subunit Hypothesis. The concept that the protein molecule consists of smaller peptide units, each with its own amino acid pattern, arose from a comparison of amino nitrogen content and lysine assays. If a protein molecule were one long peptide chain, it would have one α -carboxyl group and one α -amino group. Chibnall has determined for a number of proteins the proportion of lysine and also the Van Slyke amino nitrogen. Since the lysine value can be con-

verted to an $\epsilon\text{-NH}_2$ value, one can calculate the $\alpha\text{-NH}_2$ by the equation

$$\sum \text{NH}_2 \text{ (total NH}_2) = \epsilon\text{-NH}_2 + \alpha\text{-NH}_2$$

or

$$\alpha\text{-NH}_2 = \sum \text{NH}_2 - \epsilon\text{-NH}_2$$

When this is done, the value of $\alpha\text{-NH}_2$ for most proteins exceeds one residue per mole and is a whole number. Chibnall suggested that this can be explained by the concept that the protein molecule is a combination of submolecules or subunits. He suggested that these subunits may be linked by some of the linkages described or by other possibilities. The general idea of a subunit was first most strikingly supported by Sanger's work which revealed the A and B chains of insulin. Chibnall's calculations, however, indicate six termini per 12,000 molecular weight, as contrasted to four in Sanger's formulation.

Order of Residues. Knowledge of sequential arrangements of amino acid residues began to appear in the early 1930's. As has been explained, arrangements of residues can be expected to be more definitive and meaningful than simple compositional values. Insofar as complete elucidation of sequences is concerned, the acquisition of such knowledge is not likely to be very rapid in the near future. One structural feature which is essentially sequential, however, can be clarified and correlated with fair rapidity. This is the terminal amino acid residue. At times the feasibility of determination applies also to its quantitative evaluation. Data of this nature, as already accumulated, apply to a considerably larger number of proteins than does fuller structural evaluation. These data are also found to correlate more fully than simple amino acid composition with comparisons of proteins, as in Table 16-1.

Some positions of residues other than terminal are mentioned in Table 16-1, but for the most part longer sequences are discussed elsewhere. In particular should be mentioned the spectacular elucidation of the residue sequence in insulin, discussed in Chap. 20. Many more termini and sequences are known than are presented in the table; the examples included serve to explain, however, several relationships.

For most proteins it is not possible to indicate the optical configuration of the designated amino acid residue. In some cases, however, the assignment or estimation has been performed with the aid of assay bacteria that are stereospecific for the L-amino acid in their nutritional requirement.

It is clear also that the N-terminal residue is more often designated than the C-terminal residue. This stems from the nature of the terminal groups, amino and carboxyl, and reflects the fact that the

TABLE 16-1. Terminal Residues of Some Proteins

Protein	N-Terminus	C-Terminus
Lysozyme	L-Lysyl	L-Leucine
Rabbit spleen lytic factor	Lysyl	
Papain	Isoleucyl	
Trypsin	Isoleucyl	
α -Chymotrypsin	Isoleucyl	
Pepsin	Leucyl	
Tobacco mosaic virus	None	Threonine
Salmine	Prolyl	
Casein	Arginyl, lysyl	
Bovine plasma albumin	Aspartyl	
Ovalbumin	None	
Edestin	6 Glycyl, 1 leucyl	
β -Lactoglobulin	3 Leucyl	
Denatured	3 Leucyl	
Insulin	2 Glycyl, 2 L-Phenylalanyl	2 Asparagine, 2 Alanine
Corticotropin	L-Seryl	Phenylalanine
Oxytocin, pig and cow	Half cystinyl	Glycinamide
Vasopressin, pig and cow	Half cystinyl	Glycinamide
Horse myoglobin	1 Glycyl	
γ -Globulin	1 Alanyl	
Hemoglobin		
Horse	6 Valyl	
Donkey	6 Valyl	
Human	5 Valyl	
Cow	2 Valyl, 2 methionyl	
Sheep	2 Valyl, 2 methionyl	
Goat	2 Valyl, 2 methionyl	
Conalbumin	Alanyl	
Ovomucoid	Alanyl	
Avidin	Alanyl	
Whole soybean	L-Lysyl	
Whole corn seed	L-Lysyl	
Whole rye seed	L-Lysyl	
Whole wheat seed	L-Lysyl	

amino group has been more subject to chemical reactions than the carboxyl. Most types of amino acid are found as terminal residue. Monoaminomonocarboxyl types might at first glance appear to predominate, but these are probably present in no greater proportion as end groups than their total proportion in protein. Lysine is liberally represented, and arginine appears also. An aspartic acid residue has

been reported for bovine plasma albumin, but this may not be the only N-terminal residue.

For some proteins no terminal residues are recorded. This can be partly explained on the basis of a large cyclic structure. In other instances, knowledge is complete enough to show clearly that these molecules are partly cyclic and partly linear. These structures may thus be considered to be ∞ -shaped. Other possible interpretations include terminal prolyl or terminal amides and other types of masking of terminal amino or carboxyl. Gramicidin S, bacitracin, oxytocin, and tyrocidine are certainly representative of cyclopeptides.

The results on β -lactoglobulin suggest that denaturation (Chap. 17) produces no exposure of additional terminal residues. It is known, however, that other nonterminal groups, such as ϵ -amino groups of lysine, may be so exposed. Further study of this relationship is called for, since it seems possible that in some cases absence of terminal residues may be a consequence of "masking" due to the native protein structure.

The terminal residues of three proteins from chicken egg, conalbumin, ovomucoid, and avidin, are the same, namely alanyl. A fourth important egg protein, lysozyme, possesses an N-terminal lysyl. The fact that three of the four proteins isolated from the same source have an N-terminal alanyl residue indicates a striking structural homogeneity. The data on the seeds of wheat, corn, and rye point to the same conclusion, inasmuch as of fifteen amino acids estimated, L-lysine was the only one that was substantially terminal.

Table 16-1 provides not only comparisons of proteins within an organism but also comparisons of proteins between organisms. When the whole protein of the seed of corn, rye, or wheat is compared, the same pattern of structure is seen in all of these cereals. At a higher level, considerable similarities are observed between hemoglobins of different species, yet definite differences have been recorded. All of these results conform to the concept of an evolutionary thread in the biogenesis of protein (Chap. 24).

In an analogous relationship, terminal isoleucyl is seen to appear in three proteolytic enzymes, trypsin, chymotrypsin, and papain. For comparison, it may be seen that no other protein in the table possesses a terminal isoleucyl. This relationship is less astonishing for bovine chymotrypsin and trypsin, which occur together, than for papain, which is found in a plant source.

As more data accumulate, it will be of interest to see if sequences of two or more residues tend to appear more often than chance would

d dictate. In this connection, the aspartyl-glutamyl sequence has been found in at least three peptides which are not closely related functionally, rabbit antibody protein, insulin, and tyrocidine.

FIBRILLAR PROTEINS

The fibrillar proteins serve primarily a structural role. In the solid state they may range in properties from hard, highly crystalline fibers such as silk to the rubberlike contractile elements of muscle. In general they do not show the unique biological properties, such as enzymic activity, exhibited by the globular proteins. Furthermore, it appears that as a consequence of their insolubility, considerable chemical modification is usually involved in the process of solution. Under such conditions the protein in solution would not be expected to be representative of the native state, and would probably be far from homogeneous. For these reasons studies of the properties of the fibrillar proteins in solution have not proven particularly attractive. On the other hand, study of these proteins in the naturally occurring solid state, primarily through the techniques of X-ray diffraction, has been very instructive in throwing light on the nature and configuration of the peptide chain.

Silk Fibroin. Perhaps the first adequate investigation of this type was that of silk fibroin. Silk fibroin is the internal core, the truly fibrillar part, of the silk fiber. This protein has a relatively simple amino acid composition, glycine comprising almost half of the residues and alanine one fourth.

Silk fibroin yields an X-ray diagram indicating a considerable degree of crystallinity. By stretching and rolling, the degree of order can be improved and the pattern sharpened. The interpretation of such patterns imposes considerable difficulty, but the best possibility appears to be a structure involving long polypeptide chains running parallel to the fiber axis (Fig. 16-1). A characteristic spacing (identity or repeat period) of 7.0 Å. is observed along this axis; in other words, the structure repeats itself at this interval along the chain. This spacing is interpreted as representing two amino acid residues, so that the fiber length per residue would be 3.5 Å.

In addition, two spacings perpendicular to the fiber axis are observed. One of these, 4.3 Å., is invariant (as is the 7 Å. spacing) for silks from various species, and presumably represents the identity period between chains, as shown in Fig. 16-1. Both of these spacings may be considered to be characteristic of fully oriented polypeptide chains. The third spacing, perpendicular to the other two (the spacing

perpendicular to the plane of the paper in Fig. 16-1), is found to vary from protein to protein and in silks from various species. Values of from 4.6 to 5.3 Å. are found for silks, and values of up to 9 or 10 Å. are found for other proteins and are thought to correspond to this spacing. It should be recognized that the H-groups and R-groups in Fig. 16-1 project in and out of the plane of the paper. The size of the R-groups would govern the closeness of approach of adjacent planes.

The third spacing, which is presumed to be a measure of the distance between such planes perpendicular to the plane of the paper, would hence be expected to vary according to the amino acid composition of the protein, and may be considered as measuring an average dimension of the R-groups.

The three spacings are termed, in order, the fiber, backbone, and side-chain spacings.

The relatively short side-chain spacing observed in silk fibroin is doubtless a result of the preponderance of amino acids with small R-groups, H in glycine and CH_3 in alanine. Indeed, it seems unlikely that the few residues in silk that have large R-groups could occur in this crystal lattice at all. The silk fiber probably contains crystalline and amorphous regions, single polypeptide chains running through perhaps several such regions, as indicated schemati-

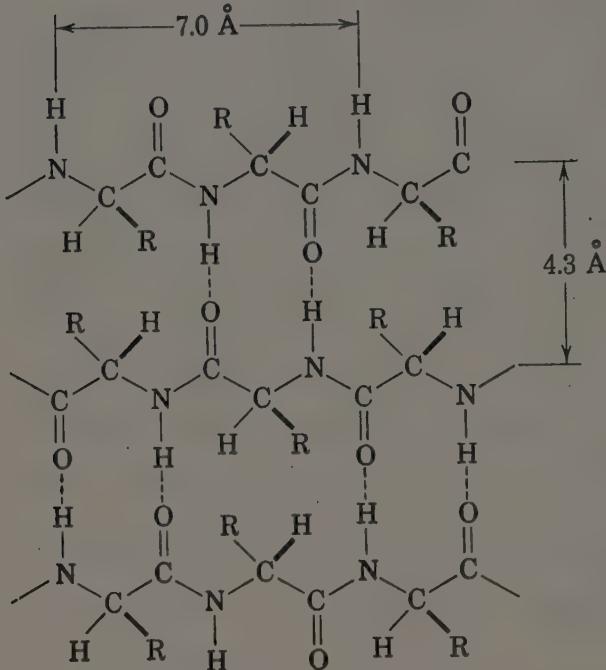


Fig. 16-1. The configuration of polypeptide chains in the extended or β -keratin configuration. The significance of the 7.0 Å. fiber axis repeat period and the 4.3 Å. backbone spacing is shown. The H-atoms and R-groups on the α -carbons are directed above and below the plane of the paper. Bold bonds denote groups directed above the plane. The variable side-chain spacing is the spacing between planes perpendicular to the plane of the figure. Shown also is a possible manner in which hydrogen bonds might serve to bond the chains together (dotted lines).

cally in Fig. 16-2. The crystalline regions may contain only glycine and alanine, the larger residues, such as tyrosine, arginine, and histidine, occurring in the amorphous regions. Evidence for this view is given by studies of the degradation of silk fibroin with hypobromite. A rapid initial reaction occurs, in which the more available amorphous regions are dissolved, leaving the crystalline portions unattacked. The latter yield the same X-ray patterns as the original silk fibroin and

apparently contain only glycine and alanine. Obviously, this result is not in accord with the ideas of Bergmann as to the regular distribution of residues along the polypeptide chain.

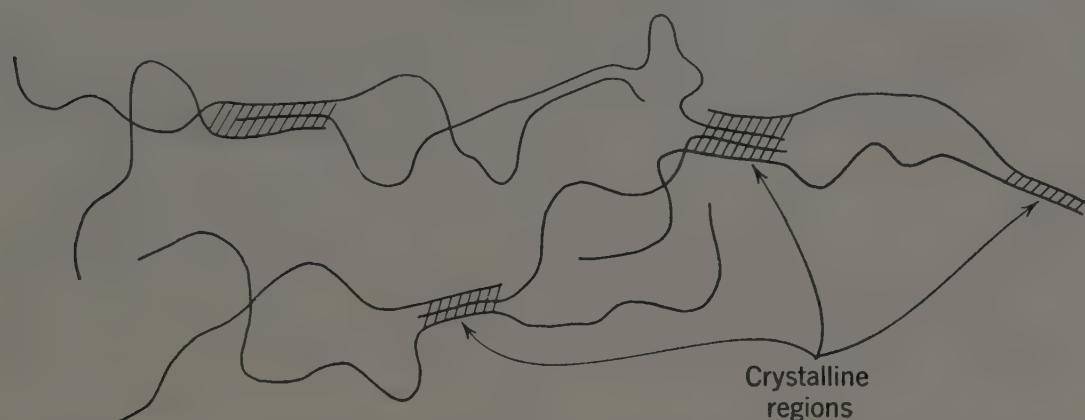


Fig. 16-2. Schematic structure of a fibrous protein, showing crystalline and amorphous regions. Single polypeptide chains may run through several crystalline and amorphous regions.

α - and β -Keratin. The crystallographic properties of the keratins are somewhat more complex (Fig. 13-7). Untreated hair or wool fibers yield an X-ray diagram which is quite different from that of silk. Upon soaking in water the fiber can be stretched to as much as double its original length, and the nature of the X-ray diagram changes to one resembling that of silk fibroin. Astbury has termed the two states of keratin the α - and β -configurations respectively. He interprets the β -configuration in substantially the same way as the silk structure, the chief difference being that the side-chain spacing is much greater, 9.8 Å instead of 4.6 to 5.2 Å, owing to the difference in amino acid composition. Keratin from wool, for example, contains little glycine and alanine but considerable cystine, glutamic acid, and arginine.

It is to be emphasized that any present interpretation of the keratin patterns, particularly the folded or α -pattern, is to be considered provisional. Many interpretations are possible, and several structures for the α -keratin configuration have been postulated. This question of the detailed configuration of polypeptide chains is considered further in the next section.

A number of other fibrillar proteins have been found to exist in configurations very similar to or identical with the α - and β -keratin structures. Most important among these are myosin (of muscle), epidermin (of epidermal tissue), and fibrin (blood clot). These form a group known as the keratin-myosin-epidermin-fibrin or KMEF group of fibrillar proteins.

Another extremely important fibrillar protein, collagen (connective

tissue), is strikingly different in its behavior, particularly with respect to X-ray diffraction. A most interesting feature of this fibrillar protein is the existence of a long-range periodicity along the fiber axis, of the order of 600 Å. This periodicity, manifested by X-ray diffraction and by electron microscopy alike, exists in collagen fibrils from a broad range of organisms. It is of interest that gelatin, a hydrolytic degradation product of collagen, yields an X-ray diffraction pattern of the same type as collagen under appropriate conditions. In spite of much intensive investigation, the structure of collagen remained obscure until 1955. The X-ray diffraction patterns suggest rather clearly some type of helical coil, but no simple peptide model is adequate. Rich and Crick (1955) have suggested that the structure consists of three helically coiled peptide chains, which in turn are coiled about one another, a coiled coil. This structure appears to account for all the known facts.

POSSIBLE CONFIGURATIONS FOR POLYPEPTIDE CHAINS

The studies just outlined on the structure of the fibrillar proteins are of great significance, not only with respect to the structures of these proteins per se, but also for the guidance they give in studies of the even more complex and possibly more important problem of structure of the globular proteins, to be discussed. Perhaps the most important result of the work on fibrillar proteins is the deduction that there are only a few characteristic configurations for the polypeptide chain, as evidenced by the close similarity of the silk fibroin and β -keratin structures and the existence of two discrete configurations for keratin. A third state, the so-called supercontracted state, also exists in the case of the keratins but there is some considerable doubt as to whether it represents a new configuration for the polypeptide chain. It is now of interest to consider, in greater detail, possible polypeptide configurations for at least the two best recognized states, the extended or β -keratin structure and the α -keratin structure or α -fold.

The Extended Configuration. A structure for silk fibroin and β -keratin has been given in Fig. 16-1. The salient features of this, and probably of any adequate configuration for the fully extended polypeptide chain, are:

(1) Alternate amino acid residues are rotated 180 degrees with respect to their neighbors, so that their R-groups project on opposite sides of the peptide chain.

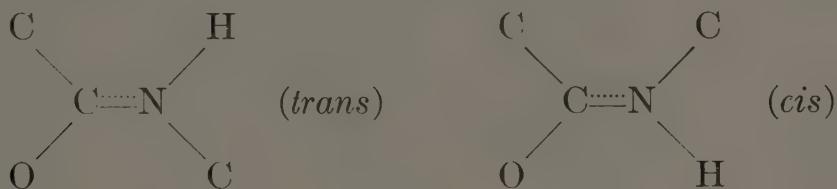
(2) The structure is maintained by interchain hydrogen bonding, that is, hydrogen bonds between NH groups of one chain and C=O groups of an adjacent parallel (or antiparallel) chain.

The characteristic dimensions of this structure are given in Fig. 16-1 and in the preceding discussion. It seems probable that the structure is essentially correct, but that it is in error with respect to detail. In particular, it is based on bond angles and bond distances taken from organic molecules other than amino acids and peptides. Scale models show that there would be crowding of the larger R-groups, so that while it might be adequate for silk, it probably is not an available configuration for polypeptide chains in general.

A much more rational basis for the development of conceivable configurations for the polypeptide chain has been provided in recent years through studies, by X-ray diffraction, of the simpler amino acids and dipeptides. Such studies have now led to important revisions in the bond angles and distances which must be expected in proteins. In particular it has been shown that:

- (1) The length of the bond joining the carbonyl carbon to the amide nitrogen is relatively short, about 1.32 Å., strongly suggesting considerable "double-bond character."
- (2) The configuration of the three valences about the amide nitrogen is nearly planar, and indeed coplanar with the valences about the carbonyl carbon. This result also suggests double-bond character for the carbon-nitrogen bond.

The conclusion that the bond between carbonyl carbon and amide nitrogen is to a considerable extent a double bond leads to a further very important deduction, namely, that two isomeric forms for the peptide units must be considered, *cis* and *trans*. These may be indicated as



It seems probable that the *trans* configuration prevails in most if not all proteins.

Utilizing these new facts, Pauling and Corey (1951) have suggested two possible extended configurations for polypeptide chains, the "pleated sheets." As the name implies, these structures lead to planar sheets possessing, however, a regular folded or pleated structure. The two forms differ in that in one all chains run the same way, that is, are parallel; in the other, alternate chains are antiparallel. These structures are shown in Fig. 16-3. They seem to be adequate in every way to explain the known facts of the β -keratin structure.

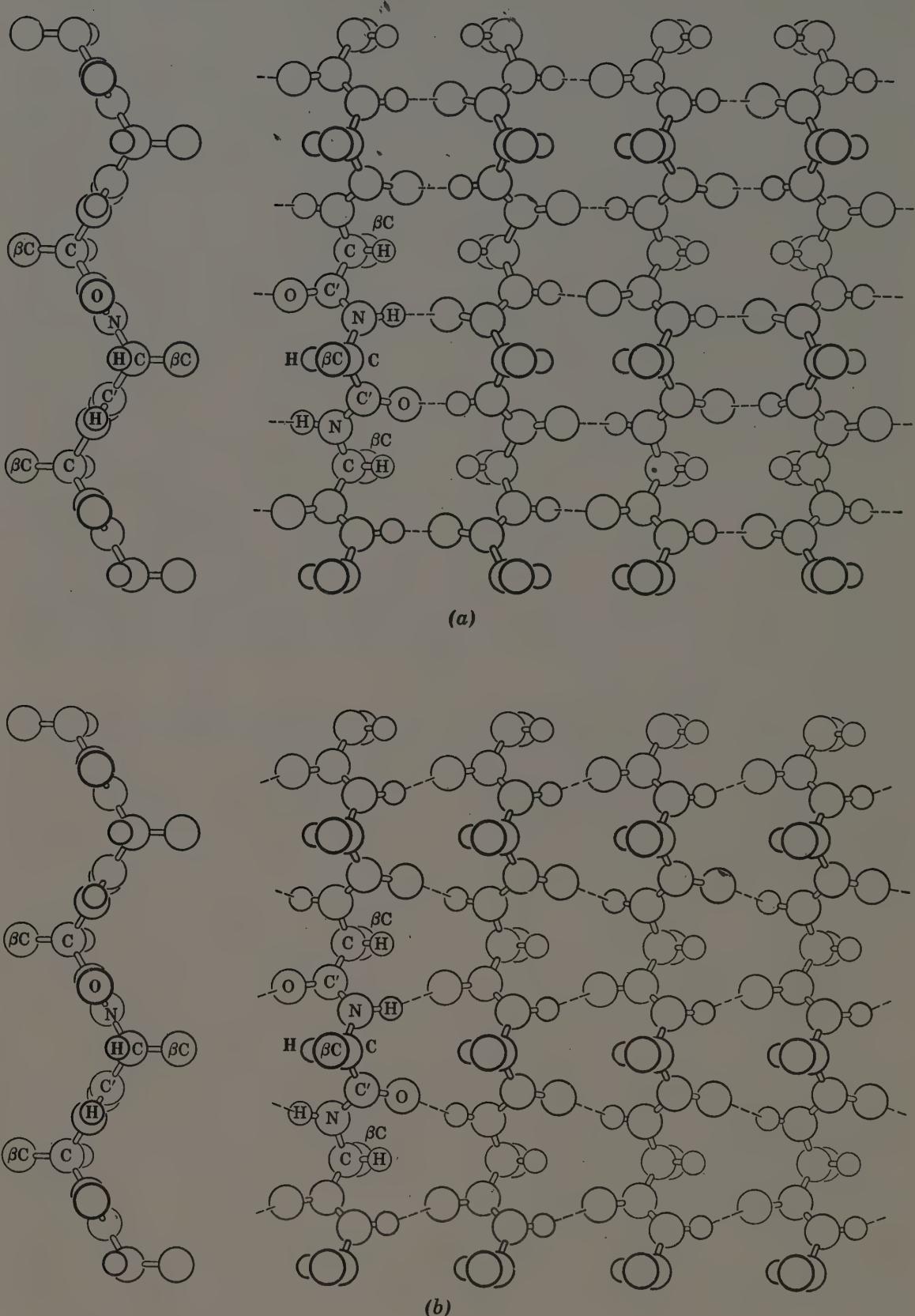


Fig. 16-3. The pleated-sheet structures: *a*, antiparallel-chain pleated sheet; *b*, parallel-chain pleated sheet. Reproduced through the courtesy of Professors Pauling and Corey.

The α -Fold. It now appears that a helical structure must be assumed for the α -keratin configuration. Indeed, it seems clear that the basic structure of extended polymeric molecules consisting of repeating units bonded in an identical manner must be generally helical. The β -keratin structure previously discussed is no exception, since it may be regarded as a special helical structure of twofold character, each successive unit being rotated by 180 degrees from its predecessor.

A helical configuration for the α -fold was suggested some years ago by Huggins (1943). In this model there were three amino acid residues per turn, alternate turns (that is, every third residue) being bonded together by intramolecular hydrogen bonds between N—H and C=O groups. A very similar configuration was suggested by Bragg (1949) for the chains in native globular proteins. It now appears, in light of the new structural considerations, that such a structure is inadequate.

Pauling and Corey have made an exhaustive survey of possible structures based on the new bond angles and bond distances, with the additional assumption that each nitrogen atom forms a hydrogen bond with a carbonyl oxygen atom of another residue, the distance between hydrogen-bonded nitrogen and oxygen being 2.72 Å., and the vector from nitrogen to oxygen being not more than 30 degrees from the N—H direction. The structure best fitting their requirements is the so-called α -helix (Fig. 16-4). In this helix, each NH group is hydrogen-bonded to the carbonyl oxygen of the third residue from it. There are 3.7 residues per turn of the helix (previous authors of helical models had assumed, for no apparent reason, that this number must be integral), and the pitch corresponds to 1.5 Å. per residue. The hydrogen bonds run approximately parallel to the fiber axis. Another helix suggested originally by these authors, having 5.1 residues per turn with hydrogen bonding

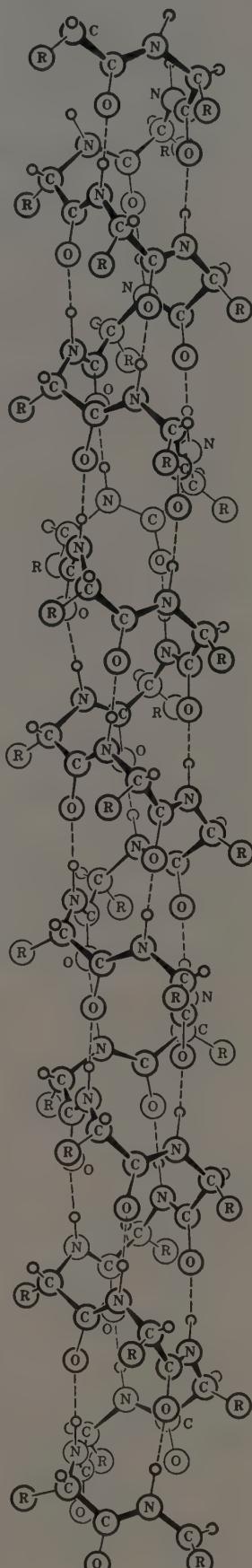


Fig. 16-4. The α -helix configuration. Reproduced through the courtesy of Professors Pauling and Corey.

between each residue and the fifth removed, appears inherently much less plausible. Other helical structures closely related to the α -helix have been suggested; the number of possibilities is limited only by the deviation one is willing to permit from the assumed values of hydrogen-bond distances and angles.

At the present time it cannot be stated with assurance that the α -helix is precisely the α -keratin configuration. There is considerable evidence, however, that it may be. It appears almost certain that it may be the configuration assumed by certain synthetic polypeptides of high molecular weight, such as polyglutamate esters.

Important advances have been made since 1945, both in the synthesis of amino acid polymers (p. 170) and in understanding of their physical properties. Particularly significant contributions have been made by workers at the Courtauld's Laboratories in England. These workers (for example, Bamford *et al.*, 1954) have shown that such polymers may be crystallized in either the α - or β - crystallographic forms, depending on the solvent system used. By means of studies of the absorption of polarized infrared radiation, it has been demonstrated that in the α -form the hydrogen bonds are oriented essentially parallel to the major axis of the peptide chain, whereas in the β -form they are predominantly perpendicular. These are precisely the orientations to be expected from the α -helix and pleated-sheet structures respectively (although admittedly various other structures would lead to the same qualitative result). Robinson and Bott (1951) have further demonstrated a difference in specific optical rotation in the two forms. Thus, two new techniques have been made available for distinguishing the two forms and for following transformations of one to the other. In this country it has been shown that poly- γ -benzyl-L-glutamate exists in a randomly coiled configuration in dichloroacetic acid (Doty *et al.*, 1954), and as a rigid rodlike molecule of dimensions corresponding to an α -helix in certain other organic solvents. A gradual transformation of the one form to the other with change in solvent, and with variation of temperature at constant solvent composition, has also been shown. Again, optical rotation was the most convenient criterion for configurational type (Doty and Yang, 1956).

Although these studies are of the greatest importance and can be expected to contribute much to our knowledge of protein structure, it must be remembered that the presence of various R-groups in actual proteins complicates the picture and may lead to some modification of the coiled structure. It should be noted that in the derivation of the α -helix no consideration was given to these side-chain groups other than the requirement that there be adequate room for them.

That a structure so derived could be correct for even synthetic poly-peptides seems remarkable.

GLOBULAR PROTEINS

In contrast to the highly insoluble fibrillar proteins are the globular proteins, soluble in water, salt, or alkali. These proteins usually occur naturally in aqueous solution, for example in plasma, egg white, and milk. Some, such as the seed globulins, may not occur in solution but are readily extractable when dissolved in aqueous media. They do not usually form fibers unless denatured. Their biological function is usually highly specific and in many cases still unknown. Because of their ready solubility, study of the physical chemistry of their solutions has proved attractive. They are probably simpler than the fibrillar proteins, in the sense that they are more homogeneous. On the other hand, their detailed structure is probably even more complex.

Molecular Weight. The size of the globular proteins was first studied quantitatively by the osmotic-pressure technique. Much of the early work is now subject to suspicion because the Donnan equilibrium effects (p. 210) were not considered and there was inadequate appreciation of the importance of making measurements at very low concentrations and carrying out appropriate extrapolations of the data to infinite dilution. Furthermore, such data give only an average molecular weight and leave the question of homogeneity entirely unanswered.

Knowledge of the structure of proteins was greatly advanced through Svedberg's development of the ultracentrifuge. Fortunately for protein chemistry, Svedberg chose to make proteins the first subjects of his investigations. The first important result which came out of this work was the demonstration that many protein solutions are relatively homogeneous in the ultracentrifugal field, that is, the solute sediments as a single component. As indicated in Chap. 13, this is not an absolute criterion of homogeneity but is certainly indicative. It now appears proper to speak of protein "molecules" rather than "colloidal particles." By combination of the ultracentrifuge and diffusion techniques Svedberg determined the molecular weight and "asymmetry" of a number of the more common proteins, and this work has been greatly extended by others.

Table 13-1 summarizes the results of such measurements on a number of proteins of interest. It will be seen that the proteins cover a wide range of molecular weights. The smallest listed there, namely ribonuclease, molecular weight about 13,000, probably represents about the lower limit. (The protamines have lower molecular

weights, but it seems likely that they should be considered as large polypeptides and not proteins.) Near the upper limit lie the hemocyanins, with molecular weights in the millions. Even larger are the viruses, which are considered by some to be giant protein molecules. The range might be indicated roughly as about 10,000 to 10 million, most of the proteins of interest having values between about 20,000 and 200,000.

Svedberg concluded that the molecular weights of proteins are not random but tend to cluster about certain values. The most probable mean figure appeared from his earlier work to be in the neighborhood of 35,000. Several proteins have values of about one half this figure, 17,000 to 18,000. Another common molecular weight range is twice 35,000 or 70,000. Svedberg termed these "molecular weight classes." More recent studies have indicated molecular weights scattered almost all over the range from 12,000 to 100,000. At least three well-defined proteins, ribonuclease, cytochrome C, and lysozyme, are now known to have molecular weights near 13,000. Ovalbumin and β -lactoglobulin do not fit any of the classes particularly well, since they have molecular weights of about 40,000, but both of these proteins are known to be electrophoretically inhomogeneous. The idea of molecular weight classes appears to be very much in the doubtful category at present.

Reversible Dissociation. S. P. L. Sørensen long ago postulated the concept of proteins as "reversible dissociation systems." According to this idea, the proteins of a system such as blood serum exist as a single component, and dissociate upon alteration of the conditions of pH, ionic strength, and so forth. On this basis, protein components isolated from such a system would be artifacts; the number and nature of the proteins derived would be a function of the conditions used. The recent development of new plasma fractionation techniques based on alcohol as a precipitant, resulting in components similar to those obtained by the classical salting-out procedures, casts doubt on the validity of these ideas.

Svedberg found definite evidence for reversible dissociation of certain protein molecules. These results were particularly striking in the case of the hemocyanins. With the hemocyanin of *Palinurus vulgaris* a single component is observed in the ultracentrifuge over the pH range 3.6 to 9.4. At higher pH a partial dissociation occurs, so that two sedimenting peaks are observed. The amount of the smaller, "slower sedimenting" component increases with pH, until at 10.8 the original component has disappeared. In other cases the dissociation pattern is more complex, but the tendency is always toward lower molecular weights at extremes of pH. These dissociations are

invariably reversible and are not accompanied by the changes in properties usually associated with denaturation. Svedberg and co-workers concluded that the molecular weights of all of the species in a given dissociation system are simple multiples of that of the smallest well-defined component.

More recent work has indicated that these dissociations are also promoted by ultraviolet light and ultrasonic waves. Divalent cations such as Ca^{++} and Mg^{++} tend to prevent dissociation or promote association.

It should be pointed out that the hemocyanins have extremely high molecular weight, so that the dissociation products are still very large, even for protein molecules. It is now amply verified that insulin, a relatively small protein, exists also as a reversibly dissociable molecule (p. 371).

Molecular Shape. Another very important result of Svedberg's work was the demonstration that the proteins of the globular class are indeed, as the name implies, very nearly spherical. In the third from the last column of Table 13-1 are given the frictional ratios f/f_0 , that is, the frictional coefficient of the molecule divided by that of an equivalent sphere. In general this ratio is seen to be not far different from unity. As shown in Chap. 13, deviation of this quotient from unity can be explained on the basis of nonspherical shape, solvation, or both. In the last two columns of the table are given calculated axial ratios, using two types of models, based on the assumption that the deviation of the frictional quotient from unity is due entirely to asymmetry. Actually some solvation must exist, but this assumption of no solvation places an upper limit on the possible asymmetry. For many proteins of interest this axial ratio does not exceed 4 or 5 for the prolate ellipsoidal (cigar-shaped) model. The outstanding exceptions are zein, the alcohol-soluble protein of corn endosperm, which is not a well-defined homogeneous protein and might be more properly considered under the fibrillar class, and fibrinogen, the precursor of fibrin, which is the structural component of the blood clot. The high asymmetry of fibrinogen is doubtless to be attributed to its special function, the formation of the fibrin clot.

The symmetry of the globular proteins finds confirmation in results obtained by other methods. The viscosity of protein solutions is very low as compared to that of other high polymers. This indicates a compact, symmetrical structure. Several proteins have been studied by the method of flow birefringence. In general these results have checked very satisfactorily with the dimensions predicted from other studies (Table 13-2).

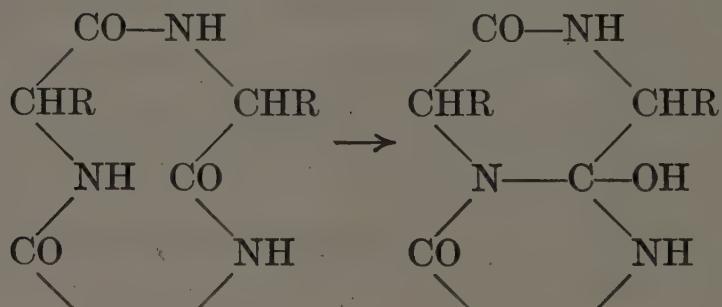
By knowing the molecular weight and the shape, it is of course possible to calculate the dimensions of a model molecule. Such calculations have been made for many proteins. The most significant thing about such calculations is the result that the shortest dimension always turns out to be much larger than the cross-sectional dimensions of a peptide chain. For example, all of the serum proteins which have been well studied by physicochemical techniques appear to have diameters, based on the prolate ellipsoidal structure, of about 40 Å.

From the foregoing it becomes apparent that the globular protein must be constructed of either (a) many relatively short (low molecular weight) polypeptide chains aggregated laterally, or (b) high molecular weight chains which are highly folded into a relatively compact, almost spherical body. The latter alternative appears more likely for several reasons. As will be seen in the following chapter, the unfolding of a highly folded structure offers an attractive explanation for protein denaturation. If the globular protein consisted of short chains, one would expect to find measurable quantities of terminal groups such as α -amino groups. By way of illustration, human serum albumin can be approximated as a prolate ellipsoid, of length 150 Å. and diameter 40 Å. Assuming β -keratin type chains spaced 3.5 Å. per residue, one sees that the chains could not exceed about 45 residues. If the structure is oblate rather than prolate, an even smaller number will be required. There would thus have to be at least 16 chains per molecule. Such a concentration of terminal groups should be easily detectable.

THE NATURE OF FOLDED STRUCTURE

Assuming a folded structure, it becomes of interest to inquire as to the nature of the folding. Many ideas have been proposed, but there is little experimental proof for any of them.

Cyclol Hypothesis. One type of folded structure, never widely accepted but of considerable historical interest, is the Wrinch cyclol structure. Wrinch (1937) suggested that in native proteins the amino acid residues are linked into rings through an N—C bond which may be considered as arising by a shift of the hydrogen from an NH group to a CO:



By means of this concept it is possible to build up many planar ring structures of high symmetry. The configuration of these rings is such that if all residues are L, the R-groups will all lie on the same side of the molecule. This provides a layer one amino acid residue thick, one surface being covered with R-groups. One of the strong arguments for this structure was the fact that it would explain the formation of monolayers when a protein is spread at an interface. However, it is now generally accepted that in the spread state peptide chains exist.

Such layers might stack up to yield essentially globular structures which could dissociate reversibly. Wrinch has given a more elaborate explanation of the globular proteins by postulating a folding of the cyclol sheets to yield cagelike polyhedra.

Certain difficulties in the cyclol structure make it very improbable. Particularly, it has been pointed out that there is not room for all of the R-groups to lie on one side of the rings. Also, estimates of the energy involved in the cyclopeptide conversion are not at all in accord with the values observed in denaturation (this calculation involves extrapolation from data on small molecules, which may or may not be justifiable). All of the experimental evidence for the cyclol structure appears to have been disproved.

Planar Folding. A folding similar to that proposed in the cyclol structure but involving hydrogen bonding between the CO and NH groups might not be subject to some of the criticisms of the cyclol structure. As one of the simplest postulates, one might assume extended chains running across the molecule and doubling back in a plane or pleated sheet, bonded laterally by interchain hydrogen bonds, as in Fig. 16-1. Coulombic forces would probably be of importance in holding this structure together; possibly also —S—S— bonds. (The considerations would in this case not be greatly different if the chains terminated at the boundary rather than doubling back.) It has been suggested that proline residues might serve the function of "hinges" at the folds. These residues do not fit smoothly into an ordinary peptide chain but tend to cause it to double back.

The fact that in the β -keratin chain, or pleated sheet, alternate R-groups project in opposite directions was pointed out in previous discussion. It has been suggested that polar R-groups might tend to lie on one side of the planes, nonpolar groups on the other. In this way one could easily understand the formation of layers or sheets with one polar and one nonpolar surface. The thickness of this layer would be just the side-chain spacing of Astbury, about 9 Å. This is about the thickness usually observed for protein monolayers in surface studies (Chap. 13).

It seems likely that if proteins were essentially sheets composed in this fashion, they would tend to combine in solution in such a way as to cover up their nonpolar surfaces. A minimal water-soluble native protein molecule, on this basis, would be analogous to a double layer (sandwich) as indicated in Fig. 16-5. It would have a thickness of about 15 to 20 Å. In cross section it might be a square, more probably round or elliptical. More layers might stack up. A four-layer structure could also exist, having two internal nonpolar interfaces

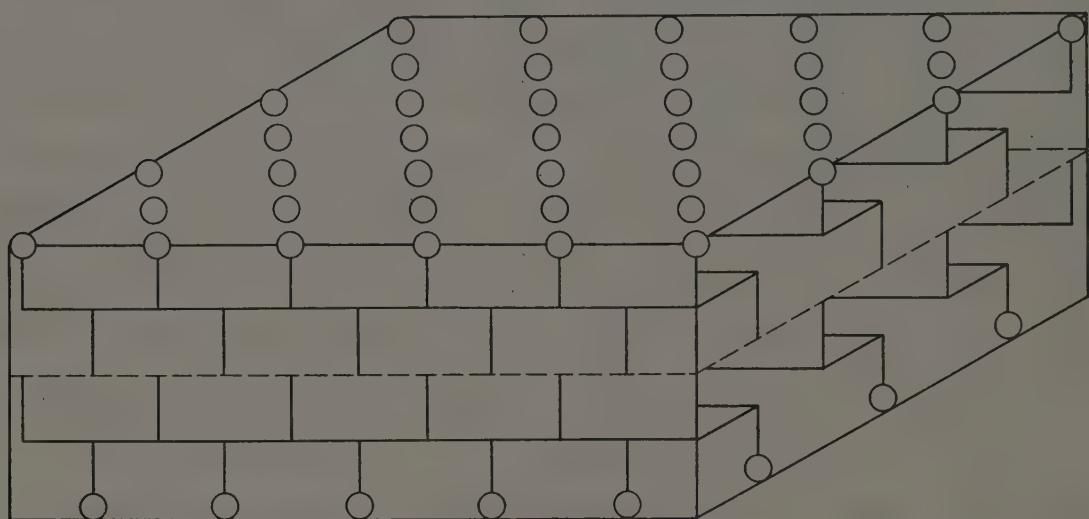


Fig. 16-5. Schematic representation of a hypothetical protein molecule consisting of two polypeptide monolayers based on a folded β -keratin configuration. Note that the alternation of polar R-groups (designated by the lines terminating with circles) with nonpolar R-groups provides an amphipathic character in the monolayers. The double layer illustrated might be considered a minimal water-soluble protein molecule, since it permits the nonpolar groups to be buried within the interior of the molecule. In other cases proteins might consist of four or some higher even number of monolayers.

and one polar interface. Both outer surfaces would be polar, so that the molecule would tend to be water-soluble. Dissociation at the polar interface would lead to a halving of the molecular weight but no other significant changes in properties (reversible dissociation).

On the basis of this type of structure one might expect most globular proteins to be approximated as flattened (disklike) ellipsoids. On the other hand, the physicochemical data for proteins have more often been interpreted on the basis of prolate models. Prolate structures could be accounted for by assuming many layers of relatively small cross section. It seems probable that in many if not most cases the physical data can be interpreted on the basis of oblate ellipsoids about as well as on the basis of prolate.

The α -Folded Structure in Globular Proteins. Bragg *et al.* (1950) have presented some evidence, based on X-ray diffraction results,

that the configuration in hemoglobin and myoglobin is actually helical, possibly the same configuration as in α -keratin. This idea appears to lack one of the more compelling features of the β -folded structure in that it is not apparent that it would provide for layers with predominantly polar and nonpolar surfaces. Such layers could arise through denaturation, however, in which case the denaturation process might be pictured as essentially an alpha-beta transformation.

Pauling and Corey suggested that their α -helix may be the basic configuration in at least some if not all globular proteins. Since this suggestion, a number of active crystallographic groups have been busily engaged in testing the hypothesis. Certain features of the X-ray diffraction patterns of certain globular proteins are in excellent agreement with this idea. There is, for example, a strong diffraction corresponding to a 1.5 Å. spacing which is found in synthetic polypeptides, various fibrillar proteins in the α -keratin configuration, and in hemoglobin. This diffraction is predicted by the α -helix but not accounted for by any other proposed configuration. Further, it has been concluded from intensive study of the diffraction patterns of hemoglobin that the molecule must contain rodlike units of dimensions compatible with those of an α -helix (p. 347). Studies of radial distribution of X-ray scattering intensity from amorphous proteins, a technique thought to yield characteristic intramolecular spacings without complication from intermolecular diffraction due to the crystal packing, are said to be in excellent accord with the α -helix. On the other hand, it has been claimed that in hemoglobin not over half of the protein can be coiled in a regular α -helix.

It is too soon to reach any definite conclusion as to the configuration of the polypeptide chains in the globular proteins. Through the diligent efforts of numerous X-ray diffractionists, together with the provocative model structures which have been proposed and the beautiful techniques for elucidation of residue sequence which are now available, complete structures may be worked out in the foreseeable future. Such a development could scarcely have been envisaged by any but the most optimistic scientist as recently as 1945.

SELECTED REFERENCES

Reviews

"The Structure of Fibrous Proteins," M. L. Huggins, *Chem. Revs.*, **32**, 195 (1943).

"X-rays and the Stoichiometry of the Proteins," W. T. Astbury, *Advances in Enzymol.*, **3**, 63 (1943).

"Protein Structure," H. B. Bull in Sahyun's *Outline of the Amino Acids and Proteins*, Reinhold, New York, 1944.

"Aspects of the Structure and Reactions of Proteins," John T. Edsall in Burk and Grummitt's *High Molecular Weight Organic Compounds*, Interscience, New York, 1949.

"The Structure of Collagen Fibrils," R. S. Bear, *Advances in Protein Chem.*, **7**, 69 (1952).

"Structure of Proteins" (a review of a symposium of the Royal Society), J. T. Edsall, *Nature*, **170**, 53 (1952).

"The Structure and Configuration of Amino Acids, Peptides and Proteins," B. W. Low in Neurath and Bailey's *The Proteins*, Vol. IA, p. 235, Academic Press, New York, 1953.

"The Crystalline Proteins: Recent X-Ray Studies and Structural Hypotheses," J. C. Kendrew, *Progr. Biophys. and Biophys. Chem.*, **4**, 244 (1954).

Original Articles

"Clupein. VIII," K. Felix and A. Mager, *Hoppe-Seyler's Z. physiol. Chem.*, **249** (1937).

"The Cyclol Hypothesis and the 'Globular Proteins,'" D. M. Wrinch, *Proc. Roy. Soc. London*, **A161**, 505 (1937).

"The Structure of Proteins," L. Pauling and C. Niemann, *J. Am. Chem. Soc.*, **61**, 1860 (1939).

"Molecular Weights, Unimolecular Layers and the General Structure of Proteins," G. T. Philippi, *Chem. Weekblad*, **36**, 266 (1939).

"The Structure of the Protein Molecule," D. G. Dervichian, *J. Chem. Phys.*, **11**, 236 (1943).

"The Empirical Formula of β -Lactoglobulin," E. Brand, L. J. Saidel, W. H. Goldwater, B. Kassell, and F. J. Ryan, *J. Am. Chem. Soc.*, **67**, 1524 (1945).

"An X-ray Study of Horse Methemoglobin," I. J. Boyes-Watson, E. Davidson, and M. F. Perutz, *Proc. Roy. Soc. London*, **A195**, 474 (1949).

"Giant Molecules," W. L. Bragg, *Nature*, **164**, 7 (1949).

"A Discussion of the Possibility of Bands of Energy Levels in Proteins," M. G. Evans and J. Gergely, *Biochim. et Biophys. Acta*, **3**, 188 (1949).

"Polypeptide Chain Configurations in Crystalline Proteins," L. Bragg, J. C. Kendrew, and M. F. Perutz, *Proc. Roy. Soc. London*, **A203**, 321 (1950).

"Infrared Spectra of Films of Native and Denatured Pepsin," L. L. Uzmann and E. R. Blout, *Nature*, **166**, 862 (1950).

"Evidence of Chain Folding of Polypeptides and Proteins," A. Elliott and E. Ambrose, *Discussions Faraday Soc.*, **No. 9**, 246 (1950).

"Observations of Proteins in Polarized Ultraviolet Light," M. Perutz, M. Jope, and R. Barer, *Discussions Faraday Soc.*, **No. 9**, 423 (1950).

"Salt Linkages in Proteins," C. Jacobsen and K. Linderstrom-Lang, *Nature*, **164**, 411 (1950).

"The 1.5A Reflection from Proteins and Polypeptides," M. F. Perutz, *Nature*, **168**, 653 (1951).

"Configurations of Polypeptide Chains with Favored Orientations around Single Bonds. Two New Pleated Sheets," L. Pauling and R. Corey, *Proc. Natl. Acad. Sci. U. S.*, **37**, 729 (1951).

"The Structure of Proteins: The Hydrogen-Bonded Helical Configurations of

the Polypeptide Chain," L. Pauling, R. B. Corey, and H. R. Branson, *Proc. Natl. Acad. Sci., U. S.*, **37**, 205 (1951).

"Some Aspects of the Chemical Statics and Dynamics of Proteins in Solution," E. A. Moelwyn-Hughes, *Proc. Roy. Soc. London*, **A205**, 60 (1951).

"The Structure of Synthetic Polypeptides. I. X-ray Investigation," C. H. Bamford, W. E. Hanby, and F. Happey, *Proc. Roy. Soc. London*, **A205**, 30 (1951).

"The Structure of Synthetic Polypeptides. II. Investigation with Polarized Infra-red Spectroscopy," E. J. Ambrose and A. Elliott, *Proc. Roy. Soc. London*, **A205**, 47 (1951).

"The Molecular Weight of Lysozyme after Reduction and Alkylation of the Disulfide Bonds," H. Fraenkel-Conrat and A. Mohammed, *J. Am. Chem. Soc.*, **73**, 625 (1951).

"Near Infrared Spectra of Compounds with Two Peptide Bonds and the Configuration of a Polypeptide Chain," S. Mizushima *et al.*, *J. Am. Chem. Soc.*, **73**, 1330 (1951).

"Optical Rotation and Chain Folding in Synthetic Polypeptides and Gelatin," C. Robinson and M. Bott, *Nature*, **168**, 325 (1951).

"X-Ray Evidence on the Molecular Structure of Globular Proteins," D. Riley and U. Arndt, *Nature*, **169**, 138 (1952).

"Structure of the Synthetic Polypeptide Poly- γ -Methyl-L-Glutamate," H. Yakel, L. Pauling, and R. Corey, *Nature*, **169**, 920 (1952).

"Structure of the Synthetic α -Polypeptides," F. H. C. Crick, *Science Progr.*, **42**, 205 (1954).

"The Configuration of Polymers of γ -Benzyl-L-Glutamate in Solution," P. Doty A. Holtzer, J. Bradbury, and E. Blout, *J. Am. Chem. Soc.*, **76**, 4493 (1954).

"Alpha and Beta Forms of Poly-L-Alanine," C. Bamford, A. Elliott, W. Hanby, and I. Trotter, *Nature*, **173**, 27 (1954).

"Structure of Poly-L-Proline," P. Cowan and S. McGavin, *Nature*, **176**, 501 (1955).

"Structure of Polyglycine," C. Bamford, L. Brown, E. Cant, E. Elliott, W. Hanby, and B. Malcolm, *Nature*, **176**, 396 (1955).

"Some Possible Modifications of the α -Helix Configuration Involving Changes of Direction and Sense of the Helix," H. Lindley, *Biochim. et Biophys. Acta*, **18**, 194 (1955).

"An Investigation of the Structure of Silk Fibroin," R. Marsh, R. Corey, and L. Pauling, *Arch. Biochem. and Biophys.*, **16**, 1 (1955).

"Optical Rotation and Polypeptide Chain Configuration in Proteins," C. Cohen, *Nature*, **175**, 129 (1955).

"The Structure of Collagen," A. Rich and F. Crick, *Nature*, **176**, 915 (1955).

"Speculations on Gene Action and Protein Specificity," D. Schwartz, *Proc. Natl. Acad. Sci. U. S.*, **41**, 300 (1955). Suggests a periodicity concept based on amino acid type.

"An Investigation of Insulin Structure by Model Building Techniques," H. Lindley and J. Rollett, *Biochim. et Biophys. Acta*, **18**, 183 (1955).

"The pH Dependence of the Deuterium Exchange of Insulin," K. Linderstrom-Lang, *Biochim. et Biophys. Acta*, **18**, 308 (1955).

"Poly- γ -Benzyl-L-Glutamate: The Helix-Coil Transition in Solution," P. Doty and J. T. Yang, *J. Am. Chem. Soc.*, **78**, 498 (1956).

Denaturation of Proteins

Manifestations of protein denaturation

Causative agents

Nature of the process

One of the most characteristic properties of proteins is their propensity to undergo gross changes through the agency of various influences, many of which are quite mild by the chemist's standards. Perhaps the most notable example is the coagulation of a protein solution, such as the white of an egg, upon heating. While not all proteins are heat-coagulable, most are.

Changes in solubility can also be brought about at ordinary temperatures in many cases by addition of certain chemicals. Concentrated aqueous urea solution (typically 4 to 8 M) is an excellent reagent for producing such changes. If egg albumin is dissolved in aqueous urea solution, no apparent change takes place; but upon removal of the urea, for example by dialysis, the protein coagulates. This indicates that the urea has induced some change in the protein that renders it insoluble in water at pH values near the isoelectric point, although it remains soluble in the aqueous urea solution. In other words, it is possible for an alteration in properties to take place without the visible manifestation of coagulation. It is now generally assumed that such a change always precedes protein coagulation, and the all-inclusive term used to cover such changes is "denaturation."

An adequate definition of the term denaturation appears impossible at the present time. The literal meaning, "change from the natural state," is certainly too broad to be helpful. Perhaps as useful a

working definition as any has been given by Neurath and co-authors (1944). According to these authors, denaturation is "any nonproteolytic modification of the unique structure of a native protein, giving rise to definite changes in chemical, physical, or biological properties." This definition serves to eliminate from consideration peptide hydrolysis; otherwise it is still all-inclusive.

Some of the known facts and ideas with regard to protein denaturation will be presented in this chapter. It should be emphasized that the subject of denaturation cannot be divorced from the considerations of protein structure treated in the preceding chapter. An adequate understanding of protein structure demands some knowledge of denaturative phenomena. Conversely, a discussion of denaturation is almost impossible without some insight into the structure of proteins.

MANIFESTATIONS OF PROTEIN DENATURATION

Chemical Evidence. Chemically, denaturation results in an enhancement in reactivity of various chemical groups over that in the native protein. On the assumption that proteins consist essentially of amino acids linked in the typical peptide linkage through their α -amino and α -carboxyl groups, it would be expected that the side chains would react chemically, at least qualitatively, as in the free amino acids. This is found to be true in the case of simple peptides; indeed, the reactivity of the side-chain groups may actually be greater than that in the free amino acid. Native proteins, however, are usually found to have side-chain groups which are not reactive chemically until the protein is either hydrolyzed or denatured.

Of the various side-chain groupings present in proteins, the sulfhydryl ($-\text{SH}$) and disulfide ($-\text{S}-\text{S}-$) groups have been studied more extensively than the others. For the $-\text{SH}$ group, color tests are available which depend on its reducing ability. The disulfide group can be determined by testing for $-\text{SH}$ reactivity following reduction, for example with thioglycolic acid, which converts $-\text{S}-\text{S}-$ to $-\text{SH}$. Such tests can be made qualitatively with nitroprusside or other reagents, or quantitatively with various oxidants, of which porphyrindin, a blue dye, is perhaps preferred. Some proteins, such as ovalbumin and serum albumin, give no test for $-\text{SH}$ in the native state but do respond following denaturation. Furthermore, these groups are reactive in the presence of agents like urea and guanidine hydrochloride. Indeed, in 5 M guanidine hydrochloride the $-\text{SH}$ groups in many proteins may be titrated quantitatively; that is, all the $-\text{SH}$ groups known to be present from amino acid analysis are reactive. Similar considerations hold in the case of $-\text{S}-\text{S}-$.

Some workers formerly held to the idea that denaturation involved a reduction of —S—S— to —SH. Such a view is no longer tenable. It would rather appear that both of these groups are merely exposed on denaturation.

Although the evidence is much less complete, it also appears that phenolic groups (tyrosine) and indole groups (tryptophan) are rendered more available upon denaturation. Evidence for the liberation of the phenolic groups is based in part on titration studies and especially on spectral studies. There are indications that in ovalbumin, for example, the phenolic groups cannot be fully titrated without the molecule undergoing irreversible denaturation.

This point raises the question of the effect of denaturation on the amphoteric properties of proteins. To a first approximation, it can be stated that the titration curves of native and denatured proteins are remarkably similar. It must be noted, however, that this similarity of titration curves is only a first approximation, and several modifying remarks seem in order. In the first place, certain groups, such as the sulfhydryl and phenolic residues, may be present in such small proportion relative to the carboxyl and amino groups that marked alterations in their availability may be scarcely noticeable in the over-all titration curve. Second, the titration curves of denatured proteins cannot usually be measured over the full *pH* range because of insolubility of the protein in the isoelectric region. Third, the titration studies on carbonyl- and ferrihemoglobin, discussed in Chap. 11, emphasize the possibility that in many cases it may not be possible to obtain the complete titration curve of a "native" protein. In those cases, it seems clear that reversible denaturation ensues at low (and possibly also high) *pH*, so that the usual titration curve must perforce be that of denatured protein at extreme *pH* values. Thus, many of the dissociable groups may in reality be unavailable in the native molecule, but they may be exposed readily through reversible alterations which effect the same result as irreversible denaturation.

Curves of electrophoretic mobility versus *pH* for denatured and native proteins are also in general nearly identical, but there are several demonstrations of shifts in the isoelectric point in the alkaline direction upon denaturation. In one or two instances it is claimed that denatured and native protein can be separated electrophoretically. Clearly, the same considerations apply to electrophoretic mobility as to titration curves, since both are manifestations of the ionic equilibria in which the protein is involved. In electrophoretic behavior there is the added complication of the effect of alterations in shape or hydra-

tion of the protein following denaturation, a question that will be considered later.

The peptide linkage also is rendered more reactive by denaturation. This is particularly true with regard to enzymic degradation, the rate of which may be many times greater for the denatured protein than for the native. It has been proposed, and there is some substantiating evidence for the view, that the initial step in the degradation of native proteins by an enzyme such as papain may be denaturation, in the sense of a change, other than proteolytic, to a more reactive form.

Needless to say, there are various possible explanations for the liberation of chemical groupings in proteins and numerous hypotheses have been proposed. Since many different types of groups are exposed (including the peptide linkage itself) under such mild conditions, denaturation cannot be a chemical reaction in the usual sense of the organic chemist. The essential feature of the process must be an opening up of a highly folded structure, and any bonds broken are probably of the relatively weak, so-called physical type rather than covalent. The end product might be the completely unfolded peptide chain, or possibly a loose, randomly coiled chain in which the reactive groups would be entirely exposed. Much more evidence for this general picture of denaturation will be seen in later sections of this chapter.

Changes in Biological Properties. With proteins possessing demonstrable biological activity, denaturation is usually accompanied by loss of such activity. For example, enzymes are typically inactivated at elevated temperatures, usually irreversibly. This inactivation can frequently be correlated with other criteria of denaturation, such as loss of water solubility and exposure of reactive groups. Thus, urease is inactivated by guanidine, and —SH groups are liberated. On the other hand, pepsin can be dissolved in aqueous urea without loss of activity and with no liberation of —SH groups.

The serological specificity of proteins is also generally altered by denaturation. In this connection it is necessary to consider separately the behavior of antibodies and protein antigens (Chap. 23). Antibodies are in many instances subject to loss of activity, that is, of ability to precipitate with their specific antigens, upon heating and upon treatment with such denaturants as urea and guanidine salts, but there are reported instances where loss of activity does not result from such treatment. Protein antigens also present confusing responses. Differences in antibody-combining ability in ovalbumin denatured in various ways have been demonstrated. But ovalbumin spread in the form of monolayers of thickness only 9 A., under which

conditions the molecule should be drastically modified, has been shown to retain the ability to combine with its specific antibody.

Tobacco mosaic virus is rendered noninfective by aqueous urea or by guanidine hydrochloride without, however, undergoing denaturation as judged by decrease in solubility.

It is thus seen that denaturation in general leads to modification of biological activity, but that the correlation is not perfect. Whether or not inactivation of an enzyme, antibody, or other biologically active protein results upon denaturation must depend in a given case on the manner in which denaturation is effected, in other words on the extent of structural alteration. Furthermore, proteins must differ as to the origin of their biological activity. In some cases the seat of the activity may be a prosthetic group which is little affected by the precise state of the protein moiety. In most cases, however, these properties are probably dependent on a highly specific structural pattern resulting from the definite folding of the peptide chains, in which case denaturation leads to inactivation.

Changes in Physical Properties. The earliest recognized alteration in proteins as the result of denaturation was the change in solubility mentioned earlier. Perhaps the most significant feature of such changes is that they are always in the direction of decreased solubility in water at the isoelectric point; never is the solubility increased. It would be appropriate to inquire at this point how such changes in properties might be explained.

Synthetic polypeptides of moderately high molecular weight have been found in general to be rather insoluble in water. It would be expected that further increase in molecular weight into the protein range would result in a further diminution in solubility. The problem becomes that of explaining the water solubility of native proteins, not the insolubility of denatured ones. The most satisfactory explanation yet offered is that in the native protein the hydrophobic groups are buried, perhaps folded into the interior of the molecule. Those groups having a strong affinity for water, and hence a solubilizing effect, must be predominantly on the surface of the molecule. This implies a more or less definite folding of the peptide chain, and loss of this definite folding, such as has already been postulated for denaturation, would be expected to lead to exposure of hydrophobic groups and to a reduction in solubility.

Another significant consequence of protein denaturation is a loss in crystallizability. By crystallizability is here meant the ability to form true three-dimensional molecular crystals, not the fibrillar-type

of orientation (see the following paragraph). Many proteins can be crystallized in the native state, but the authors know of no case in which a protein shown to be denatured by any other property has been crystallized. It would appear that crystallizability might be one of the most sensitive criteria for absence of denaturation.

Denatured proteins in the solid state yield X-ray diffraction patterns analogous to the contracted α -keratin pattern, and on stretching yield the typical β -type fibrillar pattern. It seems probable that denaturation must take place before fibers can be formed from the globular-type proteins. These results are in good accord with the unfolding concept, and are in fact to be expected on that basis.

As to the effect of denaturation on the size of the protein kinetic unit, the "molecular weight," here again there is much apparent conflict in the evidence. In the first place, not all types of denaturation can be studied, since it is necessary that the proteins be retained in solution for study. By osmotic-pressure measurements, urea denaturation appears to cause no change in the molecular weight of serum albumin or ovalbumin, or of hog or sheep hemoglobin, while horse and ox hemoglobin appear to be split into two units. Another possible approach to this problem lies in studying the effect of concentration on the equilibrium (where attainable) between native and denatured forms. This approach is based on application of the law of mass action. In this way it has been concluded that the denaturation of hemoglobin and trypsin by salicylate is not accompanied by change in size. Probably no rule can be laid down, other than to say that where a change in molecular weight results, it is in the direction of a reduction or a degradation. Aggregation of the denatured units may take place, but this should more properly be considered as the second phase in the coagulation process. In individual cases, whether or not reduction in molecular weight results will depend on the protein studied and on the manner in which denaturation is induced.

There are numerous reports of an increase in intrinsic viscosity of proteins accompanying denaturation. An increase in intrinsic viscosity implies an increase in effective (hydrodynamic) solute volume, which may result either from an increase in molecular asymmetry or in degree of hydration, or both (Chap. 13). Interpretation of the results is difficult, in fact impossible, without independent measurements of other properties. The change is at least in accord with the concept of an unfolding. Combined studies by sedimentation velocity and diffusion have been made in a few cases, the results again indicating an increase in frictional coefficient, implying an increase in molecular

asymmetry. Again, however, the results are complicated by the question of hydration, and the measurements are tedious and time-consuming.

Table 17-1 summarizes some of the more pertinent data on changes in asymmetry of proteins upon denaturation, as determined by such

TABLE 17-1. Changes in Shape Accompanying Protein Denaturation¹

Protein	Denaturing Agent	Ratio Long Axis: Short Axis ²		
		Native	Denatured	Calculated ³
Serum albumin, horse	1.5 M Urea	3.3	4.1	f/f_0
	3 M Urea		5.3	f/f_0
	6 M Urea		13.3	f/f_0
	8 M Urea		13.3	η
	8 M Guanadine hydrochloride		16.7	η
Serum globulin, horse	5 M Urea	5.2	9.3	η
	8 M Urea		15.4	η
	5.6 M Guanidine hydrochloride		15.0	η
Myogen	Urea	3.0	49	f/f_0
Egg albumin	Heat	3.1	5.5	η
	Urea		7.1	η
Hemoglobin	Acids	2.7 ⁴	6.4 ⁴	f/f_0
Helix pomatia hemocyanin ⁵	pH	2.7 ⁶	8.3 ⁷	f/f_0
			11.0 ⁸	f/f_0

¹ From H. Neurath, J. P. Greenstein, F. W. Putnam, and J. O. Erickson, *Chem. Revs.*, **34**, 157 (1944).

² Values have been calculated for prolate ellipsoids of revolution, assuming 33 per cent hydration.

³ f/f_0 = calculation from the dissymmetry constant with the Perrin equation.
 η = calculation from viscosity data with the Simha equation.

⁴ Oblate ellipsoid of revolution.

⁵ Asymmetries calculated from f/f_0 and double refraction of flow.

⁶ $M = 8.9 \times 10^6$.

⁷ $M = 4.31 \times 10^6$.

⁸ $M = 1.03 \times 10^6$.

methods. It is to be emphasized again that while the measured alterations in either intrinsic viscosity or frictional coefficient are real, the interpretation in terms of elongation of the molecule (increase in axial ratio) is highly provisional. The calculations in this table are based on the assumption of a constant degree of hydration of 33 per cent. The results could equally well be interpreted on the basis of a

constant axial ratio, with an increase in hydration upon denaturation. There is some reason to suspect that the latter assumption may be nearer the truth in the case of serum albumin in urea. The change upon denaturation in this instance might be pictured as an expansion or loosening of the molecule, with the immobilization of large quantities of water within the structure.

Additional evidence for a gross change in molecular size or shape upon denaturation is obtained from streaming birefringence measurements. Ovalbumin in the native state yields no streaming birefringence, but upon denaturation by any of a variety of means, the solutions become strongly birefringent at high shear rates. The rotary diffusion constants obtained under conditions yielding minimum aggregation are similar, whether denaturation is induced by heat, urea, or surface-active ions. Assuming the molecule to be a prolate ellipsoid, the molecular lengths calculated are of the order 600 Å. In principle these results could also be interpreted on the basis of a strong increase in hydration with little increase in molecular axial ratio, but certain facts suggest that the first interpretation may be substantially the correct one. The decrease in rotary diffusion constant with increasing degree of aggregation is rather minor, a result that is best explained on the basis of parallel aggregation of long rods. (Aggregation of nearly spherical units would cause this constant to decrease markedly with aggregation.) Also, the degree of hydration necessary to explain the low rotary diffusion constants obtained on denaturation appears to be much larger than could be accounted for by the increase in intrinsic viscosity observed. This argument is weakened by the fact that, as yet, intrinsic viscosity and streaming birefringence measurements have not been made in the same solvent systems.

For serum albumin, the streaming birefringence results are quite different, little decrease in rotary diffusion constant taking place upon denaturation. This is in accord with the earlier suggestion that in this case denaturation involves primarily an expansion of the molecule rather than an elongation. Clearly, the present status of the question of the effect of denaturation upon the shape and configuration of protein molecules leaves much to be desired. One major problem in these studies is the tendency of denatured proteins to aggregate. This may entirely mask changes in molecular properties, thus leading to erroneous conclusions.

Finally, mention should be made of the effect of denaturation upon the specific optical rotation of globular proteins. There is now a large body of evidence to indicate that the specific rotation always increases in absolute magnitude upon denaturation. The rotation of

proteins is in all reported cases negative or levo; hence, the value becomes more negative upon denaturation. Fortunately, the optical rotation of most globular proteins is not highly dependent upon such conditions as *pH* or ionic strength. The changes observed, then, must manifest a change in polypeptide configuration. The effect is large, the native proteins normally having rotations in the range —30 to —60 degrees, the denatured proteins around —100 degrees. The study of this phenomenon deserves much more attention than it has received in the past. Studies of rotatory dispersion (dependence of optical rotation on wave length of the light) promise to be even more illuminating.

CAUSATIVE AGENTS

Doubtless the most important factor governing the rate of denaturation is the temperature. Whereas the rate of most chemical reactions is increased by a factor of two or three for a rise in temperature of 10° , the rate of denaturation may increase by as much as 600-fold. So marked is the effect of temperature, it was long thought that for a given protein there actually exists a threshold temperature below which denaturation will not take place. This view is no longer tenable; rather, it appears that the reaction is merely one possessing an unusually high temperature coefficient. The implications of this fact will be discussed in the next section.

Also apparently essential for the denaturation process is water. In the anhydrous condition proteins are fairly stable. (It should be remembered that most so-called dry protein preparations may contain at least 5 to 10 per cent moisture, enough to permit denaturation. Removal of the last water may not be possible without denaturation resulting in the process.) Presumably water is essential to provide mobility, and possibly also to aid in the rupture of cross-linkages of the hydrogen-bond type between peptide folds.

The *pH* of the solution is an important factor governing denaturation rate. For most proteins there appears to exist a *pH* range of stability outside which it is not safe to operate if denaturation is to be avoided. Extremes of *pH*, either high or low, tend to throw a large net charge on the protein, according to considerations presented in Chap. 11. This high net charge would be expected to lead to repulsions between various parts of the molecule, and might easily lead to expansion (unfolding) of the structure.

Chemical Agents. Many chemical reagents are known to accelerate denaturation. Urea, in high concentrations (typically 6 *M*) in aqueous solution, is one of the best known and most used. Certain guanidine

salts are even more active than urea, and in more dilute solution. Interestingly, different guanidine salts show varying degrees of activity. For example, the bromide, iodide, and thiocyanate salts are active at lower concentrations than the chloride, insofar as sulphydryl liberation is concerned; the sulfate, carbonate, and acetate are apparently inactive. Table 17-2 compares the ability of urea, guanidine hydrochloride, and some of their derivatives to liberate —SH groups in several proteins. Sodium salicylate, acetamide, and formamide are other often-used denaturing agents. Even some inorganic ions, such as iodide, have a distinct denaturing action. The synthetic detergents have been shown to be extremely active as denaturing agents; from the point of view of effective concentration they are the most active chemical agents known in this regard, with the possible exception of H⁺ and OH⁻ ions.

Organic solvents such as alcohol and acetone in appreciable concentrations in water are apt to cause denaturation. It was long felt that such solvents could not be used at all in protein fractionation, but it is possible to use such precipitants, provided one works at low temperatures (Chap. 15).

No very definite picture can yet be presented as to how these varied agents act. It seems probable that urea functions by virtue of its affinity (hydrogen bonding) for the peptide linkage; it might tend to open up the structure by a "solubilizing action" on this grouping. The other amides probably operate similarly. The ionic agents might exert their effect by upsetting the delicate charge balance of the protein molecule, since there is now much evidence for the binding of anions by proteins, especially the soaps and synthetic detergents. Possibly there is also a tendency for detergents as well as organic solvents to exert a solubilizing action on inner hydrophobic layers. Such considerations are little more than speculation at present.

Physical Agents. The most important of all physical agents, heat, has been discussed. Spreading of proteins at an interface usually results in denaturation. Studies of such surface films indicate that the protein molecules must be unfolded, at least in part, in the process of spreading. From the practical standpoint it is important that excessive foaming of protein solutions be avoided if denaturation is not desired.

Ultraviolet irradiation has been found to cause a decrease in the solubility of proteins and inactivation of certain enzymes. The mechanism is not understood, but some rupture of peptide bonds may be involved. High pressures, of the order of several thousand kilo-

TABLE 17-2. Effect of Urea, Guanidine Hydrochloride, and Their Derivatives on Various Proteins¹

Agent	Concen- tration ²	Protein Sulphydryl as Cysteine						Urease %
		Egg Albumin %	Globin %	Serum Albumin %	TMV %	Edes- tin %	Insu- lin %	
Urea	10	1.00	0.19	0	0.70	0.34	0.07	0.64
N-Methylurea	10	1.02	0		0	0	0	0
O-Methylisourea hydrochloride	10	1.05	0		0.17		0	
Guanidine hydro- chloride	6	1.28	0.56	0.34	0.76	0.51	0.18	2.84
Methylguanidine hydrochloride	6	1.19	0.56			0.16	0.07	1.16
as-Dimethylguanidine hydrochloride	6	0.75	0			0	0	0

¹ From H. Neurath, J. P. Greenstein, F. W. Putnam, and J. O. Erickson, *Chem. Revs.*, **34**, 157 (1944).² Millimoles of agent per cc.

grams per square centimeter, can cause denaturation as judged by liberation of —SH groups and inactivation.

Ultrasonic waves are potent in producing denaturation. This may result largely from the high pressures and temperatures which are presumed to exist transiently in localized regions of the solution. There are indications, however, that more fundamental mechanisms are also involved.

Finally, the influence of the protein concentration should be mentioned. It has long been observed that proteins are often more stable at higher concentrations than when very dilute. In part this might be due to the importance of surface catalysis; in a very dilute solution relatively more of the protein would be in the surface and (presumably) denatured.

NATURE OF THE PROCESS

From the foregoing discussion it is to be concluded that protein denaturation is a transition of the physical kind, not a chemical reaction involving covalent bonds. It seems clear that fundamentally it involves an alteration in the folded structure of the peptide chains of the protein molecule, together with an associated alteration in the spatial relations of the various amino acid residues and their interactions with one another and with solvent. To further elucidate the character of the changes involved, two approaches offer particular promise, namely thermodynamics and especially kinetics. The status of knowledge in these two areas will now be surveyed.

Thermodynamics of Protein Denaturation. Classical thermodynamics may be considered as dealing only with differences between end states, and not with the mechanism whereby a system passes from the one state to the other. It is thus capable of telling us something of the difference between the native and the denatured state, but not how the transition takes place. There is the further limitation that the methods of thermodynamics, at the present stage of development, are in the main applicable only to reversible processes. This immediately raises the question of the extent to which protein denaturation is a reversible process, a question which has received much attention and discussion.

It would appear that denaturation, as manifested by a gross change in solubility, liberation of sulphydryl groups, or loss of crystallizability, is not reversible. This result is perhaps to be anticipated if drastic unfolding of the peptide chains takes place. It seems doubtful, on an intuitive basis, that the various residues could ever find their way back into precisely the same positions they occupied originally. A

few proteins, however, have been shown to regain their original solubility and even their biological properties after having lost them; trypsin, hemoglobin, and lactoglobulin can be cited as important examples. That the regenerated "native" protein is identical in every respect with the original can obviously not be proved. It has been said that full reversal of denaturation is akin to unscrambling an egg.

More subtle alterations which are reversible are possible in some proteins. Serum albumin undergoes drastic changes in specific rotation and intrinsic viscosity upon lowering the *pH*, and these alterations may be readily reversed. In concentrated urea this protein exhibits a similar reversible change. The possibility exists that many proteins may undergo such reversible alterations. Reversibility doubtless depends on how far the alteration process has proceeded. This brings up the question as to whether there are steps or stages in denaturation. For years denaturation was termed an "all or none" phenomenon, implying that there could be no intermediate stages. In a sense this dictum may still be tenable, in that even the slightest alteration, the first stage in the chain of events, results in denaturation. Nevertheless, it now seems probable that there must be graded stages between the definite native state and the fully unfolded primitive peptide chain. How far the protein passes along this path in a given instance must depend on the protein and the conditions. The farther down the path, the less likely reversal would be.

Where denaturation is reversible, it is possible in principle to determine an equilibrium constant *K* for the reaction. From this, the standard free energy change ΔF° associated with the process is calculable from the relation

$$-\Delta F^\circ = RT \ln K \quad (1)$$

From the dependence of the equilibrium constant on temperature, or from calorimetric studies, the heat or enthalpy change ΔH° of the process can be determined. Combination of these through the fundamental thermodynamic relation

$$\Delta H^\circ = \Delta F^\circ + T\Delta S^\circ \quad (2)$$

yields the entropy change. The ΔH° of the process may be considered crudely as measuring the extent of net rupture or formation of bonds in the process. The entropy change measures the change in "thermodynamic probability," a concept which is more difficult to picture. In protein denaturation, the very limited data available indicate that both of these quantities are large and positive. The most plausible explanation would seem to be that the process involves the rupture

of many relatively weak bonds or attractions, most probably hydrogen bonds, with a consequent increase in the configurational freedom of the peptide chains. It must be emphasized, however, that alternative explanations exist, and a final solution to the problem will almost certainly require much information in addition to that given by thermodynamics.

Kinetics of Protein Denaturation. The science of chemical kinetics is capable of yielding information on the actual mechanism whereby a process proceeds. This science is concerned with the quantitative effect of all conceivable variables on the rate at which a reaction takes place.

The key to precise kinetic analysis of a reaction is the method or methods employed to judge the extent of reaction. In the present case, the most used criterion has been the formation of precipitate or the formation of material which is precipitable under standard conditions. The conditions employed for such precipitation are customarily the addition of a suitable buffer of pH near the isoelectric point, usually together with a precipitating salt such as magnesium or ammonium sulfate. Loss of biological activity is a commonly used criterion in the cases of enzymes and antibodies. Properties such as intrinsic viscosity, optical rotation, and liberation of titratable sulfhydryl groups are also useful, but have not been so widely employed. In general, because of the complex nature of protein denaturation and the possibility that various steps or degrees of denaturation may be involved, it would seem most desirable to employ not one but two or several independent criteria in kinetic studies. Where such experiments have been carried out, the concordance of results is not as good as is to be desired.

The first kinetic property of interest is the "order" of the reaction. The order of a reaction with respect to any given reactant is the exponent n to which the concentration of that reactant enters in the rate expression

$$\text{Rate} = kC^n$$

The *total* order is the sum of all such exponents where all conceivable constituents involved in the reaction are included. The rate of protein denaturation is generally found to be approximately proportional to the first power of the protein concentration, that is, the reaction is first order in protein. The results, however, do not agree precisely with the first-order law, and in some cases the deviations are rather serious. Such complications are perhaps not surprising in view of the complexity of proteins and of the reaction. Heterogeneity of the

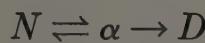
protein sample and the existence of various steps in the denaturation process are the two most likely explanations of the observed deviations.

The importance of this question of reaction order is that if the process is truly first order in protein, it may be inferred that it is an intramolecular process; that is, a given protein molecule does not have to combine or interact with another of its kind in order to undergo reaction. It must be remembered, however, that this does not rule out combination with other constituents of the medium, such as hydrogen or other ions or the solvent (water) itself. The pronounced pH dependence of protein denaturation emphasizes that hydrogen (or hydroxyl) ions cannot be ignored. It was suggested earlier in this chapter that the pH dependence may result from the coulombic repulsion arising on the protein molecule at high charge, in either acid or alkaline medium. Another possible explanation is that key interactions, such as hydrogen bonds, are disrupted by the combination with hydrogen ions. Further precise studies of the pH and ionic-strength dependence of the kinetics of denaturation should lead to a clarification of this problem.

The effect of temperature is important in that it permits evaluation of the "heat of activation" ΔH^* . According to modern theories of reaction kinetics, the rate of a reaction depends on the magnitude of an equilibrium constant representing the equilibrium between stable reactant and reactant in an activated form (the transition state). The value of this theory is that thermodynamic principles can be applied to this assumed equilibrium, so that the kinetic results can be interpreted in terms of the changes which take place in this activation process. It is thus possible to derive, through equations similar to equations 1 and 2, the free energy of activation ΔF^* and the entropy of activation ΔS^* . Again, it is generally found that these values are both large and positive for protein denaturation. In fact, the high positive heat of activation ΔH^* follows immediately from the high temperature dependence of denaturation, previously emphasized. As in the interpretation of the thermodynamics of the denaturation reaction, the interpretation of these quantities is not straightforward. It has been pointed out that a large part of both the heat of activation and the entropy of activation is involved in the interaction with protons or hydroxyl ions. In some cases where correction for such effects has been made, the values of ΔS^* and ΔH^* are found to be small. In the denaturation of hemoglobin, ΔH^* is of the order 100,000 cal. per mole in an alkaline medium, but only about 10,000 cal. per mole below pH 5, emphasizing the importance of ionic equilibria.

Studies of the effect of pressure on rate lead to deductions as to the change in molecular volume in the activation process. It has been demonstrated in several cases that high hydrostatic pressures, of the order 10,000 lb. per sq. in., retard the rate of denaturation markedly. One can deduce from such results that the activation process in protein denaturation involves an increase in volume, perhaps an expansion of the protein structure.

Some Possible Mechanisms. Numerous workers have been led to the conclusion that there is a reversibly sensitized form of the protein which can give rise to an irreversibly denatured form. Perhaps the first suggestion of this sort was made by Williams and Lundgren (1939), based on their studies of thyroglobulin in the ultracentrifuge, and formulated as



where N , α , and D refer respectively to native, reversibly sensitized or activated, and denatured forms. It was mentioned before that denaturation usually results in an enhanced susceptibility of protein to proteolytic attack. The Linderstrom-Lang school has suggested that (reversible) denaturation may in fact be the first stage in the attack of proteolytic enzymes, and have formulated the reaction



Excellent evidence for the reversible formation of an intermediate denatured protein has been presented by L. K. Christensen (1949) in the case of denaturation of β -lactoglobulin by urea. It was found that the apparent rate of denaturation, as judged by several criteria, including precipitability and optical rotation, was faster at 0° than at elevated temperature. However, the denaturation at low temperature was almost quantitatively reversible, the degree of reversibility at a given time being less the higher the temperature. The results indicate that the reversibly denatured form is favored by lower temperature; that is, the equilibrium has a negative temperature coefficient, but the rate of the irreversible step has a normal positive temperature coefficient.

Wright and Schomaker (1948) deduced a very different mechanism from kinetic studies of the inactivation of antibodies, notably diphtheria antitoxin. They concluded that there is a reversibly inactivated but *protected* form P , the mechanism being



where I is the irreversibly inactivated form. A similar mechanism

may pertain in the heat denaturation of plasma albumin at low pH. In this case, the rate of denaturation decreases with decreasing pH below pH 4, in the same range in which reversible isomerization of the protein occurs (Gibbs, 1954; Levy and Warner, 1954).

A particularly comprehensive and important comparative study of the urea and guanidine denaturation of ovalbumin and plasma albumin has been made by means of optical rotation and viscosity studies (Kauzmann *et al.*, 1953). The results show a remarkable degree of correlation between the two properties. The most salient feature appears to be that the changes in ovalbumin are relatively slow and irreversible, the rate depending on urea concentration, whereas with plasma albumin both properties change almost instantaneously and reversibly. These observations suggest that the molecule of ovalbumin is relatively rigid, alterations taking place only slowly and irreversibly, but the plasma albumin molecule must be somewhat "pliable" or adaptable. These authors have suggested that the unique properties of plasma albumin may result from the relatively high concentration of —S—S— cross-linkages which might (a) prevent the native molecule from coiling into a fully compact, rigid structure, and (b) restrain the partially unfolded or expanded molecule from completely irreversible unfolding. It is of interest in this connection that the specific rotation of native plasma albumin is somewhat anomalous, being about —60 degrees as compared to —30 to —40 degrees for most native proteins, and the rotation of all denatured proteins appears to be approximately —100 degrees. In a sense, from this point of view even native plasma albumin is partially denatured, or better, somewhat imperfectly coiled. Such suggestions are, however, still in the realm of speculation.

Though the notion that denaturation is essentially an unfolding or uncoiling of a highly folded polypeptide chain seems relatively well established, it is not yet possible to define clearly the nature of this unfolding. The problem is doubly difficult because the detailed nature of the folded structure in the native proteins is still uncertain (Chap. 16). A highly speculative scheme based on a planar-folded molecule is shown in Fig. 17-1. The possibility that denaturation is essentially an $\alpha \rightarrow \beta$ -type transformation, analogous to the transformation first observed in the fibrous keratins, has been repeatedly brought forward. There are, indeed, many compelling arguments for this point of view. Strictly speaking, however, one could not expect the denatured form to exist in the β -configuration in solution, but only in the final precipitated state, since stabilization by lateral intermolecular hydrogen bonds would appear essential. More properly, one should probably

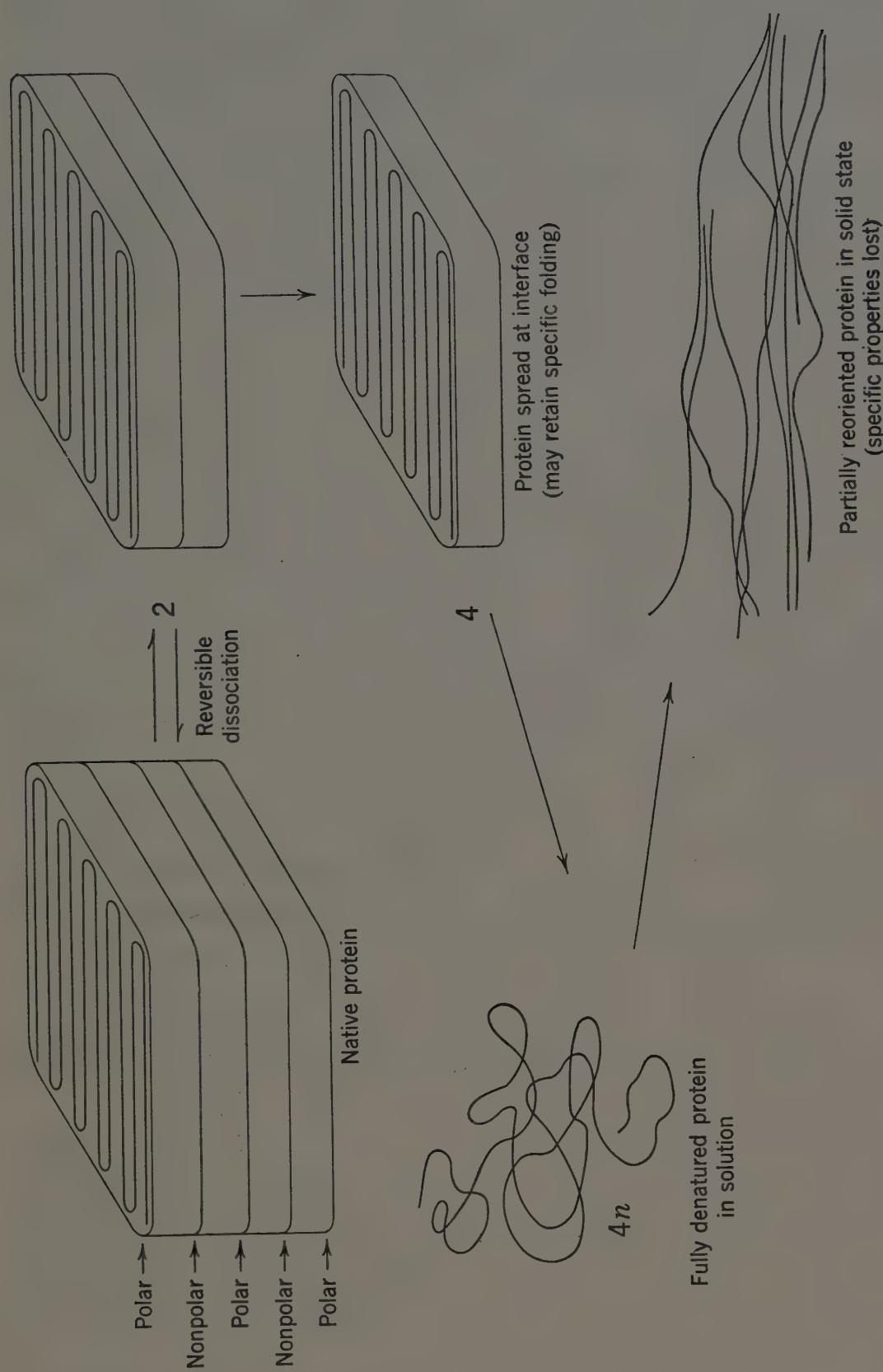


Fig. 17-1. Schematic representation of possible steps in denaturation of a protein. Based on hypothetical layer sandwich structure four for native proteins.

think in terms of a transition from the α -form, perhaps an α -helix, to a somewhat random coil.

A comparable reversible transition has now been demonstrated in synthetic polypeptides, the position of the equilibrium being dependent on pH , temperature, and solvent composition. A shift in optical

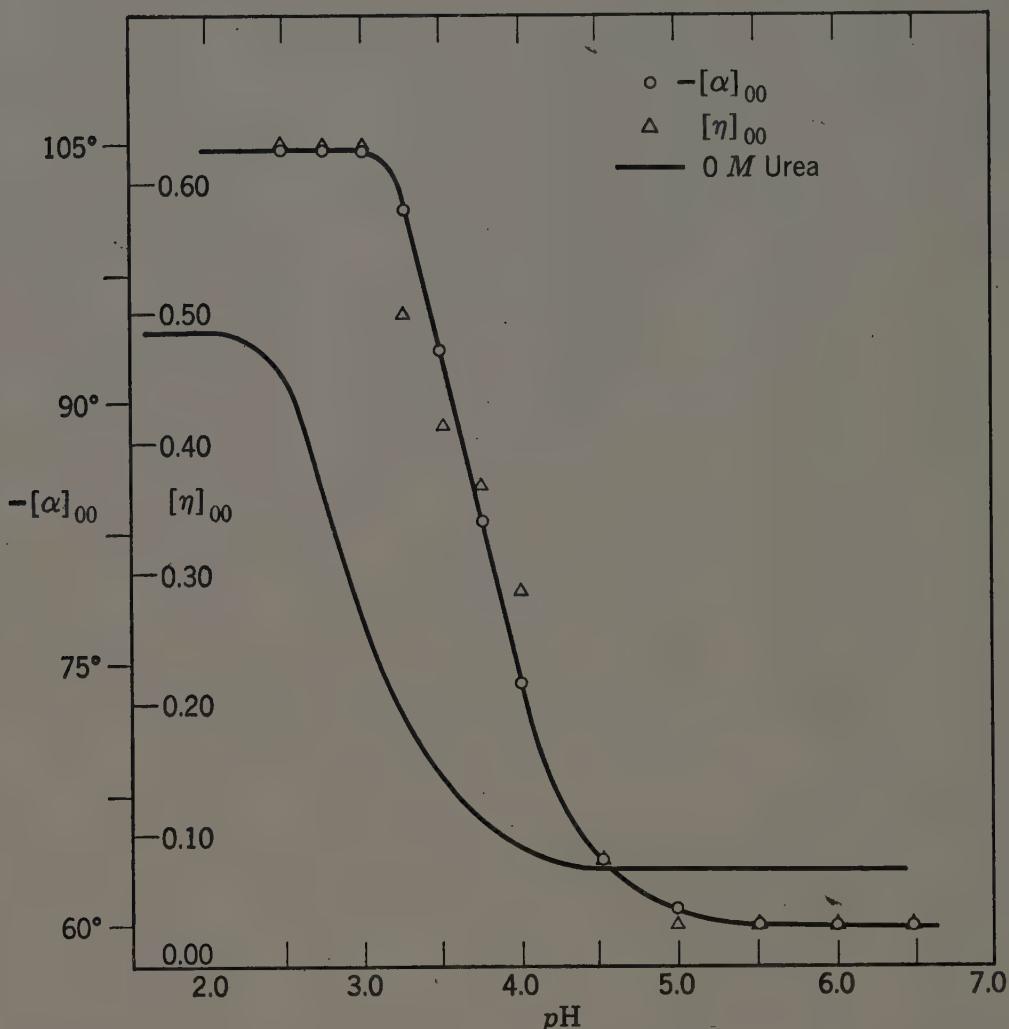


Fig. 17-2. Dependence of intrinsic viscosity and optical rotation of bovine plasma albumin on pH in the presence of 2.0 M urea, and dependence of rotation on pH in the absence of urea. From M. Sterman and J. Foster, *J. Am. Chem. Soc.*, **78**, 3652 (1956).

rotation was observed, comparable to that always found in protein denaturation. It has been calculated theoretically that in polyglycine, for example, there should be a shift in specific rotation of approximately 52 degrees in going from an α -helix to a random coil.* Some evidence based on rotatory dispersion† is also in accord with this picture.

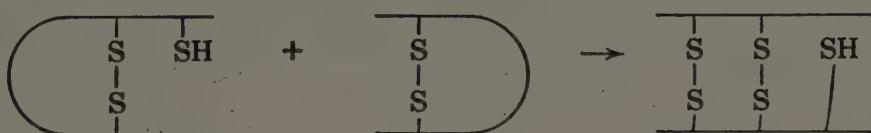
* An increase of 40 to 50 degrees is observed in plasma albumin in high urea concentration or at low pH .

† Dependence of optical rotation on wave length of light.

A very stimulating theoretical consideration of the thermodynamics of the destruction of helically coiled peptide chains should be mentioned (Schellman, 1955). These calculations, involving admittedly a number of reasonable assumptions, predict a destruction of helices by either high urea concentration or high temperature. They further predict the existence of "threshold" temperatures and urea concentrations, below which the helix is stable and above which it is unstable. It should be pointed out, however, that there is nothing in the helical model which is unique in this regard. It can be anticipated that any structure stabilized by a large number of intramolecular hydrogen bonds will lead to similar deductions, at least qualitatively.

Finally, studies on the rate of exchange of the hydrogen atoms in proteins with deuterium in heavy water should be mentioned (Hvidt and Linderstrom-Lang, 1954). Results indicate that the peptide hydrogens of native globular proteins undergo exchange only very slowly, presumably because of their involvement in hydrogen bonding. Limited results to date indicate that this may not be so in denatured proteins, yielding further evidence that denaturation involves rupture of the intramolecular network of hydrogen bonds.

From the point of view of a transition of the type under discussion here, the insolubility of the denatured protein might logically be accounted for on the basis of the tendency of the unfolded molecule to participate in intermolecular hydrogen bonding. This explanation is somewhat different from, though closely related to, that outlined previously. Mention should be made of a third possibility which has received some attention in recent years, namely, that insolubility is due to the formation of a highly cross-linked structure through disulfide exchange, which might be illustrated schematically by the reaction



Thus, intramolecular —S—S— bonds are converted to intermolecular bonds. (See, for example, Hospelhorn, *et al.*, 1954; Halwer, 1954.) Although this is doubtless a contributory factor in some cases, it also appears that aggregation and insolubilization can take place in the absence of such reactions.

Summary. Today the term denaturation seems far too vague to be of particular value, but it cannot be discarded until other and more descriptive terms can be substituted. In a fundamental way, denaturation of globular proteins seems clearly to involve a change from a definite, compact, hydrogen-bonded configuration to one that is less

compact and less definite. In some instances this may involve only a slight loosening of the structure; in extreme cases it may lead to an unfolded or randomly coiled primitive polypeptide chain. This fundamental change is manifested in a variety of observable alterations in physicochemical and biological properties. New ideas on protein structure and new techniques now available may be expected to lead to rapid expansion of knowledge in this important area of protein behavior.

SELECTED REFERENCES

Reviews

"The Coagulation of Proteins," M. L. Anson in Schmidt's *Chemistry of the Amino Acids and Proteins*, p. 407, Charles C Thomas, Springfield, Illinois, 1938.

"The Chemistry of Protein Denaturation," H. Neurath, J. P. Greenstein, F. W. Putnam, and J. O. Erickson, *Chem. Revs.*, **34**, 157 (1944).

"Protein Denaturation and the Properties of Protein Groups," M. L. Anson, *Advances in Protein Chem.*, **2**, 361 (1945).

"Protein Denaturation," F. W. Putnam in Neurath and Bailey's *The Proteins*, Vol. IB, p. 807, Academic Press, New York, 1953.

"Denaturation of Proteins and Enzymes," W. Kauzmann in McElroy and Glass's *The Mechanism of Enzyme Action*, p. 70, Johns Hopkins Press, Baltimore, 1954.

"Recent Studies on the Reversible Denaturation of Proteins," M. Joly, *Progr. Biophysics and Biophys. Chem.*, **5**, 168 (1955).

Original Articles

"Denaturation of Proteins by Urea and Related Substances," F. G. Hopkins, *Nature*, **126**, 328, 383 (1930).

"Osmotic Pressure, Molecular Weight and Stability of Amandin and Excelsin and Certain Other Proteins," N. F. Burk, *J. Biol. Chem.*, **120**, 63 (1937).

"Denaturation of Proteins by Synthetic Detergents and Bile Salts," M. L. Anson, *J. Gen. Physiol.*, **23**, 239 (1939).

"Ultracentrifugal Analysis and Stability in Protein Systems," J. Williams and H. Lundgren, *J. Phys. Chem.*, **43**, 989 (1939).

"Denaturation of Proteins and Its Apparent Reversal. I. Horse Serum Albumin," H. Neurath, G. R. Cooper, and J. O. Erickson, *J. Biol. Chem.*, **142**, 249 (1942).

"The State of Tyrosine in Egg Albumin and in Insulin as Determined by Spectrophotometric Titration," J. Crammer and A. Neuberger, *Biochem. J.*, **37**, 302 (1943).

"The Molecular Structure of Fibers Made from Native Egg Albumin," K. J. Palmer and J. A. Galvin, *J. Am. Chem. Soc.*, **65**, 2187 (1943).

"Conversion of Globular to Oriented Fibrous Proteins," G. C. Nutting, F. R. Senti, and M. J. Copley, *Science*, **99**, 328 (1944).

"Electrophoretic Properties of Denatured Serum Albumin," F. W. Putnam and H. Neurath, *J. Biol. Chem.*, **160**, 239 (1945).

- "Electrophoretic Study of Kinetics at Neutrality of Heat Denaturation of β -Lactoglobulin," D. Briggs and R. Hull, *J. Am. Chem. Soc.*, **67**, 2007 (1945).
- "The Combination of Fatty Acids and Related Compounds with Serum Albumin. I. Stabilization Against Heat Denaturation," P. Boyer *et al.*, *J. Biol. Chem.*, **162**, 181 (1946). "II. Stabilization Against Urea and Guanidine Denaturation," P. Boyer, G. Ballou, and J. Lück, *J. Biol. Chem.*, **162**, 199 (1946).
- "Pressure and Protein Denaturation," F. H. Johnson and D. H. Campbell, *J. Biol. Chem.*, **163**, 689 (1946).
- "Absolute Reaction Kinetics of Tobacco Mosaic Virus and a Proposed Theory of Denaturation," G. A. Boyd, *J. Phys. & Colloid Chem.*, **52**, 1146 (1948).
- "Studies on the Denaturation of Antibody. III. Kinetic Aspects of the Inactivation of Diphtheria Antitoxin by Urea," G. Wright and V. Schomaker, *J. Am. Chem. Soc.*, **70**, 356 (1948).
- "Trypsin Splitting and Denaturation of β -Lactoglobulin," L. K. Christensen, *Nature*, **163**, 1003 (1949).
- "The Influence of Buffer Composition, pH and Aggregation on the Thermal Denaturation of Tobacco Mosaic Virus," D. Fraser and F. H. Johnson, *Arch. Biochem.*, **24**, 338 (1949).
- "The Activated Complex of the Protein Molecule," D. S. Kahn, *J. Phys. & Colloid Chem.*, **53**, 977 (1949).
- "Streaming Birefringence of Denatured Ovalbumin," J. F. Foster and E. G. Samsa, *Science*, **112**, 473 (1950).
- "Reversible Denaturation of Chymotrypsinogen," M. A. Eisenberg and G. W. Schwert, *Federation Proc.*, **9**, 167 (1950).
- "Denaturation of Plant Proteins. III. Optical Rotation, Viscosity and Reducing Capacity," B. Jirgensons, *J. Polymer Sci.*, **5**, 179 (1950).
- "Influence of Reagents and Conditions on the Quantum Yield for the Inactivation-Denaturation of Chymotrypsin," A. D. McLaren and P. Finkelstein, *J. Am. Chem. Soc.*, **72**, 5423 (1950).
- "The Deamination of Egg Albumin. III. Quantitative Immunochemical Studies on Crystalline Egg Albumin and Its Denatured and Deaminated Derivatives," P. H. Maurer and M. Heidelberger, *J. Am. Chem. Soc.*, **73**, 2076 (1951).
- "The Kinetics and Mechanism of the Heat Inactivation of Pepsin," E. J. Casey and K. J. Laidler, *J. Am. Chem. Soc.*, **73**, 1455 (1951).
- "Behavior of a Synthetic Polypeptide Analogous to Protein Denaturation," A. Elliott, *Nature*, **170**, 1066 (1952).
- "Kinetic Study of the Reversible Acid Denaturation of Ferrihemoglobin," J. Steinhardt and E. Zaiser, *J. Am. Chem. Soc.*, **75**, 1599 (1953).
- "Thermal Denaturation of Proteins as an Order-Disorder Transition in a High Polymer," J. Colvin, *Arch. Biochem. and Biophys.*, **46**, 385 (1953).
- "The Kinetics of Protein Denaturation. I-V," W. Kauzmann *et al.*, *J. Am. Chem. Soc.*, **75**, 5139 (1953).
- "A Reversible Reaction of Bovine Serum Albumin," H. Gutfreund and J. Sturtevant, *J. Am. Chem. Soc.*, **75**, 5447 (1953).
- "The Globular-Fibrous Protein Transformation," E. Barbu and M. Joly, *Discussions Faraday Soc.*, **No. 13**, 77 (1953).
- "Changes in the Viscosity of Protein Solutions During Heat Denaturation," K. Strachitskii and K. Firfarova, *Biokhimiya*, **18**, 235 (1953).
- "The Influence of Oxygen on Irreversible Denaturation of β -Lactoglobulin," G. Johansen, *Compt. rend. trav. lab. Carlsberg, Sér. chim.*, **28**, 335 (1953).

"Exchange of Hydrogen Atoms in Insulin with Deuterium Atoms in Aqueous Solutions," A. Hvidt and K. Linderstrom-Lang, *Biochim. et Biophys. Acta*, **14**, 574 (1954).

"Denaturation of Hemoglobin by Alkali," F. Haurowitz, R. Hardin, and M. Dicks, *J. Phys. Chem.*, **58**, 103 (1954).

"Denaturation of Bovine Plasma Albumin," M. Levy and R. Warner, *J. Phys. Chem.*, **58**, 106 (1954).

"Albumin from Heated Human Plasma. II. Denaturation Kinetics: Influence of pH and Temperature," R. Gibbs, *Arch. Biochem. and Biophys.*, **51**, 277 (1954).

"The Stepwise Nature of Protein Denaturation," S. Nakagawa, T. Kaminaga, and S. Araya, *J. Biochem. Japan*, **41**, 371 (1954).

"Conformation Changes of Proteins," R. Lumry and H. Eyring, *J. Phys. Chem.*, **58**, 110 (1954).

"Sulfhydryl-Dependent Aggregation Accompanying the Denaturation of Bovine Plasma Albumin by Urea," V. Hospelhorn, B. Cross, and E. Jensen, *J. Am. Chem. Soc.*, **76**, 2827 (1954).

"The Expansion of Unimolecular Films of Ovalbumin," J. Kaplan and M. Fraser, *J. Biol. Chem.*, **210**, 57 (1954).

"Disulfide Cross-Links in Denatured Ovalbumin," M. Halwer, *J. Am. Chem. Soc.*, **76**, 183 (1954).

"The Inhibition of Trypsin. III. Influence of Urea," T. Viswanatha and I. Liener, *J. Biol. Chem.*, **215**, 777 (1955).

"Kinetics of Protein Denaturation. I. Thermal Denaturation of β -Lactoglobulin," J. Stauff and E. Uhlein, *Kolloid-Z.*, **143**, 1 (1955).

"Molecular Weight of Ovalbumin and of Bovine Serum Albumin in Urea Solution," H. McKenzie, M. Smith, and R. Wake, *Nature*, **176**, 738 (1955).

"The Sedimentation Behavior of Bovine Serum Albumin in Acid Solutions," P. Bro, S. Singer, and J. Sturtevant, *J. Am. Chem. Soc.*, **77**, 4924 (1955).

"Sedimentation and Viscosity Studies of Bovine Serum Albumin in Urea Solution," P. Charlwood, *Can. J. Chem.*, **33**, 1043 (1955).

"The Stability of Hydrogen-Bonded Peptide Structures in Aqueous Solution," J. Schellman, *Compt. rend. trav. lab. Carlsberg, Sér. chim.*, **29**, No. 15 (1955).

"Effect of pH and Salt on the Sedimentation Behavior, Viscosity and Osmotic Pressure of Conalbumin Solutions," R. Phelps and J. Cann, *Arch. Biochem. and Biophys.*, **61**, 51 (1956).

"Effect of Urea on Trypsin and Alpha-Chymotrypsin," J. Harris, *Nature*, **177**, 471 (1956).

"Swelling of Protein Molecules in Solution and the Alpha-Beta Transformation," T. Hill, *J. Phys. Chem.*, **60**, 358 (1956).

"The Optical Rotary Power of Polyamino Acids and Proteins," D. Fitts and J. Kirkwood, *J. Am. Chem. Soc.*, **78**, 2650 (1956).

"Conformation Changes in Bovine Plasma Albumin Associated with Hydrogen Ion and Urea Binding. I and II," M. Sterman and J. F. Foster, *J. Am. Chem. Soc.*, **78**, 3652, 3656 (1956).

"Heat Denaturation of Serum Albumin in Presence of Perfluorooctanoic Acid," E. Ellenbogen and P. Maurer, *Science*, **124**, 266 (1956).

Blood Proteins

Electrophoretic analysis of plasma and serum

Alcohol fractionation

Properties of the plasma proteins

Hemoglobin

In the preceding chapters some of the unique properties of proteins have been discussed, with speculations as to their detailed structure. The information behind this discussion has been derived from careful studies on a relatively few apparently well-defined proteins. It is the purpose of the next two chapters to discuss the origin, methods of preparation, and properties of some of these important proteins. In later chapters the biological importance of some of these proteins, as well as others, will be mentioned.

The blood is one of the most important and complex of vital fluids. Its primary function is transport. For present purposes blood may be considered as consisting of a suspension of erythrocytes and leucocytes in a protein solution, the plasma. The erythrocytes consist essentially of a concentrated solution of the respiratory protein hemoglobin, but they also contain enzymes of the glycolytic system, catalase, carbonic anhydrase, and presumably many other proteins.

When fresh, carefully drawn blood containing an anticoagulant (citrate or oxalate salt)* is centrifuged, the erythrocytes and leucocytes

* These anticoagulants act by virtue of their ability to precipitate or tie up the Ca^{++} ions essential in the coagulation process. A more recent innovation utilizes cation-exchange resins such as Dowex 50 for this purpose. The blood is passed directly from the donor's vein over the resin, minimizing the possibility of even the early stages of the clotting process taking place.

are removed, leaving a relatively clear straw-colored fluid, the plasma. In addition to ions (predominantly Na^+ , Cl^- , H_2PO_4^- , $\text{HPO}_4^{=}$, HCO_3^- , and $\text{CO}_3^{=}$) and other solutes of low molecular weight, the plasma contains approximately 6.5 per cent protein. If the blood is allowed to clot normally, one obtains serum, which differs from plasma chiefly in that fibrinogen, the clot-forming protein, is absent. Small amounts of other proteins may also be lost by adsorption on the clot.

Early studies showed that the protein of serum is not homogeneous. Dilution with water, when accompanied by a lowering of $p\text{H}$, yields a protein precipitate which is salt-soluble (globulin); other proteins are left in solution. The more soluble protein is predominately albumin, although some of it has one of the characteristics of a globulin, namely, precipitability from solutions by half-saturation with ammonium sulfate. This protein fraction has been termed pseudoglobulin to distinguish it from the more typical globulin, sometimes called euglobulin, which is insoluble, in absence of salt, at its isoelectric point.

Studies by means of the salting-out technique also indicated the presence of two types of globulin and an albumin in serum. In addition, there is in plasma the even more readily salted-out protein, fibrinogen. One classical scheme utilized quarter-saturation with ammonium sulfate to precipitate primarily fibrinogen, third-saturation to precipitate euglobulins, and half-saturation to precipitate pseudoglobulins, leaving albumin in solution. The precipitates obtained by such means are far from homogeneous.

ELECTROPHORETIC ANALYSIS OF PLASMA AND SERUM

With the introduction of the Tiselius electrophoretic method it became apparent that the protein composition of plasma is even more complex than indicated by salting-out data. Furthermore, a tool was now available to make possible more intelligent attempts at fractionation of the plasma proteins. A typical electrophoretic pattern of plasma is shown in Fig. 12-6. The albumin migrates as a single fairly homogeneous component, labeled A . The globulin portion is found to consist of several electrophoretic components; these have been named in order of decreasing mobility, under the conditions usually used in plasma electrophoresis ($p\text{H}$ near 8), α -, β -, and γ -globulin, as indicated in the figure. The component between the β - and γ -globulins is essentially absent in serum and corresponds to the fibrinogen. The peak labeled δ in the ascending pattern was originally thought to correspond to a fourth type of globulin. It is now known that it is an anomaly, consisting of a gradient in buffer-salt concentra-

tion as well as in concentration of all protein species. A smaller peak in the descending pattern, labeled ϵ , is also an anomaly but is invariably smaller than δ , owing to the fact that it involves only a salt gradient (Chap. 12).

Another anomaly frequently observed in plasma patterns is a sharp line or streak following the α -globulin boundary. This " β -anomaly," the explanation for which is not yet entirely clear, is associated with the development of a turbid region between the β -globulin and fibrinogen boundaries. Since the turbidity and the "spike" develop rather late in a run, after resolution is well advanced, runs can often be carried out successfully before the spike develops.

The globulin regions are not homogeneous, even electrophoretically. It is frequently possible to resolve the α - and β -globulins into distinct peaks. These are labeled with subscript numerals in the order of decreasing mobility, as α_1 - and α_2 -globulin, α_1 - having the higher mobility.

Electrophoretic studies have been carried out on plasmas from many species. In general there is a striking similarity, the same components being present in plasma from man as from the hog or chicken. There are differences, however, in the relative amounts of the components (Table 18-1).

TABLE 18-1. Electrophoretic Distribution of Proteins in Plasmas of Various Species

(Determinations by per cent)

Species	Albumin	Globulins			Fibrinogen
		α	β	γ	
Man ¹	55.2	14.0	13.4	11.0	6.5
Cow ²	34.3	16.6	12.1	15.4	21.6
Swine ³	38.1	18.8	13.4	17.8	11.9
Horse ⁴	29.8	20.5	21.9	11.2	15.8
Monkey ⁴	50.0	15.8	16.1	9.0	8.4
Sheep ⁴	43.7	16.5	15.0	15.0	9.7
Rabbit ⁴	63.3	11.5	13.0	4.3	7.9
Dog ⁴	39.6	24.9	13.0	9.3	13.3
Cat ⁴	41.4	33.0	8.7	12.5	5.2

¹ S. H. Armstrong, M. Budka, and K. Morrison, *J. Am. Chem. Soc.*, **69**, 416 (1947). Veronal buffer.

² K. R. Hogness, J. Giffee, and V. Koenig, *Arch. Biochem.*, **10**, 281 (1946). Veronal buffer.

³ J. F. Foster, R. Friedell, D. Catron, and M. Dieckmann, *Iowa State College, J. Sci.*, **24**, 521 (1950). Phosphate buffer.

⁴ H. F. Deutsch and M. B. Goodloe, *J. Biol. Chem.*, **161**, 1 (1945). Veronal buffer.

By use of the Tiselius technique it has been shown that the proteins precipitated by third-saturation with ammonium sulfate are largely γ -globulin. Between 0.34 and 0.40 saturation, γ -, β -, and α -globulins are precipitated; between 0.40 and 0.50, β - and α -globulins.

Variability of Plasma Composition. In addition to the variations between species, the plasma composition within a given species is subject to considerable variability. The electrophoretic technique has

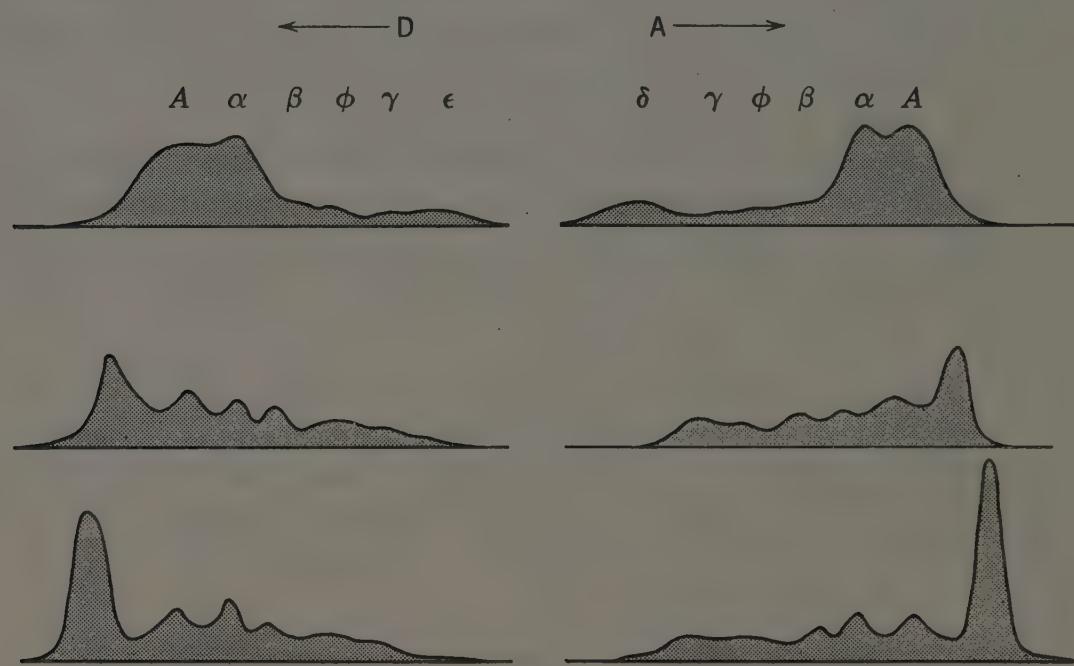


Fig. 18-1. Change in the electrophoretic pattern of plasma of pigs from birth to weaning. Top, at birth; center, at 24 hours post partum; bottom, 8 weeks (weaning).

made possible extensive studies of this variability. Clinically, much work has been done on the changes in plasma pattern associated with various diseases. Electrophoresis can serve as an aid in clinical diagnosis. In large part these studies have been empirical, although some of the variations can be given good physiological interpretations. This field has been reviewed by Gutman (1948).

Animals on near-starvation diets show pronounced abnormalities, the albumin and γ -globulin components being markedly lowered. But wide variations in rations, so long as the diet is at least marginal, are without significant effect on composition. There appear to be definite individual variations that cannot be attributed to diet or to other environmental influences and are therefore presumably genetic. Furthermore, there are important changes with age. The new-born of many species, particularly the cow, pig, and horse, are very poor in plasma γ -globulin. In the pig the γ -globulin may rise from less than 5 per cent to as high as 30 to 40 per cent of the total plasma protein

between birth and 12 hours *post partum*. These pronounced changes have been attributed to the ingestion of antibody proteins by the young animals from the maternal colostrum. Fig. 18-1 shows the change in electrophoretic composition of the plasma of young pigs. In the human the situation is quite different, the concentration of γ -globulin in the young being higher than in the adult. In this case antibody transmission is thought to take place via the placenta rather than postnatally.

In the bovine species a distinct globulin, called *fetuin*, is present to the extent of about 20 per cent of the plasma protein in the new-born, falling to a value of 5 per cent at six months. This protein may comprise as much as one third of the total serum protein in the fetal stage.

ALCOHOL FRACTIONATION

By far the most elaborate and successful procedures for the fractionation of plasma are the ethanol-water methods* developed by Cohn and co-workers. The original immediate objective of this work was to provide purified albumin for the treatment of shock on the battlefield. An attempt was made to recover all of the proteins, however, and many other useful fractions have come out of this work. The technique involves the separation of the plasma proteins into a relatively few main fractions, usually six. These can then be subfractionated to yield more highly purified proteins. Table 18-2 gives one set of conditions for the separations, together with the electrophoretic composition of the principal fractions. The yields are seen to be quantitative (there is some apparent loss in albumin and gain in the other components, presumably due to inadequacies in the electrophoretic analysis), and the separations obtained are remarkably good.

Figure 18-2 summarizes this fractionation work pictorially. Starting from the center, the next circle gives the quantitative breakdown into fractions and subfractions; the next circle gives the nature of the protein or other characteristic substances present in the fraction; the next circle identifies the protein electrophoretically; the next indicates the general biological function of the proteins, where known; the outer circle gives the specific uses to which the fractions resulting from the alcohol fractionation scheme have been applied.

* The ethanol techniques are being supplemented by separations based on adsorption, for example on BaSO_4 and $\text{Zn}(\text{OH})_2$, and precipitation with heavy-metal ions. Such separations are apt to be more specific.

TABLE 18-2. Separation and Electrophoretic Composition of Principal Fractions of Human Plasma¹

Electrophoretic Composition Estimated in Plasma,
gm./liter of plasma

Fraction	pH	Ionic Strength	Temp., °C.	Mole Fract. Ethanol	Protein Conc., gm./l.	Globulin			Fibrinogen	Total
						Albumin	α	β		
Plasma	7.4	0.16	60.3	36.3	9.2	10.6	7.2	65.8
I	7.2	0.14	-3	0.027	51.1	0.2	0.3	0.5	0.3	3.4
II + III	6.8	0.09	-5	0.091	30.0	0.8	1.1	9.1	7.0	19.0
IV-1	5.2	0.09	-5	0.062	15.8	0	4.5	0.5	0.1	5.1
IV-4	5.8	0.09	-5	0.163	10.1	0.9	2.7	2.2	0	5.8
V	4.8	0.11	-5	0.163	0.2	29.9	1.3	0.3	0	31.5
VI	0.8	0.2	<0.1	0
Total	32.6	10.1	12.6	7.4	3.1
										65.8

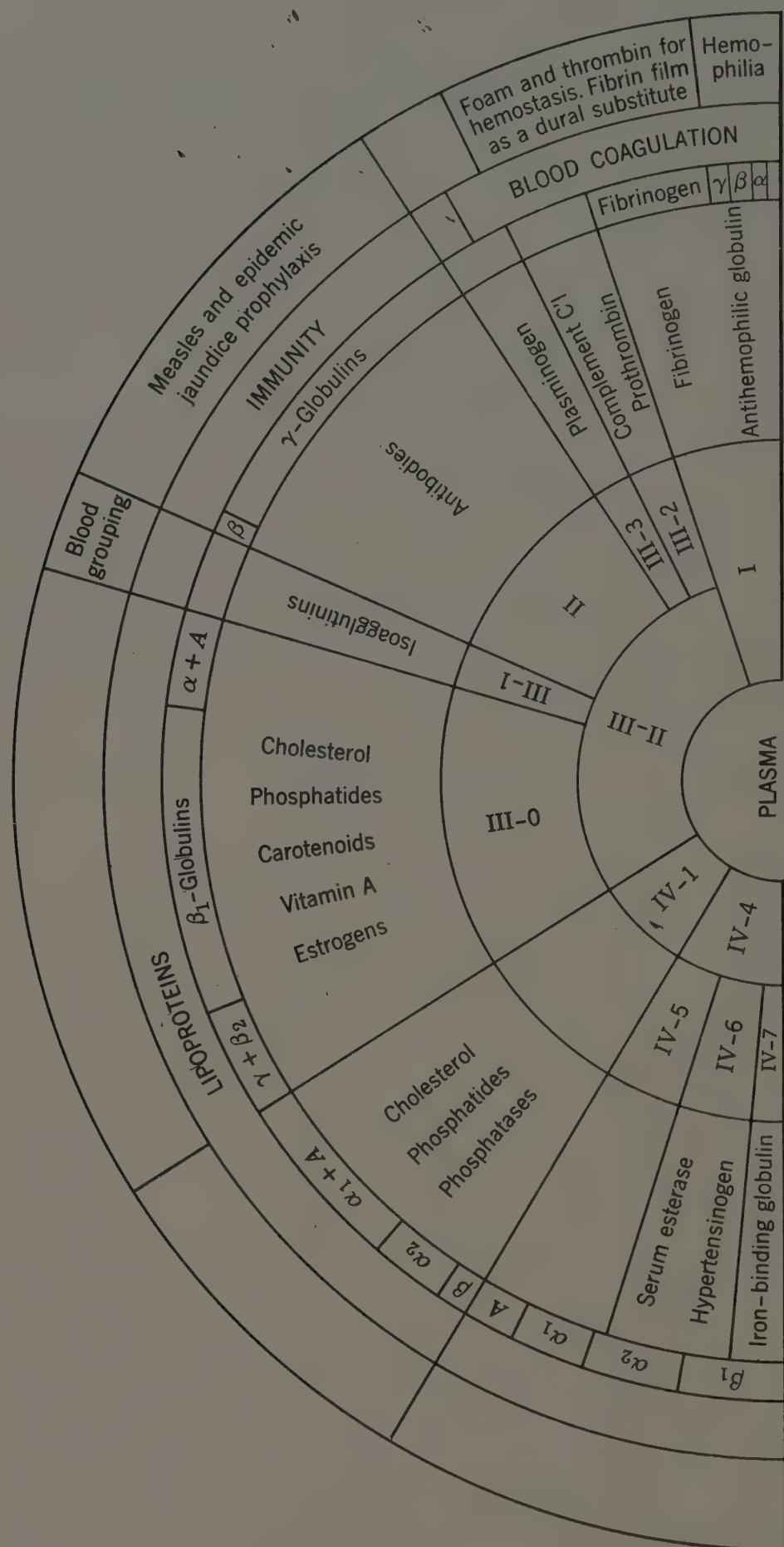
¹ From J. T. Edsall, *Advances in Protein Chem.*, **3**, 383 (1947).

PROPERTIES OF THE PLASMA PROTEINS

Albumin. The albumin component is the most readily crystallized of the plasma proteins. Albumins from several species, especially equine and human, have been crystallized by salting-out with ammonium sulfate. More recently it has been found possible to crystallize human and bovine serum albumin from ethanol-water media provided a trace of decanol, or other aliphatic alcohol containing six or more carbon atoms, is added (toluene and benzene have also been found effective). The function of the additives is not clear, but they are apparently bound and incorporated into the crystals.

The amino acid composition of human albumin is given in Table 18-3, along with the composition of some of the other plasma proteins. These data are probably the best available on these proteins and are fairly complete. Bovine albumin is quite similar in composition to human with regard to acidic, basic, and sulfur-containing amino acids, but differs appreciably in the simple aliphatic, aromatic, and hydroxy acids. Neither human nor bovine albumin contains much carbohydrate. One horse serum albumin fraction containing carbohydrate (5.5 per cent) and another without carbohydrate can be distinguished; both have been crystallized. Further data on the titratable groups present, as well as titration curves, were presented in Chap. 11 (Table 11-3 and Fig. 11-1).

Interesting features of the amino acid composition of plasma albumins include the rather high content of cystine and the low value for cysteine. On the basis of a molecular weight of 65,000, there are approximately 16 to 18 cystine residues per molecule, presumably introducing this number of —S—S— cross-linkages. The cysteine value corresponds to only one, and possibly slightly less than one, residue per molecule. By direct reaction with mercury, organic mercurials, or silver ion, only two thirds of an —SH group per molecule are found. Under proper conditions, Hg^{++} ion yields a dimer by cross-linking two albumin molecules through their —SH groups. This dimer can be crystallized and the monomeric albumin subsequently obtained by removal of mercury. The resultant albumin contains one —SH group per molecule, and has been termed mercaptalbumin. This clearly indicates ordinary crystalline plasma albumin to be heterogeneous, at least with respect to availability of the —SH group. No other difference between the two fractions has yet been found. It is of interest that in the presence of high concentrations of denaturing agents such as urea, one free —SH group per molecule is found in ordinary crystallized albumin, but upon removal of the agent the



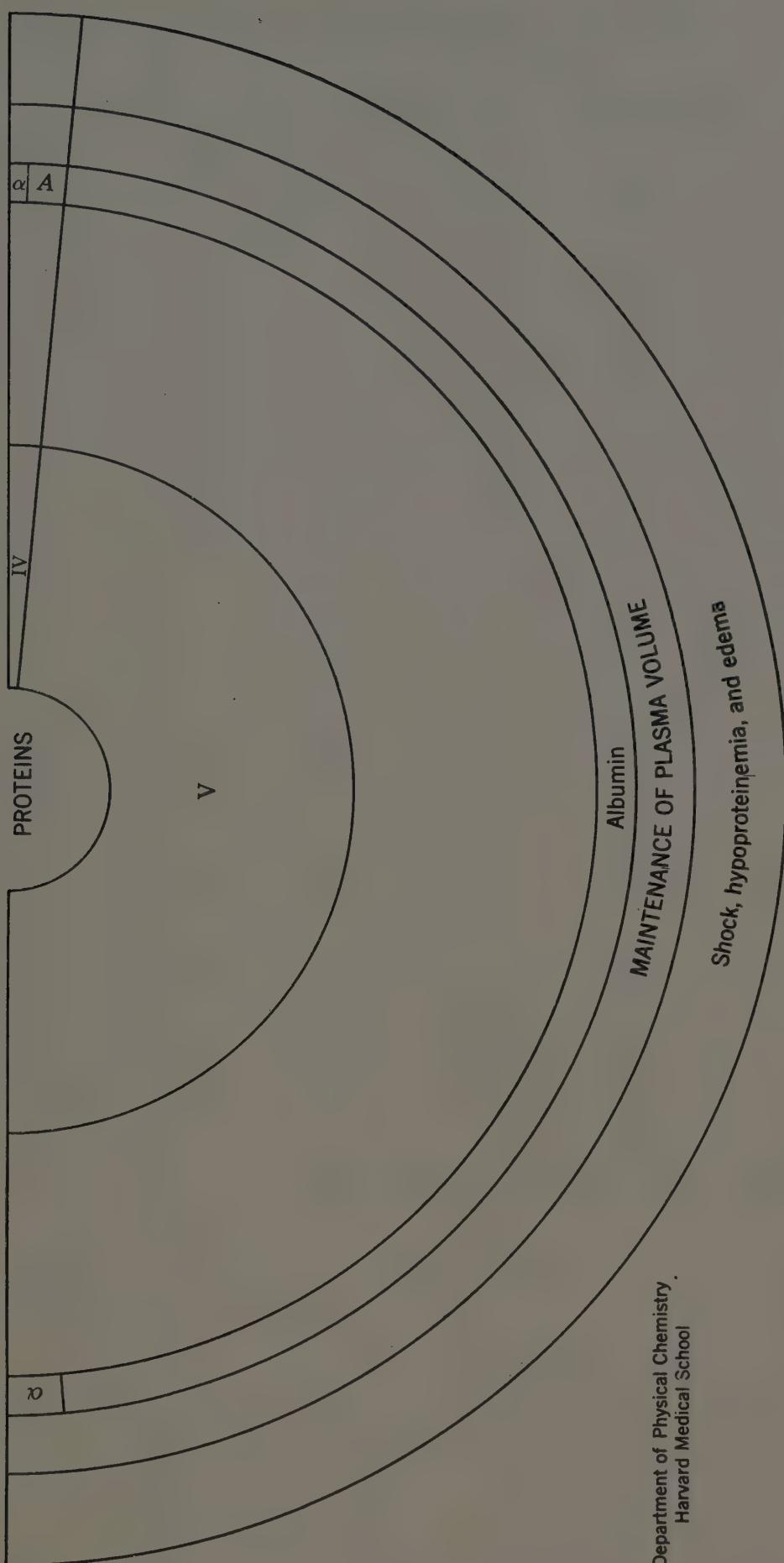


Fig. 18-2. The plasma proteins, their natural functions, clinical uses, and separation into fractions. Reproduced from J. T. Edsall, *Advances in Protein Chem.*, 3, 383 (1947).

TABLE 18-3. Amino Acid Composition of Human Plasma Proteins¹

(Grams amino acid obtained per 100 gm. protein)

Constituent	<i>A</i>	γ	β^2	α^2	Fibrinogen ³
Total N	15.95	16.03	15.24	...	16.9
Total S	1.96	1.02	1.32	...	1.26
Free α -amino	0.18	0.11
Amide N	0.88	1.11
Glycine	1.6	4.2	5.6	3.1	5.6
Alanine
Valine	7.7	9.7	7.0	5.2	4.4
Leucine	11.0	9.3	7.9	14.2	7.1
Isoleucine	1.7	2.7	5.0	1.7	4.8
Proline	5.1	8.1	7.1	4.7	5.7
Phenylalanine	7.8	4.6	4.7	4.6	4.2
Cysteine	0.7	0.7	3.5	1.5	0.4
Half cystine	5.6	2.4	2.3
Methionine	1.3	1.1	1.7	1.4	2.5
Tryptophan	(0.2)	2.9	2.0	1.9	3.3
Arginine	6.2	4.8	6.8	7.7	7.9
Histidine	3.5	2.5	2.8	2.8	2.8
Lysine	12.3	8.1	6.6	8.9	8.3
Aspartic acid	10.4	8.8	9.8	9.0	13.6
Glutamic acid	17.4	11.8	14.5	21.6	14.3
Serine	3.7	11.4	7.1	5.0	9.2
Threonine	5.0	8.4	6.1	4.9	6.6
Tyrosine	4.7	6.8	6.0	4.5	5.8
Total	105.9	108.3	104.2	102.7	108.8

¹ From the work of various authors, as compiled by J. T. Edsall, *Advances in Protein Chem.*, **3**, 464 (1947).

² Gross fractions containing many different proteins.

³ Sample was 87 per cent fibrinogen as determined by clottability.

system reverts to the original condition. It would appear that for some reason a fraction of the molecules tend to mask this group in their compact native configuration.

Other evidence for heterogeneity in crystalline plasma albumin preparations is available. Electrophoretically, though some such preparations appear homogeneous in the pH range 7 to 8, they always exhibit two or more boundaries near the isoelectric point (pH 4.0 to 4.7). It has been shown that under certain conditions two major peaks are obtained, the relative areas varying with pH in such a manner

as to indicate an equilibrium between two isomeric forms of the protein (Aoki and Foster, 1956). Below pH 3.5 the protein again appears to be homogeneous. In the ultracentrifuge most preparations show the presence of a small amount, usually 5 to 10 per cent, of a faster sedimenting component, possibly a dimer.

Another peculiar property of plasma albumins is their remarkable tendency to combine with anions. Even chloride ion is bound to an appreciable degree at ionic strengths as low as 0.01. As a consequence, the isoionic point (pH 5.1 to 5.4, depending on ionic strength) is appreciably higher than the isoelectric point (pH 4.2 to 4.7) (Chap. 14). The binding of carboxylic acids has been shown to have a stabilizing action toward heat and urea denaturation. Thus, although human plasma albumin coagulates within a few seconds at 65° in the absence of salt, it is stable for many hours in presence of 0.15 M sodium caprylate. Advantage has been taken of this fact in the sterilization of plasma albumin solutions for intravenous injection. Also, it is possible to purify albumin by differentially coagulating other proteins in the presence of caprylate, the stabilizing action apparently being limited to the albumin fraction. This stabilizing action is limited to ions of intermediate size, acetate being relatively ineffective and soaps and detergents serving as denaturing agents (as in the case of other proteins). It has been found that dyes and such compounds as the sulfa drugs are readily bound by albumin. The possible implications of this binding ability with respect to the biological function of plasma albumin are manifold.

The molecular weight of 65,000 for plasma albumins has now been found by a variety of methods, including X-ray diffraction, osmotic pressure, and sedimentation-diffusion. Light-scattering usually yields a somewhat higher value, typically 70,000 to 75,000, presumably because of the presence of dimer. The frictional ratio, 1.28, and intrinsic viscosity, 0.035, indicate the molecule to be either somewhat asymmetric or swollen. The data have been interpreted on the basis of a prolate ellipsoidal model of length about 150 A. and diameter 38 A. Low-angle X-ray scattering indicates the molecule to be less asymmetric than this.

The expansion of the plasma albumin molecule in acid solution has previously been mentioned. It also expands reversibly in concentrated urea and in solutions of guanidine salts. In both cases, there is a parallel increase in optical rotation, the value rising from about -60 degrees for the native protein to -100 degrees for the expanded form. It has been suggested that this results from an unfolding of helically coiled peptide chains to an essentially randomly

coiled form. The large number of —S—S— cross-links may be important in restraining the expansion and preventing irreversible denaturation. The extent to which this molecular flexibility is responsible for the other unusual properties of plasma albumin is not yet clear. This protein is the object of much active research in a variety of laboratories at present, and some clarification of such questions can be anticipated in the near future.

Finally, the physiological function of plasma albumin should be considered. Although it is the major protein constituent of the plasma, no obvious functional role has appeared other than that of regulation of the osmotic pressure of the blood with relation to the surrounding body fluids. Albumin has a relatively low molecular weight and a high net charge at normal blood pH (7.2). As a consequence (and as a result of the Donnan equilibrium) it is responsible for most of the osmotic pressure of plasma. It is this property which made concentrated albumin solutions useful in the treatment of shock on the battlefields of World War II. For the same reason, albumin has proved effective in reducing (though only temporarily) massive edema in patients suffering from nephritis. Other roles which have been suggested for this protein are that it provides a nutritional reserve and transports various small molecules and ions.

Fibrinogen and the Clotting Mechanism. Fibrinogen is the soluble precursor of fibrin, the clot-forming protein. It is present in man to the extent of approximately 4 per cent of the plasma proteins, and to a considerably higher extent in swine and bovine plasmas. As with the other blood proteins, the composition is quite variable from individual to individual.*

The amino acid composition of fibrinogen is given in Table 18-3. Fibrinogen is a typical globulin from the standpoint of solubility behavior, being very insoluble at the isoelectric point in the absence of salt, and the most readily salted-out of any of the plasma proteins. It is thus normally removed first in the ethanol fractionation scheme (fraction I).

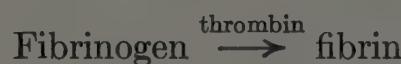
The best accepted molecular weight is now about 350,000 to 400,000 and the protein is evidently highly asymmetric. The intrinsic viscosity, 0.25, and frictional properties, as well as angular dependence of light-scattering, can best be explained on the basis of an elongated rod (or prolate ellipsoid) of length 600 to 700 Å. and diameter about

* One of the authors has found as high as 30 per cent fibrinogen in the plasma of cattle suffering severe hemorrhage, as determined both by electrophoresis and by clotting with thrombin. This high level is presumably compensatory, an attempt on the part of the animal to accomplish normal clotting.

40 A. Flow-birefringence results are in excellent agreement with these dimensions. The unusual asymmetry is understandable from a functional standpoint, that of formation of the fibrin gel or clot, in view of earlier discussion of gelation.

To date fibrinogen has not been crystallized,* but fractions containing as much as 98 per cent fibrinogen (as judged by clottability) have been prepared.

The clotting process can be represented schematically by the equation



The reaction normally appears to be irreversible (however, see footnote on p. 341). The catalyst, the enzyme *thrombin*, is present in plasma only in the form of its precursor, *prothrombin*, a protein of molecular weight 140,000, which is concentrated in the Cohn scheme in fraction II + III. Prothrombin has been prepared in a particularly high state of purity from bovine plasma. It appears that one molecule of prothrombin yields two of thrombin. The prothrombin → thrombin conversion requires a protein (possibly an enzyme) *thromboplastin* or *thrombokinase*, which is present in many tissues but particularly in lung tissue. Thromboplastin is a lipoprotein (possibly almost 50 per cent lipid or phospholipid) of enormous size (a molecular weight of 170,000,000 is claimed). In addition, calcium ion is required in the prothrombin → thrombin conversion. The importance of other factors is claimed: an accelerator factor, *Ac-globulin*, which occurs in plasma in an inactive form, and *serum prothrombin conversion accelerator* (SPCA). These two factors may or may not be identical.

The actual fibrinogen → fibrin transformation is very much more involved than is implied by the simple reaction equation given above. In recent years great strides have been made in the elucidation of this process, but some details are not yet clear. It has been clearly established, however, that thrombin is essentially a mild (and presumably very specific) proteolytic enzyme. Although it seems to be without effect on other protein substrates, it is capable of splitting certain synthetic substrates. Significantly, other proteolytic enzymes such as papain, though less active than thrombin in this regard, can nevertheless cause the clotting of fibrinogen. It has been estimated that thrombin is capable of converting 10^6 times its weight of fibrinogen to fibrin.

The first step in the conversion of fibrinogen to fibrin involves the

* Its crystallization has been claimed, but this has not been confirmed and remains very doubtful.

splitting from the fibrinogen molecule of small peptides amounting to approximately 3 per cent of the total residues of the parent molecule. The resultant change in molecular weight and shape of the fibrinogen molecule is negligible, but new end groups are created. Whereas

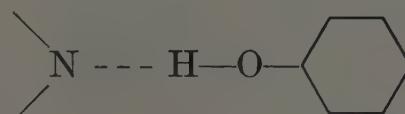


Fig. 18-3. Electron micrograph of a fibrin clot ($\times 25,200$). Courtesy of Dr. K. R. Porter, The Rockefeller Institute for Medical Research.

fibrinogen apparently contains three N-terminal residues, two tyrosine, and one glutamic acid, "activated fibrinogen" possesses two tyrosine termini (presumably the same two) plus four terminal glycine residues. Further, the amino acid composition of the "fibrinopeptides" is such as to markedly reduce the number of net negative charges on the parent molecule upon splitting. As a consequence of these modifica-

tions, the fibrinogen molecule develops a strong propensity for aggregation or polymerization.

The nature of this polymerization process has been explored extensively. Ordinarily this step proceeds rapidly to clot formation, but under unfavorable conditions the polymerization can be so retarded that it is possible to study the "intermediate polymers" that arise. Even better, certain inhibitors such as hexamethyleneglycol and urea permit limited polymerization without clot formation. It has been demonstrated that the polymerization process consists essentially of a parallel, overlapping alignment of the elongated activated fibrinogen molecules. The polymers are evidently joined through forces other than primary valence bonds, hence the polymerization is reversible upon dilution.* Presumably hydrogen bonds, plus probably electrostatic and van der Waals attractions, are responsible. There is evidence that the important hydrogen bonds may involve the imidazole groups of histidine and the phenolic groups of tyrosine. Presumably the former must be in the basic form, the latter in the acidic form, permitting bonds of the type



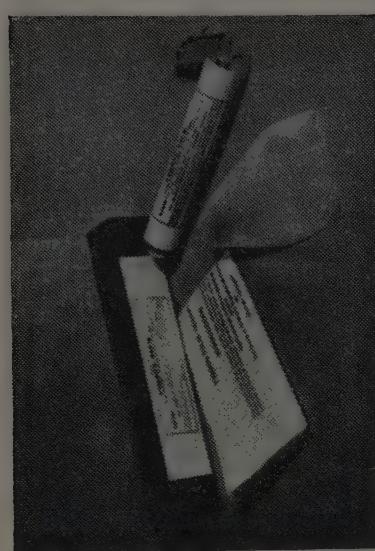
This would explain the fact that clotting takes place readily only in the *pH* range 5.3 to 10.

The rate of clot formation depends markedly on *pH*, usually being optimal somewhere around *pH* 7 to 8 depending on the ionic strength. The nature of the clots formed also depends upon *pH*. In the range of optimal clotting rate and above, the clots are transparent, gelatinous, friable, and elastic. At lower *pH* they are opaque, and can be readily compacted with expulsion of solvent. The two types are termed "fine" and "coarse," respectively.

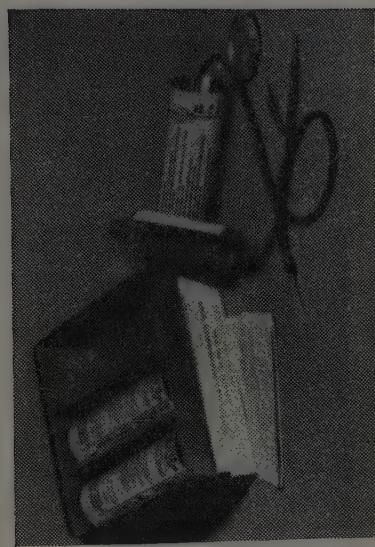
Fibrin and thrombin have found important uses clinically. Fibrin can be formed into plastic film which has found application in surgery, for example as a dural substitute. When fibrinogen solutions are whipped into a foam and then clotted and dried, a spongelike material is obtained. This sponge, when soaked in a solution of purified thrombin and applied to wounds, produces effective and rapid hemostasis. Fibrin film, fibrin foam, and thrombin, as packaged for clinical use, are shown in Fig. 18-4.

*It has been demonstrated that even the final fibrin clot, under suitable conditions, can be redissolved in the form of active fibrinogen molecules by strong urea solution, but this cannot usually be accomplished, possibly because of the formation of cross-linking —S—S— bridges through air oxidation or disulfide exchange.

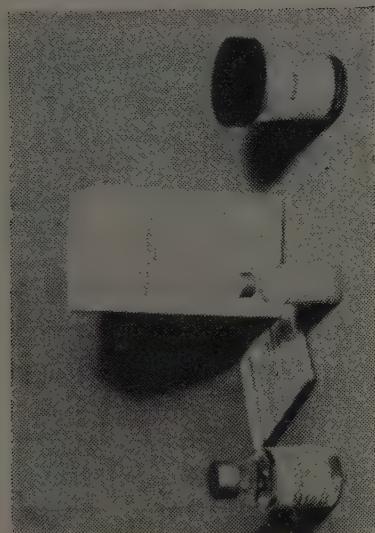




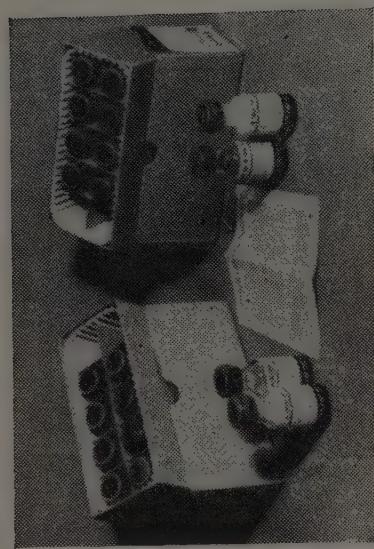
Fibrin film,
fraction I and fraction III-2



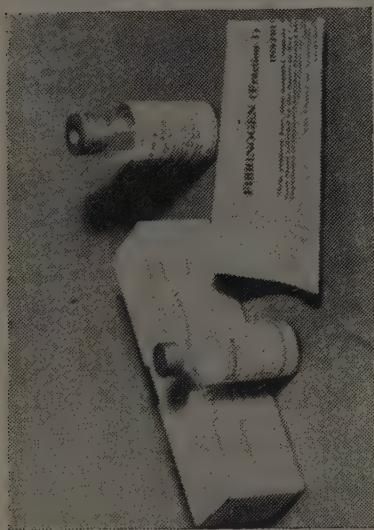
Albumin,
fraction V



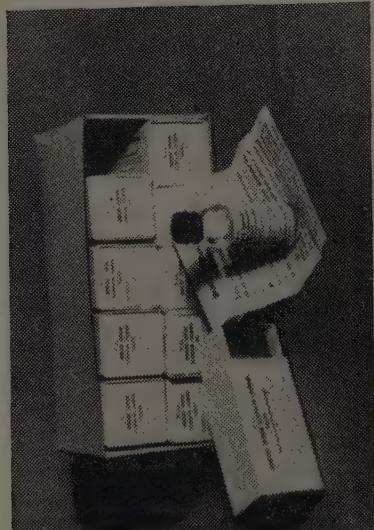
Fibrin foam and thrombin,
fraction I and fraction III-2



Isoagglutinins,
fraction III-1



Fibrinogen and thrombin,
fraction I and fraction III-2



Serum γ -globulin,
fraction II

Fig. 18-4. Products from the fractionation of human blood plasma. Courtesy of the late Dr. E. J. Cohn and the University Laboratory of Physical Chemistry, Harvard.

γ -Globulins. The globulins of plasma comprise an unknown but large number of proteins. Few if any of these proteins have been prepared in anything approaching pure form.

A γ -globulin fraction has been prepared, by means of the alcohol technique, in a high state of purity insofar as removal of the other electrophoretic components of plasma is concerned; but such preparations are far from pure, even electrophoretically. An extremely broad migrating peak is obtained. This indicates the presence of a large number of components differing only slightly in mobility. The technique of reversible boundary spreading also shows this fraction to be highly inhomogeneous. Some subfractionation has been accomplished by the electrophoresis-convection technique.

Functionally, too, this fraction is far from homogeneous, in that it contains antibodies to the various infectious diseases to which immunity has been developed in the animal. Whether all of the γ -globulin consists of specific antibodies or whether there is inactive, nonspecific globulin present is not clearly known. The various antibodies are so closely related chemically that it does not appear possible to separate them by the usual techniques. Theoretically they should be separable through isolation from the specific precipitates formed upon addition of various antigens (Chap. 23).

Further evidence for the inhomogeneity of γ -globulin is obtained in the ultracentrifuge, about 75 per cent of the material having a sedimentation constant of 7S, the remainder sedimenting more rapidly, about 10S. The main component is calculated to have a molecular weight of about 150,000 and dimensions of about 235 by 44 A. (prolate ellipsoidal model); the component of higher sedimentation constant may have a molecular weight about twice as great. By means of the ultracentrifuge it has been demonstrated that γ -globulin can be degraded with pepsin to molecules of about half the usual molecular weight without serious impairment of antibody activity. This finding is of great significance with regard to protein structure and the nature of serological reactions.

α - and β -Globulins. Perhaps even more heterogeneous than the γ -globulin component are the α - and β -components. Included in these electrophoretic fractions is a broad group of proteins, several of which possess known enzymatic or other biological activity. As previously mentioned, these can be resolved, under suitable conditions of electrophoresis, into subgroups called α_1 and α_2 , β_1 and β_2 .

Practically all of the lipids of plasma are found to be associated with the α - and β -globulin components. Carefully purified albumin and γ -globulin are essentially lipid-free. α_1 -Globulins and β_1 -globu-

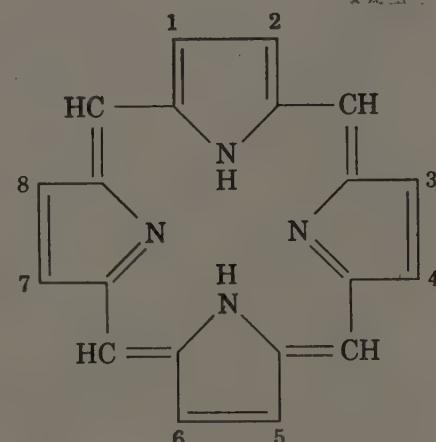
lins that are rich in lipid have been isolated. These have molecular weights of approximately 200,000 and 1 million respectively. The β -lipoprotein contains 75 per cent lipid, almost half of which is cholesterol, but much phospholipid is also present. This protein was first noticed, and originally named "X-protein," because of its unusual behavior in ultracentrifuge studies of plasma. Owing to its high lipid content it has a density near that of water, hence its sedimentation constant is strongly dependent on the salt concentration (density) of the medium. In spite of its dominant lipid content it is soluble in dilute salt solution. These lipoproteins are easily denatured, with loss of lipid, by exposure to high alcohol concentration or by freezing. It appears that one important function of plasma proteins is solubilization and transport of lipid.

There are also α_1 -, α_2 -, β_1 -, and β_2 -globulins that are essentially free of lipid. Several interesting proteins or protein fractions have been isolated from this group. One of these is a β_1 -globulin which has an unusual tendency to combine with metal ions, especially iron (either ferrous or ferric) but also with copper and zinc. This protein, siderophilin, has been prepared by ethanol fractionation (fraction IV) and crystallized. It is calculated that plasma contains approximately 2.4 gm. per liter of the protein, capable of combining 3 mg. of iron.

A glycoprotein having the electrophoretic mobility of an α_1 -globulin has been isolated from fraction VI of plasma. It is homogeneous both electrophoretically and in the ultracentrifuge, and has been crystallized.

HEMOGLOBIN

Heme Proteins. The most important protein constituent of mammalian erythrocytes, hemoglobin, is one of the more readily crystallizable proteins and one which has played an important role in the development of protein chemistry. Hemoglobin is representative of a broad class of iron-containing respiratory proteins known as the *heme proteins*, some form of which is found in almost all aerobic organisms. These proteins expedite in some way, either through storage or transport of oxygen or through catalysis, the oxidative processes of the organism. They are characterized by an iron-porphyrin prosthetic group. Porphyrins are derivatives of porphin, shown here.



The various porphyrins differ according to the nature of the substituents at the numbered positions.⁵ In the ferroheme complex the two nitrogen-bonded hydrogens are replaced by the metal ion. The net charge on the structure is zero if ferrous iron is incorporated, owing to the replacement of the two protons, or plus one if ferric iron is involved.

Preparation. No elaborate purification process is required prior to the crystallization of hemoglobins, since in a given normal adult individual the protein normally appears to be quite homogeneous. Electrophoretically the whole cell contents yield, after centrifugation to remove stroma and cell debris, a single component with only a trace of a much faster constituent, which is possibly residual plasma protein. Many hemoglobins have been crystallized, the ease and method of crystallization and the crystal habit differing widely from species to species. Some hemoglobins are insoluble at the isoelectric point and may be crystallized from water, for example by removal of salt through electrodialysis. In other cases the protein is more soluble and more elaborate procedures are involved.

Properties. Hemoglobin is readily split by acidification, into its components *heme* or *hematin*, the iron-porphyrin prosthetic group, and *globin*, the protein moiety. Heme contains iron in the ferrous state and tends to go over very readily to the more stable and readily crystallizable *hemin*, the ferric analogue of heme. If the separation is made with care, so that excessive denaturation of the globin does not result, the hematin may be recombined with the protein to yield a hemoglobin similar to, though probably not identical with, the native protein.

The combination between heme and globin is called *methemoglobin* or *ferrihemoglobin*. The reconstituted methemoglobin is probably not identical with that obtained by oxidizing hemoglobin directly to methemoglobin. This latter oxidation can be carried out readily with various mild oxidizing agents, but not with molecular oxygen. The reversible combination with oxygen is a distinct process and does not involve oxidation of the ferrous iron; rather the oxygen is held by means of secondary forces, the product being termed *oxyhemoglobin*. One mol. of oxygen is combined per mol. of iron in this structure.

Significant differences in amino acid compositions and in terminal residues have been demonstrated for hemoglobins from various species. The iron content of hemoglobin preparations is invariably near 0.35 per cent, from which one can calculate a minimal molecular weight of 16,700. This is probably one of the few cases where a calculated minimal molecular weight of a protein is trustworthy. In all cases

the true molecular weight of hemoglobin appears to be four times the minimal value, that is, there are actually four heme residues per molecule. A molecular weight of very nearly 67,000 for horse hemoglobin has been obtained from sedimentation and diffusion measurements, from osmotic pressure, and more recently from X-ray diffraction results. Other hemoglobins have very nearly this same molecular weight.

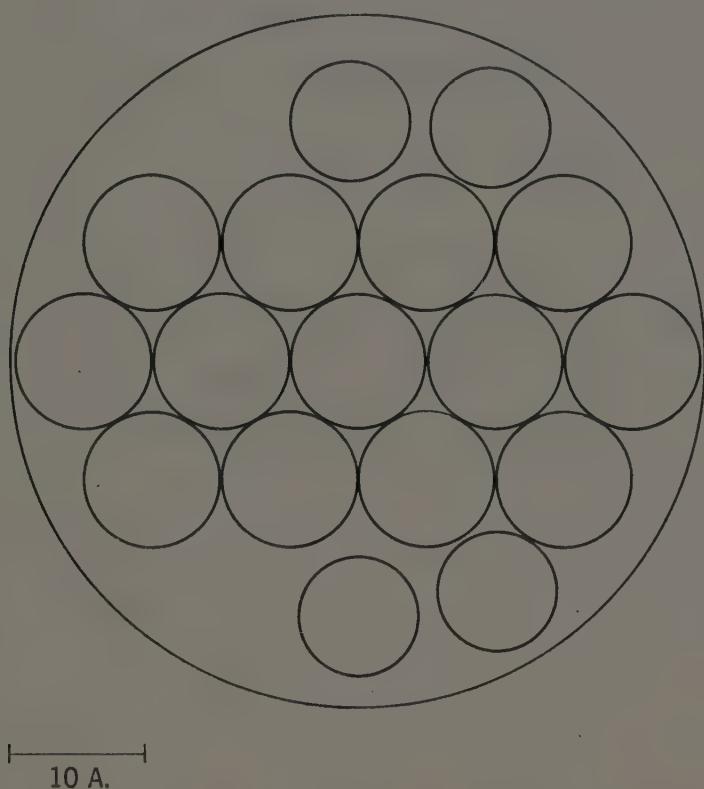


Fig. 18-5. Schematic representation of cross section of hemoglobin molecule. Circles represent rodlike units of high electron density running in the A direction. From W. L. Bragg, E. R. Howells, and M. F. Perutz, *Acta Cryst.*, 5, 136 (1952).

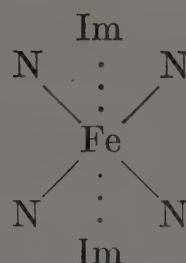
The investigation of horse methemoglobin crystals by means of X-ray crystallography has led to some of the most far-reaching conclusions yet drawn with regard to the structure of globular proteins. These results are in accord with, although they do not entirely prove, the following structure for the hemoglobin molecule. In outline the molecule may be approximated as a slightly prolate ellipsoid of length 65 Å. and width 55 Å. The crystals are heavily hydrated, but the size and shape of the molecule does not alter much with variation in water content; most of the water must exist intermolecularly, and the molecule per se must be rather rigid. The electron density appears to be concentrated in rodlike units running parallel to the long axis (the A axis of the monoclinic crystal). These rodlike units appear

to exist in layers, as indicated schematically in Fig. 18-5. The diameter of these units, approximately 10 Å., is such as to suggest the possibility of their representing coiled peptide chains of the α -helix type. It should be pointed out, however, that the Fourier projections indicate that at best they could be only poorly formed helices with frequent bends or turns. Calculations show that they could not continue as straight, uninterrupted helices for more than about 16 Å. or one fourth the length of the molecule.

It has been demonstrated that the hemoglobin molecule dissociates into two apparently similar subunits (molecular weight 34,000) upon denaturation by urea. The globin moiety of hemoglobin has also been shown to have a molecular weight of only about half that of the parent molecule. More recently evidence has been presented, based on sedimentation and diffusion, that human hemoglobin dissociates reversibly below pH 3.5 into half-size subunits, and horse hemoglobin into four subunits. In the later study it was shown that oxidation does not further reduce the molecular weight, indicating that there are no —S—S— cross-links between peptide chains. Human hemoglobin has been found to contain five valyl N-terminal residues, horse hemoglobin six (Table 16-1).

The hemoglobins have isoelectric points near 7.0. The isoelectric points of the globins, which are very rich in histidine, are usually on the alkaline side of 7.0; they are classified as histones. The amphoteric properties of hemoglobin and oxyhemoglobin are of the greatest interest. While the titration curves of the two proteins are similar they differ significantly in that the curve for oxyhemoglobin is shifted about 0.2 pH units to the acid side in the physiological region (pH 7.0 to 7.5), and in a somewhat less degree to the alkaline side between pH 5 and 6. In other words, under physiological conditions, hemoglobin is rendered a stronger acid upon combination with oxygen. Conversely, the oxygen affinity of hemoglobin increases with increasing pH above pH 6, and decreases with increasing pH below this value. These effects, called the Bohr effects after their discoverer, are of the greatest significance physiologically, though a detailed discussion of the physiological aspects of hemoglobin is beyond the range of the present treatment.

These so-called "linked functions" result from the fact that groups involved in oxygen binding and acid-base equilibrium are not entirely independent. This effect has been explained as being due to two imidazole groups, contributed by histidine residues of the protein, which serve to bind the iron-porphyrin group to the protein in the following way.



The four nitrogens belong to the porphyrin ring and lie in the plane of the paper, and the two imidazole groups, attached to the protein, lie above and below this plane.

In the case of unoxygentated hemoglobin, the dashes in this structural representation should be considered as indicating spatial relationships only, since it appears that the Fe^{++} ion is held by ionic forces with little or no covalent character. This is based on the high paramagnetic susceptibility of hemoglobin, which indicates the presence of several unpaired electrons, probably four per heme unit. Upon combining with oxygen, the bonds become essentially covalent, as indicated by a transition from paramagnetic to diamagnetic behavior. Of the two imidazole groups, one is much more remote from the Fe^{++} ion and should probably not be considered as bonded to it at all. The oxygen apparently combines at this sixth site around the Fe^{++} , thus shielding the nonbonded imidazole group and shifting its pK from about 5.25 to 5.75. At the same time, the formation of a true covalent-type bond to the other imidazole group results in a more pronounced shift in the acid direction, about 7.9 to 6.7.

The four heme groups in hemoglobin are not independent. Oxygenation of one of them leads to an enhancement of the oxygen affinity of the others. Evidence has been brought forward to indicate that the four heme groupings are intrinsically identical, and that there is complete equivalence with respect to their interactions and hence, presumably, with respect to their spatial relationships.

Hemoglobin also combines with carbon monoxide, hydrogen sulfide, and many other compounds. The mechanism of combination is believed to be very similar to that for oxygen, and in some cases the binding is even more tenacious.

Hemoglobins from different species, though having many properties in common, differ with respect to serological specificity, physical properties, and even in amino acid composition. There are pronounced differences in solubility and in crystal habit, which can be used as a means of identification. Similarly there are well-established differences between the hemoglobins of young and adult in the same species.

An intriguing example of the genetic control of protein composition has been brought to light in the case of the hemoglobin of persons

afflicted with a disease known as "sickle-cell anemia." This form of anemia, which is known to be a heritable trait and is prevalent among the members of the Negro race, is characterized by the fact that the erythrocytes of the victims change in shape when the blood is subjected to a diminished oxygen partial pressure. The hemoglobin from such individuals differs from the normal in having a somewhat higher isoelectric point, and the two forms can be separated electrophoretically. It has been concluded that the difference arises from the probable fact that the abnormal hemoglobin possesses 2 to 4 more cationic amino acid residues per molecule than the normal. A number of other "abnormal" hemoglobins identified in man have been shown to be under genetic control.

SELECTED REFERENCES

Books and Reviews

- "Blood, Blood Derivatives, and Blood Substitutes," E. J. Cohn, *Proc. Am. Phil. Soc.*, **88**, 159 (1944).
- "The Plasma Proteins and Their Fractionation," J. T. Edsall, *Advances in Protein Chem.*, **3**, 383 (1947).
- "Heme Proteins," J. Wyman, Jr., *Advances in Protein Chem.*, **4**, 407 (1948).
- "The Plasma Proteins in Disease," A. B. Gutman, *Advances in Protein Chem.*, **4**, 155 (1948).
- "Chemistry and Clinical Uses of the Protein Components Involved in Blood Clotting," J. T. Edsall, *Ergeb. Physiol., biol. chem. u. exptl. Pharmakol.*, **46**, 354 (1950).
- "Symposium on Chemistry of Prothrombin and Fibrinogen" (review papers by various outstanding workers in the field), *Physiol. Revs.*, **34**, 711-760 (1954).
- "The Proteins of Blood Plasma and Lymph," W. L. Hughes in Neurath and Bailey's *The Proteins*, Vol. IIB, p. 663, Academic Press, New York, 1954.
- "Nature of the Fibrinogen-Thrombin Reaction," K. Bailey and F. Bettelheim, *Brit. Med. Bull.*, **11**, 50 (1955).

Original Articles

- "Preparation and Properties of Serum and Plasma Proteins. I. Size and Charge of Proteins Separating upon Equilibration across Membranes with Ammonium Sulfate Solutions of Controlled pH, Ionic Strength and Temperature," E. J. Cohn, T. McMeekin, J. Oncley, J. Newell, and W. Hughes, *J. Am. Chem. Soc.*, **62**, 3386 (1940).
- "Crystallization of a Carbohydrate-containing Albumin from Horse Serum," T. McMeekin, *J. Am. Chem. Soc.*, **62**, 3393 (1940).
- Fetuvin, a New Globulin Isolated from Calf Serum*, K. O. Pedersen and The Svedberg, Almquist & Wiksell, Uppsala, 1944.
- "An Electrophoretic Survey of Various Animal Plasmas," H. Deutsch and M. Goodloe, *J. Biol. Chem.*, **161**, 1 (1945).

"Effect of Protein Depletion on Plasma Proteins in the Dog Measured by Electrophoretic Analysis," B. F. Chow, J. B. Allison, W. H. Cole, and R. D. Seeley, *Proc. Soc. Exptl. Biol. Med.*, **60**, 14 (1945).

"Physical-Chemical Characteristics of Certain of the Proteins of Normal Human Plasma," J. L. Oncley, G. Scatchard, and A. Brown, *J. Phys. & Colloid Chem.*, **51**, 184 (1947).

"Electrophoretic Studies on Human Serum Albumin," C. Hoch-Ligeti and H. Hoch, *Biochem. J.*, **43**, 556 (1948).

"The Free Amino Groups of Hemoglobins," R. Porter and F. Sanger, *Biochem. J.*, **42**, 287 (1948).

"Sickle Cell Anemia: A Molecular Disease," L. Pauling, H. A. Itano, S. J. Singer, and I. C. Wells, *Science*, **110**, 543 (1949).

"Preparation and Properties of an Acid Glycoprotein from Human Plasma," K. Schmid, *J. Am. Chem. Soc.*, **72**, 2816 (1950).

"Composition and Properties of Human Serum β -Lipoprotein," J. L. Oncley, F. R. N. Gurd, and M. Melin, *J. Am. Chem. Soc.*, **72**, 458 (1950).

"A System for the Separation of the Components of Human Blood: Quantitative Procedures for the Separation of the Protein Components of Human Plasma," E. J. Cohn, *et al.*, *J. Am. Chem. Soc.*, **72**, 465 (1950).

"Further studies on the Oxygen Equilibrium of Hemoglobin," D. W. Allen, K. F. Guthe, and J. Wyman, Jr., *J. Biol. Chem.*, **187**, 393 (1950).

"Composition of Baby Pig Plasma and Sow's Whey during Lactation," J. F. Foster, R. W. Friedell, D. Catron, and M. R. Dieckmann, *Arch. Biochem. and Biophys.*, **31**, 104 (1951).

"An Intermediate in the Conversion of Fibrinogen to Fibrin," S. Shulman, P. Ehrlich, and J. D. Ferry, *J. Am. Chem. Soc.*, **73**, 1388 (1951).

"Light Scattering Studies on Fibrinogen," R. Steiner and K. Laki, *Arch. Biochem. and Biophys.*, **73**, 882 (1951).

"The Combining Power of Hemoglobin for Alkyl Isocyanides, and the Nature of the Heme-Heme Interaction in Hemoglobin," R. St. George and L. Pauling, *Science*, **114**, 629 (1951).

"X-ray and Solubility Studies of the Hemoglobin of Sickle Cell Anemia Patients," M. Perutz, A. Liquori, and F. Eirich, *Nature*, **167**, 929 (1951).

"Dimerization of Serum Albumin in Presence of Mercurials. I," H. Edelhoch, E. Katchalski, R. Maybury, W. Hughes, Jr., and J. Edsall, *J. Am. Chem. Soc.*, **75**, 5058 (1953).

"Transformation of Fibrinogen into Fibrin. I and II," E. Mihalyi, *J. Biol. Chem.*, **209**, 723, 733 (1954).

"Polymerization of Fibrinogen," J. D. Ferry, *Physiol. Revs.*, **34**, 753 (1954).

"The Site of Attack by Thrombin on Fibrinogen," I. Tinoco, Jr., and J. D. Ferry, *J. Am. Chem. Soc.*, **76**, 5573 (1954).

"The Optical Rotation of Human Serum Albumin and γ -Globulin," B. Jirgensons and S. Sirotsky, *J. Am. Chem. Soc.*, **76**, 1367 (1954).

"Ability of Normal and Sickle-Cell Hemoglobin to Form Complexes," C. Anderson and S. Griffiths, *Nature*, **174**, 928 (1954).

"Electrical Birefringence of Fibrinogen and Activated Fibrinogen," I. Tinoco, Jr., *J. Am. Chem. Soc.*, **77**, 3476 (1955).

"Dissociation of Human Hemoglobin at Low pH," E. Field and J. O'Brien, *Biochem. J.*, **60**, 656 (1955).

"The Isolation of a Rat Plasma Albumin Which is Electrophoretically Homogeneous," J. D. Ferry, *J. Am. Chem. Soc.*, **77**, 3476 (1955).

geneous at Low pH Values," A. Keltz and J. Mehl, *J. Am. Chem. Soc.*, **77**, 5764 (1955).

"Subunits in the Molecule of Bovine Plasma Albumin," M. Reichmann and J. Colvin, *Can. J. Chem.*, **33**, 163 (1955).

"Sulphydryl Groups in Hemoglobins," V. Ingram, *Biochem. J.*, **59**, 653 (1955).

"The N-Terminal Residues of Bovine Plasma Clot," W. Middlebrook, *Biochem. J.*, **59**, 146 (1955).

"Abnormal Human Serum Globulins," F. Putnam, *Science*, **122**, 275 (1955).

"A Study of the Oxygen Equilibrium of Horse Hemoglobin," S. Takashima, *J. Am. Chem. Soc.*, **77**, 6173 (1955).

"Adsorption of Bovine Serum Albumin on Glass," H. Bull, *Biochim. et Biophys. Acta*, **19**, 464 (1956).

"Heterogeneity of Serum Albumins," R. Benesch and R. Benesch, *Federation Proc.*, **15**, 218 (1956).

"The Number of Subunits in the Molecule of Horse Hemoglobin," M. Reichmann and J. Colvin, *Can. J. Chem.*, **34**, 411 (1956).

"Hemoglobin J," O. Thorup, H. Itano, M. Wheby, and B. Leavell, *Science*, **123**, 887 (1956).

"Characteristic Electrophoretic Patterns of Plasma Proteins of Orders of Amphibia and Reptilia," H. Dessauer and W. Fox, *Science*, **124**, 225 (1956).

"Electrophoretic Demonstration of the Isomerization of Bovine Plasma Albumin at Low pH," K. Aoki and J. F. Foster, *J. Am. Chem. Soc.*, **78**, 3538 (1956).

"Ultracentrifugal Studies of Major Non-lipide Electrophoretic Components of Normal Human Serum," G. Wallenius, R. Trautman, H. Kunkel, and E. Franklin, *J. Biol. Chem.*, **225**, 253 (1957).

19

Some Notable Protein Systems

Egg white proteins

Milk proteins

Muscle proteins

Plant proteins

This chapter is concerned with the properties of some of the other important proteins and protein systems, study of which has contributed immeasurably to our present knowledge of protein chemistry. Proteins such as ovalbumin (of egg white), β -lactoglobulin (of milk), and edestin (of hemp seed) have been favorites of the protein chemist because of their relative ease of crystallization, availability, and supposed purity. The systems from which they are obtained, while not possessing the great biological interest of blood plasma, are nevertheless of considerable interest.

EGG WHITE PROTEINS

In Fig. 19-1 is shown a typical electrophoretic pattern of whole egg white at pH 7.8. The principal components, indicated by letters in the diagram, are resolved perhaps better at this pH than at any other. The per cent composition as estimated from the electrophoretic diagram is given in Table 19-1. Limited attempts have been made to evolve a comprehensive fractionation scheme analogous to that developed for the plasma proteins.

Ovalbumin. The leading boundaries in Fig. 19-1, designated A_1 and A_2 , correspond to the water-soluble or albumin fraction. This

readily crystallizable protein, long considered a "paragon of virtue" among proteins, is obviously not homogeneous, even by the electrophoretic criterion. The two electrophoretic components persist in

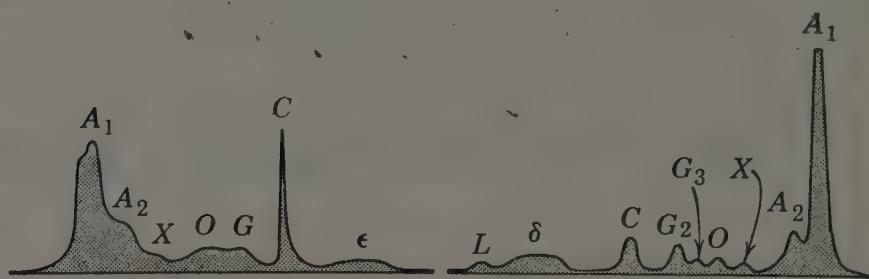


Fig. 19-1. Electrophoretic pattern of whole egg white at pH 7.8 in a phosphate buffer. The lysozyme component, *L*, migrates cathodically; all of the other protein components are anionic under these conditions. The *L* peak in the descending pattern had migrated out of the field at the time of exposure. From R. H. Forsythe and J. F. Foster, *J. Biol. Chem.*, **184**, 377 (1950).

preparations recrystallized many times. Indeed, a third component of still slower mobility is claimed (labeled *X* in the figure). Together the ovalbumin components comprise some 65 per cent of the total protein in the white.

TABLE 19-1. Protein Composition of Egg White

Component	Per Cent of Total Protein ¹
Ovalbumin	64.3
Conalbumin	13.6
Ovomucoid	9.1
Lysozyme (<i>G</i> ₁)	3.4
Globulins (<i>G</i> ₂ plus <i>G</i> ₃)	8.6
Mucin	1.1 ²
Avidin	0.06 ²

¹ Determined by electrophoretic analysis. From R. H. Forsythe and J. F. Foster, *J. Biol. Chem.*, **184**, 377 (1950).

² These components are not detected in the electrophoretic pattern. The values given were determined by chemical means.

Ovalbumin may be prepared readily by removing the globulins by precipitation in half-saturated ammonium sulfate, and then crystallizing by adding ammonium sulfate to incipient turbidity and reducing the *pH* to approximately the isoelectric point of the albumin, *pH* 4.6. Even after repeated recrystallization, ovalbumin invariably contains about 2 per cent carbohydrate; accordingly, it is usually classed as a glycoprotein. Its molecular weight is approximately 44,000, as determined by sedimentation and diffusion, and it has a frictional coefficient of only 1.16, indicating little deviation from sphericity.

Ovalbumin is one of the most unstable of all proteins. Some loss of solubility occurs during lyophilization, and there is further loss in storage at 0°. In spite of this, ovalbumin has been one of the favorite subjects of investigation by protein chemists, chiefly because of its ready availability and ease of preparation. This situation has been altered more recently with improved methods for preparation of serum albumin.

Ovalbumin contains all of the common amino acids; in particular, adequate amounts of all that are dietary essentials. Nutritionally it is rivaled by only one other protein, as far as is known, namely lactalbumin (β -lactoglobulin). The nutritional excellence of these proteins is perhaps to be expected on theological grounds in view of their principal if not sole function, namely nutrition of the young animal.

In fresh egg white, component A₁ comprises some 80 per cent of the total albumin. On storage there is a gradual decrease in A₁ and increase in A₂. These very difficultly separable proteins are now thought to differ only in their phosphorus content, A₁ containing two esterified phosphoric acid residues and A₂ only one. Indeed A₁ can be converted into A₂ by means of a phosphatase-type enzyme (Perlmann, 1950). The A₃ component may correspond to the material with all of the phosphate removed.

A crystalline modified form of ovalbumin, called *plakalbumin*, results from the action of *Bacillus subtilis* on ovalbumin. Its formation results from the removal, from the ovalbumin molecule, of 6 or 7 amino acid residues in the form of low-molecular polypeptides. Interestingly enough, plakalbumin consists of two components, P₁ and P₂, in the same proportions as A₁ and A₂ in the parent ovalbumin. These are almost certainly related in the same way as A₁ and A₂.

Conalbumin. Following the crystallization of albumin there remains in solution other highly soluble protein. Part of this is heat-coagulable; hence, by definition it is an albumin although it does not crystallize with the remainder of the albumin. This protein has been termed conalbumin.

A common method of preparation of conalbumin has involved heat coagulation of the supernatant solution following removal of the globulins and ovalbumin by salting-out procedures. Heat coagulation is hardly suitable for the preparation of a native protein. When so prepared, the protein is not homogeneous but consists of at least two electrophoretic components. Nearly pure (electrophoretically) conalbumin has been prepared by low-temperature ethanol fractionation

procedures but in relatively poor yields. Although this protein long resisted attempts at crystallization, it has now been crystallized.

Most of the yellow color (riboflavin) of egg white is bound to the conalbumin, or more probably a fraction thereof. Conalbumin has a strong affinity for metallic ions, especially iron, which yields a red complex. By depriving microorganisms of iron, it is able to prevent the growth of certain bacteria. Curiously, it has been claimed that conalbumin is immunologically identical with the serum albumin of the chicken.

Mucin. The gelatinous properties of egg white are attributed to a carbohydrate-rich protein fraction termed *ovomucin*. Closely similar glycoproteins are found in other viscous biological fluids and secretions. This ovomucin fraction, together with considerable of the globulin, can be removed by dilution of egg white with water, by reduction of the pH to 6.0 or below, or by a combination of the two procedures. It may also be removed directly from whole egg white by centrifugation in a Sharples-type centrifuge (50,000 times gravity). This indicates that it is present in a very highly aggregated, supposedly fibrous state.

By any of these means mucin is obtained as a slimy precipitate that cannot be dissolved completely. It contains only about 12 per cent nitrogen, owing to the large amount of carbohydrate. The latter is supposedly a high molecular weight polymer of a hexosamine. Mucin preparations are doubtless far from homogeneous. This component does not manifest itself in the electrophoretic pattern of egg white, the patterns before and after its removal being almost indistinguishable.

Globulins. Electrophoretically there are at least three distinguishable globulin components in egg white, G₁, G₂, and G₃. Of these, G₁ appears to be identical with lysozyme. The other components have not been thoroughly studied. The globulin fraction has been prepared by means of ammonium sulfate precipitation, but this procedure leads to no subfractionation.

Doubtless, if one may extrapolate from the findings in the case of plasma, there must be many interesting proteins represented in this fraction. Application of the ethanol fractionation techniques should be fruitful.

Lysozyme. One of the best-defined proteins of egg white is lysozyme, apparently the G₁ electrophoretic component. It is unusual in that it can be crystallized directly from the whole white. This is accomplished by adding salt to about 5 per cent and adjusting the pH to about 9.6. Lysozyme is one of the smallest known proteins,

its molecular weight being only about 13,000 to 14,000 as determined by a variety of methods.

This protein is interesting biologically because of its ability to break down (lyse) the cells of certain bacteria, whence its name. The lysing reaction probably involves the hydrolysis of a complex polysaccharide in the bacterial cell wall. Lysozymes are present in other biological fluids, notably tears, but egg white is one of the richest sources.

The isoelectric point of lysozyme obtained from egg white is very high, near pH 11. In egg white at the normal pH, 7.2 to 7.5, it is cationic, whereas the other proteins are anionic. As a consequence it migrates backward in the electrophoretic pattern (Fig. 19-1). One method of preparation, adsorption on bentonite, depends on the fact that lysozyme is the only cationic protein in egg white. Due to its relatively low molecular weight, ease of preparation, ready crystallizability in several forms (as various salts), and enzymatic activity, lysozyme appears to be a most promising protein for intensive investigations of protein structure.

Ovomucoid. Following the crystallization of ovalbumin by salting-out and the removal of conalbumin by heat coagulation, there remains in solution another nitrogen-containing fraction. In spite of its atypical properties (in addition to being non-heat-coagulable, it is not precipitated by trichloroacetic acid or other general protein precipitants), this fraction is considered to be a protein. It is nondialyzable, hence evidently of high enough molecular weight to be classed as a protein. It contains about 10 per cent carbohydrate, which is composed of glucosamine, mannose, and galactose. Ovomucoid is usually isolated after removal of all other protein, by adding alcohol or acetone to a very high concentration, usually 3 volumes of precipitant to 1 of solution.

Electrophoretically this protein migrates as a single component, but it shows pronounced reversible boundary spreading, indicating inhomogeneity. It is of considerable biological interest because the ability of egg white to inhibit trypsin action (the trypsin inhibitor) is evidently concentrated in this fraction.

MILK PROTEINS

Milk provides another interesting and important protein system. The protein content of milk varies widely with the species, ranging from as low as about 1.6 per cent in the human to as high as 10 per cent in some other mammals. Cow's milk ranges from about 3.0 to 4.0 per cent. The fact that the amino acid composition of milk

proteins is very well balanced nutritionally is not at all surprising from an evolutionary standpoint, in view of the highly specialized function of this fluid.

Casein. The major protein constituent of the milk of all species is a phosphoprotein, or phosphoglycoprotein, called casein. The remarks in this section will pertain specifically to bovine casein, on which most of the work has been done; however, caseins from different species are remarkably similar both in gross chemical composition and in physical properties. It has even been claimed that caseins from widely different species are not differentiable immunologically.

Casein does not occur in true solution, but rather in the form of large aggregates or micelles of indefinite size, ranging up to 2000 Å. or more in diameter and averaging perhaps 1000 Å. The nature of the binding forces involved in these micelles is not clear, but evidently they involve calcium ions and probably phosphate groups. The aggregates have been termed "calcium phosphocaseinate." One view is that they consist of calcium phosphate stabilized by a monolayer of calcium caseinate.

Upon lowering the *pH* of milk to about the isoelectric point, the casein suspension flocculates, yielding the familiar milk curd. The calcium and phosphate ions are largely split off and remain in the supernatant liquid, the whey. Casein can also be precipitated by adding salt or alcohol, by electrodialysis, which reduces the *pH*, or by means of the specific casein-coagulating enzyme *rennin*. It has also been removed by direct high-speed centrifugation in a Sharples-type centrifuge, yielding a product which more nearly resembles the native material in that the bound calcium and phosphate are not removed.

As a result of the loss of calcium, the solubility properties of acid-precipitated casein are markedly altered. It is only slightly soluble in water at its isoelectric point (*pH* 4.6), but solubility increases rapidly on either side of this *pH* and the resultant solutions are reasonably clear. A large proportion of the material appears to have a molecular or particle weight of 75,000 to 100,000, but it is very heterogeneous. Furthermore, the sedimentation behavior in the ultracentrifuge depends on the method of preparation. It seems possible that strong aggregation tendencies may still exist, and one does not feel entirely justified in speaking of casein "molecules."

Electrophoretically, casein is not homogeneous but exhibits two well resolved main components, designated α - and β - . These differ somewhat in isoelectric point, namely 4.1 and 4.5 at 0.1 ionic strength, and may be separated by isoelectric precipitation. Neither fraction

is entirely homogeneous electrophoretically. A third component, γ -casein, has been isolated by alcohol fractionation.

Casein is produced commercially in large volumes, and is one of the most important proteins from the industrial point of view.

β -Lactoglobulin. Following the removal of casein by any of the methods mentioned, there remains a reasonably clear protein solution, the whey. One fraction, comprising some two thirds of the protein in normal whey, is a readily crystallizable globulin designated β -lactoglobulin. Formerly a crystalline albumin, lactalbumin, was claimed, but it is now thought that in reality this was an impure form of β -lactoglobulin. Because of its assumed homogeneity (it is one of the few proteins for which adherence to the phase rule in solubility studies is claimed), it has been a favorite subject for protein investigators.

It has now been found that although bovine β -lactoglobulin is homogeneous electrophoretically near its isoelectric point (pH 5.2), it can be resolved into two or more components at pH values on either side of this pH , 4.8 and 6.5. Again at pH 8.4 it is electrophoretically homogeneous. The relative amounts of the components may be altered by repeated recrystallization, apparently without alteration in the crystal structure. The so-called β_1 component, the component of lower mobility at pH 4.8, has been prepared in an electrophoretically homogeneous state by alcohol fractionation.

The molecular weight of whole β -lactoglobulin is very nearly 40,000 as determined by various methods, including sedimentation and diffusion and X-ray diffraction analysis of its crystals. Since its frictional coefficient is only 1.26, it evidently does not deviate greatly from sphericity.

Immune Lactoglobulin. Colostral whey is much richer in total protein than whey produced in the later stages of lactation. Figure 19-2 shows that the distribution of protein components is also markedly different, the colostral whey being characterized by the presence of relatively large amounts of slow-moving protein. This latter has been demonstrated, in the case of the cow, to contain immune bodies similar to those of the maternal plasma, a point which is of great significance with regard to the passive immunity transmitted to the young.

Following parturition, the electrophoretic pattern of the whey changes within a few days to the "normal" pattern; however, this too contains immunologically active globulins of both the euglobulin and pseudoglobulin types. Each of these has been isolated in relatively pure (electrophoretically) form in significant quantities, total-

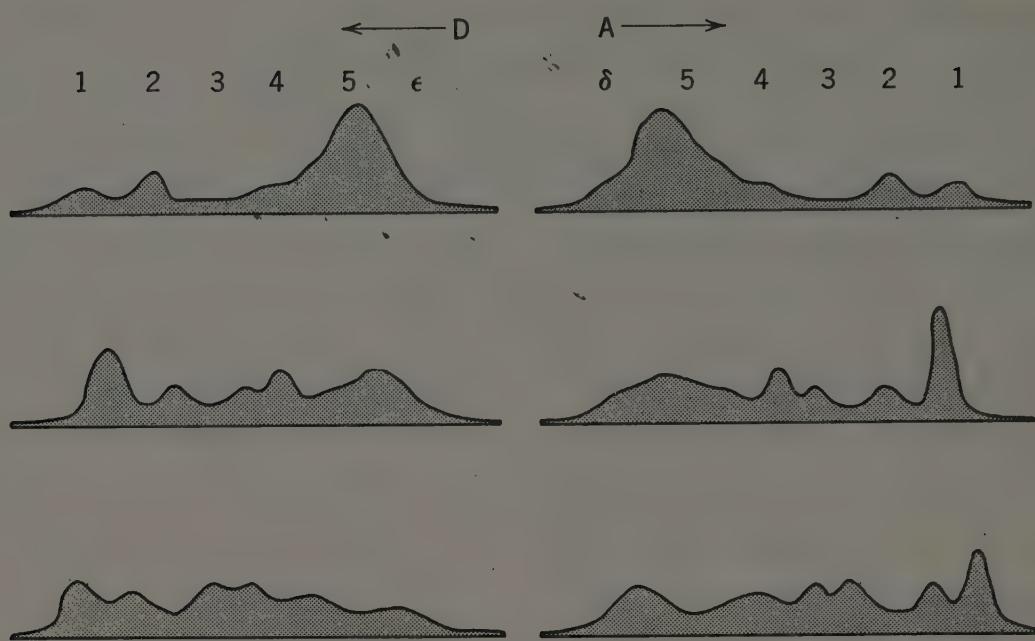


Fig. 19-2. Electrophoretic patterns of swine whey at three stages of lactation: top, immediately following parturition; center, 24 hours post partum; bottom, weaning. Note the large amount of slow-moving protein, component 5, at parturition. From J. F. Foster, R. Friedell, D. Catron, and M. Dieckmann, *Arch. Biochem. and Biophys.*, **31**, 104 (1951).

ling about 10 per cent of the whey protein. Each of these proteins is separate and distinct from the β -lactoglobulin.

MUSCLE PROTEINS

The dry-matter of muscle, in common with most biological tissues, is composed primarily of protein. Since muscle gives the higher living organisms one of their most distinguishing properties, self-locomotion, it is not surprising that the proteins of muscle have attracted much interest. Since one is here dealing with a tissue and a large part of the protein is structural in nature, the problem of extraction is considerable. Nevertheless, it is possible to extract as much as 90 per cent of the total protein of fresh muscle, using aqueous salt solutions.

Myosin, Actin, and Actomyosin. The extraction of muscle with moderately concentrated salt solutions, for example potassium chloride, yields a globulin of most unusual properties. Solutions of this protein have a high viscosity and show intense double refraction of flow, both properties indicating that the protein molecule is highly asymmetric. This protein was originally called myosin, but more recently it has been found to consist of two very distinct proteins. One of these has been crystallized and also named *myosin*. The other component has been called *actin*, and the complex is now termed *actomyosin*.

This introduces some confusion into the nomenclature of these proteins. In this discussion we will follow the newer terminology, using the term myosin to apply to the one component of actomyosin.

Myosin yields clear solutions in water, and may be crystallized at a pH of approximately 6.5 in the presence of 0.025 M potassium chloride. These are not actually three-dimensional crystals but are of the anisotropic liquid-crystal type, consisting of elongated myosin rodlets oriented laterally. Myosin is found to be homogeneous electrophoretically, with an isoelectric point of approximately 5.4 when the only cation present is potassium ion. When divalent cations such as Ca^{++} or Mg^{++} are added, the isoelectric point is found to shift upward very markedly, indicating a strong tendency for this protein to bind such cations. It is also unusually sensitive to even monovalent cations.

Myosin has a molecular weight of about one million, as determined by means of sedimentation and diffusion. It is highly asymmetric, its length being presumably of the order 2000 Å. Urea apparently has both a folding and a disaggregating influence on myosin. Short treatments with concentrated urea lead to a molecule of approximately the same molecular weight but of much lower asymmetry. Prolonged treatment leads to disaggregation.

Actin is unusual in that it can apparently exist in two forms, a globular form called *G-actin* and a fibrillar form called *F-actin*. In contrast to that of myosin the molecule of G-actin is relatively small, molecular weight about 75,000, and it is nearly spherical. In the presence of appreciable concentrations of salt, in particular divalent cations, G-actin tends to aggregate or polymerize into the fibrillar form, cations being bound in the process.

If solutions of myosin and F-actin are mixed, there is a sudden rise of viscosity and the solution exhibits strong streaming birefringence. The electron microscope indicates the presence of very long threads, not unlike the fibrils of muscle. Since the F-actin threads are very long even before the addition of myosin, this increase in viscosity has been interpreted as resulting from an increased stiffness in the complex. Combination of G-actin and myosin does not yield solutions with this characteristic high viscosity and intense double refraction of flow.

Addition of adenosinetriphosphate (ATP) to actomyosin solution, under the proper conditions of ionic concentration, leads to a very sharp reduction in viscosity and loss of birefringence of flow. This phenomenon is thought to be closely related to muscular contraction itself. Since ATP is a high-energy compound built up in respiration

and glycolysis, this phenomenon provides a possible means for utilizing metabolic energy in muscular work.

Other Muscle Proteins. Besides the proteins already mentioned that are presumed to be directly concerned in the contraction of muscle, many others are present. During activity, muscle is the site of intense metabolism, and it is therefore not surprising that many enzymes are present. Several of these have been isolated, perhaps the best known being *phosphorylase*, which has been prepared in the crystalline form. It is the enzyme which catalyzes the breakdown of high molecular weight polysaccharide (glycogen) to yield glucose-1-phosphate, the process being known as phosphorolysis. The ability of phosphorylase to catalyze the reverse reaction, namely the synthesis of polysaccharide from glucose-1-phosphate, has also been demonstrated.

Extraction of muscle with water yields a complex of proteins called *myogen*. Solubility studies of myogen as a function of pH indicate the presence of at least eight protein species, having different isoelectric points. Two crystalline fractions have been prepared, which are called *myogen A* and *myogen B*. Crystalline myogen A appears to be homogeneous, both electrophoretically and in the ultracentrifuge. Nevertheless, it has been found to have two different enzymic activities, *aldolase* and *glycerophosphate dehydrogenase*. These activities are not associated with the same protein, since it has been found possible to separate the glycerophosphate dehydrogenase activity by fractional crystallization. This is an extremely interesting example of the inadequacy of crystallization, electrophoresis, and ultracentrifugal analysis as criteria of homogeneity.

Muscle also contains a heme pigment similar to hemoglobin, called *myoglobin*. The molecular weight of myoglobin is about one fourth that of hemoglobin, and it contains only a single iron-porphyrin group.

PLANT PROTEINS

Seed Globulins. Although generally less proteinaceous than typical animal tissues, plant tissues are rich sources of proteins. In large part these proteins function structurally and are consequently difficult to extract. In the case of the dicotyledonous seeds, however, an appreciable amount of protein can be readily extracted by means of dilute salt solution, and a large number of interesting proteins, predominantly of the globulin type, have been crystallized from such extracts. These are thought to serve as food reserves, though several crystalline enzymes have been prepared from such sources. It seems

reasonable to assume that many of the crystallized seed proteins may have as yet unsuspected enzymatic activity.

A few of the more noted seed globulins should be mentioned. Perhaps the best known is *edestin*, which can be prepared readily in crystalline form by extraction of hemp seed with dilute saline, followed by adjustment of the pH and dialysis. This protein has a molecular weight of approximately 310,000, as determined by sedimentation and diffusion analyses, and is nearly spherical in shape (axial ratio less than 4). Upon denaturation with urea it is split into fragments having a molecular weight of about 50,000. Edestin has received much attention because it is easy to prepare and to crystallize. Consequently, analytical data on its amino acid composition is reasonably complete.

The very interesting seed globulin *urease* is of much interest because of the relatively simple reaction which it catalyzes, namely, the hydrolysis of urea. It has not been studied as exhaustively from the physicochemical point of view as has edestin, but it is known to have a molecular weight of nearly 500,000 and to be approximately spherical.

Excelsin and *pomelin* have been prepared in crystalline form from the Brazil nut and orange seed, respectively. The former has a molecular weight of approximately 300,000, as determined both by sedimentation and diffusion analyses and by X-ray diffraction. Crystalline globulins have also been prepared from the coconut and from tobacco seed, as well as from many other seeds.

Cereal Proteins. The cereals have not proven to be such rich sources of readily crystallizable protein. The germ (embryo) of the cereals contains some salt-soluble protein, to be sure, and possibly even small amounts of albumins. The bulk of the protein of the cereal grain exists in the storage organ, the endosperm. The endosperm proteins of corn and wheat especially have been the subject of much study, and those of rye somewhat less. The endosperm proteins are found to fall roughly into two classes, the *prolamins* (alcohol-soluble) and the *glutelins* (soluble in alkali). The distinction between these classes is difficult to make; indeed, there may be no sharp line of demarcation. The prolamins are also soluble in alkali. Extraction of corn by alkali yields some of each class, but neither is extracted quantitatively.

It has been suggested that these may be basically the same proteins, one being modified in some way, either chemically or physically. The amino acid assays available indicate this idea to be untenable, although such assays are very meager, especially in the case of the

glutelins. Others hold to the view that in the endosperm there exists a continuous gradation (spectrum) of proteins; in such case the nature of the product obtained in a given extract would depend on the conditions used.

The best known prolamin is *zein*, obtained from corn (*Zea mays*). This protein is prepared and used commercially in considerable volume. Zein is, for all practical purposes, insoluble in either water or ethanol, but it is soluble in mixtures of ethanol in water ranging between the limits of roughly 50 and 90 per cent. The optimum solubility occurs at approximately 70 per cent by weight ethanol. Zein as prepared from either corn or commercial corn gluten is obviously inhomogeneous, and although much effort has been expended in attempts at fractionation, it is very doubtful if anything approaching a homogeneous preparation has resulted. Study of zein, and of other proteins insoluble in water and salt solution, is handicapped by the inapplicability of the electrophoretic technique. In the case of zein, attempts have been made to circumvent this difficulty by carrying out electrophoresis in ethanol and water, in water at high pH, and in aqueous detergent solutions. All such approaches are subject to theoretical drawbacks, and the results are subject to some question.

Zein first became of interest in connection with nutritional studies because of the almost complete if not total absence of the essential amino acids tryptophan and lysine. As a consequence of this interest, its composition has been much studied and for a long time was perhaps more completely known than that of any other protein. This position has been lost as a result of more recent studies using new and better techniques on such proteins as insulin, β -lactoglobulin and the plasma proteins.

Zein is very rich in leucine and isoleucine, 25 to 30 per cent; proline, 10 per cent; and glutamic acid, 30 to 35 per cent, present predominantly as the amide. It has very few free carboxyl groups and cationic groups, hence the peculiar solubility properties (see p. 240). High-protein corn contains relatively more zein than low-protein corn; in fact, it appears that practically no zein is present in corn kernels up to a certain protein level, above which essentially all of the protein formed is zein. There is also evidence that the protein is formed late in the maturation period.

Zein preparations have been reported to have molecular weights over the range of about 20,000 to 50,000. Although by no means a typical fibrillar protein, the zein molecule has very high asymmetry, its axial ratio being perhaps 20:1. The significance of these studies

is questionable in view of the obvious inhomogeneity of the preparations.

Gliadin, the prolamin of wheat, has also been the subject of much study. Its amino acid composition is somewhat similar to that of zein, though it is even richer in glutamic acid, and it is also evidently far from homogeneous. The mean molecular weight is somewhat lower than that of zein, and gliadin is somewhat more soluble. It can be dissolved, for example, in concentrated solutions of sodium salicylate, in which medium zein is virtually insoluble.

SELECTED REFERENCES

Books and Reviews

- "The Proteins of Plants," H. B. Vickery, *Physiol. Revs.*, **25**, 347 (1945).
Chemistry of Muscular Contraction, A. Szent-Gyorgyi, Academic Press, New York, 1947.
"Immune Proteins in Bovine Milk and Colostrum," E. L. Smith, *J. Dairy Sci.*, **31**, 127 (1948).
"Milk Proteins," T. L. McMeekin and B. D. Polis, *Advances in Protein Chem.*, **5**, 202 (1949).
"Plant Proteins," J. W. H. Lugg, *Advances in Protein Chem.*, **5**, 230 (1949).
"The Plant Protein," J. Bonner in *Plant Biochemistry*, p. 245, Academic Press, New York, 1950.
"Egg Proteins," H. L. Fevold, *Advances in Protein Chem.*, **6**, 187 (1951).
"Egg Proteins," R. Warner in Neurath and Bailey's *The Proteins*, Vol. IIA, p. 435, Academic Press, New York, 1954.
"Milk Proteins," T. McMeekin in Neurath and Bailey's *The Proteins*, Vol. IIA, p. 389, Academic Press, New York, 1954.
"Muscle," K. Bailey in Neurath and Bailey's *The Proteins*, Vol. IIB, p. 951, Academic Press, New York, 1954.
"Seed Proteins," S. Brohult and E. Sandegren in Neurath and Bailey's *The Proteins*, Vol. IIA, p. 487, Academic Press, New York, 1954.

Original Articles

- "An Electrophoretic Study of the Proteins of Egg White," L. G. Longsworth, R. K. Cannan, and D. A. MacInnes, *J. Am. Chem. Soc.*, **62**, 2580 (1940).
"Separation of α - and β -Casein," R. C. Warner, *J. Am. Chem. Soc.*, **66**, 1725 (1944).
"An Electrophoretic Study of Gliadin," G. W. Schwert, F. W. Putnam, and D. R. Briggs, *Arch. Biochem.*, **4**, 371 (1944).
"Studies on the Fractionation of Zein," R. A. Gortner and R. T. MacDonald, *Cereal Chem.*, **21**, 324 (1944).
"Identification of the Bacteria-inhibiting, Iron-binding Protein of Egg White as Conalbumin," G. Alderton, W. H. Ward, and H. L. Fevold, *Arch. Biochem.*, **11**, 9 (1946).
"The Immune Proteins of Bovine Colostrum and Plasma," E. L. Smith, *J. Biol. Chem.*, **164**, 345 (1946).

- "Electrophoretic Inhomogeneity of Crystalline β -Lactoglobulin," C. H. Li, *J. Am. Chem. Soc.*, **68**, 2746 (1946).
- "Relationship of Zein to the Total Protein in Corn," D. W. Hansen, B. Brimhall, and G. F. Sprague, *Cereal Chem.*, **23**, 329 (1946).
- "Separation and Characterization of Conalbumin," J. A. Bain and H. F. Deutsch, *J. Biol. Chem.*, **171**, 531 (1947).
- "Zein Solutions as Association-Dissociation Systems," B. Scalpet, *J. Am. Chem. Soc.*, **69**, 1602 (1947).
- "Mechanism of Muscular Contraction," A. Szent-Gyorgyi, *Science*, **110**, 411 (1949).
- "Crystalline Aldolase," J. F. Taylor, A. A. Green, and G. T. Cori, *J. Biol. Chem.*, **173**, 591 (1949).
- "The Amino Acid Composition of Aldolase and D-Glyceraldehyde Phosphate Dehydrogenase," S. F. Velick and E. Ronzoni, *J. Biol. Chem.*, **173**, 627 (1949).
- "A Physical-Chemical Study of Zein in Aqueous Detergent Solutions," J. F. Foster, *J. Phys. & Colloid Chem.*, **53**, 175 (1949).
- "Enzymic Dephosphorylation of Ovalbumin and Plakalbumin," G. E. Perlmann, *Nature*, **166**, 870 (1950).
- "Egg White Proteins. I. Electrophoretic Studies on Whole White," R. H. Forsythe and J. F. Foster, *J. Biol. Chem.*, **184**, 377 (1950). "II. An Ethanol Fractionation Scheme," *J. Biol. Chem.*, **184**, 385 (1950).
- "Separation of γ -Casein," N. J. Hipp, M. L. Groves, J. H. Custer, and T. L. McMeekin, *J. Am. Chem. Soc.*, **72**, 4928 (1950).
- "Isolation of an Electrophoretically Homogeneous Crystalline Component of β -Lactoglobulin," B. D. Polis, H. W. Schmukler, J. H. Custer, and T. L. McMeekin, *J. Am. Chem. Soc.*, **72**, 4965 (1950).
- "Isolation of a Crystalline Albumin from Milk," B. D. Polis, H. W. Shmukler, and J. H. Custer, *J. Biol. Chem.*, **187**, 349 (1950).
- "Serological Relationship of Bovine Whey Albumin to Serum Albumins," E. Coulson and H. Stevens, *J. Biol. Chem.*, **187**, 355 (1950).
- "Some Physical Properties of Hen Egg Conalbumin," R. A. Fuller and D. R. Briggs, *J. Am. Chem. Soc.*, **78**, 5253 (1956).

20

Hormonal Proteins

Amino acids and derivatives

Insulin

Glucagon

Adrenocorticotropic hormone

Oxytocin and vasopressin

Other protein hormones

Hormones are "chemical messengers" of the body. Produced in special glands, they are transported to sites where they are capable of exerting their physiological functions in some biochemical fashion. Hormones are known to occur in plants and in lower animals, but have been found in the greatest profusion in the higher animals. This is thought to be the result of a selective evolution, leading to the elaboration of mobilizing substances for organisms so constructed that they require many special physiological processes to integrate the action of many cells. Numerous protein hormones are known.

Chemically, the hormones consist mainly of two classes, those that are protein in nature and those that are primarily steroids. Estradiol and testosterone are representatives of the latter class. The hormone insulin falls into the protein classification; in fact, insulin has often been used as a typical protein for physical and chemical studies. Such important hormones as epinephrin (adrenalin) are closely related to tyrosine. Epinephrin is thus indirectly a protein derivative. A borderline situation is found in the case of another amino acid which has hormonal properties, thyroxine. This amino acid exhibits the physiological properties of the thyroid gland. It is found as an amino

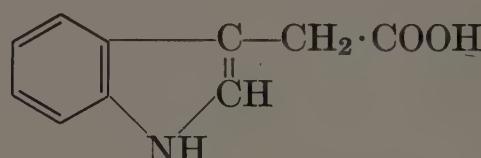
acid in a large protein molecule, thyroglobulin, which is believed by many to be the storage form of the hormone. Many of the hormones found in the pituitary gland, though proteinaceous, are often referred to as polypeptides because of the fact that their size is less than that of typical proteins.

The coexistence of two classes of hormones apparently so different in structure as the proteins and steroids becomes more comprehensible when one recalls that, physiologically, many if not all of the steroid hormones are under the control of master hormones, which are themselves proteins. This indicates that metabolically a steroid hormone and its master protein hormone bear a close relationship of some sort.

The hormones to be considered here are those in the protein category. First to be discussed are the simplest in this classification, the amino acid derivative type.

AMINO ACIDS AND DERIVATIVES

Indoleacetic Acid. One of the simplest of the organic molecules recognized as natural hormones is indoleacetic acid (*heteroauxin*), a widespread and powerful plant-growth hormone. Indoleacetic acid was a forerunner of present synthetic plant hormones such as 2,4-dichlorophenoxyacetic acid (2,4-D). In the concentrations at which indoleacetic acid is found physiologically, these hormones are true stimulants. The weed-killing action of indoleacetic acid, as well as that of the synthetic plant hormones, is believed to be not so much a killing as an overstimulation; the treated plant literally grows itself to death, probably as the result of unbalanced cellular elongation and proliferation. Indoleacetic acid has the formula



and can readily be seen to be a relative of tryptophan. It has been shown, in fact, that plants are capable of manufacturing indoleacetic acid from that amino acid.

Histamine. This simple amino acid derivative functions as a gastric hormone. Histamine can arise by decarboxylation of histidine

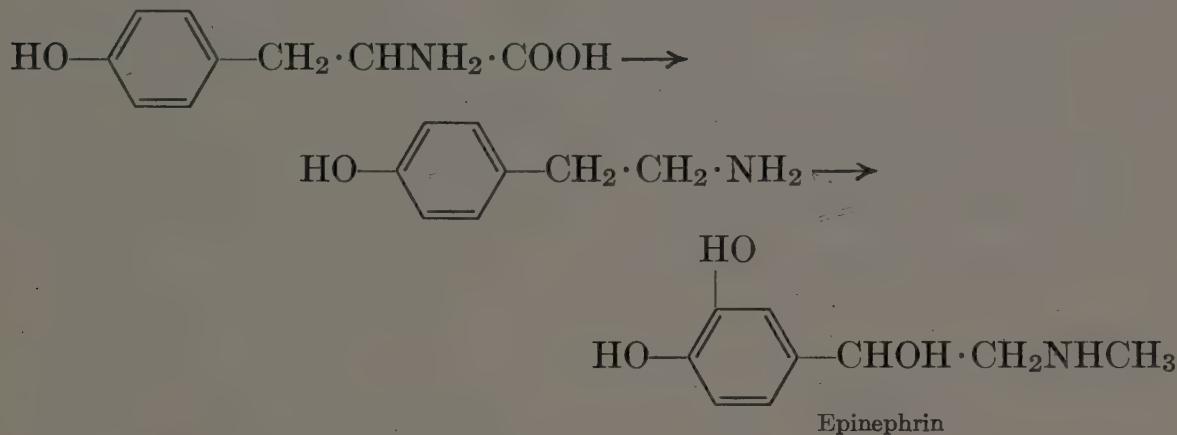


To the extent that it has been possible to study the reaction physiologically, the simple, unqualified mechanism illustrated does not appear to represent the true process.

Histamine is not only a hormone stimulating acid secretion; it is also the direct causative agent in such states as allergy and shock. A large development in the pharmaceutical industry has evolved from the fabrication of molecules intended to function at the enzyme centers where histamine presumably is effective. The resulting pharmaceutical compounds are known as the *antihistamines*.

It is probably histamine that stimulates the gastric processes and initiates peristaltic movement when such condiments as soy sauce are ingested. Abel, the discoverer of epinephrin, showed long ago that on hydrolysis of many proteins, histamine is formed from the histidine which is liberated.

Epinephrin. It is well established from physiological and tracer studies that the mother substance of epinephrin is tyrosine. The first step is a decarboxylation of tyrosine to tyramine,



with subsequent oxidation of both the side chain and the benzene ring, and methylation of the amino group. Tyramine, incidentally, has to a lesser degree many of the properties of histamine, and there is evidence for the point of view that allergic responses, such as sneezing and urticaria, are the result of the formation of tyramine and other amines from amino acids, as well as of the very potent histamine.

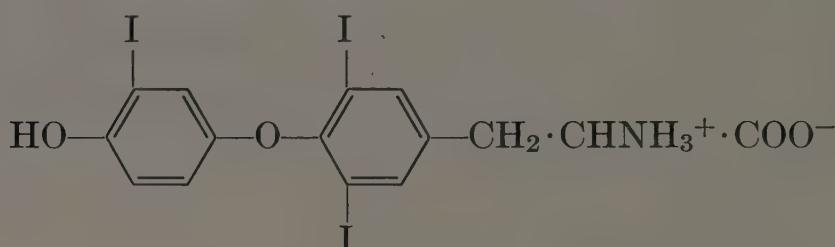
Thyroxine. Thyroxine has been shown to be formed from two molecules of diiodotyrosine, which in turn is derived from tyrosine (Chap. 7).

Thyroxine may be isolated from the glandular protein, thyroglobulin. An insufficient amount of the hormone may result in dwarfism. Some evidence exists for the ability of the hormone to stimulate gigantism in various animals. Insufficient dietary iodine results in insufficient thyroxine, and leads to the physiological condition in which the gland

enlarges in an attempt to compensate for this deficiency, the condition known as goiter.

Early in the century iodinated proteins were marketed in France as a substitute for thyroid powder, then difficultly available. This material did not at that time receive serious attention, probably because it was not considered to have a rational basis. In the early 1930's it was demonstrated that thyroxine is formed by iodination of such proteins as casein. The process has been studied and developed to a point of high yield, and thyroxine may now be manufactured from either iodinated protein or diiodotyrosine with comparative economy.

In 1952, it was learned in the laboratories of Pitt-Rivers and of Roche that the thyroid gland contains another active iodoamino acid, triiodothyronine:



L-Triiodothyronine was found to have three times as much thyroid activity as L-thyroxine. This explained the old observation that whole thyroid gland possesses more activity than can be accounted for by the thyroxine content alone.

INSULIN

The best known of the endocrinial proteins is insulin which is responsible in part for the regulation of carbohydrate metabolism. This hormone is secreted by the pancreas. Insulin deficiency is manifested by the pathological state of diabetes, and may be treated by dosing with the hormone. Administration of insulin to a normal animal results in hypoglycemia. Epinephrin has an opposite effect, and the action of these two hormones may thus be seen to be regulatory with respect to the important factor of maintenance of proper sugar level. In early isolation studies it was not possible to crystallize insulin until zinc, which had been removed in the isolation process, had been returned to the preparation. Available data indicate that the zinc combines in stoichiometric proportions (approximately 0.5 per cent by weight).

The molecular weight of insulin is approximately 40,000, as determined by ultracentrifuge and diffusion, and the molecule is nearly spherical. It has been observed, however, that insulin is dissociated

or disaggregated readily through the agency of certain compounds, such as synthetic detergents. More recently it has been demonstrated that the sedimentation constant is very dependent on the *pH* and ionic strength of the medium.

These results have been interpreted on the basis of a fast, reversible association-dissociation equilibrium between units of molecular weight 12,000 and dimers, trimers, and possibly tetramers of this unit. It is now clear, however, that the 12,000 molecular weight units are further dissociable in nonaqueous media, into units of only 6,000 molecular weight which represent the true fundamental units of insulin. The interesting philosophical question arises as to whether biological activity resides in the unit per se or only in the polymers that exist in aqueous media.

The associating forces involved in the polymerization of insulin are apparently van der Waals-type forces, the free energy of association amounting to approximately 20,000 cal. per molecule. Dissociation is due to electrostatic repulsion, and consequently takes place to an increasing extent with increasing charge, resulting from extremes in *pH* and decreasing ionic strength.

Heating insulin in an acid solution forms fibrils having lengths of the order of magnitude 10,000 Å. Regenerated insulin can be obtained from these fibrils by treatment with alkali, and is found to be indistinguishable from native insulin on the basis of crystallizability, biological activity, availability of reactive groups, ability to form fibrils, and ultracentrifuge pattern. The ready reversibility of this fibril-formation process indicates that it involves only minor structural changes. Most probably the process is primarily an aggregation of corpuscular units end-to-end.

The isoelectric point of insulin is near *pH* 5.5. The protein is unusual in that it is not precipitated by moderately high concentrations of ethanol, use of which fact is made in its preparation. Perhaps this solubility in relatively high ethanol concentrations should be expected in view of the composition of insulin. It contains approximately 30 per cent leucine, a figure which is reminiscent of the prolamins. There may be a relationship between this high leucine content and the tendency of insulin to aggregate.

No other protein molecule of comparable size is understood in its structural detail with as much certainty as is insulin. The tremendous labors of Sanger and associates (Brown *et al.*, 1955) have yielded a proposal for the entire sequential arrangement of amino acid residues in insulin. In addition, this accomplishment contributed greatly to exciting the renaissance in peptide-protein structure analysis which

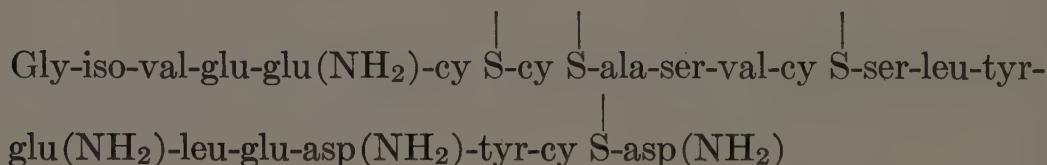
began about 1945. A detailed treatment of the work is accordingly presented at this point.

Sanger's work began at a time when the molecular weight of insulin was believed to be about 12,000. Phenylalanine had been shown by Jensen and Evans, by the phenyl isocyanate technique, to be N-terminal. Sanger discovered glycine and phenylalanine both as N-terminal residues by application of the DNFB technique. In order to estimate the proportions of these end-groups, he ran recovery experiments on the partially hydrolyzable DNP-glycine and upon DNP-phenylalanine. From these he concluded that each molecule of 12,000 weight had two glycyl chains and two phenylalanyl chains. Published cystine analyses and analogous theoretical treatments suggested that the chains were held together by disulfide cross-linkages. Sanger accordingly broke these chains apart by oxidation of the cystine residues to cysteic acid residues with performic acid. It was then possible to fractionate two main types of product on columns, chain A with glycyl terminal, and chain B with phenylalanyl terminal.

Chain B was studied first by the fragmentation technique described in Chap. 9. First, however, partial hydrolysis of the type that had been used in studies by Felix (Chap. 9) was employed to produce large fragments. Typical conditions for this stage of hydrolysis were 4 days at 37° in 11 N hydrochloric acid. The fragments were separated on ion-exchange resins, by adsorption on charcoal, and by ionophoresis. After these group separations, further fractionation was achieved by paper chromatography.

By application of the fragmentation technique on DNP-peptides, it was possible to deduce the sequences in individual peptides. The next step was to reassemble these theoretically. The gaps left by this reconstruction were filled in by Sanger and Tuppy by carrying out other fragmentations with proteolytic enzymes. The thinking involved is best understood by reference to Table 20-1. It will be seen from this table that no discordant sequences were found.

The sequence in chain A was elucidated in a similar manner and was shown to be:



Certain reservations with respect to complete acceptance of these structures have been voiced in the literature. These criticisms include statements that the degree of homogeneity of the insulin used was not

TABLE 20-1. Peptides Fragmented from Insulin Chain B¹ by Four Hydrolytic Agents

Agent	Peptide	Sequence	Structure
Acid	Phe-val-asn-glu-his-leu-cy S-gly	Tyr-leu-val-cy S-gly	Thr-pro-lys-ala
	Ser-his-leu-val-glu-ala	Gly-glu-arg-gly	
	Ala-leu	Gly-phe	
	Ala-leu-tyr		
Pepsin	Phe-val-asn-glu-his-leu-cy S-gly-ser-his-leu	Leu-val-cy S-gly-glu-arg-gly-phe	Tyr-thr-pro-lys-ala
	Val-glu-al-a-leu		
Chymo-trypsin	His-leu-cy S-gly-ser-his-leu	Leu-val-cy S-gly-glu-arg-gly-phe-phe	
	Val-glu-al-a-leu-tyr		
Trypsin		Gly-phe-phe-tyr-thr-pro-lys	Ala

¹ Structure Phe-val-asn-glu-his-leu-cy S-gly-ser-his-leu-val-glu-ala-leu-tyr-leu-val-cy S-gly-glu-arg-gly-phe-phe-tyr-thr-pro-lys-ala.

established, that discordant sequences may have been lost on fractionation (quantitative aspects of recovery are not assessed), and that some of the enzymes used are known to bring about rearrangements of sequence in other studies. A number of the details of the proposed structure have been confirmed by other workers, although it must be said that a number of other workers have found results that cannot be immediately reconciled with the structural proposal. The sequences may, however, be entirely correct; it is difficult to see how they may be wrong in any large degree. The entire work is an experimental step forward and has proved to be a strong stimulus to others.

Comparison of insulins from different species was made possible by the availability of homogeneous samples of insulin prepared by the Craig countercurrent method (Chap. 8). The results are shown in Table 20-2. The two preparations from beef have identical amino

TABLE 20-2. Amino Acid Compositions of
Insulins from Different Species¹

Amino Acid	Number of Residues			
	Beef A	Beef B	Pork	Sheep
Alanine	3	3	2	3
Arginine	1	1	1	1
Aspartic acid	3	3	3	3
Cystine	3	3	3	3
Glutamic acid	7	7	7	7
Glycine	4	4	4	5
Histidine	2	2	2	2
Isoleucine	1	1	2	1
Leucine	6	6	6	6
Lysine	1	1	1	1
Phenylalanine	3	3	3	3
Proline	1	1	1	1
Serine	3	3	3	2
Threonine	1	1	2	1
Tyrosine	4	4	4	4
Valine	5	5	4	5
Ammonia	6	5	6	6

¹ From E. J. Harfenist, *J. Am. Chem. Soc.*, **75**, 5528 (1953).

acid compositions, which certify the reliability of the analyses employed. The ability to separate two such large molecules with so small a structural difference testifies to the power of countercurrent analysis.

When beef, pork, and sheep insulins are compared, the contents of alanine, glycine, isoleucine, serine, threonine, and valine are seen to

differ. For four of these amino acids, porcine insulin differs from that of beef and sheep. For two of the amino acids, the sheep insulin differs.

It is a matter of some significance that the analysis of beef insulin A is in entire accord with the structure of insulin as proposed by Sanger. The structural findings for insulin lack any discordances, and this report has been cited as evidence that insulin is one protein which lacks heterogeneity. (The coexistence of beef A and beef B insulins can be explained as due to hydrolysis of a single amide group, as in an asparagine residue, in A during preparation.) This inference ignores the fact that the essentially qualitative chromatography used for structural assignment could have overlooked minor discordances, and also ignores the fact that the insulin employed had already undergone some purification in its preparation.

No question arises as to differences in structure of insulin from different species. Brown, Sanger, and Kitai (1955) have found that the sequences in the phenylalanyl chains are the same for the major components of insulin from pigs, sheep, or cattle. In the glycyl chain, however, there are minor differences from one species to another (the differences are italicized):

Cattle	Gly-iso-val-glu-glu(NH ₂)-cy S-cy S- <i>ala-ser-val</i> ...
Pig	Gly-iso-val-glu-glu(NH ₂)-cy S-cy S- <i>thr-ser-iso</i> ...
Sheep	Gly-iso-val-glu-glu(NH ₂)-cy S-cy S- <i>ala-gly-val</i> ...

The crystalline pig and sheep insulins were heterogeneous.

GLUCAGON

The sequential structure of glucagon, a second hormone from the pancreas, was described in 1956. This hormone is hyperglycemic and glycogenolytic in its action. The sequence was reported to be

His-ser-glu(NH₂)-gly-thr-phe-thr-ser-asp-tyr-ser-lys-tyr-leu-asp-ser-arg-arg-ala-glu(NH₂)-asp-phe-val-glu(NH₂)-try-leu-met-(aspNH₂)-thr

ADRENOCORTICOTROPIC HORMONE

The anterior lobe of the pituitary gland produces at least six or eight important hormones. One of the best known of these is the adrenocorticotropic hormone (ACTH). Evidently the chief function of this hormone, as the name partly suggests, is control over the

adrenal cortex, and thereby control over the production of many steroid hormones. This protein has been isolated from the glands of several species. It is isoelectric at pH approximately 4.7 and has a molecular weight of 20,000, as determined by sedimentation.

ACTH exists in a family of closely related molecules. Eight of these molecular species were isolated in initial studies. Although the activity of ACTH is destroyed by tryptic proteolysis, the hormonal activity is remarkably resistant to peptic hydrolysis. It was found to be possible to prepare fragments of increased specific activity by either peptic hydrolysis or mild treatment with acetic acid. These corticotropins have been studied structurally by White and co-workers and by Bell and associates (Bell, 1954). One of these active fragments is reported to have the structure

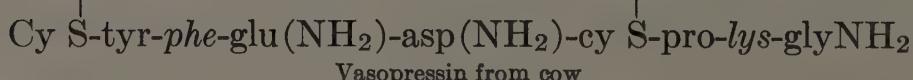
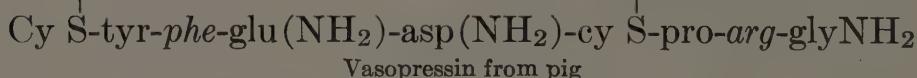
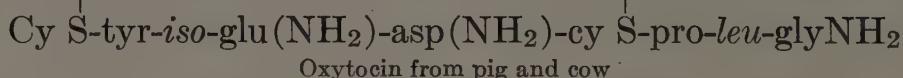
Ser-tyr-ser-met-glu-his-phe-arg-try-gly-lys-pro-val-gly-lys-lys-arg-arg-pro-val-lys-val-tyr-pro-ala-(gly, glu, asp)-asp-glu(NH₂)-leu-ala-glu-ala-phe-pro-leu-glu-phe

It is of interest that the pro-val sequence occurs twice. Basic amino acid residues display a considerable degree of sequestering, as in lys-lys-arg-arg. Basic "cores" of protein molecules have long been recognized. The lys-arg sequence has also been noted before. The asp-glu sequence has been observed in insulin, tyrocidine A, and γ -globulin.

OXYTOCIN AND VASOPRESSIN

Oxytocin and vasopressin are found as hormones of the posterior lobe of the pituitary gland. The former causes contraction of smooth musculature, whereas the latter raises blood pressure.

The chemistry of these hormones has been elucidated by du Vigneaud and associates (du Vigneaud *et al.*, 1953). Following determination of the structures by a variety of analytical procedures, this group of workers synthesized oxytocin (1953). The structures are therefore virtually certain.



The structures presented are cyclic. The left-hand cysteine residue bears a free amino group, but there is no free carboxyl because of the amide structure in the glycine residue. In addition, two molecules of ammonia cover the side chains of glutamyl and aspartyl residues.

The variation in structure from one hormone to the other is most striking. The two residues, per molecule, which exhibit differences are italicized. Oxytocin is identical for both species. Vasopressin differs from oxytocin by having phenylalanine instead of isoleucine, and a basic amino acid instead of leucine. Although vasopressin differs from one species to the other, this difference is small, arginine and lysine each being a basic amino acid. The small structural difference, permitting what we recognize as large differences in biological activity, is notable. A comparison of the variation within species and between species alongside the pattern for insulin is of particular interest.

OTHER PROTEIN HORMONES

Other protein hormones are recognized by virtue of their physiological activity. These include growth hormone, follicle-stimulating hormone (FSH), interstitial cell-stimulating hormone (ICSH), prolactin, secretin, and gonadotropin. The reader is referred to the review by Li (1954) for discussion of these less clearly elucidated hormones.

SELECTED REFERENCES

Review

"Protein Hormones," C. H. Li in Neurath and Bailey's *The Proteins*, Vol. IIA, p. 595, Academic Press, New York, 1954.

Original Articles

"Chemistry of Thyroxine. III. Constitution and Synthesis of Thyroxine," C. R. Harington and G. Barger, *Biochem. J.*, **21**, 169 (1927).

"The Biological Synthesis of Radioactive Adrenalin from Phenylalanine," S. Gurin and A. M. Delluva, *J. Biol. Chem.*, **170**, 545 (1947).

"Oxidation of Diiodotyrosine Derivatives," R. Pitt-Rivers, *Biochem. J.*, **43**, 223 (1948).

"Mechanism of Histamine Formation," E. Geiger, *Arch. Biochem.*, **17**, 391 (1948).

"Crystallization of Hypophyseal Growth Hormone," C. H. Li, H. M. Evans, and M. E. Simpson, *Science*, **108**, 624 (1948).

"The Chemical Nature and Formation of Auxin in the *Avena* Coleoptile," S. G. Wildman and J. Bonner, *Am. J. Botany*, **35**, 740 (1948).

- "Decrease in the Formation of Adrenocorticotropicin Following Administration of Epinephrine," C. A. Genzell, *Endocrinology*, **50**, 399 (1952).
- "Countercurrent Distribution Studies with Insulin," E. J. Harfenist and L. C. Craig, *J. Am. Chem. Soc.*, **74**, 3083 (1952).
- "The Amino-Acid Sequence in the Glycyl Chain of Insulin," F. Sanger and E. O. P. Thompson, *Biochem. J.*, **53**, 366 (1953).
- "3:5:3'-Triiodothyronine. 2. Physiological Activity," J. Gross and R. Pitt-Rivers, *Biochem. J.*, **53**, 652 (1953).
- "Structure of Insulin," C. Robinson, *Nature*, **172**, 773 (1953).
- "Reversible Dissociation of Insulin," E. Fredericq, *Nature*, **171**, 570 (1953).
- "The Sequence of Amino Acids in Oxytocin, with a Proposal for the Structure of Oxytocin," V. du Vigneaud, C. Ressler, and S. Trippett, *J. Biol. Chem.*, **205**, 949 (1953).
- "The Amino Acid Compositions of Insulins Isolated from Beef, Pork and Sheep Glands," E. J. Harfenist, *J. Am. Chem. Soc.*, **75**, 5528 (1953).
- "The Synthesis of an Octapeptide Amide with the Hormonal Activity of Oxytocin," V. du Vigneaud, C. Ressler, J. M. Swan, C. W. Roberts, P. G. Katsoyannis, and S. Gordon, *J. Am. Chem. Soc.*, **75**, 4880 (1953).
- "Purification and Structure of β -Corticotropin," P. H. Bell, *J. Am. Chem. Soc.*, **76**, 5565 (1954).
- "Nouvelles données sur la présence de la 3:3':5'-triiodothyronine et de la 3:3'-diiodothyronine dans la thyroglobuline et à l'état d'hormones libres dans le corps thyroïde," J. Roche, R. Michel, and W. Wolf, *Compt. rend. soc. biol.*, **149**, 1604 (1955).
- "The Structure of Pig and Sheep Insulins," H. Brown, F. Sanger, and R. Kitai, *Biochem. J.*, **60**, 556 (1955).
- "The Amino Acid Sequence of Glucagon," W. W. Bromer, L. G. Sinn, A. Staub, and O. K. Behrens, *J. Am. Chem. Soc.*, **78**, 3858 (1956).
- "Sur le métabolisme de la 3:3'-diiodothyronine," J. Roche, R. Michel, N. Etling, and J. Nunez, *Biochim. et Biophys. Acta*, **19**, 490 (1956).
- "Influence of Zinc on the Heterogeneity of Insulin," E. Fredericq, *Arch. intern. physiol. et biochim.*, **64**, 132 (1956).
- "Synthesis of a Polypeptide with ACTH-like structure," R. A. Boissonas, S. Guttmann, J.-P. Waller, and P. A. Jaquenoud, *Experientia*, **12**, 446 (1956).
- "Oxypressin, a Synthetic Octapeptide Amide with Hormonal Properties," P. G. Katsoyannis, *J. Am. Chem. Soc.*, **79**, 109 (1957).

21

Enzymes

Significance of enzymes in biology

Relationship of enzymes to protein chemistry

Occurrence and classification

Purification and crystallization

Mechanism of enzyme action

Zymogens

Activators, coenzymes, and antienzymes

Inhibitors

Structure of enzymes

SIGNIFICANCE OF ENZYMES IN BIOLOGY

Enzymes are biological catalysts. As the primary function of a catalyst is increase of the rate of a reaction, so the function of an enzyme is increase of the rate of a reaction in an organism. The enzymes are found in living systems. They may be extracted therefrom in many cases, and their action may be studied in experiments on isolated materials. A great deal has been learned of the behavior and nature of these materials by the traditional experimental approach. It should be borne in mind, however, that the value of an enzyme to an organism is to be measured in terms of its contribution to an organized, integrated pattern of action of all the enzymes, *extracellular* and *intracellular*. The enzymes make it possible for a wide variety of essential reactions to occur at the temperature of the organism itself. Many of these same reactions would occur far too

slowly in the absence of such catalysts, or would require temperatures intolerably high for most organisms on this planet.

No less remarkable than the enzymes themselves is the fact that the living cell is capable of regulating the balanced interplay of the numerous enzymic activities found within it. Some insight into this "mechanism of mechanisms" is obtainable from work on the kinetics of enzyme-catalyzed reactions. Investigation of the physical chemistry involved has been of two principal types, investigation of equilibria and of rates of reaction. Equilibrium reactions represent the kind that may most conveniently be studied in glassware. It is increasingly recognized that the kinetics of enzyme-catalyzed reactions deserve particular attention even though this type of treatment has not always been as conveniently approachable as equilibrium studies. The relative importance of kinetics was recognized by Haldane in 1930 when he wrote, "the key to a knowledge of enzymes is a study of reaction velocities, not of equilibria."

One fact often cited to substantiate the not infrequent statement that enzymes do not perform as ideal catalysts is that the enzymes are capable of promoting reactions which appear not to occur at all in their absence. This objection is theoretically answerable; in the absence of the enzyme the reaction may actually proceed, but so slowly that its rate is not appreciable. It has often been stated, also, that enzymes are not ideal catalysts because many of them are used up during the associated reaction. This loss of enzyme is to be attributed to metabolic decay. Whether or not an enzyme is used up in the reaction is of little importance biologically; biological systems are not true equilibrium systems, and enzymes are continually being elaborated.

RELATIONSHIP OF ENZYMES TO PROTEIN CHEMISTRY

The subject of enzymes belongs within the field of protein chemistry for two main reasons. One of these is the protein or peptide nature of many of the substrates (Chap. 22), and the other is the protein nature of the enzymes themselves. Enzymes can in fact be classified in the same two main categories as proteins. Many, such as pepsin and trypsin, are reported to be simple proteins; many others, carboxylase and D-amino acid oxidase for example, are conjugated proteins containing prosthetic groups. In this field of knowledge the prosthetic groups, such as cocarboxylase and riboflavin adenine dinucleotide, are known as *coenzymes*.

It appears possible to generalize, with some assurance, that those enzymes which are simple proteins are primarily hydrolytic in their

action, whereas enzymes which are conjugated possess mainly non-hydrolytic catalytic functions. Pepsin and trypsin do not contain the non-amino acid type of prosthetic group, but it is believed that, like oxidative enzymes, they must have "active centers." These are presumably sites in an appropriate configuration on the surface of the simple protein molecule.

The history of our knowledge of the composition of the enzymes records many revised judgments. Buchner and others succeeded in extracting enzymes from cellular material, but they recognized activities rather than substances. The mere fact that such activities could be separated from the living material resolved a famous controversy between Pasteur and Liebig. Much time ensued before any substantial number of investigators even dared to attempt to isolate the substances with which such activity might be associated. It was rather widely considered, following Buchner's work, that these somewhat mysterious "enzymes" were associated with proteins that functioned principally as a type of carrier. The most active preparations which it was possible to obtain did not, however, exhibit the usual qualitative protein color tests. This led to the commonly held opinion that the enzymes were not related compositionally to the proteinaceous material from which they were originally isolated. It is now easy to recognize that what was really shown was that enzymes are so powerful that their physiological activity may yet be demonstrated in dilutions which cannot be detected by sensitive protein color tests.

The decisive experiments were those of Sumner, who successfully crystallized the enzyme urease and demonstrated that its properties as an enzyme were associated unmistakably with the protein. Although it was some time before traditional thinking became widely rechanneled, dozens of enzymes have been crystallized since Sumner's historic announcement in 1926. Without exception, each of these has been found to be protein.

It cannot be said that the exact nature of enzymes has been elucidated. Some of the most challenging and interesting information is yet to be unearthed. Particularly is this true of the problem of designating the structures of the enzymes so that their behavior can be explained in a fully chemical manner.

OCCURRENCE AND CLASSIFICATION

The occurrence of enzymes is universal in biological systems. It seems likely from the evidence at hand that a major portion or all of the chemical reactions which occur in organisms are enzyme-controlled.

The occurrence of enzymes is frequently considered as being either intracellular or extracellular. Proteolytic enzymes are represented in each of these classes.

Complete classification of all the known enzymes is not within the province of this book. The classification of proteolytic enzymes is definitely within the area that requires consideration here, and this is presented in Chap. 22. Biochemists usually classify all enzymes into (1) the hydrolytic type, of which the proteolytic enzymes represent a major group, and (2) the nonhydrolytic or desmolytic enzymes. The latter group includes especially those enzymes that are concerned with oxidation-reduction processes.

PURIFICATION AND CRYSTALLIZATION

The purification and crystallization of proteolytic enzymes is synonymous with the purification and crystallization of proteins. In fact, very much of the present knowledge with regard to proteins as a whole has been derived from experiments with enzymes. Most notable of such projects have been the long-term efforts of Northrop and associates, who not only adapted, developed, and applied methods to the purification and crystallization of proteolytic enzymes but also provided the field with outstandingly useful criteria of purity.

An example of Northrop's methodology may be given by the detailing of the method of crystallization of pepsin from bovine gastric juice. For this, the contents of the fourth pouch of the cow's stomach were filtered, saturated with ammonium sulfate, filtered, and the precipitate dissolved in dilute hydrochloric acid. The solution was cooled to -10° and precipitated with acetone by bringing the content of acetone in solution to about 60 per cent. This mixture was then centrifuged, and the supernatant liquid precipitated by increasing the acetone concentration to over 75 per cent. The precipitate was then dissolved in hydrochloric acid solution and precipitated by adding magnesium sulfate to greater than half-saturation of the solution. The magnesium sulfate precipitation was repeated three to five times. The precipitate was dissolved in a small volume of sodium acetate solution, and reprecipitated at pH 3.0 with half-saturated magnesium sulfate solution. The new precipitate was dissolved in water with the addition of a little alkali, and again thrown down by acidification with sulfuric acid to pH 3.0. The suspension was again filtered, and the precipitate redissolved in three volumes of water at 45° . Acid was added at this temperature, to just below the precipitation point. The liquid was cooled slowly, after which the crystals of pepsin were filtered off.

MECHANISM OF ENZYME ACTION

This discussion may appropriately begin with a review of some of the salient features of catalysis in general. First, it should be remembered that catalysts cannot cause reactions to take place; they can only increase the rate of reaction. In the case of enzymic catalysis, the rate of the uncatalyzed reaction is frequently so slow as to be negligible, so that for all practical purposes one can speak of the reaction as "being brought about" by the enzyme. Second, it is fundamental that a catalyst cannot speed one reaction without also speeding the reverse reaction; that is, the true position of equilibrium in a reversible process is not altered. Most though not all of the reactions catalyzed by enzymes are such that the position of equilibrium is far over to one side, and the reverse reaction is therefore not observed. The result has been a tendency to consider enzymes as catalyzing only the forward reaction and not the reverse reaction. Much confusion has crept into the literature of enzymology because of this point.

In two respects, enzymes excel in comparison with most if not all nonbiological catalysts: (1) they are unusually effective catalysts, and (2) they are much more specific. Any theory for the mechanism of action of enzymes must first of all account for these facts.

It is generally conceded that direct combination between the enzyme and its substrate is a necessary prelude to reaction. This idea is supported by many indirect lines of evidence, as will be seen. Furthermore, it is in accord with the theories of ordinary catalysis. Two extreme types of catalyzed reactions are recognized, *homogeneous* and *heterogeneous*, depending on whether the catalyst can be considered as existing in the same phase as the substrate (that is, in solution) or as a separate phase. In the first case, catalysis is assumed to result from the formation of an intermediate compound or complex between reactant and catalyst. In heterogeneous catalysis, the reactant or reactants are presumed to be adsorbed on the surface of the catalyst.

Energetically, the function of a catalyst is to reduce the height of an energy barrier over which the reaction must proceed (energy of activation). Combination with the catalyst can perform this function by distorting (weakening) the bonds in one or both of the reactants, by forcing reactants together in a favorable position for reaction, or by some more subtle mechanism.

Catalysis by proteins (enzymes) is probably best considered as a case intermediate between the two extremes, having some of the

character of both homogeneous and heterogeneous catalysis. Enzymes may usually be considered as existing in true solution, but the molecules are so large as to possess large, specific surfaces.

Kinetics of Enzymic Reactions. It is through studies of the kinetics of a reaction that the mechanism of reaction is deduced. Our knowledge of the mode of action of enzymes, presently still quite limited, is based on such kinetic studies, and in particular on the effect of enzyme concentration, substrate concentration, pH, temperature, and other variables on the rate of reaction.

The first question of interest in kinetic studies is the order of the reaction (p. 317). For enzymic reactions, the order with respect to each substrate must be considered if more than one is involved, and in addition the order with respect to enzyme. Although a catalyst is not consumed in a reaction, and is therefore not a reactant in the usual sense of the word, it is certainly an important factor in governing the rate. The rate of an enzymic process is usually rather rigorously proportional to the enzyme concentration. Doubling the concentration of enzyme very nearly halves the time required for a given amount of reaction to take place. The order with respect to substrate is more complex. At very low substrate concentrations, or high enzyme concentrations, the dependence is normally first order; that is, rate is proportional to substrate concentration. At high substrate concentrations, however, the initial rate usually becomes independent of substrate concentration (zero order). In other words, for a given enzyme concentration there is a limiting rate of formation of product, beyond which one cannot go no matter how much substrate is added.

The variation of order in substrate is readily understandable if it is assumed that actual combination of enzyme with substrate takes place in the activation process. At high substrate concentrations the enzyme is saturated at all times, and any additional substrate added must "wait its turn." A quantitative description of the situation was first given by Michaelis and Menten. They assumed that the enzyme and substrate form and are in equilibrium with a dissociable complex, which complex may also break down to yield product plus enzyme. The mechanism may be given as



where E denotes enzyme, S substrate, and P product. Michaelis and Menten assumed that a true equilibrium constant exists for the first step, and expressed it as a dissociation equilibrium, the constant K_m being defined as

$$K_m = [E][S]/[ES] \quad (2)$$

On the basis of these assumptions it is a relatively simple problem to show that the velocity v at a given substrate concentration is related to substrate concentration, to K_m , and to the limiting velocity V , through the relation

$$v = \frac{V[S]}{K_m + [S]} \quad (3)$$

By means of this equation it is possible, from known values of velocity at various substrate concentrations, to determine both the limiting velocity and the value of K_m . Several graphical procedures

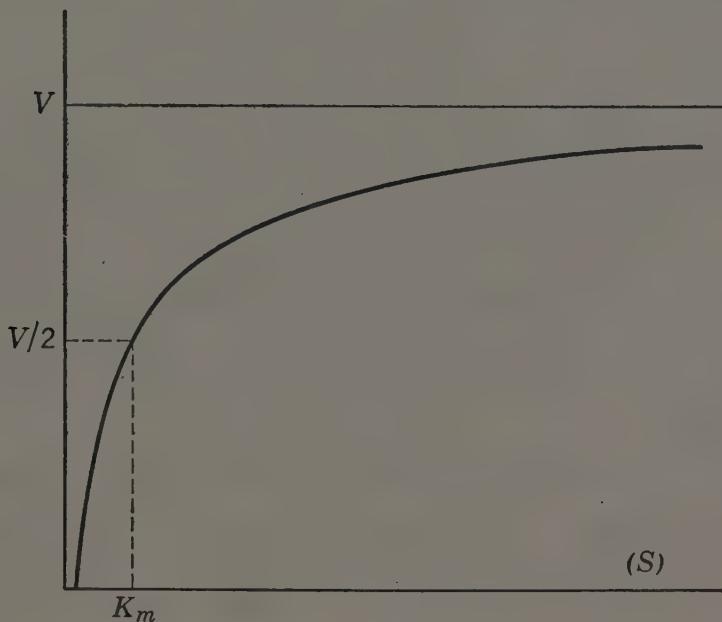


Fig. 21-1. Curve for the Michaelis-Menten equation. From E. Baldwin, *Dynamic Aspects of Biochemistry*, 2nd ed., University Press, Cambridge, 1949, p. 31.

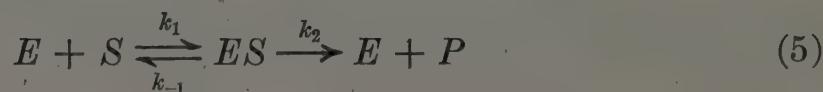
are available for this purpose, most popular of which is the Lineweaver-Burk method. In this treatment, the reciprocal of the velocity is plotted as a function of the reciprocal of $[S]$, since the Michaelis-Menten equation may be put in the form

$$\frac{1}{v} = \frac{K_m}{V} \left[\frac{1}{[S]} \right] + \frac{1}{V} \quad (4)$$

Such plots are linear, the slope being equal to K_m/V , and the intercept at $1/[S]$ equals zero (that is, $[S]$ equals infinity) being equal to $1/V$ (see Fig. 21-1).

While the Michaelis-Menten equation fits the observed results in many if not most enzymic reactions, it now appears that the result is in part fortuitous. It is inherent in their theory that the reversible reactions involved in the first stage are very rapid compared to the

second reaction leading to product. This now seems most doubtful. More generally, it is necessary to assume a pattern of the form



where the three rate constants k_1 , k_{-1} , and k_2 may be of any magnitude. In this case true equilibrium does not exist; rather, there is a "steady-state" concentration of ES which is dependent on all three reactions. The derived kinetic expression turns out, however, to be identical with the Michaelis-Menten expression, the only difference being that the constant K_m must be reinterpreted as

$$K_m = \frac{k_{-1} + k_2}{k_1} \quad (6)$$

The kinetics of enzyme reactions depend in general on such other factors as pH , temperature, and the presence of various substances such as coenzymes and inhibitors. These factors are considered in following sections. In addition, water doubtless also participates in many if not most enzymic reactions. Unfortunately, little can be learned about the kinetic significance of water from studies in aqueous solution.

Effect of pH. The rate of enzymic reactions is markedly affected by pH . Usually there is a pH range of maximum activity, which may be narrow or broad depending on the enzyme. The actual value of the optimal pH varies widely from enzyme to enzyme, but usually is near the pH at which the enzyme functions in nature. Pepsin, which normally operates in an acid medium, has an optimum near pH 2; trypsin is optimally active near pH 8. Presumably, for optimal activity the enzyme must possess just the right net charge, and possibly also the correct distribution of charges. However, the exact optimum depends also to some extent on the nature of the substrate. This is particularly true in the case of the proteases, where the substrates are proteins and peptides whose charge also depends markedly upon pH .

The pH effect is further complicated by the fact that enzymes, being proteins, are inherently unstable and are usually inactivated (denatured) by extremes in pH . The optimal pH for stability does not necessarily correspond to the pH of optimal activity.

Several studies have shown that the pH effect can be adequately accounted for by considering only two key ionizable groups. One of the most interesting examples is that of fumarase, elucidated so elegantly by Alberty and co-workers (1955). In this case it appears

that two imidazolium groups are critical. For optimal activity, one of these must be in the acidic (cationic) form and one in the basic form. It can readily be seen that the relative proportion of this active form will be maximal at a pH near 6 to 7, and that it will diminish sharply both above and below this range.

Temperature Effect. The rate of most reactions is increased by increase in temperature, and up to a point enzyme-catalyzed reactions are no exception to this rule. On the other hand, being proteins, enzymes are usually subject to heat denaturation. If the activity of an enzyme is plotted as a function of temperature, it is generally found that the curve goes through a maximum and then decreases, usually precipitously, as denaturation enters the picture. This latter effect, generally irreversible, takes place for many enzymes in the temperature range 50 to 60°, although some are stable at higher temperatures. It is frequently observed that enzymes are more stable in the presence of their substrates, presumably owing to combination therewith.

In principle it is possible to learn much regarding the mechanism of a reaction through study of the temperature effect. The temperature coefficient of a reaction provides, through the Arrhenius theory, a measure of the total energy of activation. By applying the newer absolute reaction rate theory of Eyring, it is possible to estimate further the entropy and free-energy changes involved in the activation process. While attempts have been made to apply this theory to enzymic reactions there are many uncertainties involved. An interesting conclusion reached from a comparative study of enzymic and nonenzymic catalysis of several reactions is that the unusual ability of enzymes resides in the fact that the formation of the activated enzyme-substrate complex involves a considerable positive entropy change.

ZYMOGENS

Precursors of biologically active proteins are known as *proteinogens*, and proteinogens which are precursors of enzymes are known frequently as *zymogens* or *proenzymes*. The work of Northrop and his school has led to enhanced knowledge of zymogens, particularly pepsinogen, chymotrypsinogen, and trypsinogen, and to the crystallization of some.

Zymogens are inactive precursors of the corresponding enzymes, and may be activated in various ways. A comparison of the composition and properties of pepsinogen and pepsin is given in Table 21-1.

The following facts have been ascertained about the formation of pepsin from pepsinogen.

- (1) The conversion of pepsinogen to pepsin is speeded by the addi-

TABLE 21-1. Some Properties of Pepsinogen and Pepsin¹

	Pepsinogen	Pepsin
N per cent	14.4	14.7
P per cent	0.09	0.09
Amino nitrogen as per cent total nitrogen	4.0	
Isoelectric point, <i>pI</i>	3.7	Immeasurably low
Molecular weight, by osmotic pressure	$42,000 \pm 3,000$	$38,000 \pm 3,000$
Optical rotation, $[\alpha]_D^{25}$ per gm. dry weight, degrees	-61	-71

¹ Adapted from J. H. Northrop, M. Kunitz, and R. M. Herriott, *Crystalline Enzymes*, 2nd ed., Columbia University Press, 1948, p. 81.

tion of pepsin or by increasing the concentration of pepsinogen. This means that the reaction is essentially autocatalytic (Fig. 21-2). It is of interest that the conversion of the zymogen into the enzyme

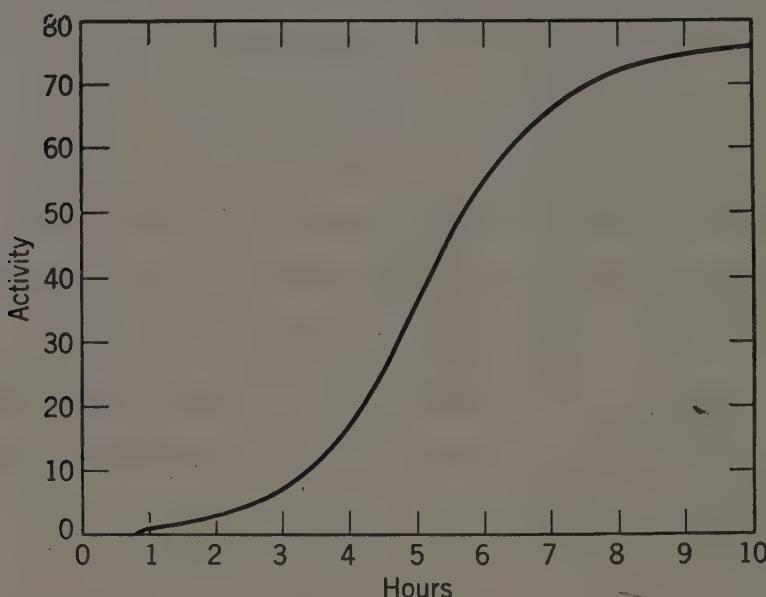
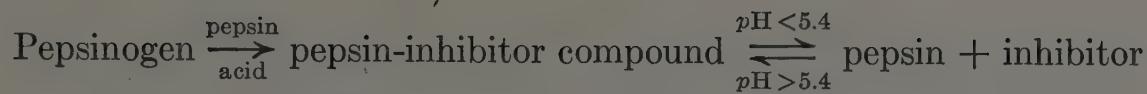


Fig. 21-2. Autocatalytic activation of crystalline trypsinogen by trypsin. From J. H. Northrop, M. Kunitz, and R. M. Herriott, *Crystalline Enzymes*, 2nd ed., Columbia University Press, 1948, p. 126.

follows a sigmoid curve, comparable to the first limb of a growth curve.

- (2) The reaction is catalyzed slowly by hydrogen ions.
- (3) The reaction has a maximum rate near the pH optimum of pepsin.
- (4) In a solution more acid than pH 5.4, or in the presence of a high concentration of salt, pepsin is found to exist in solution combined with an inhibitor. At pH's more acid than 5.4 the combination of inhibitor and enzyme dissociates into the two components. The overall sequence is illustrated as follows.



It has also been found that the type of enzyme obtainable from a specific proenzyme is a function of the zymogen itself. If, for example, chicken pepsinogen is activated by swine pepsin, the new pepsin formed is chicken pepsin. In the reverse situation, swine pepsinogen activated by either chicken pepsin or swine pepsin yields only swine pepsin. The two enzymes may be easily distinguished by either their rate of inactivation, at pH 8 to 9, or their immunological behavior.

Similar yet different situations surround the relationships involving chymotrypsinogen and trypsinogen. The relationship of trypsinogen to trypsin and the corresponding specific inhibitor has been studied in detail. Trypsin catalyzes not only the conversion of trypsinogen to trypsin, but also a second reaction in which trypsinogen is converted to an inert protein. The conversion to trypsin from the proenzyme is accelerated by a concentrated solution of magnesium or ammonium sulfate, or by enterokinase as well as by trypsin.

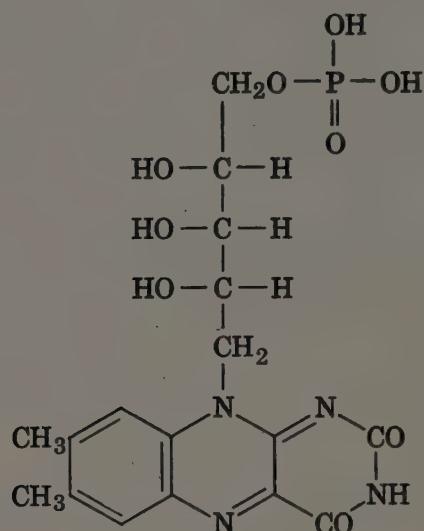
ACTIVATORS, COENZYMES, AND ANTIENZYMES

Substances of a nonspecific nature, and generally of small molecular weight, may function as *activators*. This is particularly true of such metal ions as manganese and magnesium. Hydrochloric acid, which serves to convert pepsinogen into pepsin, is sometimes spoken of as an activator.

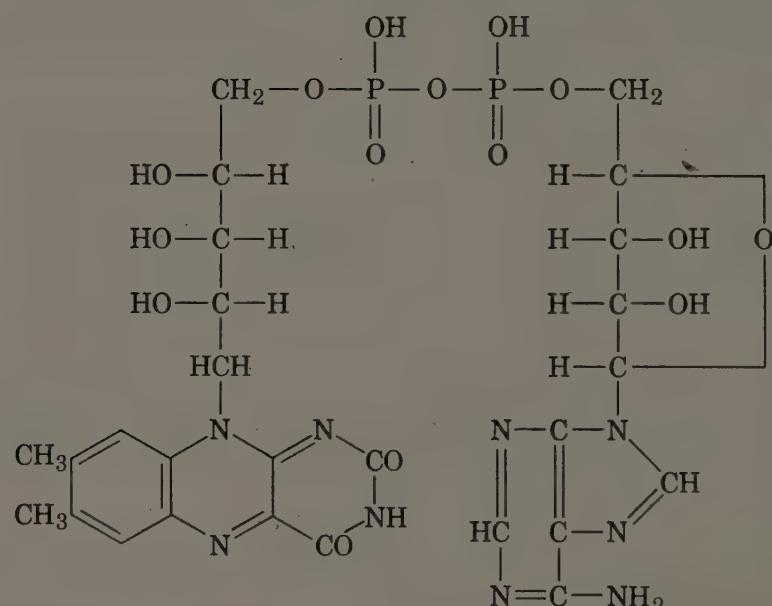
In the field of the peptidases, not only manganese ions and magnesium ions have been shown to be of importance but also zinc and cobalt. None of these activators function universally, but each has a partially specific enhancing effect. Of special interest is the fact that enzyme preparations studied by M. Johnson and co-workers were found to have greatly enhanced activity toward peptides containing D-amino acid residues after activation with manganous salts. It has been shown, furthermore, that L- and D-peptidases may be fractionated and separately activated.

The *coenzymes*, in contrast to activators, are molecules of very specific action and organic constitution. Their molecular weights range typically from 100 to several hundred. The prosthetic group of the conjugated protein is the coenzyme of an enzyme system. While such coenzymes appear to be lacking in many of the hydrolytic enzymes, such as pepsin and trypsin, they are of outstanding importance in other enzyme systems. Significant examples include the iron-porphyrin prosthetic group of the closely related series of oxidases, catalase, peroxidase, and cytochromes. All of these enzymes are

important in the oxidation-reduction processes of respiration. Many, perhaps all, vitamins function as moieties of coenzymes. Amino acid oxidases, for example, are known for both the L-form of amino acids and the D-form. The coenzyme of the mammalian L-amino acid oxidase is riboflavin phosphate. The coenzyme of D-amino acid oxidase is a prosthetic group, riboflavin adenine dinucleotide, which occurs in a number of enzymes. Each of these coenzymes is combined with a specific protein molecule.

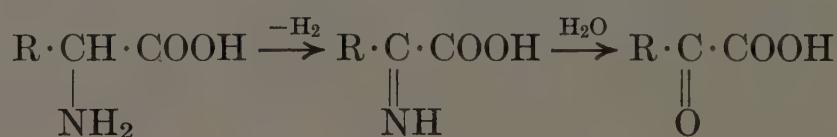


Riboflavin phosphate

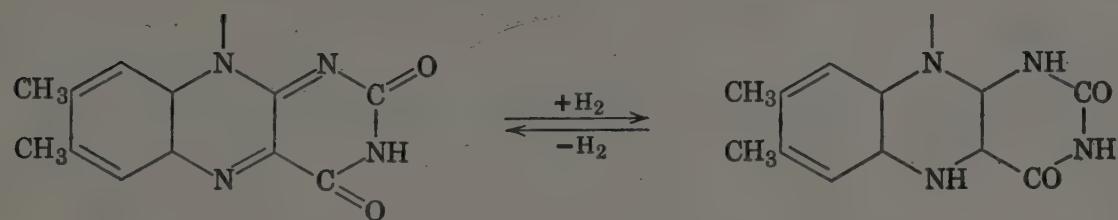


Riboflavin adenine dinucleotide

In the oxidation of D-amino acids, the first step of the sequence involves loss of hydrogen:



This reaction is catalyzed by D-amino acid oxidase, which readily accepts H₂, the riboflavin moiety undergoing the transformation:



Two main types of material have been considered as *antienzymes*. One of these is that of the antibodies which may be derived by employment of the enzyme as an antigen (Chap. 23). Although such preparations are difficult, antienzymes have been obtained in this manner for luciferase, the enzyme responsible for the oxidation of luciferin (the process which results in the light of fireflies), as well as for pepsin and urease.

Also known are specific natural polypeptides and proteins that inhibit proteolytic enzymes, for example the pepsin and trypsin inhibitors. The antitrypsin obtainable from the pancreas has been crystallized and has been shown to have a molecular weight of 40,000. Antitrypsins appear to be fairly widespread phylogenetically; they have been isolated from soybean and egg white, and the presence of such material has been demonstrated also in blood serum.

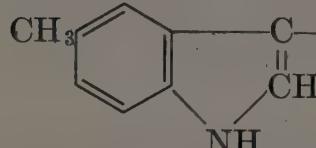
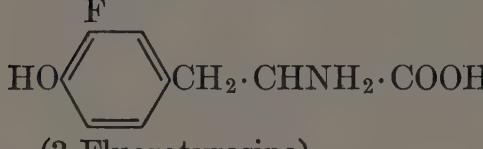
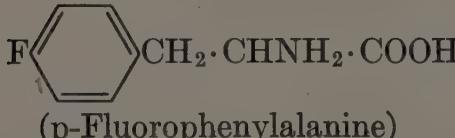
INHIBITORS

Although the antienzymes are certainly inhibitors, the latter term is usually applied to compounds of smaller structure. Virtually all of the reagents that precipitate or denature proteins, such as heavy metals, act as inhibitors of enzymes. Many of these reagents react with sulfhydryl groups, which are in turn essential to the action of a large number of enzymes, proteolytic or otherwise. A second group of inhibitors consists of those which act upon prosthetic groups. As prime examples may be cited the strong physiological poisons, hydrogen sulfide, and hydrocyanic acid, which react with the iron atom of the enzymes containing iron-porphyrin complexes. Another important type of inhibitor is the kind which competes with the substrate. These are known as *competitive inhibitors*.

Knowledge concerning such materials has grown tremendously from the initial impetus provided by sulfanilamide, the prototype of competitive inhibitors. Sulfanilamide competes with the structurally closely related *p*-aminobenzoic acid, which is essential in the synthesis of folic acid. Competitive inhibitors have been particularly useful

both in the development of therapeutically active substances, as in the sulfa drugs, and in the clarification of the fundamental knowledge of the normal substrates of various enzyme-controlled biosynthetic pathways.

TABLE 21-2. Some Antiamino Acids and Their Characteristics

Antiamino Acid	Metabolite	Organism Inhibited	Typical Inhibition Index ¹
$\text{CH}_3\text{O}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CHNH}_2\cdot\text{COOH}$ (Methoxinine)	Methionine	<i>E. coli</i> , <i>S. aureus</i>	500
$\text{CH}_3\cdot\text{CH}_2\cdot\text{S}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CHNH}_2\cdot\text{COOH}$ (Ethionine)	Methionine	Rat, <i>E. coli</i>	10
 (5-Methyltryptophan)	Tryptophan	<i>E. coli</i>	1000
 (3-Fluorotyrosine)	Tyrosine	<i>N. crassa</i>	7
 (p-Fluorophenylalanine)	Phenylalanine	<i>L. arabinosus</i>	2
 (Thienylalanine)	Phenylalanine	<i>S. cerevisiae</i>	2
$\text{HOOC}\cdot\text{CHOH}\cdot\text{CHNH}_2\cdot\text{COOH}$ (Hydroxyaspartic acid)	Aspartic acid	<i>E. coli</i>	12

¹ Antimetabolite/metabolite.

Antiamino Acids. Compounds which function successfully as competitive inhibitors are often found to be closely related to the structures with which they compete. Competitive inhibitors against amino acids are known by the more restrictive term *antiamino acids*. The antiamino acids have not been as successful therapeutically as the anti-vitamins, but they have been useful in studies of metabolism. Some of the antiamino acids are presented in Table 21-2.

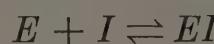
Competitive inhibitors have been studied in enzyme systems, but

usually within the integrated enzyme systems of organisms. This situation makes it possible to trace some of the steps in biosynthesis, as pointed out earlier (p. 108). Antiamino acids have also been studied principally on whole cells.

Other effects of interest are found with antiamino acids. Although *p*-fluorophenylalanine is the highly effective antiamino acid indicated by the low inhibition index in Table 21-2, a small proportion of cells in a subculture are resistant and soon dominate the population. Another interesting effect observed with the *p*-fluorophenylalanine in the presence of low levels of phenylalanine is that of growth stimulation. This may be explained on the basis of partial replacement of the metabolite amino acid; evidence for incorporation of the unnatural amino acid ethionine into rat protein has also been obtained with the tracer-containing compound.

Competitive Inhibition. Competitive inhibition exists when the degree of inhibition effected by a given concentration of inhibitor decreases as the concentration of substrate is increased. In other words, the two components, substrate and inhibitor, appear to compete. It seems probable that the inhibitor is combined by enzyme at the site normally occupied by substrate, so that there is actual competition for this site. In general agreement with this idea, it is found that competitive inhibitors usually bear a close structural similarity to the substrate.

It is assumed in this case that an equilibrium exists between enzyme *E* and inhibitor *I*, of the form



and that an equilibrium constant can be formulated analogous to the Michaelis-Menten constant for substrate combination. The situation can be treated in the same manner as in the uninhibited case, the derivation being only slightly more complex. The result is that the rate *v* at a given substrate concentration (*S*) and a given inhibitor concentration (*I*) is given by

$$v = \frac{V(S)}{K_m(1 + (I)/K_I) + (S)} \quad (7)$$

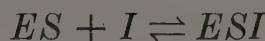
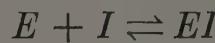
or in reciprocal form

$$\frac{1}{v} = \frac{K_m}{V} \left[1 + \frac{(I)}{K_I} \right] \frac{1}{(S)} + \frac{1}{V} \quad (8)$$

This is of the same form as in the uninhibited case, being the equation of a straight line when $1/v$ is plotted versus $1/(S)$. As (*I*) is varied,

a family of straight lines is obtained, all of the same intercept but differing in slope, the slope being $(K_m/V)[1 + (I)/K_I]$. This, then, is the test for competitive inhibition.

Noncompetitive Inhibition. In this case, it is presumed that the inhibitor reacts with an entirely different site than that occupied by the inhibitor. In accord with this notion, it is observed that inhibitors of this type usually bear little if any structural resemblance to the substrate. The inhibition equilibria may be represented as



since both E and ES may react with the inhibitor. Further, it is assumed that ESI is inactive, that is, does not yield product directly, and that the equilibrium constants for the two inhibition equilibria are identical, K_I . In this case the resulting equation is

$$v = \frac{V(S)}{[1 + (I)/K_I][K_m + (S)]} \quad (9)$$

or in reciprocal form

$$\frac{1}{v} = \left[1 + \frac{(I)}{K_I} \right] \left[\frac{K_m}{V} \left(\frac{1}{(S)} \right) + \frac{1}{V} \right] \quad (10)$$

It is now seen that both the slope and intercept of plots of $1/v$ versus $1/(S)$ are dependent on concentration of inhibitor.

STRUCTURE OF ENZYMES

It is natural that many workers should be attracted to studying enzyme structure by means of peptide analysis. The enzymes which have been most investigated in this fashion are proteolytic enzymes. A particular reason for focusing attention on this class is the availability of inactive protease precursors, resulting from the brilliant work of Northrop and associates. One of the dominant objectives of protein chemistry is discovery of the structure that imparts enzymic activity to an otherwise inert protein. Knowledge of structural differences between an inactive precursor and an active enzyme directly derived from the precursor, may be expected at least to set the stage for the discovery sought.

The terminal structures of a number of enzymes are presented in Table 21-3.

The data for chymotrypsin do not indicate which C-terminus is associated with either N-terminus. The existence of N-isoleucyl in

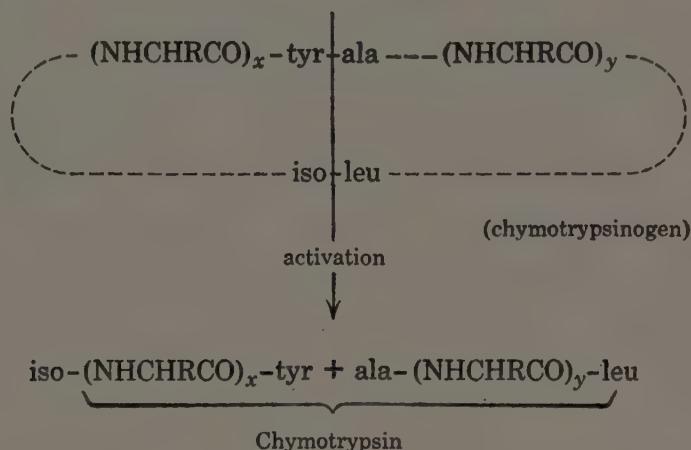
three proteases is striking in the comparative absence of this terminus for other proteins (p. 285). An evolutionary relationship at the molecular level is thus suggested (p. 432). The penultimate residue for papain is different, however, from the corresponding residue for the

TABLE 21-3. Terminal Sequences of Some Proteases

Protease	Sequence
Carboxypeptidase (bovine)	Asp(NH ₂)-ser...
Chymotrypsin (α , β , or γ)	Ala...]...tyr Iso-val...]...leu
Papain	Iso-pro-glu...
Pepsin	Leu-gly-asp-asp...
Trypsin	Iso-val...

other two. This fact does not necessarily indicate a large structural difference; the only structural variation in the bovine and porcine oxytocins and vasopressins is in two residues among eight. Further structural data on these proteases will be of particular interest.

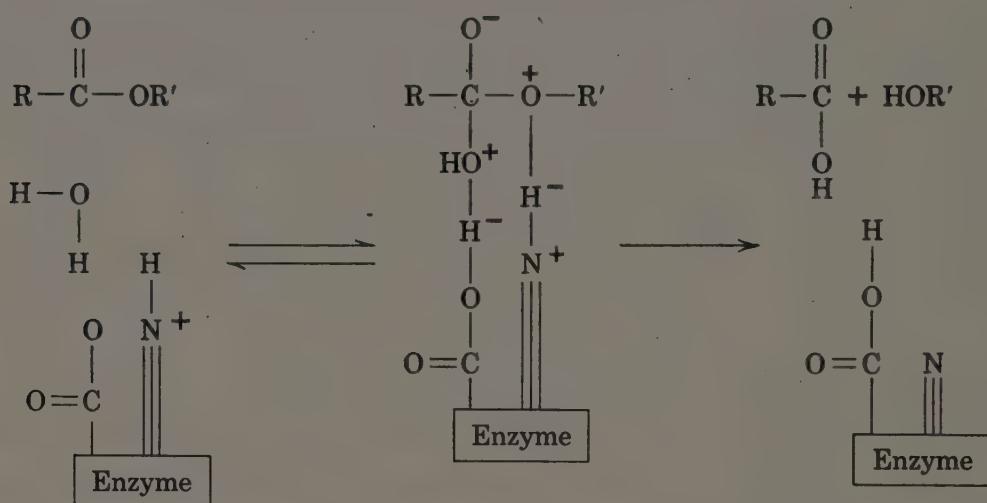
Differences in structure between proproteases (enzyme precursors) and their active molecular offspring have been found. Desnuelle found no termini in chymotrypsinogen. It has accordingly been assumed that the molecule is cyclic, but other explanations for the absence of assignable termini are possible. Upon activation, alanine and isoleucine appear as N-terminal residues, and tyrosine and leucine as C-terminal residues. This process could result from the opening of a bicyclic chymotrypsinogen, or from the simultaneous splitting of a cycle in two places. One partly conjectural picture of the latter type of process is



Studies of residue sequence have not yet led to a picture of "active centers" in enzymes. It may be that large areas of surface are involved. In oxidation-reduction reactions, it seems possible that combination of substrate may not take place directly on the surface

of the protein. There is considerable evidence to suggest that coordination to a metal is involved in some cases, the metal acting as a link between protein and substrate. A similar possibility has been suggested for certain of the peptidases. In other cases, particularly those involving oxidases, some of the "specificity" is known to reside in the coenzyme.

Another interesting suggestion is that the unusual efficiency of enzymes of the hydrolytic type, especially of esterases and peptidases, is due to a push-pull type of mechanism. It is postulated, for example, that two key groupings are involved, one of which supplies a proton, the other a hydroxyl ion. This may be illustrated in the case of an esterase, as follows:



Proteins, possessing as they do both free amino and free carboxyl groups, could easily be imagined to function in such a manner. The fact that fumarase action appears to involve two imidazole groups, one each in the acidic and basic forms, has been mentioned (p. 386). This has been interpreted on the basis of a "push-pull" mechanism. The mechanism in itself does not provide an answer to the question of the unusual ability of proteins to catalyze reactions.

Enzyme activation appears to be accompanied by an increase in entropy, as was indicated in connection with the discussion of temperature effects. It has been suggested that this might result from an increase in the randomness of the polypeptide configuration during activation, in a sense a slight reversible denaturation. This suggestion merits much further consideration and investigation.

A final possible mechanism arises from the suggestion that there may exist in proteins a relatively high degree of electronic mobility perpendicular to the peptide chains. It is imagined that the interchain hydrogen bonds (assuming configurations of the β -keratin or pleated

sheet-type) might be shortened to such an extent that there would be overlap of the outer electronic orbitals in the O, H, and N atoms involved. There is little actual evidence for such a suggestion, but it should not be disregarded. Chemical reactions are basically electronic rearrangements, and this mechanism would provide a means whereby electrons could be readily moved from one part of the substrate molecule to another, or even between one or more substrates.

SELECTED REFERENCES

Books and Reviews

"Solubility Curves of Pure Protein and of Mixtures and Solid Solutions of Proteins," M. Kunitz and J. H. Northrop, *Cold Spring Harbor Symposia Quant. Biol.*, **6**, 325 (1938).

"Metabolite Antagonists," R. O. Roblin, Jr., *Chem. Revs.*, **38**, 255 (1946).

Crystalline Enzymes, 2nd ed., J. H. Northrop, M. Kunitz, and R. M. Herriott, Columbia University Press, New York, 1948.

"Metabolic Inhibitors," R. J. Winzler, *Ann. Rev. Biochem.*, **18**, 535 (1949).

Outlines of Enzyme Chemistry, J. B. Neilands and P. K. Stumpf, John Wiley & Sons, New York, 1955.

Original Articles

"Die Kinetik der Invertwirkung," L. Michaelis and M. L. Menten, *Biochem. Z.*, **49**, 333 (1913).

"Über Pankreas-Amylase," R. Willstätter, E. Waldschmidt-Leitz, and A. R. F. Hesse, *Hoppe-Seyler's Z. physiol. Chem.*, **126**, 143 (1923).

"The Isolation and Crystallization of the Enzyme Urease," J. B. Sumner, *J. Biol. Chem.*, **69**, 435 (1926).

"Crystalline Trypsin. I. Isolation and Tests of Purity," J. H. Northrop and M. Kunitz, *J. Gen. Physiol.*, **16**, 267 (1932).

"Crystalline Trypsin. II. General Properties," J. H. Northrop and M. Kunitz, *J. Gen. Physiol.*, **16**, 295 (1932).

"The Determination of Enzyme Dissociation Constants," H. Lineweaver and D. Burk, *J. Am. Chem. Soc.*, **56**, 658 (1934).

"Evidence of the Physiological Specificity of Methionine in Regard to the Methylthiol Group: The Synthesis of S-Ethylhomocysteine (Ethionine) and a Study of Its Availability for Growth," H. M. Dyer, *J. Biol. Chem.*, **124**, 519 (1938).

"Kinetics of the Formation of Chymotrypsin from Crystalline Chymotrypsinogen and of Trypsin," M. Kunitz, *Enzymologia*, **7**, 1 (1939).

"The Kinetics of the Enzyme-Substrate Compound of Peroxidase," B. Chance, *J. Biol. Chem.*, **151**, 553 (1943).

"Studies in Chemotherapy. VIII. Methionine and Purine Antagonists and Their Relation to the Sulfonamides," R. O. Roblin, Jr., J. O. Lampen, J. P. English, Q. P. Cole, and J. R. Vaughan, Jr., *J. Am. Chem. Soc.*, **67**, 290 (1945).

"L- und D-Dipeptidasen," R. Merten, *Biochem. Z.*, **318**, 167 (1947).

"The Competitive Inhibition of the Metabolism of Alpha-Amino Acids by

Their Halogenated Analogs," H. K. Mitchell and C. Niemann, *J. Am. Chem. Soc.*, **69**, 1232 (1947).

"The Effect of the Soybean Trypsin Inhibitor on the Enzymatic Release of Amino Acids from Autoclaved Soybean Meal," I. E. Liener and H. L. Fevold, *Arch. Biochem.*, **21**, 395 (1949).

"Separation of Dehydropeptidase and Analogous L- and D-Peptidases," V. E. Price, A. Meister, J. B. Gilbert, and J. P. Greenstein, *J. Biol. Chem.*, **181**, 535 (1949).

"Étude des extrémités N-terminales du trypsinogène et de la trypsine de boeuf," M. Rovery, C. Fabre, and P. Desnuelle, *Biochim. et Biophys. Acta*, **9**, 702 (1952).

"Carboxyl Terminal Groups of Proteolytic Enzymes. I. The Activation of Chymotrypsinogen to α -Chymotrypsin," J. A. Gladner and H. Neurath, *J. Biol. Chem.*, **205**, 345 (1953).

"The Syntheses and Microbiological Properties of Acetylenic Amino Acids. Propargylglycine and 2-Amino-3-methyl-4-pentynoic Acid," H. Gershon, J. Shapira, J. S. Meek, and K. Dittmer, *J. Am. Chem. Soc.*, **76**, 3484 (1954).

"N-Terminal Group in Chymotrypsinogen," F. R. Bettelheim, *J. Biol. Chem.*, **212**, 235 (1955).

"The Effect of pH on Fumarase Activity in Acetate Buffer," C. Frieden and R. Alberty, *J. Biol. Chem.*, **212**, 859 (1955).

"Structural Changes in the Activation of Chymotrypsinogen and Trypsinogen. Effect of Urea on Chymotrypsinogen and δ -Chymotrypsin," H. Neurath, J. A. Rupley, and W. J. Dreyer, *Arch. Biochem. and Biophys.*, **65**, 243 (1956).

Peptides and Proteins as Substrates

Action pattern on synthetic substrates

Action pattern of carboxypeptidase

Classification of proteases

Action pattern on large molecules

Enzymic synthesis of peptide bonds

Attempts to reduce the problems of biological specificity to the molecular level have received attention in a number of ways. Outstanding are studies on immune reactions, enzyme specificity, and virus chemistry. Much effort has been expended in an attempt to understand the relationship of proteolytic enzymes to arrangement of amino acid residues in substrates. Much of the early work in this field of endeavor was performed by Fischer and by Abderhalden.

Fischer, Abderhalden, and their contemporaries were limited in the scope of their studies by the infeasibility of preparing a sufficient number of peptide substrates. The carbobenzoxy synthesis of Bergmann and Zervas, however, provided a rich variety of peptides containing all types of amino acid residue. With this powerful tool, inquiry was reinstated. Many suppositions of long standing were found to be groundless or to require modification, and new facts and correlations emerged.

Despite maturation of knowledge in this field with the study of synthetic substrates, the primary interest of the biochemist has stemmed from the more fundamental question of the way in which proteolytic enzymes attack the larger substrates that typify those

found in nature. This type of investigation became practicable in the late 1940's with the development of methods of terminal residue analysis. These newer methods of analysis have been extensively applied to elucidating the pattern of action of proteolytic enzymes on proteins and large peptides.

The shift in emphasis from synthetic to natural substrates does not involve an abandonment of the earlier approach. Each type of study supports the other. Despite the productivity of the carbobenzoxy synthesis in providing new peptides, only a very small fraction of the theoretical total number of L-tetrapeptides, for example, has been synthesized. Studies on natural substrates now point toward the desirability of synthesizing a relatively small number of critically selected peptides for investigation. With integrated research activity on both small and large substrates, new advances can be expected.

ACTION PATTERN ON SYNTHETIC SUBSTRATES

Specificity. The details of the action of proteolytic enzymes on their substrates have often been referred to collectively as "specificity." The expert recognizes many qualifications to the word as it is otherwise used. The interactions of complex enzyme systems may lead, at the biological level, to what may properly be called specificity. Among individual substrates, however, "all-or-none" effects are not the general rule. One can recognize different rates of reaction on different closely related substrates. Accordingly, these relationships have been spoken of more aptly as preference (of enzyme) or susceptibility (of substrate).

The scope of this qualified "specificity" appears to vary with the enzyme. Carboxypeptidase and chymotrypsin, for example, are more sharply limited in their range of action than pepsin or papain. Conditions of the reaction such as buffer concentration affect the result markedly in some cases; a range of conditions has seldom been explored in proteolytic experiments.

The factors which constitute an evaluation of specificity are brought into focus when one examines the relationship between proteolytic enzymes and their substrates. The peptide substrates represent the one large class of currently available compounds which comprise a series of very closely related substrates. For this reason they have supplied the raw material from which specificity could be evaluated.

Terminology. A frequently used synonym for the large class of peptide-hydrolyzing enzymes known as proteolytic enzymes is *protease*. Proteases, classically, are of two main types: *proteinase*, which acts upon protein molecules, and *peptidase*, which acts upon peptides.

Inasmuch as all proteins are in a sense peptides, this classifying terminology does not provide mutually exclusive categories.

One of the principal results of the studies made possible by the carbobenzoxy synthesis of peptides was the demonstration that the traditional classification was not fully valid. When substrates were synthesized with a suitable arrangement of properly chosen residues, proteinases such as pepsin and chymotrypsin were found to be active on simple derivatives of small peptides. Trypsin, for example, is now known to act upon benzoyllysineamide to form benzoyllysine and ammonia.

On the other hand, peptidases are not without action on proteins. If this were not true, carboxypeptidase could not be used at all for structural studies on proteins (p. 147). These persistent ambiguities in the use of the terms proteinase and peptidase underline the difficulties of distinguishing peptides and proteins (p. 132).

Types of Substrate Influence. The principal kinds of structural function which have been shown to influence protease activity are:

- (1) Terminal groups such as amino or α -carboxyl. A subgroup of enzymes has been classified on this basis as *aminopeptidase* and *carboxypeptidase*. Evidence for a *prolidase* and a *prolinase* has been obtained.
- (2) Type of amino acid residue; for example, basic or acidic, monoaminomonocarboxylic or benzenoid. Early classification of enzymes on the basis of whether or not they function with basic or acidic substrates has partly withstood the test of time. Such activity can be understood best in terms of charge effects.

(3) Individual amino acid residue. This aspect of the structure was formerly referred to as a determinant of specificity, but is now recognized as a determinant of susceptibility. The difference in rates when the amino acid residue differs is often substantial.

(4) Positional effects of these structures relative to bond affected. A particular residue may determine whether or not an adjacent peptide bond is hydrolyzed rapidly under the given experimental conditions. However, this effect is normally manifested on only one of the two adjacent peptide bonds. As can be seen in Table 22-2, there is at least one known instance in which two enzymes (chymotrypsin and carboxypeptidase) can act on the same type of substrate, but in which the activity is expressed on one side of the residue for one enzyme and on the other side for the other enzyme. The configurational relationship of the residue to the peptide chain is also critical.

Antipodal Specificity. Comparison of substrates containing L-amino acid residues and those containing D-amino acid residues has revealed

a dominance of proteolysis of the L-forms, that is, the same forms found in the enzymes themselves. This specificity is found to apply to most proteases that have been studied. (There appear to be, however, some special proteases which act preferentially upon D-forms.)

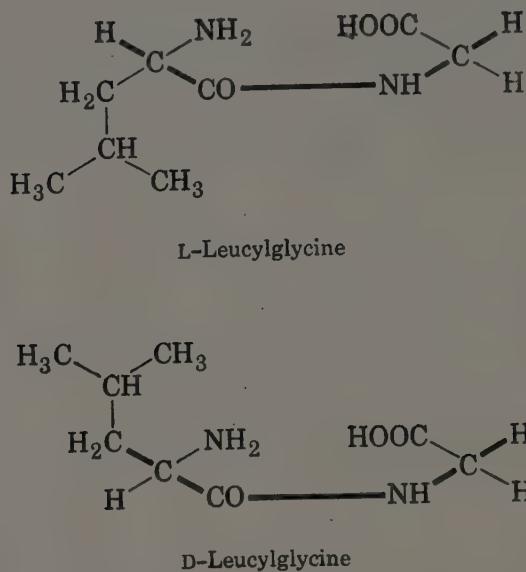


Fig. 22-1. According to the polyaffinity concept, essential groups in the substrate comprise a hexagonal plane that is capable of functional interaction with fixed groups in the enzyme. In the L-substrate, the isobutyl side chain is away from the enzyme attracted from above the hexagon. In the D-substrate, the voluminous isobutyl group intervenes and fruitful interaction is prevented. With smaller side chains, such as methyl in D-alanine residues, only partial interference is found. From M. Bergmann, L. Zervas, J. S. Fruton, F. Schneider, and H. Schleich, *J. Biol. Chem.*, **109**, 336 (1935).

The antipodal specificity has been explained by projectional considerations of three dimensional substrate molecules. In these, the side chains of the residues of L-forms are directed away from the enzyme, and permit enzyme-substrate attraction to function so that proteolysis may occur. It has long been known that the peptides containing corresponding D-forms are not readily hydrolyzed. This has been pictured as a steric interference resulting from intervention of the side chain between enzyme and substrate. Bergmann found that the extent of inhibition of hydrolysis was a function of the size of the side chain in the D-amino acid residue, being greater with larger side chains. Such relationships are illustrated in Fig. 22-1.

From results of this sort the inference has been drawn that the enzyme must interact with the substrate at several points. This concept, which has been called the "polyaffinity theory," has received experimental support from a number of directions. It may be looked upon as proof for an enzyme-substrate interaction, comparable in compulsion to the conclusions from Michaelis-Menten kinetics. The

polyaffinity theory has proved to be of value beyond the immediate borders of protein biochemistry. Ogston, for example, has extended this concept to a clarification of knowledge on citric acid metabolism.

Information of the type presented has been accumulated by subjecting individual peptides of closely related types to the action of proteases. A more detailed treatment is given for but one of the proteases, bovine carboxypeptidase.

ACTION PATTERN OF CARBOXYPEPTIDASE

Exemplary substrates tested in a number of laboratories are found in Table 22-1.

TABLE 22-1. Carboxypeptidase specificity

Substrate	Hydrolysis (+ or -)
Carbobenzoxyglycyl-L-phenylalanine	+
Carbobenzoxyglycyl-D-phenylalanine	-
Carbobenzoxyglycyl-L-alanine	+ (slow)
Carbobenzoxy-L-glutamyl-L-phenylalanine	+
Carbobenzoxy-L-glutamyl-L-phenylalanineamide	-
Glycylglycine	-
Benzoylglycylglycine	+
Chloroacetyl-L-tyrosine	+
Chloroacetyl-N-methyl-L-tyrosine	-
DL-Leucylglycyl-L-tyrosine	+

Results expressed in this table and others lead to the following conclusions, some of which are closely analogous to conclusions reached with other proteases:

- (1) The enzyme exhibits antipodal specificity in its attack upon the substrate, the L-form of the latter being favored.
- (2) The terminal amino acid residue bearing the carboxyl group is most rapidly attacked if the residue is that of phenylalanine or tyrosine. This is the property which has frequently been referred to as specificity but is more accurately represented as susceptibility (Greenstein). That the differences are quantitative rather than qualitative may be seen in the fact that the analogous alanine derivative is also attacked, but slowly.
- (3) The terminal free carboxyl group is essential. The enzyme received its name from this fact.
- (4) A primary amino group on the same carbon atom as a peptide bond inhibits hydrolysis.
- (5) The peptide hydrogen must be unsubstituted. This type of

requirement applies to many proteases, but exceptions to it have been found, some of these with carboxypeptidase.

(6) In peptides larger than dipeptides, the bond adjacent to the carboxylic terminal residue is the one hydrolyzed.

CLASSIFICATION OF PROTEASES

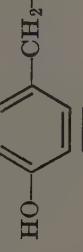
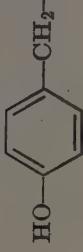
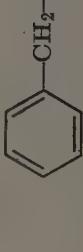
Bergmann, Fruton, Smith, Neurath, and others have accumulated information on the types of structures which are attacked by proteolytic enzymes. Table 22-2 summarizes some of the more salient results. The table represents a modification of one of Bergmann's condensed tables. The information was accumulated from experiments of the type detailed for carboxypeptidase.

One of the main points inferrible from Table 22-2 is that the action of the so-called proteinases is not restricted to molecules the size of those of a protein, as had earlier been supposed. Trypsin, for example, acts rapidly upon a substrate containing but one amino acid residue, that of arginine or lysine. The α -amino and α -carboxyl groups must, however, be masked by substituents. The enzyme thus acts upon interior linkages, a phenomenon which helps to explain why earlier workers could easily have concluded that the normal substrate was protein, in which interior linkages are dominant. It is also essential that the residue be of the correct type. This requirement alone appears to be adequate to the interactions in which pepsin is active. In the case of this enzyme, the carboxyl group of an N-substituted dipeptide may be free, providing the C-terminus is phenylalanine or tyrosine.

At one place in the table, chymotrypsin, the enzyme is seen to split ester linkages as well as peptide bonds. This type of activity has been observed by Neurath and by others for several of the proteolytic enzymes. In some cases the esterase activity is hundreds of times as great as the rate on more usual substrates. This phenomenon has led to the belief that these enzymes play a role with ester linkages in protein structure. There is some additional circumstantial evidence for ester linkages in protein, including the fact that phosphate ester linkages are found. The reality of these latter is supported by the isolation of phosphoserine from protein.

Although for each enzyme a typical R-group is illustrated, other amino acid residues participate in susceptible peptide linkages. Evidence exists also for other enzymes. There is some indication of the existence of carboxypeptidases which act typically on amino acid residues other than the ones given in the table. The residues in the

TABLE 22-2. The Specificity of Proteolytic Enzymes

Protease	Typical Substrate	R-Group of Substrate	Peptide Chain of Substrate ³	Cysteine Activation
Pepsin	Cbz-o-L-glutamyl-L-tyrosine ¹	HO—  —CH ₂ — or  —CH ₂ —	...CO—[NH·CHR·COOH]	0
Cathepsin I	Cbz-o-L-glutamyl-L-phenylalanine		...COOH + H ₂ N·CHR·COOH	0
Trypsin		H ₂ N(CH ₂) ₄ — or H ₂ N(C(NH)NH(CH ₂) ₃ —	...CO·NH·CHR·CO NH ₂ or ...CO·NH·CHR·CO· OR ₁ ...	0 +
Cathepsin II	Bz-L-lysine-amide ²			
Papain	Bz-L-arginine-amide			
Chymotrypsin	Bz-L-tyrosine ethyl ester	HO—  —CH ₂ — HO—  —CH ₂ —	...CO·NH·CHR·CO·NH... or ...CO·NH·CHR·CO· OR ₁ ...	
				0
Leucylpeptidase	L-Leucylglycine	CH ₃ CH ₃ — CH—CH ₂ —	CO·NH·CHR·COOH + H ₂ N... or HOR ₁	
Cathepsin III	L-Leucinamide	CH ₃	H ₂ N·CHR·CO NH ₂ ...	0
Carboxypeptidase	Cbz-o-glycyl-L-tyrosine	HO—  —CH ₂ —	...CO NH·CHR·COOH	0
Cathepsin IV	Cbz-o-glycyl-L-phenylalanine	HO—  —CH ₂ —	...COOH + H ₂ N·CHR·COOH	+

¹ Cbz = carbobenzoxy. ² Bz = benzoyl.

³ Broken line indicates bond split.

table are susceptible particularly to beef pancreatic carboxypeptidase. This point merits further investigation.

The table indicates that the enzymes catalyze attainment of equilibria involving peptide bonds within relatively specific structures. The enzymes thus catalyze both hydrolysis and synthesis. This behavior is in accord with the theoretical considerations (Chap. 21) but evidence was first amply provided by Bergmann and Fraenkel-Conrat. Proteases may indeed be looked upon primarily as transfer enzymes, with synthesis and hydrolysis constituting special cases of transfer action.

Table 22-2 also illustrates that for each type of digestive protease there is a corresponding cathepsin (intracellular protease). These cathepsins have action patterns similar to those of the extracellular proteases, although they differ in some respects (activatability by sulfhydryl reagent, for example). Bergmann was led to suggest that the roster of enzymes in the gastrointestinal tract accompanied evolution from unicellular to multicellular creatures.

Some of the unsolved problems in this field are exemplified by the fact that pepsins from pig, cow, sheep, and chicken hydrolyze carbobenzoxy-L-glutamyl-L-tyrosine, but pepsin from salmon does not.

ACTION PATTERN ON LARGE MOLECULES

The action of proteolytic enzymes on proteins and large peptides has received major attention since the late 1940's. Some of these studies have been quantitative, others essentially qualitative. In the course of Sanger's structural study of insulin, much information of the latter type has been provided. A sample of such information is given in Table 22-3. The results are quantitative only to the extent that major and minor degrees of splitting are distinguished.

The table reveals that many residues are exposed other than those which would be expected from studies on synthetic substrates.

Comparison of Results on Natural and Synthetic Substrates. Integration of many of the studies on large substrates (Sanger's work is one such investigation) permits the following conclusions:

- (1) Some of the specificities or strong preferences observed with synthetic substrates are reflected in the larger natural molecules. Frequently, for example, C-terminal tyrosine or phenylalanine is exposed by chymotrypsin in proteins as it is in synthetic substrates.
- (2) Failure of particular residues in synthetic substrates to respond to a given enzyme is often reflected in the large substrate.
- (3) In many other instances, stability observed in small model substrates is not reflected in the large substrates. These instances may

TABLE 22-3. Proteolysis of Insulin Fraction A

Bonds Split by	Gly-iso-val-glu-glu-CysO ₃ H-CysO ₃ H-ala-ser-val-CysO ₃ H-ser-leu-tyr-glu-asp-tyr-CysO ₃ H-asp	Chymotrypsin (C)	Pepsin (Pe)	Papain (Pa)	Major sites of action	Minor sites of action
	NH ₂	C				
	NH ₂	C	Pe	Pa		
	NH ₂	C	Pe	Pa		
	NH ₂		Pe	Pa		
	NH ₂			Pa		
	NH ₂					

be explained by the fact that only a minor proportion of the conceivable types of peptide have been synthesized and tested. In addition, half of the observed exposures might be incidental to the preference expressed for the other residue linked to the one opened. For example, in Table 22-3 the opening of C-terminal leucine might not be anticipated from studies of models. It could result, however, from the preference of pepsin for amino tyrosine in the -leu-tyr- sequence.

(4) In still other instances, stability of linkages involving particular residues in large molecules is in contrast to the susceptibility of similar small substrates. Some of the preceding reasons might explain this difference. This category would be explainable especially on the basis that large molecules could contain structural rigidity and inaccessibility that are necessarily absent from small molecules.

(5) Most striking is the similarity of action pattern of different proteases on the same substrate. This similarity has been apparent particularly in studies in which relatively full accounting of fissions is possible (Hurst and Fox, 1956). Again, Table 22-3 provides an illustration of this phenomenon, which was, however, first noted in other studies. Papain from plants and pepsin from animals typically act upon different types of substrate (Table 22-2), but of eight linkages split by papain and of seven split by pepsin, six are split by both papain and pepsin.

Concerted Activity of Digestive Proteases. The proteases of the human gastrointestinal tract constitute what might be poetically called a digestive "disassembly line." The first important proteolytic action occurs under the influence of pepsin, in the stomach. Classically, pepsin has been considered to be an enzyme which fragments large proteinaceous molecules to smaller ones. A similar action has been ascribed to trypsin, found in the secretion of the pancreas. Along with the trypsin is found *erepsin*, which is a mixture of carboxypeptidase, aminopeptidase, dipeptidase, and other peptidases. The picture involves, first, fragmentation by proteinases, followed by further breakdown of the fragments by peptidases, each protease acting on the type of substrate that characterizes its activity. The picture is one of sequential action as well as of integrated activity.

No known combination of enzymes acting either alone or simultaneously has been shown to degrade protein entirely to amino acids *in vitro*. The practical limit has been found to be breakage of about two thirds of the peptide bonds. This incomplete hydrolyzability may be explained in terms of the specificity of various enzymes, which is not broad enough to hydrolyze all bonds, or perhaps it may be attributed to partial resynthesis (p. 406). The extent of break-

down by any one protease cannot be accounted for on the basis of tabulated specificities alone (see Beloff and Anfinsen, 1948).

ENZYMIC SYNTHESIS OF PEPTIDE BONDS

Inasmuch as protein synthesis is central to growth, and peptide bonds are the backbone of a protein molecule, the synthesis of peptide bonds by enzymes has been a matter of much biological interest. Although the problem is not fully defined, much less solved, several subproblems have been stated. These include such questions as how is the energy requirement met, how are specific end products formed, what is the role of nucleic acids, and do proteolytic enzymes participate?

A number of hypotheses for the biological synthesis of peptide bonds have been brought forward. As stated before, a successful hypothesis must account for the energy requirements of the synthesis. Studies of proteinaceous materials and proteolytic enzymes in glassware have led to hydrolytic breakdown, because the equilibria favor hydrolysis. The free energy necessary for synthesis of a peptide bond has been determined from thermodynamic data on reactants yielding products containing peptide bonds. From such data and from the fundamental thermodynamic relation

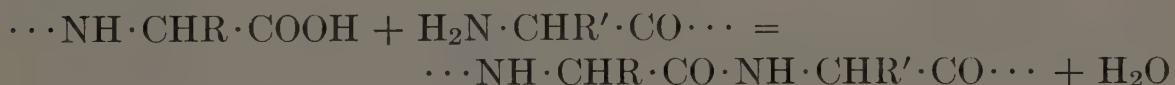
$$\Delta F^\circ = \Delta H^\circ - T\Delta S^\circ$$

it has been possible to calculate the free energy necessary as being in the range of 2000 to 4000 cal. per mole. From the equation

$$\Delta F^\circ = -RT \ln K = -RT \ln \frac{\text{peptide}}{(\text{amino acid}_1)(\text{amino acid}_2)}$$

the corresponding equilibrium constant has been shown to require that the peptide bond equilibrium be about 99 per cent on the side of hydrolysis (99 per cent expressed as a K value) for dipeptides.

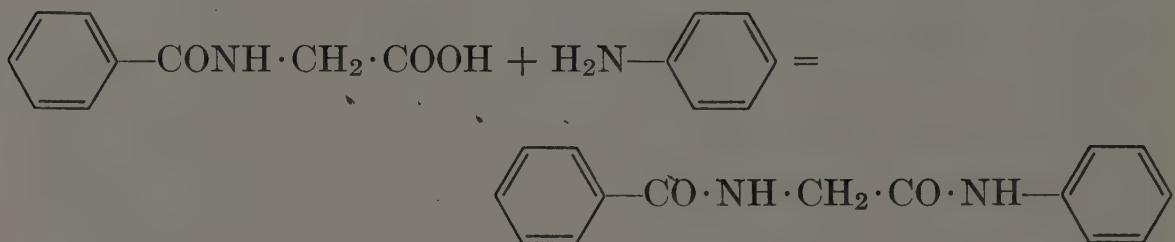
For the reaction



it is therefore necessary biologically that energy be introduced into the system.

Many attempts have been made to devise experimental models of enzymic synthesis of peptide bonds, of such a nature that the energy considerations are cared for. A most successful model, in an experimental sense, was devised by Bergmann and Fraenkel-Conrat. They selected as reactants compounds which would yield products with

solubilities less than that required by the equilibrium constant. For the reaction



the equilibrium constant is given by

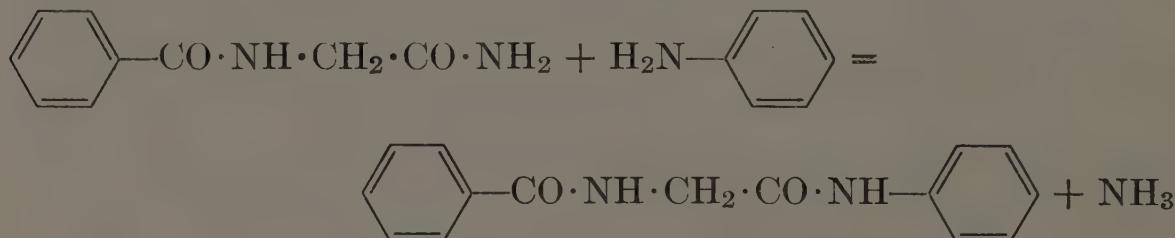
$$K = \frac{\text{hippurylanilide}}{(\text{hippuric acid}) (\text{aniline})}$$

In the presence of papain this reaction proceeds to the right because of the low solubility of the anilide. Hippurylanilide separates until either the enzyme has been dissipated or the concentrations of the reactants drop to equilibrium values. Such anilide syntheses have been performed for a variety of amino acid residues. It has been possible to replace the aniline by properly chosen amino acid derivatives so that both groups participating in peptide bond formation are amino acid residues. Other sulfhydryl-activatable proteases and chymotrypsin also function in this manner.

The solubility of the product determines the extent to which peptide bond synthesis will take place at equilibrium. The extents of the reactions that do proceed, however, have been shown to be primarily dependent upon the rates of such reactions; that is, equilibrium is not attained in most studies.

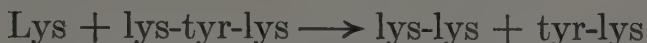
This type of model has demonstrated experimentally that proteases catalyze *in vitro* synthesis as well as hydrolysis of peptide bonds, has provided information on factors influencing enzyme reaction, and has contributed, with relative convenience, to knowledge of the specificity of peptide bond equilibria.

Transpeptidation. Bergmann and Fraenkel-Conrat recognized, in their first experiments on anilide formation, that hippurylamide reacts faster with aniline than does hippuric acid:



They formulated this reaction as a direct transfer without intermediate hydrolysis.

Fruton in the United States and Waley and Watson in England have particularly studied extensions of this discovery. Waley and Watson reported many such reactions, of which



is one example.

These direct transfers are known as *transpeptidation*. A large body of evidence for transfer activity of enzymes active on polysaccharides in a similar way has also been accumulated. Hydrolysis by such enzymes may be considered to be special instances of transfer, in which water is involved. Transpeptidation may be of significance in a metabolism that is recognized as dynamic. The evidence for transpeptidation from *in vitro* experiments is considerably more extensive than for transpeptidation *in vivo*. The *in vitro* operation of this mechanism assumes much importance when proteases are employed to prepare fragments of peptides for assignment of sequences of residues. Concordance of results with various proteases in sequence studies suggests, however, that transpeptidation is not an influence to be concerned about.

SELECTED REFERENCES

Reviews

- "The Enzymic Synthesis of Protein," H. Wasteneys and H. Borsook, *Physiol. Revs.*, **10**, 35 (1930).
- "A Classification of Proteolytic Enzymes," M. Bergmann, *Advances in Enzymol.*, **2**, 49 (1942).
- "The Enzymatic Properties of Peptidases," M. J. Johnson and J. Berger, *Advances in Enzymol.*, **2**, 69 (1942).
- "Peptides, Peptidases and Diketopiperazines," J. P. Greenstein in Schmidt's *Chemistry of the Amino Acids and Proteins*, p. 1080, Charles C Thomas, Springfield, Illinois, 1944.
- "Catalytic Action of the Metal Peptidases," E. L. Smith, *Federation Proc.*, **8**, 581 (1949).
- "Proteins, Lipids and Nucleic Acids in Cell Structures and Functions," A. Claude, *Advances in Protein Chem.*, **5**, 423 (1949).
- "The Mode of Action of the Crystalline Pancreatic Proteolytic Enzymes," H. Neurath and G. W. Schwert, *Chem. Revs.*, **46**, 69 (1950).

Original Articles

- "The Individuality of the Trypsin Complex. I. Comparative Studies on the Behavior of Various Substrates toward Trypsin-Kinase Preparations at Different pH Values," E. Abderhalden and O. Herrmann, *Fermentforschung*, **10**, 474 (1929).
- "Enzymic Histochemistry. II. Peptidase Distribution in the Root and Cotyle-

don of Malt Grain," K. Linderstrom-Lang and H. Holter, *Hoppe-Seyler's Z. physiol. Chem.*, **204**, 15 (1932).

"The Influence of the Oxygen Tension on the Reversal of Proteolysis (Protein Synthesis) in Certain Malignant Tumors and Normal Tissues," C. Voegtlin, M. E. Maver, and J. M. Johnson, *J. Pharmacol. Exptl. Therap.*, **48**, 241 (1933).

"The Role of Specificity in the Enzymic Synthesis of Proteins—Synthesis with Intracellular Enzymes," M. Bergmann and H. Fraenkel-Conrat, *J. Biol. Chem.*, **119**, 707 (1937).

"Bacterial Proteases. XII. Dipeptidases of Anaerobic Bacteria," E. Maschmann, *Biochem. Z.*, **302**, 332 (1939).

"The Specificity of Carboxypeptidase," K. Hofmann and M. Bergmann, *J. Biol. Chem.*, **134**, 225 (1940).

"Specificity of Salmon Pepsin," J. S. Fruton and M. Bergmann, *J. Biol. Chem.*, **136**, 559 (1940).

"Investigation of the Occurrence of Polypeptidases in the Central and Peripheral Nervous Systems," E. Abderhalden and G. Caesar, *Fermentforschung*, **16**, 255 (1940).

"Antipodal Specificity in the Inhibition of Growth of *Lactobacillus arabinosus* by Amino Acids," M. Fling and S. W. Fox, *J. Biol. Chem.*, **160**, 329 (1945).

"Degradation of Cystine Peptides by Tissues. IV. Dehydropeptidase Activity in Normal and Neoplastic Tissues," J. P. Greenstein and F. M. Leuthardt, *J. Natl. Cancer Inst.*, **6**, 197 (1946).

"A Manometric Method for Determining the Kinetics of an Enzymatic Hydrolysis of Peptides," P. C. Zamecnik and M. L. Stephenson, *J. Biol. Chem.*, **169**, 349 (1947).

"The Proteolytic Enzymes of Animal Tissues. VII. A Peptidase of Calf Thymus," J. S. Fruton, V. A. Smith, and P. E. Driscoll, *J. Biol. Chem.*, **173**, 457 (1948).

"The Products of Proteolysis of Some Purified Proteins," A. Beloff and C. B. Anfinsen, *J. Biol. Chem.*, **176**, 863 (1948).

"Structural Requirements of Specific Inhibitors for α -Chymotrypsin," S. Kaufman and H. Neurath, *J. Biol. Chem.*, **181**, 623 (1949).

"The Specificity of Leucine Aminopeptidase. II. Optical and Side-Chain Specificity," E. L. Smith and W. J. Polglase, *J. Biol. Chem.*, **180**, 1209 (1949).

"A New Series of Substrates for the Evaluation of Chymotrypsin Activity," B. M. Iselin, H. T. Huang, R. V. MacAllister, and C. Niemann, *J. Am. Chem. Soc.*, **72**, 1729 (1950).

"Chemical Preparation of Asymmetrically Labeled Citric Acid," P. E. Wilcox, C. Heidelberger, and V. R. Potter, *J. Am. Chem. Soc.*, **72**, 5019 (1950).

"Enzymic Synthesis of Peptide Bonds. II. "Preferences" of Papain within the Monoaminomonocarboxylic Acid Series," S. W. Fox, C. W. Pettinga, J. S. Halverson, and H. Wax, *Arch. Biochem.*, **25**, 21 (1950).

"On a New Enzymatic Peptide Synthesis," M. Brenner, H. R. Mueller, and R. W. Pfister, *Helv. Chim. Acta*, **33**, 568 (1950).

"Enzymatische Peptidsynthese. Isolierung von enzymatisch gebildetem L-Methionyl-L-methionin und L-Methionyl-L-methionyl-L-methionin; Vergleich mit synthetischen Produkten," M. Brenner and R. W. Pfister, *Helv. Chim. Acta*, **34**, 2085 (1951).

"Synthesis of Polymeric Peptides in Proteinase-catalyzed Transamidation Reactions," J. S. Fruton, W. R. Hearn, V. M. Ingram, D. S. Wiggans, and M. Winitz, *J. Biol. Chem.*, **204**, 891 (1953).

"Enzymic Synthesis of Peptide Bonds. VI. The Influence of Residue Type on Papain-catalyzed Reactions of Some Benzoylamino Acids with Some Amino Acid Anilides," S. W. Fox, M. Winitz, and C. W. Pettinga, *J. Am. Chem. Soc.*, **75**, 5539 (1953).

"Trypsin-catalyzed Transpeptidations," S. G. Waley and J. Watson, *Biochem. J.*, **57**, 529 (1954).

"Eine neue Umlagerungsreaktion und ein neues Prinzip zum Aufbau von Peptidketten," M. Brenner, J. P. Zimmermann, J. Wehrmüller, P. Quitt, and Iphigenia Photaki, *Experientia*, **11**, 399 (1955).

"Specificity of Cathepsin C," N. Izumiya and J. S. Fruton, *J. Biol. Chem.*, **218**, 59 (1956).

"The Course of Proteolysis of Lysozyme," T. L. Hurst and S. W. Fox, *Arch. Biochem. and Biophys.*, **63**, 352 (1956).

Additional Proteins with Fundamental Biological Functions

- Immunological considerations
- Specificity of immunological reactions
- Mechanism of antibody formation
- Toxins (toxoids)
- Nature of antibodies
- Nucleoproteins
- Nature of viruses

IMMUNOLOGICAL CONSIDERATIONS

The science of immunology contains a wealth of information on the specificity of protein structure and corresponding reactions. The relationship in this area between reactions *in vitro* and reactions in organisms represents in many cases a smaller leap than the corresponding extrapolation from knowledge of enzyme reactions to living systems. Each line of investigation has in its own way made it possible to learn much about the specific manifestations of proteins.

Immunochemistry is a bridge built from the shores of immunology and protein chemistry. Interest in this subject has flourished for many decades, principally because the immunological reactions have been of outstanding therapeutic importance, both in an artificial way and in terms of the natural reactions of the organism. The immunological mechanisms represent a main line of defense of the mammalian body against pathogenic invaders, proteins or microorganisms. This property of protection, however, does not always correlate fully with

the kinds of serological reaction that can be observed from studies in glassware.

Glossary. The terminology of immunology is highly specialized. Accordingly, a glossary is included here.

ANTIGEN. Antigens are substances which incite the formation of antibodies, and are also capable of reaction with antibodies. They are thus capable of performing two functions: (1) they are responsible for the stimulation of antibody production, and (2) in serological testing they are an essential component of test systems, in which they react with the corresponding antibody. Most known antigens are proteins, some are polysaccharides, and probably all are macromolecules. Antigens are sometimes referred to as foreign proteins, which signifies that they are proteins not normally found in the body they invade. The active principles of the pollens responsible for hay fever are antigenic foreign proteinaceous substances. This particular type is known also as the *allergen*.

ANTIBODIES. Antibodies are specific proteins that form in the organism in response to the introduction of an antigen or foreign microorganism. Many antibodies are capable of reacting with their complementary antigens in one or more demonstrable ways. Some antibodies for which such behavior has not been demonstrable *in vitro* are nevertheless capable of conferring protection upon the animal which produces them, or in another animal to which they are transferred.

IMMUNITY. Immunity is an increased resistance to infection, often demonstrably associated with the presence of antibodies. When the immunity is developed within the invaded animal itself, it is known as *active immunity*. When the immunity is conferred by transfer of antibodies from another animal, it is known as *passive immunity*. Active immunity is characteristically more permanent than passive immunity.

ANTISERUM. Antiserum is the blood serum which contains the antibodies.

HAPten. Hapten is a word much used in the field of immunochemistry. Studies of synthetic antigens were particularly pioneered by Landsteiner, who was also responsible for our initial knowledge of blood groups. The hapten is the determinant group in an antigen. The term hapten originally referred to chemical groups conjugated to proteins *in vitro*. The reactivity of antigens, either artificial or natural, has been found to reside predominantly in the hapten, a term sometimes applied to prosthetic groups. The hapten, then, is in a sense comparable to the active center of an enzyme.

PRECIPITIN. When a test antigen and the corresponding or *homologous* antibody react to form a precipitate, the resultant reaction is known as a precipitin reaction.

LYSIS. Lysis refers to the dissolution of cells. It requires not only the particulate antigen, such as foreign red blood cells, and the homologous antibodies, but also a complex third component, known as *complement*. This material, present in normal sera, is not increased in the process of immunization. It is inactivated by heating for an hour at 60°.

Classification of Immunological Reactions. Harington (1940) has classified immunological reactions as shown in Table 23-1. The first

TABLE 23-1. Classification of Immunological Reactions

State of Antigen	Reaction Observed	Example of Antigen
Dissolved	Precipitation	Foreign protein
Dissolved	Neutralization	Bacterial toxin
Particulate	Agglutination	Pathogen, sperm, etc.
Particulate	Lysis	Foreign erythrocytes

two kinds of reaction listed in the table emphasize the distinction, previously made, between an observed precipitation reaction and neutralization of a toxic protein such as may be liberated by a pathogenic bacterium. Largely because of its ready experimental manipulation, the precipitin type of reaction has especially come into the purview of immunochemistry and of protein chemistry. In agglutination, the precipitate consists of cells which are clumped or *agglutinated* by the antibody. Such behavior is not restricted to pathogenic bacteria or viruses; it has also been observed for sperm of various species in which extracts of the eggs of the same species perform the function of agglutination.

SPECIFICITY OF IMMUNOLOGICAL REACTIONS

Immunological specificity may be expressed at various levels. It may be studied in terms of the species involved (reactions characteristically occurring within the species but less strongly or not at all between species), in terms of the individual proteins, and most intimately, in terms of the specific portion of the protein involved in the antigen, the hapten.

Species Specificity. Experimental studies constructed to evaluate species specificity have consisted of two main steps. In the first step, the blood protein of the species, such as that of the ox, is used to elicit antisera in the rabbit. In the second step, the resultant antisera are tested against blood proteins of a variety of species. The results

of many such studies indicate that the strength or extent of reaction conforms roughly to the phylogenetic proximity of the particular species to that of, in this example, the ox. One such study has given the results indicated in Table 23-2.

TABLE 23-2. Reactions to Ox Protein Antisera¹

(Maximum reaction = 100, no reaction = 0)

Blood Protein Source	Rabbit Antisera for Ox Protein		
	No. 1	No. 2	No. 3
Ox	100	100	100
Sheep	66	50	40
Goat	50	50	40
Pig	16	8	0
Horse	16	8	1
Dog	16	8	0
Man	8	8	0
Wild rat	1	1	

¹ Reprinted by permission of the publishers from Karl Landsteiner, M.D., *The Specificity of Serological Reactions*, Revised Edition, Cambridge, Mass.: Harvard University Press, Copyright, 1945, by The President and Fellows of Harvard College.

It should be emphasized that if rabbits are used for the production of antisera, they may show considerable serological dissimilarity between the blood proteins of rats and mice. Rabbits, however, are members of the rodent family along with rats and mice, and this may explain the fact that such marked differences are not found when rabbit sera are employed to differentiate two similar species of birds, such as chicken and goose. This is what Landsteiner has referred to as "a case of faulty perspective." This type of situation illustrates one of many factors that must be controlled in serological studies in order to permit meaningful comparisons.

Protein Specificity. The kinds of specificity and partial specificities illustrated may be explainable in terms of specific protein structures. When proteins of different origin, as from different tissues, are employed in studies similar to the one above, specificity more deserving of the name is encountered. It is reasonable to suppose, for instance, that serum albumins of various species differ in relatively minor ways, and that the kind of partial "specificities" observed are to be explained on such a structural basis (Table 23-3). If one compares antigenically a nucleoprotein of a given species with a chromoprotein of another species, complete specificity will be the rule. The individual nature of blood albumin and of blood globulin is nicely demonstrated by serological means.

Hapten Specificity. Much clarification of the nature of protein specificity became possible as a result of the development of synthetic immunochemistry. This field was pioneered by Landsteiner, and further developed by workers such as Heidelberger, Harington, Haurowitz, Pauling, and Pressman. Natural proteins have been converted by alterations such as those represented by nitration, iodination, and acylation. An especially useful reaction for the production of "synthetic" antigens has been the coupling of diazotized haptens to the protein that is modified. This reaction is known to occur principally at the carbon *ortho* to the hydroxyl in tyrosine residues.

In many of the chemically altered proteins which are powerfully antigenic, the hapten has been shown to be the primary determinant of a new specificity that differs, in the main, from that observed for the unaltered protein. The original antigenic specificity of the unmodified protein is either lost or greatly blanketed. This type of study parallels investigation of natural antigens which indicate that there, too, the haptens or prosthetic groups are dominant in determination of antigenicity. Such groups in natural proteins include glucose residues, the relatively voluminous tyrosine side chain, and nucleic acid portions of nucleoproteins. Examples of the specificity observable in closely related haptens are given in Table 23-3.

TABLE 23-3. Hapten Specificity¹

Immune Seras to Hapten	Antigen Source							
	<i>p</i> -Aminobenzoic acid	<i>m</i> -Aminobenzoic acid	<i>o</i> -Aminobenzoic acid	<i>p</i> -Aminophenyl- arsenic acid	<i>p</i> -Aminophenyl arsenic acid	Sulfanilic acid	<i>o</i> -Aminocinnamic acid	Aniline
<i>p</i> -Aminobenzoic acid	++	+	+	0	0	0	0	+
<i>o</i> -Aminobenzoic acid	0	0	+	+	0	0	0	+
<i>p</i> -Aminophenyl arsenic acid	0	0	0	+++	0	0	0	0
Aniline	0	0	0	0	0	0	+	+
<i>p</i> -Nitroaniline	0	0	0	0	0	0	+	+
<i>p</i> -Toluidine	0	0	0	0	0	0	++	++

¹ Reprinted by permission of the publishers from Karl Landsteiner, M.D., *The Specificity of Serological Reactions*, Revised Edition. Cambridge, Mass.: Harvard University Press, Copyright, 1945, by The President and Fellows of Harvard College.

² ± = weak reaction.

When the hapten is one of four closely related dipeptides, the results are similar, and it can be seen that serological specificity results from minor structural differences in the hapten, as shown in Table 23-4. The terminal amino acid residue is the primary determinant of the antigenicity, and the specificity is determined to a lesser degree by

the second amino acid residue. This is quite comparable to similar effects of protease substrates.

Harington showed that by employing such haptens as thyroxine or aspirin he could develop in rabbits antisera which would not only give precipitin reactions with the corresponding antigenic proteins

TABLE 23-4. Hapten Specificity of Peptides¹

Immune Sera	Antigen Source			
	Glycyl-glycine	Glycyl-leucine	Leucyl-glycine	Leucyl-leucine
Glycylglycine	++± ²	0	0	0
Glycylleucine	+	+++	0	+
Leucylglycine	+	0	+++	0
Leucylleucine	0	+	0	++

¹ Reprinted by permission of the publishers from Karl Landsteiner, M.D., *The Specificity of Serological Reactions*, Revised Edition, Cambridge, Mass.: Harvard University Press, Copyright, 1945, by The President and Fellows of Harvard College.

² ± = weak reaction.

but would also nullify in part the physiological effects of the smaller molecules. The antipyretic (fever-lowering) effect of aspirin, which could be nicely followed normally, was in large part eliminated by injection into experimental animals of antisera developed to "aspirylprotein."

It is clear from much of this discussion that in many cases it is necessary to evaluate results which exhibit not complete specificity, but intermediate or varying degrees of specificity. In these situations the term specificity itself no longer rigorously applies. In immuno-chemistry, as in immunology, other kindred qualifications must be borne in mind. It is found, for instance, that individuals of the same species may show widely different immunological responses. There is good reason to believe that the antibody-forming capacity of an individual is largely expressible in terms of his genetic constitution. The same individual may also show widely different antibody responses on different occasions.

MECHANISM OF ANTIBODY FORMATION

Boyd, one of the leading immunologists, stated in 1947 that "at the present time the exact mechanism by which antibodies are formed is anybody's guess." Seven years later Boyd reviewed the status of the problem and essentially brought the same old suggestions for antibody formation more up to date. A completely factual selection from the available hypotheses did not appear to be possible from the

evidence at hand at either date. While a number of hypotheses have come and gone, three of the suggested mechanisms that have received favorable attention are presented in this section.

The biological site of antibody formation is also vague. In the past, the various hypotheses with regard to the anatomical origin could be classified into two main types, cellular and humoral. The first of these, which seems to be the most widely accepted today, indicates that antibodies are formed principally in cells; in mammals this is believed to be particularly the cells of the reticulo-endothelial system. Although the cellular type of antibody production appears to be the kind which leads to the heaviest precipitin reactions, one cannot rule out the possibility that weaker analogous responses may be generated in another way.

One of the most widely accepted ideas of the mechanism of antibody formation is that expressed in different ways by Haurowitz (1949), Mudd, and Alexander. In this concept new globulin molecules are synthesized under the directing influence of the antigenic molecules in cells. This concept is primarily *synthetic* in nature. It carries with it the implication that the antibody molecule formed is in some way complementary to the antigen molecule.

The second type of antibody formation, expressed by Pauling, relies not upon a synthesis of new molecules, but rather upon rearranged configurations of already formed globulin molecules. This theory pictures rearrangement of the globulin molecule to conform in a complementary fashion to surface regions of the antigen molecule. Although at first glance this idea appears to be incompatible with the preceding hypothesis, the configuration of the globulin molecule is in part a manifestation of the amino acid arrangement, and if one allows for a dynamic interchange of residues in protein during several days the two concepts of antibody formation are not mutually exclusive.

Either of these first two mechanisms relies upon the concept of *intermolecular complementarity*. In this picture, the antibody molecule adapts itself to the antigenic molecule either by guided synthesis or by rearrangement of preformed α -globulin molecules.

The concept of complementarity is closely related to that which is invoked to explain gene duplication, that is, the templet hypothesis (p. 437). This idea is one for which the essence must be credited to Paul Ehrlich. Ehrlich's ideas have been developed particularly by Pauling and by Haurowitz. One of the problems in this field is the definition of shadowy complementary structures into the precise three-dimensional arrangements of the amino acid residues most directly involved.

A third mechanism of antibody production is that proposed by Burnet. In this mechanism, proteases responsible for the synthesis of globulin molecules are modified by virtue of their action in destroying antigens introduced. As a result of this defensive behavior, the "trained" proteinases are now capable of synthesizing new molecules, the antibody molecules. This mechanism is not as simple as the two more popularly regarded modes. Though it is acknowledged that most scientists prefer simple explanations, it must also be admitted that nature often prefers devices of such complexity as to astound the investigator when he first uncovers them.

One of the many perplexing problems in immunology is the persistence of an active antibody response for years after the antigen has been introduced. Burnet's mechanism of antibody formation has seemed to be most in accord with this fact. Haurowitz and co-workers (Crampton *et al.*, 1952) have studied the persistence of C¹⁴-anthranilazoövalbumin injected into rabbits. Radioactivity in the organism was followed for months. Haurowitz concluded that such antigens are stored indefinitely in the mitochondrial proteins.

The natural antigens are virtually all macromolecules of the size of proteins or polysaccharides. Gelatin and insulin are exceptional in that they do not function directly as antigens. Gelatin is low in the content of amino acids known to contribute to the antigenicity, namely, the aromatic amino acids. Insulin, however, has a considerable content of tyrosine. Each of these proteins has been shown to be capable of functioning indirectly in antigen systems. Harington coupled hapten groups to them and obtained antibodies to the new proteins. He found that these antibodies would react also, although weakly, with the unmodified proteins.

Antigenic polysaccharides have been isolated from a number of bacterial types. These have been studied particularly in the case of the pneumococci. Such polysaccharides precipitate with antisera in dilutions up to several million.

TOXINS (TOXOIDS)

As has already been pointed out, the bacterial toxins are antigenic proteins. The toxins of diphtheria, tetanus, and botulism, have particularly been studied, and purified. The properties of two of these toxins are given in Table 23-5. It has been observed that each of the three toxins mentioned is active at a concentration equivalent to one molecule per cell, and it has been calculated for the botulinus toxin that only 20 million molecules are necessary to kill a mouse. The importance of the integrity of the protein structure of two of

these toxins has been demonstrated by the fact that the tetanus and diphtheria molecules lose their activity rapidly when they undergo proteolysis by enzymes. Botulinus toxin, however, is quite resistant to the action of pepsin and trypsin.

TABLE 23-5. Properties of Toxins¹

	Diphtheria Toxins	Botulinus Toxin, Type A
Nitrogen, per cent	16.0	14.1
Sulfur	0.75	+
Phosphorus	0.05	0.045
Amino nitrogen	0.98	-
Cystine	-	0.5
Methionine	-	0.8
Tyrosine	9.5	9.5
Tryptophan	1.4	+
Arginine	3.8	3.2
Histidine	2.4	1.0
Lysine	5.3	7.9
Isoelectric point	4.1	5.6
Sedimentation constant	4.6S	17.3S
Diffusion constant, $\times 10^7$	6.0	2.14
Frictional coefficient, f/f_0	1.22	1.76
Molecular weight	72,000	900,000

¹ From A. M. Pappenheimer, Jr., *Federation Proc.*, **6**, 479 (1947).

Another class of antigens related to toxins in behavior, and of particular interest, are the proteinaceous substances which are responsible for allergic reactions. For victims of food allergy these must be very numerous, and they are relatively uncharacterized. The allergens of such common irritants as pollens of ragweed and timothy grass have been isolated and characterized as polypeptides.

The toxins are the causative agents in such infectious diseases as measles and diphtheria. Passive immunity can be imparted to a human by transferring to the human the antibody fraction of the blood of another animal (horse or rabbit) sufficiently exposed to the disease to have developed antibodies. The medical advantage in employing such protein preparations is avoidance of the discomfort and danger of the direct infection. The disadvantages are principally two: (1) the foreign nature of the blood protein of the intermediate animal may cause severe *anaphylactic* reaction, or shock; (2) the duration of passive immunity is short as compared to that of active immunity.

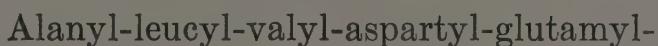
The most effective single solution to these problems has been the *toxoid*. The toxoid is toxin carefully altered by suitable gentle treat-

ment, as with formaldehyde. The fortunate consequence of this treatment is that the toxicity to humans is destroyed, yet the antigenic power of the original toxic protein molecule persists. The effect of formaldehyde on proteins is partly understood (p. 62), but the exact nature of the subtle change in the toxin → toxoid conversion is not understood, and this stands as one of the more important problems in this field.

NATURE OF ANTIBODIES

Antibody protein has been shown to be derived in almost all cases from the γ -globulin fraction of blood. A notable exception is the monkey antipoliomyelitis fraction, which is β -globulin. Northrop has succeeded in purifying diphtheria antitoxin by precipitating it from antisera with diphtheria toxin and separating the two.

Much of the available structural information on antibodies has been accumulated by Porter. Porter (1950) studied the N-terminal sequence of antiovalbumin raised in rabbits, and also the sequence in an inactive γ -globulin fraction from the same source. In each case the N-terminal pentapeptide residue proved to be



To the extent that these studies reveal the structure of the large antibody molecule, they seem to be most consistent with Pauling's theory of the formation of antibodies. The vast majority of the sequences are unaccounted for, as Porter has pointed out, and the biologically active portion of the molecule may be elsewhere in the chain and different from the sequence in the corresponding portion of the inactive molecule. Theoretically any of 19⁵ pentapeptides might occupy the terminal position, if protein synthesis is haphazard. Considerations of the evolution of protein molecules suggest that synthesis is far from haphazard. Porter's findings may well be merely an assignment of terminal structure to rabbit γ -globulin, or to rabbit antiovalbumin, however this is related in its origin to the γ -globulin.

Emil Smith and co-workers (1955) carried out a study which points more convincingly to the same conclusion supporting the Pauling theory. They analyzed the antibody protein stimulated in rabbits by specific polysaccharides of the types I, VII, VIII, and XIV pneumococci. No significant differences in the amino acid contents of the four antibody proteins were observed. Small differences in composition or sequence may have escaped notice, particularly inasmuch as only very small differences could be anticipated from an evolutionary

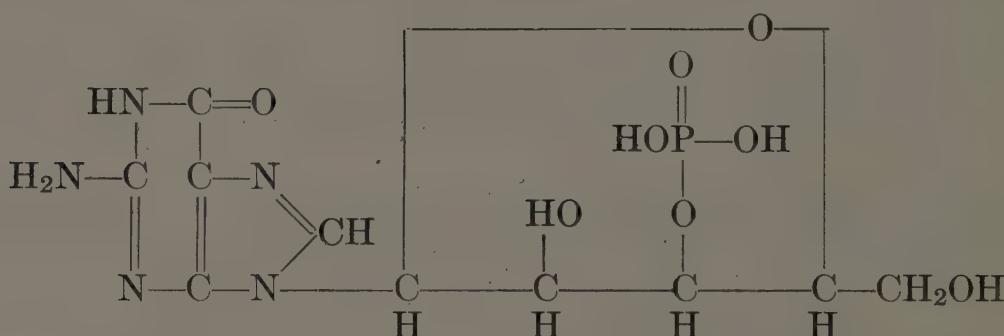
point of view. In the absence of demonstration of such small differences as yet, the point of view that they do not exist, and that changes in conformation are instead critical, must be favored.

NUCLEOPROTEINS

Of the various types of conjugated protein, the nucleoproteins are of outstanding function and interest. Their name derives from the fact that they were discovered in the nuclei of cells. They are found also in other particulate components of the cell. All viruses that have been studied have been found to be nucleoproteins. The viruses are treated in a separate section of this chapter, inasmuch as these nucleoproteins are the most thoroughly understood from the viewpoint of protein chemistry. Nucleoprotein is a significant part of chromosomes, and it has long been inferred that the genes are either nucleoprotein or the nucleic acid prosthetic group. The cytological location of the units is largely responsible for such beliefs. These opinions are based on the conjecture that the nucleoproteins operate as subtle directive influences in the synthesis of proteins centrally related to reproduction. Other hypotheses state that they function as cytological anchors, stabilizing the proteins of hereditary function by virtue of their high insolubility, or as energetic anchors by inhibiting hydrolytic catabolism of their contained proteins.

The nucleic acids are complex prosthetic groups which hydrolyze to purines, pyrimidines, pentose sugars, and phosphoric acid. The contained nitrogen bases are varied and are of kinds found elsewhere, but the pentoses are unique. They are either D-ribose or 2-D-deoxyribose. The nucleic acids extracted from cells are accordingly known as *ribonucleic acid*, RNA, and *deoxyribonucleic acid*, DNA.

Typical of isolated nucleic acid components is the nucleotide, guanylic acid:



Nucleic acids are long-chain macromolecules, composed of similar units.

Only nucleic acid preparations have the power of causing specific heritable transformations in bacteria. For this and other reasons it

is believed by most that ultimate biological control of reproduction mechanisms resides in the nucleic acids. Biological control at most other levels is mediated by proteins (enzymes, hormones); it therefore seems likely that nucleic acids function through some type of influence on protein.

NATURE OF VIRUSES

Of the various viruses, the tobacco mosaic viruses have been studied particularly. These plant pathogens can be obtained in quantity with relative ease and are stable for years. Numerous strains of TMV, believed to be closely related genetically, have been analyzed for amino acid content and for the contents of nucleic acid components. Of sixteen so analyzed, all have been shown to contain the same proportions of purine and pyrimidine bases. The amino acid contents differ considerably, as shown for two examples in Table 23-6.

TABLE 23-6. Amino Acid Contents of Two Strains of TMV

(Grams of amino acid residue per 100 gm. of virus)

Amino Acid	TMV	HR
Alanine	4.1	5.1
Arginine	8.8	
Aspartic acid	11.7	10.9
Cysteine	0.6	
Glutamic acid	9.9	13.6
Glycine	1.4	1.0
Histidine	0.0	0.6
Isoleucine	5.7	5.1
Leucine	8.0	
Lysine	1.3	
Methionine	0.0	1.9
Phenylalanine	7.5	4.8
Proline	4.9	
Serine	6.0	4.7
Threonine	8.4	7.0
Tryptophan	1.9	1.3
Tyrosine	3.4	6.1
Valine	7.8	5.3

Knight (1954) has concluded that definite differences in protein composition can accompany mutation of a virus. Although these differences are generally in the proportions of amino acids present, they can also be differences of kind. The HR strain of TMV was found, for example, to possess two amino acids, histidine and methionine, entirely lacking in the other strains. The strains most

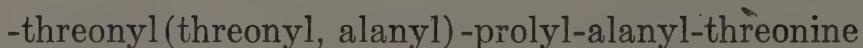
closely related to their derivation were found to be most similar in composition.

A small part of the terminal sequence of amino acid residues in TMV protein has also been analyzed. Fraenkel-Conrat reported no N-terminal residue, whereas threonine has been established as a C-terminus (Niu and Fraenkel-Conrat, 1955). Fraenkel-Conrat assigned a molecular weight of 16,000 to the subunit of TMV protein, and this figure has been confirmed by Ginoza and Atkinson (1956) with measurements of dye binding. This value is comparable to that of other protein molecules when the protein molecule is defined as that mass of protein bearing a single terminal residue of any one type, such as lysozyme and hemoglobin.

A C-terminal residue in a protein molecule lacking an N-terminal residue suggests a 6-shaped molecule, and Niu and Fraenkel-Conrat (1955) have proposed as a structure



The nature of the joint in the loop is not known, but linkage between the ω -carboxyl of an interior glutamyl residue and the otherwise N-terminal group could provide the necessary structure. This C-terminal hexapeptide structure has been found to describe TMV as well as the M and YA strains of TMV. The C-terminal hexapeptide of the HR strain, however, was found to include two stepwise differences in



It is of interest to compare also the stepwise differences in bovine and porcine oxytocin and vasopressin, and insulins from various species.

Fraenkel-Conrat and Williams (1955) have succeeded in separating the protein and the nucleic acid from TMV gently enough so that they were able to recombine the components and observe infective activity in the recombination. This type of experiment has also opened the exciting possibility of hybridizing proteins and nucleic acids from genetically distinct TMV's. The first experiments of this sort led to the conclusion that the nucleic acid component plays the dominant role in determining the character of the hybrid, but the protein was believed to contribute to the characteristics.

SELECTED REFERENCES

Books and Reviews

- "Protein Constitution and Immunological Behavior," M. Heidelberger, *Cold Spring Harbor Symposia Quant. Biol.*, **6**, 369 (1938).
- "Synthetic Immunochemistry," C. R. Harington, *J. Chem. Soc.*, **1940**, 119.
- "Chromosomes and Nucleoproteins," A. E. Mirsky, *Advances in Enzymol.*, **3**, 1 (1943).
- "Nucleoproteins," J. P. Greenstein, *Advances in Protein Chem.*, **1**, 209 (1944).
- Specificity of Serological Reactions*, revised ed., K. Landsteiner, Harvard University Press, Cambridge, 1944.
- "Aspects of Immunochemistry," M. Stacey, *Quart. Revs. London*, **1**, 179, 213 (1947).
- "Immunochemistry," H. P. Treffers, *Ann. Rev. Microbiol.*, **1**, 263 (1947).
- "Bacterial Toxins," A. M. Pappenheimer, Jr., *Federation Proc.*, **6**, 479 (1947).
- Fundamentals of Immunology*, 2nd ed., W. C. Boyd, Interscience, New York, 1947.
- Experimental Immunochemistry*, E. A. Kabat and M. M. Mayer, Charles C Thomas, Springfield, Illinois, 1948.
- "Fertilization and Immunity," A. Tyler, *Physiol. Revs.*, **28**, 180 (1948).
- "Biological Problems and Immunochemistry," F. Haurowitz, *Quart. Rev. Biol.*, **24**, 93 (1949).
- "The Proteins of Immune Reactions," W. C. Boyd in Neurath and Bailey's *The Proteins*, Vol. II B, Academic Press, New York, 1954.
- "Nucleoproteins and Viruses," R. Markham and J. D. Smith in Neurath and Bailey's *The Proteins*, Vol. II A, Academic Press, New York, 1954.
- "The Chemical Constitution of Viruses," C. A. Knight, *Advances in Virus Research*, **2**, 153 (1954).
- "Quantitative Cytochemical Methods," T. Caspersson, *Experientia*, **11**, 45 (1955).

Original Articles

- "Specificity of Agglutinins and Precipitins," K. Landsteiner and J. van der Scheer, *J. Exptl. Med.*, **40**, 91 (1924).
- "Serological Specificity of Peptides," K. Landsteiner and J. van der Scheer, *J. Exptl. Med.*, **55**, 781 (1932).
- "The Serological Differentiation of Steric Isomers," K. Landsteiner and J. van der Scheer, *Proc. Soc. Exptl. Biol. Med.*, **29**, 1261 (1932).
- "Molecular Weight of Antibodies," M. Heidelberger and K. O. Pedersen, *J. Exptl. Med.*, **65**, 393 (1937).
- "Synthetic Immunochemistry. III. Preparation and Antigenic Properties of Thyoxy Derivatives of Proteins and Physiological Effects of Their Antisera," R. F. Clutton, C. R. Harington, and M. E. Yuill, *Biochem. J.*, **32**, 1119 (1938).
- "Precipitation of Iodized Proteins of Different Iodine Contents with Specific Antisera," F. Haurowitz and G. Appel, *Z. Immunitätsforsch.*, **95**, 478 (1939).
- "Synthetic Immunochemistry. V. Observations on Antisera Against Aspirin-protein Complexes," G. C. Butler, C. R. Harington, and M. E. Yuill, *Biochem. J.*, **34**, 838 (1940).
- "Theory of the Structure and Process of Formation of Antibodies," L. Pauling, *J. Am. Chem. Soc.*, **62**, 2643 (1940).

"The Antigenic Structure of Hemolytic Streptococci of Lancefield Group A. VI. The Immunological Behavior of the Nucleoproteins Extracted at Neutrality," S. Mudd and D. B. Lackman, *J. Immunol.*, **39**, 495 (1940).

"Some Properties of Antibodies for Bacterial Exotoxins, Endotoxins, and Polysaccharides," P. Graber, *Bull. soc. chim. biol.*, **23** (1941).

"Purification and Crystallization of Diphtheria Antitoxin," J. H. Northrop, *J. Gen. Physiol.*, **25**, 465 (1942).

"The Specific Groups of Antibodies," F. Haurowitz, M. Vardar, and P. Schwerin, *J. Immunol.*, **43**, 327 (1942).

"Serological Properties of Simple Substances. III. The Composition of Precipitates of Antibodies and Polyhaptenic Simple Substances: The Valence of Antibodies," L. Pauling, D. Pressman, and C. Ikeda, *J. Am. Chem. Soc.*, **64**, 3010 (1942).

"Quantitative Microestimation of Antibodies in the Serums of Man and other Animals," M. Heidelberger and C. F. C. MacPherson, *Science*, **97**, 405 (1943).

"Serological Properties of Simple Substances. VI. The Precipitation of a Mixture of Two Specific Antisera by a Dihaptenic Substance Containing the Two Corresponding Haptenic Groups; Evidence for the Framework Theory of Serological Precipitation," L. Pauling, D. Pressman, and D. H. Campbell, *J. Am. Chem. Soc.*, **66**, 330 (1944).

"Serological Differences between Native and Denatured Proteins," F. Haurowitz and F. Bursa, *Rev. fac. sci. univ. Istanbul*, **10B**, 283 (1945).

"Purification and Crystallization of *Clostridium botulinum* Type A Toxin," C. Lamanna, O. E. McElory, and H. W. Eklund, *Science*, **103**, 613 (1946).

"Chemical Study of Rabbit Anti-ovalbumin," R. R. Porter, *Biochem. J.*, **46**, 473 (1950).

"The Use of Radioactive Isotopes in Immunological Investigation. 5. Quantitative Aspects of the Antibody-Antigen Reaction," T. E. Banks, G. E. Francis, W. Mulligan, and A. Wormall, *Biochem. J.*, **48**, 180 (1951).

"Persistence of C¹⁴-Anthranil-azo-ovalbumin in Injected Rabbits," C. F. Crampton, H. H. Reller, and F. Haurowitz, *Proc. Soc. Exptl. Biol. Med.*, **80**, 448 (1952).

"N-Terminal Groups of Normal Human Gamma Globulin and of Myeloma Proteins," F. W. Putnam, *J. Am. Chem. Soc.*, **75**, 2785 (1953).

"Studies on the Chemistry of the Transforming Activity. I. Resistance to Physical and Chemical Agents," S. Zamenhof, H. E. Alexander, and G. Leidy, *J. Exptl. Med.*, **98**, 373 (1953).

"X-Ray Diffraction by an Abnormal Protein (B8) Associated with Tobacco Mosaic Virus," R. E. Franklin and B. Commoner, *Nature*, **175**, 1076 (1955).

"Amino Acid Composition of Four Rabbit Antibodies," E. L. Smith, M. L. McFadden, A. Stockell, and V. Buettner-Janusch, *J. Biol. Chem.*, **214**, 197 (1955).

"C-Terminal Amino Acid Sequences of Four Strains of Tobacco Mosaic Virus," C.-I. Niu and H. Fraenkel-Conrat, *Arch. Biochem. and Biophys.*, **59**, 538 (1955).

"Reconstitution of Active Tobacco Mosaic Virus from Its Inactive Protein and Nucleic Acid Components," H. Fraenkel-Conrat and R. C. Williams, *Proc. Natl. Acad. Sci. U. S.*, **41**, 690 (1955).

"The Distribution of Polio Antibodies in Serum Proteins," S. N. Timasheff, A. W. Moyer, R. A. Brown, and J. G. Kirkwood, *Proc. Natl. Acad. Sci. U. S.*, **42**, 228 (1956).

"The Binding of Safranine O by Tobacco Mosaic Virus," W. Ginoza and D. E. Atkinson, *J. Am. Chem. Soc.*, **78**, 2401 (1956).

Origin, Evolution, and Biosynthesis of Protein

Primordial synthesis of amino acids

Primordial synthesis of protein

Evolution of protein molecules

Biosynthesis of protein

The origin of protein can be discussed from either of at least two points of view. The way in which protein arose on this terrestrial globe can be considered. One can also experiment with and interpret the way in which protein is formed in the organism from the most elemental substances ingested. A great amount of the latter type of information for synthesis of the amino acid units has already been accumulated (Chap. 7). There is reason to believe that these two points of view, far from being independent, are basically similar. Biological and evolutionary thinking suggests that the manner in which organisms synthesize protein and other substances must in some measure reflect the way in which these substances came into existence originally.

It is of course not easy to arrive at a secure conclusion on how the biochemical world and life began. It is possible, however, to perform experiments which test the postulates in this field and to develop a theory which, as Rubey (1955) has said, "limits the range of permissible speculation." It is possible also to assemble from structural studies precise information on the course of earlier evolution of protein molecules. A number of useful inferences can be drawn from these analyses. For such reasons as these, this chapter will treat

the biosynthesis of protein after discussing the concepts of primordial synthesis.

PRIMORDIAL SYNTHESIS OF AMINO ACIDS

A number of prebiological syntheses of amino acids have been visualized under conditions which have yielded amino acids in the laboratory. Amino acids are relatively stable compounds, and it is hardly surprising that many sets of conditions should cause various mixtures of compounds rich in carbon, hydrogen, oxygen, and nitrogen to be converted to these substances. Of the various possibilities, those that are in accord with geological knowledge of prebiological eras and those which explain current biochemical syntheses of amino acids and of other substances, are perhaps the most plausible.

The nature of the primitive atmosphere from which the amino acids may have been derived is, as might be expected, a somewhat controversial matter. Oparin and Urey (1952) suggested an atmosphere of methane, hydrogen, water, and ammonia. Rubey (1955) and Revelle (1955) tend toward the belief that the available data, particularly on the gaseous effluvia of volcanoes, support instead a prebiological atmosphere composed of carbon dioxide, nitrogen, water, and hydrogen sulfide. Other atmospheres have been proposed; most or all of them differ from the current atmosphere in lacking oxygen. The similarities in components of the suggested atmospheres are greater than the differences.

Miller (1955) discharged an electric current through the gaseous mixture proposed by Urey, and obtained therefrom glycine, alanine, aspartic acid, sarcosine, α -amino-*n*-butyric acid, and α -aminoisobutyric acid, as well as other substances. He showed that amino acids could result from hydrolysis of intermediate aminonitriles formed (as per a type of Strecker reaction, Chap. 5). The ideas of Pflüger of many decades earlier were based on the production of nitrogen compounds from prebiological cyanogen. More recently Varner and Burrell (1955) showed that cyanogen could on hydrolysis give a number of compounds, including urea.

A type of reaction which could explain the origin of photosynthesis is the irradiation of solutions of paraformaldehyde and potassium nitrate by exposure to sunlight. Bahadur (1954) obtained evidence for serine, aspartic acid, and asparagine.

Yet another mode of synthesizing amino acids under presumed primordial conditions was uncovered by one of the authors during attempts to polymerize amino acids to protein thermally. Heating of ammonium hydrogen malate yielded aspartic acid, α -alanine,

β -alanine, and polymers of such units. The individual amino acids were recovered upon hydrolysis. When the ammonium ion was replaced by urea, the nucleic acid intermediate ureidosuccinic acid also resulted. Considerable significance was attached to the fact that the first intermediate was malate, inasmuch as current biochemical processes proceed through this Krebs cycle intermediate (Chap. 7).

PRIMORDIAL SYNTHESIS OF PROTEIN

Suggestions for the primordial formation of protein are not as profuse as those for the formation of the amino acids. A number of writers have emphasized the importance of understanding the formation of protein and of enzyme as a fundamental problem in the origin of life. Two suggestions for such formation are (1) the heating of amino acids over the boiling point of water at the prevailing pressure in the primitive atmosphere, and (2) the effecting of such reaction at elevated temperature in the presence of phosphoric acid. Theoretically, either set of conditions could overcome the thermodynamic infeasibility of the synthesis of the peptide bond (Chap. 22).

Solution to the related problem of specific sequence of residues in the first protein molecule can be visualized. In theory, the individual reactant molecules would determine their own arrangement. This is suggested by systematic experiments in the enzymic synthesis of anilides (Chap. 22). In such experiments, only certain substituted amino acids participate in net synthesis, whereas other combinations involve degradation of one of the reactants.

The related problem of the origin of optical activity in nature can be solved in any of a number of suggested ways. One of these is the proposal that the first synthesis of a replicating complex of molecules included but one protein molecule, which necessarily could not be racemic. It is not to be expected that each amino acid residue would be of the same L configuration, but this difficulty can be met by a possibility suggested by Bernal—the presence of an optically active surface at which synthesis proceeds. Inasmuch as quartz is found in two crystal forms that are mirror images of each other, this type of surface has often been invoked.

The problem of optical activity may belong to the era after life began. An observation consistent with this possibility is the occurrence of D-amino acids as well as L-amino acids, in the proteins of some microbes representing a low level in phylogeny. The development of a monoconfigurational life from one composed of D- and L-types could result from spontaneous resolution of a primordial nutrient. Spontaneous resolution was first demonstrated by Pasteur with

ammonium hydrogen DL-malate, and subsequently confirmed by others. If, for example, D-aspartic acid were to crystallize spontaneously from a nutrient solution containing DL-aspartic acid, L-aspartic acid could remain in solution. Any conversion products of aspartic acid, such as alanine and protein, would then also tend to consist of L-amino acids. Another possible evolutionary origin of nature's emphasis on L-amino acids was suggested by Langenbeck. In a large number of molecules, it is to be expected that there would be a minute, polarimetrically inappreciable fraction of L- or D- in excess of the other type. It is to be expected also that there would result an evolutionary pressure for a single form. Langenbeck states that an optically inactive world is unstable. Presumably an enzyme composed of units of a single configuration will act more efficiently on a substrate molecule also composed of units of a single configuration.

A thermal origin of protein was offered because it can simultaneously explain the origin of protein and of other kinds of biochemical substance, and because taxonomic studies on thermophilic blue-green algae suggested to Copeland (1936) that these algae are the most primitive and that life originated in hot water. Perhaps as other energetic modes of the origin of protein are brought forward, comparable related attributes will appear for the new schemes.

The problem of the original life can be approached either experimentally, from conditions presumed to hold before life began, or by back-extrapolation of the process of evolution. The latter provides us with an analyzable body of evidence.

EVOLUTION OF PROTEIN MOLECULES

The evidence available suggests to most workers that the biochemical substances, such as vitamins B, and processes of today are similar to those of the most primitive organisms. This evidence is based especially on analyses of what are believed to be lineal descendants of the most primitive organisms. St. Gyorgyi and Krebs, for example, have stated that the biochemical similarities throughout phylogeny lead to the conclusion that all life is one.

Purely by analogy, one should expect that the amino acids would be very much alike between current organisms and between current and primordial organisms. This conclusion has been subjected to scrutiny. Attempts have been made to find in primitive organisms proteins lacking common amino acids. Reports of such absences have appeared but have not been confirmed. The only tenable interpretation at present is that for all organisms represented in evolution, virtually all of the common amino acids are present.

Statistical comparison of the amino acid contents of twenty-four proteins from diverse sources has shown that these proteins are more alike than they are independent. This result is consistent with the interpretation that proteins have evolved from a common primordial type or types.

Analyses of the unfractionated proteins of a number of plants indicate also a dominant common type of protein containing N-terminal lysine. This protein type can be considered to be primitive.

The possibility that variations in protein structure are a fundamental phenomenon in biology was clearly seen by Macallum, who said in 1926, "In the living complex of every cell the basic elements, proteins, composed as they are each of a variable number of amino acids and in variable proportions of these, cannot be predicated as having a uniform composition even absolutely in similar cells in the same organism, for thus mutation would never obtain in a species." Recognition of the principle that evolution must proceed at the molecular level is implicit with many scientists, particularly biologists, but the value of this principle resides in its application rather than in its passive recognition. What are some of the consequences of such application?

Microheterogeneity. When one adopts the point of view that was expressed by Macallum, microheterogeneity becomes understandable as a manifestation of Darwinian biological evolution at an underlying chemical level.

The first question that must be asked is, "What is the evidence for microheterogeneity?" In working with simpler molecules than those of protein, chemists have found it to be pragmatically beneficial to assume that a compound under study was capable of purification to a single type. In the majority of cases this assumption has proved to be fruitful (pertinent exceptions are such families of molecules as tocopherols, which could, however, be readily separated into individuals).

Two explanations for microheterogeneity have received discussion: (1) the fractionation itself produces molecular species which were not present originally; (2) the diversity is a natural one. Colvin, Smith, and Cook (1954) have carefully reviewed numerous cases and have decided in favor of the natural explanation. It had been pointed out even earlier that vitamins (tocopherol, vitamin A, vitamin B₆), polypeptide antibiotics (gramicidins, tyrocidines, polymixins), and others also occur in families of molecules, and that protein families, or microheterogeneity, should be considered to be a special case of the more general phenomenon. Furthermore, the polypeptide antibiotics,

which most closely resemble the proteins, consist of families of individuals differing in amino acid composition. These differences cannot be attributed to the gentle methods used for their fractionation.

The fact of microheterogeneity can be explained as merely an accumulation of synthetic mistakes. If these "mistakes" serve an evolutionary purpose, they constitute a chemical basis for our existence and become of paramount importance. Application of the Darwin-Macallum concept can be explained by Fig. 24-1.

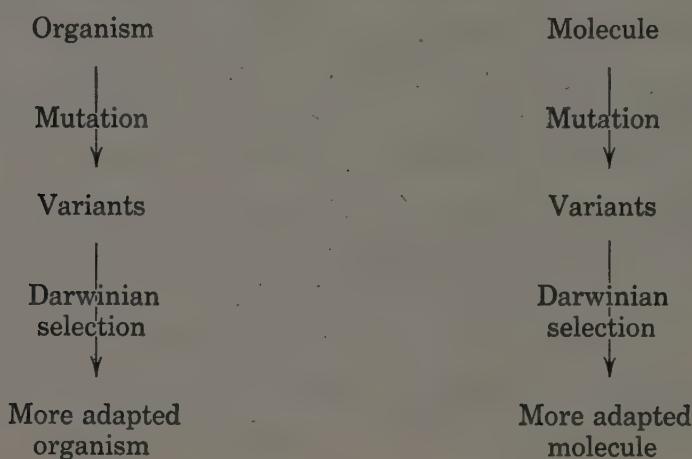


Fig. 24-1. Darwinian evolution at the molecular and biological levels.

Variation and selection are seen to occur for both organisms and molecules. These processes are interwoven (see Fox, 1953, 1955). The first step at each level is the step of variation to yield mutants from which selection can occur. In this frame of reference, microheterogeneity of protein and families of polypeptide molecules are inevitable consequences of the first, or variation, step in Darwinian evolution.

Continuous Changes in Structure. Another consequence of the evolutionary interpretation is that it made possible some prediction about regularity in protein structure. The evolutionary concept suggested gradual small changes in structure. In such a spectrum, maintenance of a periodic structure (Chap. 16) in all proteins appeared to be impossible. Information which is being accumulated does indicate small stepwise changes in protein molecules within a species and between species. Examples of these are oxytocin and vasopressin, insulin, and TMV protein.

Number of Types of Protein. It is common in textbooks and classes of biochemistry to emphasize the astronomical number of protein isomers which are theoretically possible. Evaluation of the diversity in a more realistic way emphasizes that the number of types

that have appeared during evolution must be an infinitesimal fraction of the total number possible. When this fact is faced, it is possible to attack related problems, such as the terrestrial origin of protein, with renewed assurance.

A secondary consequence of this recognition is the greater probability of understanding how nucleic acid, with its theoretically somewhat limited diversity of structures, can determine the arrangements of residues in proteins. Nucleic acid diversity alone perhaps does not permit determination of the theoretical total of protein structures, but regulation of the actual number is far more within the realm of possibility.

Experiments with papain-controlled anilide syntheses suggest that certain combinations of amino acid residues are more favored, at least in the presence of the enzyme. If this phenomenon can be extrapolated to the biosynthesis of protein, it may be possible to designate the chemical and physical properties which restrict or favor some of the possible reactions. The suggestion that no particular two-residue sequences are favored had been made from scrutiny of protein structures, but this conclusion has been withdrawn. A conclusive solution to this last problem, based on analytical evaluation alone, should now or soon be possible.

The strikingly limited diversity of protein, as indicated by some data, does not pose a special problem of explaining the tremendous diversity in organisms. Organisms can be thought of as representing a kind of molecular ecology, in which they express not so much the component molecules as the interactions of the component molecules. The interactions of a limited number of protein types with a limited number of protein types or of nucleic acid types, or with other macromolecular types, exponentializes the diversities possible.

Classification of Protein. Classification of proteins, particularly of simple proteins (those hydrolyzing to amino acids only), has suffered a long history with little affection. Part of the trouble has been that too many proteins did not fit well into the categories which were constructed. For example, difficulty has been encountered in attempting to decide whether the euglobulins and pseudoglobulins of the blood should be classified as albumins or globulins. The thought has long been entertained by many that the use of solubility criteria to distinguish types of protein is more arbitrary than useful. Recognition of an evolutionary status for protein suggests, however, that certain structural characteristics may have sufficient applicability to permit a more definite type of classification. An investigation of this possibility should lead to a genealogical chart for proteins. When the

factors governing the way in which one protein is related to others become clear, this possibility of classifying protein by structural type (N-lysyl, N-isoleucyvalyl, and so forth) will be closer to evaluation and to reality. In any event, realization of the evolutionary origin of proteins and evaluation of the scope of gradual structural variation will help to explain the difficulties in the way of applying the historical classification of protein.

BIOSYNTHESIS OF PROTEIN

In the province of biosynthesis of protein the signs have perhaps been even less clear and plain to read than those along the path of evolution of protein. Knowledge of the mode of biosynthesis of the amino acid units has, however, been brought to a remarkably high state of development (Chap. 7). In the entire area of inquiry into the origin and biosynthesis of amino acids and the origin, evolution, and biosynthesis of protein, knowledge of primordial and biological synthesis of protein is least fully represented. Perhaps either will be satisfactorily understood only when the other is in hand.

The problems of protein synthesis are among the most central and challenging in the field of biochemistry. Although many experiments have been performed in order to solve these problems and much has been written by way of interpretation, agreement has not been reached.

A greater degree of agreement has been attained in the definition of the component problems. The questions of energetics and specificity of protein synthesis have received the most consideration.

It has been pointed out that the formation of a dipeptide from two amino acids requires approximately 3000 cal. per mole. It seems probable that the energy demand per peptide linkage decreases with increasing size of the molecule. Hence it has been suggested that the energy requirement in the coupling of two dipeptides to yield a tetrapeptide is less than 3000 cal. Nevertheless, the over-all free-energy requirement in the synthesis of protein from free amino acids must be enormous. By fundamental thermodynamic principles this energy demand is entirely independent of the mechanism of synthesis. The requirement must be met by some form of coupling, direct or indirect, with an energy-yielding process. At the physiological level it seems probable that this energy source is carbohydrate metabolism. No other biological reservoir of energy appears to be sufficient to provide the necessary quantities.

It has been proposed that phosphorylamino acids are the high-energy intermediate derivatives which participate in protein synthesis. Such compounds have been synthesized by purely chemical means, but

they have not been shown to exist in nature. Physiological experiments in which synthesis of protein is concomitant with breakdown of adenosinetriphosphate (ATP) to adenosinemonophosphate (AMP) indirectly render the consideration of such intermediates more promising than they would be otherwise. Another possibility which has been considered is the intermediacy of γ -glutamyl amino acids.

The mode of energy input for synthesis of protein is not lacking in proposals, only in evidence. The same can be said for the other main problem of protein synthesis, the regulation of order of residues to yield specific products. The concept most often suggested as a solution to this problem is that of the *templet* (template) *hypothesis*. The term has been taken from the field of mechanics to signify the guiding of an order in pattern by a form. In this application, nucleic acid, particularly RNA, is visualized as the form on which proteins are assembled in specific order from the amino acids.

Although such a scheme poses some severe chemical difficulties in matching unassembled residues to an already formed surface, this picture meets in a simple fashion several requirements of the biological problem of an inheritance mechanism. The occurrence of nucleic acid and protein in chromosomes is accounted for on the basis of a nucleoprotein structure which embraces a specific arrangement of residues. In an elaboration of this concept, Hinshelwood has suggested that nucleic acids order proteins, which subsequently order new nucleic acids in the form of the original. This process is known as *reciprocity*; a mechanism with the same end results is necessary to explain the biological phenomenon of reproduction. An occasional variation at some point in the form would constitute a mutation. Evidence which can be interpreted as in accord with this mechanism is at hand. The most ardent proponents of this mechanism do not claim it to be proved or fully developed in all its aspects. The organization of protein molecules into three-dimensionally specific arrangements cannot be explained by this unaided superficial mechanism, for example.

Another possibility, inferred from studies of synthesis of anilides, is that proteolytic enzymes can order the arrangement of residues by virtue of the fact that the substrates contribute to the action pattern of the enzyme. The concept that proteolytic enzymes can direct the synthesis of protein in a reversal of hydrolysis is an idea which was championed by Bergmann. This concept lost favor as it was realized that proteases are not sharply specific, and with the belief that each step in protein synthesis requires a specific protease for its catalysis. Of all classes of enzymes, only the proteases act upon substrates which resemble themselves structurally. The model experiments suggest that

the amino acid residues themselves can dominate in a self-regulatory way the selective synthesis of the peptide chain. This picture is equivalent to the action of a new protease at each step in protein synthesis, and the need for a specific enzyme at each step is thus obviated. Furthermore, it may be inferred that the self-regulatory contribution of the intermediate peptides, coupled with the directive action of associated nucleic acids, is sufficient to impart specificity.

One criticism of the concept of templet action of nucleic acids is that the possible structural diversity of nucleic acids is not sufficiently great to provide the necessary diversity in proteins. The interaction of amino acid self-regulation and nucleic acid templet action should provide a much larger number of possibilities than either alone.

An additional property of nucleic acid is its ability to form nearly insoluble precipitates with protein. Inasmuch as the heredity-bearing materials of the cell seem to be nucleoprotein, the stability of heredity patterns can be better understood as stemming from materials which are not in solution and are therefore less subject to dynamic change. The development of a slow evolution from materials of suitable physical properties would seem to be a minimum utility of nucleic acid. Evidence from studies of transforming principle alone suggest that nucleic acid plays, as well, an important role in determining specificities. Inasmuch as in the "daily business" of the cell specific proteins likewise organize relatively specific actions, it may well be that proteins and nucleic acids in concert determine the ultimate manifestations of the genes.

The question of whether the proteolytic enzymes participate in protein synthesis has not been closed either positively or negatively. No reason has been brought forward to exclude their participation along with a templet or in conjunction with nucleic acid, although such an exclusion has at times been assumed.

It is possible that a great deal of the slowness in finally solving the problems of protein biosynthesis is due to failure to visualize their complexity as a whole and a tendency to focus attention on too narrow a segment of knowledge. To quote Kacser (1956), ". . . genetic mechanisms should not be looked for in the properties of particular substances but in the way the whole system is organized."

The rate of development of new knowledge is accelerating. Evidently what is needed first is recognition by a substantial number of experts of the potential value of cultivating an area of inquiry, even though the majority consider the major discernible problems in that area to be unsolvable. Experiences with science in the past two decades have taught us that many of the most optimistic predictions

will be fulfilled under such circumstances. In the area of the present discussion it is easy to overemphasize the ignorance and the problems. On the positive side, it is at least possible to discern the dim outlines of how our complex protein-centered life of today could have arisen from very simple beginnings, and how the multitudes on multitudes of living cells may be in their own ways re-enacting a smaller or larger part of the biochemical history which they have inherited.

SELECTED REFERENCES

Books and Reviews

- "The Paleochemistry of the Body Fluids and Tissues," A. B. Macallum, *Physiol. Revs.*, **6**, 316 (1926).
- "On the Nature and Origin of Proteins," R. J. Block, *Yale J. Biol. and Med.*, **7**, 235 (1934-35).
- The Planets*, H. C. Urey, Yale University Press, New Haven, 1952.
- "Biosynthesis of Proteins," P. N. Campbell and T. S. Work, *Nature*, **171**, 997 (1953).
- "Peptide and Protein Synthesis. Protein Turnover," H. Tarver in Neurath and Bailey's *The Proteins*, Vol. IIB, p. 1199, Academic Press, New York, 1954.
- "The Microheterogeneity of Proteins," J. Ross Colvin, D. B. Smith, and W. H. Cook, *Chem. Revs.*, **54**, 687 (1954).
- "Potentialities of Protein Isomerism," I. Asimov, *J. Chem. Educ.*, **31**, 125 (1954).
- "Development of the Hydrosphere and Atmosphere, with Special Reference to Probable Composition of the Early Atmosphere," W. W. Rubey, *Geol. Soc. Amer., Spec. Paper*, **62**, 631 (1955).
- "On the History of the Oceans," R. Revelle, *J. Marine Research Sears Foundation*, **14**, 446 (1955).
- "Perspectives in Evolution," H. F. Blum, *Am. Scientist*, **43**, 595 (1955).
- "The Biosynthesis of Peptide Bonds," J. S. Fruton in Graff's *Essays in Biochemistry*, p. 106, John Wiley & Sons, New York, 1956.
- "The Biosynthesis of Peptides and Proteins," H. Borsook, *J. Cellular Comp. Physiol.*, **47**, 35 (suppl. May 1956).
- "Molecular Organization of Genetic Material," H. Kacser, *Science*, **124**, 151 (1956).
- "Protein Structure in Relation to Function and Biosynthesis," C. B. Anfinsen and R. R. Redfield, *Advances in Protein Chem.*, **11**, 1 (1956).
- "Progress in Developing Chemical Concepts of Genetic Phenomena," N. H. Horowitz, *Federation Proc.*, **15**, 818 (1956).
- "Virus Composition and Structure," W. M. Stanley, *Federation Proc.*, **15**, 812 (1956).
- "Evolution of Protein Molecules and Thermal Synthesis of Biochemical Substances," S. W. Fox, *Am. Scientist*, **44**, 347 (1956).
- "Spontaneous Generation of Anabolic Pathways, Protein, and Nucleic Acid," S. W. Fox, A. Vegotsky, K. Harada, and P. D. Hoagland, *Ann. N. Y. Acad. Sci.*, **69**, art. 2, 328-337 (1957).

Original Articles

- "Yellowstone Thermal Myxophyceae," J. J. Copeland, *Ann. N. Y. Acad. Sci.*, **36**, 1 (1936).
- "Evidence for Intermediates in Protein Synthesis," D. Steinberg and C. B. Anfinsen, *J. Biol. Chem.*, **199**, 25 (1952).
- "A Correlation of Observations Suggesting a Familial Mode of Molecular Evolution as a Concomitant of Biological Evolution," S. W. Fox, *Am. Naturalist*, **87**, 253 (1953).
- "Genetical Implications of the Structure of Deoxyribonucleic Acid," J. D. Watson and F. H. C. Crick, *Nature*, **171**, 964 (1953).
- "Transformation of Type Specificity of Meningococci. Change in Heritable Type Induced by Type-Specific Extracts Containing Desoxyribonucleic Acid," H. E. Alexander and W. Redman, *J. Exptl. Med.*, **97**, 797 (1953).
- "Polypeptide Formation by Reaction of N-Carboxyamino Acid Anhydrides in Buffered Aqueous Solutions," R. R. Becker and M. Stahmann, *J. Biol. Chem.*, **204**, 737 (1953).
- "The Synthesis of Aldolase and Glyceraldehyde-3-Phosphate Dehydrogenase in the Rabbit," M. V. Simpson and S. F. Velick, *J. Biol. Chem.*, **208**, 61 (1954).
- "Photosynthesis of Amino Acids from Paraformaldehyde and Potassium Nitrate," K. Bahadur, *Nature*, **173**, 1141 (1954).
- "Relation of Energy Processes to the Incorporation of Amino Acids into Proteins of the Ehrlich Ascites Carcinoma," M. Rabinovitz, M. E. Olson, and D. M. Greenberg, *J. Biol. Chem.*, **213**, 1 (1955).
- "Prebiological Chemistry of Nitrogen Compounds of Metabolic Importance," J. E. Varner and R. C. Burrell, *Euclides Madrid*, **15**, 1 (1955).
- "Amino Acid Studies in the Transformation of Proteins of the Hen's Egg to Tissue Proteins During Incubation," C. O. Rupe and C. J. Farmer, *J. Biol. Chem.*, **213**, 899 (1955).
- "Production of Some Organic Compounds under Possible Primitive Earth Conditions," S. L. Miller, *J. Am. Chem. Soc.*, **77**, 2351 (1955).
- "A Statistical Evaluation of the Kinship of Protein Molecules," S. W. Fox and P. G. Homeyer, *Am. Naturalist*, **89**, 163 (1955).
- "On Biochemical Origins and Optical Activity," S. W. Fox, J. E. Johnson, and A. Vegotsky, *Science*, **124**, 923 (1956).
- "Participation of Free Amino Acids in Protein Synthesis," R. B. Loftfield and A. Harris, *J. Biol. Chem.*, **219**, 151 (1956).
- "Vitamin B₆ and Protein Synthesis," H. C. Lichstein, *J. Biol. Chem.*, **219**, 27 (1956).

Index

The authors are indebted to Mr. James van Luik for helpful advice in the preparation of this index.

- Abderhalden, E., 50, 106, 132, 145, 146, 149, 152, 399
Abel, J., 369
Abramson, H. A., 194
Absorption spectrum, shift associated with acid-binding in carbonyl-hemoglobin, 189
Accessory amino acids, 125
Accuracy, 99
Acetamide, denaturation of proteins by, 313
Acetylaminomalonic ester, in synthesis, 69
Acetylation, 51
Acetyl-coenzyme A, 110
Acid-binding capacity, of proteins, 180, 181
Acidic amino acids, 14
Acidic hydrolysis, 75
ACTH, *see* Adrenocorticotropic hormone
Actin, 360
Activators, 389
Activity, thermodynamic, 208
Activity coefficient, 244-245
 of hydrogen ion, 177
Actomyosin, 360-362
Adenosinetriphosphate, 112
 action on actomyosin, 361-362
 in protein synthesis, 437
Adrenalin, *see* Epinephrin
Adrenocorticotropic hormone, 375
Adsorption, in protein purification, 275, 357
 in relation to zone electrophoresis, 205
Agar, use of in zone electrophoresis, 203
Agglutination, 416
Aggregation, associated with denaturation, 311
 use of light-scattering in detecting, 212
Alanine, isolation, 79
 metabolism, 120
 structure, 8
 substituted, 69
 β -Alanine, 111
Alberty, R., 386
Albumins, 234, 240, 270-271
 acid-binding capacity of, 181
 amino acid composition of, 336
 combination with detergent ions, 252, 337
 denaturation of, 310, 314, 316, 319-320, 322
 distribution in various plasma fractions, 332, 334-335
 ion-binding by, 337
 molecular weight, 337
 plasma, 202, 224, 241, 249-250, 275, 298, 309-311, 314, 328-329, 343
 properties of, 333-338
 sedimentation and diffusion, 218

- Albumins, terminal residues of, 285
 titratable groups in, 182
 titration curves of, 178
- Alcohol, effect of on solubility of proteins, 243, 246-247
 effect of on titration behavior of amino acids, 37
 solubility of amino acids in, 38
 solubility of prolamins in, 363
- Aldehyde condensations, in synthesis, 68
- Aldolase, 362
 electrophoretic behavior of, 253
- Alexander, J., 420
- Alkali solubility of prolamins, 363
- Alkaline hydrolysis, 75
- Alkaloids, from amino acids, 115
 in resolution, 73
- Allergen, 415, 422
- Allergenic peptides, 140
- All-or-none character of denaturation, 316
- All-or-none combination of protein with detergent ions, 251
- Amino acid composition of proteins, 278
- Amino acid ester hydrochlorides, 12
- Amino acid esters, distillation, 56
- Amino Acid Manufacturers, 81
- Amino acid residue, 6
 reactions in protein, 59
 terminal, 147
- Amino acids, acetylation, 51
 acylation, 50
 alkylation, 49
 amphoteric properties, 20-37
 anabolism, 108
 arylation, 50
 assay, 85-87
 assay with dairy bacilli, 96
 assay without hydrolysis of protein, 87
 biochemical unity, 433
 biological significance of individual, 9, 11
 browning reaction, 59
 catabolism, 109
 chromatography, 89
 classification, 14
 colorimetric assay, 94
 commercial production, 81
 comprehensive assay, 87
 contents in proteins, 103
 conversion to carbohydrate, 110
 conversion to fat, 110
 conversion to vitamins, 110
 criteria of acceptability, 7
- Amino acids, cyclodehydration, 58
 deamination, 110
 decarboxylation, 57
 decomposition during hydrolysis, 86
 definition, 5
 derivatives for characterization, 7
 detoxicant, 9
 dissociation constants, 26
 distribution in proteins, 14
 dynamic aspects, 107
 economical methods for obtaining, 82
 electrostriction in solutions of, 29
 enthalpy of ionization, 30
 essential, 9, 11, 13, 106, 124
 esterification, 56
 formol reaction, 52
 glycogenic, 9, 11
 history, 17
 indispensable, 13, 124
 individual assay, 87
 individual metabolism, 117
 isoelectric point and solubility, 38
 isolation, 65, 77
 by groups, 77
 melting point, 30
 metabolic significance, 105
 microbiological assay, 95
 nutritional availability from protein, 123
 nutritive significance, 105
 optical activity of, 41
 optical isomers, 45
 oxidation, to aldehydes, 54
 to keto acids, 54
 to nitriles, 54
 preparation, 65
 preparative methods, 81
 primordial synthesis, 430
 purity needed for nutritional studies, 66
 reaction, with complex salts, 56
 with isocyanates, 51
 with metal ions, 56
 with ninhydrin, 52
 with nitrous acid, 51
 with organic acids, 55
 with pyridine and acetic anhydride, 58
 reactions of, 49
 relationship of side chain to reactivity, 12-13
 reliability of assays, 96
 requirement by rats, 126
 requirements, 125
 resolution of, 45, 72

- Amino acids, routine assay, 87
selective precipitation, 56
separation by zone electrophoresis, 204
solubilities of, 39
solubility in alcohol, 38
Stein-Moore assay, 90
stoichiometry of nutritional requirements, 124
structural features, 6
synthesis of, 65
 through α -halogen acid, 66
table of characteristics, 8-11
terminology, 6
variations with side chains, 7
- D-Amino acid oxidase, 44, 390
- D-Amino acids, nutritional availability, 126-127
occurrence in antibiotics, 44
- Amino acyl halides, 57
- α -Aminoadipic acid, 16, 121
 structure of, 15
- Amino alcohol oxidation, in synthesis of amino acids, 70
- α -Aminobutyric acid, 16
 structure of, 15
- Amino group, in proteins, 182
 reactions of, 59
- Aminopeptidase, 401
- Ammonium groups, titration of in proteins, 176, 180
- Ammonium reineckate, 56, 80
- Ammonium rhodanilate, 56
- Ammonium sulfate, in separation of globulins, 328
- Amphoteric properties, effect of ionic strength, 185-187, 249
of amino acids, 20-37
of hemoglobin, 348
of water, 21-22
- Anabolism, 108
- Anderson, G. W., 167, 170
- Anfinsen, C., 409
- Angiotonin, 137
- Anilides, for resolution, 74
 synthesis by proteases, 409
- Animal protein factor, 124
- Animals, synthetic abilities, 109
- Anion binding, 192
 by proteins, 187, 247-253
- Anomalies, in electrophoresis, 200, 328-329
- δ -Anomaly, 200, 203
- ϵ -Anomaly, 200, 203
- Anserine, 134
 structure, 136
- Anthranilic acid, 119, 123
- Antiamino acids, 392
- Antibiotic peptides, 137
- Antibiotics, as D-amino acid derivatives, 45-46, 115
- Antibodies, 227, 344, 415
 persistence of, 421
 proteins as, 3
- Antibody-antigen combination, 255, 307-308
- Antibody formation, mechanism, 420-421
- Antibody protein, amino acid composition, 423
- Anticoagulants, for blood, 327
- Antienzymes, 389
- Antigens, 415
 molecular size, 421
- Antiovalbumin, N-terminal pentapeptide, 423
- Antipodal specificity, 73, 401, 402
- Antiserum, 415
- Antitoxin, diphtheria, 319-320
- Antitrypsins, 391
- Aoki, K., 337
- Arginase, 112
- Arginine, 16, 118
 effect of pH on optical rotation, 42
 isolation, 79
 metabolism, 120
 structure, 8
 synthesis, 72
- Arginine peptides, synthesis, 167
- Argininosuccinic acid, 112, 118
- Arginyl peptides, synthesis, 167
- Armstrong, S. H., 329
- Aromatic biosynthesis, 117
- Arrhenius, S., 20, 23
- Arrhenius theory, 387
- Asparagine, 80
- Aspartic acid, 118, 179, 278, 280
 effect of pH on optical rotation, 42
 isolation, 80
 metabolism, 120
 solubility of, 38-39
 structure of, 8
 synthesis of, 70
 titration curve of, 33
- Aspartylaspartic acid, titration curve of, 33
- Aspergillic acid, 139
- Aspirin, as hapten, 419
- Aspirylprotein, 419
- Assay organisms, requirements, 97
- Assays, comparisons, 100
 distinction from analysis, 86

- Assays, literature survey, 86
 statistical evaluation, 99
- Astbury, W. T., 229, 231, 299
- Asymmetry, changes associated with denaturation, 309-311
 of plasma albumin, 337
 of protein molecules, in relation to gelation, 256-257
- Atkinson, D., 426
- ATP, *see* Adenosinetriphosphate
- Avidin, 285, 354
- Axial ratio, 219, 221, 222
 relation to frictional ratio, 219
see also Asymmetry
- Azobenzene-*p*-sulfonic acid, in isolation of alanine, 79
- Bacillin, 139
- Bacillus subtilis*, action on ovalbumin, 355
- Bacitracin, 138
- Bacitracin A, structure, 139
- Backbone spacing, in β -keratin, 288
- Bahadur, K., 430
- Bahn, A., 132, 145
- Baldwin, E., 385
- Bamford, C. H., 294
- Barbieri, J., 17
- Barger, G., 50, 144, 145
- Barium salts, precipitability, 79
- Barton-Wright, E. C., 101
- Base-binding capacity, of proteins, 180, 181
- Basic groups, reactivity in proteins, 60
- Battersby, A. R., 154
- Beadle, G. W., 95, 107
- Bear, R. S., 230, 231
- Beckmann rearrangement, in synthesis, 72
- Beer's law, 94
- Bell, P. H., 376
- Beloff, A., 409
- Ben-Ishai, D., 169
- Bentonite, in preparation of lysozyme, 357
- Benzenoid amino acids, 14
- Benzenoid residues, reactivity, 61
- Benzoylaminomalonic ester, in synthesis, 69
- Benzylpenicillin, 138
- Berger, A., 169
- Bergmann, M., 132, 146, 163, 167, 399, 402, 404, 406, 409, 410, 437
- Bergmann's periodicity hypothesis, 283, 289
- Bergmann-Zervas carbobenzoxy synthesis, 164
- Bergmann-Zervas-Schneider degradation, 151
- Bergmann-Zervas synthesis, 132
- Bernal, J. D., 431
- Betaines, 49
- Biochemical mutants, 95
- Biochemical unity, amino acids, 432
- Biological activity, groups essential for, 62
 loss of in denaturation, 307-308
- Biological criteria of protein purity, 264
- Birefringence of flow, *see* Streaming birefringence
- Biuret reaction, 264, 280
- Block, R. J., 79, 86, 101
- Blood, composition of, 327-328
- Blood plasma, electrophoresis of, 204, 262, 328-330
- Blout, E., 294
- Bohr effect, 348
- Boiling point elevation, 209
- Bolling, D., 86, 101
- Bopp, F., 17
- Borsook, H., 121
- Bott, M., 294
- Boundaries, observation of, 196-199
- Boyd, W., 419
- Braconnot, H., 17
- Bradbury, J., 294
- Bragg, W. L., 277, 293, 300, 347
- Bragg law, 229
- Brice, B. A., 211
- British Antilewisite, 61
- Brockmann, H., 146
- p*-Bromophenylmercapturic acid, 116
- Bronsted, J. N., 32
- Bronsted treatment of acids and bases, 20-25
- Brown, H., 371, 375
- Browning reaction, 59
- Buchner, E. H., 381
- Budka, M., 329
- Buffering power, 25
- Burnet, F. M., 421
- Burrell, R. C., 430
- Buston, H. W., 17
- Calcium ion, in casein, 358
 in coagulation of blood, 327
- Calcium salts, precipitability, 79
- Canaline, structure, 15
- Canavanine, structure, 15
- Cannan, R. K., 179, 182

- Caprylate, stabilization of plasma albumin by, 337
Carbamyl phosphate, 118
Carbobenzoxy synthesis, 164
 decarbobenzoxylation step, 169
N-Carbobenzoxyamino acid anhydrides, 170
 preparation of, 171
Carbohydrates, from amino acids, 110
 in mucin, 356
 in ovomucoid, 357
 in plasma albumin, 333
Carbon monoxide, 349
Carbonic anhydrase, 327
Carboxyl amides in protein, 60
Carboxyl groups, 179, 182
 pK of in proteins, 176
 reactivity of in proteins, 60
Carboxypeptidase, 147, 401, 405
 in resolution, 74
 specificity of, 403
Carnosine, 134
 structural assignment, 145
 structure, 136
Carpenter, F. H., 167
Casein, 224, 279, 358-359
 acid-binding capacity of, 181
 electrophoretic behavior, 358-359
 terminal residues of, 285
Catalase, 327
Cathepsins, 405-406
Cations, interaction with proteins, 248, 297
Catron, D., 329, 360
Cereal protein, 60, 363
Charge, effect on titration behavior of proteins, 185-186
Charge distribution in proteins, 183-184
Chemical analysis in assessing protein purity, 261-262
Chibnall's subunit hypothesis, 283-284
Chloroacetyl amino acids, 74
p-Chloromercuribenzoate, 61
Chlorophyll, 279
Cholesterol, 345
Choline, 121
Christensen, L. K., 319
Chromatography, 89, 204
 of proteins, 275
Chromoproteins, 279
Chymotrypsin, 405
 structural changes in production from chymotrypsinogen, 395
 terminal residues of, 285
Chymotrypsinogen, structural changes in conversion to chymotrypsin, 395
Citrate, as anticoagulant, 327
Citrulline, 16, 118
 structure, 15
 synthesis, 72
Clinical uses of plasma proteins, 341-343
Clotting, of blood, 327
Clotting mechanism, 338-341
Clupein, 2
 structural proposal in 1937, 145
Coagulation of proteins, 304, 358
Coenzymes, 380, 389
Cofactors, in amino acid nutrition, 124
Cohn, E. J., 175, 181, 247, 271, 273, 331, 339, 342-343
Colamine, 122
Cole, S. W., 17
Collagen, 235, 289-290
Colligative properties, 209
Colloidal aggregates, 207
Colorimetry, 94
Colostral transmission of antibodies, 330-331, 359-360
Colostrum, 359-360
Colvin, J., 260, 433
Competitive inhibition, 393
Complement, 416
Complementarity, in immunochemistry, 420
Composition, as related to protein solubility, 240-241
 of proteins, variability, 260-261
Conalbumin, 285, 354-356
Concentration gradient, 195-196, 217
Consden, R., 145, 154, 155
Convection, destruction of boundaries by, 196
Cook, W. H., 433
Copeland, J. J., 432
Corey, R. B., 283, 291, 293, 301
Corn, protein composition of, 364
Corn seed protein, terminal residues in, 285
Corticotropin, 285
Coulomb's law, 235
Coulombic forces in protein interactions, 255
Coulombic interactions, within proteins, 282, 299, 318
Countercurrent distribution, 133
Courtaulds Laboratories, 294
Craig, L. C., 133, 138, 154, 155, 374
Cramer, E., 17

- Crampton, C. F., 421
 Creatine, 120, 121
 Crick, F., 290
 Cross, B., 323
 Crystallinity, as a criterion of protein homogeneity, 263
 Crystallizability, loss of, in denaturation, 308-309
 Curtius, T., 70, 162
 Curtius azide procedure, 165
 Cyanoacetic ester synthesis, 70
 Cyclic structure, in peptides and proteins, 286
 Cyclol hypothesis, 298-299
 Cystathionine, 108, 121
 Cysteic acid, 120
 Cysteine, amphoteric properties of, 27
 effect of pH on optical rotation, 43
 in detoxication, 116
 metabolism, 120
 structure, 8
 Cystine, 16, 121
 isolation, 80
 solubility of, 38-39
 synthesis of peptides, 166
 Cystine-cysteine, interconversion, 13
 Cystine residues, in proteins, 280-282
 Cytochrome C, 218, 296
 Cytoplasmic protein, extraction of, 265
 Dairy bacilli, in assay, 96
 Dakin, butanol method, 78
 Dakin-West reaction, 58
 De Broglie waves, 225
 Debye-Hückel theory, 37, 185-186, 244-246
 Decanol, in crystallization of plasma albumin, 333
 Decomposition of amino acids on heating, 38
 5-Dehydroquinic acid, 119
 Delluva, A. M., 107
 Denaturation, 227, 261, 301, 345, 355
 associated with proteolysis, 307
 associated with titration of ovalbumin, 188
 by guanidine, 305, 308, 310, 312-313
 by iodide, 313
 by organic solvents, 313
 by urea, 304, 308-309
 changes in optical activity in, 311-312, 316, 319-320, 322, 337
 effect on electrophoretic behavior, 306
 exposure of terminal residues in, 286
 kinetics of, 317, 320
 Denaturation, liberation of sulfhydryl groups in, 305-306, 313-314
 mechanism, 319-323
 order of reaction, 317-318
 solubility changes associated with, 308
 Denatured proteins, 60
 gelation of, 256
 Deoxyribonucleic acid, 424
 Desnuelle, P., 395
 Detergent ions, interaction with proteins, 250-252, 272, 313
 Detoxication, 116
 Deutsch, H. F., 329
 Dextran, as reference material in zone electrophoresis, 204
 Dialysis, 194, 266
 Diaminopimelic acid, 16
 structure, 15
 Diastereoisomer, 73
 Diastereomer, 73
 Dibromotyrosine, structure, 15
 3,4-Dichlorobenzenesulfonic acid, 80
 Dicyclohexylcarbodiimide method, 169
 Dieckmann, M., 329, 360
 Dielectric constant, 235
 control of in protein fractionation, 271
 effect on titration behavior of proteins, 185-188
 of amino acid solutions, 29
 Dielectric dispersion, 221
 Diffusion, 196
 Diffusion coefficient, 216, 309
 definition and measurement of, 217
 Dihydroxyphenylalanine, structure, 15
 Diiodothyronine, 16
 Diiodotyrosine, solubility of, 39
 structure, 8
 Diketopiperazines, 58, 280
 formation from amino acid esters, 30
 in synthesis, 68
 Dimerization of plasma albumin, 333, 337
 2,4-Dinitrochlorobenzene, 50, 149
 2,4-Dinitrofluorobenzene, 50, 148
 Dintzis, H. M., 267
 Diphtheria antitoxin, purification via precipitation, 423
 Dipolar ampholyte, 28
 Dipolar forces, 236-238
 Dipolarion, 27
 Dipole moment, 221
 Dissociation, of hemoglobin, 348
 Dissociation constant, definition, 22-23
 of amino acids, 26

- Dissociation constant, of important acids, 24
Dissociation equilibria, of amino acids, 28
Disulfide linkage, 266, 280, 299, 305-306, 320, 323, 333, 338, 348
oxidation of, 281
Djenkolic acid, 16
structure, 15
DNA, 424
DNFB, 50, 148
DNP-glycine, hydrolyzability, 149
DNP-proline, hydrolyzability, 149
Donnan equilibrium, 210, 295, 338
Dopa, 16
Doty, P., 294
Double-bond character, in peptide linkage, 291
Dreschel, E., 17
Drying of proteins, 268-270
Dunn, M. S., 81
Du Vigneaud, V., 2, 160, 170, 375
Dyes, binding of by proteins, 247-250

Edema, 338
Edestin, 270, 285, 314, 353
acid-binding capacity of, 181
properties of, 363
Edman, P., 146, 152
Edman method, 152
Edsall, J. T., 26, 29, 32, 184, 332, 334-335, 336
Egg albumin, *see* Ovalbumin
Egg white, electrophoretic behavior, 353-354
Ehrlich, F., 17; 420
Ehrlich's reagent, 95
Einstein, A., 210, 220
Elastin, 235
Electrical transport, 78
Electrodialysis, 267
Electrolytes, effect on solubility of amino acids, 38
Electron microscopy, 225-227, 290, 361
Electroosmosis, 204
Electrophoresis, as a criterion of protein homogeneity, 262
ionic gradients in, 200
of blood plasma, 328-330
of γ -globulin, 344
of plasma albumin, 336-337
preparative, 273
Electrophoresis apparatus, 201, 273
Electrophoresis-convection, 274, 344
Electrophoretic behavior, of casein, 358-359
Electrophoretic behavior, of β -lactoglobulin, 359
of egg white, 353-354
of native and denatured proteins, 306
Electrophoretic mobility, effect of ionic strength, 193-194
relation to radius, 193
Electrophoretic separation of proteins, 195, 204
Electrostatic interactions, 185-187, 249
Electrostatic repulsion, in protein expansion, 187
Electrostriction, 238, 282
in amino acid solutions, 29
Electroviscous effect, 215
Elemycin, 139
Elliott, A., 294
Elution, 89
Enantiography, in electrophoretic patterns, 201, 203
Endosperm proteins, 363
Energy of activation, in protein denaturation, 318-319
Enthalpy of ionization, 30
relation to protein titration behavior, 180
Entropy changes associated with denaturation, 316
Entropy of activation, in denaturation, 318
Enzyme activation, 396
Enzyme inhibitors, 391
Enzyme kinetics, 384
Enzyme preference, 400
Enzymes, 227, 261, 379
active centers, 381, 395
catalytic action, 380
effect of pH, 386
effect of temperature, 387
energetic function, 383
extracellular, 379
gross classification, 382
intracellular, 379
limiting velocity, 385
mechanism of action, 383
metal coordination, 396
proteins as, 3
proteolytic, 280
push-pull mechanism, 396
recognition of protein nature, 381
specificity, 383
structure, 394
Enzyme-substrate interaction, 255, 384
Enzymic hydrolysis, 76
heterogeneous, 383
homogeneous, 383

- Enzymic hydrolysis, order of, 384
 Ephedrine, 115
 Epidermin, 289
 Epinephrin, 369
 biosynthesis, 114
 Epstein, J., 182
 Erepzin, 76
 Ergothioneine, 49
 Erickson, J. O., 310
 Essential, qualifications of the term, 125
 Essential amino acids, 106, 124
 Essentiality, degrees of, 125
 Ester linkage, in protein, 282
 Ethanolamine, 122
 Ether extraction of protein, 266
 Ethionine, 392
 Euglobulin, 328
 Evans, E. A., Jr., 372
 Excelsin, 314, 363
 Expansion of protein molecules, 187, 311, 337-338
 Extraction of muscle proteins, 360
 Eyring, H., 387
 Families of peptides, 140
 Fat, from amino acids, 110
 Fatty acids, salts of, interaction with proteins, 249-250
 Feather rachis, 230
 Feeding, proteins in, 2
 Felix, K., 2, 133, 145, 372
 Fermentation *L. casei* factor, 136
 Ferrihemoglobin, 346
 Fetuin, 331
 Fiber spacing, in β -keratin, 288
 Fibrillar proteins, 230, 231
 Fibrin, 256, 289, 340-343
 Fibrinogen, 202, 212, 224, 241, 256, 297, 328-329, 332
 amino acid composition of, 336
 molecular weight and shape, 338-339
 properties of, 338-343
 sedimentation and diffusion results, 218
 Fibrinopeptides, 340
 Fick laws of diffusion, 217
 Fischer, E., 3, 17-18, 57, 65, 132, 161-163, 279, 399
 Fischer ester distillation method, 77
 Flavianic acid, 55, 79, 80, 248
 Flow birefringence, *see* Streaming birefringence
 p-Fluorophenylalanine, 392
 growth stimulation, 393
 3-Fluorotyrosine, 392
 Foaming, denaturation of proteins by, 313
 Fodor, A., 163
 Folic acids, 111, 136
 Foreman dicarboxylic amino acid separation, 79, 88
 Formaldehyde, effect on titration behavior of amino acids, 37
 see also Formol titration
 Formamide, denaturation of proteins by, 313
 Formol reaction, 52
 Formol titration, 37, 180
 Forsythe, R. H., 354
 Foster, J. F., 322, 329, 337, 354, 360
 Fox, S. W., 144, 156, 408, 434
 Fractionation of proteins, 255, 331-332, 354-357
 Fraenkel-Conrat, H., 1, 406, 409, 410, 426
 Free energy, electrostatic, 185-186
 of denaturation, 316
 of solution, 239-240
 standard, of ion binding, 185-186
 Freeze-drying, 268-269
 Freezing point depression, 209
 French, D., 43
 Frensdorff, H. K., 320
 Frictional coefficient, 215, 219, 309
 Frictional resistance, relation to electrophoretic mobility, 192-193
 Friedell, R., 329, 360
 Frontal analysis, 91
 Fruton, J. S., 132, 402, 404, 411
 Fugitt, C. H., 248-249
 Fumarase, kinetics of action, 386
 Gagnon, P. E., 70
 Gaussian distribution, 196, 203, 217
 Geiger, W., 171
 Gel, nature of, 256-257
 Gelatin, 241, 250, 256
 Gibbs, R., 320
 Gibbs phase rule, 253-254, 263
 Giffee, J., 329
 Ginoza, W., 426
 Gish, D. T., 167
 Gliadin, 218, 365
 Globin, 314, 346
 Globulin, horse serum, denaturation of, 310
 Ac-Globulin, 339
 Globulins, 234, 240, 267, 270-271
 egg white, 354, 356
 electrophoretic resolution of, 202
 plasma, 275, 328, 332, 334-335, 344-345

- Globulins, plasma sedimentation and diffusion results, 218
- α -Globulins, 202, 241, 328-329, 332, 334-335, 344-345
amino acid composition of, 336
- β -Globulins, 202, 241, 328-329, 332, 334-335, 344-345
amino acid composition of, 336
- γ -Globulin, 202-203, 224, 241, 250, 262, 275, 322, 328-329, 343-344
amino acid composition of, 336
as antibody fraction, 423
N-terminal pentapeptide, 423
terminal residues in, 285
- Glucagon, assignment of sequential structure, 375
- Glutamic acid, 118, 179, 278, 280
effect of pH on optical rotation, 42-43
in detoxication, 116
isolation, 80
linkage in glutathione, 37
metabolism, 120
solubility of, 38-39
structure, 8
synthesis, 70
- Glutamic acid semialdehyde, 118
- Glutamine, 80
in detoxication, 116
- Glutathione, abnormal glutamyl linkage, 136
functions, 136
occurrence, 136
structure assignment, 154
titration behavior of, 36, 38
- Glutelins, 234, 363-364
- Glycerophosphate dehydrogenase, 362
- Glycine, copper complex, 56
in detoxication, 116
isolation, 80
metabolism, 120
selective precipitation, 56
structure, 8
titration curve of, 33
- Glycine derivatives, aldehyde condensations, 68
- Glycocoll, 13
- Glycolytic enzymes, 327
- Glycoproteins, 268, 279, 345
- Glyoxylic acid reaction, 264
- Goodloe, M. B., 329
- Gordon, A. H., 154-155
- Gorin, M. H., 194
- Gramicidin, 137-138
 D -amino acids in, 138
- Gramicidin A, 138
composition, 141
- Gramicidin B, 138
composition, 141
- Gramicidin S, 138
structure assignment, 153, 154
structure reported in 1947, 145
- Gravimetric assays, 88
- Greenstein, J. P., 33, 74, 310, 403
- Gronwall, A., 242-243
- GSH, 127
- GSSG, 137
- Guanidine, denaturation of proteins with, 305, 308, 310, 312-313, 337
- Guanidinium groups, titration of in proteins, 176, 179, 182
- Gurin, S., 107
- Gutman, A. B., 330
- Hair, X-ray diffraction pattern, 230
- Haldane, J. B. S., 380
- α -Halogen acid synthesis, 66
- Halwer, M., 211, 323
- Hanby, W., 294
- Hapten, 415
specificity, 418
- Hardy, W. B., 252
- Harington, C. R., 169, 418-419
- Harris, M., 248-249
- Haurowitz, F., 418, 420-421
- Hedin, S. G., 17
- Heidelberger, M., 418
- Helical structures, for peptide chain, 293-295, 323
- Helix-coil transition, 294, 322-323
- Hematin, 346
- Heme, 120, 346
- Heme proteins, 345-346
- Hemin, 346
- Hemocyanin, 224, 296
denaturation of, 310
- Hemoglobin, 120, 227, 231, 278-279, 301, 309, 327, 345-350
acid-binding capacity of, 181
amphoteric properties, 348
carboxy, solubility of, 244
denaturation of, 310, 316
dissociation of, 348
molecular weight, 346-347
preparation of, 346
sedimentation and diffusion results, 218
terminal residues in various species, 285
titration behavior of, 188-189
- Hemostasis, 341
- Herriott, R. M., 140, 388
- Hess, G. P., 169

- Heteroauxin, 115, 367
 Heterogeneity of proteins *see* Homogeneity of proteins
 Hexamethyleneglycol, inhibition of clotting by, 341
 Hiller, A., 17
 Hinshelwood, C., 437
 Hippuric acid, in synthesis, 68
 Histamine, 57, 367
 physiological properties, 369
 Histidine, 57
 effect of *pH* on optical rotation, 42
 isolation, 80
 structure, 8
 titration curve of, 33
 Histidylhistidine, titration curve of, 33
 Hofmann, K., 167
 Hofmeister, F., 279
 Hogness, K. R., 329
 Holtzer, A., 294
 Homocysteine, 108, 121
 Homogeneity of proteins, 212, 215, 216, 217, 260, 317-318, 336, 344, 359, 362, 364
 reversible boundary spreading as a criterion of, 205
 Homogenization, in protein preparation, 265
 Homogentisic acid, 122
 Homoserine, 108
 Hopkins, F. G., 17
 Hordein, 218
 Hormonal proteins, 367
 Hormones, 113
 protein derivatives, 367
 proteins as, 3
 Horowitz, N. H., 121
 Hospelhorn, V., 323
 Howells, E. R., 347
 Huggins, M. L., 293
 Humin, 75
 Hurst, T., 156, 408
 Hvidt, A., 323
 Hydantoin, in synthesis, 68
 Hydration, 219, 221, 238, 309-310, 347
 Hydrodynamic volume, 220, 309
 Hydrogen bonds, 236-238
 in formation of blood clot, 341
 in proteins, 282-283, 290-294, 312-313, 317-318, 323, 341
 Hydrogen sulfide, 349
 Hydrolysis, acidic, 75
 alkaline, 75
 enzymic, 76
 estimation of rate or extent, 76
 of proteins, 75, 280
 Hydrolyzate, parenteral, 106
 Hydroxyamino acids, 14
 3-Hydroxyanthranilic acid, 111
 Hydroxyaspartic acid, 392
 Hydroxyglutamic acid, 16
 Hydroxylsine, structure, 8
 Hydroxyproline, 118
 isolation, 80
 selective precipitation, 56
 solubility of, 38
 structure, 8
 Hypertensin, 137
 Hypotonic solution, 265
 Ideal solutions, 208-209
 Imidazole groups, in hemoglobin, 348-349
 in proteins, 182, 341
 Immune lactoglobulin, 359
 Immunity, 344, 415
 active, 415
 passive, 415
 Immunochromatography, and proteins, 414
 Immunological behavior of proteins, 356, 358
 Immunological reactions, classification, 416
 Immunology, glossary, 415
 Indispensable amino acids, 124
 Indole, 119, 122-123
 Indoleacetic acid, 115, 367
 in corn, 115
 Indole groups, 306
 Indole nucleus, reactivity, 62
 Inductive effects, in proteins, 184-185
 Inflammatory peptides, 140
 Infrared absorption, polarized, 294
 Inheritance, 2
 Inhibitors, 391
 competitive, 391
 Insulin, 231, 241, 250, 255, 278, 284-285
 acid-binding capacity of, 181
 denaturation of, 314
 fibrils, 371
 fragmentation in sequence assignment, 372
 isoelectric point, 371
 masking of titratable groups in, 188
 molecular weight, 371
 sedimentation and diffusion results, 218
 terminal residues in, 285
 titratable groups in, 182
 Insulin chain B, assignment of sequence in fragments, 373

- Insulins, comparative amino acid composition, 374
- Intermolecular complementarity, 420
- Intrinsic dissociation constant, definition, 186
- Intrinsic ionization constant, 183
- Inversion, 45, 74
- Iodide, denaturation of proteins by, 313
- Iodoacetamide, 61
- Iodoacetic acid, 61
- Iodogorgoic acid, solubility of, 38
- Ion-binding, by plasma albumin, 337
effect on isoelectric point, 252–253, 337
- Ion exchange, 78
in purification of proteins, 267
- Ionic attraction, 235
- Ionic gradients, in electrophoresis, 200
- Ionic strength, definition, 242
effect, on electrophoretic mobility, 193–194
on protein titration behavior, 185–186
on solubility, 241–245
- Iron, in hemoglobin, 345–346, 349
- Isocyanates, reaction with amino acids, 51
- Isoelectric point, 192, 251, 348
definition, 30
effect of ion binding on, 252–253, 337
in relation to solubility of amino acids, 38
shift in on denaturation, 306
- Isoionic point, calculation of, 31
distinction from isoelectric point, 30
distribution of protein species, 183
effect of ion binding on, 252–253, 337
of amino acids, 26
of proteins, determination of, 179
- Isolation, amino acids, 65, 77
- Isoleucine, 278
structure, 10
- Isotope dilution, 88
- Isotope swamping, 94
- Jacobsen, C., 282
- Jensen, E., 323
- Jensen, H., 372
- Jirgensons, B., 42, 44
- Johnson, M., 389
- Kacser, H., 438
- Kauzmann, W., 320
- Kendall, E. C., 17
- Kendrew, J. C., 300
- Keratin, 234, 235, 241, 266, 282
 α - and β -forms, 229, 288–295, 299, 320
- Ketene, acetylation with, 51
- α -Keto acid synthesis, 70
- Kibrick, A., 179, 182
- Kinetics of protein denaturation, 317–320
- Kirkwood, J., 222, 274
- Kitai, R., 375
- Kjeldahl nitrogen analysis, 265
- Klett Manufacturing Co., 201, 273
- Knight, C. A., 425
- Koenig, V., 329
- Kögl hypothesis, 46
- Kossel, A., 17
- Kossel, silver salt method, 79
- Krebs, H. A., 432
- Krebs cycle, 117
amino acids from, 118
- Kunitz, M., 140, 388
- Kynurenone, 111
- Lactalbumin, 218, 355, 359
- Lactoglobulin, immune, 359
- β -Lactoglobulin, 212, 218, 231, 257, 278, 296, 353, 355, 359
denaturation of, 316, 319
electrophoretic behavior, 359
solubility of, 242–243, 246
terminal residues in, 285–286
- Lambert law, 98
- Lamm, O., 197
- Landsteiner, K., 417–419
- Langenbeck, W., 432
- Langmuir, I., 228
- Lauffer, M. A., 225, 226
- Leuchs, H., 171
- Leuchs anhydride, 170
- Leucine, 278
contamination by methionine, 80
effect of pH on optical rotation, 42
isolation, 80
structure, 10
- D-Leucine, in gramicidin, 138
- Leucylpeptidase, 405
- Levintow, L., 74
- Levy, M., 320
- Lewis, H. B., 121
- Licheniformin, 139
- Liebig, J., 17
- Light-scattering, 210–212, 337
- Linderstrom-Lang, K., 282, 319, 323
- Lineweaver-Burk plot, 385
- Linked functions, 348
- Lipids, removal from proteins, 266

- Lipoproteins, plasma, 334-335, 339, 344-345
 Liquid-junction potential, in pH measurements, 177
 Loeb, J., 175, 247
 London dispersion forces, 239, 255
 Longsworth, L. G., 198-199
 Lundgren, H., 251, 319
 Lutz, O., 42, 44
 Lycomarasmin, structure, 140
 Lyophilization, 268-269, 271
 Lysine, absence in zein, 364
 isolation, 80
 metabolism, 121
 side-chain reactivity, 59
 structure, 10
 synthesis, 72
 Lysis, 416
 Lysozyme, 212, 281, 296, 354, 356-357
 acid-binding capacity of, 181
 sedimentation and diffusion results, 218
 terminal residues in, 285
 titratable groups in, 182
 Macallum, A. B., 433
 Malonic ester syntheses, 69
 Mandelkerh, L., 219
 Martin, A. J. P., 154-155
 McCoy, R. H., 17
 Mechanism of denaturation, 319-323
 Medicine, proteins in, 3
 Melanin, 113
 Melanoidins, 59, 113
 Melting point, of amino acids, 30
 Mendel, L. B., 106
 Menten, M. L., 384
 Mercaptalbumin, 333
 Mercury, reaction with plasma albumin, 333
 Merrifield, R. B., 124, 140
 Metabolic requirements, 105
 Metal coordination, in enzymes, 396
 Metal ions, binding by conalbumin, 356
 combination with plasma, proteins, 333, 345
 use of in protein fractionations, 272-273
 Methemoglobin, 346
 Methionine, 108
 isolation, 80
 metabolism, 121
 structure, 10
 synthesis, 68
 Methoxinine, 392
 5-Methyltryptophan, 392
 Meyer, C. E., 17
 Michaelis, L., 384
 Michaelis-Menten concepts, 384, 386, 393
 Microbiological assay, 95, 100
 Microheterogeneity, of proteins, 260, 433-434
 Microorganisms, anabolic pathways, 109
 Milk proteins, nutritional quality, 357-358
 Miller, S. L., 430
 Millon's reagent, 95, 264
 Mobility, electrophoretic, 191, 193
 Molecular shape, 208, 215
 Molecular weight, and shape of proteins, 218, 295-298
 changes associated with denaturation, 309
 distinction between weight and number averages, 212
 from sedimentation equilibrium, 214-215
 of plasma albumin, 337
 of proteins, measurement of, 208, 210
 Molecular weight classes, 296
 Monoaminomonocarboxylic acids, 14
 Monobromotyrosine, structure, 15
 Monoiodohistidine, 16
 Monoiodotyrosine, 16
 Mörner, K. A. H., 17
 Morrison, K., 329
 Moyer, L. S., 194
 Mucin, egg white, 354, 356
 Mudd, S., 420
 Mueller, J. H., 17
 Mukherjee, D. H., 17
 Muscle proteins, extraction of, 360
 Mutants, 95
 Myogen, 362
 denaturation of, 310
 Myoglobin, 218, 285, 301, 362
 Myosin, 224, 289, 314, 360
 Native proteins, 60
 Neurath, H., 132, 310, 404
Neurospora crassa mutants, 95
 Nicotine, 115
 Nicotinic acid, 111
 Nielsen, L., 274
 Ninhydrin, 52
 reaction, 52, 53
 use of, in analysis, 53
 Niu, C. I., 426
 Noncompetitive inhibition, 394
 Norleucine, 16

- Northrop, J. H., 2, 140, 382, 387-388, 394, 423
Nucleic acid, 272, 424
Nucleoproteins, 279, 424
 biological function, 424
Number average molecular weight, 212
Nutrition, proteins in, 3
Nutritional properties of zein, 364
Nutritional quality of milk proteins, 357-358
Nutritional requirements, 105

Observation of boundaries, methods for, 196-199
Ogston, A. G., 403
Olcott, H. S., 86
Oncley, J. L., 213, 219, 221
Oparin, A. I., 430
Optical activity, 39
 changes in associated with denaturation, 311-312, 316, 319-320, 322, 337
 of amino acids, 39-46
 origin in life, 431
Optical isomers, physiological manifestation, 45
Optical methods for observation of boundaries, 196-199
Optical rotation, of proteins and peptides, 294
Order of reaction, in protein denaturation, 317-318
Organic solvents, denaturation of proteins by, 313
 in protein fractionation, 271, 357
 use of in drying proteins, 268
Ornithine, 16, 112, 118
 structure, 15
 synthesis, 72
L-Ornithine, in tyrocidine, 138
Ornithuric acid, 117
Osborne, T. B., 106
Osmotic pressure, 208-210, 309, 337
 of blood plasma, 338
Ovalbumin, 212, 241, 248, 251, 275, 296, 309
 acid-binding capacity of, 181
 combination with detergent ions, 251-252
 denaturation of, 310-311, 314, 320
 masking of phenolic groups in, 188
 properties of, 353-355
 sedimentation and diffusion results, 218
 terminal residues, 285
Ovalbumin, titratable groups in, 182
 titration curves of, 179
Ovomucin, 354, 356
Ovomucoid, 285, 354, 357
Oxalate, as anticoagulant, 327
Oxaloacetic acid, 120
Oxidation of disulfide linkages, 281
Oxonium ion, 21
Oxyhemoglobin, 346
Oxytocin, 2, 155, 285
 structure, 375

Paladini, A. C., 155
Palmer, A. H., 179, 182
Pantothenic acid, 111
Papain, 73, 76, 405
 terminal residues, 285
Paper electrophoresis, 203
Pappenheimer, A. M., 422
Paramagnetism, 349
Partial specific volume, 214
Partition chromatography, 93
Partition ratios, 134
Partography, 93
Pasteur, L., 381
Pauling, L., 283, 291, 293, 301, 418, 420, 423
Penicillin, 138, 229
Pepsin, 76, 218, 405
 details of crystallization, 382
 low isoelectric point of, 32
 production from pepsinogen, 387-388
 properties, 388
 terminal residues, 285
Pepsin inhibitor, 140, 388
Pepsinogen, 387
 conversion to pepsin, 387-388
 properties, 388
Peptic degradation of γ -globulin, 344
Peptidase, definition, 400
Peptide, 6
 differentiation from protein, 131
 terminology, 132
Peptide analysis methods, comparative features, 157
Peptide antibiotics, families, 141
Peptide bond, energetics of formation, 409
 in proteins, 279-280
 synthesis, 409
Peptide chain, changes in folding, 315, 317, 320-324
 helical structures, 293-295, 323
Peptide linkage, 6
 double-bond character, 291
 exposure of on denaturation, 307

- Peptides, allergenic, 140
 as objects of study, 3
 assignment of structure, 144
 families of, 140
 fractionation, 131-132
 by countercurrent distribution, 133
 by electrophoresis, 133
 by paper chromatography, 133
 on ion exchange columns, 133
 hydrolytic rates, 77
 infrared absorption, 294
 large, by synthesis, 163
 naturally occurring, 131
 occurrence in families, 140
 progression of properties with size, 163
 quantitative analysis, of dynamic changes, 156
 of residue sequence, 156
 selective precipitation, 132
 separation by zone electrophoresis, 204
 synthesis, 160
 titration behavior of, 33-37
 Peptide structure, assignment, by
 amino group reaction, 147
 by application of amino condensations, 148
 by Edman method, 152
 by enzymic methods, 147
 by Grignard reaction, 151
 by physical properties, 148
 by stepwise degradation, 151
 by thiohydantoin formation, 151
 by thiothiazolidone method, 153
 by titration, 147
 from studies of fragments, 155
 with acyl halides, 150
 with aryl isocyanate, 149
 with benzylamine, 150
 with dinitrofluorobenzene, 148
 with hypobromite, 150
 with nitrous acid, 153
 with trinitrotoluene, 148
 by racemization, 147
 examples of assignment, 154
 Peptide synthesis, by carbobenzoxy method, 164
 by dicyclohexylcarbodiimide, 169
 by hydrolysis of diketopiperazines, 162
 by N-carbobenzoxy amino acid anhydrides, 170
 by phosphite amide anhydride modification, 170
 by tosyl derivatives, 164
 Peptide synthesis, objectives, 160
 through aminoacyl halides, 160
 through haloacyl halides, 162
 use of Curtius azide modification, 165
 Pertzoff, V. A., 43
 Perutz, M. F., 231, 300, 347
 Pervaporation, 269-270
 Pflüger, E., 430
 pH, control of in protein fractionation, 270
 effect, on casein, 358
 on clotting reaction, 341
 on optical activity of amino acids, 41
 on solubility of proteins, 241-242
 on protein denaturation, 312, 318, 322, 337
 measurement of in protein titration studies, 177
 Phase rule, applied to protein purity, 254
 Phenolic groups, 306
 evidence for masking of in ovalbumin, 188
 in proteins, 179, 182, 306, 341
 Phenylalanine, 119
 metabolism, 122
 structure, 10
 synthesis, 68
 D-Phenylalanine, in tyrocidine, 138
 Phenylhydantoins, in determination of structure of peptides, 146
 Phenylisothiocyanate, 152
 Phenylpyruvic acid, 119
 Phenylthiohydantoins, 51
 Philpot-Svensson optical system, 198-199, 217
 Phosphate ester, in ovalbumin, 355
 Phosphate linkage, in proteins, 282
 Phosphite amide synthesis, 170
 Phospholipid, 345
 Phosphoprotein, 279, 358
 Phosphorylase, 362
 Photometer, 94
 Phthalimidomalic ester, in synthesis, 69
 Phthaloylamino acids, 51
 Picric acid, 248, 268
 Picrolonic acid, 248
 Pilocarpine, 116
 Pinhey, K. G., 36, 145
 Pipsyl chloride, 93
 Pirie, N. W., 36, 145
 Pitt-Rivers, R., 370
 pK values, of titratable groups in proteins, 176

- pK_a*, definition, 23
 of acids, 24
 of amino acids, 26
- pK₀*, definition, 185-186
- Plakalbumin, 355
- Planar folding, of peptide chains, 299-300
- Plasma, blood, 296, 328
 osmotic pressure of, 338
 swine, electrophoretic pattern, 202
- Plasma albumin, *see* *Albumins*
- Plasma proteins, clinical uses, 341-343
 distribution, 329-331
 function of, 334-335
- Pleated sheet structure, 291-292, 299
- Polarimetry, use of in assessing purity of amino acids, 44
- Polarization of fluorescence, 224
- Polyaffinity theory, 402
- Polyamino acids, 170
- Polybasic acids, dissociation properties of, 182-183
- Polyglycine, 320
- Polymerization, in clotting process, 341
 of amino acids, 170
- Polymixins, 139
- Polypeptide, 6
- Pomelin, 363
- Porphyrindin, 60
- Porphyrin ring, 345
- Porter, R. R., 423
- Potassium trioxolatochromate, 56, 80
- Potential gradient, 192-193
- Potter, V. R., 108
- Precipitation, of denatured proteins, 317, 319
 of proteins, 271-273
- Precipitin, 416
- Precipitin reaction, 264
- Precision, 99
- Preference, 400
- Preparative electrophoresis, 273
- Prephenic acid, 119
- Pressman, D., 418
- Pressure, effect of on protein denaturation, 313-315, 319
- Prolamins, 234, 241, 363-364
 solubility in alcohol, 363
 solubility in alkali, 363
- Polidase, 401
- Prolinase, 401
- Proline, 118
 isolation, 81
 selective precipitation, 56
 solubility of, 38
 structure, 10
- Proline, synthesis, 72
- Prolylamino acids, synthesis, 172
- Prosthetic groups, 380
 in proteins, 279
- Protamines, 255, 295-296
 basic character of, 32
- Protease, classification, 404
 definition, 400
- Protease action, on insulin fraction A, 407
 on large substrates, 406
- Proteases, as transfer enzymes, 406
 concerted action in digestion, 408
 terminal sequences, 395
- Protein, biosynthesis of, 429, 436
 catabolism, 109
 cereal, 60
 composition, 101
 continuity of structures, 434
 correspondence of nucleic acid diversity, 435
 evolution, 429
 number of types, 435
 origin, 429
 prebiological synthesis, 431
- Protein biosynthesis, and ATP, 437
 energetic requirement, 436
 γ -glutamyl amino acids in, 437
 phosphoryl amino acids in, 436
- Protein denaturation, *see* Denaturation
- Protein purification, by ion exchange, 267
- Protein purity, 261-262, 264
- Protein research, interdisciplinary nature, 3
- Protein sequential structure, assignment from smaller peptides, 155
- Protein solubility, effect of ionic interaction, 241-253
 of organic solvents, 246-247
- Protein specificity, immunological, 417
- Protein synthesis, 409
- Proteinase, definition, 400
- Proteins, amino acid composition, 278
 amino acid sequence in, 284-287
 as enzymes, 379
 as hormones, 367
 base-binding capacity, 180-181
 basic groups, 60
 biological significance, 2
 classification, 435
 criteria of homogeneity, 205
 denatured, 60
 determination by ultraviolet absorption, 265
 distribution of charge forms, 183

- Proteins, drying of, 268
 extraction by ultrasonic vibrations, 265
 fractionation using organic solvents, 271, 357
 homogeneity of, *see* Homogeneity
 hydrolysis, 75
 in therapy, 3
 inductive effects in, 184-185
 industrial uses, 3
 masking of phenolic groups in, 188
 molecular expansion, 187, 311, 337-338
 native, 60
 occurrence in families, 140
 optical activity, 294
 preparation, 265
 purity, based on phase rule, 254
 based on solubility, 263-264
 relationship to peptides, 6
 separation by electrophoresis, 195, 204
 similarity of composition, 432
 solubility, *see* Solubility
 titration behavior and ionic strength, 185-186
 variability of composition, 260-261
 Proteolysis, 307, 319, 344, 355
 by thrombin, 339-340
 Proteolytic enzymes, action patterns, 405
 classification, 400
 Prothrombin, 339
 Protons, in solution, 21
 Pseudoglobulin, 328
 PTC, 152
 Pteroylglutamic acid, 136
 Push-pull mechanism, 396
 Putnam, F. W., 310
 Pyrrolidyl amino acids, 14
 Pyrroline-5-carboxylic acid, 118
 Quill, porcupine, 230
 Racemization, during hydrolysis, 75
 of amino acids, 45
 Radius of molecule, relation to electrophoretic mobility, 193
 Raman spectra, of amino acids, 29
 Raoult law, 208
 Rapid-flow technique, 189
 Reciprocity, 437
 Recovery experiments, in assay, 99
 Reduction, of disulfide linkage, 305-306
 solubilization of keratins by, 266
 Refractive gradient, observation of, 197-199
 Refractive increment, 211
 Reinecke salt, 248
 Reliability, in amino acid assays, 96
 Rennin, 358
 Replication, 1
 Resolution, of racemic amino acids, 45, 72
 Requirements, assay organisms, 97
 metabolic, 105
 nutritional, 105
 Revelle, R., 430
 Reversibility, of clotting process, 341
 of denaturation, 315-316, 319
 Reversible boundary spreading, 262
 Reversible dissociation, of proteins, 296, 300
 R_f , definition, 93
 Riboflavin, in egg white, 356
 Riboflavin adenine nucleotide, 390
 Riboflavin phosphate, 390
 Ribonuclease, 218, 281, 295-296
 Ribonucleic acid, 424
 Rice, O. K., 240
 Rich, A., 290
 Rigidity, of gels, 256
 Ritthausen, H., 17
 RNA, 424
 Robinson, C., 294
 Roche, J., 370
 Rose, W. C., 17, 106
 Rotary diffusion coefficient, 221
 Rotatory dispersion, 312, 322
 Rubey, W. W., 429-430
 Rugo, H. J., 230
 Rye seed protein, terminal residues in, 285
 Salicylate, as solvent for gliadin, 365
 denaturation of proteins by, 313
 Salmine, 285
 Salt, effect on solubility of proteins, 241-245
 concentration, in protein fractionation, 270-271
 Salting-in effect, 243, 245, 247, 271
 Salting-out effect, 243, 245, 328
 use of in separating albumins and globulins, 271, 328
 Sanger, F., 50, 144-145, 149, 284, 371-372, 375
 Scale-line displacement, 197
 Scanning technique, 197-198
 Schellman, J., 320, 323
 Scheraga, H., 219
 Schleich, H., 402
 Schlieren (Toepler) principle, 196

- Schmidt, C. L. A., 7
Schneider, F., 402
Schoenheimer, R., 106, 163, 164
Schomaker, V., 319
Schotten-Baumann reaction (acylation), 50, 94, 161, 164
Schryver, S. B., 17
Schultze, E., 17
Scleroproteins, 235
Secretin, composition, 139
Sedimentation velocity, 215-216, 309
Seed proteins, 362-363
Selective adsorption, 89
Semipermeable membrane, 210, 268
use in dialysis, 194, 266
Separation of proteins, by electrophoresis, 195, 204
Sequence of amino acid residues, in proteins, 284-287
Sequential analysis, 2
Serine, in proteins, 282
metabolism, 122
structure, 10
synthesis, 70
Serum, blood, 328
Serum prothrombin conversion accelerator, 339
Sheehan, J. C., 169
Shikimic acid, 119
Shock, treatment of with plasma albumin, 338
Shumaker, J., 222
Sickle-cell anemia, 349-350
Side-chain spacing, in β -keratin, 288
Siderophilin, 345
Silica gel, in chromatography of proteins, 275
in zone electrophoresis, 203
Silk fibroin, 229, 287-290
partial sequence in 1933, 145
Simpson, R. B., 320
Skatole, 122
Skin, nitration, 62
Smith, D. B., 433
Smith, E., 132, 404, 423
Solubility, as a criterion of protein purity, 263-264
changes in associated with protein denaturation, 308
of amino acids, 39
of proteins, effect of alcohol, 243, 246-247
effect of ionic strength, 241-245
effect of ion interaction, 241-253
effect of organic solvents, 246-247
effect of pH, 241-242
Solubility, in relation to composition, 240-241
Solutions, ideal, 208-209
Solvation, 208, 215, 219, 221, 239
Solvents, organic, effect on protein solubility, 246-247
Sørensen, S. P. L., 30, 38, 175, 252, 296
Sørensen's formol titration, 52
Soybean meal, 123
Soybean protein, 285
Specialized Instruments Corp., 214
Species specificity, 416
Specificity, 400
Spectrophotometer, 94
Speiser, R., 211
St. Gyorgyi, A., 432
Stanley, W., 1-2
Starch, use of in zone electrophoresis, 203
Statistical evaluation, 99
Steinhardt, J., 188-189, 248-249
Stein-Moore assay, 90, 99-100
Sterman, M., 322
Stokes-Einstein theory, 193
Streaming birefringence, 221, 222-224, 297, 311, 339, 360-361
Strecker synthesis, 67
Strepogenin, activity, 140
structure, 140
Structure determination, by X-ray diffraction, 229
Substrate, influence of structural features, 401
natural vs. synthetic, 406
peptide, 399
protein, 399
Substrate specificity, 400
Substrate susceptibility, 400
Subtilins, 139
Sulphydryl groups, in plasma albumin, 333
in protein, 60, 179, 333
liberation of in denaturation, 305-306, 313-314
Sulfur-containing amino acids, 14
Sumner, J. B., 381
Surface monolayers, 228, 301, 307-308, 313
Susceptibility, 400
Svedberg, T., 175, 197, 207, 212, 216, 295-297
Svedberg unit, sedimentation velocity, 215
Synge, R. L. M., 77, 145, 154-155
Synthesis of peptides, 160

- Synthetic abilities, 105
 Synthetic peptides, 280, 322
- Tactoids, 257-258
 Tanford, C., 177-178, 182, 185
 Taurine, 120
 Teas, H. J., 108
 Temperature, effect on protein denaturation, 312, 319
 effect on solubility of proteins, 270
 Templet hypothesis, 437
 Terminal residues, in proteins, 284-287, 340
 of hemoglobin, 348
 Teropterin, 136
 Tetraethyl pyrophosphite, 170
 Thermodynamics of denaturation, 315-317
 Thienylalanine, 392
 Thiothiazolidone method, 153
 Threonine, optical isomers of, 44
 structure, 10
 synthesis, 70
 Thrombin, 339-340, 343
 Thromboplastin (thrombokinase), 339
 Thyroglobulin, 16, 319
 Thyroxine, 113
 as hapten, 419
 biosynthesis, 114
 physiological properties, 369
 production, by iodination of protein, 370
 from diiodotyrosine, 370
 solubility of, 38
 structure, 10
 Time factor, in nutrition, 123
 Tiselius, A., 196, 200, 202, 275, 328
 Titration, of wool protein with various acids, 248-249
 Titration behavior, comparative, of native and denatured proteins, 306
 Titration curves, of proteins, 176-179
 interpretation of, 179-184
 properties of, 25
 relation to electrophoretic mobility, 192
 TMV, 1, 224, 250, 257-258, 307-308, 314, 425
 dimensions of, 225
 hybridization, 426
 terminal residues, 285
 TMV protein, 426
 C-terminal structure, 426
 subunit weight, 426
 TMV's, amino acid compositions, 425
- Tobacco mosaic virus, *see* TMV
 Tobacco necrosis virus, 226-227
 Toepler schlieren principle, 196
 Toxin, 421
 botulinus, 421
 diphtheria, 421
 disadvantages in therapy, 422
 tetanus, 421
 Toxins, as disease agents, 422
 Toxoid, 421-422
 Tracers, in metabolic studies, 107
 Transforming principles, 424
 Transpeptidation, 410
 Trichloroacetic acid, precipitation of proteins by, 250, 265, 268, 357
 Triiodothyronine, 370
 structure, 16
 Triketohydridene hydrate, 52
 Trotter, I., 294
 Trypsin, 76, 389, 405
 denaturation of, 316
 proteolysis of small substrate, 401
 terminal residues, 285
 Trypsin inhibitor, 140, 357
 Trypsinogen, 389
 Tryptophan, 111, 119, 261, 278, 306
 absence of in zein, 364
 biosynthesis, 123
 destruction during hydrolysis, 76
 effect of pH on optical rotation, 42
 extraction, 78
 isolation, 81
 metabolism, 122
 structure, 10
 Tuppy, H., 372
 Turbidity, relation to molecular weight, 211
 Tutin, F., 145
 Tyndall effect, 210
 Tyrocidine, 137-138
 Tyrocidine A, 138
 structure, 141, 155
 Tyrocidine B, structure, 141
 Tyrosine, 119, 306
 amphoteric properties, 27
 isolation, 81
 metabolism, 122
 reactivity in proteins, 60
 solubility of, 38-39
 structure, 10
 Tyrothricin, 138
- Ultracentrifugation, as a criterion of protein homogeneity, 262-263
 Ultrafiltration, 268
 Ultramicroscope, 210

- Ultrasonic vibrations, effect on proteins, 315
in protein extraction, 265
Ultraviolet absorption, determination of proteins by, 265
Unit cell, 229
University Laboratory of Physical Chemistry, Harvard, 213, 269, 272, 334-335, 342-343
Unsaturated acid synthesis, 70
Urea, 118
denaturation of proteins with, 304, 308-310, 312, 316, 319, 322, 333, 337, 363
disaggregation of myosin by, 361
formation, 111
hydrolysis by urease, 363
Urea cycle, 112
Urease, 314, 363
crystallization, 381
Ureidosuccinic acid, prebiological synthesis, 431
Ureotelic metabolism, 113
Urey, H., 430
Uric acid, 113
Uricotelic metabolism, 113

Valine, structure, 10
D-Valine, in gramicidin, 138
Van der Waals forces, 239, 252, 255-256
Van Slyke, D., 17, 52
Van Slyke amino nitrogen analysis, 283-284
Van Slyke reaction, 52
van't Hoff, J., 209
Vapor pressure, lowering of in solutions, 209
Varner, J. E., 430
Vasopressin, 155, 285
structure, 375
Velick, S. F., 253
Veronal buffer, in electrophoresis, 203
Vickery, H. B., 7
Viruses (nucleo-), proteins as, 3
Viscosity, intrinsic, 220-221
changes in accompanying denaturation, 309-311, 320, 322
solute, 216
Vitalism, 1
Vitamin B₁₂, 124

Vitamins, from amino acids, 110
Wagner, M. L., 182
Walden inversion, 74
Waley, S. G., 411
Warner, R., 320
Water, amphoteric properties of, 21-22
in relation to protein denaturation, 312
Watson, J., 411
Watson, M. T., 320
Wave length of light, in light scattering, 211
Weight average molecular weight, 212
Weyl, T., 17
Wheat seed protein, terminal residues in, 285
Whey proteins, 359-360
White, W. F., 375
Williams, J. W., 319
Williams, R. C., 1, 426
Wöhler, F., 1-2
Wollaston, W. H., 18
Woolley, D. W., 124, 140
Wool protein, 248-249
Wright, G., 319
Wrinch, D., 298-299
Wyckoff, R. W. G., 226-227

Xanthoproteic reaction, 264
X-protein, 345
X-ray diffraction, 228, 287, 289-290, 300-301, 337, 347-348, 359, 363
evidence for dipolarionic form of amino acids, 30
of denatured proteins, 309
of protein gels, 257-258
X-rays, scattering of, 212

Yang, J. T., 294
Young, R. W., 170

Zaiser, E., 188-189
Zein, 218, 224, 241, 250, 261, 297
acid-binding capacity of, 181
properties of, 364-365
Zervas, L., 163, 399, 402
Zone electrophoresis, 204
Zwitterion, 27
Zymogens, 387

547.75
F833i

36623

Fox

Introduction to protein chemistry

547.75
F833i

36623

Fox

Introduction to protein chemistry

DATE	ISSUED TO

THE MASTER'S COLLEGE
POWELL LIBRARY
SANTA CLARITA, CA 91321

THE MASTER'S COLLEGE

547.75 F833i

MAIN

Fox, Sidney W./Introduction to protein c



3 3540 00007 1419