Molecular Conformation of Bovine A1 Basic Protein, A Coiling Macromolecule in Aqueous Solution†

W. R. Krigbaum* and Thomas S. Hsu‡

ABSTRACT: Aqueous solutions of bovine A1 protein, the major component of the basic protein fraction of myelin, were studied by small angle X-ray diffraction. The experimentally determined molecular weight, 17,800, is within 3% of that corresponding to the amino acid sequence, 18,395, and the radius of gyration was found to be 46.3 Å. No equivalent scattering particle of uniform electron density could be found which was compatible with all parameters evaluated from the diffraction measurements. The possibility of a coiled shape was therefore investigated using a worm-like chain model. This yielded a contour length of 439 Å and a persistence length of 15.7 Å. The radius of gyration of this model chain, 47.1 Å, is in quite reasonable accord with the experimental value. The latter, after correction for excluded volume effects and finite chain length, yields for the characteristic ratio, $\overline{r_o^2}/n_p l_p^2$, 5.4. This may be compared with the value, 6.1, obtained after applying a correction for finite chain length to the viscosity data given by Tanford et al. for 12 proteins in 6 *M* guanidine hydrochloride and 0.1 *M* β-mercaptoethanol. These two experimental values fall in the expected order, since the 15% glycine content of the A1 protein is considerably higher than the average for other proteins, which is about 8%. The corresponding values predicted from conformational calculations by Miller et al. for random copolymers of the L-alanine-glycine type are 5.9 (15% glycine) and 7.0 (8% glycine). We conclude that the A1 protein exists predominately, if not exclusively, as a random coil in aqueous solution.

The myelin membrane of the central nervous system contains several types of polypeptide chains (Reynolds and Green, 1973). The acid-extractable material, which constitutes 25–30% of the total membrane protein, is a single species, the A1 protein. In addition to providing an example of a structural membrane protein of known sequence, the A1 protein is of special physiological interest as an active factor in experimental allergic encephalomyelitis (EAE), an auto-immune demyelinating disease. Hence, the A1 protein may be used to investigate delayed hypersensitivity phenomena. Further, EAE has been proposed as a possible animal model for multiple sclerosis in humans.

For the reasons cited above, the A1 protein has been actively investigated, particularly by Eylar and collaborators. Oshiro and Eylar (1970) showed that the A1 protein derived from bovine brain and spinal cord is identical. The sequences (Eylar et al., 1971) of bovine and human A1 proteins were found to be quite similar, consisting of 170 and 172 residues, respectively, and differing in only 11 residues. The bovine protein contains 31 lysine and arginine residues, and 10 histidine residues, but only 11 glutamic and aspartic acid residues. Hence, the molecule will bear a substantial positive charge at neutral pH. There are no cystine residues present in either protein. Hashim and Eylar (1969) found that pepsin digestion of the A1 molecule abolishes antibody activity, although the reaction mixture retains full encephalitogenic activity. Investigation of a series of synthetic peptides prepared by the Merrifield technique enabled Eylar et al. (1970) to establish that the active region is a nine residue sequence containing the only tryptophan present in the A1 molecule: -Phe-Ser-Trp-Gly-Ala-Glu-Gly-Gln-(Lys or Arg)-, and that the Trp—Gln(Lys or Asn) relationship is essential for encephalitogenic activity. A second region of the A1 molecule is active in rabbits, but not in guinea pigs.

Eylar and Thompson (1969) concluded from sedimentation and viscosity data that the A1 molecule takes the general shape of an open coil in aqueous solution. Both Palmer and Dawson (1969), and Oshiro and Eylar (1970), interpreted their optical rotatory dispersion spectra as indicating virtually no helix or β structure. On the other hand, this method is not highly reliable in the range of small amounts of secondary structure.

We have undertaken a small angle X-ray study of the A1 molecule in aqueous solution to obtain more detailed information concerning its conformation.

Experimental Section

Samples of the A1 protein were obtained from Merck Chemical Company through the courtesy of Dr. Eylar. The buffer solution used as a solvent was 0.050 *M* $NH_4C_2H_3O_2$, 0.0175 *M* $HC_2H_3O_2$, and 0.02 *M* NaCl (pH 5.2). The solutions were exhaustively dialyzed against buffer, and their concentrations were determined by measuring the optical density at the absorption peak near 280 mμ, taking for the extinction coefficient of a 1% solution, with 1-cm path length, 5.89 (Eylar and Thompson, 1969). The partial specific volume, $\bar{\nu}_2$, of the protein in buffer solution, as measured using an Anton Paar DMA-02c precision density meter, is 0.700 ± 0.005. This is somewhat lower than the value, 0.72, calculated using the method of Cohn and Edsall (1943).

Diffraction measurements were performed using a medium resolution Kratky camera, nickel filtered Cu Kα radiation, and a proportional counter with pulse height analyzer as the detection system. Slit optics were used, with a 90-μ entrance slit, 270-μ receiving slit, and a sample to detector

† From the Gross Chemical Laboratory, Duke University, Durham, North Carolina 27706. *Received November 1, 1974.* This work was supported by U.S. Public Health Service Grant HE 12157-0351 GPIV from the National Institutes of Health.

‡ Present address: Camille Dreyfus Polymer Research Laboratory, Research Triangle Park, Durham, North Carolina.

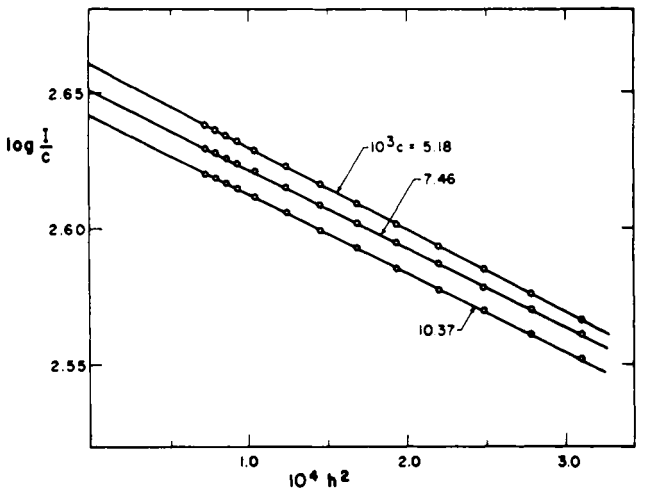

FIGURE 1: Guinier plot of the data at smallest angles for three concentrations of the A1 protein.

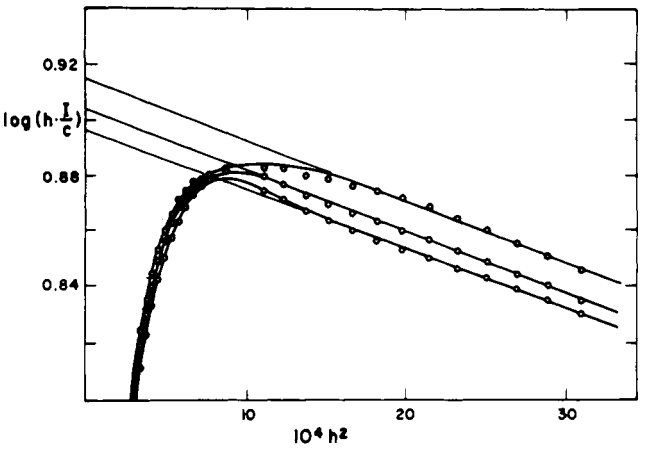

FIGURE 2: Plot for determining the radius of gyration of the cross-section (eq 4). Concentrations are in inverse order to those shown in Figure 1.

distance, d, of 22.0 cm. These settings resulted in a theoretical resolution of approximately 900 Å. One reads directly the vertical displacement, m (cm), of the detector slit from the primary beam, and this may be converted to the corresponding Bragg angle, θ, through use of the relation

$$m/a = \tan 2\theta \cong 2\theta \quad (1)$$

The scattered intensities were placed on an absolute scale by comparison with the scattering from a calibrated Lupolene platelet (Kratky et al., 1966; Pilz and Kratky, 1967) kindly provided by Dr. O. Kratky. Slit desmearing was performed using a modification of the computer program developed by Heine and Roppert (1962).

Experimental Parameters. The treatment of Guinier (1939) indicates that, for sufficiently small scattering angles, the angular dependence of the excess scattered intensity can be expressed in the form

$$\ln (I/c) = \ln (I_0/c) - (h^2R^2/3) \quad (2)$$

where $h = (4\pi/\lambda) \sin \theta$ is the magnitude of the reciprocal lattice vector, λ is the wavelength of the radiation, and R is the radius of gyration of the scattering particle. Hence, R can be evaluated from the initial slope of a plot of ln (I/c) vs. h^2. If the intercepts for different concentrations are extrapolated to infinite dilution, yielding $(I_0/c)_0$, the molecular weight of the scattering particle can be calculated from the relation

$$M = (d^2/i_eNP_0D(\Delta z)^2)\left(\frac{I_0}{c}\right)_0 \quad (3)$$

Here i_e is the Thomson scattering factor for one electron, N is the Avogadro number, P_0 is the intensity per centimeter of length of the primary beam in the plane of the receiving slit, D is the sample thickness (cm), and c is the solute concentration in grams per milliliter. The electron density difference in moles per gram, Δz, is given by $(z_2 - \bar{v}_2d_1)$, where z_2 is the number of moles of electrons per gram of solute, $\bar{v}_2$ is the partial specific volume of solute, and d_1 is the electron density of the solvent in moles per milliliter.

For filled particles having an elongated shape, the radius of gyration of the cross-section, R_x, can be determined from the slope of the linear portion of a plot of ln (hI/c) vs. h^2

$$\ln (hI/c) = \ln (hI_0/c) - h^2R^2/2 \quad (4)$$

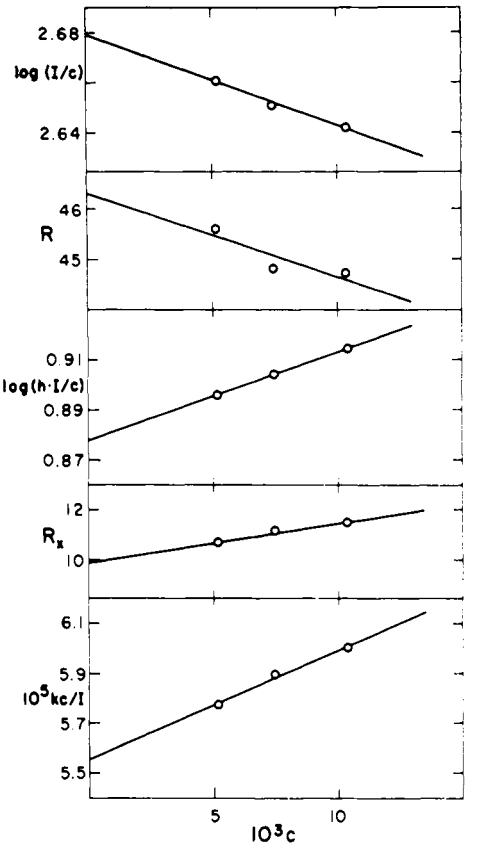

FIGURE 3: Concentration dependence of several experimental parameters extrapolated to infinite dilution.

Data obtained at the smallest angles for three concentrations of the A1 protein are shown as a Guinier plot in Figure 1. We were concerned that the A1 protein might bear sufficient charge at pH 5.2 so that interparticle interference effects would be troublesome. However, since the inner portion of Figure 1 shows no downward curvature, we conclude that the ionic strength of the buffer is sufficiently high, and the protein concentration is sufficiently low, to suppress interparticle interference effects. A similar plot of log (hI/c) vs. h^2 appears in Figure 2, while Figure 3 illustrates the concentration dependence of these intercepts and slopes. Extrapolation to infinite dilution yields M = 17,800, R = 46.3 Å, and R_x = 9.9 Å. The experimentally determined

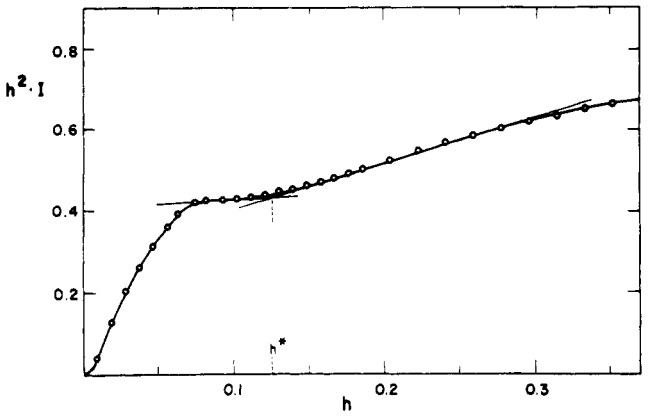

FIGURE 4: Plot of small angle data to determine the persistence length (see text).

molecular weight is in good agreement with that corresponding to the sequence of the bovine A1 protein, 18,395. We should also observe that the cross-sectional curves in Figure 2 do not have a normal appearance, and the concentration dependence of R_x shown in Figure 3 is also unusual.

Additional parameters may be determined through use of the invariant, Q, as described by Porod (1951a,b). The invariant may be evaluated using slit-smeared intensities, designated $\tilde{I}$

$$\tilde{Q} = \int_0^\infty h\tilde{I}dh \tag{5}$$

Since intensity measurements must be terminated when the excess intensity becomes too small to measure with adequate precision, the integral can only be evaluated if the outer part of the scattering curve exhibits a Porod region in which $\tilde{I}$ varies as h^{-3}. If the scattering particle has some degree of internal structure, the scattered intensity may contain an unwanted constant term, i.e., $\tilde{I} = k_1h^{-3} + k_2$. In this case $h^3\tilde{I}$ is first plotted against h^3, and k_1 is evaluated by extrapolating the linear region found at high angles back to $h^3 = 0$. The area of outer tail of the $h\tilde{I}$ curve is then evaluated analytically, taking $\tilde{I} = k_1h^{-3}$. The value obtained for the invariant in this manner is $\tilde{Q} = 0.397$. The volume per molecule, V, evaluated from the relation

$$V = 4\pi^2I_0/\tilde{Q} \tag{6}$$

is 4.73×10^4 Å^3. The average area of the cross-sectional surface, s_x, given by

$$s_x = 4\pi(hI)_0/\tilde{Q} \tag{7}$$

is 218 Å. The surface per unit volume, s, calculated from

$$\bar{s} = 4v_1k_1/\tilde{Q} \tag{8}$$

where v_1 is the volume fraction of solvent, is $\bar{s} = 0.411$ Å^{-1}, so that the total surface area per molecule, $\bar{s}V$, is 1.94×10^4 Å^2.

Interpretation of the Scattering Behavior. Our initial problem is to determine from the diffraction data whether the A1 protein molecule in solution is globular or coiling. We note the volume obtained from the invariant, $V = 4.73 \times 10^4$ Å^3, is much larger than the $\bar{v}M$ product, 2.10×10^4 Å^3. This large apparent swelling ratio suggests an open coil form. Nevertheless, we have attempted to find a filled particle of uniform electron density which would be equivalent in scattering behavior. Comparison of the observed radius of gyration, $R = 46.0$ Å, with the radius of gyration, 17.4 Å, of a sphere having the observed volume, V, indicates that the equivalent scattering particle must be quite anisotropic. The same conclusion is reached concerning the two cross-sectional dimensions from the difference between the observed cross-sectional radius of gyration, $R_x = 9.9$ Å, and that of a circular cross-section of area s_x, 5.9 Å. Comparison of the experimental scattering curve, plotted as log I vs. log h, with corresponding plots of normalized intensities calculated for particles of particular shape and various axial ratios, likewise indicates a very asymmetric shape. However, among the ellipsoids, cylinders, and rectangular prisms, no equivalent particle could be found which was compatible with all of the experimental parameters.

We therefore examined the possibility that the A1 molecule exists as an open coil in aqueous solution, using as a model the wormlike chain with finite persistence length. The quantity $(hI_0/c)_0$ can be obtained by extrapolating the intercepts of the ln (hI/c) vs. h^2 plot (eq 4) to infinite dilution. The mass per unit length, μ, of a coiling chain can then be evaluated from the relation

$$\mu = (d^2/i_eNP_0D(\Delta z)^2)(hI_0/c)_0 \tag{9}$$

yielding $\mu = 40.5$ g/(mol Å). The contour length of the chain, $L = M/\mu$, is found to be 439 Å.

The Debye scattering function for random coil exhibits, for h values just beyond the Guinier region, an $I \sim h^{-2}$ dependence. A chain with finite persistence length will undergo a transition with increasing scattering angle from coil behavior, $I \sim h^{-2}$, to rodlike behavior, $I \sim h^{-1}$. The persistence length can be evaluated from the h value (designated h^*) at which this transition occurs. Figure 4 illustrates a plot of h^2I vs. h for a solution of the A1 protein. The curve is seen to rise to a nearly horizontal plateau (corresponding to the $I \sim h^{-2}$ region), followed by a transition to a linearly increasing region (the $I \sim h^{-1}$ rod-like domain). The persistence length, a, may be evaluated from h^* through use of the relation

$$a = A/h^* \tag{10}$$

where A is a numerical constant. For flexible chains obeying Kuhn statistics, $A = 1.91$. Heine et al. (1961) have performed Monte Carlo calculations for worm-like chains of short to medium length. They report that A decreases from 2.3 to 1.5 as the number of persistence lengths, x^*, in the chain increases from 10 toward infinity. The reason for the discrepancy between the latter limiting value and that for a flexible Kuhn chain is not immediately evident. We estimate, from a plot of the dependence of A upon x^*, that the appropriate value for the A1 protein is $A = 1.94$. Use of this value yields for the persistence length $a = 15.7$ Å. Hence, according to this model the A1 protein can be represented as a chain of $x^* = L/a = 28$ persistence lengths, each consisting of $170/28 = 6.1$ residues. The root-mean-square chain displacement length for such a worm-like chain, as calculated from

$$\overline{r^2} = 2a(L - a + ae^{-L/a}) \tag{11}$$

is $(\overline{r^2})^{1/2} = 115$ Å. Since $R^2 = r^2/6$, this corresponds to a radius of gyration, R, of 47.1 Å, in good agreement with the experimental value, $R = 46.3$ Å.

The fact that Figure 4 exhibits clearly defined coil and rod-like regions, and the agreement between the experimental value of R and that predicted for a worm-like chain having the observed contour length and persistence length, argues strongly in favor of a coiling form for the A1 mole-

cule. In an attempt to provide additional verification of this conclusion, the characteristic ratio, $(r_o^2/n_p l_p^2)$, was estimated for the A1 protein. Here n_p is the number of residues in the chain, and l_p = 3.80 Å is the length of the trans peptide unit. For this purpose, the experimental value of $\overline{r^2}$ must be corrected for both excluded volume effects and the finite value of n_p. In order to estimate the chain expansion factor, α, the osmotic second virial coefficient, A_2, was evaluated according to the relation

$$k(c/I_0) = (1/M) + 2A_2c + \ldots \quad (12)$$

as shown in the lower portion of Figure 3, yielding $A_2 = 22 \times 10^{-4}$ ml mol/g^2. This relatively large value undoubtedly reflects the polyelectrolyte behavior of the A1 protein in aqueous solution. Using $[\eta]$ = 9.27 ml/g, as reported for the A1 protein by Eylar and Thompson (1961), we estimate $\alpha^2 = 1.03$ from the relation of Orofino and Flory (1957)

$$A_2M/[\eta] = 188 \ln [1 + 0.886(\alpha^2 - 1)] \quad (13)$$

The A1 protein consists of 170 residues. Conformational calculations of Brandt and Flory (1965) for poly-L-alanine type homopolymers indicate that R_o^2/n_p for a chain of 170 residues will have attained approximately 90% of the value characteristic of an infinitely long chain. The A1 protein contains 25 glycine residues, or 15 residue percent, whereas the glycine content averaged over a large number of proteins is only 8% (Dayhoff and Hunt, 1972). Miller et al. (1967) have shown that, for random L-alanine-glycine type copolymers, the $\overline{r_o^2}/n_p$ ratio approaches the long chain limit more rapidly as the glycine content increases. Unfortunately, they do not furnish a curve for the chain length dependence of the (R_o^2/n_p) ratio. If we accept as an estimate that R_o^2/n_p will have achieved 95% of its asymptotic value for a 170 residue chain containing 15% glycine, then the characteristic ratio $(r_o^2/n_p l_p^2)$ is found to be 5.4 for the A1 protein. This stands in reasonable accord with the theoretical value, 5.9, predicted by Miller et al. (1967) for a random copolymer containing 15% glycine.

The A1 protein appears to be one member of the relatively small group of proteins which are coiling in the native state. However, the characteristic ratios of other proteins can be determined experimentally by performing measurements in a solvent which breaks any disulfide cross-links present, and completely denatures the protein to random coil form. Tanford et al. (1966, 1967) have collected viscosities measured for 12 different proteins in a solvent which was 6 M in guanidine hydrochloride and 0.1 M in β-mercaptoethanol. These authors calculated $\overline{r^2}$ using the relation of Flory (1953)

$$(\overline{r^2})^{3/2} = [\eta]M/\Phi \quad (14)$$

assigning $\Phi = 2.1 \times 10^{23}$ (with $[\eta]$ in ml/g). They then applied an excluded volume correction, using the extrapolation procedure of Kurata and Stockmayer (1963) or Stockmayer and Fixman (1963). Their final result, $\overline{r_o^2} = 70n_p$, corresponds to a characteristic ratio of 4.9.

Clearly, $[\eta]/M^{1/2}$ will increase with molecular weight due to both excluded volume effects and the increase of $(\overline{r_o^2}/n_p)$ toward its asymptotic limit. The effect of the latter variation is to decrease the $[\eta]/M^{1/2}$ ratio by a factor which increases quite rapidly with decreasing molecular weight. Tanford et al. (1966) recognized this fact, but they attempted to circumvent the correction by eliminating insulin, the protein of lowest molecular weight, from consideration. We have estimated finite chain length corrections for each

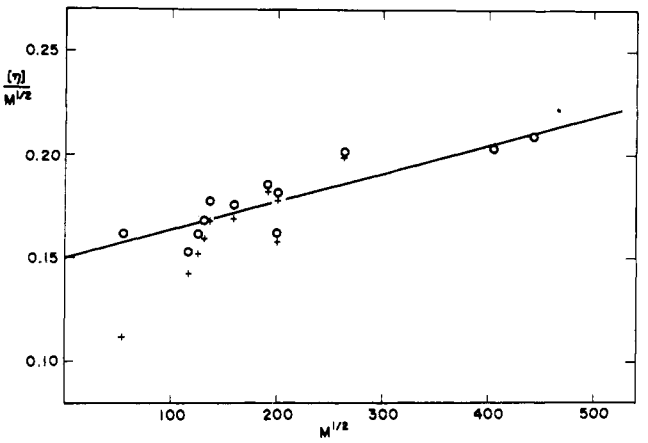

FIGURE 5: Viscosity data of Tanford et al. (1966, 1967) for proteins in guanidine hydrochloride (+), and after corrections or finite chain length (○), plotted according to the Stockmayer-Fixman procedure.

of their proteins using two different procedures. In the first, the chain length dependence of $(\overline{r_o^2}/n_p)$ calculated for random copolymers by Miller et al. was used, assigning to all polymers an average glycine content of 8%. The second method involves the assumption that the number of residues per persistence length found for the A1 protein, 6.1, is applicable to other proteins as well. The number of persistence lengths, x^*, in each protein chain can then be determined from its n_p value, and the ratio of (r_o^2/n_p) to its asymptotic value can then be evaluated using eq 10 for the worm-like chain. There was little difference between the correction factors obtained by these two procedures. Figure 5 illustrates the data given by Tanford et al., and the values after correction for finite chain length, plotted according to the Stockmayer-Fixman procedure. The corrected data can be presented reasonably well by a linear relationship, whereas the data prior to correction appear to be better represented by a curve concave downward. Following Tanford et al. in their assignment $\Phi = 2.1 \times 10^{23}$, we obtain a characteristic ratio of 6.1 for these 12 proteins. The glycine contents of these proteins range from 2 to 10%, but deviations from the line drawn in Figure 5 do not appear to correlate in any regular manner with glycine content. For comparison, the characteristic ratio predicted by Miller et al. (1967) for a random copolymer containing 8% glycine is 7.0.

In view of the higher glycine content (15%) of the A1 protein, the two experimental values for the characteristic ratio, 5.4 for the A1 protein and 6.1 for the 12 proteins of Tanford et al., fall in the expected order. The corresponding theoretical values, 5.9 and 7.0, are 10–15% higher. However, one must bear in mind that the conformational calculations were not intended to represent proteins, but homopolypeptides of glycine and L-alanine and their random copolymers. When the uncertainties in both the experimental and predicted ratios are taken into consideration, the agreement is quite good.

From the concordance between the two experimental values for the characteristic ratio, and of these with the values predicted by conformational analysis, and from the accord between the experimental R value and that predicted from the persistence length and chain contour length using the worm-like chain model, we conclude that the A1 molecule exists predominately, if not exclusively, in the form of a random coil in aqueous solution. This conclusion is in agreement with that reached by Palmer and Dawson (1969) and Oshiro and Eylar (1970) from optical rotatory

dispersion studies. Hence, the A1 protein is one of the small number of proteins which are randomly coiled in aqueous solution, and this is perhaps the first conformational study of such a protein by diffraction methods. Holtzer et al. (1965) performed a light scattering study of tropomyosin B in a denaturing solvent, 5 *M* guanidine hydrochloride; however, the studies are not comparable with ours, since these authors left the disulfide cross-links intact.

References

Brandt, D. A., and Flory, P. J. (1965), *J. Am. Chem. Soc. 87,* 2791.

Cohn, E. J., and Edsall, J. T. (1943), Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, New York, N.Y., Reinhold, p 375.

Dayhoff, M. D., and Hunt, L. T. (1972), in Atlas of Protein Sequence and Structure, Vol. 5, Dayhoff, M. D., Ed., Silver Spring, Md., National Biochemical Research Foundation, p 355.

Eylar, E. H., Brostoff, S., Hashim, G., Coccam, J., and Burnett, P. (1971), *J. Biol. Chem. 246,* 5770.

Eylar, E. H., Caccam, J., Jackson, J., Westall, F., and Robinson, A. (1970), *Science 168,* 1220.

Eylar, E. H., and Thompson, M. (1969), *Arch. Biochem. Biophys. 129,* 468.

Flory, P. J. (1953), Principles of Polymer Chemistry, Ithaca, N.Y., Cornell University Press, p 616.

Guinier, A. (1939), *Ann. Phys. (Leipzig) 12,* 161.

Hashim, G. A., and Eylar, E. H. (1969), *Arch. Biochem. Biophys. 135,* 324.

Heine, S., Kratky, O., Porod, G., and Schmitz, P. J. (1961), *Makromol. Chem. 44–46,* 682.

Heine, S., and Roppert, J. (1962), *Acta Phys. Austriaca 15,* 1948.

Holtzer, A., Clark, R., and Lowey, S. (1965), *Biochemistry 4,* 2401.

Kratky, O., Pilz, I., and Schmitz, P. J. (1966), *J. Colloid Interface Sci. 21,* 24.

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

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

Orofino, T. A., and Flory, P. J. (1957), *J. Chem. Phys. 26,* 1716.

Oshiro, Y., and Eylar, E. C. (1970), *Arch. Biochem. Biophys. 138,* 606.

Palmer, F. B., and Dawson, R. M. C. (1969), *Biochem. J. 111,* 629.

Pilz, I., and Kratky, O. (1967), *J. Colloid Interface Sci. 14,* 211.

Porod, G. (1951a), *Kolloid-Z, 124,* 83.

Porod, G. (1951b), *Kolloid-Z. 125,* 51.

Reynolds, J. A., and Green, H. O. (1973), *J. Biol. Chem. 248,* 1207.

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

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

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