Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model

Edward P. O'Brien*†, Guy Ziv‡, Gilad Haran‡, Bernard R. Brooks†, and D. Thirumalai*§

*Biophysics Program, Institute for Physical Science and Technology and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742; †Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892; and ‡Chemical Physics Department, Weizmann Institute of Science, Rehovot 76100, Israel

Edited by José N. Onuchic, University of California, San Diego, La Jolla, CA, and approved June 3, 2008 (received for review March 2, 2008)

Interactions between denaturants and proteins are commonly used to probe the structures of the denatured state ensemble and their stabilities. Osmolytes, a class of small intracellular organic molecules found in all taxa, also profoundly affect the equilibrium properties of proteins. We introduce the molecular transfer model, which combines simulations in the absence of denaturants or osmolytes, and Tanford's transfer model to predict the dependence of equilibrium properties of proteins at finite concentration of osmolytes. The calculated changes in the thermodynamic quantities (probability of being in the native basin of attraction, *m* values, FRET efficiency, and structures of the denatured state ensemble) with GdmCl concentration [C] for the protein L and cold shock protein CspTm compare well with experiments. The radii of gyration of the subpopulation of unfolded molecules for both proteins decrease (i.e., they undergo a collapse transition) as [C] decreases. Although global folding is cooperative, residual secondary structures persist at high denaturant concentrations. The temperature dependence of the specific heat shows that the folding temperature (T_F) changes linearly as urea and trimethylamine *N*-oxide (TMAO) concentrations increase. The increase in T_F in TMAO can be as large as 20°C, whereas urea decreases T_F by as much as 35°C. The stabilities of protein L and CspTm also increase linearly with the concentration of osmolytes (proline, sorbitol, sucrose, TMAO, and sarcosine).

denatured state ensemble | FRET and SAXS experiments | protein L | cold shock protein | protein collapse

To function proteins fold (1), whereas misfolding is linked to a number of conformational diseases (2, 3), thus making it important to determine the factors that control stability of proteins (1) and their assembly mechanisms (4–6). A molecular understanding of protein folding requires quantitative estimates of the energetic changes (7, 8) in the folding reaction and characterization of the populated structures along the folding pathways. A large number of studies have dissected the interactions that contribute to the stability of proteins (1, 7–15).

In contrast, only relatively recently has there been a concerted effort to determine the structures of the denatured state ensemble (DSE) (16) whose experimental resolution is difficult because of fluctuations in the unfolded structures. In particular, it is difficult to determine the properties of the DSE under conditions in which the native state is stable because the population of the unfolded structures is low (17). Single-molecule FRET experiments have begun to investigate the variations in the global properties of the DSE under native conditions (18–20). Despite these intense efforts, structural characterization of the DSE and its link to global thermodynamic properties and the folding process is lacking.

Denaturants, such as urea and guanidinium chloride (GdmCl), destabilize proteins. In contrast, osmolytes that protect cells against environmental stresses such as high temperature, desiccation, and pressure can stabilize proteins (21). Thus, a complete understanding of the stability of proteins and a description of the structures in the diverse DSEs requires experimental and theoretical studies that provide a quantitative description of the effects of both osmolytes and denaturants.

From a theoretical perspective, significant advances in our understanding of how proteins fold have come from molecular simulations by using coarse-grained (CG) off-lattice models (22–27). However, the CG models only probe the folding of proteins by changing temperature, making it difficult to compare the predictions directly with many experiments that use denaturants. In principle, all-atom simulations of proteins in aqueous denaturant solutions can be used to calculate the conformational properties of proteins. However, the difficulty in adequately sampling the protein conformational space makes most of these simulations inherently nonergodic (28). Here, we overcome these problems by combining Tanford's transfer model (TM) (29, 30) with simulations using an off-lattice side chain representation of polypeptide chains (26) to predict the dependence of the size of the protein, fraction of molecules in the native state, and FRET efficiencies as a function of the concentration ([C]) of denaturants and osmolytes. We present a method that combines molecular simulations of a protein of interest at [C] = 0, and the experimental transfer free energies (31, 32) to predict the thermodynamic averages at [C] ≠ 0. In the process, we have greatly expanded the power and scope of CG off-lattice models (23, 25, 27) in predicting the outcomes of experiments. Applications of the resulting molecular transfer model (MTM) to protein L and cold shock protein (CspTm) (Fig. 1*A*) show that calculated changes in the fraction of folded conformations, and the average FRET efficiency as a function of [GdmCl] are in excellent agreement with experiments (20, 33, 34). The stability in the presence of glycine betaine, proline, sucrose, sarcosine, sorbitol, and trimethylamine *N*-oxide (TMAO) for the two proteins increases linearly as [C] increases. The heat capacity changes in proteins in denaturants and osmolytes are interpreted in terms of changes in the folding landscape. Our results also give plausible explanations for the inability of scattering methods to directly infer protein collapse at low [C].

Results

MTM Accurately Captures Denaturant-Induced Unfolding of Protein L and CspTm. To establish the efficacy of the MTM, we calculate a number of quantities that can be directly compared with data from ensemble and single-molecule experiments (18–20, 33, 34). As with most molecular force fields, the absolute interaction energies in the C_α-SCM at [C] = 0 are not accurate. We set the temperature ($T = T_S$) so that the calculated free energy of stability of the native state $\Delta G_{NU}(T_S)$, with respect to the unfolded structures, and the mea-

Author contributions: E.P.O. and D.T. designed research; E.P.O., G.Z., B.R.B., and D.T. performed research; B.R.B. contributed new reagents/analytic tools; E.P.O., G.Z., G.H., B.R.B., and D.T. analyzed data; and E.P.O., G.H., and D.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

§To whom correspondence should be addressed. E-mail: thirum@umd.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802113105/DCSupplemental.

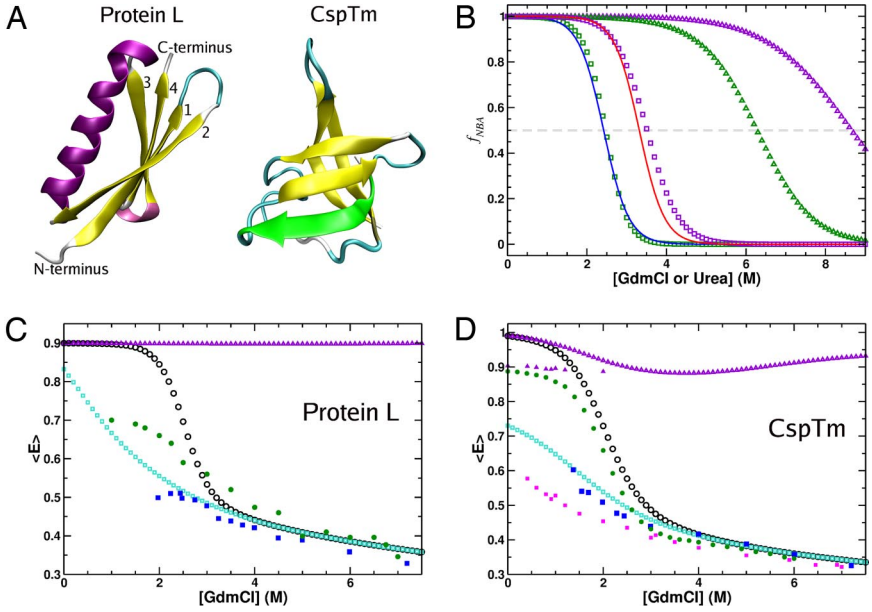

Fig. 1. Native structures and comparison of calculated and experimental results. (*A*) The numbers in protein L label the strands starting from the N terminus. The N-terminal β-strand in CspTm is green. (*B*) The fraction of molecules in the NBA (f_{NBA}) as a function of GdmCl (green squares) and urea (green triangles) for protein L. Results for CspTm in GdmCl and urea are shown in violet squares and violet triangles, respectively. Blue line is the result of f_{NBA} ([C]) for protein L (35). Results in red line is for CspTm (36). Dashed line shows $f_{NBA} = 0.5$. (*C*) The dependence of $\langle E \rangle$ for protein L (open circles) versus GdmCl concentration. Open triangles show $\langle E \rangle$ for the native state and the squares are for the DSE. The experimental values for the average $\langle E \rangle$ and $\langle E \rangle$ of the DSE are shown as green circles (33) and blue squares (34), respectively. (*D*) Results for CspTm by using the same notation as in *C*. The filled blue squares are experimental results from ref. 34. Filled green circles, violet triangles, and magenta squares correspond to experimental measurements of $\langle E \rangle$, the NBA $\langle E \rangle$, and the DSE $\langle E \rangle$, respectively (20). To account for the destabilization of CspTm because of the attachment of dyes, we set $T_S = 341$ K, which gives C_m in agreement with experiment (18). In *C* and *D* we use $R_o = 55$ Å (see Eq. **3** in *SI Text*). Changes in R_o with [C] cause small corrections to $\langle E \rangle$.

sured $\Delta G_{NU}(T_E)$ at $T = T_E$ coincide. In the absence of denaturants, $T_S = 328$ K and $T_E = 295$ K and $\Delta G_{NU}(T_S) = \Delta G_{NU}(T_E) = -4.6$ kcal/mol (35) for protein L (Fig. 1*A*). For CspTm (Fig. 1*A*) $T_S = 326$ K and $T_E = 298$ K and $\Delta G_{NU}(T_S) = \Delta G_{NU}(T_E) = -6.3$ kcal/mol (36). By adjusting T_S appropriately, we find that the dependence of the calculated fraction of molecules in the native basin of attraction (NBA), f_{NBA}, as a function of [C] for GdmCl is in excellent agreement with experiments (Fig. 1*B*). The values for C_m, the midpoint concentration at which $f_{NBA} = 0.5$, for both proteins also reproduce the measured values accurately (Table 1).

Measured and Predicted FRET Efficiencies Are in Good Agreement. In an attempt to characterize the nature of unfolded states of proteins under folding conditions (low denaturant concentrations) several groups have used single-molecule FRET spectroscopy (18, 20, 33, 34). By attaching fluorescent dyes at two points [typically, but not always (20), located at the termini of the protein], the average FRET efficiency $\langle E \rangle$ as a function of [GdmCl] has been measured for protein L and CspTm. We calculated $\langle E \rangle$ as a function of [GdmCl] for protein L (Fig. 1*C*) and CspTm (Fig. 1*D*). The discrepancies between different experiments not withstanding (18, 20, 33, 34), the simulated and the measured $\langle E \rangle$ for protein L and CspTm, for the subpopulation of unfolded states, are in excellent agreement (Fig. 1 *C* and *D*) with each other. The average FRET efficiency, that weights the subpopulations of folded and unfolded states, reflects the cooperativity observed in f_{NBA} (Fig. 1*B*). The values of $\langle E \rangle$ for the structures in the NBA are roughly constant as [GdmCl] changes (Fig. 1 *C* and *D*). Even though the simulated value of $\langle E \rangle (= 0.9)$ for protein L at zero [GdmCl] agrees with the calculated FRET efficiency by using Protein Data Bank (PDB) coordinates (PDB ID code 1HZ6), it is larger than the measured value, which is in the range from 0.7 to 0.8. The discrepancy could also arise because the present simulations do not explicitly include the dyes with flexible linkers which can have a large effect (34). Despite the difference at [C] = 0, our simulations accurately reproduce the experimental measurements.

Table 1. Calculated thermodynamic parameters for protein L and CspTm

	Protein L		CspTm	
Osmolyte	*m* value*	ΔG_{NU} [0] †	*m* value	ΔG_{NU} [0]
GdmCl	2.4‡	−6.0§	1.7¶	−5.8‖
Urea	0.9	−5.7	0.7	−6.1
Betaine	0.2	−4.8	0.2	−6.3
Proline	−0.1	−4.7	0.1	−6.3
Sorbitol	−0.1	−4.7	−0.3	−6.3
Sucrose	−0.2	−4.7	−0.4	−6.3
TMAO	−0.2	−4.7	−0.5	−6.3
Sarcosine	−0.2	−4.7	−0.5	−6.3

*Units in kcal M^{-1} mol^{-1}.
†Native state stability, in kcal mol^{-1} units, at 0 M, using the linear extrapolation method (50).
‡Experimental value is 1.9 kcal mol^{-1} M^{-1} (35).
§Two-state fit to experimental data gives ΔG_{NU} [0] = −4.6 kcal mol^{-1} to −6.0 kcal mol^{-1} (35, 51).
¶Experimental value is 1.9 ± 0.08 kcal mol^{-1} M^{-1} (36).
‖The experimental value is ΔG_{NU} [0] = −6.3 ± 0.3 kcal mol^{-1} (36).

Changes in R_g Depend on the Nature of Cosolvents. The R_g distribution $[P(R_g)]$ for protein L in urea, at the folding (or melting) temperature $T_F = 356$ K, shows the expected bimodal behavior (Fig. 2). At 0 M, there is a sharp peak in $P(R_g)$ at $\overline{R_g^N}$ (the value in the native state) ≈12 Å, whereas at 6 M urea a relatively broad ensemble of conformations, with larger R_g values (>12 Å), is populated (Fig. 2*A*). The distribution $P(R_g)$ at 6 M urea compares favorably with recent all-atom simulations (see figure 10 in ref. 34). In 6 M TMAO the peak height at $R_g \sim 12$ Å increases, which reflects its stabilizing influence. The average $\overline{R_g}$ for protein L expands continuously as urea concentration increases from 0 to 6 M (Fig. 2*B*). Decomposition of the ensemble of structures into the DSE subpopulation shows that $\overline{R_g^{DSE}}$ expands from 21.6 Å at 0 M urea to 24 Å at 6 M urea whereas $\overline{R_g^N}$ is independent of urea concentration (Fig. 2*B*). At physiological concentrations (≈1 M) the change in $\overline{R_g}$ induced by TMAO is small (Fig. 2*B*). Just as for urea, the value of $\overline{R_g^N}$ remains constant at all TMAO concentrations (Fig. 2*B*).

There are substantial changes in the size of protein L and CspTm in aqueous GdmCl solution (Fig. 2*C*). (*i*) For both proteins, the precipitous change in $\overline{R_g}$ occurs at [C] ~ C_m which suggests that global unfolding is accompanied by expansion of the proteins (compare Figs. 1 *C* and *D* and 2*C*). Unfolding in GdmCl is considerably more cooperative than in urea (data not shown). (*ii*) In contrast to protein L, whose $\overline{R_g^N}$ is nearly independent of the concentration of GdmCl (Fig. 2*C*), $\overline{R_g^N}$ for CspTm increases marginally when [C] exceeds ≈2.5 M. Moder-

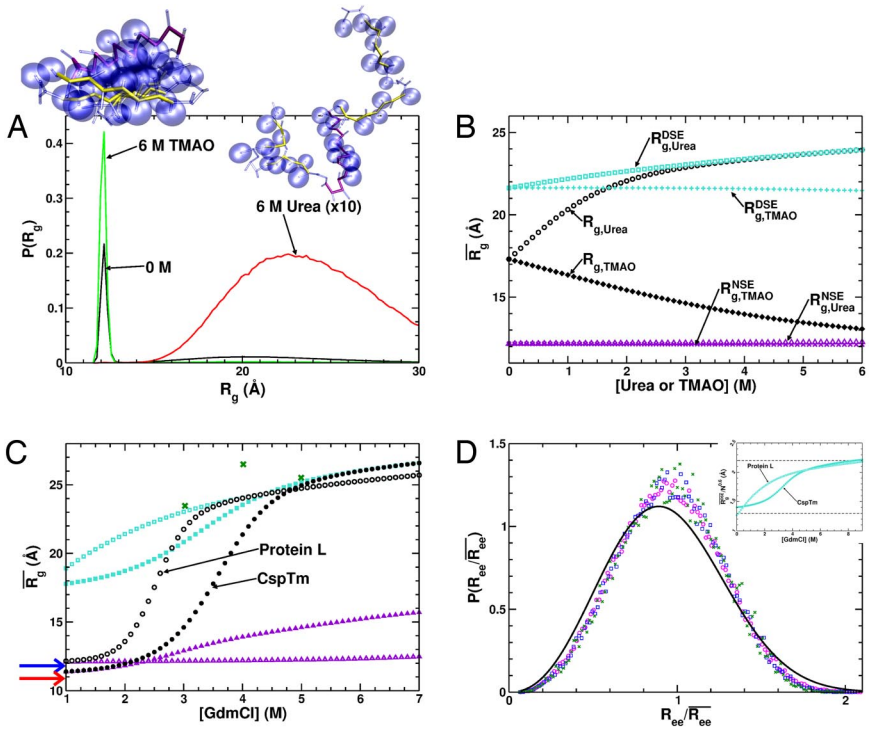

Fig. 2. $P(R_g)$ and $\overline{R_g}$. (*A*) The distribution $P(R_g)$ for protein L at 0 M and 6 M TMAO and urea at $T_F = 356$K, the melting temperature at 0 M. For $P(R_g)$ at 0 M, the areas under native and denatured ensembles are equal. The $P(R_g)$ at 6 M urea is multiplied by 10. The structure on the left corresponds to the C_α-SCM representation of the native state and the one to the right is an example of a conformation in the DSE. For clarity, only hydrophobic side chains are displayed as blue spheres. (*B*) The average $\overline{R_g}$ of protein L as a function of urea (open black circles) and TMAO (black diamonds) concentration at T_F. The values of $\overline{R_g}$ of the NBA is in violet squares for urea and plus signs for TMAO. The results for $\overline{R_g^{DSE}}$ in urea and TMAO are in turquoise triangles and x symbols, respectively. By using Flory theory $\overline{R_g}$ at [C] = 0 is $0.5(\overline{R_g^N} + \overline{R_g^D}) \approx$ 17.7 Å, which agrees with the simulations. (*C*) The $\overline{R_g}$ for protein L (open black circles) and CspTm (filled black circles) as a function of GdmCl concentration at a temperature of 328 K (protein L) and 326 K (CspTm). $\overline{R_g}$ for the NBA is in triangles and DSE is in squares for protein L (open symbols) and CspTm (filled symbols). Blue and red arrows show $\overline{R_g}$ computed from crystal structures for protein L and CspTm, respectively, The green X's are $\overline{R_g}$ from SAXS experiments (42) for protein L with His tag ($N = 79$). (*D*) The DSE distribution $P(R_{ee}^{DSE}/\overline{R_{ee}^{DSE}})$ for protein L in 5, 7, and 9 M GdmCl at 328 K. The solid black line is the theoretical universal curve for a self-avoiding polymer chain. (*Inset*) The effective Kuhn length $a_D([C], T) = \overline{R_g^{DSE}}/N^{0.6}$ versus GdmCl concentration. The dashed lines show the range of experimentally measured Kuhn lengths (40).

ate denaturant-induced increase in $\overline{R_g^N}$ at high concentrations of GdmCl indicates that packing is somewhat compromised in CspTm, arising from enhanced fluctuations in the N-terminal β-strand (Fig. 1*A* and see below). (*iii*) The values of $\overline{R_g^{DSE}}$ for both proteins increase nearly continuously as [C] increases. In CspTm, there may be an inflection point at [C] $\sim$ 2.5 M, which coincides with the onset of a modest increase in $\overline{R_g^N}$. (*iv*) At high [C], $\overline{R_g^{DSE}} \approx$25.5 Å for protein L and $\overline{R_g^{DSE}} \approx$26.5 Å for CspTm (Fig. 2*C*). These values are in near quantitative agreement with the analysis of FRET efficiency by using a highly simplified Gaussian model for the end-to-end distribution function (20, 33, 34).

Dissecting Denaturant-Induced Loss of Secondary and Tertiary Structures. The native structure of protein L has a β-sheet composed of two β-hairpins formed by strands 1 and 2, and strands 3 and 4 that interact with a central helix (Fig. 1*A*). The loss in the β-strand contacts in GdmCl and urea mirror the overall unfolding of the protein (compare Fig. 3*A* and Fig. 1*B*). Chain expansion and the loss of secondary and tertiary contacts occur at nearly similar concentrations (see Figs. 1*B*, 2*C*, and 3*A*). For protein L, at high denaturant concentrations there is near complete loss of β-strand content, whereas residual helical content persists (Fig. 3*A*).

Comparison of the plots (Fig. 3*B*) of the tertiary contacts involving the secondary structural elements (SSEs) and the total number of contacts in protein L as a function of urea concentration shows that most of the curves overlap. These results (Fig. 3*B*) show that the loss of secondary and tertiary interactions occurs cooperatively. The fluctuations of the various SSEs $\sigma^2_{Q_i} = \langle Q_i^2 \rangle - \langle Q_i \rangle^2$, as a function of urea concentration (Fig. 3*C*) show that the strands 1 and 4, that join the two β-hairpins together to form the full β-sheet, have the most cooperative transition (Fig. 3*C*). These strands, which are far apart in sequence space, form the longest-range contacts in the NBA. Similarly, contacts involving the two hairpins S12 and S34 also unfold cooperatively. Thus, SSEs that form long-range contacts in the NBA unfold most cooperatively.

Heat Capacity of Proteins Are Greatly Altered by Osmolytes. The temperature dependence of the heat capacity (C_V) for protein L and CspTm shows that, as urea concentration increases from 0 M to 8 M, the curves shift to the left (Figs. 4 *A* and *B*). In contrast, in the presence of the osmolyte TMAO the curves move to the right (Fig. 4 *A* and *B*). For proteins that fold in an apparent two-state manner the peak in C_V can be identified with the folding temperature, T_F. The decrease in the folding temperature $\Delta T_F([C]) \equiv T_F([C]) - T_F(0)$ as the concentration of urea increases from 0 to 8 M can be as large as 35°C. As the concentration of TMAO increases from 0 to 8 M, $\Delta T_F([C])$ increases by as much as 12°C for protein L and $\approx$20°C for CspTm. These results (Fig. 4 *A* and *B*) indicate that there are large variations in thermal stability of CspTm and protein L as the concentrations of urea and TMAO are increased.

In contrast to the behavior of C_V for protein L (Fig. 4*A*), the peak heights and the widths change significantly for CspTm in urea and TMAO (Fig. 4*B*). For CspTm the maximum in C_V goes from 6.5 kcal °C^{-1} M^{-1} at 0 M to $\approx$9.0 kcal °C^{-1} M^{-1} in 8 M TMAO and $\approx$5.0 kcal °C^{-1} M^{-1} in 8 M urea. The maximum in C_V for protein L, however, changes by only $\approx$0.2 kcal °C^{-1} M^{-1} under these same solution conditions (Fig. 4*A*).

Protein Stability Changes Linearly as Denaturant and Osmolyte Concentrations Increase. ***Denaturants.*** Although the changes in native state stability $\Delta G_{NU}([C])$ as a function of [C] for protein L (Fig. 4*C*) and CspTm (Fig. 4*D*) at $T \approx$ 328 K show evidence for nonlinearity in some of the curves, the free-energy change can be approximately fit by using $\Delta G_{NU}([C]) = \Delta G_{NU}(0) - m[C]$ (37, 38). The m values show that GdmCl is significantly more efficient in denaturing protein L and CspTm than urea (Table 1). As a result, the denaturation midpoint C_m for protein L, obtained by using $\Delta G_{NU}([C_m]) = 0$, is 2.4 M in aqueous GdmCl and 6.3 M in aqueous urea.

The calculated (2.4 kcal mol^{-1} M^{-1} for protein L and 1.7 kcal mol^{-1} M^{-1} for CspTm) and measured (1.9 kcal mol^{-1} M^{-1}) GdmCl m values for protein L and CspTm (Table 1) are in excellent agreement. The predicted m value for betaine is relatively small ($m \simeq$ 0.2 kcal mol^{-1} M^{-1}), which implies that betaine only marginally affects the stability of CspTm and protein L (Table 1 and Fig. 4 *C* and *D*). Therefore, the efficiency of denaturation follows

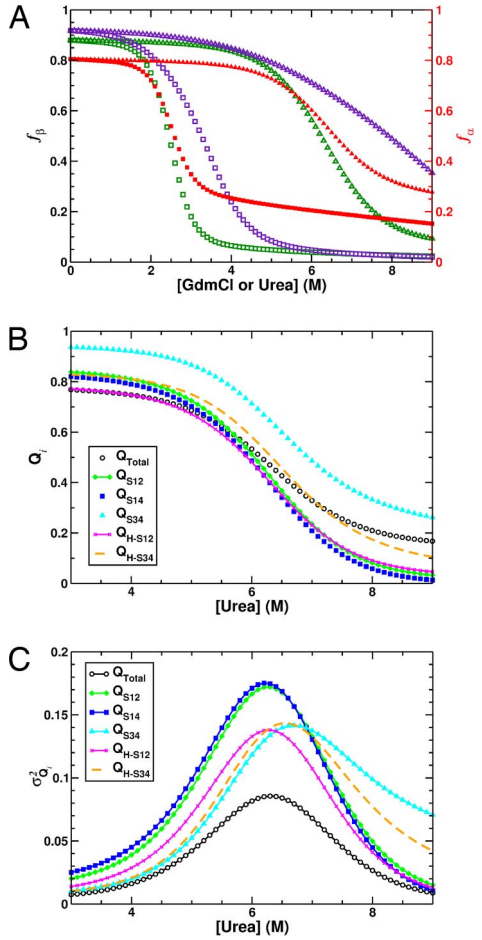

Fig. 3. Changes in the secondary structural elements of protein L as a function of urea and GdmCl concentration at 328 K. (*A*) The dependence of β-sheet (green and violet symbols) and helix (red symbols) content of protein L and CspTm on the concentration of GdmCl and urea by using the same notation as Fig. 1*B*. (*B*) The dependence of the fraction of native contacts in urea for protein L. The fraction of native contacts for the entire protein is denoted Q_T, between strands 1 and 2 as Q_{S12}, between strands 1 and 4 as Q_{S14}, between strands 3 and 4 as Q_{S34}, between strands 1, 2, and the helix as Q_{H-S12}, and between strands 3, 4, and the helix as Q_{H-S34}. (*C*) Variance in the fraction of native contacts versus urea concentration.

the trend GdmCl > urea > betaine. The predictions for aqueous urea and betaine await future experiments.

Osmolytes. The stability changes for osmolytes (proline, sorbitol, sucrose, TMAO, and sarcosine) for protein L (Fig. 4*C*) and CspTm (Fig. 4*D*) at $T \simeq 328$ K vary linearly over a broad range of concentrations. The extracted m values for all these osmolytes vary only moderately for protein L ($m = -(0.1$ to $0.2)$ kcal mol^{-1} M^{-1}) and for CspTm ($m = -(0.5$ to $0.3)$ kcal mol^{-1} M^{-1}; see Table 1). The nearly constant m values for the osmolytes is consistent with experiments that have found that m values for TMAO and sarcosine are roughly the same for barstar (39). As a result of the small m values the osmolytes increase the stability of the small proteins only modestly (≈1 kcal/mol).

Discussion

Flory Theory, Simulations, and Experiments for R_g and the End-to-End Distance Distribution $P(R_{ee})$. The $\overline{R_g}$ values of proteins scales as $\overline{R_g^D} = a_D([C], T)N^\nu$ (where ν, the Flory exponent, is $\nu \simeq 0.59$) (40). The Kuhn length $a_D([C], T)$ reflects the quality of the solvent, which depends on $[C]$, T, and the protein sequence, is found to be a constant $a_D \approx 2$ Å (40). (However, see supporting information (SI) Fig. S1.)

Analysis of the folded structures of proteins shows that $\overline{R_g^N} = a_N N^{1/3}$ with $a_N \approx 3$ Å (41). For protein L ($N = 64$) and CspTm ($N = 66$), we expect that $\overline{R_g^N} \approx 12$ Å and 12.1 Å, respectively. Direct calculation of $\overline{R_g^N}$ by using coordinates from the structures of protein L and CspTm give 12 Å and 11 Å, respectively.

If $a_D([C], T) \sim a_D = 2$ Å is a constant, then Flory theory predicts that $\overline{R_g^D} \sim 23.3$ Å for protein L ($n = 64$), which is in excellent agreement with the simulation results (Fig. 2*B*). Small angle x-ray scattering (SAXS) measurements of protein L with a histidine tag, resulting in $N = 79$ (40), show that $\overline{R_g} = 26 \pm 1.5$ Å and 25 ± 1.5 Å at 4 M and 5 M GdmCl, respectively. From Flory theory we expect $\overline{R_g^D} \approx 27.8$ Å. The agreement between theory, simulations, and SAXS data show that, as far as $\overline{R_g}$ is concerned, protein L behaves as a random coil at high GdmCl concentrations.

In apparent contrast to SAXS measurements (42), our simulations and analysis of FRET data show that protein L (18, 19, 33, 34) and CspTm (20) collapse at low [C]. The differences could arise for the following reasons. (*i*) At $[C] < [C_m]$ almost all of the scattering intensity arises from the folded state, just as at $[C] > [C_m]$ the scattering is dominated by the conformations in the DSE. Thus, it is unlikely that SAXS measurements can resolve the small contributions of $\overline{R_g^{DSE}}$ at low values of [C]. (*ii*) At a fixed T, the "nonuniversal" Kuhn length $a_D([C], T)$ should be [C]-dependent. The Kuhn length $a_D([C], T) \to a_D$ only when $x = [C]/C_m \gg 1$ so that interresidue attractive interactions are negligible, and hence the conformational characteristics of proteins are determined solely by excluded volume interactions. To ascertain the variations of the Kuhn length as [C] changes we computed $a_D([C], T) = \overline{R_g^{DSE}}/N^\nu$, which increase from ≈1.3 Å to ≈2.2 Å (see Fig. 2*D Inset*). Recent, SAXS experiments (see figure 3b in ref. 43) also show that $\overline{R_g}$ for the 159-residue *Escherichia coli* dihydrofolate reductase continues to increase as urea concentration increases in the range 4.5–8 M which can be rationalized in terms of a [C]-dependent Kuhn length. (*iii*) There are large changes in the distribution $P(R_g^{DSE})$ as [C] changes (Fig. S2). If proteins are random coils at high [C] then $P(R_{ee}^{DSE})$, for sufficiently large $y = R_{ee}^{DSE}/\overline{R_{ee}^{DSE}}$ should be given by the universal curve $P(y) = c_1 y^{2+\theta} \exp(-c_2 y^{1/(1-\nu)})$ (44), where $\theta = (\gamma - 1)/2 \approx 1/3$, $c_1 = 3.7$ and $c_2 = 1.2$ (see *SI Text*). The simulation results show that, to an excellent approximation, this is indeed the case for $P(R_{ee}^{DSE}/\overline{R_{ee}^{DSE}})$ (Fig. 2*D*) for $y > 1.5$ and [C] > 5 M GdmCl (see also Fig. S3). Thus only at high [C], when the residual intrapeptide attraction is negligible, the random-coil nature of proteins emerges, whereas at low [C] there are substantial deviations from the self-avoiding $P(y)$ (in Fig. S2).

The incorrect assumption that $a_D([C], T)$ is a constant (or equivalently that $\overline{R_g^{DSE}}$ is [C] independent) when analyzing experimental results (see Fig. S4 for further discussion), and the limited data at [C] beyond the transition region (42) make it difficult to infer protein collapse by using SAXS measurements. In addition, it has been suggested (34) that interprotein interactions could also have affected the SAXs measurements. At the very least, the protein L measurements have to be extended beyond 5 M GdmCl to decipher the changes in $\overline{R_g^{DSE}}$.

Structural Interpretation of the Heat Capacity Curves. The origin of the contrasting behaviors in C_V between protein L and CspTm in urea and TMAO (Fig. 4 *A* and *B*) is reflected in the free-energy surfaces (FESs) at T_F. The two-dimensional FES, expressed in terms of the potential energy (E_P) and the root-mean-square deviation (Δ) from the native state, of protein L has two distinct basins at all osmolyte concentrations (data not shown). On the other hand, CspTm displays three distinct basins at 0 M (Fig. 4*E*). The basin centered at $\Delta \sim 3$ Å corresponds to conformations that closely resemble the crystal structure. The basin, at $\Delta \sim 9$ Å, corresponds to conformations in which the N-terminal strand (Fig. 1*A*) is disordered but the rest of the barrel is intact. The basin centered at $\Delta \sim 22$ Å consists of mostly random coil conformations that have little β-sheet content. At 8 M TMAO the basin of

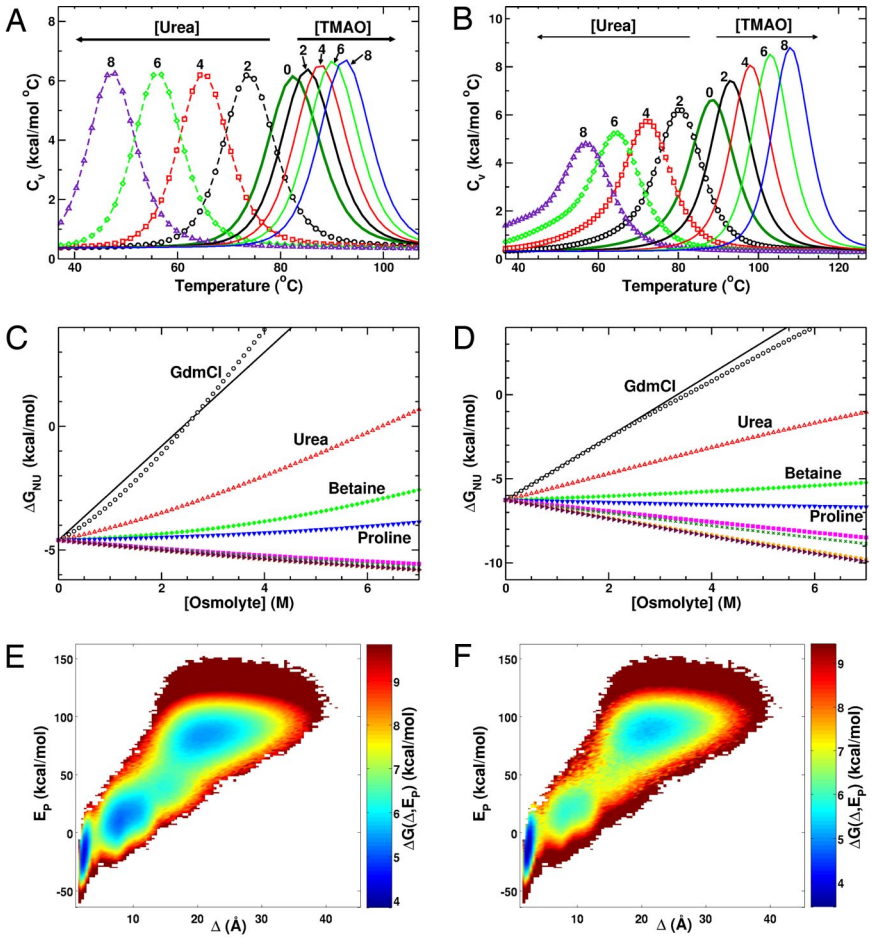

Fig. 4. Thermodynamic properties of protein L and CspTm in denaturant and osmolyte solutions. (*A*) Heat capacity of protein L versus temperature as a function of urea and TMAO concentration. Numbers above the maxima of each trace give the osmolyte concentration in molar units. Curves to the left of the 0 M trace correspond to increasing urea concentrations, whereas those to the right represent increasing TMAO concentrations. (*B*) Results for CspTm by using the same notation as in *A*. (*C*) The stability of the native state ensemble of protein L as a function of concentration of various osmolytes at 328 K. The data corresponding to GdmCl, urea, betaine, and proline are labeled. The variation of ΔG_{NU} in aqueous sorbitol, sucrose, sarcosine, and TMAO solutions are similar, and are unlabeled. The solid black line is the experimental result for GdmCl denaturation (35). (*D*) Same as *C* except the results are for CspTm and experimental results are taken from ref. 36. (*E*) The free-energy surface of CspTm as a function of the root-mean-square deviation relative to the crystal structure (Δ) and the potential energy (E_P) at 0 M and T_F (361 K). (*F*) The same as *E*, except at 8 M TMAO and 381 K.

attraction centered at $\Delta \sim 9$ Å at 0 M is significantly destabilized (Fig. 4*F*) resulting in a sharper transition in C_V (Fig. 4*B*). In contrast, urea expands the area of the denatured basin in the FES (data not shown), which in turn leads to a reduction in the height of C_V and an increase in the width of the transition.

Conclusions

By using converged simulations in the absence of denaturants and osmolytes, together with the measured transfer free energies, the MTM accurately predicts the dependence of any thermodynamic property at arbitrary denaturant or osmolyte concentration. The striking agreement between the computed and measured GdmCl-induced changes in the FRET efficiencies for protein L and CspTm attests to the success of the MTM. The structures of the denatured states, as measured by the residual secondary and tertiary structure content, can be greatly perturbed by adjusting the osmolyte concentration. As a consequence, the folding trajectories may change significantly depending on the initial conditions. Predictions for urea-induced changes in the DSE and the profound differences between the heat capacity changes in urea and TMAO between protein L and CspTm are amenable to experimental tests. More generally, the MTM provides a structural interpretation of the cooperative thermal melting of proteins in osmolytes. In addition, we have made a number of testable predictions for the changes in equilibrium properties of these small single-domain proteins in osmolytes. The present theory sets the stage for using the MTM not only in the context of the C_α-side chain model (C_α-SCM), but also in conjunction with all-atom Go models for which exhaustive sampling can be carried out.

Methods

C_α-SCM for Proteins. We use the coarse-grained C_α-SCM (for details, see *SI Text* and Tables S1 and S2) in which each residue in the polypeptide chain is represented by using two interaction sites, one that is centered on the α-carbon atom and another that is located at the center of mass of the side chain (26).

Molecular Transfer Model (MTM). The energy of a protein conformation at nonzero [C] is taken to be a sum of the potential energy E_P of the protein (see *SI Text*) and the transfer free energy $\Delta G_{tr}([C])$ based on TM. According to the TM, the free energy of transferring a protein to osmolyte solution is equal to the sum of the transfer free energies (TFEs) of the individual groups (side chain and backbone moieties) that are solvent exposed. The free-energy cost of transferring the *i*th protein conformation from water to aqueous osmolyte solution at concentration [C] is written as

$$\Delta G_{tr}(i,[C]) = \sum_{k=1}^{N_{SC}} \delta g_{tr,k}^{SC}([C]) n_k (\alpha_{i,k}^{SC}/\alpha_{k,Gly-k-Gly}^{SC}) + \sum_{k=1}^{N_{BB}} \delta g_{tr}^{BB}([C]) n_k (\alpha_{i,k}^{BB}/\alpha_{k,Gly-k-Gly}^{BB}) \quad [1]$$

where the sums are over the different amino acid types in the protein, n_k is the number of amino acid residues of type k, $\delta g_{tr,k}^{SC}$ and $\delta g_{tr,k}^{BB}$ are the transfer free energies of the side chain and backbone group of amino acid type k (Table S3), respectively (30, 32). For denaturants $\delta g_{tr} < 0$, that is, transfer is thermodynamically favorable for the peptide backbone and many types of amino acid side chains (9, 29, 45). The transfer of some of these substituents to an osmolyte solution results in $\delta g_{tr} > 0$ (45). The solvent accessible surface areas of the side chain and backbone group of amino acid type k are $\alpha_{i,k}^{SC}$ and $\alpha_{i,k}^{BB}$, respectively, and $\alpha_{k,\mathrm{Gly}-k-\mathrm{Gly}}^{SC}$ is the solvent-accessible surface area of the side chain and backbone in the tripeptide *Gly–k–Gly*.

BIOPHYSICS

To combine experimentally measured $\delta g_{tr,k}$'s with simulations at [C] = 0, we introduce the primary equation of MTM, that has the form of the Weighted Histogram Analysis Method (46–48), namely,

$$\langle A([C_i],T)\rangle = Z([C_i],T)^{-1}\sum_{k=1}^{R}\sum_{t=1}^{n_k}\frac{A_{k,t}e^{-\beta(E_P(k,t,[0])+\Delta G_{tr}(k,t,[C_i]))}}{\sum_{m=1}^{R} n_m e^{f_m-\beta_m E_m(k,t,[0])}}, \qquad [2]$$

where $Z([C_i],T)$ is the partition function. Thus, if $Z([0],T)$ is computed and the transfer free energy of each protein conformation is known, then any thermodynamic property, at arbitrary $[C_i]$, can be predicted. In Eq. **2**, R is the number of independent simulated trajectories, n_k is the number of conformations from the kth simulation, $A_{k,t}$ is the value of property A for the tth conformation, $\beta = 1/k_BT$, where k_B is Boltzmann's constant and T is the temperature. The potential energy of the tth conformation from the kth simulation in the presence of osmolyte i at concentration C_i is $E(k, t, [C_i]) = E_P(k, t, [0]) + \Delta G_{tr}(k, t, [C_i])$, where $E_P(k, t, [0])$ is the corresponding value at 0 M. The free-energy cost of transferring the tth conformation in the kth simulation from 0 M to $[C_i]$ M is $\Delta G_{tr}(k, t, [C_i])$. In the denominator of Eq. **2**, n_m and f_m are, respectively, the number of conformations and the free energy in the mth simulation.

The values $\alpha_{k,\text{Gly}-k-\text{Gly}}$ for the side chain and backbone groups (Eq. **1**) are listed in Table S4. For the osmolytes considered here (urea, glycine betaine, proline, sucrose, sarcosine, sorbitol, and TMAO) we use the TFEs given in ref. 45, and for aqueous GdmCl we use the transfer free energies listed in ref. 9. We extrapolate to osmolyte concentrations that were not experimentally measured by fitting the TFE data to a straight line (49) (see *SI Text* for details).

ACKNOWLEDGMENTS. We thank Ben Schuler and Ngo Toan for useful discussions. This work was supported in part by National Science Foundation Grants 05-14056 (to D.T.), the US-Israeli Binational Science Foundation Grant 2002371 (to D.T. and G.H.), National Institutes of Health Grant 1R01GM080515 (to G.H.), and a National Institutes of Health GPP Biophysics Fellowship (to E.O.), and the Intramural Research Program of the National Heart, Lung, and Blood Institute of the National Institutes of Health (B.B.).

1. Fersht AR (1999) *Structure and Mechanism in Protein Sci: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York), 2nd Ed.
2. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–336.
3. Thirumalai D, Klimov DK, Dima RI (2003) Emerging ideas on the molecular basis of protein and peptide aggregation. *Curr Opin Struct Biol* 13:146–159.
4. Onuchic JN, Wolynes PG (2004) Current opinion in structural biology. *Curr Opin Struct Biol* 14:70–75.
5. Thirumalai D, Hyeon C (2005) RNA and protein folding: Common themes and variations. *Biochemistry* 44:4957–4970.
6. Schaeffer RD, Fersht AR, Daggett V (2008) Combining experiment and simulation in protein folding: Closing the gap for small model systems. *Curr Opin Struct Biol* 18:4–9.
7. Baldwin RL (2007) Energetics of protein folding. *J Mol Biol* 371:283–301.
8. Auton M, Bolen DW (2007) Application of the transfer model to understand how naturally occurring osmolytes affect protein stability. *Methods Enzymol* 428:397–418.
9. Pace CN (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131:266–280.
10. Scholtz JM, Barrick D, York EJ, Steward JM, Baldwin RL (1995) Urea unfolding of peptide helices as a model for interpreting protein unfolding. *Proc Natl Acad Sci USA* 92:185–189.
11. Lee JC, Timasheff SN (1981) The stabilization of proteins by sucrose. *J Biol Chem* 256:7193–7201.
12. Robinson DR, Jencks (1965) WP Effect of compounds of urea-guanidinium class on activity coefficient of acetyltetraglycine ethyl ester and related compounds. *J Am Chem Soc* 87:2462–2469.
13. Schellman JA (2003) Protein stability in mixed solvents: A balance of contact interaction and excluded volume. *Biophys J* 85:108–125.
14. Wills PR, Winzor DJ (1993) Thermodynamic analysis of ''preferential solvation'' in protein solutions. *Biopolymers* 33:1627–1629.
15. Tanford C (1970) Protein denaturation. Part c. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24:1–95.
16. Klein-Seetharaman J, *et al.* (2002) Long-range interactions within a nonnative protein. *Science* 295:1719–1722.
17. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule fret. *Curr Opin Struct Biol* 18:16–26.
18. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
19. Rhoades E, Cohen M, Schuler B, Haran G (2004) Two-state folding observed in individual protein molecules. *J Am Chem Soc* 126:14686–14687.
20. Hoffman A, *et al.* (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104:105–110.
21. Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN (1982) Living with water-stress - evolution of osmolyte systems. *Science* 217:1214–1222.
22. Honeycutt JD, Thirumalai D (1990) Metastability of the folded states of globular-proteins. *Proc Natl Acad Sci USA* 87:3526–3529.
23. Klimov DK, Thirumalai D (1999) Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr Opin Struct Biol* 9:197–207.
24. Cheung MS, Finke JM, Callahan B, Onuchic JN (2003) Exploring the interplay between topology and secondary structural formation in the protein folding problem. *J Phys Chem B* 107:11193–11200.
25. Klimov DK, Thirumalai D (1998) Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J Chem Phys* 109:4119–4125.
26. Klimov DK, Thirumalai D (2000) Mechanisms and kinetics of beta-hairpin formation. *Proc Natl Acad Sci USA* 97:2544–2549.
27. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and ''en-route'' intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298:937–953.
28. Tobi D, Elber R, Thirumalai D (2003) The dominant interaction between peptide and urea is electrostatic in nature: A molecular dynamics simulation study. *Biopolymers* 68:359–369.
29. Nozaki Y, Tanford C (1963) Solubility of amino acids and related compounds in aqueous urea solutions. *J Biol Chem* 238:4074–4081.
30. Tanford C (1964) Isothermal unfolding of globular proteins in aqueous urea solutions. *J Am Chem Soc* 86:2050–2059.
31. Auton M, Bolen DW (2005) Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc Natl Acad Sci USA* 102:15065–15068.
32. Auton M, Holthauzen LMF, Bolen DW (2007) Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc Natl Acad Sci USA* 104:15317–15322.
33. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci USA* 103:11539–11543.
34. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA (2007) Characterizing the unfolded states of proteins using single-molecule fret spectroscopy and molecular simulations. *Proc Natl Acad Sci USA* 104:1528–1533.
35. Kim DE, Fisher C, Baker D (2000) A breakdown of symmetry in the folding transition state of protein l. *J Mol Biol* 298:971–984.
36. Perl D, *et al.* (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nature* 5:229–235.
37. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor-2. 1. Evidence for a 2-state transition. *Biochemistry* 30:10428–10435.
38. Santoro MM, Liu YF, Khan SMA, Hou LX, Bolen DW (1992) Increased thermal-stability of proteins in the presence of naturally-occurring osmolytes. *Biochemistry* 31:5278–5283.
39. Pradeep L, Udgaonkar JB (2004) Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J Biol Chem* 279:40303–40313.
40. Kohn JE, *et al.* (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101:12491–12496.
41. Dima RI, Thirumalai D (2004) Asymmetry in the shapes of folded and denatured states of proteins. *J Phys Chem B* 108:6564–6570.
42. Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D (1999) Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat Struct Biol* 6:554–556.
43. Arai M, *et al.* (2007) Microsecond hydrophobic collapse in the folding of Escherichia coli dihydrofolate reductase an alpha/beta-type protein. *J Mol Biol* 368:219–229.
44. Fisher ME (1966) Shape of a self-avoiding walk or polymer chain. *J Chem Phys* 44:616–622.
45. Auton M, Bolen DW (2004) Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry* 43:1329–1342.
46. Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data analysis. *Phys Rev Lett* 63:1195–1198.
47. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J Comput Chem* 13:1011–1021.
48. Shea J, Nochomovitz YD, Guo Z, Brooks CL (1998) Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist β-barrel model. *J Chem Phys* 109:2895–2903.
49. Makhatadze GI (1999) Thermodynamics of protein interactions with urea and guanidinium hydrochloride. *J Phys Chem B* 103:4781–4785.
50. Santoro MM, Bolen DW (1988) Unfolding free-energy changes determined by the linear extrapolation method. 1. unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27:8063–8068.
51. Yi Q, Scalley ML, Simons KT, Gladwin ST, Baker D (2007) Characterization of the free energy spectrum of peptostreptococcal protein l. *Fold Des* 2:271–280.