The CHEMISTRY and FUNCTION of PROTEINS

FELIX HAUROWITZ

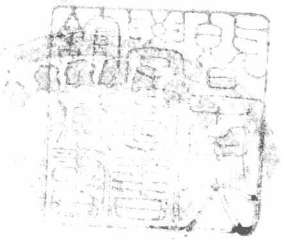
Academic Press

The CHEMISTRY and FUNCTION of PROTEINS

SECOND EDITION

By Felix Haurowitz

Distinguished Service Professor of Chemistry
Indiana University, Bloomington, Indiana



1963

ACADEMIC PRESS • New York and London

ACADEMIC PRESS INC.
111 Fifth Avenue, New York 3, New York

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
Berkeley Square House, London W.1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 63-16962

PRINTED IN THE UNITED STATES OF AMERICA

Foreword

The first edition of this book was published 1950. It was written as a textbook for a one-semester course on proteins for graduate students. The author's intent was to write a book which would be understood not only by graduate students of chemistry, but also by students of the biological sciences, who sometimes have a poor background in physical chemistry. The main topics discussed were (1) protein structure, (2) biological activity of proteins, and (3) biosynthesis of proteins. The book was so well accepted by the scientific world that it soon had to be reprinted. It was also translated into Russian and Japanese. These successes encouraged the author to write this second edition.

During the last decade, work on proteins has increased enormously. Chromatography, electrophoresis, and their variations have enabled biochemists to solve the amino acid sequence in a number of proteins and thus to obtain a deeper insight into structural problems. Very little was known about these methods when the first edition was written. Sanger had not yet published his classical work on the structure of insulin. The situation in protein chemistry was characterized by the following sentence (first edition, p. 116): "While the sequence of amino acids in shorter peptide chains can be determined, it would be hopeless to endeavor to fully characterize peptide chains containing 100 or more amino acids." Evidently these words were too pessimistic. Most of the space in Chapters I to VII of this book is devoted to the methods used in work on the *structure* of proteins.

A second part of the book (Chapters VIII to XV) is concerned with the properties of those proteins which have been investigated more thoroughly than others. It would have been preferable to classify these proteins systematically and to discuss them according to such a logical scheme. Since we do not yet have a satisfactory classification scheme, the system used in the first edition of this book was retained. The proteins were classified according to their occurrence in nature and their function. The latter principle is of particular value since it enables us to correlate structure with *function*. We know at present much more about the structural bases of enzyme action, of hormonal effects, and of antigen-antibody interaction than was known twelve years ago.

Our knowledge of amino acid supply and amino acid biosynthesis has

increased considerably since 1950. Work of this type is based on metabolic experiments with isotopically labeled amino acids, their precursors and intermediates; it is thus quite different from work on the structure of proteins. For this reason, it was decided to limit the discussion of amino acid metabolism to a short section in Chapter XVI.

Nevertheless, it seemed desirable to discuss in some detail the *biosynthesis* of proteins from amino acids, since the structure of proteins is the result of the biosynthetic mechanism. In writing on biosynthesis it was clear to the author that he was reporting on a problem which is in a state of rapid flux. However, the correlation between structure and biosynthesis of proteins is so important that it seemed useful to report on our present views, realizing that they are ephemeral and that they will change in the near future. It was the author's particular endeavor in this part of the book to distinguish experimental facts from hypotheses which, although frequently ingenious, still need further experimental support.

Although it was attempted to give the reader a survey on the entire field of protein chemistry, it was obviously impossible to cite all papers which deal with protein chemistry. It would have been easy to quote only the most recent publications from which the interested reader can go back to the original sources. This would have involved the omission of the names of Kossel, Sörensen, Michaelis, and other pioneers in the development of protein chemistry. It was difficult to strike the right balance in this respect. If good review articles on the subject under discussion were available, the reader was frequently referred to these. Most of the papers quoted were published during the last decade, i.e., after the publication of the first edition of this book, and a large number of these appeared in 1960 or later.

In view of the extent of the bibliography it was decided to condense it by using special abbreviations for those journals which were most frequently quoted, and by listing at most two authors for each reference. A list of the abbreviations for the journals will be found on pages ix and x. The finding of the references cited has been facilitated by the use of bold print for their numbers and by arranging them according to the sections of each chapter. Further condensation of the volume of the book was accomplished by the extensive use of small print for problems of secondary importance. The rapid progress of research, particularly in the field treated in the last chapter, necessitated additions and changes in the period between the submission of the manuscript for publication and the final completion of the page proofs. An attempt has been made to include all of the important literature published in 1962.

Since the emphasis of this second edition of the book is on problems of structure and function, the title was changed from "Chemistry and

Biology of Proteins" to "The Chemistry and Function of Proteins." The enormous increase in the literature on proteins made it desirable to submit most of the chapters of this book to experts in the respective fields. An appeal to several colleagues resulted in a response which far surpassed the author's expectation. The following colleagues were kind enough to read different chapters or sections and to send or give to the author their detailed written comments: Dr. R. S. Baer (Boston University), Dr. Wm. W. Bromer (Eli Lilly and Co.), Dr. H. B. Bull (University of Iowa), Dr. E. O. Davidson (Eli Lilly and Co.), Dr. J. T. Edsall (Harvard University), Dr. J. F. Foster (Purdue University), Dr. E. Gold (Indiana University), Dr. J. Gross (Massachusetts General Hospital), Dr. F. R. N. Gurd (Indiana University), Dr. R. L. Hill (Duke University), Dr. W. J. Kleinschmidt (Eli Lilly and Co.), Dr. I. M. Klotz (Northwestern University), Dr. H. R. Mahler (Indiana University), Dr. M. M. Marsh (Eli Lilly and Co.), Dr. W. F. H. M. Mommaerts (University of California, Los Angeles), Dr. E. A. Peterson (National Institutes of Health), Dr. H. V. Rickenberg (Indiana University), Dr. A. H. Sievert (Lake Bluff, Illinois), Dr. H. Wm. Sievert (Abbott Laboratories), Dr. H. A. Sober (National Institutes of Health), Dr. Ch. Tanford (Duke University), Dr. J. A. Thoma (Indiana University), Dr. G. Vidaver (Indiana University), Dr. D. Wetlaufer (University of Minnesota), and Dr. R. J. Winzler (University of Illinois). The author is also indebted to Mr. J. L. Groff for help in the preparation of the manuscript, and to Dr. L. Stewart for help in reading the proofs. The author's appeal to the readers of the first edition of the book elicited a number of valuable comments. Particular thanks are due to Dr. K. H. Gustavson (Stockholm) and Dr. R. Hotchkiss (The Rockefeller Institute) for detailed suggestions for improvements. Because of this generous help, use could be made of a series of most valuable critical remarks and suggestions, and many errors could be avoided.

In spite of the ample and generous aid of so many colleagues, it is unavoidable that a book on such a wide area of intensive research and rapid progress should contain errors. The only person responsible for them and for the mode of presentation is the author.

FELIX HAUROWITZ

Bloomington, Indiana
April, 1963

List of Abbreviations

The following abbreviations are used for amino acyl residues in protein or peptide molecules.

Ala = alanine
Arg = arginine
Asp = aspartic acid
AspN or AsN = asparagine
Cys = cystine, cysteine
Glu = glutamic acid
GluN or GlN = glutamine
Gly = glycine
His = histidine
Hypro = hydroxyproline
Ileu = isoleucine
Met = methionine
Phe = phenylalanine
Pro = proline
Ser = serine
Thr = threonine
Try = tryptophan
Tyr = tyrosine
Val = valine

Other abbreviations frequently used in articles on protein chemistry are listed below.

Å = angstrom unit = 0.1 mμ = 10^{-8} cm
A (in polynucleotides) = adenosine monophosphate residue
ABB = *Archives of Biochemistry and Biophysics*
ADP = adenosine diphosphate
ATP = adenosine triphosphate
BBA = *Biochemica et Biophysica Acta*
BGG = bovine γ-globulin
BJ = *Biochemical Journal* (*London*)
BSA = Bovine serum albumin
C = curie = 10^3 mC(millicurie) = 10^6 μC (microcurie)
C (in nucleic acids) = cytidine monophosphate residue
cal = calorie = 10^{-3} kcal
Cbo or Cbzo = carbobenzoxy residue
cpm = counts per minute
DAB = diaminobutyric acid
DFP = diisopropylfluorophosphate = diisopropyl-phosphofluoridate
DIT = diiodotyrosine
DNA = deoxyribonucleic acid
DNFB = 2,4-dinitrofluorobenzene
DNP = dinitrophenyl
DNPr = deoxyribonucleoprotein
DPN = diphosphopyridine nucleotide = NAD
DOC = deoxycholate
Dopa = dihydroxyphenylalanine
ETP = electron transport particle

ETS = electron transport system
ΔF = free energy change
FSH = follicle-stimulating hormone
g = relative centrifugal force (see p. 71)
G (in nucleic acids) = guanosine monophosphate residue
GABA = γ-aminobutyric acid
ΔH = enthalpy change
Hb = hemoglobin
HDL = high density lipoproteins
ICSH = Interstitial cell-stimulating hormone
JACS = *Journal of the American Chemical Society*
JBC = *Journal of Biological Chemistry*
k_1, k_2, . . . = velocity constants
K_1, K_2, . . . = equilibrium constants
kcal = kilocalorie = 1000 cal
μg = microgram = 10^{-6} gram
MIT = monoiodotyrosine
mp = melting point
mRNA = messenger-RNA
MSH = melanophore-stimulating hormone
N = normal (concentration)
NAD = nicotinamide-adenine dinucleotide = DPN
NADP = nicotinamide-adenine dinucleotide phosphate = TPN
NEM = *N*-ethylmaleimide
PCMB or *p*CMB = *p*-chloromercuribenzoate
pipsyl = *p*-iodophenylsulfonyl
pK = $-\log$ K (see p. 95)
PTC = plasma thromboplastin component
R = gas constant = 1.98 cal per mole per degree = 8.314 joules per mole per degree = 0.08204 liter atmospheres per mole per degree
RNA = ribonucleic acid
RNPr = ribonucleoprotein
rpm = revolutions per minute
S = Svedberg unit (see page 73)
ΔS = entropy change
SDS = sodium dodecylsulfate
sRNA = soluble RNA
T = temperature in degrees Kelvin = °C + 273
T (in nucleic acids) = thymidine monophosphate residue
T_4 = thyroxine (see p. 348)
TCA = trichloroacetic acid
TMV = tobacco mosaic virus
TPN = triphosphopyridine nucleotide = NADP
U (in nucleic acids) = uridine monophosphate residue
UV = ultraviolet

Contents

Chapter I

Role of Proteins in Biology

Proteins have a particular significance in biology in that they constitute one of the indispensable components of living matter. To be sure, living organisms also contain carbohydrates and lipids, frequently in even greater abundance than proteins. Thus, green plants, which are rather poor in proteins, are rich in cellulose, a carbohydrate. There are, however, essential differences between proteins and most of the other cellular constituents. Wherever the phenomena of *growth* and *reproduction* are seen, proteins and nucleic acids are primarily involved. In animal and plant cells multiplication is initiated by the nucleus, in which proteins and nucleic acids are closely associated. In bacteria, where there is no typical nucleus, the bulk of the living substance is likewise formed by proteins and nucleoproteins. Both components are also found in viruses, the simplest of which are free of lipids and carbohydrates.

As the name of the proteins indicates (Greek *protos*, the first), they have been considered for many years the primary component of living matter. The great importance of the nucleic acids was recognized much later and has been in the foreground during the last decade. Many scientists seem to consider nucleic acids more important than proteins, and attribute to them an almost exclusive role in the genetic determination of inheritable characteristics. We must not forget, however, that growth and reproduction, like all processes which involve primarily the formation or degradation of proteins and nucleic acids, depend on the presence of certain *enzymes*, and that all these enzymes are proteins. Accordingly, nucleic acids cannot be formed in the absence of enzyme proteins. The presence of both proteins and nucleic acids is required for growth and multiplication.

The basic principle of the *chemical structure* of proteins is quite simple. They consist of long chains of amino acids linked to each other by peptide bonds. Complications arise, however, (1) from the presence of about 20 different types of amino acid residues in the peptide chains, (2) from the great length of these chains, which may consist of several hundred amino acids, and (3) from a particular *conformation* of the peptide chains, i.e., their specific folding, which results in a definite three-dimen-

sional pattern. Even if all proteins consisted of straight peptide chains, free of folds and turns, an almost infinite number of proteins could be formed merely by changes in the order of the 20 amino acids in the long chains. Since each of these chains could assume an unlimited number of conformations, it is not surprising that each animal or plant species has its own species-specific protein. In some instances we are even able to find variations among the proteins of individuals of the same species. Only during the last decade have details of the amino acid sequence and of the conformation of some of the proteins been revealed.

It is understandable that the *conformation* of a protein is of the greatest importance for its chemical and physical properties as well as for its biological functions. The solubility of the proteins, their serological behavior, and their enzymatic and hormonal activities depend on the molecular groups present within the *surface* of the molecules, and hence on the spatial arrangement and the mode of folding of the peptide chains. It is, therefore, one of the principal endeavors of protein chemistry to elucidate the three-dimensional structure of the protein macromolecules and the distribution of functional groups within the molecules.

If the natural conformation of a protein molecule is destroyed or modified by physical or chemical operations, many of the chemical and physicochemical properties are changed, and the biological functions of the original *native protein* are altered or lost. We designate the modified product of such reactions as *denatured protein.* In using these terms we must be careful to avoid oversimplifications and fallacies. The designation *native* does not necessarily imply that the pure protein which we call native is identical with the form in which the same protein exists in the living cell. It is unavoidable that even the most careful methods of preparation destroy some of the weak bonds which link the protein molecule to neighboring molecules of other types inside the living cell. It is quite possible that some of the proteins occur in living organisms as giant "supermolecules" whose peptide chains form bridges between adjacent cells or extend over a series of cells. The pure crystalline protein preparations in a labeled bottle may not always have much in common with the protein molecules present in living tissues. It is also clear that denaturation can take place in many different ways and that we are usually dealing with a mixture of denatured products formed from one and the same native protein.

In spite of the dangers involved in the *preparation* and *purification* of proteins, protein chemists have succeeded in isolating pure, frequently crystalline, proteins which still manifest the same enzymatic or hormonal activities as the crude extracts of tissues or organs. Therefore we have every reason to assume that enzymes such as crystalline urease, isolated

first by Sumner (1926), and crystalline pepsin, prepared by Northrop (1929), are essentially identical with the enzyme molecules in the living cells, and that the same is true for the pituitary hormones vasopressin and oxytocin isolated from the posterior lobe of the hypophysis by du Vigneaud (1952). Although it is quite possible that even these well-defined molecules occur in the cells more or less loosely bound to other proteins, their isolation does not seem to affect their biological activity significantly.

General problems of protein structure are discussed in the first part of this book (Chapters II–VII). A general survey of the isolation and purification of proteins (Chapter II) is followed by description of their degradation to amino acids (Chapter III) and of methods for the determination of amino acid sequences (Chapter IV). The physical chemistry of the proteins is discussed in the subsequent chapters which deal with the size and shape (Chapter V), electrochemistry and hydration (Chapter VI), and internal structure or conformation (Chapter VII), of the proteins.

In the second part of the book (Chapters VIII–XIV) the most important types of proteins are discussed. Since an enormous number of proteins is known, and since their properties vary considerably, many attempts have been made to classify proteins systematically. Most of these *classifications* are unsatisfactory and have been abandoned. One of the classifications, proposed by a committee of American and British biochemists, was based on the solubility of proteins in different solvents. Proteins soluble in 50%-saturated ammonium sulfate solution were called *albumins;* those precipitated by the same solvent were referred to as *globulins.* The latter class was subdivided into euglobulins, insoluble in salt-free water, and pseudoglobulins, soluble under the same conditions. However, the solubility of proteins in salt solutions depends not only on the concentration of the salts, but also on pH, temperature, and other factors. We do not yet know which of the chemical structures in the protein molecules are responsible for the differences in solubility.

In view of the present lack of a structural basis for the classification of proteins it seemed best to arrange the proteins in this text according to their biological distribution and function. The first group is formed by the soluble proteins of the blood plasma, milk, eggs, and seeds, and by the protamines and histones (Chapter VIII). Another group contains the structural proteins of the connective tissue, epidermis, and muscle (Chapter IX). Chapter X presents a survey on intermolecular forces acting between proteins and other molecules, and is written as an introduction to Chapter XI, which deals with the so-called conjugated proteins. These are complexes formed by the combination of proteins

with nonprotein substances. Although hemoglobin and some of the other conjugated proteins occur as such *in vivo*, we still do not know whether this is also true for the nucleoproteins. They may be formed *in vitro* by combination of proteins with nucleic acids isolated from the same biological material. Proteins with enzymatic activity are discussed in Chapter XII, protein hormones in Chapter XIII, and toxins in Chapter XIV. It is unavoidable that Chapters VIII–XIV overlap to a certain extent. For example, many of the enzymes belong to the conjugated proteins. Moreover, it is impossible to draw a sharp borderline between some of the genuine proteins which contain very small amounts of carbohydrate and the typical conjugated glycoproteins.

Chapter XV is devoted to the antibody proteins. Analyses of antibodies and of their mode of formation have provided us with important knowledge on the mechanism of immunological reactions and have also given us valuable information on the specificity and the biosynthesis of animal proteins.

The two final chapters of the first edition of this book dealt with protein biosynthesis. In the first of these, the problem of the supply of indispensable *amino acids* and the mechanisms of amino acid biosynthesis were discussed. During the 12 years which separate this edition from its predecessor, however, the principal problems of amino acid supply have been solved. Most of the pathways of amino acid biosynthesis have been considerably clarified. Today the problems of protein chemistry are quite different from those of the chemistry of the amino acids. For this reason amino acid biosynthesis and amino acid supply will not be treated in this text. Readers interested in amino acid chemistry are referred to Greenstein's fundamental treatise on amino acids (1) and to an authoritative book on their metabolism (2).

The last chapter of the preceding edition of this book was devoted to a discussion of the mechanism of protein biosynthesis. Our views have undergone dramatic changes since 1950 and are at present in a state of flux. Each month brings new reports on important discoveries, particularly on the mutual interaction of proteins and nucleic acids. The problem which is currently the focus of interest is the role of the nucleic acids in protein biosynthesis. Most biochemists believe that the "four-letter code" of the nucleic acids, which contain only 4 different nucleotides, determines the specific sequence of 20 different types of amino acids present in the protein macromolecules. The nucleic acids are considered to be the material which transmits genetic characteristics from generation to generation. Whatever the mechanism of this process may be, it involves the presence of enzymes which act as catalysts in the synthesis of both proteins and nucleic acids. As far as we know, all these

enzymes are proteins. The presence of these enzymes is just as indispensable for growth and replication as is the presence of nucleic acids. We are still far from an understanding of these intimate relations between nucleic acids and proteins. Therefore, our interpretations are only tentative and their validity is ephemeral.

In Chapter XVI an attempt has been made to draw a picture of the rapidly changing views on the mechanism of protein biosynthesis and on its dependence on the genetic role of the nucleic acids. This picture is based on a limited number of experimental facts and on their interpretation by means of ingenious theories and speculations. Since the latter undergo rapid changes, it is unavoidable that parts of this picture will be obsolete when this book is published.

The development of protein research in the near future can be predicted with more confidence. In the next few years we will learn the complete amino acid sequence of most of the well-known "pure proteins." We will find many more examples of genetically determined deviations in the amino acid sequence. We will learn whether the primary amino acid sequence alone determines the conformation of the peptide chains or whether proteins of identical primary structure can occur in two or more different conformations, and can have different types of folding of their peptide chains. Much more insight into the three-dimensional structure of the peptide chains will be gained. This, in turn, will give us more information on the "active sites" or "active patches" on the surface of globular enzymes, hormones, antigens, and antibodies, and also more information about the basis of their biological activities. There is hardly any doubt that these predictable discoveries will open new vistas and will thus lead to advances into unknown areas of molecular biology where the boundaries between biochemistry, biophysics, and the morphological sciences disappear.

REFERENCES

1. J. P. Greenstein and M. Winnitz, "Chemistry of Amino Acids," Wiley, New York, 1961. **2.** A. Meister, "Biochemistry of the Amino Acids," Academic Press, New York, 1957.