PROTEINS

A Guide to Study by Physical and Chemical Methods

By Rudy H. Haschemeyer and Audrey E. V. Haschemeyer

PROTEINS

A Guide to Study by Physical and Chemical Methods

RUDY H. HASCHEMEYER

Cornell University Medical School

AUDREY E. V. HASCHEMEYER

Hunter College of the City University of New York

A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS, New York · London · Sydney · Toronto

Library of Congress Cataloging in Publication Data

Haschemeyer, Rudy Harm, 1930–
Proteins: a guide to study by physical and chemical methods.

"A Wiley-Interscience publication."
Includes bibliographical references.
1. Proteins. 2. Biological chemistry—Technique.
I. Haschemeyer, Audrey E. V., 1936– joint author.
II. Title.

QP551.H37 574.1'9245 72-13134
ISBN 0-471-35850-9

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

Proteins comprise the structural building blocks as well as most of the functional machinery of living organisms. There are few areas of biological or biomedical research today that do not demand a working knowledge of the chemical and physical properties of these complex macromolecules. Thus the advanced study of proteins beyond the level of the general biochemistry course has become increasingly essential. This book evolved from many years of teaching courses in proteins to graduate students in biochemistry and biology and to advanced medical students. To meet the needs of these students, and of researchers in all branches of the biomedical sciences, we have attempted to present a comprehensive and critical guide to the physical and chemical methods available for the study of proteins. An evaluation of the current state of knowledge of protein structure, including the important areas of protein folding and conformational equilibria, is also presented.

The book is organized for a two semester course or, if physical methods alone are to be emphasized, a one semester course based on Part II. A general theoretical background is presented for each of the physical methods discussed to indicate both the potential and the limitations of the technique in its application to macromolecules. More rigorous derivations of equations are presented where we have found them to be valuable in teaching the material. Primary emphasis, however, is given to the usefulness of the various techniques for real problems involving proteins and their contributions to our overall understanding of protein structure and function. To facilitate access to further information on each topic a listing of important general references and reviews, and selected references of interest in the current literature is provided.

We wish to express our appreciation to the colleagues and students who offered valuable criticism and advice on portions of the manuscript. Special thanks are due to Dr. Robert W. Woody for his review of the complete manuscript. All responsibility for the opinions expressed herein and any errors that may occur, however, rests with the authors. We will welcome communications from interested scientists and students.

RUDOLPH HASCHEMEYER
AUDREY H. HASCHEMEYER

New York, New York
February 1973

Contents

PART ONE

I Introduction

In all areas of biological and medical research today there is increasing need for knowledge about proteins. These complex macromolecules, with particle weights ranging from the thousands to the millions, comprise in essence the working machinery of life. They cannot claim the central position held by the nucleic acids as the carriers of heredity, yet certain proteins are responsible for the control of expression of hereditary information from its first transcription from the gene to its final translation into new polypeptide chains. Hundreds of proteins have been identified in the category of enzymes, catalyzing a myriad of complex biochemical reactions. Others are responsible for the basic structural framework of living organisms; in higher animals these structural proteins include collagen of bones, cartilage, and tendons; keratin of hair and nails; elastin of blood vessels and ligaments; and myosin of muscle.

The name protein (Greek, *proteios*, of the first rank) was first used in 1838 by Mulder following a suggestion by Berzelius. Mulder was also among the first to do a systematic study of the elemental composition of proteins. Most proteins were found to contain 50 to 55% carbon, 6 to 7% hydrogen, 20 to 23% oxygen, and 12 to 19% nitrogen. Protein determinations based on nitrogen (assuming an average content of 16%) came to be used for analysis of tissues and food samples. Sulfur (0.2–3.0%) was found to occur in proteins, as was phosphorus in some cases (as high as 3%). Trace elements identified in certain proteins (e.g., 0.34% iron in hemoglobin) permitted calculation of minimum molecular weights. These results gave the first indication that proteins have large molecular weights compared with other organic substances known at that time. Decomposition to smaller molecular weight units could be achieved through hydrolysis catalyzed by acids, alkalis, or certain biological preparations (containing proteolytic enzymes). During the late 1800s, amino acids were identified as the basic building units

of proteins. Eventually 20 different amino acids were shown to occur as components of most proteins; a number of others were found in special cases.

Analysis of the more common amino acids proved that all but proline (which is actually an imino acid) have structures consisting of a carbon atom (the α-carbon) and four substituent groups: a carboxyl group, an amino group, a hydrogen atom, and an R group which differs among the various amino acids. Pasteur had shown in 1851 that amino acids are optically active, and this property was soon correlated with the asymmetry of the α-carbon atom resulting when all four substituents of the α-carbon are different (glycine is not optically active because R = hydrogen). Although amino acids were found to differ in the direction in which they rotate polarized light, it was eventually established that all amino acids which occur in proteins have the same configuration (denoted L) with regard to the arrangement of groups about the α-carbon atom.

The nature of the bond which links amino acids together in proteins was elucidated independently by Emil Fischer and Franz Hofmeister in 1902. They proposed that water is eliminated between the α-carboxyl group of one amino acid and the amino group of another to produce an amide linkage. The condensation of two amino acids to form a peptide is illustrated by the general equation.

$$\underset{\text{Amino acid}}{NH_2\overset{R}{\overset{|}{C}}HCOOH} + \underset{\text{Amino acid}}{NH_2\overset{R'}{\overset{|}{C}}HCOOH} \rightarrow \underset{\text{Peptide}}{NH_2\overset{R}{\overset{|}{C}}HCO{-}NH\overset{R'}{\overset{|}{C}}HCOOH} + H_2O \qquad (1\text{-}1)$$

Proteins isolated from different sources and by different techniques were found to vary considerably in their properties (e.g., solubility), and in 1908 an attempt was made to develop a scheme of classification. First, two general groups were differentiated: the simple proteins, which yield only amino acids upon hydrolysis, and the conjugated proteins, which contain prosthetic groups or other substances that are released upon hydrolysis. The proteins categorized as "simple" (though they are certainly quite complex structurally) were further subdivided into several groups. These include the albumins which are readily soluble in water; the globulins which are insoluble or sparingly soluble in water but are soluble in dilute neutral salts; the glutelins which are soluble in dilute acid or alkali; the prolamines which are soluble in 70 to 80% alcohol but insoluble in either water or absolute alcohol alone; the albuminoids or scleroproteins which are fibrous insoluble animal proteins; the histones which are basic proteins containing a high percentage of the basic amino acids (lysine, histidine, and arginine) and are soluble in water; and the

protamines which are low molecular weight, basic proteins and are soluble in water. The conjugated proteins are classified according to their nonprotein moieties or prosthetic groups, which may be nucleic acid (nucleoproteins), carbohydrate (mucoproteins and glycoproteins), lipid (lipoproteins), highly colored prosthetic groups (chromoproteins such as the hemoglobins, cytochromes, and flavoproteins), or metals (metalloproteins, a group that includes many enzymes). Although the division between groups is not sharp (as between albumins and globulins) and despite the fact that proteins may belong to more than one group (e.g., hemoglobin is both a chromoprotein and a metalloprotein), some of the terminology of this classification scheme is still in use.

Early attempts at protein fractionation were largely limited to methods that took advantage of the different solubility properties of proteins. Egg albumin from egg white was successfully crystallized by Hofmeister in 1889, but most proteins proved difficult to purify or to crystallize. In the 1940s the development of chromatography led to improved methods of purification of proteins and analysis of their components. Separation of acylated amino acids on the basis of their different partition coefficients between water and an immiscible organic solvent was achieved by Neuberger in 1938. In 1941 Martin and Synge used silica gel column chromatography for the separation of acylated amino acids, and in 1944 Consden, Gordon, and Martin obtained separation of free amino acids by paper chromatography. These methods made it possible to determine the amino acid composition of proteins quickly and conveniently. The techniques of chromatography were later extended to include separation of proteins themselves by either absorption, ion exchange, or molecular sieving.

With the automation of chromatographic methods, rapid and convenient determination of amino acid composition of extremely small amounts of proteins became possible. Table 1-1 shows the amino acid composition of a variety of proteins. Most proteins have fairly similar amino acid compositions even though they differ greatly in physical and biological properties. Thus an important next step was the determination of the order of the amino acids in the polypeptide chains of the protein. Sanger showed in 1947 that the compound 1-fluoro-2,4-dinitrobenzene could be reacted with free α-amino groups at the end of polypeptide chains to form a linkage stable to acid hydrolysis. This reaction permitted the identification of the *N*-terminal amino acid residues of proteins and contributed to the first sequence determination, that of the protein hormone insulin. Insulin was shown to consist of two chains joined by disulfide bonds, one containing 21 amino acid residues, the other containing 30 residues (Fig. 1-1). Since that time the complete or partial sequences of hundreds of proteins have been determined.

A major contribution to the study of the three-dimensional structure and physical properties of proteins was the development of the ultracentrifuge

Table 1-1 Amino acid composition of proteins [from a compilation by D. M. Kirschenbaum, *Anal. Biochem.*, **44**, 159 (1971)].

Amino acid	Glycine	Alanine	Valine	Leucine	Isoleucine	Proline	Serine	Threonine	Aspartic acid	Glutamic acid	Half-cystine	Methionine	Lysine	Arginine	Histidine	Phenylalanine	Tyrosine	Tryptophan	Amide ammonia	MW × 10^{-3}
Albumin (bovine)	15	44	35	58	14	28	26	32	54	80	36	4	58	23	17	26	19	2	27	65
Carbonic anhydrase B (porcine)	18	20	16	23	15	17	29	11	34	21	0	0	23	6	14	9	11	6	17	30
Casein (human)	3	7	19	26	13	39	9	9	11	39	0	3	11	3	5	5	7	1	32	24
Chymotrypsinogen B (porcine)	22	22	25	17	11	14	28	17	20	15	10	2	6	8	3	8	4	13	—	26
Creatine kinase (human)	66	48	46	64	25	45	41	31	73	75	7	19	66	32	28	26	20	6	53	81
Deoxyribonuclease A (bovine)	9	23	25	23	12	9	30	15	34	20	4	4	9	12	6	12	16	3	—	31
Enolase (salmon)	96	104	61	73	61	37	56	37	121	95	14	14	91	34	22	33	28	3	—	106
Ferredoxin (*Chromatium*)	5	3	6	3	6	5	4	6	8	16	9	1	2	2	2	0	3	0	—	10
β-Galactosidase (*E. coli*)	84	94	72	112	44	71	69	70	122	145	17	21	27	73	36	45	34	32	—	135
Glyceraldehyde-3-*P* dehydrogenase (human)	134	127	98	78	73	56	79	80	162	81	8	28	102	38	39	55	45	—	—	140
Glycerol-3-*P* dehydrogenase (rabbit)	30	25	25	24	22	12	9	11	22	33	9	5	22	7	7	12	3	2	—	30
Invertase (*Neurospora*)	158	124	122	114	57	93	141	137	189	130	9	20	55	54	26	67	63	44	—	187
Methionyl-tRNA synthetase (*E. coli*)	114	143	94	136	80	78	83	64	158	171	24	38	94	78	35	77	49	33	—	173
Penicillinase (*B. cereus*)	20	33	18	20	17	9	13	24	34	24	—	4	21	11	4	8	11	5	28	32
Prolactin (ovine)	11	9	10	22	11	11	15	9	22	22	6	7	9	11	8	6	7	2	—	23
Rhodopsin (bovine)	21	25	15	20	10	16	15	25	19	25	6	8	8	6	4	22	10	—	—	28
D-Serine dehydratase (*E. coli*)	45	47	27	46	17	14	27	20	31	49	5	12	19	17	11	18	13	4	—	46
Thrombin (bovine)	30	18	19	33	13	18	18	16	32	35	8	4	26	24	8	14	11	10	—	34
Thyroglobulin (human)	394	383	310	476	155	347	484	283	400	697	244	62	195	275	80	259	110	117	549	670
Urease (jackbean)	64	61	45	57	51	41	39	45	72	62	11	19	39	31	20	19	16	5	45	75

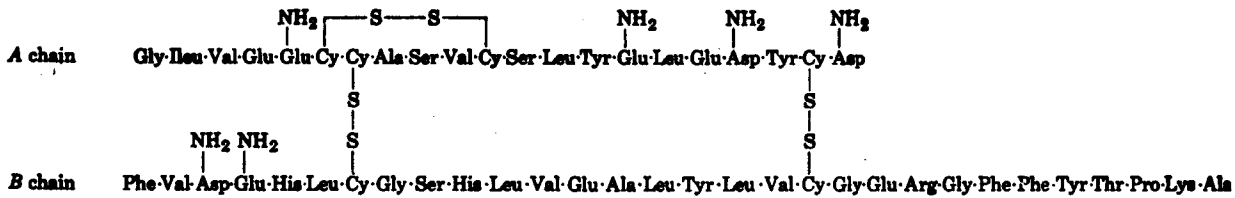

Fig. 1-1 The primary structure of beef insulin. (From F. Sanger, in *Currents in Biochem. Res.*, D. E. Green, ed., Wiley, New York, 1956.)

by Svedberg in the 1930s. From observations on rates of migration of proteins under high centrifugal fields, Svedberg and his associates demonstrated that different proteins have characteristic molecular sizes and shapes. Since this early work the technique of ultracentrifugation has been developed to become a powerful tool for many biological purposes. The use of electrophoresis for separation and characterization of proteins on the basis of migration in an electric field was introduced by Tiselius in 1933. The application to proteins of other physical techniques, such as the measurement of light scattering, viscosity, and diffusion, followed shortly thereafter.

During the past 10 or 20 years tremendous strides have been made in elucidating the three-dimensional structure of proteins. Structural analysis of crystals of amino acids and small polypeptides led to the determination of the spatial configuration of the peptide bond, and Pauling and Corey's

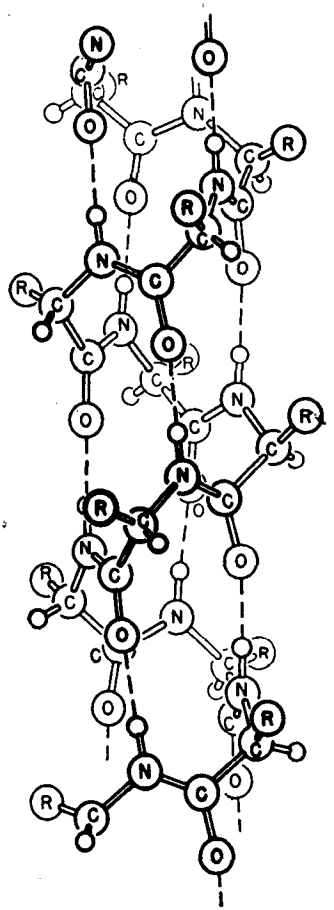

Fig. 1-2 The right-handed α-helix. (From R. B. Corey and L. Pauling, *Proc. Intern. Wool Textile Research Conf.*, Australia, Part B, 1955.)

proposal in 1951 of the α-helix as a basic model for polypeptides (Fig. 1-2). In the α-helix the polypeptide backbone is folded in such a way that the carbonyl group of one amino acid is hydrogen-bonded with the amide hydrogen of an amino acid further up the helix. Further investigation has largely substantiated these ideas, and it is now known that many proteins contain α-helical segments as well as other types of repeating hydrogen-bonded structures. Attempts to determine the complete three-dimensional structure of a protein by means of X-ray crystallography began in the middle 1930s and have reached an impressive productivity today. The first structure solved at high resolution was that of sperm whale myoglobin by J. C. Kendrew and his associates (1961), and these results were effectively applied to the study of a much larger but related molecule, horse hemoglobin, by M. F. Perutz's group. In the case of myoglobin, the positions of all non-hydrogen atoms in the molecule have been determined to yield the structure shown in Fig. 1-3. Other structures solved at high resolution in the late 1960s

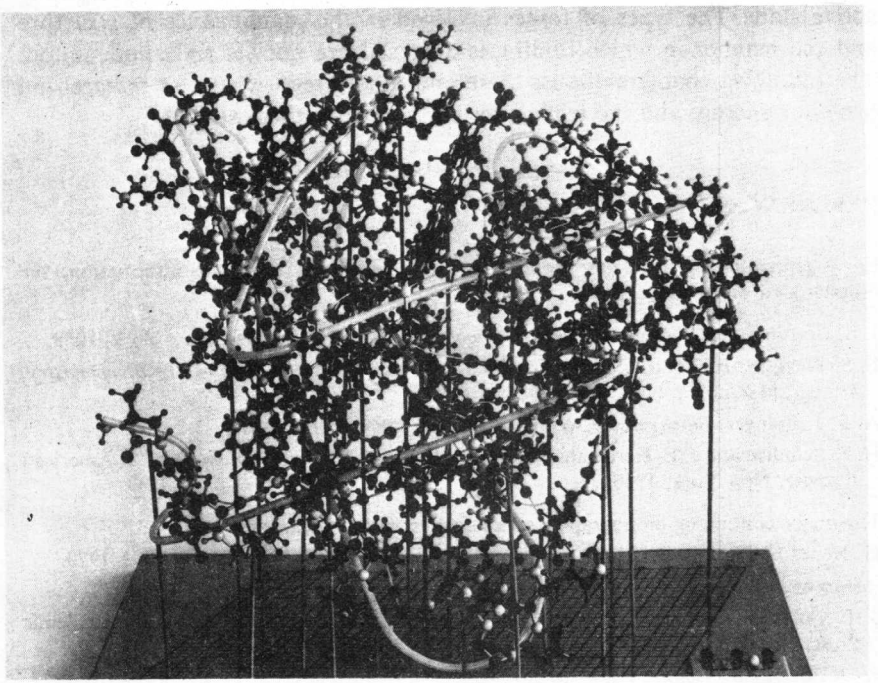

Fig. 1-3 Sperm whale myoglobin—a wire atomic model with a white cord marking the path of the α-helices. (Courtesy of J. C. Kendrew.)