methanol and the combined filtrates evaporated. Thin layer chromatography of the residue, 10 mg., revealed the presence of both isodrimenin and confertifolin. Crystallization of the solid mixture from hexane and sublimation (120° (1 mm.)) afforded needles of confertifolin, m.p. 152.5–153.5° (lit.[2b] m.p. 152°), m.m.p. 152–153°; infrared spectrum identical with that of an authentic sample.[26]

Drimenol (IIIa).—A solution of 21 mg. of the diol IIIb, prepared by the reduction of drimenin,[2b] and 3 mg. of sodium acetate in 5 ml. of acetic anhydride was heated on the steam bath for 0.5 hr. The solution was evaporated under vacuum and a solution of the residue in 6 ml. of sodium-dried tetrahydrofuran and 15 ml. of liquid ammonia was treated with 18 mg. of lithium wire and stirred for 1.5 hr. The solvents were allowed to evaporate at room temperature, the residue dissolved in ether, and the resulting solution washed with water, dried (magnesium sulfate), and evaporated. A solution of the residual oil, 20 mg. in 7 ml. of 5% sodium hydroxide and 10 ml. of 95% ethanol, was refluxed for 1 hr. The mixture was concentrated and extracted with ether. The extract was dried (magnesium sulfate), evaporated, and the residual oil chromatographed on a 5-mm. thick silica plate. The chromatogram was developed with 19:1 benzene–ethyl acetate, the positions of the products determined by their fluorescence under ultraviolet light and the products isolated by the extraction of individual portions of the silica plate with the aforementioned solvent pair. Filtration and evaporation of one eluate and sublimation (70° (1 mm.)) of the resultant oil, 9 mg., gave 8 mg. of a solid whose crystallization from hexane afforded drimenol, m.p., m.m.p. 94–95° (lit.[2a] m.p. 97–98°) spectra: infrared [(KBr) OH 3.15 (m, broad), C=C 5.98 (w)μ] identical with that of an authentic sample[26]; p.m.r., 9-proton singlet 0.88 p.p.m. (3 C—Me); 3-proton multiplet 1.80 p.p.m. (C_8-Me); 2-proton multiplet 3.70–3.86 p.p.m. (hydroxymethyl); 1-proton multiplet 5.45–5.62 p.p.m. (C_7-H).

[CONTRIBUTION FROM THE DEPARTMENT OF BIOCHEMISTRY, DUKE UNIVERSITY MEDICAL CENTER, DURHAM, NORTH CAROLINA]

Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions[1]

BY CHARLES TANFORD

RECEIVED NOVEMBER 20, 1963

The theoretical equations for the unfolding of globular proteins, presented in an earlier paper, are applied to the unfolding produced by the action of urea. The course of unfolding is shown to depend primarily on interactions with the solvent, which can be determined from solubility studies on model compounds. Since appropriate solubility studies have been carried out in urea solutions, it is possible to make a quantitative comparison between theoretical and experimental unfolding curves in urea. The theoretical calculation depends on an estimate of the fraction of hydrophobic and peptide groups which are buried in the native structure and exposed during unfolding. With reasonable estimates of this fraction, it is found that, for most proteins, the observed course of unfolding in urea is incompatible with calculations based on an all-or-none reaction. Stable intermediate states must occur between the native and fully unfolded conformations. The number of such intermediate states is small for β-lactoglobulin: for this protein the possibility of an all-or-none process is not excluded entirely. In ribonuclease the data suggest the presence of about three regions able to unfold independently. In larger protein molecules the number of intermediate states may be even larger. The urea concentration at the midpoint of the experimental transition curve gives a value for the difference in free energy ($\Delta F_{u,H_2O}°$) between native and unfolded proteins in water. This result does not depend on whether stable intermediate forms occur during unfolding. The value of $\Delta F_{u,H_2O}°$ at 25°, for several proteins, is found to lie in the neighborhood of 100–200 cal./mole per amino acid residue.

It is the purpose of this paper to account quantitatively for the unfolding of globular proteins by aqueous urea solutions, using the general theoretical treatment of unfolding proposed in an earlier paper.[2] It will be seen that the calculations required can be made with considerably greater certainty than those presented in the earlier paper, because fewer assumptions are necessary. In the earlier paper the objective was to calculate the difference in free energy between a native globular protein conformation and an unfolded conformation of the same protein. It was necessary to make assumptions about the detailed structure within the globular form, and about the degree of flexibility in the unfolded form. For the present calculations, a much less detailed model is necessary; only the effect of solvent composition on the free energy of unfolding is required, and to make this calculation one only has to know which parts of the molecule are in contact with the solvent in both the globular and unfolded forms. The actual structure within the globular conformation does not enter into the calculation.

As was true of the earlier paper, the quantitative estimates of free energies of interaction with solvent are based on solubility studies of model compounds. In this paper, the solubilities of amino acids and related compounds in aqueous urea solutions, determined in this laboratory,[3] will be employed.

Theory

A schematic representation of the native, compactly folded conformation of a globular protein is shown in Fig. 1. It is a linear chain of –NH–CH–CO– groups (rectangles) from which side chains (circles) project. Some of the peptide units and side chains, shown in black, are in contact with the solvent, but many more, shown in white, are assumed to be within the globular structure and to have no contact with solvent. It will be assumed that only a single conformation of this kind exists in the water–urea solvent system at a given temperature, and that practically all molecules possess this conformation when no urea is present.

A similar schematic representation of a highly unfolded protein molecule also is shown in Fig. 1. The polypeptide chain in this state of the molecule is assumed to be flexible, so that the relative positions of various parts of the chain do not remain fixed: the figure may be taken to represent a time average. From the point of view of the present paper, the important feature of the unfolded molecule, as contrasted with the native form, is that nearly all of its parts are in

(1) Supported by research grant G-17477 from the National Science Foundation and by research grants AM-04576 and GM-K6-14,222 from the National Institutes of Health, U. S. Public Health Service.

(2) C. Tanford, *J. Am. Chem. Soc.*, **84**, 4240 (1962).

(3) Y. Nozaki and C. Tanford, *J. Biol. Chem.*, **238**, 4074 (1963).

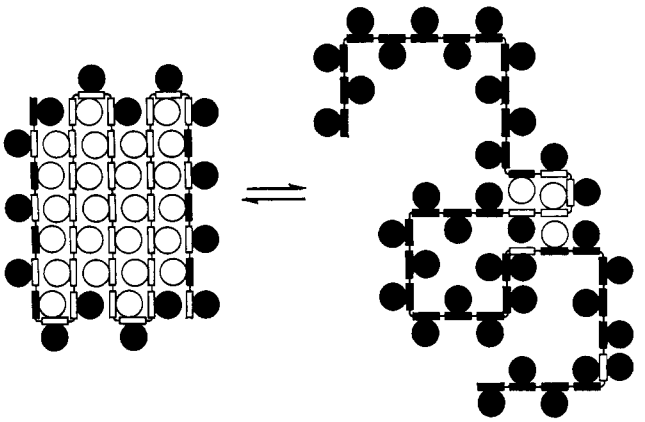

Fig. 1.—Schematic diagrams of the native and unfolded states of a protein molecule. Rectangles represent peptide groups and circles represent side chains, which in a real protein would of course vary in size. The rectangles and circles shown in black represent groups of contact with the surrounding solvent, whereas those shown in white are shielded from the solvent by other parts of the protein molecule.

contact with solvent. The possibility that a few of the peptide groups or side chains are still grouped together, with little or no solvent contacts, is easily incorporated into the model, and a few of the parts of the unfolded molecule are accordingly shown in the figure in white, indicating that they are not in contact with solvent.

It will be assumed again that only a single unfolded state exists in the water–urea solvent system, and that all protein molecules tend to adopt this state at a sufficiently high urea concentration. It is considered probable that different unfolded states of the molecule will exist in other solvent systems. They may differ in the degree of flexibility and in the number and nature of the groups which are still shielded from solvent. But such alternative forms are assumed not to have any stability during the course of urea denaturation.

It will also be assumed at first that the transition between native and unfolded states is an all-or-none process, which means that no appreciable number of molecules exists (at any concentration of urea) in which the conformation is partially unfolded, *i.e.*, intermediate between the two representations shown in Fig. 1. The possible existence of intermediate states, and their effect on the calculated progress of the unfolding reaction, will be considered towards the end of the paper.

With these assumptions, the unfolding reaction at any concentration of urea, provided that it is reversible, may be represented as an equilibrium between unique native (N) and unfolded (U) molecules, $N \rightleftharpoons U$, which may be described by an equilibrium constant K and a standard free energy of unfolding (ΔF_u°), such that

$$-\Delta F_u^\circ = RT \ln K = RT \ln ([U]/[N]) \quad (1)$$

square brackets representing thermodynamic activities. At sufficiently low protein concentrations, these activities may be replaced by concentrations, and it will be assumed hereafter that they represent concentrations. Since the reaction involves no change in the number of molecules, the concentration units are immaterial.

The parameters ΔF_u° and K of eq. 1 are variables whose values depend on the urea concentration. The values which these parameters take on in the absence of urea will be designated as $\Delta F_{u,H_2O}^\circ$ and K_{H_2O}.

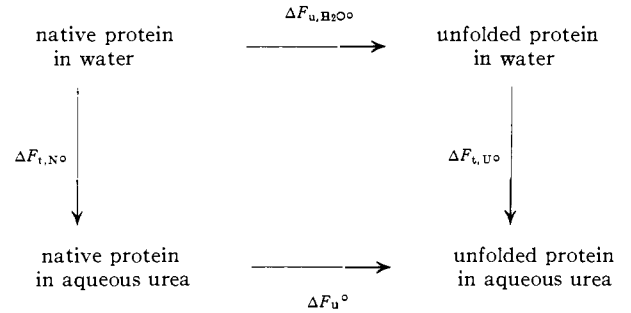

Fig. 2.—Free energy diagram demonstrating the validity of eq. 2.

A previous paper[2] has described the calculation of $\Delta F_{u,H_2O}^\circ$. It was found that estimates of some of the factors which contribute to it are at present quite uncertain. Estimates of the contribution of hydrophobic interactions to $\Delta F_{u,H_2O}^\circ$ were made in the earlier paper with some confidence, but these contributions alone are much larger than $\Delta F_{u,H_2O}^\circ$ itself. They are opposed by the contribution to $\Delta F_{u,H_2O}^\circ$ which arises from the greater flexibility of unfolded over native molecules, the magnitude of which cannot be estimated with any reliability. A moderately large contribution to $\Delta F_{u,H_2O}^\circ$ may also arise from the relative strengths of intramolecular hydrogen bonds in the native conformation, as compared to hydrogen bonds between the protein molecule and the solvent in the unfolded state. To calculate this contribution would require a knowledge of the precise hydrogen bonds within conformation N, and their free energies relative to hydrogen bonds to water. Information on these subjects is sparse or nonexistent.

The same difficulties are encountered in any attempt to calculate ΔF_u° directly in an aqueous urea solution but the problem becomes much simpler if one regards $\Delta F_{u,H_2O}^\circ$ as an unknown parameter and is content to calculate the difference between ΔF_u° and $\Delta F_{u,H_2O}^\circ$ at any concentration of urea. We shall call this quantity $\delta\Delta F_u^\circ$. As is shown by Fig. 2

$$\delta\Delta F_u^\circ = \Delta F_u^\circ - \Delta F_{u,H_2O}^\circ = \Delta F_{t,U}^\circ - \Delta F_{t,N}^\circ \quad (2)$$

where $\Delta F_{t,U}^\circ$ is the standard free energy of transfer of the protein in its unfolded state from water to a urea solution of given concentration, and $\Delta F_{t,N}^\circ$ is the same quantity for the protein in its native conformation. These free energies of transfer depend only on interactions with the solvent. This means, furthermore, that only those parts of the protein molecule which are in contact with the solvent are involved. The nature and strength of hydrogen bonds within the native structure is not involved, since these bonds remain unchanged in the transfer of conformation N from water to a urea solution. Both transfer reactions also involve no change in flexibility, so that this unknown factor also does not affect the ΔF_t° values.

It has been shown elsewhere[3] that the standard free energy of transfer of any simple substance, from water to any other solvent, can be determined from solubilities of the substance in the two solvent media. From data obtained in this way for the transfer of amino acids and related compounds from water to absolute ethanol, and from water to urea solutions, it appears that the free energy of interaction between an uncharged amino acid side chain and the solvent is, as a good approxi-

mation, independent of the kind of molecule to which the side chain is attached. It is therefore likely to be a good approximation to assume that uncharged amino acid side chains of a protein molecule, if they are in contact with the solvent, make contributions to ΔF_t° which do not depend on the over-all conformation of the protein molecule, and which, moreover, can be estimated directly from literature data on amino acids or other suitable model compounds.

That the same principle applies to charged groups of proteins is inherent in the normal analysis of hydrogen ion titration curves of proteins. Experimental titration data can be accommodated largely by a theoretical treatment which assumes that all titratable groups of a protein molecule have the same intrinsic free energy of dissociation as similar groups on model compounds, and that the only interaction between titratable groups is an electrostatic one.[4,5] If a group on a protein molecule has the same intrinsic free energy of dissociation as it would have on a model compound, then the free energies of interaction with the solvent must also be the same. Moreover, the theory which is used to calculate (usually quite successfully) the electrostatic interaction between protein charges is identical with that which is used to calculate similar interactions in amino acids and other model compounds.[6] The theory includes as a parameter an effective distance of closest approach of solvent to a charged group, which parameter would be expected to depend on the nature of the interaction between the solvent and the charged group. It is found that the parameter must generally be assigned the same value in proteins and in model compounds.[6] It should accordingly be excellent to assume that the contribution which direct solvent interactions of a particular charged group make to $\Delta F_{t,U}^\circ$ is the same as the contribution which they make to $\Delta F_{t,N}^\circ$ and that the contribution could be estimated (although this will be seen to be unnecessary) from solubility studies on appropriate model compounds. Charged groups of course make a further contribution to the free energy of transfer through long-range electrostatic interactions. A term which takes these interactions into account will be introduced below (eq. 6).

Initial solubility studies in urea solutions, on substances containing peptide groups,[3] indicate that the interaction between a peptide group and the solvent is more strongly dependent on the nature of the molecule which contains the group than is true for uncharged amino acid side chains, or for charged groups. It is likely that the average contribution of each peptide group to $\Delta F_{t,U}^\circ$ will, at a given concentration of urea, be the same for every unfolded protein, but from presently existing data one would get a different numerical value for this contribution, depending on the model compound which is chosen. It is also possible that peptide groups exposed to the solvent in the native structure (if any exist) may make a somewhat different contribution to $\Delta F_{t,N}^\circ$ than that which is applicable to an unfolded conformation. As a first approximation, a single value will be chosen for the contribution of an exposed peptide group, regardless of protein conformation, with the understanding that the choice is subject to correction as more data on model compounds are accumulated.[7]

These considerations suggest that it is possible to consider the interaction between a protein molecule and its solvent as the sum of individual contributions from the constituent groups of the protein molecule–peptide groups and the individual kinds of side chains (or parts thereof). Each kind of distinguishable group may be designated by a different subscript, i. A unique value for the contribution which such a group makes to the total free energy of transfer at any urea concentration is decided on and designated $\Delta f_{t,i}$ If now there are $n_{i,N}$ groups of type i in contact with the solvent in the native conformation, and $n_{i,U}$ groups in contact with the solvent in the unfolded state, we obtain

$$\delta\Delta F_u^\circ = \Delta F_{t,U}^\circ - \Delta F_{t,N}^\circ = \sum_i (n_{i,U} - n_{i,N})\Delta f_{t,i} \quad (3)$$

The maximum contribution which any kind of group can make to $\delta\Delta F_u^\circ$ is $n_i^\circ \Delta f_{t,i}$, where n_i° is the total number of groups of type i which the protein molecule possesses. It is convenient to introduce a parameter α_i to designate the fraction of the maximum contribution which any kind of group actually makes to the sum of eq. 3. Thus

$$\alpha_i = (n_{i,U} - n_{i,N})/n_i^\circ \quad (4)$$

and

$$\delta\Delta F_u^\circ = \sum_i \alpha_i n_i^\circ \Delta f_{t,i} \quad (5)$$

It may be noted at this point that the schematic representations of Fig. 1, in which all groups are designated as fully accessible to solvent or not at all accessible, are oversimplified. It is certainly possible that some groups are located so that they interact with solvent over part of their surface area only. Such groups would be considered here as partly accessible to solvent, and the existence of such groups might then lead to nonintegral values for $n_{i,U}$ or $n_{i,N}$. Adjustability of the parameter α_i thus allows for the existence of incomplete solvent contacts.

The theoretical treatment outlined so far does not provide for the contribution of long-range intramolecular electrostatic interactions to ΔF_u°. Such interactions can be included[2] by considering the equations given above as applying to a hypothetical discharged state of the protein molecule. The expression for ΔF_u° of the charged molecule is obtained then by adding the work of charging (W) to the free energy of each conformation in every solvent. We obtain

$$\delta\Delta F_u^\circ = \sum_i \alpha_i n_i^\circ \Delta f_{t,i} + \delta\Delta W \quad (6)$$

where

$$\delta\Delta W = (W_U - W_{U,H_2O}) - (W_N - W_{N,H_2O}) \quad (7)$$

(4) K. Linderstrøm-Lang, *Compt. rend. trav. lab. Carlsberg*, **15**, No. 7 (1924).

(5) C. Tanford, *Advan. Protein Chem.*, **17**, 69 (1962).

(6) C. Tanford, *J. Am. Chem. Soc.*, **79**, 5348 (1957).

(7) To a lesser extent, all the calculations of this paper must be considered subject to revision. Wetlaufer, *et al.*,[8] have suggested, for instance, that the contribution which a nonpolar amino acid side chain makes to ΔF_t°, when the side chain is part of an appropriate hydrocarbon molecule, may differ by about 10% from the contribution which it makes to ΔF_t° for the amino acid itself. If this difference primarily reflects a secondary effect of the charged groups of the amino acid, then the values obtained from hydrocarbon solubilities would more appropriately be equated with the contribution expected from a nonpolar amino acid side chain on a protein molecule.

(8) D. B. Wetlaufer, S. K. Malik, L. Stoller, and R. L. Coffin, *J. Am. Chem. Soc.*, **86**, 508 (1964).

the subscripts having the same meaning as before. The work of charging in the unfolded state is likely to be sufficiently small in most applications to permit neglect of the term $(W_U - W_{U,H_2O})$. In the absence of exact information on the location of charges, W_N for the native protein is best calculated for an appropriate equivalent sphere, with the net molecular charge Z spread evenly over the surface.[9] The difference between W_N and W_{N,H_2O} is then simply a function of the difference between the dielectric constants of the solvents, and we get

$$\delta\Delta W = \frac{\mathfrak{N}Z^2\epsilon^2}{2R}\left(1 - \frac{\kappa R}{1 + \kappa a}\right)\left(\frac{1}{D_{H_2O}} - \frac{1}{D}\right) \quad (8)$$

where $\mathfrak{N}$ is Avogadro's number, Z is the molecular charge in protonic units, ϵ is the protonic charge in e.s.u., R is the radius of the equivalent sphere, and a the radius to which the centers of ions are excluded, D_{H_2O} and D are the relevant dielectric constants, and κ is the usual Debye-Hückel parameter proportional to the square root of the ionic strength.

Procedure for Numerical Calculations

In making numerical calculations of $\delta\Delta F_u^\circ$ for any given protein, it is convenient to classify the component parts of the protein molecule into categories, based on their chemical properties. In doing so, we believe it realistic to separate the charged groups of acidic and basic amino acids from the hydrocarbon chains which join them to the polypeptide backbone. The reason for this was explained in our earlier paper.[2] It appears probable that the charged groups of native globular proteins are nearly always at the surface, in contact with solvent. The hydrocarbon chains to which the charged groups are attached may, however, often be within the globular interior, not in contact with solvent.

The various kinds of chemical groups are then classified as follows.

Charged groups include amino and carboxyl terminal groups, terminal groups of arginine, lysine, histidine, glutamic acid, and aspartic acid side chains. The number of each kind of charged group depends on pH, and the listing given here is that applicable to approximately neutral solutions.

Hydrophobic side chains include alanine, valine, leucine, isoleucine, proline, phenylalanine, tyrosine, tryptophan, methionine, cystine, cysteine, the hydrocarbon portions of arginine, lysine, aspartic, and glutamic acids. It is of course somewhat arbitrary to include side chains which contain polar groups (*e.g.*, tyrosine, cystine) in this category. The justification for it will be given below, when values of α_i are discussed.

Uncharged polar side chains include asparagine, glutamine, serine, threonine. When acidic and basic side chains are at a pH where they would be uncharged, they may be included in this category under certain circumstances, as will be discussed below.

The Peptide Groups.—The number of peptide groups is one less than the total number of residues. Glycyl residues contribute only a peptide group, and no side chain.

(9) C. Tanford, "Physical Chemistry of Macromolecules," John Wiley and Sons, Inc., New York, N. Y., 1961, Chapter 7.

In the absence of precise information on the native structure of an individual protein and on the exact extent of openness of unfolded structures, the assignment of α_i values must depend on generalizations deduced from the known conformation of myoglobin[10] and from indirect evidence obtained with other proteins. The following generalizations have been made.

(1) The unfolded molecule has 75–100% of all its component parts in contact with solvent. This generalization is based on the experience that disulfide bonds, phenolic groups, and other parts of protein molecules which frequently react anomalously in native proteins, generally lose their anomalous behavior when the protein has been unfolded in urea.

(2) Charged groups are virtually always in contact with solvent, both in the native and unfolded states of the molecule. This generalization has been known for a long time on the basis of experimental protein titration curves. It means that, for all charged groups, $n_{i,U} = n_{i,N} = n_i^\circ$, so that $\alpha_i = 0$. Thus charged groups do not contribute to the summation on the right-hand side of eq. 6, and it is not necessary to have values of $\Delta f_{t,i}$ for such groups. (Charged groups do of course contribute to $\delta\Delta F_u^\circ$ through the term $\delta\Delta W$.)

(3) Uncharged polar side chains (excluding those which are the uncharged forms of acidic or basic side chains) tend to be accessible to the solvent in the native conformation, leading to low values of α_i. We have arbitrarily taken $\alpha_i = 0.25$. The total contribution of uncharged polar side chains to $\delta\Delta F_u^\circ$ is small regardless of the choice of α_i (see Fig. 3), so that it is not worth while to allow a range of values for α_i.

The values of $\Delta f_{t,i}$ for asparagine, glutamine, and threonine side chains have been measured.[3] Since $\Delta f_{t,i}$ for alanine side chains is negligibly small, the value for serine has also been assumed negligible.

(4) Uncharged polar side chains which are the uncharged forms of acidic or basic side chains normally are treated as if they were charged side chains. If no conformational change has occurred in going from a pH where these groups are charged to a pH where they are uncharged, then the terminal polar groups must still all be in contact with solvent. If evidence for a conformational transition exists, groups of this kind which are uncharged at the point at which the transition occurs would be included with other uncharged polar side chains. In β-lactoglobulin there is thought to be a buried carboxyl group. For such a group, $\alpha_i = 1$. (We have assumed that $\Delta f_{t,i}$ for the group is intermediate between the values given earlier for glutamine and asparagine side chains.)

(5) About 75% of the larger hydrophobic parts of the molecule have been assumed inaccessible to the solvent in the native conformation. Allowing for the possibility that some hydrophobic parts may still be inaccessible to solvent after unfolding, we have made alternative calculations, based on $\alpha_i = 0.5$ and $\alpha_i = 0.75$ for these groups. Tyrosine, tryptophan, and the sulfur-containing side chains are included with the hydrophobic parts of the molecule because they have often been found inaccessible to solvent by a variety of studies.

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

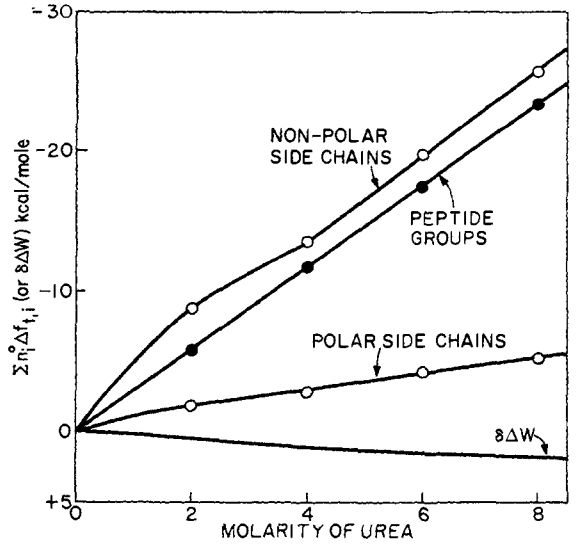

Fig. 3.—Maximum possible contribution to $\delta\Delta F_u^\circ$ for β-lactoglobulin, from peptide groups, nonpolar side chains, polar side chains, and long-range electrostatic interactions. The electrostatic free energy is calculated for pH 3, ionic strength 0.02.

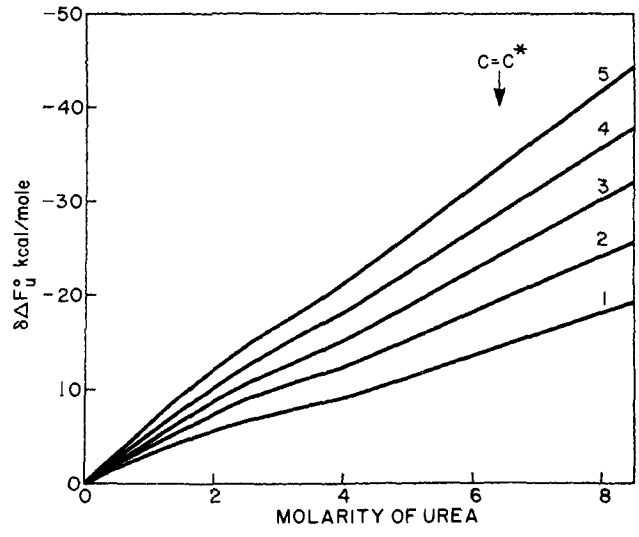

Fig. 4.—$\delta\Delta F_u^\circ$ for β-lactoglobulin, for different choices of the parameters α_i. The value of α_i for polar side chains is 0.25 for every curve. The values of α_i for nonpolar side chains and for peptide groups are shown in Table II.

$\Delta f_{t,i}$ values for about half the hydrophobic side chains have been measured.[3] For those which have not yet been determined, we have made reasonable estimates: for isoleucine, the same value as for leucine; for valine and proline, 0.67 of the value for leucine; for cysteine or half-cystine, the same value as for methionine (see Table IV, ref. 3.); for the hydrocarbon part of lysine, 0.33 of the value for leucine; for the hydrocarbon part of histidine, 0.50 of the experimental value for uncharged histidine. For the hydrocarbon parts of arginine and glutamic and aspartic acids, we have assumed $\Delta f_{t,i}$ to be negligibly small.

(6) About 75% of the peptide groups have been assumed inaccessible to solvent in the native conformation. Again, allowing for the possibility that some peptide groups may still be inaccessible after unfolding, we have made alternate calculations based on $\alpha_i = 0.5$ and $\alpha_i = 0.75$.

The value of $\Delta f_{t,i}$ which should be chosen for the exposed peptide groups of a protein molecule (chiefly those in the unfolded conformation) is uncertain. Of the alternative values given in Table VIII of ref. 3, we have chosen those based on solubilities of acetyltetraglycine ethyl ester, determined by Robinson and Jencks,[11] because the peptide groups of this molecule should closely resemble those of an unfolded protein molecule. To allow for the still existing uncertainty, we have made additional alternate calculations using $\alpha_i = 0.25$ and $\alpha_i = 1$ for peptide groups. This additional range allowed for α_i has the same effect on the calculation as allowing a range of values for $\Delta f_{t,i}$.[12]

An actual calculation for β-lactoglobulin is illustrated by Table I and Fig. 3 and 4. The maximum possible contribution which the groups of each category can make to $\delta\Delta F_u^\circ$ is shown in Table I. This quantity is equal to $\Sigma\, n_i^\circ \Delta f_{t,i}$, the sum extending over all groups in a given category, and only the amino acid composition of the protein is needed to make the calculation. Charged groups are not included because, as shown earlier, they do not contribute to $\delta\Delta F_u^\circ$. The data are plotted as a function of urea concentration in Fig. 3.

Table I and Fig. 3 also show the contributions to $\delta\Delta F_u^\circ$ made by the electrostatic term $\delta\Delta W$, as calculated by eq. 8. Because the dielectric constant of water increases when urea is added, this term is positive, stabilizing the native structure. The electrostatic term, however, is seen to be very small in comparison with the other terms in $\delta\Delta F_u^\circ$, all of which are negative.

Different possible values of α_i are now introduced. The different combinations we have used are given in Table II, and they lead to the values of $\delta\Delta F_u^\circ$ which are shown, as a function of urea concentration, in Fig. 4.

Application to Experimental Studies

If the unfolding of a globular protein by urea actually represents an all-or-none reaction between unique native and unfolded forms, then any experimental parameter which depends on conformation can be used to determine the equilibrium constant K (eq. 1) as a function of urea concentration (C). All experimental methods must lead to identical values of K. (We shall postpone for the moment the question of whether any truly all-or-none reactions of this kind occur.)

It is not possible to calculate the equilibrium constant K directly by the procedure given in this paper, because the free energy of stabilization of the native protein in water ($\Delta F_{u,H_2O}^\circ$) is unknown, and because we have been compelled to allow a range of possible values for the structural parameters α_i which reflect the extent of exposure of parts of the protein molecule to the solvent.

It is possible, however, to proceed in the reverse direction and to use the experimental values of K to determine what values of $\Delta F_{u,H_2O}^\circ$ and of α_i are compatible with the experimental data. In doing so, we

(11) D. R. Robinson and W. P. Jencks, *J. Biol. Chem.*, **238**, PC 1558 (1963).

(12) It should be noted that $\Delta f_{t,i}$ values for peptide groups, taken from ref. 3, represent free energies of transfer for a glycyl residue, which has one hydrogen atom more than a peptide group. However, the $\Delta f_{t,i}$ values of side chains in ref. 3 represent the effect of substituting the side chain for a hydrogen atom. When the $\Delta f_{t,i}$ values for peptide groups and side chains are added, therefore, we get correctly the free energy of transfer of an entire residue.

TABLE I

CALCULATION OF $\delta\Delta F_u^\circ$ FOR β-LACTOGLOBULIN[a,b]

Type of group	Number per molecule (n_i°)[c]	$\Delta f_{t,i}$ cal./mole, 4 M urea	$\Delta f_{t,i}$ cal./mole, 8 M urea	$n_i^\circ \Delta f_{t,i}$, 4 M urea	$n_i^\circ \Delta f_{t,i}$, 8 M urea
Ala side chain	14	+15	+10	+210	+140
Val side chain	10	−105	−200	−1050	−2000
Leu side chain	22	−155	−295	−3410	−6490
Ileu side chain	10	−155	−295	−1550	−2950
Pro side chain	8	−105	−200	−840	−1600
Phe side chain	4	−330	−600	−1320	−2400
Tyr side chain	4	−395	−735	−1580	−2940
Trp side chain	2	−505	−920	−1010	−1840
Met side chain	4	−225	−415	−900	−1660
Half-cys side chain	5	−225	−415	−1125	−2075
Lys hydrocarbon part	15	−50	−100	−750	−1500
His hydrocarbon part	2	−80	−130	−160	−260
Total hydrophobic groups				−13,480	−25,580
Thr side chain	8	−60	−115	−480	−920
Asn side chain	5	−225	−430	−1120	−2150
Gln side chain	8	−130	−230	−1040	−1840
Buried COOH group	1	−180	−330	−180	−330
Total polar groups				−2820	−5240
Peptide groups	161	−72	−145	−11,590	−23,340
Electrostatic free energy contribution ($\delta\Delta W$)				+1150	+1845

[a] At pH 3, ionic strength 0.02, 25°, at two urea concentrations. [b] To obtain a value for $\delta\Delta F_u^\circ$ the data of this table must be combined with suitable values of the parameters α_i (see eq. 6 and Fig 4). Calculations were made for 2 *M* and 6 *M* urea as well as for 4 *M* and 8 *M*, and Fig. 3 is based on all four calculations. [c] From amino acid analyses of W. G. Gordon, J. J. Basch, and E. B. Kalan, *J. Biol. Chem.*, **236**, 2908 (1961); K. A. Piez, E. W. Davie, J. E. Folk, and J. A. Gladner, *ibid.*, **236**, 2912 (1961).

TABLE II

CALCULATED VALUES OF $\Delta F_{u,H_2O}^\circ$ AND OF ν FOR β-LACTOGLOBULIN AT 25° [a]

Curve no. (Fig. 4)	Value of α_i: Nonpolar groups	Value of α_i: Peptide groups		$\Delta F_{u,H_2O}^\circ$, kcal./mole	ν[b]
1	0.5	0.25		14.5	26
2	0.5	0.5	[c]	19.5	34
	0.75	0.25			
3	0.5	0.75	[c]	24.1	42
	0.75	0.5			
4	0.5	1.00	[c]	28.8	50
	0.75	0.75			
5	0.75	1.00		33.7	56

[a] Results obtained by applying eq. 10 and 11 to the data of Fig. 4. The experimental value for the midpoint of the unfolding reaction is $C^* = 6.4$ *M*. [b] The values of ν are calculated for an all-or-none unfolding process. [c] Where two sets of α_i are bracketed, the $\delta\Delta F_u^\circ$ values differ by only a few per cent. Both the curves of Fig. 4 and the data tabulated here represent arithmetic means.

must recognize that experimental values of K can be obtained with good accuracy only when K is in the range of about 0.2 to 5.0. We shall therefore confine ourselves to two experimental parameters which can be determined readily within the narrow range of urea concentration where K lies within these limits. These parameters are: (1) the concentration C^* of urea at which K is exactly equal to unity, and (2) the concentration dependence of K at this point

$$\nu = (\partial \ln K/\partial \ln C)_{C=C^*} \quad (9)$$

For example, for β-lactoglobulin at pH 3, ionic strength 0.02, and 25° (assuming the data to represent an all-or-none reaction), we have found $C^* = 6.4$ *M* and $\nu = 22$.[13]

Corresponding to the concentration C^* there is a value of $\delta\Delta F_u^\circ$ for each choice of the α_i parameters, which can be read off a plot such as is shown for β-lactoglobulin in Fig. 4. Since, at $C = C^*$, $K = 1$, we have $\Delta F_u^\circ = 0$, and from eq. 2

$$(\delta\Delta F_u^\circ)_{C=C^*} = -\Delta F_{u,H_2O}^\circ \quad (10)$$

Thus the calculation leads, for each choice of α_i, to a unique value for $\Delta F_{u,H_2O}^\circ$.

Equally important is the fact that knowledge of C^* alone also leads, for each choice of α_i, to a unique value for the parameter ν, which can be compared to the experimental value for this parameter. Since $\Delta F_{u,H_2O}^\circ$ is independent of C, eq. 1, 2, and 9 show that

$$RT\nu = -[\partial(\delta\Delta F_u^\circ)/\partial \ln C]_{C=C^*} = -C^*[\partial(\delta\Delta F_u^\circ)/\partial C]_{C=C^*} \quad (11)$$

The derivative on the right-hand side of eq. 11 can be evaluated directly from the slope of plots such as are shown in Fig. 4.

All values of $\Delta F_{u,H_2O}^\circ$ and of ν obtained in this way for β-lactoglobulin are tabulated in Table II, and it is seen at once that only one choice of α_i gives a value of ν which lies close to the experimental value. The value of $\Delta F_{u,H_2O}^\circ$ which corresponds to it is 14.5 kcal./mole.

The Existence of Intermediate States[14]

It has so far been assumed that the unfolding reaction involves an all-or-none transition between unique native and unfolded conformations. It is possible, however, that intermediate, partially unfolded states may exist. A detailed general treatment of this topic will be presented at a later date. To avoid the lengthy statistical analysis required for a general treatment, we shall consider here just one simple possibility: that the protein molecule consists of r regions which can unfold independently. We shall assume that the change in experimental parameters which accompanies unfolding of each region is the same, so that the observed equilibrium constant (which we shall designate as K') will reflect the ratio of unfolded to folded regions rather than whole molecules. We shall also assume that the free energy of unfolding of each region ($\Delta F_u^{\circ\prime}$) is the same, *i.e.*, equal to $1/r$ of the free energy of unfolding of the whole molecule. We then have

$$-RT \ln K' = \Delta F_u^{\circ\prime} = \Delta F_u^\circ/r = (\Delta F_{u,H_2O}^\circ + \delta\Delta F_u^\circ)/r \quad (12)$$

Equation 12 shows that intermediate states of this kind have no effect on the relation between the free energy of unfolding and the concentration C^* at which the observed equilibrium constant is unity. Setting $K' = 1$ again leads to eq. 10, so that the calculated

(13) C. Tanford and P. K. De, *J. Biol. Chem.*, **236**, 1711 (1961).

(14) It is important to distinguish between transient intermediate states, the existence of which is deduced from kinetic studies and stable intermediate states. We are concerned here only with the latter kind. In order to affect the thermodynamics of unfolding, an intermediate must be present at some concentration of the unfolding agent in sufficient concentration to make a substantial contribution to the observable parameters (such as optical rotation),

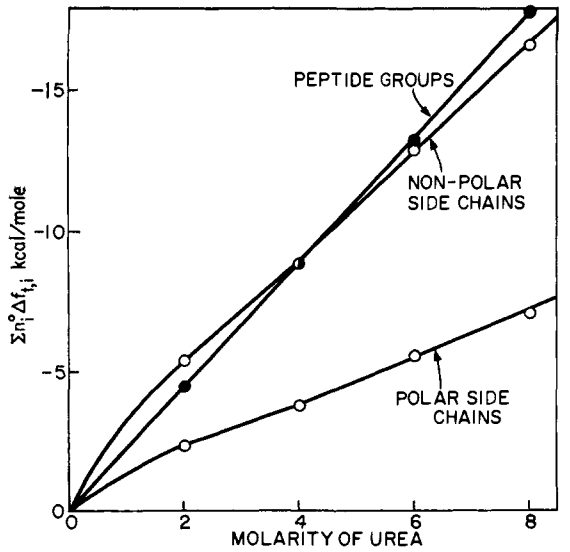

Fig. 5.—Maximum possible contribution to $\delta\Delta F_u^\circ$ for ribonuclease, from peptide groups, nonpolar side chains, and polar side chains. The electrostatic free energy is assumed to make a negligible contribution.

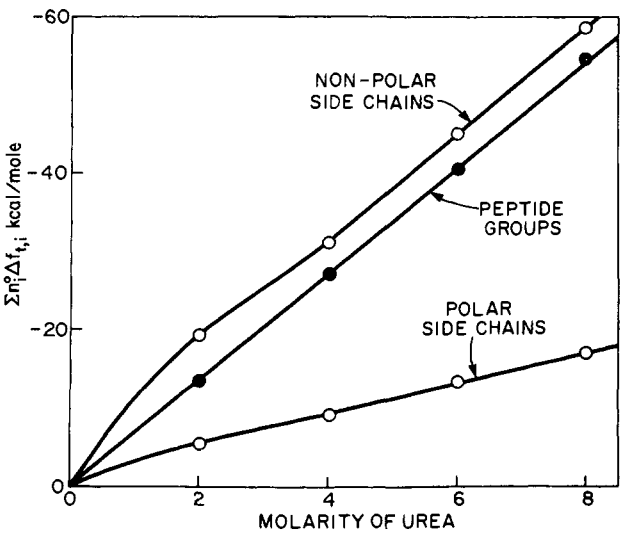

Fig. 6.—Maximum possible contribution to $\delta\Delta F_u^\circ$ for pepsinogen, from peptide groups, nonpolar side chains, and polar side chains. The electrostatic free energy is assumed to make a negligible contribution.

value of $\delta\Delta F_u^\circ$, at $C = C^*$, again gives the value of $\Delta F_{u,H_2O}^\circ$ for the whole molecule.

The existence of intermediates, however, does affect the concentration dependence of the equilibrium constant, defined as in eq. 9. Thus where

$$\nu' = (\partial \ln K'/\partial \ln C)_{C=C^*} \qquad (13)$$

eq. 12 and 11 lead to

$$RT\nu' = -(1/r)[\partial(\delta\Delta F_u^\circ)/\partial \ln C]_{C=C^*} = RT\nu/r \qquad (14)$$

Equation 14 shows that the predicted value of ν', for any given choice of α_i (and for the corresponding unique value of $\Delta F_{u,H_2O}^\circ$), can fall far below the value of ν calculated on the basis of an all-or-none transition.

A similar calculation, based on the same premises, with the one change that the r parts of the molecule must unfold in an obligatory order, is more complicated, but leads to a similar result: we still have $\delta\Delta F_{u,C^*}^\circ = -\Delta F_{u,H_2O}^\circ$ and $\nu' < \nu$.

In principle, it is possible to determine unequivocally whether intermediate states of this kind occur. One way would be to determine by calorimetry a value of ΔH_u° for the reaction leading to complete unfolding. If the reaction is all-or-none, this ΔH_u° will be identical with an observed ΔH based on the temperature dependence of the equilibrium constant. If intermediate states exist, the latter ΔH will be smaller. The pertinent data, however, generally are not available. In the absence of such data, it must be concluded that experimental values of ν can generally be smaller than those calculated on the basis of an all-or-none reaction. With reference to the calculations for β-lactoglobulin given in Table II, we must conclude that all the choices of α_i which are listed and all the corresponding $\Delta F_{u,H_2O}^\circ$ values are compatible with experiment if the unfolding reaction is not necessarily an all-or-none reaction.

Microheterogeneity

Experimental values of ν may be smaller than values calculated on the basis of an all-or-none reaction for another reason. If the native protein is heterogeneous, possessing several distinct species each with somewhat different values of $\Delta F_{u,H_2O}^\circ$ and $\delta\Delta F_u^\circ$, then each species may be half unfolded at a somewhat different value of C. The over-all unfolding curve then will be flatter than the unfolding curve for any individual species, and spuriously low values of ν will be obtained.

Such microheterogeneity could not have affected the experimental data for β-lactoglobulin discussed above. This protein, when taken from pooled milk, is known to be a mixture of two genetically different forms, β-lactoglobulins A and B. We have found, however, that the unfolding curve for pure form A by urea is virtually superimposable on the curve obtained previously for the protein derived from pooled milk.[15]

Results for Individual Proteins

The comparison of experimental data for β-lactoglobulin with calculations made according to the procedure outlined above has already been presented, as an illustration of the procedure. Similar calculations have been made for four other proteins: ribonuclease, pepsinogen, serum albumin, and one of the fragments (fragment I) obtained from papain cleavage of rabbit γ-globulin. Figures 5–8 show the maximum possible contributions ($\Sigma n_i^\circ \Delta f_{t,i}$) which the various kinds of groups could make to $\delta\Delta F_u^\circ$ for each protein. The calculations are based on amino acid compositions given in the literature.[16–21] The number of asparagine and glutamine residues of pepsinogen and serum albumin were obtained by assigning the reported values for ammonia to aspartic and glutamic acids in the same proportion as the total content of these acids. In the case of γ-globulin fragment I, the

(15) N. Pace, unpublished data.
(16) C. H. W. Hirs, W. H. Stein, and S. Moore, *J. Biol. Chem.*, **235**, 633 (1960).
(17) R. Arnon and G. E. Perlmann, *ibid.*, **238**, 653 (1963).
(18) H. van Vunakis and R. M. Herriott, *Biochim. Biophys. Acta*, **33**, 600 (1957).
(19) G. R. Tristram, "The Proteins," Vol. 1A, H. Neurath and K. Bailey, Ed., Academic Press, New York, N. Y., 1953, Chapter 3.
(20) J. B. Fleischman, R. R. Porter, and E. M. Press, *Biochem. J.*, **88**, 220 (1963).
(21) W. J. Mandy, M. K. Stambaugh, and A. Nisonoff, *Science*, **140**, 901 (1963).

TABLE III

CALCULATED VALUES FOR $\Delta F_{u,H_2O}°$ AND FOR ν[a]

Value of α_i		Ribonuclease		Pepsinogen		Serum albumin		γ-Globulin fragment I	
Nonpolar groups	Peptide groups	$\Delta F_{u,H_2O}°$, kcal./mole	ν	$\Delta F_{u,H_2O}°$, kcal./mole	ν	$\Delta F_{u,H_2O}°$, kcal./mole	ν	$\Delta F_{u,H_2O}°$, kcal./mole	ν
0.5	0.25	11.2	18	22	32	52	83	56	91
0.5	0.5	14.4	23	29	42	68	109	73	119
0.75	0.25								
0.5	0.75	17.7	29	35	53	84	135	90	149
0.75	0.5								
0.5	1.00	21.1	34	41	63	99	161	108	179
0.75	0.75								
0.75	1.00	24.4	40	47	73	114	188	126	209

[a] Based on the data of Fig. 5–8 and on the experimental urea concentration (C^*) at the midpoint of the unfolding reaction. See eq. 10 and 11. The calculated values of ν are those applicable to an all-or-none unfolding process.

experimental titration curve[22] was used to determine the number of free COOH groups, and they were assigned to aspartic and glutamic acids in the same proportion as the total content of these acids.

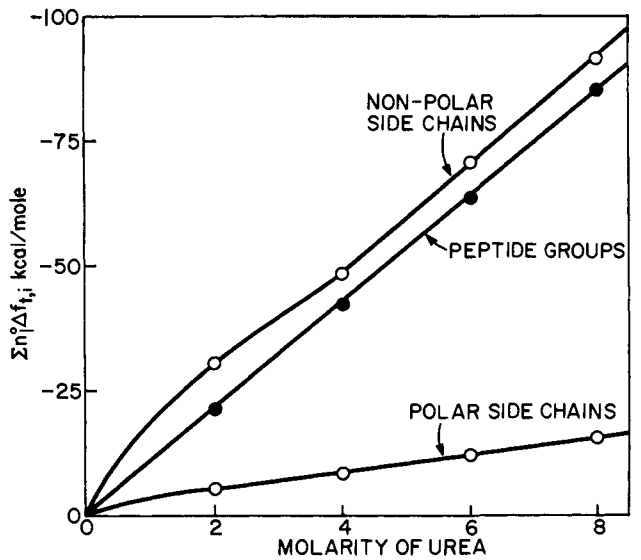

Fig. 7.—Maximum possible contribution to $\delta\Delta F_u°$ for bovine serum albumin, from peptide groups, nonpolar side chains, and polar side chains. The electrostatic free energy is assumed to make a negligible contribution.

From Fig. 5–8, curves such as those of Fig. 4 were calculated for different values of α_i, as was done for β-lactoglobulin. These curves are not shown, but the $\Delta F_{u,H_2O}°$ and ν-values calculated from them on the basis of experimental values of C^* are given in Table III.

Ribonuclease.—Experimental data on the isothermal unfolding of ribonuclease[23,24] by urea are not entirely consistent. They indicate that $C^* \simeq 6$ M, and this is the figure which has been used in the calculations. Foss and Schellman[23] have data at two urea concentrations in the usable range. Assuming a linear relation between log K and log C, one obtains by eq. 9 or 12 the value $\nu_{obsd} = 4.5$.[25] An alternative value, $\nu_{obsd} = 8$, can be obtained from the concentration dependence of the forward and reverse rate constants given by Nelson and Hummel.[24]

(22) Y. Nozaki, unpublished data.

(23) J. G. Foss and J. A. Schellman, *J. Phys. Chem.*, **63**, 2007 (1959).

(24) C. A. Nelson and J. P. Hummel, *J. Biol. Chem.*, **237**, 1567 (1962).

(25) Where clarity demands it, we shall distinguish between experimental values of $\partial \ln K/\partial \ln C$ (designated ν_{obsd}) and values calculated either for an all-or-none reaction (designated ν) or for a reaction with intermediate stable conformations (designated ν').

It is evident from Table III that even the lowest values of α_i which have been used lead to a calculated value of ν much larger than the experimental value. This means either that even lower values of α_i must be applicable to ribonuclease, which is considered improbable, or that there are stable intermediate stages between the native and completely unfolded conformations. (The possibility of conformational heterogeneity would appear to be remote in the case of ribonuclease.)

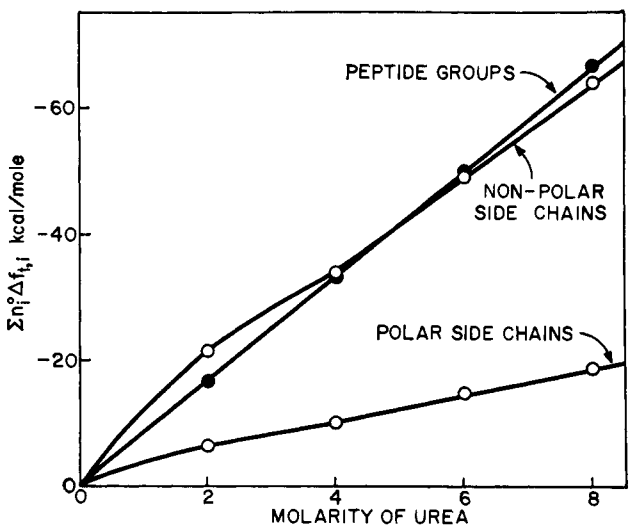

Fig. 8.—Maximum possible contribution to $\delta\Delta F_u°$ for γ-globulin fragment I, from peptide groups, nonpolar side chains, and polar side chains. The electrostatic free energy is assumed to make a negligible contribution.

Pepsinogen.—The data of Perlmann[26] lead to $C^* = 3.6$ M, which is a considerably lower concentration of urea than is required for unfolding of the other proteins which are discussed in this paper. This leads to lower calculated values of $\delta\Delta F_u°$ and of ν than would be obtained if C^* were higher. In spite of this, the ν-values listed in Table III are still much higher than the experimental value, which is $\nu_{obsd} = 5$. The existence of stable, partially unfolded, intermediate states is again indicated.

Serum Albumin.—Data for the unfolding of bovine serum albumin by urea have been taken from the work of Kauzmann and Simpson.[27] At 25° we have estimated $C^* = 5.8$ M and $\nu_{obsd} = 5$. Because the molecular weight of serum albumin is so large, the

(26) G. E. Perlmann, *J. Mol. Biol.*, **6**, 452 (1963).

(27) W. Kauzmann and R. B. Simpson, *J. Am. Chem. Soc.*, **75**, 5154 (1953).

calculated values of ν for an all-or-none transition, as shown in Table III, become correspondingly large. Foster[28] has suggested on the basis of other data that serum albumin might consist of four similar folded regions, which can unfold independently. As indicated by eq. 13, this would reduce the calculated values of ν by a factor of four. Even this, however, is not enough to bring ν down to the range of the observed value. More recent data from Foster's laboratory[29] indicate that serum albumin is a microheterogeneous protein, and that could account for the additional discrepancy between the observed value of ν and that calculated for an all-or-none reaction in Table III.

Fragment I of γ-Globulin.—From the data of Buckley, *et al.*,[30] assuming that their final rotation values represent an equilibrium curve, we have estimated $C^* = 8.4\ M$ and $\nu_{obsd} = 10$. Since γ-globulin is inherently a heterogeneous protein, and since this heterogeneity should reside in the part of the molecule which is contained in fragment I,[31] the calculated values of ν should be much higher than the observed value, and Table III shows that this is indeed so. It is not possible to determine whether the existence of stable intermediate states is an additional factor involved in the relatively low experimental values of ν.

Discussion

The calculations of this paper are limited chiefly by our inability to assign definite values to the parameters α_i. The final results of the calculations, as given in Tables II and III, have therefore been presented perforce as alternative values, for $\Delta F_{u,H_2O}°$ and for ν, for different choices of α_i.

The values of $\Delta F_{u,H_2O}°$ listed in the tables were obtained by combining calculated values of $\delta\Delta F_u°$ with the experimental urea concentrations at the midpoints of the transitions between the native globular form of each protein and the unfolded form. These values have been shown to represent, unequivocally, the difference in free energy between the original native and the final unfolded forms of the molecule, in the original aqueous solvent, regardless of whether stable intermediate forms exist. (If the original native protein consists of a mixture of closely similar "microheterogeneous" molecules, the result is an average.) Thus, if the first and last entries of the tables represent realistic lower and upper limits to the values of the α_i parameters, then lower and upper limits of the free energies of stabilization of the native conformations can be assigned. Such values are tabulated in Table IV.

The values given in Table IV tend to be larger than similar data derived from other measurements. Harrington and Schellman,[32] for example, give a free energy of unfolding of 5.5 kcal./mole for ribonuclease in aqueous solution, on the basis of thermal unfolding. This value is 50% less than the minimum value of $\Delta F_{u,H_2O}°$ for this protein given in Table IV. In fact, this is not a discrepancy. Harrington and Schellman's data are based entirely on the apparent equilibrium constant and its temperature dependence. Their free energy of unfolding is therefore that applicable to any subunit of the whole molecule which can unfold independently, *i.e.*, it corresponds to the free energy change designated here as $\Delta F_{u,H_2O}°'$ (eq. 12 and adjacent discussion), rather than $\Delta F_{u,H_2O}°$. The difference between the two free energies of unfolding therefore provides an independent confirmation of the conclusion reached earlier on the basis of experimental values of ν, that stable intermediate states exist in the unfolding of ribonuclease.[33]

TABLE IV
FREE ENERGIES OF STABILIZATION OF NATIVE PROTEINS IN AQUEOUS SOLUTION AT 25°

	$\Delta F_{u,H_2O}°$ (cal./mole)			
	Per mole of protein		Per amino acid residue	
Compd.	Min.	Max.	Min.	Max.
Ribonuclease	11,200	24,400	90	197
β-Lactoglobulin	14,500	33,700	90	208
Pepsinogen	22,000	47,000	59	125
Serum albumin	52,000	131,000	89	222
γ-Globulin fragment I	56,000	126,000	121	272

We have pointed out above that intermediate states of two kinds may occur in the unfolding process. Several parts of the protein molecule may be able to unfold independently, so that the intermediate states represent molecules in which any one, any two, etc., of these regions are unfolded. Alternatively, there may be an obligatory order for unfolding of different regions of the molecule. Assuming that the first of these alternatives is the correct one, and that each independent region has $\Delta F_{u,H_2O}°' = \Delta F_{u,H_2O}°/r$, we can calculate r in two ways; firstly, as $r = \Delta F_{u,H_2O}°/\Delta F_{u,H_2O}°'$, and secondly, by eq. 14, as $r = \nu/\nu_{obsd} = \nu/\nu'$. The result of the calculation depends of course on the choice of the α_i. If for ribonuclease we choose $\alpha_i = 0.5$ for both nonpolar and peptide groups (second line of Table III) and use Harrington and Schellman's value[32] for $\Delta F_{u,H_2O}°'$ and Nelson and Hummel's value[34] for ν_{obsd}, we get $r = 2.6$ by the first method and $r = 2.9$ by the second. These values are in excellent agreement with the value $r = 3$, estimated from enthalpy data by Harrington and Schellman.[33] With other choices of the α_i, the estimated r values would range from $r = 2$ (using the first line of Table III) to $r = 5$ (using the last line of Table III).[34]

The comparison made in this paper between the steepness of experimental unfolding curves, as represented by the parameter $\nu_{obsd} = \partial \ln K/\partial \ln C$, with calculated values for the same parameter, indicates that the

(28) J. F. Foster, "The Plasma Proteins," F. W. Putnam, Ed., Academic Press, New York, N. Y., 1960, Chapter 6.

(29) M. Sogami and J. F. Foster, *J. Biol. Chem.*, **238**, PC 2245 (1963).

(30) C. E. Buckley, III, P. Whitney, and C. Tanford, *Proc. Natl. Acad. Sci. U. S.*, **50**, 827 (1963).

(31) R. R. Porter, "Basic Problems in Neoplastic Disease," Columbia University Press, New York, N. Y., 1962, p. 177.

(32) W. F. Harrington and J. A. Schellman, *Compt. rend. trav. lab. Carlsberg*, **30**, 21 (1956).

(33) Harrington and Schellman[32] were aware of the fact that their results represented $\Delta F_{u,H_2O}°'$, rather than $\Delta F_{u,H_2O}°$, and pointed this out in their paper. By comparing their observed value of $\Delta H_{u,H_2O}°'$, derived from the temperature dependence of K, with an assumed value of $\Delta H_{u,H_2O}°$, they concluded that their data represented independent unfolding of about one third of the ribonuclease molecule. *I.e.*, in the terminology of this paper, they arrive at a value of 3 for the parameter r.

(34) The significance of these values for the number of intermediates (r) is of course limited by the simple assumption which has been made, that $\Delta F_u°$ for each region is the same. There is actually every reason to expect that different regions of a molecule such as ribonuclease, which can unfold independently, will have somewhat different stabilities. This would lead to additional spreading out of the transition. It would make $\nu_{obsd} < \nu'$ of eq. 14, and would thus decrease the value of r which would be estimated from the experimental data. For ribonuclease, however, $r = 2$ represents a minimum value, so that the range of possible values of r, $r = 2$ to 5, is not really affected. The likelihood of dissimilar independent regions would appear however to make lower values of r ($r = 2$ or 3) more probable than the higher values within this range.

existence of stable intermediate forms between native and unfolded protein molecules must be a general phenomenon. Of the proteins for which calculations have been made, only β-lactoglobulin could possibly be considered as unfolding by means of an all-or-none reaction in which intermediate states have only a transient existence. And even for β-lactoglobulin, this possibility exists only if the minimum α_i values are assigned.

It is worthwhile at this point to note that values of α_i smaller than those represented by the top lines of Tables II to IV are considered improbable when urea is the unfolding agent. Smaller values of the α_i parameters would require that the final state of unfolding attained in urea is one in which the major part of the molecule is in fact still tightly folded. This would mean, in a molecule of the size of serum albumin or fragment I of γ-globulin, that folded regions as large as the entire ribonuclease or β-lactoglobulin molecule could still exist. If this were so, then the optical rotatory properties (for example) of unfolded proteins would be expected to vary from one protein to another by as much as the similar properties of the smaller native proteins vary from each other. Moreover, inaccessible disulfide bonds or anomalous tyrosyl residues would be expected. In fact, urea-unfolded molecules have not been found to exhibit such variations, except in instances where the changes produced by urea clearly have not been completed at the highest experimentally attainable urea concentrations.

Small values of α_i would also result if most of the peptide groups and side chains of the native molecule, in its equilibrium conformation, were freely accessible to solvent. This possibility is considered highly improbable.

It should be noted in conclusion that there is evidence to suggest that guanidine resembles urea as an unfolding agent, except that it is effective at lower concentrations. Thus the theoretical treatment of this paper could be applied equally well to unfolding by guanidine, if the requisite solubility data were available. It is possible, however, that other frequently used "denaturing agents" such as acids, bases, ethanol, etc., may have a significantly different mode of action. It is quite likely that the final product obtained by reaction with these reagents is an incompletely unfolded molecule, with specific local intramolecular interactions maintained. If a reagent is to behave similarly to urea, it must solubilize both hydrophobic and peptide groups, and must also maintain a high dielectric constant so as to avoid intramolecular ion-pair formation.

[CONTRIBUTION FROM THE NATIONAL INSTITUTE OF ARTHRITIS AND METABOLIC DISEASES, NATIONAL INSTITUTES OF HEALTH, PUBLIC HEALTH SERVICE, U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE, BETHESDA, MARYLAND]

Association Constants of Methylmercuric and Mercuric Ions with Nucleosides

BY RICHARD B. SIMPSON

RECEIVED AUGUST 8, 1963

Spectra of methylmercuric hydroxide or mercuric hydroxide with nucleosides at various pH values were used to determine the sites of mercury binding and the association constants. In addition to the expected binding to the nitrogen sites having titratable hydrogens, there is binding to the primary amine nitrogens, as deduced independently by Eichhorn and Clark. Guanosine and inorganic mercury evidently polymerize slowly. The affinity of calf thymus DNA for inorganic mercury is at least an order of magnitude greater than that of the nucleosides.

In 1952 Katz[1] added mercuric chloride to a nucleic acid solution and discovered a remarkable reversible decrease in the viscosity, as well as changes in other properties, indicating some kind of configurational rearrangement. The molecular structural changes are still a subject of investigation and conjecture. At Katz's suggestion, we decided to determine the sites of binding and association constants of methylmercuric and mercuric ions with the nucleosides, since this information might prove useful in explaining the nucleic acid reaction. A study had already been made by Davidson and co-workers[2,3] at an acid pH. The present investigation covers the whole pH range.

Experimental

A 15% solution of methylmercuric hydroxide was a gift from the Morton Chemical Co. of Woodstock, Illinois. 7-Methylguanosine was a product of Cyclo Chemical Co. Spectra were measured on a Cary recording spectrophotometer with 1 mm., 1 cm., and 5 cm. cells (the shortest cells for the highest methylmercuric or mercuric concentrations and *vice versa*). Spectra of each nucleoside were recorded at several values of pH with various concentrations of methylmercuric hydroxide or mercuric nitrate. Usually, only perchloric acid or sodium hydroxide was added, but sometimes low concentrations of such weakly complexing buffers as fluoride, acetate, phosphate, or borate were used.

Although methylmercuric or mercuric compounds were added to the "blank" cells or, in some difference spectra, to one of a pair of extra cells in the light paths, the absorbance of concentrated solutions of mercury or methylmercury is so high at short wave lengths that this region of the nucleoside difference spectra is less reliable for calculations. The molar absorption coefficient (ϵ_M) of CH_3HgOH is approximately 1.8 at 250 mμ and 0.35 at 260 mμ, and ϵ_M of CH_3HgClO_4 is about twice as large as that of CH_3HgOH.

Results

Sites of Binding.—The recorded spectra, in general, represent absorption by molecules with partial mercuration at one or more sites. After some trial and error, it was found that on the basis of a hypothetical but chemically reasonable set of sites the spectra could be analyzed into component difference spectra, each of which corresponded to mercuration at one site only.

Some of the difference spectra may be identified with nitrogen sites of known pK by means of the pH dependence. The effect of pH on the mercuration at a particular site, *i.e.*, on the height of the difference spec-

(1) S. Katz, *J. Am. Chem. Soc.*, **74**, 2238 (1952).

(2) T. Yamane and N. Davidson, *ibid.*, **83**, 2599 (1961).

(3) R. Ferreira, E. Ben-Zvi, T. Yamane, J. Vasilevskis, and N. Davidson, Advances in the Chemistry of the Coordination Compounds," Macmillan Co., New York, N. Y., 1961, pp. 457–462.