JOURNAL
OF THE AMERICAN CHEMICAL SOCIETY

VOLUME 89, NUMBER 4 FEBRUARY 15, 1967

Physical and Inorganic Chemistry

Proteins as Random Coils. I. Intrinsic Viscosities and Sedimentation Coefficients in Concentrated Guanidine Hydrochloride[1a]

Charles Tanford, Kazuo Kawahara,[1b] and Savo Lapanje[1c]

Contribution from the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27706. *Received August 22, 1966*

Abstract: Measurements have been made of the intrinsic viscosities and sedimentation coefficients of protein polypeptide chains in concentrated guanidine hydrochloride, in the presence of β-mercaptoethanol to rupture disulfide bonds where they exist. The results show that both properties depend on molecular weight exactly as predicted for randomly coiled linear polymer chains. It is therefore concluded that protein polypeptide chains, in the solvent medium employed, are true random coils, retaining no elements of their original native conformation. Intrinsic viscosities of proteins which possess disulfide bonds were also measured in concentrated guanidine hydrochloride in the absence of β-mercaptoethanol. They were found to be substantially smaller than those obtained in the presence of the reducing agent. It is likely that this reflects solely the physical restrictions imposed by the disulfide cross links on an otherwise randomly coiled chain, but exact theoretical equations by which to prove this suggestion are not available.

Native protein molecules are known to be folded into well-defined, usually essentially rigid three-dimensional structures. For most proteins this structure is compact and globular, as exemplified by the known structures of myoglobin[2] and lysozyme.[3] In a few proteins the native structure is rod-like, or it consists of a rod-like structure with globular appendages. Myosin[4,5] is an example of the latter.

The native structure of a protein is usually stable over a fairly wide range of external conditions, but it can be disrupted by sufficiently drastic changes in physical or chemical environment. This process is known as *denaturation*. In many instances it is a reversible process, showing that the structural changes which occur represent changes in the thermodynamic stabilities of various possible conformations of the protein molecule.

As we have stated, denaturation can be brought about both by physical and chemical changes in the environment, and many studies of the process, using a variety of denaturing agents, have been reported in recent years.[6] Some of these are quite detailed investigations in which both the thermodynamics and kinetics of the process, native state ⇌ denatured state, have been studied.[7-10] However, relatively little work has been done to define

(1) (a) This work was supported by grants from the National Science Foundation and from the National Institutes of Health, United States Public Health Service. (b) On leave of absence from the Department of Polymer Science, Osaka University, Osaka, Japan. (c) On leave of absence from the Department of Chemistry, University of Ljubljana, Ljubljana, Yugoslavia.

(2) J. C. Kendrew, H. C. Watson, B. E. Strandberg, R. E. Dickerson, D. C. Phillips, and V. C. Shore, *Nature*, **190**, 666 (1961).

(3) C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarima, *ibid.*, **206**, 757 (1965).

(4) A. Holtzer, S. Lowey, and T. M. Schuster, "The Molecular Basis of Neoplasia," University of Texas Press, Austin, Texas, 1962, p 259.

(5) E. F. Woods, S. Himmelfarb, and W. F. Harrington, *J. Biol. Chem.*, **238**, 2374 (1963).

(6) For a recent review of this field see P. Joly, "Physico-Chemical Basis for Protein Denaturation," Academic Press Inc., New York, N. Y., 1966.

(7) J. Hermans, Jr., and H. A. Scheraga, *J. Am. Chem. Soc.*, **83**, 3283 (1961); R. A. Scott and H. A. Scheraga, *ibid.*, **85**, 3866 (1963).

(8) A. J. Sophianopoulos and B. J. Weiss, *Biochemistry*, **3**, 1920 (1964).

(9) C. Tanford, R. H. Pain, and N. S. Otchin, *J. Mol. Biol.*, **15**, 489 (1966).

(10) J. F. Brandts, *J. Am. Chem. Soc.*, **86**, 4291 (1964).

the nature of the product of this reaction. Is the entire native structure disrupted, or only a part of it? Do the parts of the molecule which are disrupted refold to a new rigid structure, or do they remain essentially structureless? The answers to these questions are generally not known with any precision. It is evident from the available data, however, that the answers to these questions will not be unique. The data suggest, on the contrary, that different denatured states are produced by different classes of denaturing agents. It has been shown, for example, that the conformation attained in urea and formamide is very different from that produced by ethanol and dioxane.[11] It has been pointed out that detergents and urea lead to different denatured states.[12] Many denatured states must retain substantial elements of the native structure. Acid-denatured serum albumin, for example, retains a large proportion of tyrosine residues inaccessible to the effects of solvent perturbation.[13]

It is clear that an interpretation of thermodynamic and kinetic data for denaturation is possible only if the denatured state can be precisely defined. Two recently proposed theoretical treatments[14-16] of thermodynamic data for denaturation, for example, can be applied in their simplest forms only if the denatured state is essentially devoid of secondary and tertiary structure, *i.e.*, if it is a "random coil," in the sense in which this term is normally used by polymer chemists.[17]

With the foregoing questions in mind, an investigation of the denatured state of proteins has been initiated in this laboratory. The first objective has been to discover whether any denaturing agent leads to the formation of the simplest of all possible denatured states, the random coil. It is by no means certain that proteins can ever attain such a state, for random coils are typically the result of dissolving polymer molecules in a good or indifferent solvent, in which the attractive forces between the polymer segments and the solvent are stronger than, or at least as strong as, the attractive forces between one polymer segment and another. To discover such a solvent for proteins must be inherently difficult, because of the variety of chemical groups which a protein molecule contains. Thus a good solvent for peptide or amide groups is likely to be a poor solvent for hydrophobic groups, and *vice versa*. Any solvent with a low dielectric constant can be expected to be a poor solvent for ionic groups.

Among solvent media commonly employed for denaturation of proteins, concentrated aqueous solutions of guanidine hydrochloride (GuHCl) generally produce larger changes in physical and chemical properties than occur in other denaturing media (urea may be equally effective for many proteins). We have therefore chosen aqueous solutions of GuHCl as the solvent medium for this initial study. In addition, we have introduced a moderate concentration of a reducing agent, β-mercaptoethanol, so as to rupture disulfide bonds and to prevent their formation by oxidation of thiol groups when they are not present in the native protein. This ensures that the primary bonds of the molecules form no cross links, so that the resulting conformation, if it is indeed a random coil, will obey relations applicable to linear polymer chains. (Some experiments showing the effects of leaving disulfide bonds intact will be described toward the end of this paper.)

Hydrodynamic properties of dissolved macromolecules are especially sensitive to conformation. Moreover, the theory of the hydrodynamic behavior of randomly coiled linear polymers is well developed[17-19] and has been thoroughly tested on solutions of linear synthetic polymers. For this reason hydrodynamic measurements were thought to provide the most critical initial test for random-coil behavior. The present paper will accordingly present the results of measurement of the viscosity and sedimentation of proteins in concentrated GuHCl solutions. Subsequent papers will consider optical rotatory dispersion, osmotic pressure, titration studies, and other aspects of the behavior of proteins under these conditions. It is hoped that all such information, taken together, will permit an unequivocal characterization of the conformation, *i.e.*, whether it is a true random coil, and will detect appreciable structured regions if such still remain.

A short preliminary account of most of the viscosity data of this paper has been published previously.[20]

Experimental Section

Proteins. The insulin used in this study was beef insulin, five times recrystallized, lot No. T-2842, kindly donated by the Eli Lilly Co., Indianapolis, Ind. Ribonuclease was purchased from Wilson Laboratories. Myoglobin was a salt-free crystalline preparation from Mann Research Laboratories. It was converted to the CO derivative by reduction with $Na_2S_2O_4$ and dialysis against CO-saturated water. β-Lactoglobulin (type A) was donated by Dr. R. Townend, of the Eastern Utilization Research and Development Division, U. S. Department of Agriculture. The protein was recrystallized before use. α-Chymotrypsinogen A was obtained from Sigma Chemical Corp. Pepsinogen was a chromatographically purified sample from Worthington Biochemical Corp. Bovine serum albumin was a crystalline product from International Chemical and Nuclear Corp. A sample from Pentex, Inc., was used for some of the measurements. The aldolase and hemoglobin preparations have been described previously.[21,22]

Other Reagents. The preparation and purification of guanidine hydrochloride have been described previously.[23,24] Other reagents used were the best available commercial products.

Preparation of Solutions. Insulin was spread in a thin layer in a shallow dish, which was covered with a film of paraffin, and allowed to stand in the refrigerator for 2 days, to equilibrate with respect to moisture content. Moisture content was then determined by heating the crystals to 107°. Solutions for measurement were prepared by adding weighed amounts of the crystals to aqueous solutions of guanidine hydrochloride. Reference solvents were obtained by adding to similar solutions the calculated amounts of water contained in the insulin crystals.

(11) C. Tanford and P. K. De, *J. Biol. Chem.*, **236**, 1711 (1961).
(12) M. L. Meyer and W. Kauzmann, *Arch. Biochem. Biophys.*, **99**, 348 (1962). See also B. Jirgensons, *J. Biol. Chem.*, **241**, 147 (1966).
(13) M. Laskowski, Jr., *Federation Proc.*, **25**, 20 (1966).
(14) C. Tanford, *J. Am. Chem. Soc.*, **86**, 2050 (1964).
(15) J. F. Brandts, *ibid.*, **86**, 4302 (1964).
(16) See also H. A. Scheraga in "The Proteins," Vol. 1, H. Neurath, Ed., Academic Press Inc., New York, N. Y., 1963, p 478.
(17) P. J. Flory, "Principles of Polymer Chemistry," Cornell University, Press, Ithaca, N. Y., 1953.
(18) C. Tanford, "Physical Chemistry of Macromolecules," John Wiley and Sons, Inc., New York, N. Y., 1961.
(19) M. Kurata and W. A. Stockmayer, *Fortschr. Hochpolymer.-Forsch.*, **3**, 196 (1963).
(20) C. Tanford, K. Kawahara, and S. Lapanje, *J. Biol. Chem.*, **241**, 1921 (1966).
(21) K. Kawahara and C. Tanford, *Biochemistry*, **5**, 1578 (1966).
(22) K. Kawahara, A. G. Kirshner, and C. Tanford, *ibid.*, **4**, 1203 (1965).
(23) K. Kawahara and C. Tanford, *J. Biol. Chem.*, **241**, 3228 (1966).
(24) Y. Nozaki and C. Tanford, *J. Am. Chem. Soc.*, **89**, 736 (1967).

For all the other proteins used in this study, relatively concentrated stock solutions were prepared, the solvent being water or a dilute solution of NaCl. Traces of insoluble material were removed by centrifugation or filtration. The stock solutions were dialyzed against suitable solvent mixtures: usually water or dilute aqueous NaCl, adjusted with HCl or KOH to give a pH near the isoionic pH of the protein. For pepsinogen, the solvent contained 0.1 *M* Tris buffer, pH 7.5. For ribonuclease, passage through a Dintzis deionizing column[25] was used instead of dialysis. Protein concentrations of stock solutions were obtained from dry weight content, measured by heating aliquots of both the protein solution and the dialysate to 107°, obtaining the protein content by appropriate subtraction.

Solutions for measurement were obtained by weighing out appropriate amounts of protein stock solutions and stock solutions of GuHCl and β-mercaptoethanol. Reference solvents were obtained by using the dialysate described in the preceding paragraph in place of the protein solution. Amounts added were adjusted so that the protein solutions had exactly the same compositions as the corresponding reference solvents, except for the dry weight of the protein which they contained.

The densities of all solutions were calculated from the composition, using the density data reported earlier[23] and assuming additivity for components other than GuHCl in the solutions. The small effect which the protein content has on density was calculated by assuming an approximate value for the partial specific volume.

Viscosity. Viscosity measurements were made in Cannon-Fenske capillary viscometers, as described previously.[26] On some occasions a dilution-type viscometer was used instead.

Sedimentation Velocity. Sedimentation velocities were measured in a Spinco Model E analytical ultracentrifuge, generally using a synthetic boundary cell, at a rotor speed of 42,040 rpm. This type of cell was preferred for two reasons. (1) A double-sectored cell was necessary in any event because the protein refractive index gradient had to be corrected for the refractive index gradient due to sedimentation of GuHCl. (2) The time required for separation of the protein peak from the upper meniscus in a standard cell would have been excessively long because of the slow rate of sedimentation. The peak would have become very diffuse before measurements of peak positions could be made.

Measurements were made using Schlieren optics. Plates were read (with extreme care) in a Gaertner two-dimensional microcomparator.

Temperature of Measurements. All measurements were made at 25.0°.

Results

The concentration of GuHCl used for almost all of the experiments of this paper was 6 *M*. This is a sufficiently high concentration to assure that the transition to the denatured state is complete in all the proteins we have used. The concentration of added β-mercaptoethanol was usually 0.1 *M*. We shall use the abbreviation RSH for β-mercaptoethanol, the nonspecific symbol "R" indicating that a variety of thiol reagents would probably have served equally well. We shall use the symbol GuHCl/RSH to designate the normal solvent mixture of approximately 6 *M* GuHCl and 0.1 *M* RSH (plus low concentrations of NaCl or Tris buffer, when these were present in the protein stock solutions), and specify concentrations of these reagents only when they differed appreciably from these values.

Time Dependency. The unfolding of proteins by GuHCl and the rupture of disulfide bonds by RSH are not instantaneous reactions. Time studies were carried out to determine the time required for the viscosity to attain its final value for each protein. Whenever the pH of the GuHCl/RSH medium was near pH 6 or above, the time required was found to be less than 1 hr. The pH of the medium was well below 6 in the experiments with β-lactoglobulin and serum albumin, and the time required for completion of the unfolding process for these proteins was much longer: about 36 hr for β-lactoglobulin (pH 4.7) and about 12 hr for serum albumin (pH 5.4). These longer times are presumably to be ascribed to the slowness of the reduction of disulfide bonds at the low pH.

An interesting anomaly was observed for solutions of serum albumin in GuHCl/RSH. The viscosity of such solutions, after rising sharply during the unfolding process, decreased slowly with time, at an approximately constant rate, for as long as it was followed, 10–20 days. The observed decrease was too small to have a significant effect on the value chosen for the intrinsic viscosity, but the change was far outside the experimental error. No explanation for this decrease is known.

Viscosities of all other proteins in GuHCl/RSH remained perfectly constant over a period of several days.

Molecular Weights. Since we are anticipating that GuHCl/RSH will come close to rupturing all noncovalent intramolecular bonds, as well as all disulfide bonds, this solvent should dissociate all proteins to their constituent polypeptide chains. To verify this prediction, molecular weights of most of the proteins in GuHCl/RSH solution were determined, using either sedimentation equilibrium or combination of the intrinsic viscosity with the sedimentation coefficient, measurements of both these quantities being reported below. The measurements for aldolase have been described in detail,[21] because we obtained absolute molecular weight values for this protein which differed somewhat from previously measured values. Measurements for other proteins were made by similar procedures, but generally with less care, because weights of the polypeptide chains of these proteins are already known, and approximate values sufficed to demonstrate that complete dissociation had occurred. Within experimental error, the expected values were obtained in each instance.

Intrinsic Viscosity. The experimental data leading to the determination of intrinsic viscosities are shown in Figure 1. The figure includes data measured in the absence of RSH (to be discussed below) in addition to the data measured in GuHCl/RSH. The data were fitted by least squares to the expression

$$\eta_{sp}/c = [\eta] + k[\eta]^2 c \qquad (1)$$

in which $[\eta]$ is the intrinsic viscosity and k is a dimensionless constant known as the Huggins constant. The concentration units are g/cc. The values of $[\eta]$ and k obtained from these data are summarized in Table I. As is evident from Figure 1, the $[\eta]$ values obtained from the data are quite precise, but the slopes, *i.e.*, the values of the Huggins constants, are subject to considerable experimental uncertainty, especially for the proteins of lowest molecular weight.

Figure 2 shows a logarithmic plot of the viscosities of Table I *vs.* the number (n) of monomer units per polypeptide chain. It is seen that there is essentially a linear relationship between log $[\eta]$ and log n. The straight line shown in the figure was determined by the method of least squares and corresponds to the relation

$$[\eta] = 0.716n^{0.66} \qquad (2)$$

(25) H. M. Dintzis, Ph.D. Thesis, Harvard University, 1952.
(26) C. Tanford and J. G. Buzzell, *J. Phys. Chem.*, **60**, 225 (1956).

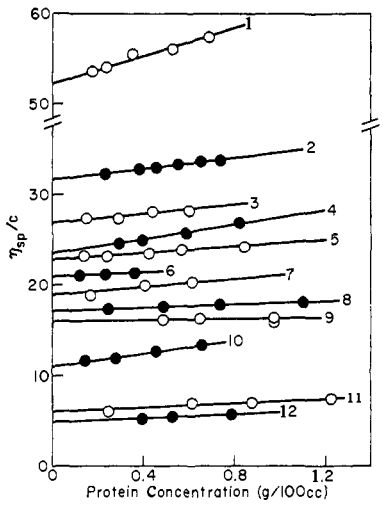

Figure 1. Viscosity data in GuHCl/RSH at 25°: 1, serum albumin, pH 5.3; 2, pepsinogen, pH 7.5; 3, chymotrypsinogen, pH 6.1; 4, serum albumin, RSH omitted, pH 5.4; 5, β-lactoglobulin, pH 4.6; 6, myoglobin, pH 7; 7, hemoglobin, pH 7; 8, β-lactoglobulin, RSH omitted, pH 5; 9, ribonuclease, pH 6.7; 10, chymotrypsinogen, RSH omitted, pH 6.1; 11, insulin, 0.2 *M* RSH, pH 5.5; 12, insulin, RSH omitted, pH 5.7. The data for aldolase have been presented previously.[21]

Equation 2 differs slightly from the corresponding equation given in the preliminary communication[20] of these results because an additional protein has been included in the analysis. There is actually an uncertainty of about 0.04 in the exponent of n, due to the scatter of the experimental points.

Table I. Intrinsic Viscosities of Polypeptide Chains in 6 *M* GuHCl, 0.1 *M* RSH, at 25°

Protein	Mol wt[a]	Residues per chain[a]	$[\eta]$, cc/g	Huggins constant, k, eq 1
Insulin	2,970	26	6.1	
Ribonuclease	13,680	124	16.0[e]	0.16[e]
Hemoglobin	15,500	144	18.9	0.59
Myoglobin	17,200	153	20.9	0.29
β-Lactoglobulin	18,400	162	22.8	0.38
Chymotrypsinogen	25,700	245	26.8	0.35
Glyceraldehyde-3-phosphate dehydrogenase[b]	36,300	331	34.5	0.67
Pepsinogen	40,000	365	31.5	0.34
Aldolase	40,000	365	35.3	0.35
Serum albumin	69,000	627	52.2	0.29
Thyroglobulin[c]	165,000	1500	82	0.50
Myosin[d]	197,000	1790	92.6	0.50

[a] For multichain proteins, the viscosity averages are used, as defined by eq 3. [b] Data in 5 *M* GuHCl, ref 27. [c] Data in 5 *M* GuHCl, ref 28. [d] Data in 5 *M* GuHCl, ref 5. See text for possible alternate values of the molecular weight. [e] More recent data with a ribonuclease sample of higher purity gave $[\eta]$ = 16.6 cc/g and k = 0.95. The difference in k shows the large experimental error in the determination of this parameter for the low molecular weight proteins.

Table I and Figure 2 contain, in addition to our data, similar data which have been obtained in other laboratories for glyceraldehyde 3-phosphate dehydrogenase,[27] thyroglobulin,[28] and myosin.[5] These proteins were studied in 5 *M* rather than 6 *M* GuHCl, and thiol groups were protected by reaction with N-ethylmaleimide rather than by the method used in the present study. These modifications in procedure should have no significant effect on the intrinsic viscosity if the proteins, with disulfide bridges broken, are random coils in both 5 and 6 *M* GuHCl. The points corresponding to these three proteins do indeed fall quite close to the straight line of Figure 2, which was determined on the basis of our own data alone.

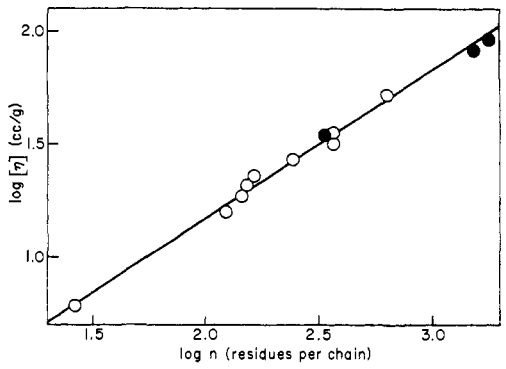

Figure 2. Intrinsic viscosity as a function of chain length. Filled circles represent data in 5 *M* GuHCl. The straight line represents eq 2.

It may be noted that the value of n assigned to myosin in Table I and Figure 2 is based on the model of Woods, *et al.*,[5] which assigns a molecular weight of about 600,000 to native myosin, and a molecular weight of about 200,000 to each of three equivalent polypeptide chains. Significantly lower molecular weights for the native protein have been reported by other workers,[4,29] and the model has also been questioned on other grounds.[30] A recent determination of the molecular weights of the polypeptide chains[31] suggests (with a native molecular weight of 520,000) that myosin may have four polypeptide chains, two with $n \simeq 420$ and two with $n \simeq 1950$. The appropriate average value of n to use in Figure 2 is given by[18]

$$\langle n \rangle^{0.66} = (n_1^{1.66} + n_2^{1.66})/(n_1 + n_2) \qquad (3)$$

with n_1 and n_2 being the lengths of two kinds of chains, present in equimolar amounts. With the values of n_1 and n_2 just given, $\langle n \rangle$ = 1625. If this value is used in place of n = 1790, the point for myosin falls exactly on the straight line of Figure 2, rather than somewhat below it, as now shown. (This is not intended as evidence for or against any of the myosin models, as exact adherence to eq 2 is not be expected, as will be discussed later.)

Sedimentation Velocity. It proved to be considerably more difficult to obtain reliable values for the sedimentation coefficient, s^0, at zero protein concentration, than for the intrinsic viscosity. There are two reasons for this, one being that the high density of the solvent and the large frictional coefficients of the protein molecules under these conditions lead to very low rates of sedimentation. The problem is aggravated by the fact that a relatively low rotor speed had to be employed because of the necessity of using synthetic boundary

(27) W. F. Harrington and G. M. Karr, *J. Mol. Biol.*, **13**, 885 (1965).
(28) B. De Crombrugghe, R. Pitt-Rivers, and H. Edelhoch, *J. Biol. Chem.*, **241**, 2766 (1966).
(29) H. Mueller, *ibid.*, **239**, 797 (1964).
(30) H. Mueller, *ibid.*, **240**, 3816 (1965).
(31) P. Dreizen, D. J. Hartshorne, and A. Stracher, *ibid.*, **241**, 443 (1966).

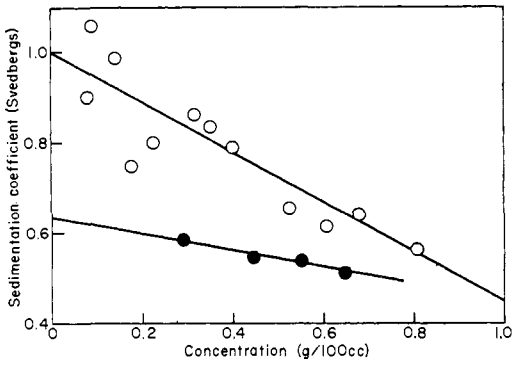

Figure 3. Representative sedimentation data. The upper curve shows data for serum albumin, the lower curve for chymotrypsinogen. The figure illustrates the poor precision of the data, which is partly responsible for the large probable error in the extrapolated s^0 values shown in Figure 5.

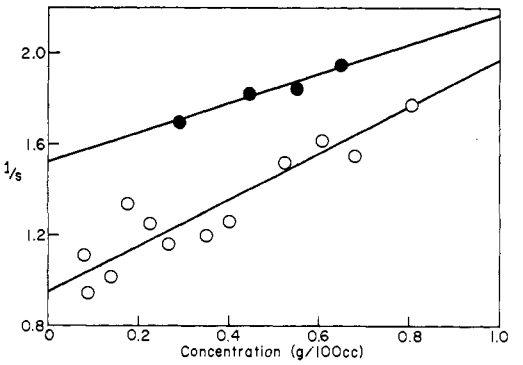

Figure 4. The data of Figure 3, plotted as $1/s$ *vs.* concentration.

cells for the measurements (see Experimental Section). Thus measured s values were less reproducible and had an over-all precision considerably lower than one is accustomed to expect from sedimentation studies of proteins in their native state.

A second reason for relatively poor accuracy in the values of s^0 lies in the very marked concentration dependence of measured s values. The effect of concentration on the measurements is here considerably greater than in the case of η_{sp}/c. Moreover, s values diminish with increase in concentration, so that the advantage gained by the larger refractive index gradient at high concentration is cancelled by the correspondingly slower movement of the refractive index gradient peak.

Because of the poor precision of the data, we have extrapolated the experimental sedimentation coefficients to zero protein concentration both by plotting s *vs.* c and by plotting $1/s$ *vs.* c, and averaged the result. Although a plot of $1/s$ *vs.* c is theoretically more justifiable, such a plot tends to magnify the errors inherent in the measurement of the very small s values which one obtains at high protein concentration. Typical data are shown in Figures 3 and 4.

In addition to the s^0 values which were measured by extrapolation of data at several protein concentrations to zero concentration, we have included in the analysis some estimated s^0 values based on a single measured s value or on a handful of scattered values. The estimates were based on estimates of the concentration dependence of s values derived from Figures 3 and 4.

All sedimentation coefficients obtained in this way are summarized in Table II, which also contains similar data for glyceraldehyde-3-phosphate dehydrogenase and myosin, determined by Harrington and coworkers.[5,27] Their data were determined in 5 M guanidine hydrochloride and have been corrected to the density and viscosity of the GuHCl/RSH mixture used in the present study.

Table II shows that the s^0 values, in contrast to the viscosities of Table I, do not vary in a systematic fashion with chain length. This is the expected result because the sedimentation velocity depends on the partial specific volume of the protein, and on preferential interaction with solvent components, as well as on chain length, *i.e.*, in a multicomponent system

$$s^0 = M(1 - \phi'\rho)/Nf \qquad (4)$$

where M is the molecular weight of the dry protein molecule, f the frictional coefficient of the solvated particle, ρ the solvent density, N Avogadro's number, and ϕ' an effective specific volume, which includes both the influence of the true partial specific volume of the protein and the influence of preferential interaction with individual solvent components. (This parameter has been discussed at some length by Eisenberg[32,33] and by Woods, *et al.*[5]) Thus the quantity which should vary systematically with molecular weight or chain length is $s^0/(1 - \phi'\rho)$.

Table II. Sedimentation Coefficients (s^0) in 6 M GuHCl, 0.1 M RSH, at 25°

Protein	Mol wt	Residues per chain	Apparent spec vol (ϕ'), cc/g	s^0_{25}, S
Ribonuclease	13,680	124	0.69	0.59[d]
Hemoglobin	15,500	144	0.74	0.45[d]
Myoglobin	17,200	153	0.74	0.46[d]
β-Lactoglobulin	18,400	162	0.74	0.61
Immunoglobulin[a] (fragment I)	25,000	230	0.73	0.57[d]
Chymotrypsinogen	25,700	245	0.71	0.65
Glyceraldehyde 3-phosphate dehydrogenase[b]	36,300	331	0.73	0.76
Pepsinogen	40,000	365	0.74	0.76
Aldolase	40,000	365	0.74	0.725
Immunoglobulin[a]	40,000	372	0.73	0.74[d]
Serum albumin	69,000	627	0.725	1.03
Myosin[c]	197,000	1790	0.71	1.88

[a] Reference 35. [b] Data in 5 M GuHCl, ref 27, corrected to the viscosity and density of 6 M GuHCl, 0.1 M RSH. [c] Data in 5 M GuHCl, ref 5, corrected to the viscosity and density of 6 M GuHCl, 0.1 M RSH. [d] Estimated s^0 values based on single or scattered measurements at finite concentrations.

An additional uncertainty is introduced thereby into the interpretation of the sedimentation data. There is considerable uncertainty concerning the values of partial specific volumes of proteins in GuHCl solutions,[34] and only limited knowledge exists on the subject of preferential interactions in this solvent. Values of ϕ' have been determined experimentally for γ-globulin[35]

(32) H. Eisenberg, *J. Chem. Phys.*, **36**, 1837 (1962).
(33) E. F. Casassa and H. Eisenberg, *Advan. Protein Chem.*, **19**, 287 (1964).
(34) F. J. Reithel and J. D. Sakura, *J. Phys. Chem.*, **67**, 2497 (1963).
(35) E. Marler, C. A. Nelson, and C. Tanford, *Biochemistry*, **3**, 279 (1964).

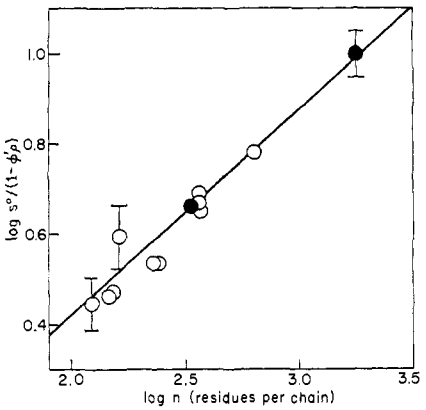

Figure 5. Sedimentation coefficients as a function of chain length. Filled circles represent data in 5 *M* GuHCl. Error bars shown for three of the points represent the probable uncertainty which must be ascribed to all of the data. The straight line represents eq 5.

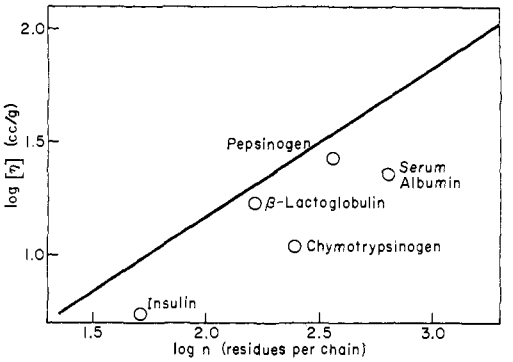

Figure 6. Effect of disulfide bonds on the intrinsic viscosity. The data were obtained in 6 *M* GuHCl, in the absence of RSH. The straight line is taken from Figure 2.

and myosin[5,36] and have been found to be of the order of 0.01 cc/g below the true partial specific volumes in dilute aqueous salt solutions. The ϕ' values shown in Table II have been obtained on this basis, and they must be considered to have an uncertainty of at least 0.01 cc/g. Because of the high solvent density, this introduces an uncertainty of nearly 10% into the values of $s^0/(1 - \phi'\rho)$.

A logarithmic plot of $s^0/(1 - \phi'\rho)$ *vs.* chain length is shown in Figure 5. The error bars shown for three of the points indicate the expected uncertainty due to a likely error of 0.05 in the determination of s^0, plus the uncertainty due to uncertainty in the value of ϕ'. A similar uncertainty applies to all of the data. We do not know why one point (that for β-lactoglobulin) falls so far out of line from the rest of the data.

The straight line of Figure 5 was determined by the method of least squares. It corresponds to the relation (with s in Svedberg units)

$$s^0/(1 - \phi'\rho) = 0.286n^{0.473} \quad (5)$$

Results in the Absence of RSH. When RSH is omitted from the solvent medium, disulfide bonds, in proteins which contain them, remain intact. This hampers the ability of the molecule to unfold and leads to substantially reduced values for the intrinsic viscosity. Data for a few proteins are shown in Figure 6. A corresponding *increase* in the value of s^0 is observed in sedimentation analysis, reflecting a decrease in the frictional coefficient. For serum albumin, for example, s^0 is 1.4 S, compared to the value of 1.03 S observed in the presence of RSH.

An additional effect of the absence of RSH from the solvent medium is that a definite time dependency of the results is obtained for all proteins which contain cysteine side chains. This is presumably a reflection of the reactivity of exposed SH groups on the protein, which can lead to formation of new disulfide bonds, either by oxidation or by interchange with existing disulfide bonds. We have already noted previously[22] an increase with time in the sedimentation coefficient of hemoglobin under these conditions, probably reflecting aggregation through new SS bridges. In the present investigation we have noted a decrease with time in the intrinsic viscosity of aldolase chains, suggesting that intrachain disulfide bonds are being formed, but an increase with time in the viscosity of serum albumin, suggesting formation of new interchain disulfide bonds by this protein's single SH group. It should be noted that the observed changes were quite slow and did not prevent an assignment of a characteristic value of $[\eta]$ for the proteins on initial solution in 6 *M* GuHCl. For serum albumin, for example, $[\eta]$ increased from 26.0 cc/g to only 30.3 cc/g in 10 days.

Discussion

Intrinsic viscosity is a measure of the effective specific volume (in units of cc/g) of the domain of a macromolecule in solution. Its dependency on the length of homologous polymer chains provides a sensitive criterion for the gross conformation of macromolecules.[18] For solid spheres which do not interact with the solvent, $[\eta]$ is very small and independent of molecular weight.[37] For long rods of constant diameter, $[\eta]$ is very large and varies roughly as (molecular weight).[1,8] For random coils in a thermodynamically-ideal solvent

$$[\eta] = Kn^{1/2} \quad (6)$$

where K is a constant which depends on the molecular weight (M_0) per monomer unit, on the bond lengths and bond angles of the links of the coiled chain, and on the freedom of rotation of the links with respect to each other, but is independent of the length of the chain, at least for long chains.

The parameter K of eq 6 would be a true constant for a series of homopolymers of varying chain length, but it can be expected to be only approximately constant for a series of randomly coiled proteins, which vary in amino acid content. The value of M_0, averaged over the protein molecule, will vary a little from protein to protein. More importantly, the side chain influences the freedom of rotation of one amino acid residue with respect to its neighbors.[38] Thus eq 6 can be expected to be obeyed only approximately by a series of randomly coiled proteins of different chain lengths. The deviations from the equation should however depend on amino acid composition, and not on chain length, so that the dependence of $[\eta]$ on $n^{1/2}$ should be closely maintained.

(36) W. W. Kielley and W. F. Harrington, *Biochim. Biophys. Acta*, **41**, 401 (1960).

(37) A. Einstein, *Ann. Physik.*, **19**, 289 (1906); **34**, 591 (1911).

(38) J. A. Schellman and C. Schellman, *Proteins*, **2**, 1 (1964).

The effect of thermodynamic nonideality, in a good solvent, is to expand the molecular domain[17,19] and, hence, the intrinsic viscosity. The effect may be expressed in terms of a linear expansion factor, α. The viscosity in a good solvent will exceed that predicted by eq 6 by the factor α.[3] The dependency of α on molecular weight is complex,[17,19] but for a limited molecular weight range, it is usually possible to approximate it as

$$\alpha = (\text{constant})n^x \quad (7)$$

with x generally less than 0.1. The intrinsic viscosity is then given by the relation

$$[\eta] = K'n^a \quad (8)$$

with a generally between 0.5 and 0.8.

As was pointed out in the introduction to this paper, the thermodynamics of interaction between a polypeptide and the solvent must be a function of the amino acid side chains. The terms "ideal" and "nonideal," when applied to randomly coiled proteins, must represent averages over all of the constituent amino acids. Clearly some variation in α for different proteins in the same solvent must be expected. Again, however, this variation should not depend in a regular way on n, so that the exponent, a, of eq 8 should not be significantly affected. The variation between different proteins should be reflected in more or less random deviations from the viscosities predicted by eq 8 with a single value of K'.

It is thus evident that the results presented in Figure 2 and eq 2 are those to be expected for polypeptide chains obeying random-coil hydrodynamics. The exponent of n classifies the solvent as a moderately good solvent. The values of Huggins constants shown in Table I are also of the order of magnitude expected for random coils, for which k is typically near 0.35.[39] (As was pointed out earlier, the precision with which k values were determined is not high.) It may be noted that myosin, which is a fibrous protein in its native state, undergoes a marked reduction in $[\eta]$ on transfer to the GuHCl solution, in contrast to the increase in $[\eta]$ which occurs for globular proteins. All distinction between the two kinds of protein is lost in GuHCl solution.

A similar conclusion is compatible with our measurements of sedimentation velocity. The theoretically predicted[17,18] dependency of frictional coefficient f on molecular weight is that f should vary as $\alpha n^{1/2}$ for random coils, where α is given by eq 7. With

$$s^0 = M_0 n(1 - \phi'\rho)/Nf \quad (9)$$

where N is Avogadro's number, we thus have $s^0/(1 - \phi'\rho)$ proportional to $n^{0.50-x}$ for random coils. From the experimental value of the coefficient a of eq 8 ($a = 0.66$), the appropriate value of x is 0.05, and $s^0/(1 - \phi'\rho)$ should therefore vary as $n^{0.45}$. The experimental result, as given by Figure 5 and eq 5, is clearly identical with the expected result within the experimental error to which the data are subject.

In view of the fact that proteins are extremely heterogeneous with respect to amino acid composition, random-coil behavior as such cannot be regarded as necessarily indicating that they are randomly coiled in the way this term is usually used, implying flexibility between monomer units all along the polymer chain. It is conceivable that some regions of the polypeptide chains still retain a rigid structure, forming one or more solid sphere-like beads. The viscosity of partially random coils, containing such beads at one or more points along the polymer chain, could still obey eq 8 approximately if the fraction of monomer units incorporated in the beads remains approximately unchanged as the chain length increases. The major distinction between this kind of a situation and a true random coil would lie in the absolute magnitude of the intrinsic viscosities or the dimensions of the coil. If only a part of the polymer chain is in a randomly coiled state, the volume of the molecular domain would be significantly reduced.

In our earlier paper,[20] this question was tentatively resolved by showing that unperturbed end-to-end distances calculated from our viscosity data lie within the range predicted for randomly coiled polypeptide chains by the theoretical treatment of Flory[40] and Brant and Flory.[41] The recent extension of the theoretical treatment to polypeptide chains containing glycine and amino acid side chains with branched β-carbon atoms[42] improves the agreement between the theoretical and observed results. It indicates that a small number of glycine residues markedly decreases the expected end-to-end distance, but that a small number of side chains with branched β-carbon atoms does not exert much influence. The result obtained by us, that the observed dimensions fall close to the *lower* limit of the range considered permissable by the Brant and Flory calculations, is thus reasonable.

The question of structured beads arises again when we consider the data in 6 M GuHCl alone, in the absence of RSH. The polypeptide chains of some of the proteins under these conditions possess cross links (disulfide bonds). Such cross links must of themselves decrease the viscosity, because of the physical restriction which they impose on an otherwise random coil. In addition, however, the disulfide bonds could upset the delicate equilibrium (in some local region) between a compact structure and a randomly coiled chain, so as to favor the former, whereas the latter is favored when the disulfide bond is broken. In other words, part of the decrease in viscosity could arise from the formation of a few small structured beads in parts of the polypeptide chain near the disulfide bonds.

Theoretical equations for the effects to be expected from the purely physical restrictions of cross links have been derived only for the case of macromolecules which are closed rings.[43,44] An extension to the more complicated assembly of rings and linear segments which result from the specific locations of cross links in the proteins here under study[45-47] is not simply made. The possible existence of structured beads in cross-linked proteins in 6 M GuHCl cannot therefore be eliminated in a rigorous way on the basis of the viscosity data alone. It should be noted, however, that the intrinsic viscosities of β-lactoglobulin (two disulfide

(39) M. L. Huggins, *J. Am. Chem. Soc.*, **64**, 2716 (1942).

(40) P. J. Flory, *Brookhaven Symp. Biol.*, **13**, 89 (1960).
(41) D. A. Brant and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2791 (1965).
(42) W. G. Miller and P. J. Flory, *ibid.*, in press.
(43) V. Bloomfield and B. H. Zimm, *J. Chem. Phys.*, **44**, 315 (1966).
(44) M. Fukatsu and M. Kurata, *ibid.*, **44**, 4539 (1966).
(45) B. S. Hartley, J. R. Brown, D. L. Kauffman, and L. B. Smillie, *Nature*, **207**, 1157 (1965).
(46) D. G. Smyth, W. H. Stein, and S. Moore, *J. Biol. Chem.*, **238**, 327 (1963).
(47) F. Sanger, E. O. P. Thompson, and R. Kitai, *Biochem. J.*, **59**, 509 (1965).

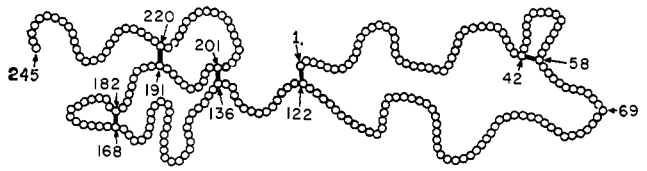

Figure 7. Location of disulfide bonds in chymotrypsinogen.[45]

bonds per chain) and pepsinogen (three disulfide bonds) are not reduced substantially when the disulfide bonds remain intact. The intrinsic viscosity of serum albumin also remains quite high when one considers the fact that it contains 17 disulfide bonds per chain. A question arises perhaps in the case of chymotrypsinogen (five disulfide bonds), for which $[\eta] = 11.0$ cc/g when disulfide bonds are intact, compared to the value of 26.8 cc/g given in Table I. The locations of the disulfide bonds of this protein are shown in Figure 7.

Additional evidence on this question is provided by optical rotatory dispension data, to be presented in a subsequent paper. These data suggest that structured beads do not exist in any of the proteins we have investigated. The effect of the presence of the disulfide bonds on the rotation is quite small, and preliminary examination of the data suggests that the effect for each protein (including chymotrypsinogen) consists simply of a fixed change in molar rotation for each disulfide bond broken.

It should be noted, on the other hand, that synthetic polypeptides such as polyleucine and polyphenylalanine retain structured regions in 6 *M* GuHCl, at least at room temperature.[48,49] Structured regions may therefore be expected in proteins, too, if long segments of a polypeptide chain consist entirely or predominantly of highly hydrophobic amino acid residues.

(48) H. E. Auer and P. Doty, *Biochemistry*, **5**, 1716 (1966).
(49) H. J. Sage and G. D. Fasman, *ibid.*, **5**, 286 (1966).

Acid–Base Titrations in Concentrated Guanidine Hydrochloride. Dissociation Constants of the Guanidinium Ion and of Some Amino Acids[1]

Yasuhiko Nozaki and Charles Tanford

Contribution from the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27706. *Received August 22, 1966*

Abstract: This paper shows that concentrated guanidine hydrochloride is a suitable medium for acid–base titrations, not significantly different from concentrated solutions of alkali metal chlorides, except at high pH, where the dissociation of the guanidinium ion becomes appreciable. The p*K* for this dissociation, in 6 *M* solution, was found to be 13.74 at 25°. The p*K*'s of a number of amino acids have been determined in 6 *M* guanidine hydrochloride and were found to be quite close to the corresponding p*K* values in dilute salt solutions. The results indicate in particular that intramolecular electrostatic interactions, between charged groups on amino acid molecules, are not greatly affected by the presence of high concentrations of guanidine hydrochloride.

It has recently been suggested in a paper from this laboratory[2] that proteins dissolved in concentrated guanidine hydrochloride (GuHCl)[3] solutions lose their characteristic native structure and assume a random conformation similar to that which simple organic polymers usually possess in solution. A detailed study of several proteins under these conditions is in progress, and some of the initial results are reported in papers which accompany this one.[4]

It is of special interest to investigate the acid–base properties of proteins in this solvent medium, as the complex interactions which introduce anomalies into the hydrogen ion equilibria of native proteins[5,6] should disappear. However, as only fragmentary information exists on the suitability of concentrated GuHCl solutions as a medium for acid–base titrations,[7] a preliminary investigation with simpler acids and bases appeared desirable. This paper reports the results of such an investigation. The results prove interesting for their own sake, as an example of acid–base equilibria in concentrated salt solutions, quite apart from their subsequent use in the interpretation of protein titration curves.

Experimental Section

Guanidine Hydrochloride. The GuHCl used in this study was prepared from commercial guanidine carbonate (Eastman). The carbonate was first recrystallized from chilled 60% aqueous ethanol, and converted to the chloride by addition of 20% HCl to a lasting pH 4. The mixture was flash evaporated at 40° until crystals began to form. It was then cooled, and the crystals were filtered off. The crystals were dissolved in hot absolute alcohol and

(1) Supported by grants from the National Science Foundation and from the National Institutes of Health, U. S. Public Health Service.
(2) C. Tanford, K. Kawahara, and S. Lapanje, *J. Biol. Chem.*, **241**, 1921 (1966).
(3) The abbreviations GuH^+ and Gu will be used to designate the guanidinium ion and the uncharged guanidine molecule, respectively.
(4) C. Tanford, K. Kawahara, and S. Lapanje, *J. Am. Chem. Soc.*, **89**, 729 (1967); Y. Nozaki and C. Tanford, *ibid.*, **89**, 742 (1967).
(5) C. Tanford, *Advan. Protein Chem.*, **17**, 69 (1962).
(6) J. Steinhardt and S. Beychok, *Proteins*, **2**, 140 (1964).
(7) J. W. Donovan, M. Laskowski, Jr., and H. A. Scheraga, *J. Mol. Biol.*, **1**, 293 (1959).