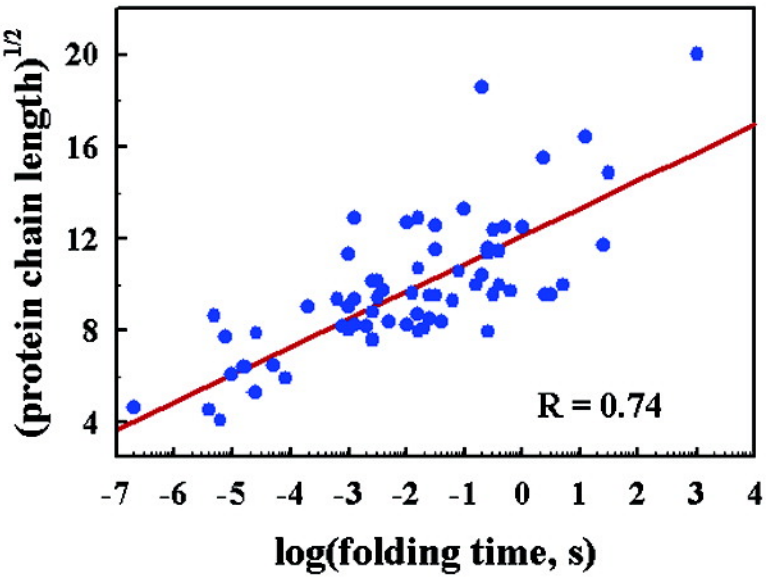
Communication

Scaling of Folding Times with Protein Size

Athi N. Naganathan, and Victor Muoz

J. Am. Chem. Soc., **2005**, 127 (2), 480-481• DOI: 10.1021/ja044449u • Publication Date (Web): 18 December 2004



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 8 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
COMMUNICATIONS

Published on Web 12/18/2004

Scaling of Folding Times with Protein Size

Athi N. Naganathan and Victor Muñoz*

Department of Chemistry and Biochemistry, and Center for Biomolecular Structure and Organization, University of Maryland, College Park, Maryland 20742

Received September 13, 2004; E-mail: vm48@umail.umd.edu

For some proteins it takes an hour to spontaneously fold into their biologically active three-dimensional (3D) structures, while others manage to do it in mere microseconds.[1] The 9-orders-of-magnitude range in observed folding times has usually been considered a manifestation of the large variability in structural patterns and amino acidic sequences found in natural proteins.[2] Here we show that folding times correlate with protein size, following a simple scaling law similar to those found in a growing variety of biological processes,[3] but operating at the molecular level.

In recent years there has been strong emphasis in trying to understand the connection between the 3D structure of a protein and its folding time. Simple structural descriptors, such as the relative contact order, that condense the structural features of proteins into a single number have been found to correlate with the folding time of small single-domain proteins.[2] Similar prediction accuracy has been obtained with a more mechanistic approach in which folding times are calculated from protein structures with statistical mechanical models of folding.[4] However, the prediction of folding times of larger multidomain proteins seems to require the additional consideration of protein size.[1,5] The obvious question that arises is: how strong a determinant of folding time is protein size alone?

Starting with the seminal work of Thirumalai,[6] several theoretical studies have emphasized that folding times should scale with the size of the protein following the simple relation $\log(\tau_F) \approx N^{\beta}$, in which the value of β should lie between $^1/_2$ and $^2/_3$.[7−9] Recently, an empirical correlation between folding times and protein size has been reported.[10] Here we further investigate this issue by looking at a compilation of experimentally determined folding times for 69 proteins/peptides with size ranging from 16 to 396 residues. This compilation includes all of the two- and three-state proteins that have been used by Ivankov et al. in a previous analysis.[1] The only exception is cytochrome b_{562} because the available data come from a very long extrapolation. Additionally, we have incorporated the folding times of two proteins with disulfide bonds (Tendamistat: 72 residues, 15 ms; and egg-white lysozyme: 129 residues, 250 ms), of the proteins BBL (40 residues, 16 μs) and FSD-1 (27 residues, 24 μs), which have been determined in our laboratory, of the designed helix bundle α3D (73 residues, 3 μs[11]), and of seven-repeat ankyrin (239 residues, 2.5 s; Doug Barrick, personal communication). When we analyze this database we find that there is a strong correlation between folding time and protein size (Figure 1). The correlation coefficient (R) is 0.74 when the logarithm of folding times is correlated to the square root of the number of residues in the protein ($N^{1/2}$) and converges to a maximum value of ∼0.78 as the exponent approaches 0 (inset of Figure 1).

The correlation between folding time and protein size is better than previously found with a smaller database.[11] This observation is a strong indication that the scaling law is robust because the database of 69 proteins increases the dynamic range significantly by adding experimental data on very small and very large proteins.

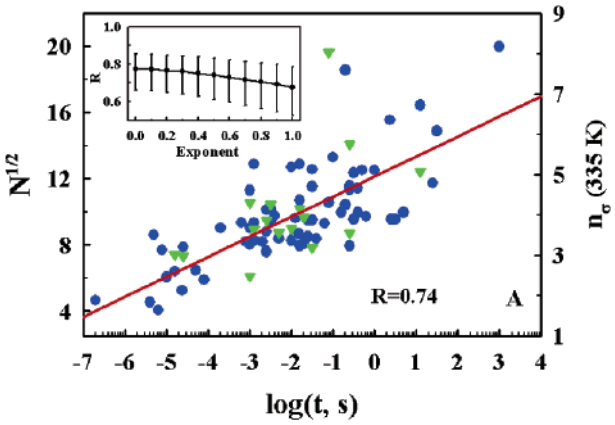


Figure 1. Linear correlation between experimental folding times and the square root of the number of residues. (Blue circles) Data from 69 proteins and peptides. The red line corresponds to the linear fit to the data. (Green triangles) Folding times directly calculated from n_σ for the subset of proteins with thermodynamic data available.[12] (Inset) Dependence of the correlation coefficient with the magnitude of the exponent of N.

In fact the correlation coefficient that we report here is almost as good as the one previously obtained by Ivankov et al. when predicting folding times with a combination of protein size and secondary structure prediction.[1] Therefore, folding times increase with protein size as expected from polymer dynamics arguments. An important practical implication of this finding is that we can predict folding times with a precision of ∼1.1 time decades by just knowing the size of a protein, a property most easily obtained experimentally.

With the scarcity of data currently available the characteristic exponent can have any value from 0.5 to 0. It is not even clear whether trying to find the exponent that maximizes the correlation coefficient is the right strategy when there is intrinsic "noise" from other factors, such as structure and sequence. However, an exponent of 0.5, as originally predicted by Thirumalai,[6] offers an intriguing thermodynamic interpretation to the scaling law. To derive this interpretation we define the parameter:

$$n_\sigma = \frac{\Delta H(T)}{\sqrt{RT^2 \Delta C_p}}$$

where the numerator corresponds to the difference in enthalpy between unfolded and folded states and the denominator accounts for the enthalpy fluctuations of the unfolded state in excess of those of the folded state. Assuming that the heat capacity of the folded protein is entirely due to nonstructural enthalpy fluctuations, the latter can be interpreted as the fraction of the enthalpy fluctuations of the unfolded chain and solvent that is associated with the folding process.[13] Thus, n_σ is related to the frequency at which the unfolded state reaches enthalpy values characteristic of the native state

10.1021/ja044449u CCC: $30.25

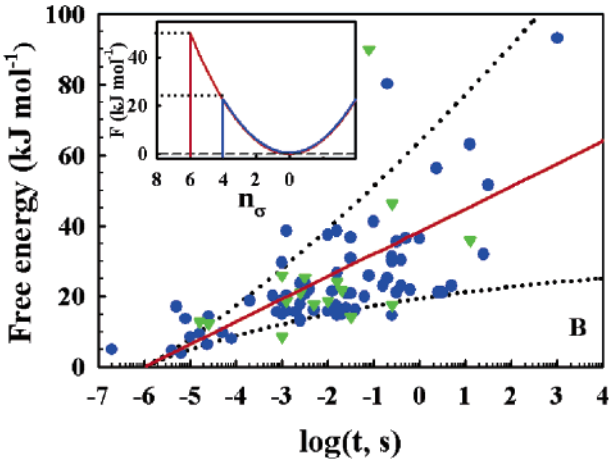


Figure 2. Free energy barriers for the proteins shown in Figure 1 calculated from the height of the harmonic potential at n_σ (335 K) standard deviations from the unfolded minimum (see inset). The red line shows the folding rates calculated with a preexponential of (1/1μs) and barrier heights as a function of protein size obtained using ΔH (333K) = 2.92 kJ/(mol·res) and ΔC_p = 58 J/(mol·K·res).[12] Dashed lines correspond to ±1 standard deviation from the red line and show that the uncertainty increases at longer folding times−larger protein sizes.

minimum. Because ΔH and ΔC_p scale linearly with protein size,[12] n_σ scales as $N^{1/2}$. Indeed, direct calculation of n_σ for the subset of proteins for which there is thermodynamic data produces the same trend observed for $N^{1/2}$ (green triangles and right scale in Figure 1).

In principle, n_σ can be calculated at any temperature, but one of them is of particular interest. The solvent contribution to ΔH varies linearly with temperature, so there is one temperature at which the solvent contributions to ΔH and ΔC_p compensate exactly. At this temperature ΔH could act as an order parameter for folding, or perhaps even as a reasonable reaction coordinate, so that it might be possible to estimate heights of folding barriers from n_σ. Under these conditions the free energy barrier can be estimated with a potential of mean-force approach: the unfolded state is represented as a harmonic well and the native state as an infinitely sharp potential (i.e. no structural fluctuations) placed at n_σ standard deviations from the minimum of the harmonic well. The height of the barrier is simply obtained from the intersect of the two potentials (inset to Figure 2).

It is not unreasonable to expect that the solvation contributions to ΔH and ΔC_p converge at similar temperatures for all proteins.[12] However, we do not know in what range of temperatures this compensation should occur. What we find is that the scaling with protein size of the barriers estimated with our procedure agrees with the experimental data when n_σ is calculated at ~335 K (Figure 2). At temperatures lower than 335 K the estimated slope is too low and at higher temperatures is too steep. Interestingly, the calculation at 335 K produces a folding speed limit (crossing with the abscissa) of ~1 μs, which is in agreement with current empirical estimates.[14] The remaining variability in folding times (i.e. differences in folding time among proteins of the same size) can be explained with only small changes in the reference temperature at which to calculate n_σ (335 ± 11 K). In other words, the effects on the folding time of protein structure and sequence are rather small from a thermodynamic standpoint.

These observations unveil surprisingly simple patterns underneath the apparent complexity of protein folding. Furthermore, our analysis suggests that the expectation for proteins of less than 50 residues (i.e. $N^{1/2} \leq 7$) is to have very marginal folding free energy barriers (i.e. <12 kJ/mol).

Acknowledgment. V.M. is a recipient of a Dreyfus New Faculty Award, a Packard Fellowship for Science and Engineering, and a Searle Scholarship. The research described here has been supported by NIH Grant GM066800-1 and NSF Grant MCB-0317294.

References

(1) Ivankov, D. N.; Finkelstein, A. V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8942−8944.
(2) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *227*, 985−994.
(3) West, G. B.; Brown, J. H.; Enquist, B. J. *Science* **1997**, *276*, 122−126.
(4) Muñoz, V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11311−11316.
(5) Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K. W.; Baker, D.; Finkelstein, A. V. *Protein Sci.* **2003**, *9*, 2057−2062.
(6) Thirumalai, D. *J. Phys.* **1995**, *5*, 1457−1467.
(7) Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6170−6175.
(8) Finkelstein, A. V.; Badretdinov, A. Y. *Folding Des.* **1998**, *3*, 67−68.
(9) Gutin, A. M.; Abkevich, V. I.; Shakhnovich, E. I. *Phys. Rev. Lett.* **1996**, *77*, 5433−5436.
(10) Li, M. S.; Klimov, D. K.; Thirumalai, D. *Polymer* **2004**, *45*, 573−579.
(11) Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C. Y.; Lahr, S. J.; Daggett, V.; Roder, H.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15486−15491.
(12) Robertson, A. D.; Murphy, K. P. *Chem. Rev.* **1997**, *97*, 1251−1267.
(13) Muñoz, V.; Sanchez-Ruiz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**. In press.
(14) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76−88.

JA044449U