The Proteins

CHEMISTRY, BIOLOGICAL ACTIVITY, AND METHODS

Edited by

HANS NEURATH
Department of Biochemistry
University of Washington
Seattle, Washington

KENNETH BAILEY
Department of Biochemistry
University of Cambridge
Cambridge, England

VOLUME I, PART B

ACADEMIC PRESS INC., PUBLISHERS
NEW YORK, 1953

Library of Congress Catalog Card Number: 52-13366

PRINTED IN THE UNITED STATES OF AMERICA

CONTRIBUTORS TO VOLUME I, PART B

JOHN T. EDSALL, *University Laboratory of Physical Chemistry Related to Medicine and Public Health, Harvard University, Boston, Massachusetts.*

IRVING M. KLOTZ, *Department of Chemistry, Northwestern University, Evanston, Illinois.*

R. R. PORTER, *National Institute for Medical Research, Mill Hill, London, England.*

FRANK W. PUTNAM, *Department of Biochemistry, University of Chicago, Chicago, Illinois.*

CONTENTS

CONTENTS OF VOLUME I, PART A

CHAPTER 7

The Size, Shape and Hydration of Protein Molecules

BY JOHN T. EDSALL

I. Introduction

Protein molecules are large and unstable; they are also extraordinarily reactive, and in extremely diverse ways. They form complexes, of vary-

ing stability, with almost all anions and cations, with lipides, with carbohydrates, and with one another. In addition to the conjugated proteins of well-defined character, such as the hemoglobins, there are innumerable proteins with a specific affinity for molecules with particular configurations, notably the enzymes with their capacity to react with their substrates and with other molecules containing the same or similar groupings in the correct arrangement. Some enzyme molecules are bound very tightly to particular coenzymes; others form very loose unstable enzyme–coenzyme complexes. Proteins are generally obtained for study either from actively metabolizing tissues or from media such as blood plasma, which are in a constant state of dynamic interchange with such tissues. The lifetime of a protein molecule in its natural surroundings is commonly short, often a matter only of days or even hours; some, however, such as collagen or mammalian hemoglobin, may persist for months, being in relatively isolated structures, set apart from the constant flux of dynamic interchange proceeding in tissues such as liver.

This chapter will deal with studies on purified proteins, separated by various fractionation procedures from the complex systems of which they form a part in nature. Such studies on separated proteins are fundamental to the progress of this field. However, the results so obtained must always be interpreted with due consideration of the natural origins of the separated protein molecules, and of their commonly great sensitivity to denaturation even when they are prepared and kept under conditions of maximum stability. When the physical chemist makes a series of measurements on the properties of a particular protein preparation, he may be studying a different kind of molecule in his last measurement from that which he studied in the first. Moreover, even the freshly prepared substance may be something quite different from anything that was present in the original tissue from which the protein preparation had been extracted.

These considerations should not discourage the investigator of proteins from proceeding with his work, but they should cause him to proceed in constant awareness of the delicacy and complexity of the structures with which he deals, of the imperfections and dangers of all fractionation procedures, and of the frequent need for revising his methods and his conclusions as better methods become available. The results presented in this chapter should be read in connection with Chap. 1 where methods of isolation are discussed in detail.

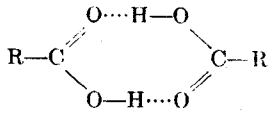
The values of molecular weight discussed in this chapter are commonly reliable to within 5 to 10 per cent; sometimes, in exceptionally favorable cases, to within 1 or 2 per cent. The absolute values of molecular dimensions, on the other hand, are as a rule considerably more uncertain.

Estimates of molecular shape are generally complicated by the difficulty of resolving effects due to solvation of the protein from those due to molecular asymmetry. In many cases, also, the data on asymmetry permit description of the molecule only in terms of an equivalent ellipsoid of revolution having approximately the same physical constants as the actual molecule. Electron micrographs provide more specific pictures, but they are taken of dried proteins, which cannot be generally identical in size and shape with the same molecules in aqueous solution. Thus our present conceptions of the shapes of protein molecules give only a rather broad fuzzy outline, like a photograph considerably out of focus; the finer details of the geometry of the molecule as yet escape us. Deeper insight is beginning to be obtained from x-ray diffraction studies on protein crystals; for the discussion of that topic, and for the detailed structure of some amino acids and peptides, the reader should turn to Chap. 4.

The studies to be presented here raise two important general questions, which deserve some discussion before the separate methods and data are considered. Just what is meant when we speak of a protein molecule? And how do we recognize a pure protein when we have prepared it, if indeed there is any such thing as a pure preparation?

1. What Is a Protein Molecule?

To many chemists the term "molecule" denotes a group of atoms linked together by covalent bonds. For the purposes of the protein chemist, such a definition is too restricted. Indeed even among very simple organic compounds, one finds structures which almost everyone would denote as molecules, but which do not obey this definition, for example, the dimers of formic acid and of the higher fatty acids.



The two single molecules are held together in this structure by hydrogen

(1) Some general references on the problem of protein size and shape may be noted here.[2-5b]

(2) E. J. Cohn and J. T. Edsall, Proteins, Amino Acids and Peptides, Reinhold Publishing Corporation, New York, 1943.

(3) J. L. Oncley, *Ann. N. Y. Acad. Sci.* **41**, 121 (1941).

(4) J. Wyman and E. N. Ingalls, *J. Biol. Chem.* **147**, 297 (1943).

(5) J. T. Edsall, *Fortschr. Chem. Forsch.* **1**, 119 (1949); a few portions of this article have, with the permission of the publishers (Springer-Verlag), been incorporated in some pages of the present chapter, especially in sec. XIX.

(5a) A. E. Alexander and P. Johnson, Colloid Science, Clarendon Press, Oxford, 1949.

(5b) C. Sadron, *Progress in Biophys. and Biophys. Chem.* (London and New York) **3**, 237 (1953).

bonds only, and to form two molecules of monomer from the dim requires the expenditure of only about 10 kcal. of energy per mole. Yet in spite of the relative instability of this structure, it involves a pattern with such a definite configuration that few chemists would hesitate to call it a molecule. The analogy with protein molecules is actually very close, since there is now strong evidence that the fundamental subunits of protein structure are largely held together by hydrogen bonds.

A comparison between the chemistry of proteins and of high polymers is perhaps instructive in this respect. The monomer units in synthetic high polymers are almost always linked together through covalent bonds. The same unit repeats itself throughout a long chain with a characteristic free terminal group at each end. To the high-polymer chemist, the molecule is, therefore, the structure which contains one such free terminal group at each end; and one of the most satisfactory methods of molecular weight determination for molecules that are not too large is to apply an analytical technique that is specific for the particular end groups of a particular polymer. Of course, aggregates of these primary molecular units can and frequently do form; but they are, in general, loose and unstable, of variable composition, and, therefore, without a precisely defined molecular weight. Thus, the polyamino acids prepared by Woodward and Schramm[6] were reported by them to give extremely high molecular weights of over a million in certain solvents, as deduced from osmotic pressure and viscosity measurements. However, studies in other solvents[7] have shown that these large structures are aggregates which can be dissociated into much smaller units by a suitable choice of solvent. The true average molecular weight was considered to be that found in the solvent giving the lowest value of all those tested, and proved to be of the order of 40,000.

On the other hand, the protein chemist takes a different attitude toward the molecule of such a substance as human or horse hemoglobin. Here the molecular weight of 67,000, as determined from measurements in water or dilute salt solutions, is taken to be a true unit, in spite of the fact that hemoglobin from several animal species dissociates into two half molecules, each with a molecular weight of 34,000, when dissolved in concentrated urea solution.[8,9] According to the criterion applied by most high-polymer chemists, the latter is the true molecule, and the larger unit is an aggregation product. However, the latter is universally

(6) R. B. Woodward and C. H. Schramm, *J. Am. Chem. Soc.* **69**, 1551 (1947).
(7) F. Eirich, E. Katchalski, J. Reisman, and P. Spitnik, unpublished results; for a review of the whole situation, see E. Katchalski, *Advances in Protein Chem.* **6**, 123 (1951) especially p. 166.
(8) N. F. Burk and D. M. Greenberg, *J. Biol. Chem.* **87**, 197 (1930).
(9) J. Steinhardt, *J. Biol. Chem.* **123**, 543 (1938).

regarded by protein chemists as the native hemoglobin molecule, even though it consists of two halves which can be split by relatively mild reagents. The smaller unit is, of course, also a definite molecule. This point of view is justified by various facts: for instance, that native hemoglobin forms very perfect crystals, showing a high degree of internal regularity in molecular structure, as judged by x-ray diffraction patterns; and also by the fact that in its combinations with other molecules, such as oxygen, it functions not as a loose aggregate but as a highly integrated structure, in which the four heme groups present in the 67,000 molecular-weight unit show strong interactions indicating that they are closely coupled.[10]

Analysis of free end groups gives no simple correlation here with other findings; horse and donkey hemoglobin are reported to contain six terminal amino groups of valine per molecule of 66,000 molecular weight; human hemoglobin to contain five valine amino groups; and cow, sheep, and goat to contain two valine and two methionine.[11] As yet these data are not easy to correlate with other known properties of hemoglobin; presumably they indicate subunits of peptide chains within the larger molecular framework, but the subunits have never been isolated as such, and it is not particularly helpful at present to think of them as molecules.

For the purposes of the discussion in this chapter, we shall generally consider a molecule as a molecular kinetic unit in solution—the unit which is observed to move in a diffusion or sedimentation experiment, and which by its presence lowers the activity of the solvent and thereby determines the osmotic pressure of the system. The significance of such measurements for protein chemistry, however, lies in the fact that the molecular kinetic unit is also generally a structural unit with a definite pattern. This is not true, for instance, of a particle of gold in a gold sol; such particles are fragments, derived from the breakdown of a larger structure, and the exact number of gold atoms in a particle is arbitrary, depending on the conditions of preparation. The nearest analog in protein chemistry is gelatin, which consists of a heterogeneous mixture of breakdown products derived from the hydrolysis of collagen. Studies of the molecular-weight distribution of the breakdown products,[12] and of the x-ray diffraction patterns given by the native collagen,[13] suggest the existence of rather compact structural units in collagen, which may be

(10) J. Wyman, *Advances in Protein Chem.* **4**, 407 (1948).

(11) R. R. Porter and F. Sanger, *Biochem. J.* **42**, 287 (1948); also *in* F. J. W. Roughton and J. C. Kendrew, eds., Haemoglobin, Butterworths Scientific Publications, London, 1949, p. 121.

(12) G. Scatchard, J. L. Oncley, J. W. Williams, and A. Brown, *J. Am. Chem. Soc.* **66**, 1980 (1944).

(13) R. S. Bear, *Advances in Protein Chem.* **7**, 69 (1952).

called molecules. There are similar indications from x-ray studies on keratin.[14] These postulated molecules are elements in a fibrous structure and have never been identified as such in solutions derived from the fibers. However, the indications of their existence serve to broaden the possible significance of the term molecule as applied to proteins and may lead in time to the isolation and characterization of these units. In this chapter, however, the discussion will be restricted to protein molecules which can be studied in solution.

2. Criteria of Purity of Protein Preparations

It is extremely difficult to demonstrate the purity of a protein preparation. Many criteria can be applied, and a truly pure protein must satisfy all of them. Thus, when studied by electrophoresis (Chap. 6), such a protein should not only migrate with a single boundary, but the degree of boundary spreading should be no greater than corresponds to the diffusion constant of the protein, as independently determined. Moreover this should be true over the entire pH range within which the protein is stable.[15] The same is true of sedimentation measurements in the ultracentrifuge and of the rate of free diffusion. Not only should there be only a single sedimenting boundary, but the form of this boundary should possess the symmetry characteristic of a single component. A similar statement holds for diffusion. Likewise, the study of rotary diffusion, discussed later in this chapter, provides additional evidence; measurements of double refraction of flow on elongated molecules, for example, may show the presence of several components of differing length. Osmotic-pressure measurements, taken alone, give only a number-average molecular weight; but if light-scattering measurements, which give a weight average, are made on the same preparation, agreement between the two results provides evidence of uniformity with respect to molecular weight.

Each of these methods gives evidence only with respect to certain gross properties of the molecules in the preparation—general size and shape, and net electric charge. All the molecules may be indistinguishable in these general properties, and yet may differ greatly among themselves in their amino acid composition and in the finer details of their structure. The solubility test of purity, discussed in Chap. 1, is in this respect a more searching criterion. Even this, however, is limited in its

(14) R. S. Bear and H. J. Rugo, *Ann. N. Y. Acad. Sci.* **55**, 627 (1950).

(15) It should be pointed out that the pH stability range of a protein is usually determined by just such criteria as uniformity of electrophoretic mobility and sedimentation constant. The important thing here is that there is a pH range of significant width over which these criteria of uniformity are satisfied.

sensitivity; several per cent of a minor component might well be present without being detected with assurance, owing to the inherent errors of solubility measurements and the difficulty of making determinations when the amount of solid phase in equilibrium with the solution is very small.

Recent advances in the detailed study of the sequence of amino acid residues in the polypeptide chains of proteins have provided another important criterion of uniformity. If two protein molecules are identical, the sequence of amino acid residues in all the peptide chains present in both proteins must be identical. This criterion has been applied to the B-chain of the insulin molecule, which contains a terminal phenylalanyl amino group, by Sanger and Tuppy;[16] and to the A-chain of insulin, which contains a terminal glycyl amino group, by Sanger and Thompson.[16a] All breakdown products derived from the partial hydrolysis of both chains were found to have structures compatible with a unique sequence of amino acid residues in each of the two original total chains. In this fundamental respect, therefore, the insulin molecules in the original preparation appeared to be all alike. No similar data of this sort for other proteins are yet available. Fundamental as such studies are, however, it must be remembered that the unique structure of a native protein depends not only on the sequence of amino acid residues in the peptide chains, but also on the geometrical configuration of the chains and the pattern in which they are folded. A given sequence is compatible with many such configurations, so that a collection of molecules uniform with respect to sequence could still differ markedly among themselves in detailed structure. A mixture of native and denatured hemoglobin or serum albumin molecules, for example, might well exemplify such a system, if denaturation simply involves a disarrangement of peptide chains from their native configuration, without any actual splitting of covalent bonds.

An analytical proof of *lack* of purity may often be derived from amino acid analysis, or from the determination of any specific reactive group in the molecule. If the molecular weight is known by any of the methods to be described in this chapter, and if accurate analysis shows that on the average less than one group of a particular type is present per molecule of protein, then it is clear that more than one type of molecule is present in the preparation. Thus the tryptophan content[17] of crystallized

(16) F. Sanger and H. Tuppy, *Biochem. J.* **49**, 463, 481 (1951).

(16*a*) F. Sanger and E. O. P. Thompson, *Biochem. J.* **53**, 353, 366 (1953); F. Sanger, E. O. P. Thompson, and H. Tuppy, *in* Symposium sur les Hormones Protéiques et Dérivées des Protéines, Paris, 1952.

(17) E. Brand, B. Kassell, and L. J. Saidel, *J. Clin. Invest.* **23**, 437 (1944).

human serum albumin was found to be only 0.6 residue per molecule of molecular weight 69,000. Even more conclusive was the finding of Hughes[18] and of Benesch and Benesch[19] that the content of titratable sulfhydryl groups in these serum albumin preparations was only 0.6–0.7 per mole. In this case the sulfhydryl-containing mercaptalbumin was crystallized by Hughes as a mercury derivative; the albumin molecules which lacked a free sulfhydryl group remained in the supernatant liquid. Thus the test which demonstrated the lack of homogeneity in the preparation also served as a clue to a method of separating the components.

In general, one may separate a protein preparation into two or more fractions by any method that does not denature the protein. If the protein is a single substance, all the fractions should be identical in amino acid content, in absorption spectrum, and in any specific activity that the preparation may possess, such as enzymatic or hormonal activity. Here again, as with other criteria discussed above, the evidence is essentially negative. The separation of two fractions of significantly different properties is a proof that the original preparation was not pure, but the failure to achieve such a separation is no sure test of purity. However, if several different methods of fractionation all fail to achieve separation, and if the protein meets the other tests of purity already described, few chemists would hesitate to pronounce it pure.

Immunological assay is often a powerful tool for the investigation of protein purity, and has been well discussed by Kabat.[20]

Recently the purity of insulin has been examined by the method of countercurrent distribution.[21] The preparations studied showed definite evidence of resolution into two components, after 900 transfers, involving distribution between 2-butanol and 1 per cent aqueous dichloroacetic acid; but the proportions of the two components differed considerably in different preparations. It will probably be difficult to extend this method to most other proteins, because of the difficulty of finding suitable solvents giving a suitable distribution coefficient and the danger of denaturing most proteins during the long series of equilibrations involved. However, the method is so powerful where it is applicable that further exploration is certainly called for.

In the case of some of the smaller proteins, at least, separation on an ion-exchange resin may be employed as an effective means of obtaining the desired component free from accompanying impurities, and also as a

(18) W. L. Hughes, Jr., *Cold Spring Harbor Symposia Quant. Biol.* **14**, 79 (1950).
(19) R. Benesch and R. E. Benesch, *Arch. Biochem.* **19**, 35 (1948).
(20) E. A. Kabat, *J. Immunol.* **47**, 513 (1943), especially pp. 520–2.
(21) E. J. Harfenist and L. C. Craig, *J. Am. Chem. Soc.* **73**, 877 (1951); **74**, 3083 (1952).

sensitive technique for examining the purity of the separated component. Preliminary studies of this sort have been reported on cytochrome c,[21a] and on ribonuclease and lysozyme,[21b] using the cation-exchange resin Amberlite IRC-50. All of these proteins have molecular weights below 18,000. It remains to be seen whether larger proteins can be purified by similar methods without denaturation.

Closely related to this procedure is the fractionation of proteins by partition chromatography. Martin and Porter[21c] fractionated ox pancreas ribonuclease, using a two-phase system composed of ammonium sulfate, water, and ethylene glycol monoethyl ether (cellosolve). At suitable compositions this separates into a lighter phase containing chiefly water and cellosolve, and a denser phase composed chiefly of water and ammonium sulfate. The supporting column of kieselguhr held the lighter (stationary) phase, and the aqueous salt solution, containing the ribonuclease, flowed down through this. A sharp resolution into two enzymatically active fractions was obtained. More recently Porter[21d] has explored other two-phase systems containing water, various glycol ethers, and concentrated phosphate buffers or other salts. Very homogeneous insulin preparations could be obtained by chromatographic fractionation in such systems, even when the starting material was quite impure. Promising results were obtained with several other proteins.

It must be admitted that few, if any, of the proteins known today meet all the complex and exacting tests of purity here briefly summarized. For instance, the studies of Alberty, Anderson, and Williams[22,23] have demonstrated electrophoretic heterogeneity in preparations of proteins which had earlier been reported as pure even by the very rigorous solubility test.[24] This evidence of imperfect purity in even the best available protein preparations should not deter the chemist from studying them as

(21*a*) S. Paléus and J. B. Neilands, *Acta Chem. Scand.* **4**, 1024 (1950).

(21*b*) C. H. W. Hirs, W. H. Stein, and S. Moore, *J. Am. Chem. Soc.* **73**, 1893 (1951). A full report of the work on ribonuclease is given by C. H. W. Hirs, S. Moore, and W. H. Stein, *J. Biol. Chem.* **200**, 493 (1953) and on lysozyme by H. H. Tallan and W. H. Stein, *ibid.* **200**, 507 (1953).

(21*c*) A. J. P. Martin and R. R. Porter, *Biochem. J.* **49**, 215 (1951).

(21*d*) R. R. Porter, *ibid.* **53**, 320 (1953).

(22) R. A. Alberty, E. A., Anderson, and J. W. Williams, *J. Phys. & Colloid Chem.* **52**, 217 (1948).

(23) E. A. Anderson and R. A. Alberty, *J. Phys. & Colloid. Chem.* **52**, 1345 (1948).

(24) It would be desirable, to make such tests even more illuminating, to carry out the electrophoretic studies, the solubility measurements, and as many other kinds of studies as possible, on the identical preparation of protein in the same laboratory at the same time. The instability of protein molecules in general, and the difficulty of reproducing different preparations exactly, would suggest the desirability of this procedure.

they can now be obtained. Progress will come by working with the best preparations available, bearing in mind their limitations as to purity, and constantly striving to develop better methods of purification, in conjunction with the characterization of the products. The proteins to be discussed in this chapter are for the most part preparations which are fairly homogeneous with respect to size and shape, and generally also with respect to electrophoretic mobility. Some have been found, at least in certain preparations, to satisfy the rigorous solubility test. Other well-known and widely studied preparations, like zein or the γ-globulins, are certainly not chemical individuals at all, but families of closely related proteins. The reader should bear these limitations in mind as the discussion proceeds.

II. The Binding of Water by Proteins

In most of this chapter we shall be concerned primarily with the properties of proteins in solution. The major component of the solvent is almost always water,[25] generally containing some dissolved ions and often various organic solutes. The molecular kinetic unit in solution is not the anhydrous protein molecule but the molecule with a considerable amount of water quite tightly bound. Often the binding of other solute molecules and ions by the protein is very strong; such interactions are discussed elsewhere in this volume by Klotz (Chap. 8), and will be mentioned only incidentally here. The binding of water by the protein, however, is of such major importance for the subject of this chapter that it requires more detailed consideration.

Several studies have now been made on the binding of water by dried proteins, as a function of the vapor pressure of water in the system.[26–30]

(25) The prolamines, such as zein, gliadin, and secalin, dissolve best in alcohol-water mixtures containing mostly alcohol, or in such organic solvents as propylene glycol. Insulin is also soluble in propylene glycol–water mixtures.

(26) H. B. Bull, *J. Am. Chem. Soc.* **66,** 1499 (1944).

(27) E. F. Mellon, A. H. Korn, and S. R. Hoover, *J. Am. Chem. Soc.* **69,** 827 (1947); **71,** 2761 (1949); E. F. Mellon, A. H. Korn, E. L. Kokes, and S. R. Hoover, *J. Am. Chem. Soc.* **73,** 1870 (1951).

(28) H. J. Frey and W. J. Moore, *J. Am. Chem. Soc.* **70,** 3644 (1948).

(29) S. W. Benson, D. A. Ellis, and R. W. Zwanzig, *J. Am. Chem. Soc.* **72,** 2102 (1950).

(30) F. Haurowitz, *J. Biol. Chem.* **193,** 443 (1951). This is a study of hemoglobin, and especially of the binding of water by the heme groups. The term "dried protein" here denotes material which has been carefully dried *in vacuo*, so that the amount of water which can still be removed by heating in an oven at 105° or above is considerably less than 1% of the weight of the protein. Air-dried proteins may still contain 8 or 10% water: this is, of course, the strongly bound water which is under discussion here.

Some typical curves, taken from the work of Bull, are shown in Fig. 1. For all proteins studied, there is a very rapid increase of water uptake as the vapor pressure of water rises from zero to about one tenth of that for pure water at the same temperature. This is followed by a plateau, with another steep rise as 100 per cent humidity is approached. The amount

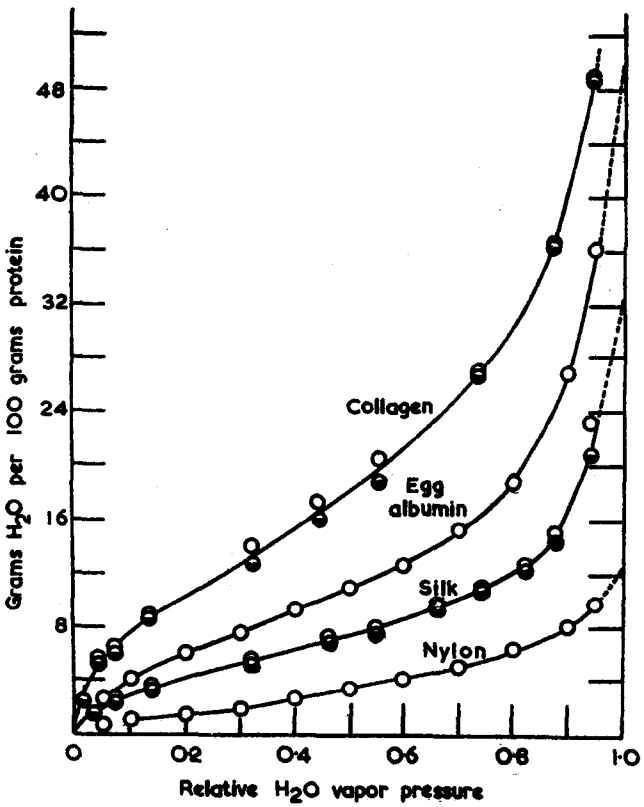


FIG. 1. Water adsorption curves for unstretched nylon, for wet (open circles) and dry (half circles) silk, for unlyophilized egg albumin, and for wet (open circles) and dry (half circles) collagen at 25°. From Bull.[26]

bound in the first steep portion of the curve varies for different proteins from about 4 to 10 g. per 100 g. dry protein. It is considerably less than the amount required to form a monolayer of water over the surface of the protein molecule. Pauling[31] has pointed out a correlation between the amount of water which is strongly bound and the number of ionic and polar groups in the amino acid side chains of the protein. Pauling con-

(31) L. Pauling, *J. Am. Chem. Soc.* **67**, 555 (1945).