J|A|C|S COMMUNICATIONS

Published on Web 01/26/2009

Protein Folding Rates and Stability: How Much Is There Beyond Size?

David De Sancho,[†] Urmi Doshi,[‡] and Victor Muñoz*[,†,‡]

Centro de Investigaciones Biológicas, Spanish National Research Council (CSIC), Ramiro de Maeztu 9, Madrid 28040, Spain, and Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 27042

Received November 11, 2008; E-mail: vmunoz@cib.csic.es

From a chemical standpoint protein folding poses an intriguing conundrum. Theory predicts general trends that have also been observed empirically. Polymer-based size scaling is one such trend.[1,2] It has been found for the changes in conformational entropy, heat capacity, and stabilization enthalpy upon folding,[3] as well as in folding rates.[4] The native 3D structure is another important factor affecting overall folding rates.[5,6] However, finer details exhibit very complex behavior seemingly lacking any specific pattern. Here we address this issue using a simple theoretical procedure to reproduce size-scaling effects in protein folding and quantify the net contribution from the other factors to experimentally determined rates and stabilities.

We employ experimental data for 52 proteins ranging from 37 to 151 residues (see Supporting Information, SI) that have been measured in similar experimental conditions (i.e., "standard" folding conditions[7]) and do not show evidence of unfolding intermediates. Thus we factor out the strong temperature effects[8] and eliminate complex folding from multistate proteins.[9] We analyze folding rates and stabilities with a simple one-dimensional free energy surface model of protein folding that directly accounts for size-scaling effects.

The model has been described in detail before.[8] Briefly, it represents changes in the thermodynamic parameters upon folding as functions of the single local order parameter nativeness (n), which ranges from 0 (fully unfolded) to 1 (fully folded):

$$\Delta S^{conf}(n) = N(-R[n\ln(n) + (1-n)\ln(1-n)] + n\Delta S_{res}^{n=1} + (1-n)\Delta S_{res}^{n=0});$$

$$\Delta H^{385K}(n) = N\Delta H_{res}^{385K}[1 + (\exp(\kappa_{\Delta H} n) - 1)/(1 - \exp(\kappa_{\Delta H}))];$$

$$\Delta C_p(n) = N\Delta C_{p,res}[1 + (\exp(\kappa_{\Delta C_p} n) - 1)/(1 - \exp(\kappa_{\Delta C_p}))];$$

with N being the number of protein residues. One-dimensional free energy surfaces are directly obtained from these functions.[8] The basic model parameters are the changes in conformational entropy (ΔS_{res}), heat capacity ($\Delta C_{p,res}$), stabilization enthalpy (ΔH_{res} at 385 K) upon folding per residue, and the exponential constants defining the enthalpy and heat capacity curvatures ($\kappa_{\Delta H}$, $\kappa_{\Delta Cp}$). The balance between entropy and enthalpy at 298 K determines protein stability. The barrier (and thus folding rate) is ultimately defined by the curvature of the stabilization enthalpy ($\kappa_{\Delta H}$). Folding rates are obtained as diffusion on the free energy surface using a size-dependent diffusion coefficient ($D = k_0/N$).

The contribution of size to folding rates and stabilities is illustrated in Figure 1. These calculations use single values of all model parameters. Most were obtained from previous empirical estimates, whereas ΔH_{res} and $\kappa_{\Delta H}$ were fitted to obtain best agreement with experiment. Figure 1A shows that size effects alone reproduce the general trends in absolute midpoint folding rates ($R = 0.61$), although the correlation fans out as the rates get slower. The R is lower than previously reported for correlations with $N^{1/\alpha}$ ($R = 0.68$ for $\alpha = 1$, see ref 4) on a larger database. However, the range of protein sizes is much narrower here. The correlation with the relative contact order (RCO[5]) is rather weak ($R = -0.52$) but grows to $R = -0.78$ when the contact order is not normalized by the number of residues (ACO[10]), further confirming a main role for size. It is important to emphasize that, in contrast to empirical correlations, the calculations in Figure 1 are absolute predictions from a simple statistical mechanical model and a set of physically sensible parameters.

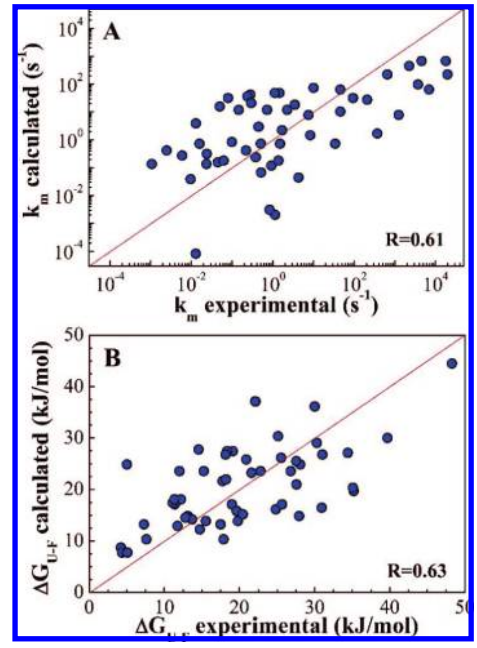

Figure 1. Comparison between experimental midpoint folding rates (A) and stabilities in water at 298 K (B) with model predictions using only size scaling. 1:1 correlation line is shown to guide the eye. The parameters are ΔS_{res} = 16.5 J/(mol·K), $\Delta C_{p,res}$ =58 J/(mol·K),[3] $\kappa_{\Delta Cp}$ = 4.3,[8] $k_0 = 8 \times 10^4$ n^2 s^{-1}, which is equivalent to a pre-exponential of $3.5 \times 10^6/N$ s^{-1} at 298 K, ΔH_{res}(385 K) = 6.4 kJ/mol and $\kappa_{\Delta H} = 3$.

More striking is the comparison between experimental folding stabilities and the model predictions (Figure 1B). Here we observe that protein stability is also largely determined by size. Therefore, the size scaling of folding entropy and enthalpy[3] is also preserved in the final stability, which is the leftover of an almost perfect cancelation between these two large opposing forces.

Size is then a primary factor in protein folding. The important question is how much do other factors contribute. We address it by fitting the model to exactly reproduce the experimental data and then analyzing the parameter variability. Particularly, we use the folding relaxation rate as a function of chemical denaturant (i.e., chevron plot) for the 52 proteins. These data include all the information relative to stability and rates as well as their chemical denaturant dependence. To describe chemical denaturation effects we employ the empirical equation:[8]

[†] Spanish National Research Council (CSIC).
[‡] University of Maryland.

10.1021/ja808843h CCC: $40.75

$$\Delta G(F_D, n) = \Delta G^{H_2O} + [(1 + C)(n^j/(n^j + C)) - 1]F_D$$

where F_D is the chemical denaturation free energy (assumed linear with denaturant concentration), and C and j are *ad hoc* parameters that determine the relative slopes of folding and unfolding limbs in the chevron plot. For simplicity, D, heat capacity, and ΔS_{res} were fixed to the empirical values of Figure 1. These parameters are expected to change less from protein to protein than the stabilization energy. We could thus fit the chevron plots with only four floating parameters: C, j, ΔH_{res}, and $\kappa_{\Delta H}$. Attributing all structure and sequence variability to the energetics (i.e., ΔH_{res} and $\kappa_{\Delta H}$) simplifies the analysis without compromising its main results.

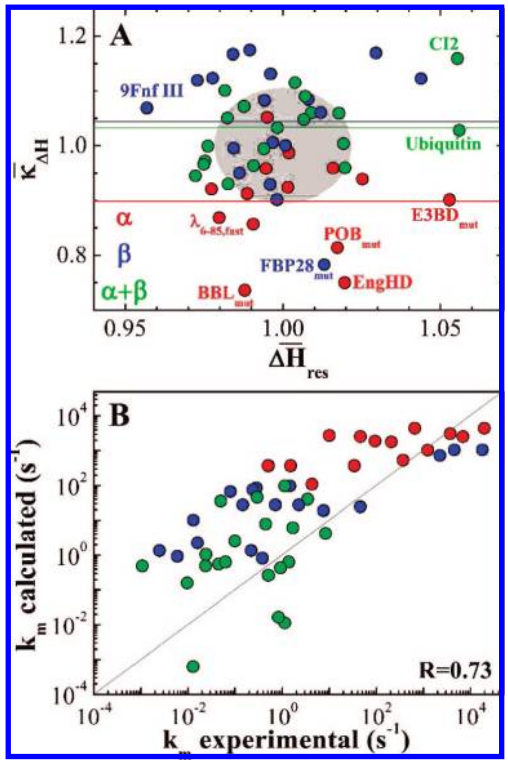

Figure 2. Analysis of protein stability and folding rates. (A) $\kappa_{\Delta H}$ and ΔH_{res} values (normalized against mean values) from fitting the 52 experimental chevron plots. Colors are assigned according to the three basic structural types. The gray area signals 1 standard deviation in each axis. (B) Correlation between experimental and calculated midpoint folding rates using an average $\kappa_{\Delta H}$ for each structural type (horizontal lines in panel B).

The results after the 52 fits (see examples in SI) are shown in Figure 2A. All experimental data were amenable to the highly constrained analysis, indicating consistency between experiments, model, and set of empirical parameters. Figure 2A shows that experimental variability in folding rates and stabilities beyond size is caused by very little changes in the two fundamental parameters of the model. Indeed, the deviations in folding rate are accounted for with $\kappa_{\Delta H}$ values that differ by only 8% on average. More remarkably, the variability in stabilization enthalpy (ΔH_{res}) is a meager 1.7% on average (Figure 2A). The plot shows the majority of dots clustering around the gray area, which signals 1 standard deviation on each parameter.

However, there are several proteins more scattered on the plot, highlighting interesting idiosyncrasies. For example, the paradigmatic two-state folder CI2 appears on the upper right of the plot indicating a proportionally higher stability and folding barrier. As an extracellular protease inhibitor, those special CI2 features could be the result of strong evolutionary selection. The intracellular protein degradation-involved ubiquitin is also more stable, whereas 9Fnf-III emerges as an unusually unstable protein, as it was originally proposed.[11] The lower part of the plot is populated by fast, downhill or near downhill folding proteins. Lower than average $\kappa_{\Delta H}$ for these proteins suggests again specific selection, supporting a biological role for downhill folding.[12]

The analysis shown in Figure 2A allows evaluating the role of other contributions to folding. A simple surface/volume correction for size scaling ($N-1.5N^{2/3}$ instead of N; see ref 13) renders slightly worse results, indicating that surface effects are not significant in the 37−151 residues size range. The influence of structure on the folding rates is noticed as a trend to lower $\kappa_{\Delta H}$ for α-helical proteins (horizontal lines in Figure 2A). Indeed, using distinct $\kappa_{\Delta H}$ values for the three structural types improves the rate prediction (Figure 2B). The agreement only is slightly better when $\kappa_{\Delta H}$ is made proportional to the *RCO* (R = 0.78), suggesting that structural effects on rates are mostly coarse-grained. The three structural types have similarly broad distributions, indicating decoupling between sequence and structure. Structural effects seem to be negligible for protein stability, with similar mean and standard deviation for ΔH_{res} for the three structural types.

We thus find that protein size is the primary factor determining folding rates, and also protein stability. Moreover, experimental deviations from the size prediction due to structure and sequence effects correspond to minute differences in the fundamental folding parameters. The reason is that folding involves exquisitely balanced large and opposing energy terms. Thus minimal perturbations translate into significant changes in rates and stability. Such sensitivity puts a powerful toolbox for engineering folding at our disposal. However, judging from the homogeneity observed in real proteins (Figure 2A), evolution seems to have selected a narrow subset of possibilities. This observation suggests the presence of strong folding constraints when building biologically functional proteins. In practical terms, our results and analysis highlight that predicting folding more precisely shall require highly accurate protein force fields and provide useful benchmarks toward developing such methods.

Acknowledgment. This work has been supported by NSF grant MCB-0317294 and Marie Curie Excellence Award MEXT-CT-2006-042334.

Supporting Information Available: Database of folding rates, stabilities, experimental conditions, *RCO* values, and complete ref 7. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Gutin, A. M.; Abkevich, V. I.; Shakhnovich, E. I. *Phys. Rev. Lett.* **1996**, *77*, 5433–5436.
(2) Li, M. S.; Klimov, D. K.; Thirumalai, D. *Polymer* **2004**, *45*, 573–579.
(3) Robertson, A. D.; Murphy, K. P. *Chem. Rev.* **1997**, *97*, 1251–1267.
(4) Naganathan, A. N.; Muñoz, V. *J. Am. Chem. Soc.* **2005**, *127*, 480–481.
(5) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *227*, 985–994.
(6) Muñoz, V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.SA.* **1999**, *96*, 11311–11316.
(7) Maxwell, K. L.; et al. *Protein Sci.* **2005**, *14*, 602–616.
(8) Naganathan, A. N.; Doshi, U.; Muñoz, V. *J. Am. Chem. Soc.* **2007**, *129*, 5673–5682.
(9) Galzitskaya, O. V.; Garbuzynskiy, S. O.; Ivankov, D. N.; Finkelstein, A. V. *Proteins* **2003**, *51*, 162–166.
(10) Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K. W.; Baker, D.; Finkelstein, A. V. *Protein Sci.* **2003**, *9*, 2057–2062.
(11) Plaxco, K. W.; Spitzfaden, C.; Campbell, I. D.; Dobson, C. M. *J. Mol. Biol.* **1997**, *270*, 763–770.
(12) Muñoz, V. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 395–412.
(13) Finkelstein, A. V.; Badretdinov, A. Y. *Fold. Des.* **1997**, *2*, 115–121.

JA808843H