ELSEVIER

Biophysical Chemistry 96 (2002) 91–101

Biophysical Chemistry

www.elsevier.com/locate/bpc

Fifty years of solvent denaturation☆

John A. Schellman*

Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

Received 10 June 2001; received in revised form 7 December 2001; accepted 7 December 2001

Abstract

The author has worked in the area of solvent denaturation and stabilization, off and on, for approximately 50 years. This paper is a personal view of the progress which has been made since 1950. The topic is limited to the development of thermodynamic molecular models for the interpretation of the unfolding and stabilization of protein structures. The story starts with the work in Kauzmann's laboratory shortly after World War II and proceeds through models for multisite binding, the linear denaturation curve, the special considerations required for understanding weak solvent exchange and a new model for the balance of solvent contact interactions and excluded volume.

Keywords: Binding; Denaturation; Excluded volume; Osmolyte; Solvent interaction; Stabilization

1. Background 1931–1946

I first encountered the phenomenon of 'chemical denaturation' in Walter Kauzmann's laboratory in 1948. My own thesis topic was on a different subject but it was clear that the major thrust of the laboratory was the study of denaturation. As is well known, these quantitative studies formed a cornerstone for all work that followed and led to the establishment of denaturation as the unfolding of structured protein chains to a random polymer state [1,2].

There was, of course, a considerable amount of previous work. A thoroughly worn copy of the excellent review of Neurath et al. [3] was evidence for its role as the prime source of information for earlier studies. Papers by Hopkins [4,49] Wu [5] and Eyring [6] provided a starting point for the evolution of ideas on the nature of the denaturation process.

Most of Kauzmann's work made use of urea or guanidinium chloride as denaturing agents. The fact that urea caused proteins to become insoluble (after rapid dilution) goes back to Spiro [7]. A number of compounds structurally related to urea and guanidine have been identified, but these are generally less effective [3]. It is the quantitative nature of the interaction of these reagents with

☆ This paper is dedicated to Rainer Jaenicke who has contributed so much to the field of protein science, not only by his important research contributions but also as a central harmonizing and organizing spirit for workers in the field.

*Tel.: +1 503 346 6093; fax: + 1 541 346 5891.

E-mail address: john@molbio.uoregon.edu (J.A. Schellman).

0301-4622/02/$ - see front matter
PII: S0301-4622(02)00009-1

proteins that is the topic of this short history. Much has been learned but we are still far from quantitative predictions.

2. Early observations 1946–1952

One of the intriguing features of the studies in the Kauzmann laboratory was the abrupt nature of the denaturation curve (kinetic or equilibrium) as a function of denaturant concentration. A protein could be 5% denatured at 4 M urea and 95% denatured at 6 M urea. In equilibrium experiments, if K_{unf} is defined as the ratio of unfolded to folded protein molecules, then plots of ln K_{unf} vs. the log of the denaturant concentration, [D], gave slopes of order 15 at low temperature, the number decreasing as temperature is increased.

$$\ln K_{unf} = \ln K_o + \bar{n}\ln[D] \quad (1)$$

Here, K_{o} is the equilibrium constant in 1 M denaturant. For kinetic experiments it was found that plots of ln K vs. ln [D] gave slopes of similar magnitude where K is the first order rate constant. For either type of event it was deduced that the unfolding of proteins involved a complex of the protein with many urea molecules. $\bar{n}$ was interpreted as an average number of denaturant molecules bound to the exposed sites of the transition state or to the denatured protein [1].

This put into quantitative form ideas presented earlier by Hopkins [4] who had found that the rate of urea denaturation decreases with temperature. He proposed that denaturation occurs via a protein–urea complex and that this complex becomes less stable as temperature increases. Simpson and Kauzmann actually found a minimum in rate and stability at approximately 25 °C. This could be fit into Hopkins' hypothesis by assuming that an intrinsic increase of unfolding with temperature competes with a diminishing stability of the urea complex. Below 25 °C the urea effect dominates, above 25 °C, the thermal effect. This very interesting observation would be explained today in terms of the temperature dependence of the hydrophobic interaction.

3. Independent binding site theory 1953–1955

About 1950 Kauzmann suggested to the author that the binding of urea to unfolded proteins was probably related to the interaction of urea with the peptide group which has a similar constitution. He proposed that a measure of the strength of the urea–peptide interaction might be found by evaluating the urea–urea interaction in aqueous solution. This could be done by looking at the non-ideality of urea solutions for which activities and enthalpies of dilution were available. At the time, I was writing a thesis far removed from this topic and nothing was done.

The publication of Pauling and Corey's H-bonded helical structures in 1951 aroused everyone's interest in H-bonds and their role in protein structure. My own work at the Carlsberg laboratory attempted to simplify the protein stability problem by studying helix stability in water. Amazingly (at the time) none of the natural peptides that were studied formed helices: A and B chains of insulin, oxidized ribonuclease with disrupted SS bridges, protamines, etc. All gave evidence of being random polypeptides [8,9]. It was time to perform the analysis suggested by Kauzmann.

Using thermal data of Gucker and Pickard [10] and activity data of Scatchard and coworkers [11], the result for the bimolecular interaction was $\Delta H = -2.1$ kcal with an association constant of $K = 0.041$. From this data the energy of the NH–O hydrogen bond was estimated to be approximately 1 kcal. The result was unexpected. It provided a rationale for the marginal lack of stability of H-bonded structures in aqueous solution. Pauling had estimated the enthalpy of formation of a peptide hydrogen bond in water as -4 to -5 kcal on the basis of diketopiperazine solubility and this value, which would indicate a high stability for peptide structures, was universally accepted at the time. The energy and stability of the peptide bond in water has been a point of controversy ever since. For many years it was popular to ascribe zero stability to the interaction but this has changed in recent years. It is worthwhile to keep in mind that regardless of the H-bond problem, urea has a virial coefficient in aqueous solution that is associated

with a significant negative enthalpy. There is an interaction.

The association constant was unexpectedly weak. More will be said about this in Section 7.

We return now to solvent denaturation theory. Using multiple binding theory and assuming independent sites it is found that the free energy of interaction of binding to n sites with an association constant K' is given by $-nRT\ln(1+K'[D])$ where $[D]$ is the molarity of the denaturant. The total free energy of unfolding is then given as [12]

$$\Delta\bar{G}_{unf}=\Delta\bar{G}^{\circ}_{unf}-nRT\ln(1+K'[D]) \quad \text{or}$$
$$\ln K_{unf}=\ln K^{\circ}_{unf}+n\ln(1+K'[D]) \qquad (2)$$

where $\Delta\bar{G}^{\circ}_{unf}$ is the free energy of unfolding in the absence of denaturant. This formula was developed shortly after the deduction of the structures of the α-helix, the two β-sheets and the DNA double helix. At the time the hydrogen bond was considered to be the major stabilizing force in the structure of macromolecules Consequently n was interpreted as Δn, the number of hydrogen bond sites exposed by the breaking of peptide hydrogen bonds as a result of unfolding. Note that Δn refers to the total number of binding sites, which are exposed, not the number of bound molecules of D. The connection with Eq. (1), however, is direct via the well-known formula

$$n\frac{\sigma\ln(1+K'[D])}{\sigma\ln[D]}=\bar{n}$$

where $\bar{n}$ is the mean number of molecules of D bound.

The same formula [Eq. (2)] was obtained by Aune and Tanford a number of years later [13]. At this point the stability of macromolecular structures was considered to arise exclusively from hydrophobic interactions so that n was defined as the number of binding sites with the emphasis on hydrophobic groups. Hydrophobic interactions are not stoichiometric like hydrogen bonds. This problem was circumvented by Tanford's group who determined free energies of transfer of protein groups from water to denaturing solvent systems and applied the result to the transfer of groups from the interior of proteins to the denaturing solvents in the unfolding reaction [14].

In recent years investigators have found it necessary to develop a more balanced point of view in which both polar and non-polar interactions are important parts of protein stability.

4. A general formula 1975

Later the binding model of denaturation was put into a completely general form [15]. It can be shown that the excess macromolecular free energy which is produced by interacting with other solutes is given by $-RT\ln\sum$ where $\sum$ is the binding polynomial discussed extensively by Wyman. The relation is quite general and is applicable to several different ligands interacting with the protein with no restrictions about independent binding. For example for two ligands A and B it is written as:

$$\sum=1+K_{10}[\mathrm{A}]+K_{01}[\mathrm{B}]+K_{20}+[\mathrm{A}]^2+K_{11}[\mathrm{A}][\mathrm{B}]+\ldots$$

where the K's are the standard 'macroscopic' equilibrium constants and the subscripts indicate, respectively, the number of A and B molecules bound. The K's contain statistical factors and may contain interaction parameters (linkages) amongst the sites.

If there is only one ligand and the n sites are independent, the binding polynomial factors to $(1+K'[D])^n$ and we regain Eq. (2). For a change from state f to state u of a protein

$$\Delta G=\Delta G^{\circ}-RT\ln\left(\sum\nolimits_u/\sum\nolimits_f\right) \quad \text{or}$$
$$\ln K=\ln K^{\circ}+\ln\left(\sum\nolimits_u/\sum\nolimits_f\right) \qquad (3)$$

For denaturation u is the unfolded state and f is the folded state. If we assume independent sites, then the binding polynomial for site j is given by $(1+K'_jC)$ and the binding polynomial for the totality of sites by $\prod_j(1+K'_jC)$. Sites, which are common to the folded and unfolded states, cancel out, leaving the binding polynomial for the sites, which are exposed, on unfolding. We then obtain, for solvent denaturation,

$$\Delta G_{unf}=\Delta G^{\circ}_{unf}-RT\sum\ln(1+K'_jC) \quad \text{or}$$
$$\ln K_{unf}=\ln K^{\circ}_{unf}+\sum\ln(1+K'_jC) \qquad (3a)$$

The sum is over the exposed sites. Assuming an average binding constant, this formula reverts to Eq. (2). The reason for the prime on K' will be evident later on in this article.

As long as one is working with simple independent site models, nothing has been added by the more general formulation. But the phenomenon has been put into a rigorous thermodynamic framework and may be used for more complicated models, for example, the effect of the binding of substrate or other molecules on unfolding. It is also applicable to other types of binding problems. If f is oxy- and u is deoxy-hemoglobin, then the binding polynomials can contain concentrations of O_2, CO_2, H_3O^+ and diphospho-glycerol phosphate for a complete description of the phenomena including the Bohr effect.

5. Linearity! 1974

In 1974, Greene and Pace published a remarkable paper in which they demonstrated that plots of the free energy of unfolding (or $\ln K$) were linear in denaturant concentration for both urea and guanidinium chloride [16]. Deviations have been found for some systems, especially for guanidinium chloride at low concentration, but most systems obey a linear relationship in the transition region within experimental error.

$$\Delta G_{unf} = \Delta G^{\circ}_{unf} - mc_3 \quad \text{or}$$
$$\ln K_{unf} = \ln K^{\circ}_{unf} + (m/RT)C_3 \tag{4}$$

C_3 is the concentration of denaturant and $-m$ is the constant slope of a plot of $\Delta\bar{G}$ vs. C. The 'Pace plot' has revolutionized the interpretation of isothermal denaturation and has become a standard method of determining the free energy of stabilization of proteins at a given temperature. It also demonstrates that the abrupt onset of unfolding over a small concentration range can be associated with a simple linear dependence on the free energy. Any linear effect, for example changes in pressure, will produce similar curves. The abruptness comes from the magnitude of Δn. The co-operativity of the process lies in the two-state nature of the protein transition, not in the multiple binding event per se. Because of the cooperativity a large number of sites are abruptly made available for contact with the solvent. Estimates of the value of $K'[\mathrm{D}]$ are quite small so that the shape of the denaturation curve has nothing to do with the steep slope of the binding curve near $K'[\mathrm{D}] = 1$ (see Fig. 1). Urea and guanidine are not 'powerful chemical denaturants' as the old literature described them but produce their effects by a very large number of weak interactions.

6. Solution theory 1978 [17]

Pace's linear relationship suggests the application of multicomponent solution theory which has been elegantly reviewed and creatively elaborated by Casassa and Eisenberg [18]. Their work was based on virial solution theory in which the free energy of a solution component is given by $\bar{G} = \bar{G}^{id} + \bar{G}^{ex}$ where $\bar{G}^{id}$ is the ideal part of the free energy and $\bar{G}^{ex}$ is the excess free energy produced by solute interactions. $\bar{G}^{ex}$, which is usually identical to a free energy of transfer, is expressed as $\bar{G}^{ex} = RTb = RT\ln\gamma$ and b is a power series in molar concentrations. For linear behavior only the first term in denaturant is required, so $b = b^{\circ}_{23}C_3$. b°_{23} is the coefficient of interaction between the protein (2) and the denaturant (3). The subscripts will be dropped from here on. The concentration of protein is kept constant in protein denaturation studies so that:

$$\Delta G_{unf} = \Delta G^{\circ}_{unf} + RT\Delta b^{\circ}C_3 \quad \text{or}$$
$$\ln K = \ln K^{\circ} - \Delta b^{\circ}C \tag{5}$$

where the notation on b and C has been simplified. From Eqs. (4) and (5) we see that $m/RT = -\Delta b^{\circ}$. The advantage of Δb° is that it is a common and important component of solution theory, not just a slope; it is directly related to properties such as preferential interaction, activity coefficients, colligative properties, etc. [19]; it is a second virial coefficient and is the basis for most statistical mechanical analyses of interaction in solution.

Viewed in the light of statistical mechanics, $\Delta b^{\circ}C$ can be represented as a sum over the contact free energies of sites exposed during unfolding. For heterogeneous, independent sites we would have:

$$\ln K_{unf} = \ln K^{\circ}_{unf} - \sum \ln(1 + K'_j C) \tag{6}$$

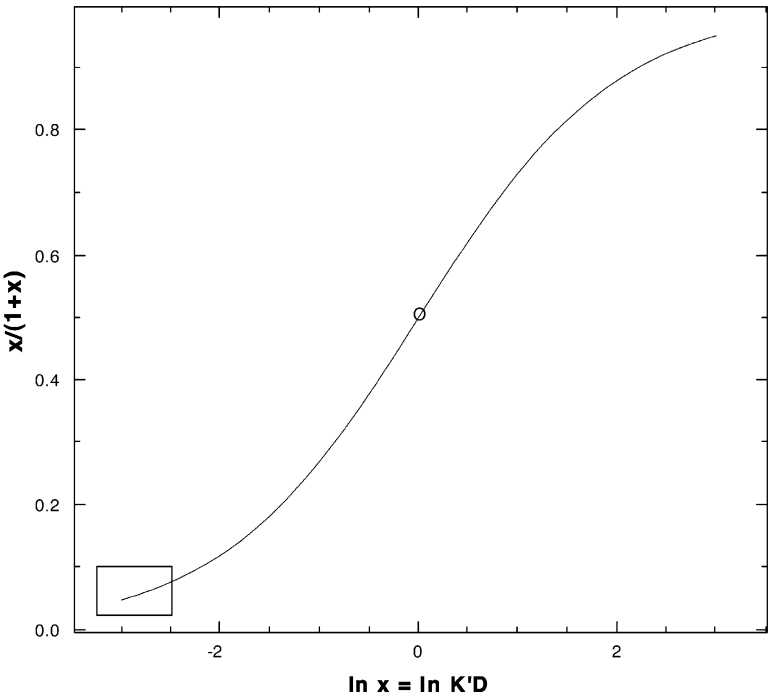

Fig. 1. The standard curve for the simple titration of a site. The small box on the lower left indicates the region that is appropriate for the binding of a molecule like urea to a protein.

Subscript j indicates a specific site on the protein. This relation can only be linear if the K's are small, so the log can be expanded.

$$\ln K_{unf} = \ln K^{\circ}_{unf} - C\sum K'_j \tag{7}$$

If the number of exposed sites, Δn, can be enumerated this can be written as:

$$\ln K_{unf} = \ln K^{\circ}_{unf} - \Delta n\langle K'\rangle C \tag{7a}$$

This is the form in which this formula is used in practical applications. $\langle K'\rangle$ is the average value over all sites.

If all the sites were identical Δn would be proportional to the area exposed to the solvent when the protein unfolds. See below for a description of this type of model. However, we know that the sites are not identical since the sites may be at anionic, cationic, polar, aliphatic, or aromatic groups of the protein. We might construe $\sum K'_j$ as being partitioned amongst various surface areas: ionic, polar, non-polar, etc. In this case there would still be proportionality with the area, but it would depend on the composition of the exposed protein surfaces. If proteins in general possessed internal surfaces of the same or similar composition, then $\langle K'\rangle$ would be the same for different proteins as would the constant of proportionality between area and Δb. This appears to be almost the case. Calculated surface areas are in general proportional to m or Δb° values determined by experiment. However, see [20].

7. K-1, binding vs. 'being there'

It is not correct to use ordinary binding theory for the interaction of denaturants and osmolytes with proteins [21]. Biochemists speak of association constants of 10^3 and 10^6 as being small. One needs a different type of discussion for K's of the order of 0.01 which is typical of the interaction of urea with a protein site. We now inquire into the meaning of such small values of K.

Imagine we have a solution of A and B molecules of the same size and that the interactions of A with A, B with B and A with B are all identical. For definiteness we assume also a mole fraction of 1/2 for A and B. Then, because of the indifference of the energy, there will be a 50% probability that a contact site on a B molecule involves an interaction with another B molecule. In a binding curve we would be at the point indicated by a circle in Fig. 1. Is this binding? It cannot be since the model we have described is the model for an ideal solution. All interactions occur by chance since all are equivalent. Binding has to be defined in terms of the excess of an interacting pair over that which occurs by mere chance. Otherwise it has no thermodynamic consequences. This intuitive conclusion is confirmed by thermodynamic analysis.

This problem does not arise for strong interactions where chance plays a very small role compared to the binding free energy, but it is very important for very weak interactions with a solvent component in high concentration. The important key is the realization that the binding of a molecule like urea to a protein site is in fact an exchange of one solvent component for another at the site. The free energy formula, Eq. (3a), is a good starting point. With two solvent components we write for the binding polynomial $({}^1K_j\chi_1+{}^3K_j\chi_3)$ rather than $(1+K'C)$. The switch from molarity to mole fraction simplifies the derivation considerably. ${}^1K$ is the binding constant for the main solvent, water, and ${}^3K$ is the binding constant for the denaturant. The contribution of a site is then $-RT\ln({}^1K_j\chi_1+{}^3K_j\chi_3)=-RT\ln K_j-RT\ln\left(\chi_1+{}^3K_j/{}^1K_j\chi_3\right)$. $-RT\ln{}^1K_j$ is the free energy of hydration of the site and is incorporated in the standard state free energy in water. ${}^3K_j/{}^1K_j\equiv K$ is the equilibrium constant for replacement of a water molecule at the site by a denaturant molecule which may be written as

$$K_j=\frac{[3]}{[1]}\frac{\chi_1}{\chi_3}$$

where [3] and [1] are the respective averages of occupation of the site by 3 and 1, respectively. Introducing K_j into the above and also putting $\chi_1=1-\chi_3$ which is true for protein concentrations small compared to that of water and denaturant, we then have:

$$\ln K_{unf}=\ln K^{\circ}_{unf}+\sum_j\ln(1+(K_j-1)\chi_3) \tag{8}$$

instead of Eq. (3a). This is the desired relation. χ_3 can be converted to C if desired, but this will be done in a more favorable context below. When K_j is much greater than 1, K_{j-1} is essentially equal to K_j and we have normal binding theory. When K_j equals 1 we have random occupation of the site and the free energy contribution goes to zero. When K_j is less than 1, we have preferential interaction with solvent water. This is easily seen by rearranging the above expression to:

$$[3]/[1]=K_j\chi_3/\chi_1 \tag{9}$$

χ_3/χ_1 is the ratio of occupations for the random case, $[3]/[1]$ is the ratio for the real, non-random case. K_j is the multiplicative factor that converts from one to the other. In the discussion of proteins in mixed solvents we will be dealing with K's slightly greater than unity. For osmolytes the K's are slightly less than unity.

We see that the constant K'_j which appeared in Eq. (2), Eq. (3a) and Eq. (6) is not a normal association constant, but has a built-in subtraction for the random replacement of the principal solvent. For other concentration units K-1 is replaced by K-c where c is a factor that depends on the particular concentration unit [22]. K-1 is very closely related to the preferential interaction coefficient for the co-solvent (denaturant or osmolyte) and has the same sign. The theory has been put on a more rigorous basis using thermodynamic activities and the Gibbs–Duhem equation [22].

8. Excluded volume

There is a further chapter in this long history, which is sufficiently new that the results have not yet been published. This section is not only a part of the long story being developed, but also a preliminary research report. Up to here the empha-

sis has been on a molecular interpretation of empirical quantities such as m values, Δb_{23}, virial coefficients, and preferential interaction that are all thermodynamically-related to one another. But there is a deficiency in the work outlined above in that it does not take into account excluded volume.

Excluded volume has played an important role in solution theory for a very long time [23]. It produces spectacular effects in solution of macromolecules [24–26] and is now being demonstrated to be of major importance in real biological systems [27,28]. On the other hand, the excluded volume effect in the interactions of small molecules with macromolecules has received relatively little attention. Physical studies of these interactions are normally expressed in terms of molalities or mole fractions, so that volume effects are not perceived directly. Presumably most workers, including myself, considered the effects to be small when expressed in concentration units that do not explicitly depend on volume. However, this is not true. Excluded volume effects can be quite important in protein solvation. This will be demonstrated here by direct calculation. The methodology will be given in a later publication. The importance of excluded volume in small molecule–large molecule interactions has recently been stressed by Wills and coworkers [29] and by Saunders et al. [30].

With the appropriate osmotic conditions the chemical potential of a protein can be written as

$$\mu_2 = \mu_2^\circ + RT\ln C_2 + RTB_2C_3 + \text{ higher terms}$$
$$\ln\gamma = B_2C_3$$

where C is the concentration of cosolvent and B_2 is the second virial coefficient. B_2 is clearly related to the empirical m and b values discussed earlier. All subscripts will be dropped hereafter since B, b, X and K all designate the interaction of protein with cosolvent, component 3. Excluded volume in liters per mole will be symbolized by X. It is well known that the second virial coefficient for a pair of molecules is given by the excluded volume, X, for a 'hard' potential, i.e. a potential which is zero outside the contact region, but jumps abruptly to a very high value at the contact distance. Thus, $B = X$ in this limit. At the opposite extreme, if two molecules interact via a weak association, we have $BC \cong -\ln(1+K'C) \cong -KC$ from the expression for weak binding. So $B = -K$ in this case. It can be shown that these effects are additive so that, in general, $B = X - K$. Note that X and K have the same units. Generalizing to the multiple binding problem, we have:

$$\Delta B = \Delta X - \sum K'_j \equiv \Delta X - n\langle K'\rangle \qquad (10)$$

The deltas arise from the fact that it is the difference in virial coefficient, excluded volume and intrinsic association which is required, and all three terms of Eq. (10) refer to the changes which occur when the protein is unfolded. The calculation of excluded volume of molecules of complex shape has always been a difficult problem. In the protein field we are very fortunate in having algorithms [31–33] and computer programs which can calculate the surface area, volumes and excluded volumes of proteins relative to small spherical molecules. For non-spherical molecules like urea and sucrose one may average over all orientations of approach and assume a sphere with the average contact distance. This solves the problem of determining X with sufficient accuracy for any protein with known conformation.

The aim of the work is to unscramble the thermodynamics in order to evaluate the real interaction terms. The steps are as follows.

1. Obtain Δb° (or m) values from the literature.
2. From Δb° calculate ΔB.
3. Determine the accessible surface area (ASA) of the native protein for the water molecule, assuming a radius of 1.5 Å. The Molecular Surface program of Connolly [34] was used for this purpose.
4. Determine the surface area of the unfolded state. This is the major limitation of the entire calculation. The surface area and excluded volume of the unfolded protein depend on the number and nature of the intramolecular contacts. Not only is this unknown but presumably changes with concentration of cosolvent following Flory's rules for good and poor solvents. Both the expansion of polypeptide chains in denaturants [35] and their contraction in stabilizing osmolytes have been observed experimentally [36]. For the ASA of the water molecule we make

use of the tables of Creamer et al. [37]. One table is for the fully unfolded peptide chain; the other is based on the average of realistic chain conformation found in unordered regions of proteins. As a best value we took the mean of these surface areas for each protein.

5. The change in surface area is determined as the difference between the surface areas of the unfolded and folded protein. A binding event is defined as the replacement of a surface water molecule by a group of the cosolvent (NH_2, CO, OH, CH_3, etc.) A water molecule is assumed to occupy 10 Å of surface area. Therefore, the area change divided by 10 Å is the number of sites n in the formulas above.
6. The excluded volume of the folded protein for various cosolvent species: urea, TMAO (trimethylamine oxide), guanidinium, Cl^- and sucrose is obtained using the volume segment of Connolly's Molecular Surface program.
7. The tables of Creamer et al. do not extend to volumes. To get the excluded volumes of the unfolded chains use is made of a numerical rule for the ratio of excluded volume to ASA. This rule is very accurate for fully extended polypeptides but a possible uncertainly remains in applying it to the ensemble of structures of Creamer et al.
8. nK' is determined from the relation $\Delta B = \Delta X - \Delta n\langle K'\rangle$ followed by the determination of K-1 and K itself, which is an average association constant over all the sites.
9. K is defined in terms of volume fraction concentration units:

$$K = \frac{[3]}{[1]}\frac{\varphi_1}{\varphi_3} \quad \text{or} \quad \frac{[3]}{[1]} = K_j\frac{\varphi_1}{\varphi_3}$$

because the volume fraction provides a better measure of the probability for site occupation. Compare with Eq. (9). φ_3/φ_1 is the ratio of occupations that would take place by random chance. K is the multiplication factor that converts φ_3/φ_1 to the actual ratio of occupations. This differs from the random case because of the difference in interaction of the two components. K may be greater or less than 1 depending on the relative occupancies of [1] and [3].

The principle results of the calculation are the values for the average intrinsic association constant, $\langle K\rangle$ and the evaluation of the relative contributions of the excluded volume and $n\langle K'\rangle$, the genuine interaction with the cosolvent. Fig. 2 presents the results for a protein denaturant (urea) and a protein stabilizer (TMAO). Each bar graph has three components. The one on the left measures the change in excluded volume which is always positive for a normal unfolding; the central one is the sum of the contributions of the contact interactions with exposed sites ($\Delta\langle K'\rangle$), which can be positive or negative; and the right one is the algebraic sum of the other two. It is proportional to the experimental m or Δb° values for the unfolding.

Fig. 2a represents the urea denaturation of ribonuclease A. The m value was taken from [16], and the structure 1RND of the Protein Data Bank was used for volume and area calculations. We note that both the excluded volume and the interaction term are much larger than the virial coefficient itself. It is the latter which is actually measured in experiments via the Pace plot. The conclusion is that the interaction with protein is much larger than has been deduced from interpretations ignoring excluded volume. For urea–ribonuclease roughly 80% of the contact free energy is used to cancel out the excluded volume. Only the residual 20% contributes to m or b°. On this basis we can define a denaturing agent as a substance that has a favorable interaction with exposed groups that is strong enough to overcome the excluded volume, which favors the folded form.

The values of $\langle K\rangle$ that come out of this analysis are remarkably uniform as can be seen in Table 1. This is evidence for the uniformity of composition of the inner surfaces of the proteins. Of the sample of proteins only staphylococcal nuclease shows deviations. This is not surprising since this protein displays an abnormally high m value and unusual stability properties. Even with the enhanced contribution of the binding component, which results from the present analysis, the population of solvent species in the first layer is not much different from that of bulk solvent. From Table 1 and the volume fraction of urea we calculate that at the midpoint of the ribonuclease transition the fraction of sites

that have contact with urea molecules is approximately 36%. For a random system, in the absence of preferential interactions, it would be equal to 32%. This is a relatively small perturbation of the solvent composition in the contact region.

An additional feature of the analysis is that K' is sufficiently small that a linear plot of ΔG_{unf} vs. C is predicted.

The right side of Fig. 2 shows the solvation behavior of ribonuclease T1 in the presence of a stabilizer, TMAO. Calculations were made with the PDB structure 9RNT. The possibility of this calculation is dependent on a very special type of experiment performed by Bolen's group [38]. These investigators prepared an altered form of ribonuclease T1 that does not fold into a stable conformation. By adding TMAO they were able to refold the protein into a form that is enzymatically active and has the solution physical properties of the folded protein. They were also able to measure the m value (negative for inducing folding), which is an essential part of the analysis outlined above. Note that though this protein is smaller than ribonuclease A, the change in excluded volume is larger because the mean radius of the TMAO molecule is larger than that of urea. The result is the dominance of the excluded volume term. The average association constant per site, $\langle K \rangle$, is 1.130 for this system, which is to be compared with the value of 1.208. Since the contribution to the free energy is proportional to

Table 1
Average $\langle K \rangle$ for five proteins

Protein	$\langle K \rangle$
Ribonuclease T1	1.208
Ribonuclease A	1.204
HEW lysozyme	1.211
Staph. Nuclease	1.260
T4 lysozyme	1.212

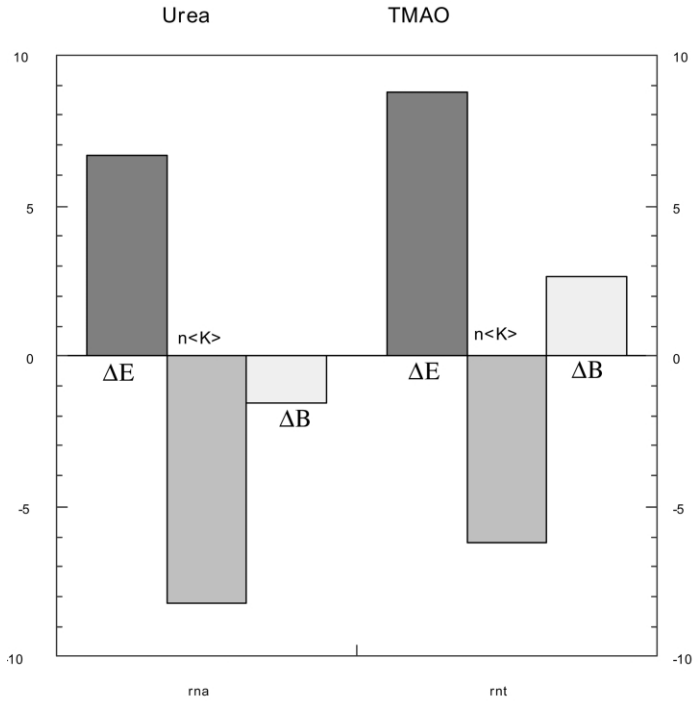

Fig. 2. The composition of the virial coefficient. The ordinate represents the magnitude of a virial component: $n\langle K \rangle$, X, or B. The units are inverse molarity. The left half of the figure deals with the interaction of urea with exposed sites of ribonuclease A. The right side shows the results for TMAO interacting with ribonuclease T1. See text for discussion.

K-1, this indicates that the TMAO interaction is on average 62% of that for urea.

Bolen's group has measured the free energies of transfer of amino acid side chains and the peptide group into TMAO solutions. With this they have analyzed the stabilization of ribonuclease T1 [39]. Their results probably indicate a larger unfavorable interaction with TMAO than that obtained in the present analysis, but direct comparisons will require careful and sophisticated analysis because of different interpretive methods.

At any rate, as far as the author is concerned, this is the present status of this long development. The interaction of folded proteins with co-solvents is a balance between excluded volume and weak binding (corrected for random occupancy). It is the algebraic sum of these two effects that controls the behavior of the protein. For denaturants the intrinsic interaction free energy is considerably larger than the observed free energy because of cancellation by excluded volume. When $n\langle K'\rangle$ is larger than the excluded volume, the cosolvent is a denaturant; when it is smaller, the co-solvent is a stabilizing osmolyte.

9. Conclusion

The work discussed above represents a narrow thread that runs through a broad and major field. There have been many other approaches with the same goal: the free energy of transfer method [40–42], the correlation of $\Delta\overline{G}_{unf}$ with exposed surface area [17,42–44], the correlation of $\Delta\overline{G}_{unf}$ with the effects of solutes on surface tension [45], the domain model which thermodynamically investigates the equilibrium of the solvation domain with the bulk medium [46,47], papers which emphasize the importance of excluded volume especially for osmolytes [29]. A recent review has discussed many of these methods [48]. The work of Saunders et al. has the same general program as Section 8 of this paper, but differs in execution [30].

The main characteristic of the work in this short history is the attempt to comprehend solvent interactions in terms of molecular events: one or more molecules of water in contact with a protein are replaced by a cosolvent molecule. On the one hand, one wants a chemist's description in terms of association, on the other, a link with molecular statistical mechanics. The bridge between these descriptions will be presented in a later publication. I would like to think the results are an intrinsic and important part of the large picture, but realize also that most of the real progress has come from scores of outstanding experimental papers. The current standards of data collection are very high. The problem is now viewed very broadly via a spectrum of co-solvents that vary from the strongest denaturants to the strongest stabilizers. If this 50-year history has demonstrated anything it is that 'details of molecular interaction' are hard to come by in complex systems. We must wait for the day when the molecular models, stability relations, the calorimetric data and transfer free energies are all comprehensible in terms of the same basic model.

Acknowledgments

I would like to thank Buzz Baldwin, Wayne Bolen, Nick Pace and Jeffrey Myers for comments on the manuscript and suggestions concerning the new material of Section 8. Fred Richards pointed out the possibility of using surface area programs to obtain excluded volumes. J.W. Ponder and M.L. Connolly provided advice in implementing this suggestion. The treatment of the unfolded chain benefited from a discussion with George Rose.

References

[1] R.B. Simpson, W. Kauzmann, J. Am. Chem. Soc. 75 (1953) 5139–5152.
[2] K.H. Frensdorff, M.T. Watson, W. Kauzmann, J. Am. Chem. Soc. 75 (1953) 5167–5172.
[3] H. Neurath, J.P. Greenstein, F.W. Putnam, J.O. Erickson, Chem. Revs. 34 (1944) 157–264.
[4] F.G. Hopkins, Nature 126 (1930) 328–330.
[5] H. Wu, Chinese. J. Physiol. 5 (1931) 321–344.
[6] H. Eyring, A.E. Stearn, Chem. Rev. 24 (1939) 253.
[7] Spiro, Zeit. für Physiol. Chem. 30 (1900) 182
[8] K. Linderstrom-Lang, J.A. Schellman, Biochim. et Biophys. Acta 15 (1954) 156–157.
[9] J.A. Schellman, Compt. rend. Lab. Carlsberg Ser. C him. 30 (1958) 439–449.
[10] F.T. Gucker, H.B. Pickard, J. Am. Chem. Soc. 62 (1940) 1464–1471.
[11] G. Scatchard, W. Hamer, S. Wood, J. Am. Chem. Soc. 60 (1938) 3061.

[12] J.A. Schellman, Compt. rend. Trav. Carlsberg. Lab (Ser. Chim.) 29 (1955) 230–259.
[13] K. Aune, C. Tanford, Biochemistry 8 (1969) 4586–4590.
[14] C. Tanford, Advan. Protein Chem. 24 (1970) 1–95.
[15] J.A. Schellman, Biopolymers 14 (1975) 999–1018.
[16] R.F.J. Greene, C.N. Pace, J. Biol. Chem. 249 (1974) 5388–5393.
[17] J.A. Schellman, Biopolymers 17 (1978) 1305–1322.
[18] E.F. Casassa, H. Eisenberg, Adv. Protein Chem. 19 (1964) 287–395.
[19] H. Eisenberg, Biological Macromolecules and Polyelectrolytes in Solution, Clarendon Press, Oxford, 1976.
[20] I.V. Baskakov, D.W. Bolen, Protein Sci. 8 (1999) 1314–1319.
[21] J.A. Schellman, Biopolymers 26 (1987) 549–559.
[22] J.A. Schellman, Biophys. Chem. 37 (1990) 121–140.
[23] J.O. Hirschfelder, C.F. Curtiss, R.B. Bird, Molecular Theory of Gases and Liquids, John Wiley & Sons, New York, 1954.
[24] B.H. Zimm, J. Chem. Phys. 14 (1946) 164–179.
[25] P.J. Flory, Principles of Polymer Chemistry, Cornell Univ. Press, Ithaca, 1953.
[26] C. Tanford, Physical Chemistry of Macromolecules, Wiley, New York, 1961.
[27] A.P. Minton, in: G.K. Ackers, M.J. Johnson (Eds.), Methods in Enzymology, Academic Press, New York, 1998, pp. 127–149.
[28] M.T. Record Jr, E.S. Courtenay, S. Cayley, Trends Biochem. Sci. 5 (1998) 190–194.
[29] P.R. Wills, W.D. Comper, D.J. Winzor, Arch. Biochem. Biophys. 300 (1993) 206–212.
[30] A.J. Saunders, P.R. Davis-Searles, D.L. Allen, G.J. Pielak, D.A. Erie, Biopolymers 53 (2000) 293–307.
[31] B. Lee, F.M. Richards, J. Mol. Biol. 55 (1971) 379–400.
[32] F.M. Richards, Ann. Rev. Biophys. Bioeng. 6 (1977) 151–176.
[33] M.L. Connolly, J. Am. Chem. Soc. 107 (1985) 1118–1124.
[34] M.L. Connolly, J. App. Cryst. 16 (1983) 548–558.
[35] S. Lapanje, C. Tanford, J. Am. Chem. Soc. 89 (1967) 5030–5033.
[36] Y. Qu, C.L. Bolen, D.W. Bolen, Proc. Natl. Acad. Sci. USA 95 (1998) 9268–9273.
[37] T.P. Creamer, R. Srinivasan, G.D. Rose, Biochemistry 36 (1997) 2832–2835.
[38] I. Baskakov, D.W. Bolen, J. Biol. Chem. 273 (1998) 4831–4834.
[39] A. Wang, D.W. Bolen, Biochemistry 36 (1997) 9101–9108.
[40] Y. Nozaki, C. Tanford, J. Biol. Chem. 238 (1963) 4074–4080.
[41] C.N. Pace, Biochemistry 40 (2001) 310–313.
[42] Y. Liu, D.W. Bolen, Biochemistry 34 (1995) 12884–12891.
[43] D. Shortle, H.S. Chan, K.A. Dill, Protein Sci. 1 (1992) 201–215.
[44] J.K. Myers, C.N. Pace, J.M. Scholtz, Protein Sci. 4 (1995) 2138–2148.
[45] Y. Kita, T. Arakawa, T.-Y. Lin, S.N. Timasheff, Biochemistry 33 (1994) 15178–15189.
[46] M.T. Record Jr, C.F. Anderson, Biophys. J. 68 (1995) 786–794.
[47] E.S. Courtenay, M.W. Capp, R.M. Saecker, M.T. Record Jr, Proteins: Struct. Funct. Genet. Suppl. 4 (2000) 72–85.
[48] P.R. Davis-Searles, A.J. Saunders, D.A. Erie, D.J. Winzor, G.J. Pielak, Ann. Rev. Biomol. Struct. 30 (2001) 271–306.
[49] F.G. Hopkins, Nature 126 (1930) 383–384.