

A molecular mechanism for osmolyte-induced protein stability

Timothy O. Street*, D. Wayne Bolen†, and George D. Rose**

*T. C. Jenkins Department of Biophysics, The Johns Hopkins University, Jenkins Hall, 3400 North Charles Street, Baltimore, MD 21218; and †Department of Human Biological Chemistry, University of Texas Medical Branch, 301 University Boulevard, 5.154 Medical Research Building, Galveston, TX 77555-1052

Communicated by Carl Frieden, Washington University School of Medicine, St. Louis, MO, July 25, 2006 (received for review June 28, 2006)

Osmolytes are small organic compounds that affect protein stability and are ubiquitous in living systems. In the equilibrium protein folding reaction, unfolded (U) \rightleftharpoons native (N), protecting osmolytes push the equilibrium toward N, whereas denaturing osmolytes push the equilibrium toward U. As yet, there is no universal molecular theory that can explain the mechanism by which osmolytes interact with the protein to affect protein stability. Here, we lay the groundwork for such a theory, starting with a key observation: the transfer free energy of protein backbone from water to a water/osmolyte solution, Δg_{tr} , is negatively correlated with an osmolyte's fractional polar surface area. Δg_{tr} measures the degree to which an osmolyte stabilizes a protein. Consequently, a straightforward interpretation of this correlation implies that the interaction between the protein backbone and osmolyte polar groups is more favorable than the corresponding interaction with nonpolar groups. Such an interpretation immediately suggests the existence of a universal mechanism involving osmolyte, backbone, and water. We test this idea by using it to construct a quantitative solvation model in which backbone/solvent interaction energy is a function of interactant polarity, and the number of energetically equivalent ways of realizing a given interaction is a function of interactant surface area. Using this model, calculated Δg_{tr} values show a strong correlation with measured values ($R = 0.99$). In addition, the model correctly predicts that protecting/denaturing osmolytes will be preferentially excluded/accumulated around the protein backbone. Taken together, these model-based results rationalize the dominant interactions observed in experimental studies of osmolyte-induced protein stabilization and denaturation.

organic osmolytes | osmolyte mechanism | protein folding

The equilibrium protein folding reaction, unfolded (U) \rightleftharpoons native (N), is not an ordinary chemical reaction because no covalent bonds are made or broken in the interconversion between N and U. Instead, protein denaturation/renaturation is just a reequilibration between the unfolded and folded populations under changed solvent conditions. Accordingly, a thermodynamic description of protein folding can be framed in terms of solvent interactions with the unfolded and native states (1–3).

Osmolytes are small organic compounds that exert a dramatic influence on the protein folding reaction, again without making or breaking covalent bonds. Protecting osmolytes push the folding equilibrium toward N, whereas denaturing osmolytes push the equilibrium toward U. Both types of osmolytes are of utmost significance. Protecting osmolytes are ubiquitous in nature, where they play a vital role in stabilizing intracellular proteins against a wide variety of adverse environmental conditions (4–7). Alternatively, urea, a denaturing osmolyte found naturally in mammalian kidney, has been a key reagent throughout the long history of solvent denaturation studies (3, 8–10).

The solution thermodynamics of protein/osmolyte mixtures has been well characterized in the literature (3, 11–19). In the emerging view (12, 17), protecting osmolytes raise the free energy of the unfolded state, favoring the folded population, whereas denaturing osmolytes lower the free energy of the unfolded state, favoring the unfolded population. Accordingly,

protecting/denaturing osmolytes interact unfavorably/favorably with the unfolded state, resulting in preferential depletion/accumulation of osmolyte proximate to the protein surface. Such osmolyte-induced behavior has been well characterized in thermodynamic terms, but thermodynamics is a descriptive science, deliberately devoid of mechanism. As yet, there is no universal theory that can account for the mechanism by which osmolytes interact with the protein to affect stability.

What specific molecular interactions in a protein–osmolyte–water solution stabilize/destabilize the unfolded state of proteins? An important clue comes from recent work showing that the osmolyte effect operates predominantly on the protein backbone, a component common to all residues (11–13). This conclusion was reached by measuring transfer free energies, Δg_{tr} , of backbone models from water to 1 M osmolyte solutions. Although side chains do play a role, it is primarily the backbone transfer free energy that determines the extent to which osmolytes either stabilize (i.e., $\Delta g_{tr} > 0$) or destabilize (i.e., $\Delta g_{tr} < 0$) the protein relative to an equivalent aqueous solution.

Thus, the backbone Δg_{tr} value is the key metric for evaluating the relative denaturing/stabilizing strength of different osmolytes. The thermodynamic reference state for this metric is given by interactions of the peptide backbone unit with solvent water. When that backbone unit is transferred from water to an aqueous osmolyte solution, the very presence of a molecule that experiences backbone interactions which differ from corresponding interactions with water either raises (for a protecting osmolyte) or lowers (for a denaturing osmolyte) the Δg_{tr} value relative to this reference state. Furthermore, given the well defined nature of the two solvent systems, the resultant Δg_{tr} will arise solely from differences between backbone/water and backbone/water/osmolyte interactions. Any molecular interpretation of osmolyte interactions must be consistent with this experimental reality.

We demonstrate that Δg_{tr} values for a wide variety of osmolytes are negatively correlated with their fractional polar surface area (SA), $f_{polar}^{osmolyte}$. The correlation suggests that polar and nonpolar osmolyte surfaces interact with the protein backbone at different energies, and that the extent of interaction is related to interactant SA. Specific instances in the literature are in known agreement with this plausible idea. For example, the polar molecule urea that has long been known to interact favorably with the amide backbone of proteins (20, 21). Also, in a related correlation, Record and colleagues (18) noted that Δg_{tr} for glycine betaine is proportional to polar SA. To quantify our idea, we develop a simple statistical mechanics backbone solvation model in which the interaction energy depends on interac-

Author contributions: T.O.S. and G.D.R. designed research; T.O.S. performed research; D.W.B. contributed key ideas; T.O.S. and G.D.R. analyzed data; and T.O.S., D.W.B., and G.D.R. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Abbreviations: TMAO, trimethylamine *N*-oxide; SA, surface area.

†To whom correspondence should be addressed. E-mail: grose@jhu.edu.

© 2006 by The National Academy of Sciences of the USA

Table 1. Solvent accessible surface areas and Δg_{tr} values of osmolytes

Osmolyte*	SA ₊ , Å ^{2†}	SA ₋ , Å ^{2†}	SA ₀ , Å ^{2†}	ΣSA, Å ^{2†}	Δg_{tr} , cal/mol [‡]
TMAO	0.0	43.2	168.4	211.6	89 ± 2
Betaine	3.6	82.7	166.7	253.0	65 ± 3
Sucrose	0.0	336.9	137.3	474.2	56 ± 6
Trehalose	0.0	340.6	145.2	485.8	54 ± 8
Sarcosine	24.5	43.3	141.6	209.4	50 ± 2
Sorbitol	0.0	233.6	97.8	331.4	43 ± 7
Proline	24.5	88.9	133.5	246.9	40 ± 8
Glycerol	0.0	142.7	84.1	226.8	22 ± 8
Urea	111.8	51.4	11.6	174.8	-41 ± 2
Guanidine	167.8	0.0	11.6	179.4	-59

*The 10 osmolytes listed in Fig. 1.

[†]Osmolyte surface areas in Å² with partial positive, negative, and neutral charge are indicated by SA₊, SA₋, and SA₀, respectively.

[‡]Total surface area in Å² = sum of SA₊, SA₋, and SA₀.

[§] Δg_{tr} is the free energy change that accompanies the transfer of a backbone unit from water to a 1 M osmolyte solution. Uncertainty in Δg_{tr} values is based on two independent measurement techniques (11). Value for guanidine was provided by S. Sarker (personal communication).

tant polarity, and the interaction degeneracy (i.e., the number of energetically equivalent ways of realizing the interaction) depends on the corresponding interactant SAs. Our goal is to learn whether this minimal model, polar interactions in a statistical mechanical framework, is sufficient to account for the diversity of experimental phenomena associated with protecting and denaturing osmolytes.

Results

Calculations described below were performed by using x-ray structures of eight stabilizing osmolytes [trimethylamine *N*-oxide (TMAO), betaine, sarcosine, proline, trehalose, sucrose, glycerol, and sorbitol], a destabilizing osmolyte (urea) and a related denaturant (guanidine). The Δg_{tr} values for these 10 compounds have been measured by two independent methods in all cases except guanidine (11) (Table 1). Comparisons between osmolyte structures and the associated water to osmolyte Δg_{tr} values (Fig. 1) indicate no evident correlation with either the total osmolyte SA or its polar SA. For example, TMAO and urea (Fig. 1 *a* and *j*) have similar total SAs but opposite effects on protein stability. Likewise, sucrose (Fig. 1*c*), an intermediate stabilizer, has greater polar SA than either TMAO, a strong stabilizer, or urea, a strong denaturant.

However, there is a clear correlation ($R = 0.88$) between Δg_{tr} and $f_{polar\ surface}^{osmolyte}$ for these osmolytes (Fig. 2). Specifically, as $f_{polar\ surface}^{osmolyte}$ increases, the osmolyte interaction with the protein backbone becomes increasingly favorable (i.e., their Δg_{tr} value decreases). This correlation suggests that the interaction between the protein backbone and osmolyte polar groups is more favorable than the corresponding interaction with nonpolar

groups. Furthermore, the correlation suggests that the probability of interaction scales with interactant SA.

A Model for Solvent Interactions with the Protein Backbone. Given the chemical heterogeneity of these osmolytes, what accounts for the observed correlation between $f_{polar\ surface}^{osmolyte}$ and Δg_{tr} ? The most direct explanation would be the existence of a universal interaction mechanism. Accordingly, we propose a quantitative model for solvent (water and osmolyte) interactions with backbone polar groups (the amide nitrogen, bearing a partial positive charge, and the carbonyl oxygen, bearing a partial negative charge). Three types of protein/solvent interactions were defined for these two backbone groups: favorable, unfavorable, and neutral, having energies of -1, 1, and 0 kcal/mol, respectively. Favorable interactions occur between polar groups with opposite charges, unfavorable interactions are between polar groups with like charges, and neutral interactions involve nonpolar groups, as illustrated for TMAO in Fig. 3. The backbone amide nitrogen is assigned to have one solvent interaction site, whereas the larger carbonyl oxygen, with its two lone pair electrons, is assigned to have two such sites. However, the overall results do not change appreciably if both backbone groups are treated equivalently (discussed below). Solvent/solvent and backbone/backbone interactions are not included in the model.

A degeneracy term was included to quantify the number of ways of realizing a given backbone/solvent interaction in terms of the area of its available participating interactant surfaces (see *Methods*). To implement this contribution, polar and nonpolar SAs were calculated for each osmolyte, with polar surface further subdivided into contributions from groups with partial positive (nitrogen) and negative (oxygen) charges (Table 1).

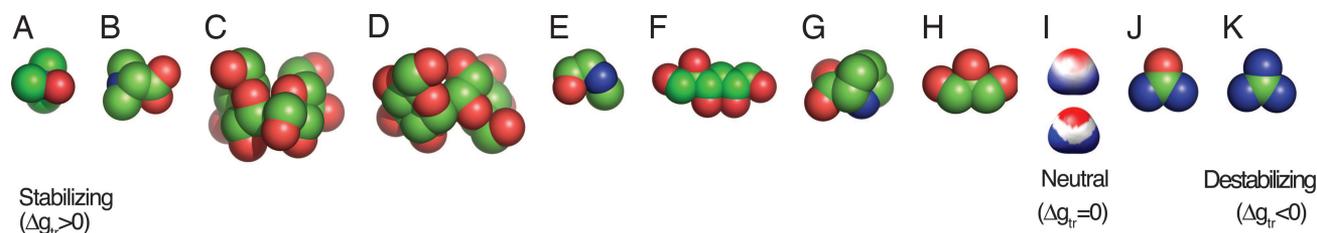


Fig. 1. Molecular structures of osmolytes. Protecting osmolytes are TMAO, betaine, sucrose, trehalose, sarcosine, sorbitol, proline, and glycerol (A–H), and denaturants are urea and guanidine (I–K). Compounds are ordered by their measured Δg_{tr} values (see Table 1), shown in space-filling representations and color-coded by atom type: oxygen (red), nitrogen (blue), and carbon (green). Water polarity is represented by its surface electrostatic potential (I, Upper), using a color saturation scale that runs from -0.07 (red) to 0.11 (blue) e/Å; white indicates neutral potential. The water surface is partitioned into discrete positive (red), negative (blue), and neutral (white) surfaces (I, Lower) using electrostatic potential cutoffs described in *Methods*.

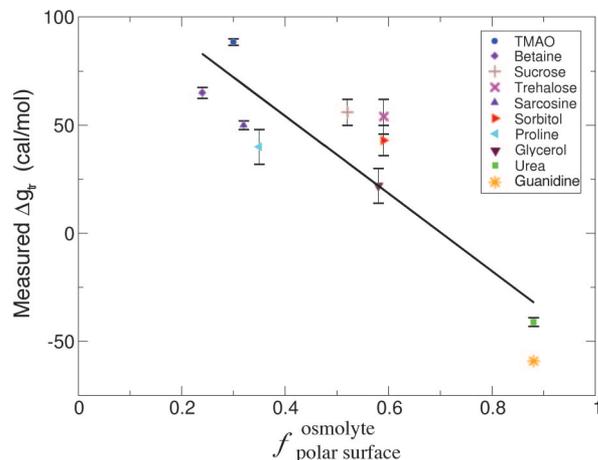


Fig. 2. The polar fraction of osmolyte surface correlates with measured Δg_{tr} values. Fractional polar SA, $f_{\text{polar surface}}^{\text{osmolyte}}$, is plotted against Δg_{tr} values from Table 1 for the 10 osmolytes in Fig. 1. The linear regression line (solid line) has a negative slope with a correlation coefficient of 0.88, indicating that backbone/osmolyte interactions become increasingly favorable as osmolytes become increasingly polar.

To treat water and osmolytes in a consistent manner, it was also necessary to subdivide a water molecule into polar and nonpolar regions. Although the decomposition of osmolyte surfaces into polar and nonpolar components is uncomplicated,

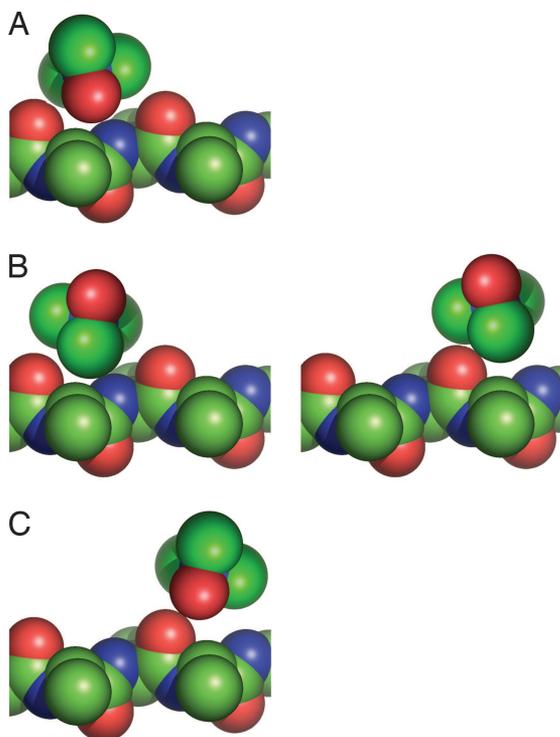


Fig. 3. Illustrating TMAO/backbone interactions. Interactions between an osmolyte, such as TMAO (upper molecule), and the protein backbone (lower structure) can be favorable, neutral, or unfavorable. Favorable interactions are between groups of opposite charge (A), neutral interactions involve at least one nonpolar group (B), and unfavorable interactions are between groups of like charge (C). Atoms are color-coded as in Fig. 1. A large fraction of the TMAO surface is nonpolar, affording more opportunities (i.e., a higher degeneracy) for this osmolyte to realize neutral interactions than either favorable or unfavorable interactions.

a similar decomposition of water requires a method to evaluate its surface charge distribution. To this end, an *ab initio* calculation of the water electrostatic potential and electron density was performed, as described in *Methods*. When this potential is mapped onto a space-filling model (22), distinct regions of positive (blue), negative (red), and neutral (white) charge are apparent (Fig. 1*t*, top structure). After applying polar and nonpolar cutoff values (see *Methods*), the water surface is found to have approximately equal regions of positive, negative, neutral charge (37%, 33%, and 29%, respectively; Fig. 1*t*, lower structure). Although the precise decomposition depends on the chosen electron density and the electrostatic potential threshold (data not shown), our overall results are quite insensitive to this decomposition, as discussed below. Consequently, the total water SA of 30 Å² (the area of a water-sized sphere of radius ≈1.5 Å) was subdivided into three equal-area regions of 10 Å² each.

Comparisons Between Model-Based and Measured Δg_{tr} Values. The model was tested by using it to calculate Δg_{tr} values and comparing them to experimentally measured values. For 1 M osmolyte concentrations, the calculated and measured Δg_{tr} values are in good agreement (Fig. 4*A*), with a correlation coefficient of 0.99, a substantial improvement from the corresponding correlation with $f_{\text{polar surface}}^{\text{osmolyte}}$ ($r = 0.88$). Although the slope of the linear regression line is less than unity (slope = 0.81), a slope of 1 with a correlation coefficient of 0.98 can be obtained if interaction energies are set to -1.5 , 1.5 , and 0 kcal/mol (for favorable, unfavorable, and neutral interactions, respectively). As a further test, Δg_{tr} values were calculated at osmolyte concentrations beyond 1 M (Fig. 4*B*, solid lines); these values compare favorably with the corresponding experimental values for sarcosine, urea, and guanidine (Fig. 4*B*, symbols), those osmolytes for which measurements beyond 1 M are available (13).

The success of the model in predicting Δg_{tr} values for a diversity of compounds across a range of concentrations is consistent with our hypothesis that interactions between the protein backbone and osmolytes are dominated by their SA and outer-group polarity. Moreover, this conclusion is insensitive to substantial changes in model parameters, as discussed below.

How Robust Are the Calculated Δg_{tr} Values? Our model has two adjustable parameters: the polar and nonpolar SAs associated with water, and the interaction energy between solvent and the protein backbone. To assess the sensitivity of calculated Δg_{tr} values to these parameters, both were varied extensively and Δg_{tr} values were recalculated.

Positive, negative, and neutral water SAs were simultaneously randomized within a 5- to 15-Å² interval, with solvent interaction energies fixed at their original values. The resulting Δg_{tr} values remain in good agreement with the measured values (Fig. 6, which is published as supporting information on the PNAS web site); ≈90% of the calculated Δg_{tr} values have correlation coefficients with measured values that exceed 0.80.

In another test, interaction energies were assigned random values in the range 0.5–5 kcal/mol (multiplied by -1 for favorable interactions), with neutral interaction energy and the polar and nonpolar water SAs held fixed at their original values. The correlation with measured Δg_{tr} values was recalculated for each new value (Fig. 7, which is published as supporting information on the PNAS web site), and again, model-based and measured Δg_{tr} values were found to be in good agreement; all correlation coefficients exceed 0.80.

As a final test, the correlation between model-based and measured Δg_{tr} values was recalculated under the alternative assumption that the backbone carbonyl oxygen provides only one solvent interaction site, not two. Water surface decomposition

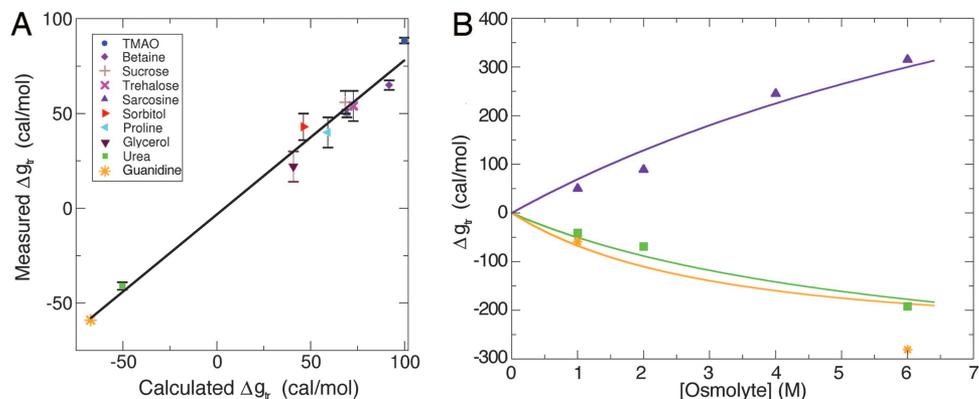


Fig. 4. Comparison between calculated and measured Δg_{tr} values for osmolytes. (A) Δg_{tr} values, calculated from the model, are plotted against experimentally determined values from Table 1. Good agreement is apparent. The linear regression line (solid line) is given by $\Delta g_{tr}^{measured} = 0.81 \Delta g_{tr}^{calculated} - 3.2$, with correlation coefficient 0.99. Data points corresponding to the 10 osmolytes in Fig. 1 are annotated in *Inset*. (B) Calculated Δg_{tr} values on higher osmolyte concentrations (>1 M) are plotted against available experimental data for sarcosine (indigo triangles), urea (green squares), and guanidine (orange asterisks). Solid lines were drawn from model-based Δg_{tr} values, extended beyond 1 M osmolyte concentrations.

and interaction energies were held fixed at their original values. Still, model-based and measured Δg_{tr} values remain highly correlated (Fig. 8, which is published as supporting information on the PNAS web site), although the correlation coefficient is reduced to 0.86.

Preferential Osmolyte Interactions with the Protein Backbone. Characteristically, stabilizing/destabilizing osmolytes are preferentially excluded/accumulated at the protein surface, respectively (14, 15, 23). To test this aspect of the model, the local concentration of osmolytes at backbone interaction sites was calculated for a 1 M osmolyte solution (see *Methods*). As shown in Fig. 5, stabilizing osmolytes are preferentially excluded from backbone polar groups, whereas denaturing osmolytes are preferentially accumulated there. In the model, the molecular basis of these preferential interactions is rooted in solvent/backbone interactions. Stabilizing osmolytes, such as TMAO, are preferentially excluded from the backbone because water is more likely than TMAO to interact favorably with backbone polar groups. Con-

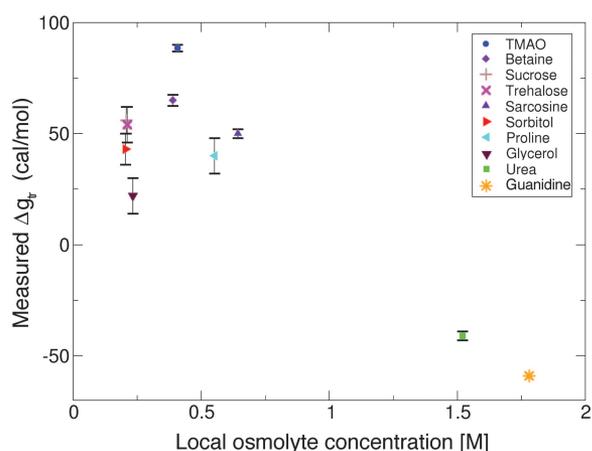


Fig. 5. Protecting/denaturing osmolytes are preferentially depleted/accumulated at the protein backbone. Concentration of osmolyte around the backbone in a 1 M osmolyte solution plotted against measured Δg_{tr} values from Table 1. Data points corresponding to the 10 osmolytes in Fig. 1 are annotated in *Inset*. The local osmolyte concentration is given by the scaled difference between $\langle O_{pref} \rangle$ and $\langle O_{bulk} \rangle$, described in *Methods*. It is apparent that the backbone concentration of protecting osmolytes ($\Delta g_{tr} > 0$) is comparatively depleted ($[\text{osmolyte}] < 1.0$ M), whereas that of denaturing osmolytes ($\Delta g_{tr} < 0$) is comparatively enriched ($[\text{osmolyte}] > 1.0$ M).

versely, destabilizing osmolytes, such as urea, are preferentially accumulated near the backbone because they have a stronger propensity to interact with the backbone than water. However, it should be noted that the model does not take nonpolar backbone regions into account, and interactions around these regions can also contribute to the actual local osmolyte concentration.

Discussion

The model presented here establishes a connection between bulk thermodynamic quantities and the molecular interactions that give rise to these quantities. In particular, it was developed to explain the experimentally determined backbone transfer free energies from water to osmolyte in terms of interactions between the protein backbone and water or osmolyte molecules. The model was validated by using it to predict significant, experimentally observed behavior. In addition, the model predicts that the free energy change for folding/unfolding will be linearly dependent on osmolyte concentration or approximately so, and for both protectants and denaturants, consistent with Pace's linear extrapolation model (9, 10). Our model also correctly predicts m values ($d\Delta g_{tr}/d[\text{osmolyte}]$) of opposite sign and approximately equal magnitude for proteins that are either forced to fold by sarcosine or denatured by urea, thereby accounting for the wide range of effects that natural osmolytes can exert on protein stability (12). Finally, Δg_{tr} values calculated by using the model correlate extremely well with experimental values ($R = 0.99$), illustrating that the relevant water–osmolyte–backbone energies are captured by the model.

Protein/osmolyte interactions are conspicuously weak, and in such cases, classical binding models are notoriously deficient (24). In particular, urea/backbone interactions would have apparent binding constants slightly greater than unity ($K_{binding} \approx 1.2$) (3), whereas protecting osmolyte/backbone interactions would have apparent binding constants slightly less than unity ($K_{binding} \approx 0.8$). In this weak-binding regime, free energy effects are ostensibly additive because neither type of osmolyte occupies a significant fraction of the backbone surface, so there is essentially no competition for backbone binding sites. Such additivity is observed in experiments (25) and is consistent with the model.

Many previous studies have related SA calculations to thermodynamic quantities associated with protein folding (26–32), motivated by the early observation that the transfer of nonpolar groups from organic to aqueous solvent is accompanied by an anomalous change in heat capacity, ΔC_p (33). The observed correlation between ΔC_p and nonpolar surface is often interpreted to mean

that water around nonpolar surfaces differs structurally and thermodynamically from water in bulk solution (34). In contrast, the model proposed here focuses on solvent interactions with the protein backbone, and SAs are only used here to estimate the binding-competent fraction of interacting molecules.

Several other types of solvent interactions affecting protein stability are neglected in our model, including crowding and excluded volume (35–37), the structure of water in osmolyte solutions (38, 39), side chain/solvent interactions, and binding situations in which a large osmolyte molecule can occlude more than one backbone unit. In addition, the model treats all nitrogens and oxygens as equally polar, an obvious simplification. However, despite such simplicity, the model captures key thermodynamic aspects of osmolyte behavior in a parameter-insensitive fashion. Therefore, it seems likely that the model's anchoring suppositions, solvent/protein interactions that depend on polarity and SA, are primarily responsible for the osmolyte effect in proteins.

Methods

Atom coordinates for TMAO, betaine, sucrose, trehalose, sarcosine, sorbitol, proline, glycerol, urea, and guanidine were obtained from the HICUP database (40). Their solvent accessible polar and nonpolar SAs were calculated by using PyMOL (41) with a probe radius of 1.4 Å (Table 1). Guanidine differs from other osmolytes investigated here because it is introduced into solution as a salt, guanidinium hydrochloride. To correct for the chloride ion associated with guanidine, a small, negatively charged surface of 30 Å² was added to the guanidine SA, although this addition does not dramatically change its calculated Δg_{tr} value (these values are –67 and –62 cal/mol with and without chloride ion addition, respectively).

The surface of a water molecule was defined at a threshold electron density of $\rho = 0.005$, giving a molecular volume of 11.5 Å³ (approximately the volume of a 1.5-Å sphere). The surface of water was decomposed into polar, nonpolar, and neutral regions by calculating the electrostatic potential and mapping it onto this surface. The thresholds used to delimit polar and nonpolar regions were defined by the boundary where the potential decays to $1/e$ of its minimum and maximum values, –0.023 and 0.037 e/Å, respectively. The neutral region was then defined as the complement to these two regions. Both the electrostatic potential and the electron density were calculated *ab initio* by using CPMD (42).

The average energy of the protein backbone in various osmolyte solutions was calculated by using a statistical mechanics model in which the backbone has three interaction sites: one at the amide nitrogen and two at the carbonyl oxygen; these sites are represented by the indices i, j , and k , respectively. At each interaction site, the solvent (either water or osmolyte) can present a positively (+), neutral (o), or negatively charged (–) surface. Accordingly, the indices i, j , and k can take the values +, o, or –, resulting in a total of $3^3 = 27$ possible microstates. Interactions between the backbone and solvent are assigned energies of –1, 1, and 0 kcal/mol corresponding to interactions between opposite, like, or uncharged groups, respectively. The interaction energies for each site are assumed to be additive and independent. These microstates and their associated energies and degeneracies are enumerated in Table 2.

The degeneracy of a solvent interaction at a particular backbone interaction site reflects the number of energetically equivalent ways of making that interaction. In our model, the degeneracy is given by the water or osmolyte SA that can participate in the interaction. Given that the three backbone interaction sites (i, j , and k) are independent, the total degeneracy (Ω_{ijk}) of a specific microstate, consisting of a +, o, or – solvent interaction at sites i, j , and k , will be the product of the degeneracies, as represented by their SAs, at individual interaction sites

Table 2. Solvent interactions with the protein backbone

N, i	Interaction site*		Microstate [†]	
	O ₁ , j	O ₂ , k	E , kcal/mol	Ω
+	+	+	–1	Ω_{+++}
+	+	o	0	Ω_{++o}
+	+	–	1	Ω_{+-}
+	o	+	0	Ω_{+o+}
+	o	o	1	Ω_{+oo}
+	o	–	2	Ω_{+o-}
+	–	+	1	Ω_{+--}
+	–	o	2	Ω_{+-o}
+	–	–	3	Ω_{+--}
o	+	+	–2	Ω_{o++}
o	+	o	–1	Ω_{o+o}
o	+	–	0	Ω_{o+-}
o	o	+	–1	Ω_{oo+}
o	o	o	0	Ω_{ooo}
o	o	–	1	Ω_{oo-}
o	–	+	0	Ω_{o-o}
o	–	o	1	Ω_{o--}
o	–	–	2	Ω_{o--}
–	+	+	–3	Ω_{-++}
–	+	o	–2	Ω_{-+o}
–	+	–	–1	Ω_{-+-}
–	o	+	–2	Ω_{-o+}
–	o	o	–1	Ω_{-oo}
–	o	–	0	Ω_{-o-}
–	–	+	–1	Ω_{--+}
–	–	o	0	Ω_{--o}
–	–	–	1	Ω_{---}

The 27 microstates and their associated energies and degeneracies.

*The single amide nitrogen and two carbonyl oxygen backbone interaction sites are indicated by N, O₁, and O₂, respectively. Solvent interactions with these sites are given by the charge of the interacting solvent surface: positive (+), negative (–), and neutral (o), and i, j , and k indices are varied over the range of values for these interactions.

[†] E is the energy of a given microstate, and Ω is its degeneracy.

$$\Omega_{ijk} = SA_i \cdot SA_j \cdot SA_k \quad [1]$$

These SAs will have a contribution from water ($SA_{w,i}$, $SA_{w,j}$, and $SA_{w,k}$) and a contribution from osmolyte ($SA_{o,i}$, $SA_{o,j}$, and $SA_{o,k}$) that depends on the osmolyte concentration. At Y molar:

$$SA_i = 55.5 \cdot SA_{w,i} + Y \cdot SA_{o,i} \quad [2]$$

$$SA_j = 55.5 \cdot SA_{w,j} + Y \cdot SA_{o,j} \quad [3]$$

$$SA_k = 55.5 \cdot SA_{w,k} + Y \cdot SA_{o,k} \quad [4]$$

In the model, +, o, and – water SAs are equal (10 Å²), and the corresponding osmolyte SAs are given in Table 1. The SA calculations treat the activity of water as a constant (i.e., that molarity of water in osmolyte solutions of varying concentration is ≈ 55.5 M), a plausible approximation at the low cosolvent concentrations used here.

The probability of any given microstate is given by

$$P_{ijk} = \frac{\Omega_{ijk} e^{-E_{ijk}/kT}}{\sum_i \sum_j \sum_k \Omega_{ijk} e^{-E_{ijk}/kT}} \quad [5]$$

where k is Boltzmann's constant and T is 298.15 K, the temperature at which Δg_{tr} values were measured experimentally. These probabilities can be used to calculate the average energy of the system

$$\langle E \rangle = \sum_i \sum_j \sum_k E_{ijk} p_{ijk}, \quad [6]$$

with Δg_{tr} values given by the difference between the average system energy at 0 and 1 M osmolyte concentrations. The average occupancy of osmolytes on the backbone interaction sites can also be calculated as

$$\langle O_{pref} \rangle = \sum_i \sum_j \sum_k p_{ijk} \left(\frac{SA_{o,i}}{SA_{o,i} + SA_{w,i}} + \frac{SA_{o,j}}{SA_{o,j} + SA_{w,j}} + \frac{SA_{o,k}}{SA_{o,k} + SA_{w,k}} \right). \quad [7]$$

This value can be compared with the expected osmolyte occupancy based solely on the bulk solution concentration (i.e., no

preferential interactions with the three backbone interaction sites)

$$\langle O_{bulk} \rangle = 3 \cdot \frac{SA_o}{SA_o + SA_w}. \quad [8]$$

The relative difference between $\langle O_{pref} \rangle$ and $\langle O_{bulk} \rangle$ yields a local osmolyte concentration when scaled to molarity.

All numerical calculations were performed in Python (www.python.org).

We thank Buzz Baldwin and two anonymous referees for suggestions. This work was supported by a Burroughs Wellcome predoctoral fellowship (to T.O.S.), National Institutes of Health Grant GM-49760 (to D.W.B.), and by the Mathers Foundation (G.D.R.).

1. Wu H (1931) *Chinese J Physiol* 5:321–344.
2. Mirsky AE, Pauling L (1936) *Proc Natl Acad Sci USA* 22:439–447.
3. Schellman JA (2002) *Biophys Chem* 96:91–101.
4. Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN (1982) *Science* 217:1214–1222.
5. Record MT, Jr, Courtenay ES, Cayley DS, Guttman HJ (1998) *Trends Biochem Sci* 23:143–148.
6. Record MT, Jr, Courtenay ES, Cayley S, Guttman HJ (1998) *Trends Biochem Sci* 23:190–194.
7. Hochachka PW, Somero GN (2002) *Biochemical Adaptation* (Oxford Univ Press, Oxford).
8. Tanford C (1968) *Adv Prot Chem* 23:121–282.
9. Greene RF, Jr, Pace CN (1974) *J Biol Chem* 249:5388–5393.
10. Santoro MM, Bolen DW (1988) *Biochemistry* 27:8063–8068.
11. Auton M, Bolen DW (2004) *Biochemistry* 43:1329–1342.
12. Auton M, Bolen DW (2005) *Proc Natl Acad Sci USA* 102:15065–15068.
13. Liu Y, Bolen DW (1995) *Biochemistry* 34:12884–12891.
14. Lee JC, Timasheff SN (1981) *J Biol Chem* 256:7193–7201.
15. Timasheff SN (1992) *Biochemistry* 31:9857–9864.
16. Makhatazde GI, Privalov PL (1992) *J Mol Biol* 226:491–505.
17. Bolen DW, Baskakov IV (2001) *J Mol Biol* 310:955–963.
18. Felitsky DJ, Cannon JG, Capp MW, Hong J, Van Wynsberghe AW, Anderson CF, Record MT, Jr (2004) *Biochemistry* 43:14732–14743.
19. Hong J, Capp MW, Anderson CF, Saecker RM, Felitsky DJ, Anderson MW, Record MT, Jr (2004) *Biochemistry* 43:14744–14758.
20. Nozaki Y, Tanford C (1963) *J Biol Chem* 238:4074–4080.
21. Roseman M, Jencks WP (1974) *J Am Chem Soc* 97:631–640.
22. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) *J Comput Chem* 25:1605–1612.
23. Lin TY, Timasheff SN (1994) *Biochemistry* 33:12695–12701.
24. Schellman JA (1987) *Biopolymers* 26:549–559.
25. Mello CC, Barrick D (2003) *Protein Sci* 12:1522–1529.
26. Choithia C (1974) *Nature* 248:338–339.
27. Eisenberg D, McLachlan AD (1986) *Nature* 319:199–203.
28. Spolar RS, Ha JH, Record MT, Jr (1989) *Proc Natl Acad Sci USA* 86:8382–8385.
29. Richards FM (1977) *Annu Rev Biophys Bioeng* 6:151–176.
30. Lee B, Richards FM (1971) *J Mol Biol* 55:379–400.
31. Hilser VJ, Gomez J, Freire E (1996) *Proteins* 26:123–133.
32. Robertson AD, Murphy KP (1997) *Chem Rev* 97:1251–1268.
33. Cohn EJ, Edsall JT (1943) *Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions* (Hafner, New York).
34. Gallagher KR, Sharp KA (2003) *J Am Chem Soc* 125:9853–9860.
35. Minton AP (1998) *Methods Enzymol* 295:127–149.
36. Schellman JA (2003) *Biophys J* 85:108–125.
37. Saunders AJ, Davis-Searles PR, Allen DL, Pielak GJ, Erie DA (2000) *Biopolymers* 53:293–307.
38. Batchelor JD, Olteanu A, Tripathy A, Pielak GJ (2004) *J Am Chem Soc* 126:1958–1961.
39. Bennion BJ, Daggett V (2004) *Proc Natl Acad Sci USA* 101:6433–6438.
40. Kleywegt GJ, Jones TA (1998) *Acta Crystallogr D* 54:1119–1131.
41. DeLano WL (2002) *The PYMOL Molecular Graphics System* (DeLano Sci, San Carlos, CA).
42. Hutter J, Alavi A, Deutsch T, Bernasconi M, Goedecker S, Marx D, Tuckerman M, Parrinello M (1997–2001) *CPMD* (Max-Planck-Institut für Festkörperforschung, Stuttgart), Vol 1997–2004.