Proteins as Random Coils. IV. Osmotic Pressures, Second Virial Coefficients, and Unperturbed Dimensions in 6 *M* Guanidine Hydrochloride[1]

Savo Lapanje[2] and Charles Tanford

Contribution from the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27706. *Received May 4, 1967*

Abstract: Osmotic pressure measurements are reported for reduced protein polypeptide chains in 6 *M* guanidine hydrochloride. Previous results have indicated that the polypeptide chains are random coils, devoid of long-range structure, in this solvent medium, and end-to-end distances have been calculated for them from viscosity measurements. In this paper these results are corrected to unperturbed dimensions with the aid of second virial coefficients obtained from the osmotic pressure data. The relation between the unperturbed end-to-end distances (in A) and the number (n) of amino acids per chain is found to be $\langle L^2 \rangle_0 = (60 \pm 10)n$, a result essentially identical with that obtained earlier from viscosity data alone by use of the Kurata–Stockmayer procedure. These dimensions are somewhat shorter than would be predicted on the basis of theoretical and experimental studies of Miller, Brant, and Flory, but the discrepancy is judged to be within the limits of error inherent in the experimental and theoretical procedures.

We have shown in earlier papers of this series[3-6] that proteins dissolved in 6 *M* guanidine hydrochloride (GuHCl) appear to be devoid of any residual noncovalent structure. In the presence of a reducing agent, which breaks any disulfide bonds which may be present, the proteins behave like randomly coiled linear polymers. The physical chemistry of such polymers is well understood. In particular, characteristic dimensions, such as the average radius of gyration or the average end-to-end distance, can be calculated from appropriate experimental data. Of the data which we have obtained, the intrinsic viscosity is best suited for this purpose. Preliminary values of the end-to-end distances of a number of reduced protein polypeptide chains, based on viscosity measurements in 6 *M* GuHCl, have been reported by us.[3]

The dimensions of randomly coiled polymer chains are strongly influenced by polymer–solvent interactions. Thus different values are obtained for the same polymer in different solvents. It is generally desirable to eliminate the long-range effects of such interactions by determining the so-called *unperturbed* dimensions, these being defined as the dimensions in a thermodynamically ideal solvent. It is particularly interesting to determine unperturbed dimensions for protein polypeptide chains, because an initial theoretical calculation of these dimensions has been reported by Flory and co-workers.[7,8] A comparison between experimental and theoretical dimensions is obviously desirable.

The best way to determine unperturbed dimensions is to base them on measurements made in a thermodynamically ideal solvent. We do not believe that such a solvent can be found for *randomly coiled* protein polypeptide chains, because such chains always contain a diverse variety of amino acids. The terms "ideal" and "nonideal" when applied to solutions of protein polypeptide chains must reflect an averaging of the interactions of all the constituent amino acid residues with the solvent. An apparently *moderately* good solvent, such as 6 *M* GuHCl is for protein polypeptide chains, presumably represents a balance between *highly* favorable residue–solvent contact in some locations and less favorable ones (or even distinctly favorable residue–residue contacts) in other locations. If an ideal solvent were to exist, the average preference would have to be for contacts between residues, so as to counteract the excluded volume due to the space occupied by the polymer chain itself. To achieve such an average preference would require some contacts between residues which would be *highly* favored, and the likelihood is that such contacts would lead partially or completely to a stable long-range structure.

The alternate procedure, and the one we have sought to employ, is to estimate the extent to which nonideality is influencing the observed results, and to calculate the unperturbed dimensions directly from the observed dimensions in 6 *M* GuHCl. This can be done on the basis of viscosity measurements alone, by using the procedure of Kurata and Stockmayer,[9] or the equivalent method of Stockmayer and Fixman.[10] Application of this procedure to the viscosities of reduced protein polypeptide chains in 6 *M* GuHCl was reported in a preliminary communication from this laboratory,[3] and was shown to lead to the result

$$\langle L^2 \rangle_0 = 70n \qquad (1)$$

where $\langle L^2 \rangle_0$ is the root-mean-square unperturbed end-to-end distance in A, and n is the number of residues per chain. The procedure of Kurata and Stockmayer is however not the preferred method for estimating the

(1) This work was supported by grants from the National Science Foundation and from the National Institutes of Health, U. S. Public Health Service.

(2) On leave of absence from the Department of Chemistry, University of Ljubljana, Yugoslavia.

(3) C. Tanford, K. Kawahara, and S. Lapanje, *J. Biol. Chem.*, **241**, 1921 (1966).

(4) C. Tanford, K. Kawahara, and S. Lapanje, *J. Am. Chem. Soc.*, **89**, 729 (1967).

(5) Y. Nozaki and C. Tanford, *ibid.*, **89**, 742 (1967).

(6) C. Tanford, K. Kawahara, S. Lapanje, T. M. Hooker, Jr., M. H. Zarlengo, A. Salahuddin, K. C. Aune, and T. Takagi, *ibid.*, **89**, (1967).

(7) D. A. Brant and P. J. Flory, *ibid.*, **87**, 2791 (1965).

(8) W. G. Miller, D. A. Brant, and P. J. Flory, *J. Mol. Biol.*, **23**, 67 (1967).

(9) M. Kurata and W. H. Stockmayer, *Fortschr. Hochpolymer Forsch.*, **3**, 196 (1963).

(10) W. H. Stockmayer and M. Fixman, *J. Polymer Sci.*, **1C**, 137 (1963).

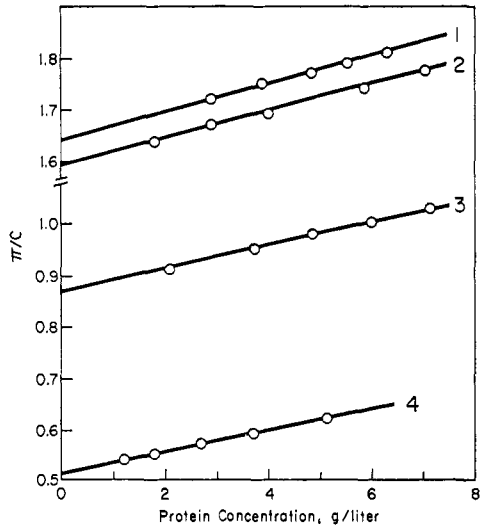

Figure 1. Representative experimental data. The units for osmotic pressure are centimeters of solvent (density 1.14 g/cc). The value of RT in the units employed (at 25°) is 2.20×10^4. Curves 1 and 2 represent duplicate runs, starting with different stock solutions, for ribonuclease. These were the only two duplicate runs which showed such a large difference in the intercept. Curves 3 and 4 are representative runs for chymotrypsinogen and aldolase.

effects of nonideality. Alternative procedures, based on independent measures of thermodynamic nonideality *per se*, are considered more reliable.

The objective of this paper is to obtain an independent estimate of nonideality by determination of second virial coefficients (A_2) from osmotic pressure measurements. These results will then be combined with our previous viscosity data to yield unperturbed dimensions.

Experimental Section

Proteins. The source of proteins used in this study and the preparation of samples for solution have been described previously.[4,6] All the measurements described here were performed between pH 6 and 8, but we do not believe that pH is an important variable because electrostatic effects must be essentially eliminated by the high salt concentration present. In the experiments with pepsinogen, Tris buffer was used to maintain the pH constant. For all other proteins pH adjustment was made by addition of HCl or KOH.

Osmotic Pressure. Osmotic pressure measurements were made with the Hewlett-Packard 503 high-speed membrane osmometer. Schleicher and Schuell's B 18 and B 19 membranes were used throughout this study. The former was used for aldolase and serum albumin, the latter for other proteins. Before use, the membranes were equilibrated with the solvent, 6 M GuHCl + 0.1 M β-mercaptoethanol, for at least 24 hr. During this time they slightly swelled.

The time necessary for the bubble in the osmometer capillary to rise from the trap to the height of the light path was an important criterion for the quality of a membrane. Sometimes this time was 10 min, often up to 20 min. If the time was higher, the membrane was replaced with another. In accord with this, in some cases equilibrium in the system was reached within 10 or 15 min; in others we had to wait for 0.5 hr or even longer. This may partly be due to sticking of the solvent to the capillary walls resulting in the sluggishness of response of the system. This phenomenon was observed regardless of how thoroughly the capillary was cleaned. This appears to be an inherent characteristic of this particular solvent in addition to its swelling action on the membranes.

An important criterion for the precision of each set of measurements with a membrane was the reproducibility of the solvent readings. Those sets of measurements were judged as satisfactory in which the fluctuations were less than 0.15 cm. In best experiments they were better than 0.05 cm. If in the beginning of a set of measurements large fluctuations of the solvent readings were observed, the measurements were discontinued, and a new membrane was installed. This was very often the case. An additional criterion is naturally the reproducibility of the solution readings. Therefore several experiments were performed with each set of solutions, and the sequence in which solutions were introduced in the system was changed in each set of measurements. We believe that in this way we avoided as much as possible the possibility of taking an apparent equilibrium due to sluggishness of response for a true equilibrium. In addition to this, for each protein at least two completely independent sets of measurements were made, *i.e.*, sets starting with a new stock solution. The estimated error for each protein is given in the table presenting the experimental data. An additional criterion for judging the accuracy of our data is the values of the molecular weights of the proteins studied, resulting from the osmotic measurements.

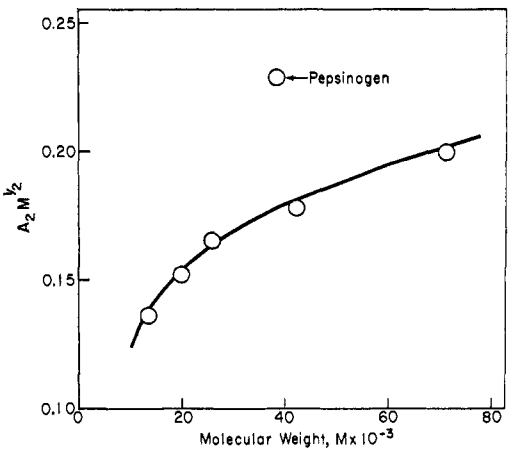

Figure 2. Plot of $A_2M^{1/2}$ as a function of molecular weight.

In some cases we extended equilibration of the membrane with the solvent over a period of a few days without noticing any particular effect. However, when the equilibration time was less than 24 hr, reproducibility of readings was less than usual which means that the swelling of membranes was not completed yet.

We also attempted in a few instances to use as membranes cellophane tubing (Visking Corp.) as used by Kupke[11] in his osmometric studies of insulin and ribonuclease in concentrated GuHCl solutions. We were not able to obtain reproducible measurements in this way.

Results

All the measurements reported here were carried out in 6 M GuHCl, containing 0.1 M β-mercaptoethanol, at 25°. The purpose of the mercaptoethanol was to make sure that all of the polypeptide chains were in the reduced condition, containing no disulfide cross-links.

Typical experimental data are shown in Figure 1. Molecular weights and second virial coefficients obtained from the data are tabulated in Table I. The molecular weights are seen to agree to within a few per cent with accepted values or values determined previously by us by sedimentation equilibrium. In Figure 2 we have plotted $A_2M^{1/2}$ as a function of molecular weight. We see that all the results (except that for pepsinogen) fall on a smooth curve, $A_2M^{1/2}$ increasing with an increase in molecular weight. This is the typical result for polymers composed of identical subunits. That such a result is obtained here suggests that the *average* thermodynamic interaction parameters which characterize each protein polypeptide chain, and which in principle can vary from one protein to another as a result of variations in amino acid content, do not in fact vary significantly. Whether the exceptional position of pepsinogen represents a real difference or an artifact is not certain. The pepsinogen solutions contained added

(11) D. W. Kupke, *Compt. Rend. Trav. Lab. Carlsberg*, **32**, 107 (1961).

buffer, not used for any other protein, and this could account for the discrepancy. Unfortunately, the discrepancy was not discovered until after the experiments had been completed, and additional measurements in the absence of buffer could not be made.

Table I. Results of Osmotic Pressure Measurements

Protein	Mol wt, exptl	$A_2 \times 10^3$, cc mole/g^2
Ribonuclease	13,600	1.16 ± 5%
β-Lactoglobulin	20,000	1.08 ± 8%
Chymotrypsinogen	25,700	1.03 ± 3%
Pepsinogen	38,400	1.17 ± 9%
Aldolase	41,900	0.87 ± 8%
Serum albumin	71,100	0.75 ± 12%

To use the second virial coefficient in the determination of unperturbed dimensions, A_2 must be related to the expansion factor α, defined as

$$\alpha^2 = \langle L^2\rangle/\langle L^2\rangle_0 \qquad (2)$$

where $\langle L^2\rangle$ is the root-mean-square end-to-end distance under nonideal condition. A variety of expressions have been used to relate A_2 to α, some of which make explicit use of *both* the second virial coefficient and the intrinsic viscosity. Others involve the use of A_2 alone, together with some measure of the unperturbed dimensions, which need not necessarily come from viscosity measurements. Among relations of the first type are the following, due to Krigbaum[12] and to Orofino and Flory,[13,14] respectively.

$$A_2M/[\eta] = 2 - 2/\alpha^3 \qquad (3)$$

$$A_2M/[\eta] = 189 \ln [1 + 0.885(\alpha^2 - 1)] \qquad (4)$$

A survey of the possible relations between A_2 and α which do not make explicit use of the intrinsic viscosity has been made recently by Berry,[15] in relation to experimental data for polystyrene. He found that the following equation

$$A_2M^{1/2} = \frac{4N(\pi/6)^{3/2}(\langle L^2\rangle_0/M)^{3/2}}{a_1}(\alpha_2 - 1) \qquad (5)$$

in which N is Avogadro's number and a_1 a universal constant equal to 134/105, though theoretically more approximate than some other relations of similar kind, was able to reproduce experimental data over a wider range of conditions, and we shall therefore use it here. (Alternative relations, such as result from combination of the Flory–Fisk[16] equation for α with the Orofino–Flory[13] equation for A_2, would not have substantially altered the results.)

None of these equations is free from uncertainty. The relations involving $A_2M/[\eta]$ seem more straightforward, but their derivation has involved an assumed value for the parameter Φ of eq 6, which, as we shall point out in the discussion, cannot really be assigned with confidence. Equation 5 (and alternate equations of similar type) require previous knowledge of a value for the very parameter, $\langle L^2\rangle_0/M$, which we wish to calculate. This turns out to be not as objectionable as might at first be supposed, because we need only an approximate value for $\langle L^2\rangle_0/M$, and use it only to estimate a nonideality correction. The main burden of the calculation falls on viscosity data, and the result will prove to be very insensitive to the choice of an approximate value of $\langle L^2\rangle_0/M$ for eq 5.

Table II gives the results of the calculations of α with eq 3–5. In using eq 5 we have considered two possible initial values for $\langle L^2\rangle_0/M$. One is based on the preliminary unperturbed dimensions as given by eq 1. With[17] $M = 110n$, this is equivalent to $\langle L^2\rangle_0/M = 0.64$. An alternative value is based on the theoretical calculations of Flory and co-workers,[7,8] and on their experimental results for synthetic polypeptides.[8,14] For polypeptide chains containing a small amount of glycine (as ours do) a figure of $\langle L^2\rangle_0/M \simeq 0.90$ is appropriate.

Table II. Calculation of the Expansion Factor α from A_2

Protein	Value of α			
	Eq 4	Eq 3	Eq 5[a]	Eq 5[b]
Ribonuclease	1.33	1.25	1.17	1.11
β-Lactoglobulin	1.32	1.24	1.18	1.12
Chymotrypsinogen	1.33	1.26	1.20	1.13
Pepsinogen	1.51	1.52	1.28	1.18
Aldolase	1.35	1.27	1.22	1.14
Serum albumin	1.35	1.27	1.24	1.16

[a] Using $\langle L^2\rangle_0/M = 0.64$. [b] Using $\langle L^2\rangle_0/M = 0.90$.

Intrinsic viscosities for the same six proteins listed in Table I have been reported earlier. The results, together with root-mean-square end-to-end distances determined from them, are shown in Table III. The values of $\langle L^2\rangle^{1/2}$ were obtained from the well-known relation[18,19]

$$[\eta]M = \Phi\langle L^2\rangle^{3/2} \qquad (6)$$

The value of the parameter Φ in this equation which we have used is the asymptotic value for high molecular weight *impermeable* coils in good solvents, which is close to 2.1×10^{23} cgs units. These dimensions may now be combined with α values of Table II to obtain unperturbed dimensions. Different values are, of course, obtained depending on which set of α values is used.

Table III. Intrinsic Viscosities and Unperturbed End-to-End Distances (in A)[a]

Protein	$[\eta]$, cc/g	$\langle L^2\rangle^{1/2}$	$\langle L^2\rangle_0^{1/2}$, min[b]	$\langle L^2\rangle_0^{1/2}$, max[b]
Ribonuclease	16.3	102	77	92
β-Lactoglobulin	22.8	126	95	112
Chymotrypsinogen	26.8	148	115	131
Pepsinogen	31.5	179	119	152
Aldolase	35.3	189	140	166
Serum albumin	52.2	258	191	222

[a] These data are based on the assumption that eq 6 applies, with $\Phi = 2.1 \times 10^{23}$ cgs units. [b] Using α values from the first and last columns of Table II.

(12) W. R. Krigbaum, *J. Polymer Sci.*, **18**, 315 (1955).
(13) T. A. Orofino and P. J. Flory, *J. Chem. Phys.*, **26**, 1067 (1957).
(14) D. A. Brant and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2788 (1965).
(15) G. C. Berry, *J. Chem. Phys.*, **44**, 4550 (1966).
(16) P. J. Flory and S. Fisk, *ibid.*, **44**, 2243 (1966).
(17) The mean residue weights of the proteins used in this study range from 105 to 112.
(18) P. J. Flory, "Principles of Polymer Chemistry," Cornell University Press, Ithaca, N. Y., 1953, Chapter 14.
(19) C. Tanford, "Physical Chemistry of Macromolecules," John Wiley and Sons, Inc., New York, N. Y., 1961, Chapter 6.

Maximum and minimum values, using the smallest and largest α values, are given in Table III.

The extreme values of $\langle L^2\rangle_0^{1/2}$ shown in the last two columns of Table III are plotted against $M^{1/2}$ in Figure 3. A reasonable fit to the expected linear relation between these quantities is obtained, and the same would be true if the intermediate values of α from Table II were used. The least-squares equations for the straight lines of Figure 3 are $\langle L^2\rangle_0^{1/2} = 0.81M^{1/2}$ and $\langle L^2\rangle_0^{1/2} = 0.68M^{1/2}$, respectively. Since there is no compelling argument in favor of any of the individual sets of α values of Table II, it is perhaps best to state the over-all result of these measurements as a range. In terms of the number of residues per chain, this becomes

$$\langle L^2\rangle_0 = (60 \pm 10)n \qquad (7)$$

which is in excellent agreement with the relation given by eq 1, which was determined[3] (with an uncertainty of about 10 to 15%) from viscosity data alone, using data for a larger number of proteins than were used in the present study.

It may be noted that the points in Figure 3 for pepsinogen ($M^{1/2} = 196$) do not deviate dramatically from those of the other proteins, in spite of the anomalous value of A_2 for this protein seen in Figure 2.

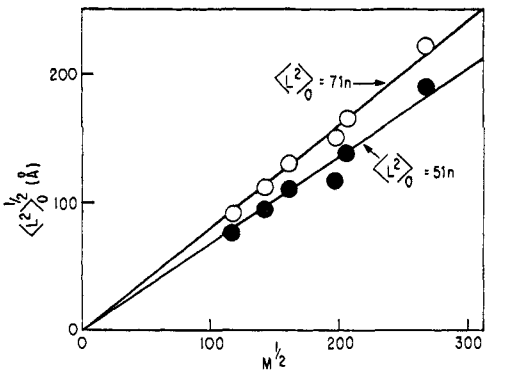

Figure 3. Unperturbed end-to-end distances, from Table III, as a function of $M^{1/2}$.

Discussion

The unperturbed dimensions of randomly coiled reduced protein polypeptide chains in 6 *M* GuHCl, as given by eq 7, are somewhat lower than the dimensions calculated on a theoretical basis by Brant and Flory[7] and by Miller, Brant, and Flory[8] and the corresponding experimental $\langle L^2\rangle_0/n$ values which they obtained for some synthetic polypeptides.[8,14] Their results may be written as $\langle L^2\rangle_0 = 130n$ for a polypeptide chain in which every residue contains a CH_2 group in the β position. They have shown that the presence of glycine reduces the end-to-end distance to $\langle L^2\rangle_0 = 95n$ when the glycine content is 10%. Additional shrinkage which could result from the presence of proline remains to be investigated. It may be noted that all of the proteins we have used contain both glycine and proline. The glycine content varies from 2 to 10% and the proline content from 3 to 6% of the total number of residues.

Neither the theoretical nor experimental results of Flory and co-workers have an absolute accuracy superior to the results presented here. Their experimental data, like ours, were obtained under nonideal conditions, and corrected to ideal conditions by use of the second virial coefficient, and this procedure, as we have shown here, involves considerable uncertainty. Their theoretical calculations involve parameters which are not known with great precision and which, to some extent, were adjusted to fit the experimental data. These considerations, plus the already mentioned possibility of an influence of the presence of proline on the unperturbed dimensions, suggest that the discrepancy between our results and those of Flory and co-workers is not significant.

An additional possibility is that the value of Φ (eq 6) which we have used is too large. (A smaller value would bring our dimensions closer to the values of Flory and co-workers.) The value we have used is by no means firmly established. It is based on experimental comparisons between viscosities and light-scattering dimensions for relatively few polymer–solvent systems. Moreover, the comparisons have always been made for polymers of higher molecular weight and larger dimensions than the protein used in this study. It has frequently been suggested that shorter chains may not be entirely impermeable to solvent.[20] Berry,[15] for instance, has provided evidence that partial free draining occurs even with polystyrene, which has a higher segment density than polypeptides, in the molecular weight range near 100,000. Partial free draining would result in smaller Φ values, and very much smaller Φ values are frequently observed with cellulose derivatives[21–23] and have been ascribed to partial free draining.

It is of interest in this connection that the results of Miller, Brant, and Flory[8] for poly-L-glutamate indicate a trend to smaller values of $\langle L^2\rangle_0/n$ for smaller molecular weights when a constant value of Φ (the same as we have employed) is used. For one sample, of molecular weight 20,000 and containing 13% glycine, they obtained $\langle L^2\rangle_0/n$ = 55 to 60 A^2, which agrees with our result for proteins.

(20) For a recent discussion, see R. Ullman, *J. Chem. Phys.*, **40**, 2193 (1964).
(21) M. L. Hunt, S. Newman, H. A. Scheraga, and P. J. Flory, *J. Phys. Chem.*, **60**, 1278 (1956).
(22) W. R. Krigbaum and L. H. Sperling, *ibid.*, **64**, 99 (1960).
(23) J. J. Hermans, *J. Polymer Sci.*, **2C**, 117 (1963).