Conformational Study and Determination of the Molecular Weight of Highly Charged Basic Proteins by Sedimentation Equilibrium and Gel Electrophoresis†

Juan Ausió and Juan A. Subirana*

ABSTRACT: We study several highly charged protamines and some related proteins from the sperm of molluscs. Circular dichroism and hydrodynamic parameters obtained from the sedimentation constant and intrinsic viscosity show that these proteins behave as random coils. However, it appears that a small amount of structure is present at basic pH. The molecular weight of these proteins is determined by several methods. When sedimentation equilibrium is used, we have found that the influence of concentration is much smaller than expected. We have also found that the highly charged nature of these proteins can be properly taken into account by using the methodology presently available [Williams, J. W., Van Holde, K. E., Baldwin, R. L., & Fujita, H. (1958) *Chem. Rev. 58*, 715; Eisenberg, H. (1976) *Biological Macromolecules and Polyelectrolytes in Solution*, Oxford University Press, London]. The calculations have been carried out in most cases by the method of Chernyak & Magretova [Chernyak, V. Ya., & Margretova, N. N. (1975) *Biochem. Biophys. Res. Commun. 65*, 990], which does not require the knowledge of the protein concentration. The overall adequacy of this approach has been ascertained by using as standards histone H1 and the protamine thynnine, both of known molecular weight and different charge densities. An electrophoretic method for the rapid estimation of the molecular weights of this type of proteins is also given. The values obtained by this method, as well as those found either with the Scheraga–Mandelkern equation or from the sedimentation and diffusion constants, agree within experimental error with the values obtained from sedimentation equilibrium.

The molecular weight of the highly basic proteins present in sperm cells is not easily determined. Conventional sodium dodecyl sulfate ($NaDodSO_4$) electrophoresis (Shapiro et al., 1967) cannot be used, since protamines precipitate under these conditions. An alternative electrophoretic method has been developed in our laboratory (Colom & Subirana, 1979), but it has some uncertainties due to the lack of standards with variable percentages of lysine and arginine. Therefore, we decided to use the method of sedimentation equilibrium in spite of the inherent difficulties due to the high charge and small size of these proteins.

In order to test the accuracy of our approach, we used as standards in these studies several basic proteins with a well-known molcular weight. Once the accuracy of the method had been demonstrated, we applied it to several protamines and related proteins present in the spermatozoa of molluscs. At the same time we were able to optimize the conditions to be used in order to determine the molecular weights electrophoretically. We have complemented this study by determining the hydrodyanmic properties and circular dichroism of some of these proteins.

Methods and Materials

Preparation of Proteins. In these studies we have used histone H1, from calf thymus, and thynnine, the tuna fish protamine, as standard proteins in order to evaluate the correctness of our methods. The sperm protein ϕ_0 from the sea cucumber *Holothuria tubulosa* was also used as a standard, since its sequence became available in the course of these studies (A. Jordan et al., unpublished results). The other proteins studied were obtained from the sperm of different species of molluscs. Their purification has been described in detail elsewhere (Subirana et al., 1973; Colom & Subirana, 1979). In the case of the mussel *Mytilus edulis*, spermatozoa were extracted with 5% perchloric acid, and the resulting mixture of proteins was fractionated by column chromatography as described below. The amino acid composition of the proteins used is given in Table I.

Column Chromatography. The sperm proteins from *M. edulis* were fractionated on a ionic exchange column of Sephadex CM-25. The size of the column was 25 × 1.5 cm, and it was operated at a flow rate of 24 mL/h. The proteins were eluted with a NaCl gradient (1–2 M) in 50 mM sodium acetate adjusted to pH 6.7. Fractions of 4 mL were collected, and their absorbance was read at 230 nm. The results obtained are shown in Figure 1. The protein in the fractions was recovered by dialysis against 0.25 N HCl and precipitation with acetone.

Gel Electrophoresis. The method of Panyim & Chalkley (1969) at pH 2.3 was used, but the final concentration of acrylamide was increased to 12% and that of *N,N'*-methylenebis(acrylamide) to 0.3%. The concentration of urea was maintained at 6 M. Polymerized iridine was used as a molecular weight standard (Barfod & Larsen, 1976).

Ultracentrifugation. We used a Beckman Spinco Model E ultracentrifuge, provided with an optical interference system, focused at the half-plane of the cells. All experiments were carried out in the range 18–22 °C. For sedimentation equilibrium we used cells provided with either 12-mm Kel-F centerpieces with six channels or 12-mm double-sector Epon–aluminum centerpieces. In most cases a cushion of FC-43 fluorocarbon oil (3-M Co.) was used. Equilibrium was attained in about 24–48 h.

The plates were measured on a Nikon 6C profile projector. The average of several measurements was used in the calculations.

Before the protein solutions were introduced into the ultracentrifuge cells, they were extensively dialyzed against the adequate solvent by using Spectrapor 3 dialysis bags (Spectrum Medical Industries, Inc.). The dialyzed solutions were than

† From the Unidad de Química Macromolecular del C.S.I.C., Escuela T.S. de Ingenieros Industriales Diagonal, 999, Barcelona (28), Spain. *Received April 5, 1982.*

0006-2960/82/0421-5910$01.25/0

Table I: Amino Acid Composition of the Proteins Used in This Study[a]

amino acid	thynnine[b]	iridine[b]	histone H1[c] (calf thymus)	ϕ_0 protein[d] *H.t.* (sea cucumber)	*L.p.* protamine[e] (squid)	*G.d.* protamine[e] (marine snail)	*C.s.* Z protein[f]	*M.e.* ϕ1 protamine[e] (mussel)	*M.e.* ϕ3 protein[g] (mussel)
Lys			28.7	19.2		5.8	34.2	21.8	50.6
His								0.4	
Arg	62.9	66.1	1.7	23.1	78.0	56.3	19.2	28.8	4.6
Asp			2.0			0.3			0.5
Thr	2.9		5.4	2.6		2.0	3.1	4.0	2.2
Ser	7.0	12.3	6.7	12.8	12.1	17.1	7.6	17.0	9.8
Glu	1.8		3.4	2.6		0.2			0.6
Pro	5.9	9.1	10.1	6.4	2.4		9.5	6.5	9.5
Gly		6.0	6.9			4.9	3.7	6.5	0.5
Ala	7.0	1.0	25.1	26.9		9.6	20.5	13.9	19.7
$^1/_2$-Cys									
Val	9.0	4.5	4.1	5.1		3.5	2.1	1.0	1.9
Met									
Ile		1.0	0.8	1.3		0.1			
Leu			4.1						
Tyr	3.4		0.5		7.3				
Phe			0.5						
M	157.4	155.0	110.8	127.1	173.6	152.9	131.0	131.2	129.4
M_c	250.1	234.4	365.4	299.8	223.1	243.9	245.6	259.2	234.8
$\bar{v}_2$	0.682	0.66	0.730	0.711	0.667	0.678	0.720	0.701	0.738

[a] The table also includes for every protein the partial specific volume $\bar{v}_2$, the average molecular weight M of an amino acid residue, and the average molecular weight per positive charge M_c. The weight of the Cl^- ions associated with the basic residues is included in these values. [b] Ando et al. (1973) [c] Hnilica (1972). [d] *H.t.* = *Holothuria tubulosa*; Jordan et al. (unpublished results). [e] *L.p.* = *Loligo pealeii*; *G.d.* = *Gibbula divaricata*; *M.e.* = *Mytilus edulis*; Subirana et al. (1973). [f] *C.s.* = *Cryptochiton stellerii*; Colom & Subirana (1979). [g] Phelan et al. (1974).

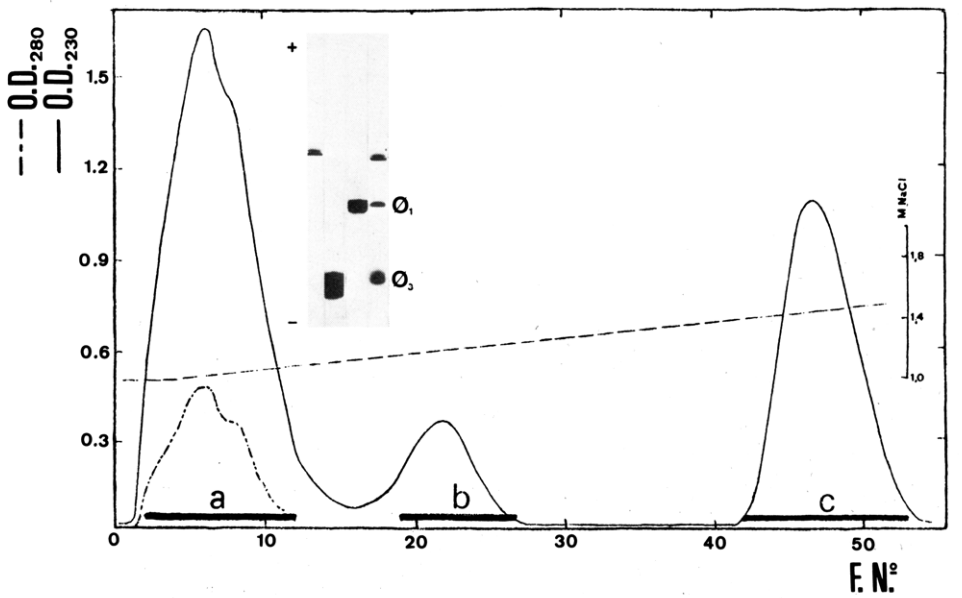

FIGURE 1: Column chromatography of whole sperm proteins from the mussel *M. edulis*, carried out under the conditions given in the text. The tubes indicated by a (histones), b (ϕ1 protamine), and c (ϕ3 protein) were pooled. The electrophoretic pattern of the protein recovered from them is shown in the insert.

centrifuged in a Sorvall centrifuge at 10000 rpm for 10 min. In general, the solvent was 0.25 M NaCl and 0.01 M NH_4Cl, pH 9.2, but in a few cases 6 M guanidinium chloride was used.

Sedimentation Equilibrium. All the proteins studied in this paper are homogeneous as shown by their electrophoretic behavior, and consequently they behave as pure thermodynamic components in the ultracentrifuge, as it will be discussed later. Therefore in the calculations there is no problem of determining average values, and any method used should give the molecular weight of the homogdisperse protein. The possibility of microheterogeneity in some of our proteins (Ando et al., 1973) cannot be excluded, but this fact would have no appreciable influence in the results obtained.

Due to the low molecular weight of the proteins studied, the low-speed method (Van Holde & Baldwin, 1958) was used in all cases. For comparative purposes, we also used the high-speed method (Yphantis, 1964) in some of the proteins which have a higher molecular weight. In two cases we also used the short column–low-speed method (Yphantis, 1960).

The calculations were carried out following the approach of Chernyak & Magretova (1975). The advantage of this method for homogeneous samples such as ours is that it does not require the initial concentration of protein in order to determine its molecular weight. In fact the experimental determination of the initial concentration is quite difficult in this case, since the proteins often tend to form insoluble aggregates. Furthermore, some of these proteins are hygroscopic and may contain low molecular weight contaminating salts. In any case these effects are small, since it was found that the total concentration determined from the fringe displacement

(assuming that 1 mg/mL gives a displacement of four fringes) was always very close to the initial concentration C_0 introduced in the cell.

This method is based on the fact that the concentration of protein y in the cell must follow an exponential equation:

$$y = y_0 + b \exp(qr^2/2) \quad (1)$$

where r is the distance to the axis of rotation and

$$q = M_2^{app}(1 - \bar{v}_2\rho_0)\omega^2/(RT) \quad (2)$$

where

$$\frac{1}{M_2^{app}} = \frac{1}{M_2} + 2A_2C_2 + \ldots \quad (3)$$

The experimental values of y are adjusted to expression 1 by searching with a minimum squares procedure the best value of q. The search was carried out in a FACOM computer. Once q is determined, the value of the apparent molecular weight M_2^{app} is determined from expression 2, where according to Eisenberg (Eisenberg, 1962, 1976; Eisenberg et al., 1978) and for a polyelectrolyte in a three-component system

$$M_2\left(\frac{\partial\rho}{\partial c_2}\right)_\mu = M_2^{app}(1 - \bar{v}_2\rho_0) = M_2[(1 - \bar{v}_2\rho_0) + \xi_3(1 - \bar{v}_3\rho_0)] = M_2(1 - \phi'\rho_0) \quad (4)$$

with

$$\xi_3 = \frac{M_3}{M_2}\left(\frac{\partial m_3}{\partial m_2}\right)_\mu = \frac{M_3}{M_c}\left(\frac{\partial m_3}{\partial m_c}\right)_\mu = \frac{M_3}{M_c}\Gamma \quad (5)$$

so that

$$\Gamma = \frac{(\bar{v}_2 - \phi')\rho_0}{(1 - \bar{v}_3\rho_0)}\frac{M_c}{M_3} \quad (6)$$

and

$$M_2^{app} = M_2\left[1 + \Gamma\left(\frac{M_3}{M_c}\right)\left[\frac{(1 - \bar{v}_3\rho_0)}{(1 - \bar{v}_2\rho_0)}\right]\right] \quad (7)$$

The components are defined according to Scatchard & Bregman (1959): water (component 1), PX_Z (component 2), and XY (component 3). Where P is a Z valent positive ion, X and Y are negative counter and positive co-ions, respectively. With this formulation $M_c = M_2/Z$. For ideal cases and when the main contribution in ξ_3 is due only to an ideal Donnan effect, then $\Gamma = -1/2$, so that eq 7 becomes the expression used by Williams et al. (1958) to correct for the secondary charge effect. In fact, in most of our calculations we will use expression 7 with $\Gamma = -1/2$ to correct the experimental values obtained from (2). We will show under Results that this procedure does not introduce any appreciable error.

In some cases a short column run (Yphantis, 1960) was carried out in parallel with the standard low-speed method. In that case the initial concentration C_0 was estimated from the expression

$$C_0 = 2(C_b - C_m)/[q(r_b^2 - r_m^2)] \quad (8)$$

where q was obtained as described above for the standard experiment. The value of the apparent molecular weight M_2^{app} was then calculated from

$$M_2^{app} = \frac{1}{C_0(r_b + r_m)/2}\left(\frac{dC_2}{dr}\right)_r \frac{RT}{\omega^2(\partial\rho/\partial c_2)\mu} \quad (9)$$

Determination of Protein Concentrations. The proteins were extensively dialyzed against distilled water. Hydrochloric acid was added to reach a 0.25 N concentration. The proteins were then precipitated with 6 volumes of acetone and dehydrated over P_2O_5 under vacuum during 48 h. The concentration was determined from the dry weight of the protein as chlorhydrate. Where it has been possible (i.e., histone H1 and thynnine), the extinction coefficient for tyrosine, ϵ_{278} = 1350 mol of Tyr L^{-1} cm^{-1} (Avilés et al., 1978), has also been used. In the latter cases agreement within ±1% was found in the results obtained when both methods were used simultaneously.

Determination of $(\partial\rho/\partial c_2)_\mu$ and Partial Specific Volumes. Determinations of $(\partial\rho/\partial c_2)_\mu$ were carried out on an Anton Paar digital density meter attached to one DMA 601 density measuring cell. Prior to the measurements, samples were extensively dialyzed at room temperature against the buffer for at least 48 h. Measurements were performed at 25 ± 0.01 °C.

Partial specific volumes were calculated from amino acid composition (Cohn & Edsall, 1941). This method is based on the ideas of Traube (1896) on the additivity of the molar volumes of atoms and atomic groups. Following the same procedure, in our calculations an HCl moiety was included in the definition of the component for every basic amino acid residue. In fact, at the pH 9.2 used, all arginines and about 80–90% of the lysines should be ionized, as shown by the titration curves obtained from charged polypeptides (Snell & Fasman, 1972). The chloride specific volume was taken to be 0.507 (Fajans & Johnson, 1942). In agreement with this procedure Applequist & Doty (1962) experimentally obtained a value of 0.722 for the specific volume of polylysine hydrochloride. This is considerably lower than the value of 0.82 used for the lysine residue alone. However, it is also significantly lower than 0.752 which is the result obtained from the separate specific volumes of lysine (0.82) and Cl^- (0.507). In our calculations we used the latter values, which gave more consistent results from the molecular weights. It should be noted that the difference in the final results is not very large. For example, in the case of *M. edulis* ϕ3, the protein with a higher lysine content, if the value of Applequist and Doty is used, $\bar{v}_2$ = 0.720 instead of 0.738 as given in Table I.

Furthermore, the value of $\bar{v}_2$ = 0.66, calculated with the same procedure, for iridine (see Table I) is also identical with the experimental $\bar{v}_2$ = 0.663 reported for iridine hydrochloride (Gehatia & Hashimoto, 1963). The convenience of using this approach will be discussed under Results.

When 6 M guanidinium chloride was used as a solvent, the values for every amino acid were adequately corrected (Lee & Timasheff, 1974), but this theoretical correction turned out to be negligible in most cases.

Sedimentation Coefficients. The sedimentation coefficients were determined from the rate of movement of the boundary maximum by using interference optics. The photographic plates were read on a Nikon profile projector. The following buffers were used, all of them containing 0.15 M NaCl: pH 2.1, 5.3 mM glycine and 14.7 mM HCl; pH 4.0, 18.3 mM glycine and 1.7 mM HCl; pH 6.0, 2.3 mM Na_2HPO_4 and 13.2 mM NaH_2PO_4; pH 7.0, 14.2 mM Na_2HPO_4 and 3.2 mM NaH_2PO_4; pH 8.3, 21 mM glycine and 1.1 mM NaOH; pH 9.2, 10 mM NH_4Cl adjusted with NH_4OH; pH 10.0, 28.8 mM glycine and 8.6 mM NaOH.

These buffers were also used in the circular dichroism experiments. The proteins were disolved in these buffer solutions and dialyzed against them for at least 24 h by using Spectrapor 3 dialysis tubing (Spectrum Medical Industries, Inc.). The

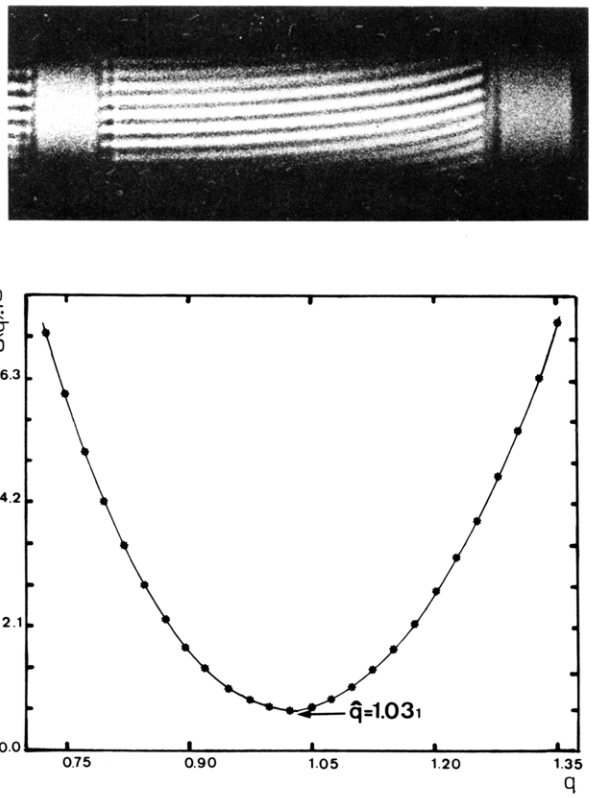

FIGURE 2: Interference pattern and the corresponding calculation to determine the optimal value of q as described in the text. The pattern shown correspond to tuna fish protamine (thynnine) in 0.25 M NaCl and 0.01 M NH_4Cl, pH 9.2, and $C_0 \simeq 0.4$ mg/mL and 44000 rpm.

protein concentration was determined from its absorption at 230 nm.

Intrinsic Viscosities. They were determined at 25 ± 0.01 °C with an Ubbelhode viscometer with a flow time of about 220 s for water solutions. When necessary, the reduced viscosity was corrected for absorption by the method of Eisenberg (1976).

Circular Dichroism. The circular dichroism spectra were obtained at 22 °C with a Mark III dichrograph (Jobin-Yvon). The cells had an optical path of 0.5 mm, and the sensitivity was set to 5×10^{-6} ΔA mm^{-1}. The results have been expressed in deg cm^2 (dmol of residue)$^{-1}$. The estimation of the amount of α helix present was carried out from the values of $[\theta]_{222}$, assuming −1000 for random coil and −25000 for 100% α helix (Bradbury et al., 1975).

Results

Sedimentation Equilibrium. The molecular weights obtained under the various conditions used by us are given on Table II. In most cases we used for the calculations the method of Chernyak & Magretova (1975), which does not require a knowledge of the original protein concentration. This procedure was very adequate since in many experiments we found that aggregates appeared in the ultracentrifuge cells and precipitated out when the run was started. As a result it was not possible to use the original concentration of protein in the calculations. The essence of this method is to find the value of q that gives the best straight line for the experimental points. The molecular weight M_2 is then calculated from expression 2 and corrected from (7) with $\Gamma = -1/2$. The value of q is determined by a computer procedure which may be represented in a plot such as the one shown in Figure 2. Once q is known, the experimental points can be represented as shown on Figure 3, where it can be seen that they fall very accurately

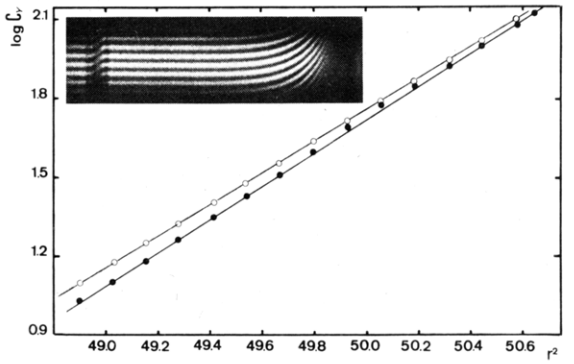

FIGURE 3: Plot of the logarithm of concentration vs. r^2 (expression 1) as directly determined from the experimental points (●) and as determined from the value of q (○). This experiment was carried out by the Yphantis method (1964) with the protein ϕ_0 from *H. tubulosa* at $C_0 \simeq 0.5$ mg/mL in 0.25 M NaCl and 0.01 M NH_4Cl, pH 9.2, at 56000 rpm. The inset shows the actual interference pattern obtained. The fringes are horizontal at the meniscus as required for the use of this method.

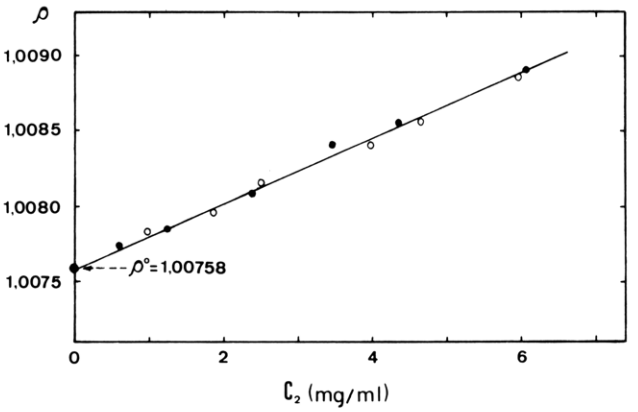

FIGURE 4: Plot of density vs. concentration for the evaluation of $(\partial\rho/\partial c_2)_\mu$ for thynnine (○) and histone H1 (●). In both cases, concentrations are expressed in grams of protein clorhydrate per liter of solution (the solvent used was 0.25 M NaCl and 0.02 M NH_4Cl, pH 9.2).

on straight lines. This result shows that the method of calculation is adequate and confirms that the protein preparations are pure. Furthermore, this method gives molecular weight values which are essentially identical with those found by the method of Yphantis, as shown in Table II. An additional proof of the correctness of our approach is given by the fact that in those cases in which the molecular weight is known from sequence determination (thynnine, histone H1, ϕ_0), the values found by us coincide with them, as shown in Table III.

In order to test more precisely the validity of the calculations carried out by us, in particular the use of $\Gamma = -1/2$ in eq 7, we determined the value of $(\partial\rho/\partial c)_\mu$ for two of the proteins with known sequence used, namely, histone H1 and thynnine. These two proteins approximately correspond to both ends of the range of properties (Arg percentage, size, $\bar{v}_2$, etc.) of the proteins studied by us. The results obtained are shown in Figure 4. The value of $(\partial\rho/\partial c)_\mu$ turned out to be the same for both proteins and equal to 0.217. Since in both cases c_2 was expressed as protein chlorhydrate, the molecular weight obtained should also correspond to the protein chlorhydrate. As a matter of fact, when this value was used directly in expressions 4 and 2, the values of M_2 obtained for histone H1 and thynnine were 23400 and 5500, respectively. They are in good agreement with the values 23994 and 5350 determined from the sequence, once the chloride ions are taken in account. They are also in good agreement with the values given in Table

Table II: Molecular Weights Obtained by Sedimentation Equilibrium[a]

protein	method	speed (rpm × 10^{-3})	C_0 (mg/mL)	M_Y^{app}	q	M_2^{app}	$\overline{M_2^{app}}$	M_2	M^{app*}	M^*
thynnine	L	36	6		0.67	3 750				
	L	40	4		0.86	3 850	3 850	*5 350*	4 600	5 350
	L	44	0.4		1.02	3 850				
histone H1	L	16	3.0		0.59	19 200				
	L	16	2.6		0.60	19 700				
	L	16	2.2		0.65	21 300				
	Y	40	1.0	18 000	3.61	18 900	19 200	*24 400*	22 800	25 900
	Y	40	0.5	19 100	3.54	18 500				
	Y	40	0.4	17 600	3.26	17 300				
	Y	40	0.25	19 900	3.60	18 900				
ϕ_0 protein from	L	32	2		0.89	6 900				
H. tubulosa	L	30	2		0.81	7 400	7 100	*9 400*	9 400	10 500
	Y	56	0.5	7 400	2.71	7 000				
L. pealeii	L	36	2		1.56	8 250				
protamine	L-S	36	2		(1.56)	8 700	8 700	*12 200*	11 100	13 000
	L	32	1.6		1.32	8 800				
	L	40	0.5		2.01	8 600				
G. divaricata	L	36	1.2		0.99	10 200				
protamine	L-S	36	1.2		(0.99)	10 200	10 200	*13 900*		
	Y	56	0.6	10 300						
	Y	52	0.4	9 800						
C. stellerii	L	18	2		0.38	9 700				
Z protein	L	30	2		0.78	7 300				
	L	40	0.5		1.64	8 450	8 350	*12 100*		
	L	40	0.4		1.62	8 350				
M. edulis ϕ3	L	44	1.3		1.62	7 200				
protein	L	44	0.6		1.53	6 900	7 100	*10 750*	12 300	16 000
	L	36	0.5		1.09	7 200				
M. edulis ϕ1	L	30	1		1.11	9 600	9 600	*13 200*		
protein										

[a] The methods used were low speed (L) (Van Holde & Baldwin, 1958), high speed (Y) (Yphantis, 1964), and low speed in a short column (L-S) (Yphantis, 1960). M_Y^{app} is the apparent molecular weight determined by the high-speed method. $\overline{M_2^{app}}$ is the average of all the values determined. M^{app*} and M^* are the values obtained in 6 M guanidinium chloride. The other symbols are defined in the text (formulas 2 and 7).

Table III: Comparison of the Molecular Weights Obtained by Different Methods[a]

protein	basic (%)	N	N_s	N_e	N_H
thynnine	62.9	34	33	ref	36
iridine	66.1	32.5		ref	32
histone H1	30.5	215	220	219	187
H. tubulosa ϕ_0 protein	42.3	78	74	78	
L. pealeii protamine	78		70	72	
G. divaricata protamine	62		91		
C. stellerii Z protein	53.4		92		
M. edulis ϕ1 protein	50.6		100	111	92
M. edulis ϕ3 protein	55.2	46	83	81	82

[a] The data on the first two columns were obtained from the references given in Table I. N = number of amino acid residues per molecule determined from the sequence or by chemical methods. N_s = number of amino acid residues per molecule determined from equilibrium sedimentation. N_H = number of amino acid residues per molecule determined from the expression of Scheraga & Mandelkern (1953) using $\beta = 2.0$. N_e = number of amino acid residues per molecule determined from the electrophoretic mobility. ref = reference.

II obtained from expression 7 with $\Gamma = -1/2$.

Using eq 6 we have also evaluated the parameter Γ. In that equation, the value 0.777 cm^3/g was assigned to ϕ' calculated through eq 4 from the value of $(\partial\rho/\partial c_2)_\mu$ experimentally determined. According to the definition of c_2 employed, the partial specific volume $\bar{v}_2$ was also calculated as chlorhydrate (see Methods and Materials). The values of v_2 and M_c shown in Table I were used in eq 6. The value of Γ turned out to be −0.43 for histone H1 and −0.59 for thynnine. These values are very close, within experimental error, to the value $\Gamma = -0.5$ expected for the ideal case. The molecular weights shown in Table II were therefore determined y using this latter value, as it was impossible to determine $(\partial\rho/\partial c_2)_\mu$ for all the other proteins studied in this paper, since only small amounts of pure proteins were easily available.

Inspection of the values reported in Table II shows that there is no appreciable effect of concentration on the value of the molecular weights determined, probably due to the low concentrations of protein used, below 2 mg/mL in most cases.

The values in Table II also show that the low-speed method gives the same result as the high-speed method in those cases in which it can be used. The two experiments carried out in short columns also give coincident results. Most individual determinations deviate less than 5% from the average value found for the molecular weight. An exception is the Z protein from *C. stellerii*, in which some values deviate about 15% from the average. This observation may be due to the presence of a higher molecular weight impurity in this sample (Colom & Subirana, 1979).

The molecular weights obtained in 6 M guanidinium chloride are also shown in Table II. These experiments were carried out in order to determine if there were any signs of aggregation in the pH 9.2 buffer used in most experiments, since this pH is close to the isoelectric point, in particular, in the lysine-rich proteins. In most cases the molecular weights obtained give slightly higher values, but in *M. edulis* ϕ3 the difference is remarkable. We interpret these results as due to an inadequate determination of the specific volumes in this

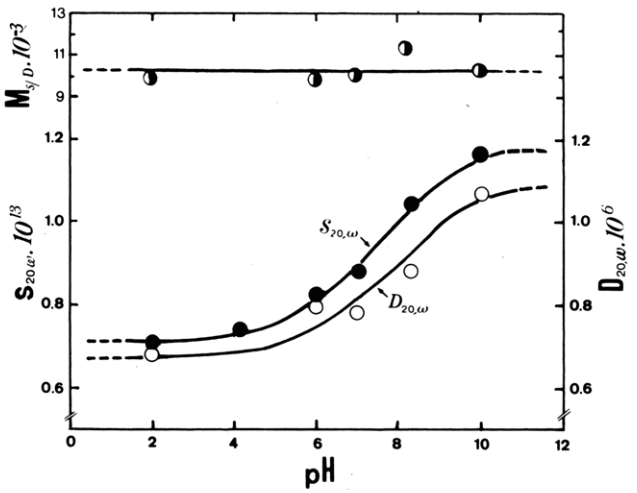

FIGURE 5: Influence of pH on the sedimentation and diffusion constants of the $\phi 3$ protein at C_0 = 5.1 mg/mL. The molecular weight calculated from these constants (Tanford, 1961) is also shown in the figure.

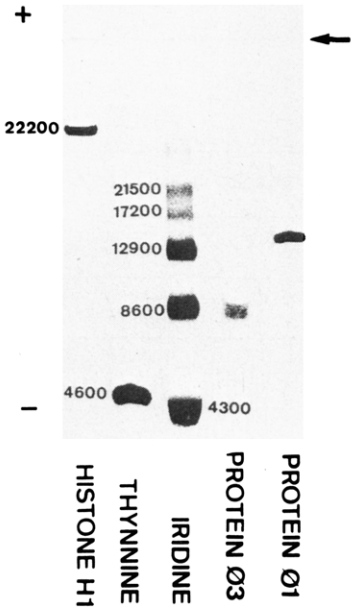

FIGURE 6: Gel electrophoretic pattern of some of the proteins studied in this work. The molecular weights (excluding the chloride ions) are shown on the figure. The arrow indicates the origin.

solvent, so that the molecular weight values obtained under these conditions are not reliable. It should be noted that the *M. edulis* $\phi 3$ protein, which shows a higher deviation, is also very rich in lysine (50.6%). In any case the values observed in this solvent indicate that formation of specific aggregates is unlikely in the pH 9.2 buffer used in most determinations.

In order to ascertain the molecular weight for the case of the $\phi 3$ protein in which the largest discrepancy appeared with the guanidinium chloride values, we also obtained a molecular weight of 10 300 by the s/D method, as shown in Figure 5. An excellent agreement with the value obtained by sedimentation equilibrium is thus found.

Electrophoresis. Once the molecular weights of the proteins had been determined by sedimentation equilibrium, it appeared of interest to compare the results obtained with those found by gel electrophoresis (Colom & Subirana, 1979). For this purpose we used the acid–urea gel system of Panyim & Chalkey (1969), with a slightly higher polyacrylamide concentration. A typical electrophoretic pattern is shown in Figure 6. No aggregates appear in any of the proteins studied.

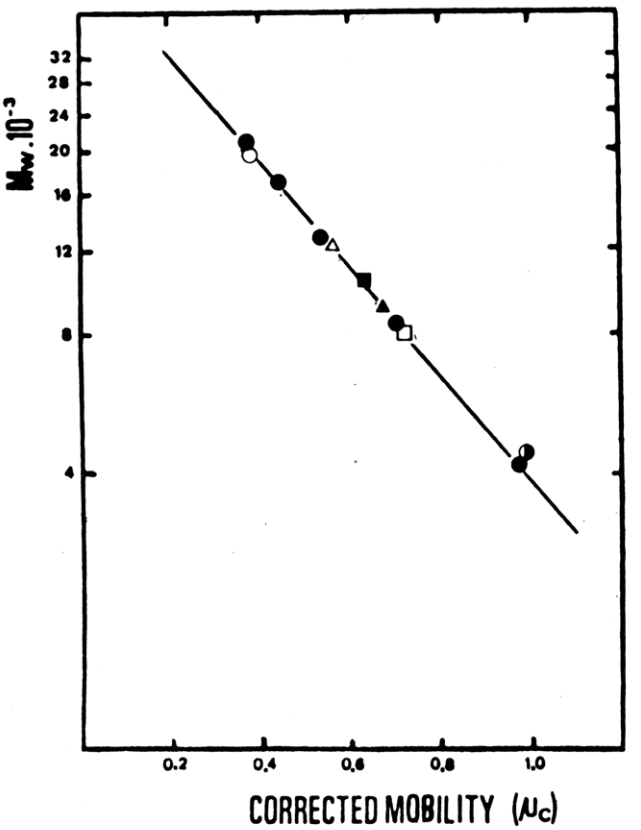

FIGURE 7: Determination of molecular weights by electrophoresis. The value for the electrophoretic mobility was corrected as described in the text (formula 10) in order to take into account the differences in charge density in the various proteins studied. The following proteins were taken as standards in order to draw the straight line represented in the figure: polymerized iridine (●), thynnine (◐), and histone H1 (○). The values for the following proteins were determined from the line: *H. tubulosa* ϕ_0 (□), squid protamine (■), *M. edulis* $\phi 1$ (△), and *M. edulis* $\phi 3$ (▲).

If under the conditions used the proteins lack any globular structure, it can be assumed that all have a similar stiff random coil structure. The influence of charge can then be empirically corrected by the expression

$$u_c = u\frac{a_r M}{a M_r} \tag{10}$$

where u is the measured mobility, a is the percent of basic amino acids, and a_r is the corresponding value for the protein used as a reference (in this case iridine, a_r = 66.1%). M and M_r are the average molecular weights per amino acid residue, calculated from the amino acid compositions. The values of M for each protein are given in Table I. Due to the different residue weights of lysine and arginine, M takes into account the different mobilities of proteins rich in either of these two amino acids.

The results obtained were represented in a semilogarithmic plot, as shown in Figure 7. The proteins used as standards fall on a straight line, thus showing that correction according to expression 10 is appropriate. It is worth noting that histone H1 falls on the same line that the iridine oligomers, in spite of the very different amino acid compositions. This observation indicates that histone H1 at the low pH (2.3) of these experiments has a conformation similar to that of the protamines. Other histones (results not shown) do not fall on the straight line of Figure 7, probably due to a preservation of globular structure in spite of the low pH used in the electrophoretic buffer.

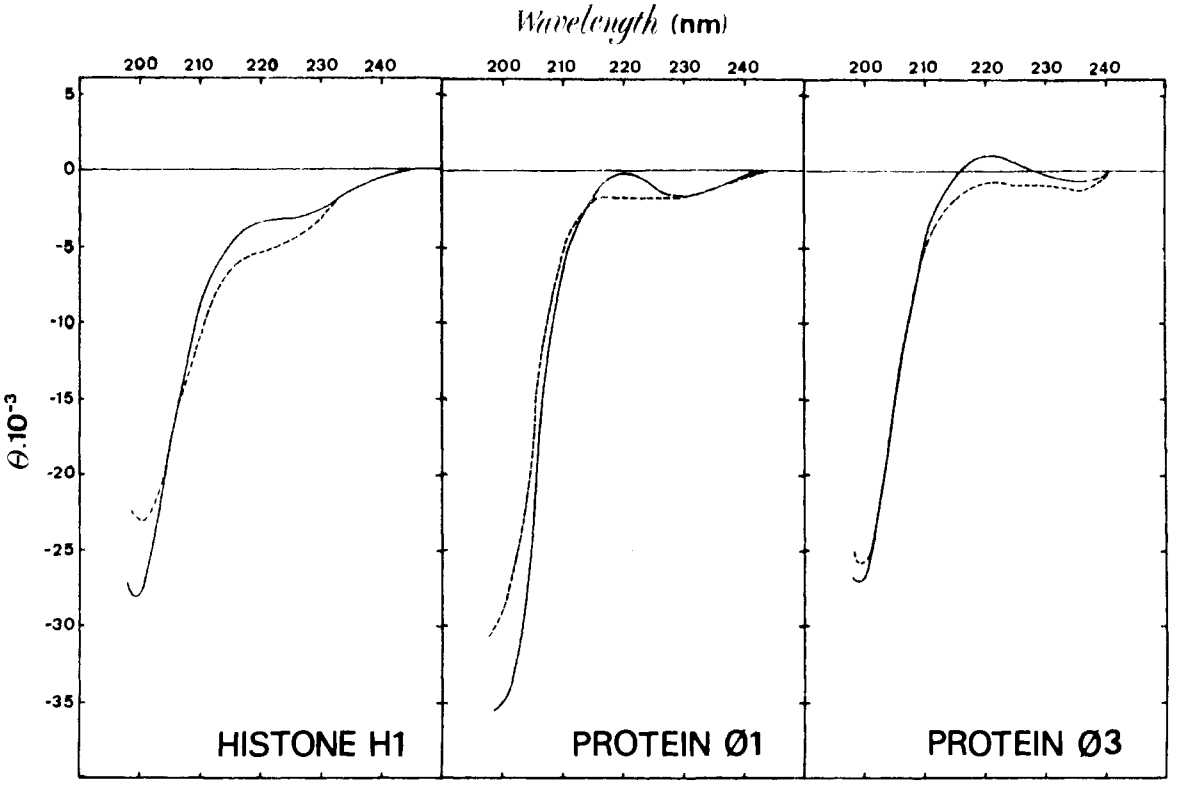

FIGURE 8: Circular dichroism spectra at pH 2 (continuous line) and 9.2 (dotted line) of the proteins indicated on the figure.

We carried out some electrophoretic runs at basic pH, about 9.2, and found that the lysine-containing proteins ran faster with respect to the arginine-containing ones. The relative difference in mobility was roughly proportional to the amount of lysine present. It appears that under these conditions there is a change in either charge, conformation, or both in the lysine-containing proteins.

The molecular weight of the other proteins was calculated from the plot shown in Figure 7 and is given in Table III. The values obtained are in very good agreement with those obtained by sedimentation equilibrium. The use of expression 10 appears to give a satisfactory correction for charge density in these basic proteins. The values obtained are therefore more accurate than those reported by Colom & Subirana (1979), who did not use any correction for charge density.

Hydrodynamic Properties. The values of the intrinsic viscosity and sedimentation coefficient are given in Table IV. From these values the molecular weight of the protein can be determined with the Scheraga–Mandelkern equation (1953). The results obtained are given in Table III, and it can be seen that there is a reasonable agreement with the values obtained by the other methods. The value of β was slightly lower than usual, but it is known that the value of this parameter depends on the conformation of the proteins. The radii of gyration of the proteins assumed to be random coils are also given in Table IV. All of them show reasonable values for such conformation.

In the case of the lysine-rich protein $\phi 3$ we also obtained the sedimentation and diffusion coefficients as a function of pH, as shown in Figure 5. It appears that there is a transition to a more compact structure as the pH is raised. The molecular weight remains constant throughout this transition.

Circular Dichroism. The spectra obtained for some of the proteins studied are given in Figure 8. All of them have the shape typical for a random coil. Only in the case of histone H1 can a small amount of structure be detected. The percentage of α helix which can be determined from these curves is very low, as shown in Table IV. However, when the ellipticity is studied as a function of pH as shown in Figure 9, it is clear that a small amount of structure appears in all of these proteins at basic pH, but the nature of this structure cannot be determined from these experiments.

Table IV: Hydrodynamic and Conformational Parameters at pH 9.2[a]

protein	$s^0_{20,w}$ (S)	$[\eta]$ (cm³/g)	R_G (Å)	α helix (%)
iridine[b]	0.74	8.16	24.1	0
thynnine	0.74[c]	8.7	25.2	0
$\phi 1$ protein	1.12	9.2	34.5	4
$\phi 3$ protein	1.10	5.2	26.7	0
histone H1	1.3[d]	13.0	47	15

[a] The value of R_G was determined from $[\eta]$ (Tanford, 1961). The amount of α helix was calculated as described by Bradbury et al. (1975). [b] Values taken from Gehatia & Hashimoto (1963). [c] Not determined; assumed to be equal to iridine. [d] Taken from Hartman et al. (1977).

Discussion

In this paper we have applied for the first time the method of sedimentation equilibrium to the determination of the molecular weight of several highly charged protamines and related proteins. The recent theoretical treatment developed for polyelectrolytes in multicomponent solutions (Eisenberg, 1976) allows us to determine the molecular weight of charged biological polymers with high accuracy. Once $(\partial\rho/\partial c_2)_\mu$ is consistently defined and determined properly and according to formula 4, there is no ambiguity in the value obtained for the molecular weight. In our case we have always worked with the protein chlorhydrates, so that the molecular weights and $\bar{v}_2$ include the contributiion of the chloride ions which neutralize the protein charges. In expression 7, usually called the correction for the "secondary charge effect", we have found that the ideal value of $\Gamma = -1/2$ is sufficiently accurate for the proteins we have studied, as shown under results. Previous

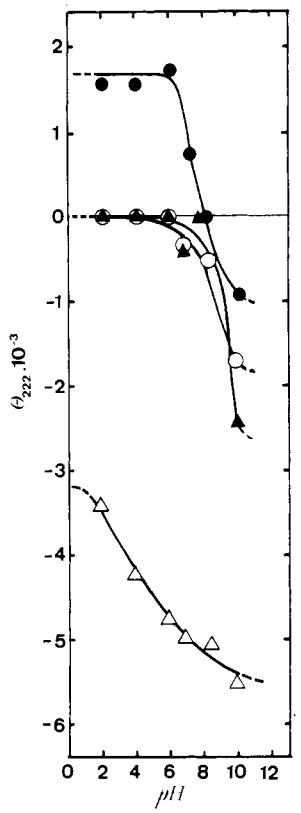

FIGURE 9: Influence of pH on the ellipticity at 222 nm of the following proteins: $\phi 3$ (●), thynnine (▲), $\phi 1$ (○), and histone H1 (△).

results obtained with histones (Teller et al., 1965; Haydon & Peacocke, 1968; Edwards & Shooter, 1969) were reasonably accurate, although they were subjected to some uncertainty due to the corrections used for what was called the "charge effect", which can now be more clearly understood by using the methodology of Eisenberg (1976).

As shown by the circular dichroism and hydrodynamic studies reported here, these proteins behave as random coils. The electrostatic effects are very important in these materials, and at first we thought that they would have a large influence on the results obtained, but it turns out that this is not so if appropriate conditions are chosen. An important finding was that the concentration of protein did not have any significant influence on the molecular weights in the range of concentrations we used, which were unusually below 2 mg/mL as shown in Table II. This was an unexpected result, given the highly charged nature of these proteins, which should result in an extremely nonideal behavior. The primary charge effect should result in a strong influence of concentration on the apparent molecular weight. The fact that such influence was not observed indicates that the simplified expression of Williams et al. (1958, eq 62) grossly overestimates this effect. In fact, these authors recognize that there is no theory which takes into account the thermodynamic nonideality of such a system. Other authors (Erlander & Senti, 1963) have also found a much lower concentration dependence than expected in different types of proteins. Furthermore, more recently several virial coefficients have been established for DNA (Borochov et al., 1981), which also has a high charge density. In 0.2 M NaCl a value of 5×10^{-7} L mol g^{-2} has been reported for the second virial coefficient A_2. Assuming a molecular weight of 10000, and choosing $c_2 = 5$ mg/mL (which are values similar to those used throughout the present work), it turns out that the concentration dependence is far below the uncertainty due to experimental error. It appears that the salt concentrations used are sufficient to screen and charges on the proteins studied as to lower the second virial coefficient A_2 in eq 3 to negligible values.

In summary the values obtained for the three standard proteins were rather accurate. Errors of 3% (thynnine), 2% (H1), and 5% (ϕ_0) were detected as shown in Table III. The molecular weights obtained for the other proteins should have errors of a similar magnitude. This degree of accuracy demonstrates that the method of calculation suggested by Chernyak & Magretova (1975) is adequate for pure proteins. This method allows the determination of molecular weights without requiring a knowledge of the original concentration of protein, and it should find a wider use for homogeneous proteins such as the ones we have studied here.

As shown in Table II, the molecular weights were also determined in 6 M guanidinium chloride as a solvent. The values M^* obtained are much higher than expected, particularly in the proteins with a higher lysine content. These results are probably due to an inadequate calculation of $\bar{v}_2$ under these conditions.

On the other hand, the molecular weights determined by electrophoresis and by hydrodynamic methods were in good agreement with the values calculated by sedimentation equilibrium as shown in Table III. The electrophoretic method used here constitutes an improvement over the method of Colom & Subirana (1979), since it allows a correction for lysine-containing proteins. In the case of the $\phi 3$ protein, we also obtained the molecular weight by the s/D method, finding an excellent agreement with the value obtained by sedimentation equilibrium.

An unexpected result of this work was that the molecular weight of the $\phi 3$ protein from *M. edulis* was found to be about twice the value determined by chemical methods (Phelan et al., 1974), as shown in Table III. This difference may be due to the microheterogeneity detected in this protein (Ausió, 1980). Sequence studies presently in progress may help to resolve the origin of this discrepancy.

Acknowledgments

Suggestions and discussion with Professor H. Eisenberg were essential for an adequate presentation of the sedimentation equilibrium results. We are thankful to N. Barfod, G. Bretzel, J. Colom, J. Cortadas, and B. Larsen for providing us with some of the proteins used in this study, to A. B. Martínez for his help with the computer programs, and to J. M. Ripol for this valuable technical assistance. We are also thankful to Professor A. M. Municio and his collaborators for their assistance in the circular dichroism experiments. The measurements of $(\partial\rho/\partial c_2)_\mu$ were carried out at the Polymer Department, Weizmann Institute of Science, with the skillful and enthusiastic technical assistance of Y. Haik.

References

Ando, T., Yamasaki, M., & Suzuki, K. (1973) *Protamines*, Springer-Verlag, New York.

Applequist, J., & Doty, P. (1962) in *Polyamino acids, Polypeptides and Proteins* (Stahman, M. A., Ed.) The University of Wisconsin Press, Madison.

Ausió, J. (1980) Ph.D. Thesis, University of Barcelona, Faculty of Biology.

Avilés, J., Chapman, G. E., Kneale, G. G., Crane-Robinson, C., & Bradbury, E. M. (1978) *Eur. J. Biochem. 88*, 363.

Barfod, N. M., & Larsen, B. (1976) *Biochim. Biophys. Acta 427*, 197.

Borochov, N., Eisenberg, H., & Kam, Z. (1981) *Biopolymers 20*, 231.

Bradbury, E. M., Cary, P. D., Chapman, G. E., Crane-Robinson, C., Danby, S. E., & Rattle, H. W. E. (1975) *Eur. J. Biochem. 52*, 605.
Chernyak, V. Ya., & Magretova, N. N. (1975) *Biochem. Biophys. Res. Commun. 65*, 990.
Cohn, E. J., & Edsall, J. T. (1941) *Proteins, Amino acids and Peptides*, p 372, Reinhold, New York.
Colom, J., & Subirana, J. A. (1979) *Biochim. Biophys. Acta 581*, 217.
Edwards, P. A., & Shooter, K. V. (1969) *Biochem. J. 144*, 227.
Eisenberg, H. (1962) *J. Chem. Phys. 36*, 1837.
Eisenberg, H. (1976) *Biological Macromolecules and Polyelectrolytes in Solution*, Oxford University Press, London.
Eisenberg, H., Haik, Y., Ifft, M., Leicht, W., Mevarech, M., & Pundak, S. (1978) in *Energetics* and *Structure* of *Halophilic Microorganisms* (Caplan, S. R., & Ginburg, M., Eds.) Elsevier/North-Holland Biomedical Press, New York.
Erlander, S. R., & Senti, F. R. (1963) *Makromol. Chem. 65*, 31.
Fajans, K., & Johnson, J. O. (1942) *J. Am. Chem. Soc. 64*, 668.
Gehatia, M., & Hashimoto, C. (1963) *Biochim. Biophys. Acta 69*, 212.
Hartman, G., Chapman, G. E., Moss, T., & Bradbury, M. E. (1977) *Eur. J. Biochem. 77*, 45.
Haydon, A. J., & Peacocke, A. R. (1968) *Biochem. J. 110*, 243.
Hnilica, L. S. (1972) *The Structure and Biological Functions of Histones*, CRC Press, Cleveland, OH.
Lee, J. C., & Timasheff, S. M. (1974) *Arch. Biochem. Biophys. 165*, 268.
Panyim, S., & Chalkley, R. (1969) *Arch. Biochem. Biophys. 130*, 337.
Phelan, J. J., Colom, J., Cozcolluela, C., Subirana, J. A., & Cole, R. D. (1974) *J. Biol. Chem. 249*, 1099.
Scatchard, G., & Bregman, J. (1959) *J. Am. Chem. Soc. 81*, 6095.
Scheraga, H. A., & Mandelkern, L. (1953) *J. Am. Chem. Soc. 75*, 179.
Shapiro, A. L., Viñuela, E., & Maizel, J. V. (1967) *Biochem. Biophys. Res. Commun. 28*, 815.
Snell, C., & Fasman, G. D. (1972) *Biopolymers 11*, 1723.
Subirana, J. A. Cozcolluela, C., Palau, J., & Unzeta, M. (1973) *Biochim. Biophys. Acta 317*, 364.
Tanford, C. (1961) in *Physical Chemistry of Macromolecules*, Wiley, New York.
Teller, D. C., Kinkade, J. M., & Cole, R. D. (1965) *Biochem. Biophys. Res. Commun. 20*, 739.
Traube, J. (1896) *Liebigs Ann. Chem. 290*, 43.
Van Holde, K. E., & Baldwin, R. L. (1958) *J. Phys. Chem. 62*, 734.
Williams, J. W., Van Holde, K. E., Baldwin, R. L., & Fujita, H. (1958) *Chem. Rev. 58*, 715.
Yphantis, D. A. (1960) *Ann. N.Y. Acad. Sci. 88*, 586.
Yphantis, D. A. (1964) *Biochemistry 3*, 297.

Detection and Characterization Using Circular Dichroism and Fluorescence Spectroscopy of a Stable Intermediate Conformation Formed in the Denaturation of Bovine Carbonic Anhydrase with Guanidinium Chloride†

Robert W. Henkens,* Barbara B. Kitchell, S. Chace Lottich, Paul J. Stein,‡ and Taffy J. Williams

ABSTRACT: Particularly stable elements of noncovalent structure in bovine carbonic anhydrase have been detected and studied. These are present in a highly populated intermediate state formed during denaturation of the enzyme with guanidinium chloride. The intermediate has been detected by analysis of the denaturation profiles, and some of its structural properties have been characterized by CD and fluorescence spectroscopy, including fluorescence polarization and lifetime measurements. Measurements have been made on the Zn^{2+}–enzyme, Co^{2+}–enzyme, and apoenzyme to ascertain the structural effects of the active-site Zn^{2+}. Kinetic measurements indicate that this intermediate is on the folding pathway from the random coil to the native state.

Of special interest in understanding the factors that determine the three-dimensional structure of a globular protein is the detection and identification of especially stable elements within the overall structure. One approach to this problem comes in studying equilibrium denaturation of the protein, since any intermediates found in the presence of strong denaturants necessarily contain stable elements of secondary or tertiary structure or both. In this paper, we report a spectroscopic analysis (CD and fluorescence)[1] of the guanidinium chloride denaturation of bovine carbonic anhydrase with the aim of detecting and studying stable elements of enzyme structure.

Bovine carbonic anhydrase is a zinc-containing protein composed of a single polypeptide chain that is folded without disulfide cross-links into a compact, globular structure with

† From the Department of Chemistry (R.W.H., P.J.S., and S.C.L.) and the Department of Surgery (B.B.K.), Duke University, Durham, North Carolina 27706, and the Naval Medical Research Institute (T.J.W.), Bethesda, Maryland 20814. *Received September 22, 1981; revised manuscript received June 17, 1982.* This material is based on work supported by the National Science Foundation under Grant PCM 7609329, National Institutes of Health Grant GM25753, and the Naval Medical Research and Development Command (Research Task No. MR 04120001.0435 and MF 58524013.1044). The opinions or assertions contained herein are personal and are not to be construed as official or reflecting the Navy Department or naval service at large.

‡ Present address: Department of Chemistry, College of St. Scholastica, Duluth, MN 55811.

[1] Abbreviations: GdmCl, guanidinium chloride; CD, circular dichroism; DNSA, (dimethylamino)naphthalenesulfonamide.

0006-2960/82/0421-5918$01.25/0