FOR THE RECORD

The sarcosine effect on protein stability: A case of nonadditivity?

BEATRIZ IBARRA-MOLERO,^{1,3} ISABEL M. PLAZA DEL PINO,¹ BADREDINE SOUHAIL,^{1,2}
HASSAN O. HAMMOU,¹ AND JOSE M. SANCHEZ-RUIZ¹

¹Facultad de Ciencias, Departamento de Química Física, Universidad de Granada, 18071 Granada, Spain

²Département de Chimie (Faculté des Sciences), Université Abdelmalek Essaadi, BP 2121, 93000 Tétouan, Morocco

(RECEIVED July 19, 1999; FINAL REVISION February 11, 2000; ACCEPTED February 11, 2000)

Abstract: We have used differential scanning calorimetry to determine the effect of low concentrations ($C = 0\text{--}2\text{ M}$) of the osmolyte sarcosine on the Gibbs energy changes (ΔG) for the unfolding of hen-egg-white lysozyme, ribonuclease A, and ubiquitin, under the same buffer and pH conditions. We have also computed this effect on the basis of the additivity assumption and using published values of the transfer Gibbs energies for the amino acid side chains and the peptide backbone unit. The values thus predicted for the slope $\partial\Delta G/\partial C$ agree with the experimental ones, but only if the unfolded state is assumed to be compact (that is, if the accessibility to solvent of the unfolded state is modeled using segments excised from native structures). The additivity-based calculations predict similar $\partial\Delta G/\partial C$ values for the three proteins studied. We point out that, to the extent that this approximate constancy of $\partial\Delta G/\partial C$ holds, osmolyte-induced increases in denaturation temperature will be larger for proteins with low unfolding enthalpy (small proteins that bury a large proportion of apolar surface). The experimental results reported here are consistent with this hypothesis.

Keywords: additivity assumption; osmolyte effects; protein stability

When subjected to denaturing stresses, some organisms produce relatively low concentrations of certain low molecular weight substances (stabilizing osmolytes) that are able to enhance protein stability without detrimental effects on protein function (Yancey et al., 1982; Arakawa & Timasheff, 1985; Santoro et al., 1992; Timasheff, 1992; Plaza del Pino & Sanchez-Ruiz, 1995). This protein stabilizing effect is phenomenologically related to the fact that these substances are preferentially excluded from the protein surface (Arakawa & Timasheff, 1985; Timasheff, 1992). From a

microscopic viewpoint, the model-compound studies of Bolen and coworkers (Liu & Bolen, 1995; Wang & Bolen, 1997) strongly suggest that the stabilizing effect is due to the interaction of the peptide backbone with the osmolyte solutions. Thus, these authors used solubility measurements to calculate the Gibbs energy changes associated with the transfer of the amino acid side chains and peptide backbone models from water to osmolyte solutions. They found that transfer Gibbs energies for many side chains are negative (i.e., the osmolyte favors the exposure to solvent of those side chains), but that this effect is overcome by the transfer Gibbs energy of the peptide backbone unit, which is positive (the osmolyte disfavor the exposure to solvent of the peptide backbone, thus stabilizing the native state). In fact, it is possible to use the transfer Gibbs energies reported by Bolen and coworkers to calculate the osmolyte effect on the Gibbs energy change for protein unfolding (that is, the osmolyte effect on the thermodynamic stability of the protein):

$$\Delta G(O) - \Delta G(W) = \Delta_T G(U) - \Delta_T G(N) \quad (1)$$

$$\Delta_T G = \sum_i n_i \cdot \alpha_i \cdot \delta g_i. \quad (2)$$

In Equation 1 above, $\Delta G(W)$ and $\Delta G(O)$ are the unfolding Gibbs energies in “water” (aqueous buffer) and in the osmolyte solution, and $\Delta_T G(X)$ stands for the transfer Gibbs energy of the state X (U or N) from water to the osmolyte solution. Equation 1 can be easily derived from a simple thermodynamic cycle, as shown by Liu and Bolen (1995). Equation 2 allows us to calculate the transfer Gibbs energy of protein states, from the transfer Gibbs energies of the several “elements” (side chains, peptide unit) in which the protein is mentally divided. In Equation 2, n_i is the number of a given element (with transfer Gibbs energy equal to δg_i) in the protein and α_i is the average accessibility to the solvent of that element; α_i values in the native state can be derived from accessible surface area (ASA) calculations, while, in the simplest calculation, α_i values in the unfolded state are taken to be equal to unity (see Liu & Bolen, 1995).

Reprint requests to: Jose M. Sanchez-Ruiz, Facultad de Ciencias, Departamento de Química Física, Campus Fuentenueva s/n, 18071 Granada, Spain; e-mail: sanchezr@goliat.ugr.es.

³Present address: Department of Chemistry, Penn State University, University Park, Pennsylvania 16802.

Abbreviations: ASA, accessible surface area; DSC, differential scanning calorimetry; HEW, hen egg white; RNase, ribonuclease A; TMAO, trimethylamine N-oxide.

Clearly, Equation 2 is based upon the additivity assumption (the transfer Gibbs energy of a given state is assumed to be equal to sum of the transfer Gibbs energies of its constituents parts). Although the assumption of additivity has a long history in protein energetics (and other fields of physical chemistry), it has never been demonstrated to hold in general (for a discussion, see Liu & Bolen, 1995) and, in fact, in a recent analysis of protein folding energetics (D'Aquino et al., 1996), the authors suggest that large deviations from additivity predictions are expected to occur for polar group hydration.

In the specific case of osmolyte effects on stability, Wang and Bolen (1997) showed that the transfer Gibbs energies of native and unfolded ribonuclease T1 from water to a 1 M solution of the osmolyte TMAO can be predicted to a good degree of approximation on the basis of the additivity assumption. On the other hand, a comparison between unfolding Gibbs energy changes derived from experimental DSC measurements (Fig. 6 in Plaza del Pino & Sanchez-Ruiz, 1995) with the predictions of Equations 1 and 2 with $\alpha_i = 1$ for the unfolded state (Table 3 in Liu & Bolen, 1995) suggests that large deviations from additivity occur in the stabilization of ribonuclease A caused by the osmolyte sarcosine. To clarify this issue, we have carried out DSC measurements on three model proteins (bovine ubiquitin, HEW lysozyme, and ribonuclease A) under the same buffer conditions, in solutions with several concentrations of sarcosine, and we have compared the osmolyte effect on protein stability with the predictions of the additivity approach (Equations 1 and 2) using different models for the accessibility of the side chains and peptide backbone in the unfolded state.

Results and discussion: For the three proteins under study, we carried out DSC experiments in solutions with concentrations 0, 1, and 2 M of sarcosine. For each protein and sarcosine concentration, at least two (but usually more) DSC experiments were performed at different pH values within the range 3–4. The buffers employed were 10 mM glycine or 25 mM formate, in both cases with 0.2 M ClK (no significant difference between the results obtained in these two buffers was detected). Good fits of the two-state model to the experimental DSC data were obtained in all cases (for details on the procedures for DSC data analysis, see Ibarra-Molero et al., 1999b) and from them values of denaturation temperature (T_m , temperature at which the unfolding Gibbs energy is zero), unfolding enthalpy at the denaturation temperature (ΔH_m) and unfolding heat capacity were obtained.

Figure 1 shows a plot of ΔH_m vs. T_m for the three proteins. For each protein, the data obtained at different sarcosine concentrations fall in the same line, although, for reasons not clear to us, the scatter of the experimental ΔH_m data is somewhat larger in the case of ubiquitin (note, however, that, for ubiquitin, a larger number of ΔH_m data, spanning a wider temperature range, have been obtained). Furthermore, the unfolding heat capacity values determined by extrapolating to the T_m the native and unfolded baselines of the DSC profiles (ubiquitin: 4.2 ± 0.6 ; RNase: 4.7 ± 1.0 ; lysozyme: 6.2 ± 0.9) agree with the slopes of the plots of ΔH_m vs. T_m (ubiquitin: 3.6 ± 0.9 ; RNase: 6.2 ± 0.9 ; lysozyme: 7.0 ± 0.6 ; all values in $\text{kJ K}^{-1} \text{mol}^{-1}$). These results, together with our previous study on sarcosine effects on folding energetics (Plaza del Pino & Sanchez-Ruiz, 1995), indicate that, within the 0–2 M range, there is no significant sarcosine effect on unfolding enthalpy and unfolding heat capacity values and, therefore, we will assume in the following calculations that these parameters are functions of

Fig. 1. Plots of unfolding enthalpy vs. temperature for the three proteins studied in this work. The symbols refer to the sarcosine concentration: ●, 0 M; ■, 1 M; ▲, 2 M. For each sarcosine concentration, values corresponding to different pH values are included.

temperature only. Nevertheless, the possible errors introduced by this assumption will be discussed in the Appendix.

Figure 2 shows plots of denaturation temperature vs. pH for the three proteins and the three sarcosine concentrations employed in

Fig. 2. Plots of denaturation temperature vs. pH for the three proteins studied in this work. The numbers alongside the lines are the values of sarcosine molarity.

this work. Since the pH range studied is small, we have fitted the T_m vs. pH dependencies with a linear equation, which has been used to obtain by interpolation the T_m values corresponding to pH 3.5. Plots of T_m (pH = 3.5) vs. sarcosine concentration for the three proteins are given in Figure 3A. Unfolding enthalpies at T_m (pH = 3.5) are then obtained from the linear ΔH_m vs. T_m dependencies shown in Figure 1 and, finally, unfolding Gibbs energy changes at 25 °C and pH 3.5 are calculated by using the integrated Gibbs–Helmholtz equation with the average ΔC_p value:

$$\Delta G = \Delta H_m \left(1 - \frac{298}{T_m} \right) + \Delta C_p \left[298 - T_m - 298 \ln \left(\frac{298}{T_m} \right) \right]. \quad (3)$$

Plots of ΔG (25 °C, pH 3.5) vs. sarcosine concentration for the three proteins are shown in Figure 3B. In this work, we will describe the osmolyte effect on thermodynamic stability by the parameter $\partial\Delta G/\partial C$, where C is the concentration of sarcosine in

Fig. 3. The effect of sarcosine concentration (C) on the denaturation temperature at pH = 3.5 (A) and the unfolding Gibbs energy (ΔG) at 25 °C and pH 3.5 for the three proteins studied in this work. ΔT_m is the difference between the T_m value at a given sarcosine concentration and the T_m value in the absence of the osmolyte (the bars shown for the ΔT_m values represent the standard uncertainty intervals derived from the linear fits to the T_m vs. pH dependencies of Fig. 2). Likewise, $\Delta\Delta G$ is the difference between the ΔG value at a given sarcosine concentration and the ΔG value in the absence of the osmolyte (the bars shown for the $\Delta\Delta G$ values represent the standard uncertainty intervals calculated as $\delta\Delta\Delta G = C \cdot \delta(\partial\Delta G/\partial C)$, where $\delta(\partial\Delta G/\partial C)$ are the values for standard errors associated to the slope $\partial\Delta G/\partial C$ determined in the Appendix). Error bars are not shown when they are smaller than the size of the symbol.

moles per liter of solution. The values of $\partial\Delta G/\partial C$ obtained from the plots in Figure 3B are (in $\text{kJ mol}^{-1} \text{M}^{-1}$): 3.0 (lysozyme), 3.8 (ribonuclease A), and 5.2 (ubiquitin). It is important to note that, although the calculation of $\partial\Delta G/\partial C$ formally involves the values of the unfolding heat capacity changes (Equation 3), in practice the final outcome of the calculation does not depend on the value used for ΔC_p , the reason being any error in ΔC_p will affect in the same manner the ΔG values at different sarcosine concentrations and, hence, it will not affect the slope of the ΔG vs. C plots (see the Appendix for a detailed discussion). As a result, the errors associated to the calculated $\partial\Delta G/\partial C$ slopes are very low; thus, in the Appendix, these errors are estimated to be of the order of the tenths of $\text{kJ mol}^{-1} \text{M}^{-1}$.

It must also be noted that the ΔG values shown in Figure 3B have been calculated using the constant- ΔC_p Gibbs–Helmholtz extrapolation (Equation 3), while, strictly, unfolding heat capacity changes appear to change somewhat with temperature (see, for instance, Makhatadze & Privalov, 1995). We have recently pointed out and discussed (Ibarra-Molero & Sanchez-Ruiz, 1996; Ibarra-Molero et al., 1999b) the fact that we may expect discrepancies of a few kJ/mol between the ΔG values at room temperature calculated by constant- ΔC_p and nonconstant- ΔC_p Gibbs–Helmholtz extrapolation. Again, however, provided that sarcosine does not significantly affect the unfolding enthalpy and heat capacity changes within the concentration range studied (as seems to be the case here; see, however, the Appendix), the discrepancy will be essentially independent of sarcosine concentration and, therefore, it will not alter the value of the slope $\partial\Delta G/\partial C$. On the other hand, the value of $\partial\Delta G/\partial C$ calculated for ubiquitin might be distorted by an additional stabilizing effect due to chloride ions. Thus, anions have been shown to have a large stabilizing effect on ubiquitin at very acidic pH (Makhatadze et al., 1998) and sarcosine solutions at acidic pH must contain a certain amount of an anion (chloride ion in our case), since sarcosine has a protonation with a pK of ~ 2 (Christensen et al., 1976). Clearly, at a given pH, the amount of chloride ion “contributed” by sarcosine will increase with sarcosine concentration and the stabilizing effect of the anion will increase the value of the slope $\partial\Delta G/\partial C$. In this work, we have tried to minimize this chloride effect by working with a comparatively high background concentration of ClK (0.2 M in all cases) and by carrying out measurements within a pH range (3–4) at which the chloride stabilizing effect is not expected to be large (see Ibarra-Molero et al., 1999a) and which is removed from the pK value for sarcosine protonation (~ 2). Nevertheless, we cannot strictly rule out the possibility that the value of $\partial\Delta G/\partial C$ obtained for ubiquitin reflects both osmolyte and chloride ion stabilizing effects and that the osmolyte effect on thermodynamic stability is actually somewhat smaller than $5.2 \text{ kJ mol}^{-1} \text{M}^{-1}$. However, as will be clear from the discussion below, a smaller sarcosine effect on ubiquitin stability will not alter the general conclusions of this work.

We have also obtained theoretical estimates of the slope $\partial\Delta G/\partial C$ on the basis of

$$\frac{\partial\Delta G}{\partial C} = \frac{\Delta G(2M \text{ sar.}) - \Delta G(0M \text{ sar.})}{2M} \quad (4)$$

where the difference between the unfolding Gibbs energy in 2 M and 0 M sarcosine is calculated using Equations 1 and 2 and the transfer Gibbs energies of the side chains and the peptide unit reported by Liu and Bolen (1995). Accessibilities (α_i values) for

the native state were obtained from ASAs (see Materials and methods) and, in our first calculation, accessibilities in the unfolded state were assumed equal to unity. The $\partial\Delta G/\partial C$ values obtained in this manner (Fig. 4A) are two to five times larger than those derived from experiment. A possible explanation for this large discrepancy is simply that additivity does not hold. Another plausible explanation is that additivity does hold to a significant extent, but the unfolded state cannot be modeled as a fully solvated chain with accessibility equal to unity ($\alpha_i = 1$). Actually, this second hypothesis appears attractive in this case, since unfolded proteins are expected to “contract” in solutions of stabilizing osmolytes, as has been experimentally shown by Qu et al. (1998). Creamer et al. (1995, 1997) have recently discussed the issue of the accessibility to solvent of unfolded states. These authors suggest that the use of tripeptide models overestimates the unfolded state ASA and propose instead the use of two limiting models that are expected to bracket the unfolded state ASA between two reliable limits. The upper limit is obtained by modeling the protein in an extended

Fig. 4. Experimental and predicted values of the sarcosine effect on unfolding Gibbs energy ($\partial\Delta G/\partial C$), as well as backbone and side-chain contributions to the predicted value. Different models for the accessibility to solvent of the unfolded state have been used. **A:** Accessibilities equal to unity. **B:** Rose's upper limit. **C:** Rose's lower limit. The meaning of the labels is as follows: B, backbone contribution; S, side-chain contributions; note that we have plotted separately the stabilizing (positive) side-chain contributions (arising from Ala, Val, Leu, Ile, and Met) and the destabilizing (negative) side-chain contributions; T, the value of ($\partial\Delta G/\partial C$) predicted theoretically (sum of the backbone and side-chain contributions); E, the value of ($\partial\Delta G/\partial C$) calculated from the experimental DSC studies (that is, the slopes of the plots in Fig. 3B). The bar shown for the ubiquitin experimental value represents an upper-limit estimate of the total standard uncertainty, calculated taking into account the possible error introduced by neglecting the osmolyte concentration effect on ΔH and ΔC_p (see the Appendix for details).

conformation, with backbone dihedrals set to $\phi = -120^\circ$ and $\Psi = -120^\circ$ and side-chain torsions set to 180° . The lower limit is derived from the ASA values in chain segments of length 17 amino acids excised in their native conformation from a set of 43 proteins. We have repeated the calculation of $\partial\Delta G/\partial C$ (Equations 1 and 2), but using the accessibilities in the unfolded state (see Materials and methods) derived from Rose's upper and lower limits; the results are compared with the experimental values in Figures 4B and 4C. Surprisingly, the experimental $\partial\Delta G/\partial C$ results are close to the values predicted by using the lower limit; actually, the experimental values for lysozyme and RNase are somewhat lower than the lower-limit prediction, while the experimental value for ubiquitin is somewhat higher (see, however, the above discussion on chloride ion effects). Thus, if additivity is assumed to hold in this case, the experimental $\partial\Delta G/\partial C$ values suggest a very compact unfolded state with comparatively little exposure to solvent.

It is also worth noting that similar $\partial\Delta G/\partial C$ values are theoretically predicted (and experimentally found) for the three proteins studied (Fig. 4). To some extent, this appears to be a consequence of cancellation between stabilizing and destabilizing contributions; thus, the peptide backbone contribution is smaller for ubiquitin (the smallest protein in the set), but so is (in absolute value) the destabilizing side-chain contribution. Differences between the proteins studied, however, are more pronounced in the plot of denaturation temperature vs. sarcosine concentration (see Fig. 3A). Thus, the values of the slope dT_m/dC are (in $K M^{-1}$): 7.1 ± 0.4 for ubiquitin, 3.4 ± 0.1 for RNase, and 2.5 ± 0.1 for lysozyme. We have previously pointed out (Plaza del Pino & Sanchez-Ruiz, 1995) that, although protein stability is often defined as the unfolding Gibbs energy at a given reference temperature ($25^\circ C$ usually), what matters for many practical or technological purposes is the denaturation temperature (what we refer to in this specific context as “operational stability”). We show in the Appendix (see Equation A.5) that the relation between the osmolyte effect on thermodynamic stability ($\partial\Delta G/\partial C$ at $25^\circ C$) and the osmolyte effect on operational stability (dT_m/dC) can be written as

$$\left(\frac{dT_m}{dC}\right) = \frac{T_m}{\Delta H_m} \frac{T_m}{298} \left(\frac{\partial\Delta G}{\partial C}\right)_{298} \quad (5)$$

which shows that, for a given $\partial\Delta G/\partial C$ value, the osmolyte effect on denaturation temperature will be higher, *the lower the unfolding enthalpy value*. This result helps to understand the large sarcosine effect on ubiquitin T_m value, since the unfolding enthalpy values for this protein are comparatively low (see Fig. 1), a fact that can be attributed to the two following factors: (1) it is a small protein, and we may expect unfolding enthalpies at a given temperature to scale roughly with protein size; (2) it buries a large proportion of apolar ASA upon folding (see Table I in Makhatadze & Privalov, 1995) and the coefficient of the buried apolar area in parametric equations for unfolding enthalpy is negative in a wide temperature range; for instance, empirical statistical analysis of the thermodynamic data set for protein unfolding suggests that the enthalpy change at $60^\circ C$ (in cal/mol) can be expressed in terms of the apolar and polar ASA buried upon folding as $-8.44 \times \Delta ASA_{ap} + 31.4 \times \Delta ASA_{pol}$ (Xie & Freire, 1992; Hilser et al., 1996).

In general, our analysis suggests that increases in denaturation temperature caused by osmolytes will be larger for proteins with low unfolding enthalpy (small proteins that bury a large proportion of apolar area upon folding), at least to the extent that the osmolyte

effect on “thermodynamic stability” ($\partial\Delta G/\partial C$) is similar for the different proteins.

Materials and methods: Bovine ubiquitin, HEW lysozyme, and ribonuclease A were purchased from Sigma (St. Louis, Missouri) and were used without further purification. Preparation of protein solutions, sarcosine solutions, and measurement of protein concentrations have been described elsewhere (Plaza del Pino & Sanchez-Ruiz, 1995; Ibarra-Molero & Sanchez-Ruiz, 1996; Ibarra-Molero et al., 1999a, 1999b). Differential scanning calorimetry experiments were performed with a VP-DSC microcalorimeter from MicroCal (Northampton, Massachusetts). The samples were degassed at room temperature prior to calorimetric experiments. Calorimetric cells (operating volume ~ 0.5 mL) were kept under an excess pressure of 30 psi to prevent degassing during the scan. The scan rate was always 1.5 K/min, and the protein concentrations were in the range 1–2 mg/mL. Further details regarding the calorimetric experiment and the calorimetric data analysis can be found elsewhere (Ibarra-Molero et al., 1999b). Accessible surface areas calculations were performed as described (Ibarra-Molero et al., 1999a). Accessibilities of side chains and peptide units in the native state of the proteins studied were calculated as the ratio between the ASA value in the native protein and the ASA value in a Gly-X-Gly tripeptide. In the case of the upper-limit and lower-limit models for the unfolded state, accessibilities were calculated in the same manner, but using the ASA values given by Creamer et al. (1995, 1997).

Acknowledgments: This work was supported by Grant PB96-1439 from the DGES (Spanish Ministry of Science and Culture) and by the “FUNDACION RAMON ARECES.”

References

- Akawa T, Timasheff SN. 1985. The stabilization of proteins by osmolytes. *Biophys J* 47:411–414.
- Christensen JJ, Hansen LD, Izatt RM. 1976. *Handbook of proton ionization heats*. New York: Wiley.
- Creamer TP, Srinivasan R, Rose GD. 1995. Modeling unfolded states of peptides and proteins. *Biochemistry* 34:16245–16250.
- Creamer TP, Srinivasan R, Rose GD. 1997. Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry* 36:2832–2835.
- D'Aquino JA, Gomez J, Hilser VJ, Lee KH, Amzel LM, Freire E. 1996. The magnitude of the backbone conformational change in protein folding. *Proteins* 25:143–156.
- Hilser VJ, Gomez J, Freire E. 1996. The enthalpy change in protein folding and binding: Refinement of parameters for structure-based calculations. *Proteins* 26:123–133.
- Ibarra-Molero B, Loladze VV, Makhatadze GI, Sanchez-Ruiz JM. 1999a. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* 25:8138–8149.
- Ibarra-Molero B, Makhatadze GI, Sanchez-Ruiz JM. 1999b. Cold denaturation of ubiquitin. *Biochim Biophys Acta* 1429:384–390.
- Ibarra-Molero B, Sanchez-Ruiz JM. 1996. A model-independent, nonlinear extrapolation procedure for the characterization of protein folding energetics from solvent denaturation data. *Biochemistry* 47:14689–14702.
- Liu Y, Bolen DW. 1995. The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes. *Biochemistry* 34:12884–12891.
- Makhatadze GI, Lopez MM, Richardson JM, Thomas ST. 1998. Anion binding to the ubiquitin molecule. *Protein Sci* 7:689–697.
- Makhatadze G, Privalov PL. 1995. Energetics of protein structure. *Adv Prot Chem* 47:307–425.
- Plaza del Pino IM, Sanchez-Ruiz JM. 1995. An osmolyte effect on the heat capacity for protein folding. *Biochemistry* 34:8621–8630.
- Qu Y, Bolen CL, Bolen DW. 1998. Osmolyte-driven contraction of a random coil protein. *Proc Natl Acad Sci USA* 95:9268–9273.
- Santoro MM, Liu Y, Khan SM, Hou LX, Bolen DW. 1992. Increased thermal stability of proteins in the presence of naturally occurring osmolytes. *Biochemistry* 31:5278–5283.
- Taylor JR. 1982. *An introduction to error analysis. The study of uncertainties in physical measurements*. Mill Valley, CA: University Science Books.
- Timasheff SN. 1992. A physicochemical basis for the selection of osmolytes by nature. In: Osmond CB, Bolis CL, Somero GN, eds. *Water and life*. Berlin: Springer Verlag, pp 70–84.
- Wang A, Bolen DW. 1997. A naturally occurring protective system in urea-rich cells: Mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry* 36:9101–9108.
- Xie D, Freire E. 1992. Structure based prediction of protein folding intermediates. *J Mol Biol* 242:62–80.
- Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN. 1982. Living with water stress: Evolution of osmolyte systems. *Science* 217:1214–1222.

Appendix: On the errors associated with the values obtained for the effect of osmolyte concentration on denaturation temperature and unfolding Gibbs energy: We will show in this Appendix that the relevant experimental quantities reported in this work (the slopes dT_m/dC and $\partial\Delta G/\partial C$) are determined with sufficient low error to justify our conclusions.

First, it must be noted that denaturation temperatures are obtained directly from the DSC transitions (or by simple interpolation in the T_m vs. pH plot of Fig. 2). Denaturation temperatures are determined accurately from DSC transitions (to better than 1 K usually) and, therefore, the fact that the slope dT_m/dC is much larger for ubiquitin ($dT_m/dC = 7.1$ K/M; i.e., an increase of about 14 K in T_m when going from 0 M to 2 M sarcosine) than, for instance, for lysozyme ($dT_m/dC = 2.5$ K/M; i.e., an increase of ~ 5 K in T_m when going from 0 M to 2 M) is beyond any reasonable doubt.

On the other hand, it may not be clear that the errors associated to the values of the slope $\partial\Delta G/\partial C$ are indeed low. Thus, the errors of the unfolding heat capacity changes reported in this paper are comparatively large (close to ± 1 kJ K $^{-1}$ mol $^{-1}$ in some cases), which implies that the errors associated to the ΔG values at 25 °C calculated by using the Gibbs–Helmholtz equation (Equation 3) are of the order of several kJ/mol, that is, comparable to the total sarcosine effect on ΔG (Fig. 3B). It may seem then that the slopes $\partial\Delta G/\partial C$ are determined in this paper with an intolerably large error. *This is not so, however, because the effect of the error in ΔC_p on the calculated ΔG is independent of sarcosine concentration and, therefore, does not affect the slope $\partial\Delta G/\partial C$.* To make this idea clear, we will start with the Gibbs–Helmholtz expression for ΔG at any temperature,

$$\Delta G(T) = \Delta H_m \left(1 - \frac{T}{T_m} \right) + \Delta C_p \left[T - T_m - T \ln \left(\frac{T}{T_m} \right) \right] \quad (A1)$$

and differentiate to obtain the expression for the slope $\partial\Delta G/\partial C$ at the temperature T in terms of the osmolyte concentration dependencies of the input parameters of the calculation (ΔH_m , T_m , and ΔC_p):

$$\begin{aligned} \left(\frac{\partial\Delta G}{\partial C} \right)_T &= \frac{d\Delta H_m}{dC} \left(1 - \frac{T}{T_m} \right) + \Delta H_m \frac{T}{T_m^2} \frac{dT_m}{dC} \\ &+ \frac{\partial\Delta C_p}{\partial C} \left[T - T_m - T \ln \left(\frac{T}{T_m} \right) \right] \\ &+ \Delta C_p \frac{dT_m}{dC} \left(\frac{T}{T_m} - 1 \right). \end{aligned} \quad (A2)$$

It must be noted that the total derivative of ΔH_m in Equation A2 ($d\Delta H_m/dC$) takes into account the “direct” effect of osmolyte concentration on ΔH , as well as the “indirect” effect caused by the fact that ΔH_m is the ΔH value at the denaturation temperature and T_m changes with osmolyte concentration. Thus, using the known rules of partial differentiation, this derivative can be written as (see Plaza del Pino & Sanchez-Ruiz, 1995)

$$\frac{d\Delta H_m}{dC} = \frac{\partial \Delta H}{\partial C} + \frac{\partial \Delta H}{\partial T} \left(\frac{dT_m}{dC} \right) = \frac{\partial \Delta H}{\partial C} + \Delta C_p \frac{dT_m}{dC} \quad (\text{A3})$$

where $\partial \Delta H / \partial C$ is the derivative of ΔH with respect to osmolyte concentration evaluated at a constant temperature equal to the T_m value. Upon substitution of Equation A3 into Equation A2, terms with ΔC_p cancel out and we obtain:

$$\left(\frac{\partial \Delta G}{\partial C} \right)_T = \frac{\Delta H_m}{T_m} \frac{T}{T_m} \frac{dT_m}{dC} + \left(1 - \frac{T}{T_m} \right) \frac{\partial \Delta H}{\partial C} + \left[T - T_m - T \ln \left(\frac{T}{T_m} \right) \right] \frac{\partial \Delta C_p}{\partial C}. \quad (\text{A4})$$

We will assume for the moment that ΔH and ΔC_p do not change with osmolyte concentration (the errors introduced by this assumption will be discussed further below). That is, we assume that $\partial \Delta H / \partial C = 0$ and $\partial \Delta C_p / \partial C = 0$ and Equation A4 reduces to

$$\left(\frac{\partial \Delta G}{\partial C} \right)_T = \frac{\Delta H_m}{T_m} \frac{T}{T_m} \frac{dT_m}{dC} \quad (\text{A5})$$

which shows that the value calculated for the slope $\partial \Delta G / \partial C$ at a given temperature (T) depends on the values for the denaturation temperature, the unfolding enthalpy at T_m , and the osmolyte concentration effect on T_m ($\partial T_m / \partial C$) used in the calculation; it does not depend, however, on the value of ΔC_p , which does not appear in Equation A5. The denaturation temperatures at pH 3.5 and 1 M sarcosine are 363.6 ± 0.4 K (ubiquitin), 329.4 ± 0.1 K (ribonuclease A), and 346.4 ± 0.2 K (lysozyme), where the values and their associated standard errors have been calculated from the linear fits to the T_m vs. pH dependencies (Fig. 2). The values and standard errors of ΔH_m at the above denaturation temperature values are 327.4 ± 5.1 kJ/mol (ubiquitin), 398.5 ± 3.1 kJ/mol (ribonuclease A), and 479.1 ± 1.8 kJ/mol (lysozyme), which are derived from the linear fits to the ΔH_m vs. T_m dependencies (Fig. 1). From the linear fits to the T_m vs. C dependencies (Fig. 3A), we obtained that the slopes dT_m/dC are 7.1 ± 0.4 K/M (ubiquitin), 3.4 ± 0.1 K/M (ribonuclease A), and 2.5 ± 0.1 K/M (lysozyme). Finally, upon substitution of the above values in Equation A5 and application of the known procedures of standard error propagation, we obtain the following values and associated errors for the slope $\partial \Delta G / \partial C$ at $T = 298$ K: 5.17 ± 0.32 kJ mol⁻¹ M⁻¹ (ubiquitin), 3.68 ± 0.03 kJ mol⁻¹ M⁻¹ (ribonuclease A), and 2.94 ± 0.12 kJ mol⁻¹ M⁻¹ (lysozyme). Two points must be noted about these values: (1) They agree with those given in the text and are calculated using the conventional procedure (Gibbs–Helmholtz extrapolation to obtain ΔG at 25 °C, followed by a linear fit to the ΔG vs. C dependence); this was to be expected since both calculations are essentially equivalent. (2) The associated standard errors are rather low; again, this was to be expected because the calculation does not use ΔC_p values and the errors associated with T_m , ΔH_m , and dT_m/dC are comparatively small.

The above calculations are to some extent equivalent to a sensitivity analysis on the effect of the value of the unfolding heat capacity change, since what we have shown is that the value calculated for the slope $\partial \Delta G / \partial C$ is not sensitive to the value used for ΔC_p , and, as a result, the errors associated to ΔC_p do not propagate to the slope $\partial \Delta G / \partial C$. Consequently, the errors associated with the calculated values of $\partial \Delta G / \partial C$ are very low. This result is confirmed by a more conventional Monte Carlo analysis of errors. Thus, for each of the proteins studied in this paper, we have generated 1,000 sets of three ΔG values at 25 °C (corresponding to the three sarcosine concentrations employed: 0 M, 1 M, and 2 M) using the following procedure: (1) Denaturation temperature values at pH = 3.5 and the three sarcosine concentrations used were randomly generated according to a Gaussian distribution with values of the mean and standard deviations equal to those previously determined from the analysis of the pH dependence of the T_m values. (2) For each of the protein studied, unfolding enthalpy values at a

given reference temperature (T_{ref} , chosen as a value in the middle of the experimental T_m range) were randomly generated according to a Gaussian distribution with the mean and standard deviation values for ΔH at the reference temperature (ΔH_{ref}) derived from the linear least-squares fittings to the ΔH vs. T_m dependencies (Fig. 1). (3) Unfolding heat capacity values were likewise randomly generated according to a Gaussian distribution with the mean and standard deviation values for ΔC_p derived from the least-squares fittings to the ΔH vs. T_m dependencies. (4) For each protein, unfolding enthalpy values at the three denaturation temperatures generated in step 1 were calculated from the ΔH_{ref} and ΔC_p values generated in steps 2 and 3 by using $\Delta H_m = \Delta H_{ref} + \Delta C_p \cdot (T_m - T_{ref})$. (5) ΔG values at 25 °C and different sarcosine concentrations (0 M, 1 M, 2 M) were obtained by using the integrated Gibbs–Helmholtz equation (Equation 3) with the T_m , ΔC_p , and ΔH_m values generated in the previous steps. Once the 1,000 sets of three ΔG values had been obtained for the three proteins, we calculated for each protein 1,000 values of the slope $\partial \Delta G / \partial C$ from the ΔG values at 0 M and 1 M sarcosine [as $(\Delta G(1 \text{ M}) - \Delta G(0 \text{ M})) / 1 \text{ M}]$ and 1,000 values from the ΔG values at 0 M and 2 M [as $(\Delta G(2 \text{ M}) - \Delta G(0 \text{ M})) / 2 \text{ M}]$. Finally, the mean and standard deviation values of the slope were obtained; the results (mean value $\pm$ standard deviation in kJ mol⁻¹ M⁻¹) are: (1) from the ΔG values at 0 M and 1 M sarcosine: 5.50 ± 0.60 (ubiquitin), 2.70 ± 0.55 (lysozyme), 3.62 ± 0.15 (ribonuclease A) and (2) from the ΔG values at 0 M and 2 M sarcosine: 5.14 ± 0.60 (ubiquitin), 2.95 ± 0.26 (lysozyme), 3.68 ± 0.03 (ribonuclease A). These $\partial \Delta G / \partial C$ values are in excellent agreement with those previously calculated in the main text and in this Appendix on the basis of Equation A5. In addition, the associated errors are indeed very low, being of the same order than those previously estimated from the error analysis based on Equation A5.

It is important to note at this stage that the fact that the errors associated with the values of the slope $\partial \Delta G / \partial C$ are very low is entirely consistent with the usual procedures of error propagation. Thus, assume for instance that we wish to calculate the value of $\partial \Delta G / \partial C$ for ubiquitin from the ΔG values at 0 M and 1 M; in this case, the slope would be numerically equal to the difference between the two ΔG values, that is, to the $\Delta \Delta G$ value for $C = 1$ M. Now, the standard error (σ_q) associated to a function of two variables, $q(x, y)$, is related to standard errors of the variables (σ_x and σ_y) through (see Eq. 9.9 on p. 177 in Taylor, 1982)

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y} \right)^2 \sigma_y^2 + 2 \frac{\partial q}{\partial x} \frac{\partial q}{\partial y} \sigma_{xy} \quad (\text{A6})$$

where σ_{xy} is the covariance of x and y [defined as $(1/N) \cdot \sum (x_i - \langle x \rangle) \cdot (y_i - \langle y \rangle)$], where $\langle x \rangle$ and $\langle y \rangle$ are the mean values of x and y . Here, q would be $\Delta \Delta G$, and the two variables (x and y) would be equal to $\Delta G(C = 0 \text{ M})$ and $\Delta G(C = 1 \text{ M})$; therefore, the function $q(x, y)$ would simply be $q = y - x$ and Equation A.6 reduces to

$$\sigma_q^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}. \quad (\text{A7})$$

From our Monte Carlo analysis, the values and standard errors of $\Delta G(C = 0 \text{ M})$ and $\Delta G(C = 1 \text{ M})$ are 30.78 ± 5.51 and 36.29 ± 5.57 kJ/mol, respectively; these σ_y and σ_x values are comparatively large, reflecting mainly the effect of the error associated to the ΔC_p value. However, the covariance is also a large number ($\sigma_{xy} = 30.61$ kJ²/mol² from our Monte Carlo values), which simply indicates that the errors associated to $\Delta G(C = 0 \text{ M})$ and $\Delta G(C = 1 \text{ M})$ are strongly correlated, as was to be expected from the fact that the error in ΔC_p affects in the same manner both ΔG values. The error in $\Delta \Delta G$ calculated using Equation A7 and the above values of σ_x , σ_y , and σ_{xy} is only 0.41 kJ/mol, a low value that is in agreement with the value we obtained previously from the analysis of the 1,000 $\Delta \Delta G$ values derived from the 1,000 Monte Carlo ΔG sets. The interpretation of this result is straightforward: since the errors in $\Delta G(C = 0 \text{ M})$ and $\Delta G(C = 1 \text{ M})$ are correlated, they cancel each other out to a significant extent in the difference $\Delta \Delta G = \Delta G(C = 1 \text{ M}) - \Delta G(C = 0 \text{ M})$ (for a discussion on covariance and error correlation, see Chapter 9 in Taylor, 1982). In general, error correlation explains why we have errors for ΔG of the order of several kJ/mol and, at the same time, very low errors

associated with the slopes $\partial\Delta G/\partial C$ and the values of $\Delta\Delta G$ (as is shown in Fig. 3).

By now, it should be clear beyond reasonable doubt that the errors associated to the values we report for the slopes $\partial\Delta G/\partial C$ are indeed very low, *at least under the assumption that ΔH and ΔC_p do not change significantly with sarcosine concentration within the range 0–2 M*. This assumption is clearly supported by (1) our previous exhaustive study on ribonuclease A unfolding in sarcosine solutions in which osmolyte effects on ΔH and ΔC_p could only be detected at sarcosine concentrations higher than those employed in this work; (2) the plots of ΔH_m vs. T_m reported in this paper for lysozyme and ribonuclease A (Fig. 1), which show an excellent agreement between unfolding parameters determined at different sarcosine concentrations within the 0–2 M range. On the other hand, we must recognize that, in the case of ubiquitin, the larger scatter of the experimental ΔH_m values (Fig. 1) casts some doubts on the correctness of the assumption. Therefore, we proceed now to estimate upper bounds to the error introduced in the ubiquitin $\partial\Delta G/\partial C$ value by assuming that ΔH and ΔC_p do not change with C on the basis of Equation A4 (a Monte Carlo analysis will be described further below). This requires that we provide estimates for the possible values of the second and third terms in the right side of Equation A4. We will begin with the term corresponding to the effect of sarcosine on ΔC_p ; i.e., the term containing the derivative $\partial\Delta C_p/\partial C$. We have calculated the value for this derivative from a linear fit to the experimental ΔC_p vs. C data (using the ΔC_p values determined by extrapolating to the T_m values the native and unfolded baselines of the DSC profiles for ubiquitin); the result is $-0.06 \pm 0.28 \text{ kJ mol}^{-1} \text{ K}^{-1} \text{ M}^{-1}$. That is, for ubiquitin the derivative $\partial\Delta C_p/\partial C$ is zero within an uncertainty of $\pm 0.28 \text{ kJ mol}^{-1} \text{ K}^{-1} \text{ M}^{-1}$. Using this standard error and the T_m value at pH 3.5 and 1 M sarcosine previously calculated, we find that the $\partial\Delta C_p/\partial C$ term in Equation A4 (third term in the right side) is zero within an uncertainty of $\pm 1.8 \text{ kJ mol}^{-1} \text{ M}^{-1}$.

To evaluate the second term in the right side of Equation A4, we need an estimate of the derivative $\partial\Delta H/\partial C$. Values of ΔH at a common temper-

ature (chosen as the T_m value for pH 3.5 and 1 M sarcosine) were obtained from the experimental ΔH_m , ΔC_p , and T_m values using the Kirchoff equation. Then, linear fits to the resulting ΔH vs. C data were performed and the slope $\partial\Delta H/\partial C$ for ubiquitin turned out to be $-0.5 \pm 6.2 \text{ kJ mol}^{-1} \text{ M}^{-1}$ that is, zero within the standard uncertainty. As a result, the $\partial\Delta H/\partial C$ term in Equation A4 (second term in the right side) is calculated to be zero within $\pm 1.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$ for ubiquitin.

We have obtained estimates of the standard errors associated with the three terms in the right side of Equation A4, and the total standard error of the slope $\partial\Delta G/\partial C$ for ubiquitin can now be calculated using the known rules of standard error propagation; the result is $\pm 2.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$. It should be noted that this uncertainty is likely to be an upper-limit estimate (because our calculated errors for the second and third terms in the right side of Equation A4 are upper-limit estimates), and the actual error could be, in fact, much smaller. Nevertheless, even this upper-limit error would not alter our main conclusion, namely, that the values predicted for $\partial\Delta G/\partial C$ on the basis of additivity agree with the experimental ones only if the unfolded state is assumed to be compact (Rose's lower limit). This is clear from Figure 4, where the error bar shown for ubiquitin corresponds to this $\pm 2.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$ uncertainty interval.

The above result may be again confirmed by a Monte Carlo analysis. Thus, we generated 1,000 sets of ΔG values, essentially as described above, but linear dependencies of ΔH_{ref} and ΔC_p with C were allowed and the corresponding slopes were generated randomly according to Gaussian distributions of zero mean and standard deviation values equal to those given above for $\partial\Delta H/\partial C$ and $\partial\Delta C_p/\partial C$. The 1,000 sets of ΔG values were then used to calculate the 1,000 values of the slope $\partial\Delta G/\partial C$ from the ΔG values at $C = 0 \text{ M}$ and $C = 1 \text{ M}$, and 1,000 values for the same slope from the ΔG values at $C = 0 \text{ M}$ and $C = 2 \text{ M}$; the corresponding mean and standard deviation values are $5.49 \pm 1.46 \text{ kJ mol}^{-1} \text{ M}^{-1}$ (from $C = 0 \text{ M}$ and $C = 1 \text{ M}$) and $5.15 \pm 1.87 \text{ kJ mol}^{-1} \text{ M}^{-1}$. Note that the standard errors obtained are in good agreement with the value ($2.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$) we derived on the basis of Equation A4.