FOR THE RECORD

Protein unfolding rates correlate as strongly as folding rates with native structure

Aron Broom,[1] Shachi Gosavi,[2] and Elizabeth M. Meiering[1]*

[1]Department of Chemistry, Guelph-Waterloo Centre for Graduate Studies in Chemistry and Biochemistry, University of Waterloo, Waterloo, Ontario, Canada N2L 1W2

[2]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, Karnataka 560065, India

Received 10 September 2014; Accepted 4 November 2014
DOI: 10.1002/pro.2606
Published online 25 November 2014 proteinscience.org

Abstract: Although the folding rates of proteins have been studied extensively, both experimentally and theoretically, and many native state topological parameters have been proposed to correlate with or predict these rates, unfolding rates have received much less attention. Moreover, unfolding rates have generally been thought either to not relate to native topology in the same manner as folding rates, perhaps depending on different topological parameters, or to be more difficult to predict. Using a dataset of 108 proteins including two-state and multistate folders, we find that both unfolding and folding rates correlate strongly, and comparably well, with well-established measures of native topology, the absolute contact order and the long range order, with correlation coefficient values of 0.75 or higher. In addition, compared to folding rates, the absolute values of unfolding rates vary more strongly with native topology, have a larger range of values, and correlate better with thermodynamic stability. Similar trends are observed for subsets of different protein structural classes. Taken together, these results suggest that choosing a scaffold for protein engineering may require a compromise between a simple topology that will fold sufficiently quickly but also unfold quickly, and a complex topology that will unfold slowly and hence have kinetic stability, but fold slowly. These observations, together with the established role of kinetic stability in determining resistance to thermal and chemical denaturation as well as proteases, have important implications for understanding fundamental aspects of protein unfolding and folding and for protein engineering and design.

Keywords: kinetic stability; protein design; scaffold selection; contact order; topology; folding rate; unfolding rate; long range order; protein engineering; structural complexity

Abbreviations: ACO, absolute contact order; LRO, long range order; RCO, relative contact order.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Science and Engineering Council (NSERC) (to E.M.M.); Grant sponsor: Government of India-DST (Ramanujan Fellowship, 5 years, wef 29/06/2010 [to S.G.]); Grant number: SR/S2/RJN-63/2009.

*Correspondence to: Elizabeth M. Meiering; Guelph-Waterloo Centre for Graduate Studies in Chemistry and Biochemistry, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 1W2. E-mail: meiering@uwaterloo.ca

Introduction

Extensive experimental[1,2] and theoretical[3–5] research has been conducted to understand protein folding rates. In seminal studies, Plaxco *et al*. found that a simple measure of the topology of the native state, the relative contact order (RCO), correlated well with folding rates for a small set of monomeric proteins that showed two-state behavior.[6] Later, they revised their conclusions to show that absolute contact order (ACO), which in addition to topology includes effects of protein length, correlated better for a larger dataset including multistate folders.[7] It has long been noted that folding rates depend on protein length, but the quantitative, physical basis for this dependence remains under investigation[8–13] and additional studies have demonstrated that consideration of both topology and length leads to improved correlations compared to topology alone.[14,15] Here, we use structural complexity as a broad term to encompass the complexity imparted by both the topology and the length of the protein, and rate, when referring to the unfolding and folding rate constants (k_u and k_f, respectively).

The correlation of ACO with folding rates suggests that while the transition state for folding lacks much of the well-defined structure of the native state, it nevertheless has a broadly similar structure and complexity, as has been suggested from theory and simulation[5,15–17] and demonstrated experimentally.[18] Many alternative measures of native state structural complexity have also been found to correlate with folding rates.[19–24] In particular, two measures have emerged as being consistently well correlated: the ACO and the long range order (LRO). LRO was found not only to correlate well overall[19] but also to correlate well across different structural classes of proteins.[14] Recently, the correlation of folding rates with ACO was explicitly derived from theory.[25]

In contrast, relatively little work on the relationship between unfolding rates and native structure has been reported, with the existing studies suggesting that while native structure should correlate with unfolding,[26] the measures of structural complexity that work well for predicting folding do not perform well for unfolding.[27–29] In particular, work by Jung *et al*. concluded that while structural complexity does correlate with both unfolding and folding rates, the predictive parameters are different.[27,28] Harihar and Selvaraj compared LRO with unfolding rates, finding a moderate correlation overall, but greatly differing correlations for the alpha, beta, and mixed structural classes.[29] In these studies, relatively small datasets of $\sim$ 25 two-state folding proteins were used. In particular, the approach in two later studies was to use the small consensus dataset compiled by Maxwell *et al*.[30] in order to avoid noise in the data resulting from experimental differences.[27,29] However, noise resulting from sequence-specific effects can also be considerable, as noted in studies of homologous proteins.[31–33] Thus, the use of a small dataset for examining a relationship with such considerable noise may be a fraught endeavor. For instance, the original very strong correlation of folding rate with RCO[6] was later found to be considerably weaker when larger datasets including multistate folders were used.[7,14]

Here, in order to clearly identify general relationships between folding/unfolding rates and structural complexity, we have used a relatively large set of kinetic data for monomeric proteins obtained using similar experimental approaches. Proteins with disulfide bonds or large prosthetic groups are excluded because they are known to cause anomalous kinetics. The dataset is largely similar to that of Garbuzynskiy *et al*., who recently elucidated relationships between protein length, stability, and folding rates.[34] Using our dataset of 108 two-state and multistate folders (see Methods, Supporting Information Table SII), we tested various measures of structural complexity, discovering that two commonly used parameters, ACO and LRO, not only correlate strongly with folding rates but also correlate strongly, and equally well, with unfolding rates. Furthermore, the results are very similar for different structural classes of proteins. Importantly, these results address the previously reported apparent differences between the structural determinants of unfolding and folding rates which may have been a consequence of the comparatively smaller datasets that obscured the true relationships. The finding that the same measures of structural complexity are equally predictive of unfolding and folding rates has important implications for fundamental understanding of the process of protein unfolding. It also suggests that for the protein engineer, a key choice needs to be made when selecting a scaffold for design to achieve the desired balance between the typically desirable properties of fast folding and slow unfolding.

Results and Discussion

To identify general trends with increased confidence, we used a previously established large dataset[30,34] augmented with additional proteins (see Methods). We analyzed this dataset using a range of measures of structural complexity found previously to be correlated with folding rates (Supporting Information Table SI). Two well-established parameters, ACO and LRO (see Methods), exhibited superior correlations, which are described in detail later. We note that the trends for unfolding rates for the full dataset shown in Figure 1 also hold in general (with some variations in statistical significance) for various subsets of the data (Table I).

Unfolding rates correlate with ACO, LRO, folding rates, and stability

Strikingly, the logarithm of both k_u and k_f have equally negative correlation coefficients with ACO and LRO [Fig. 1(A–D), Table I], which suggests

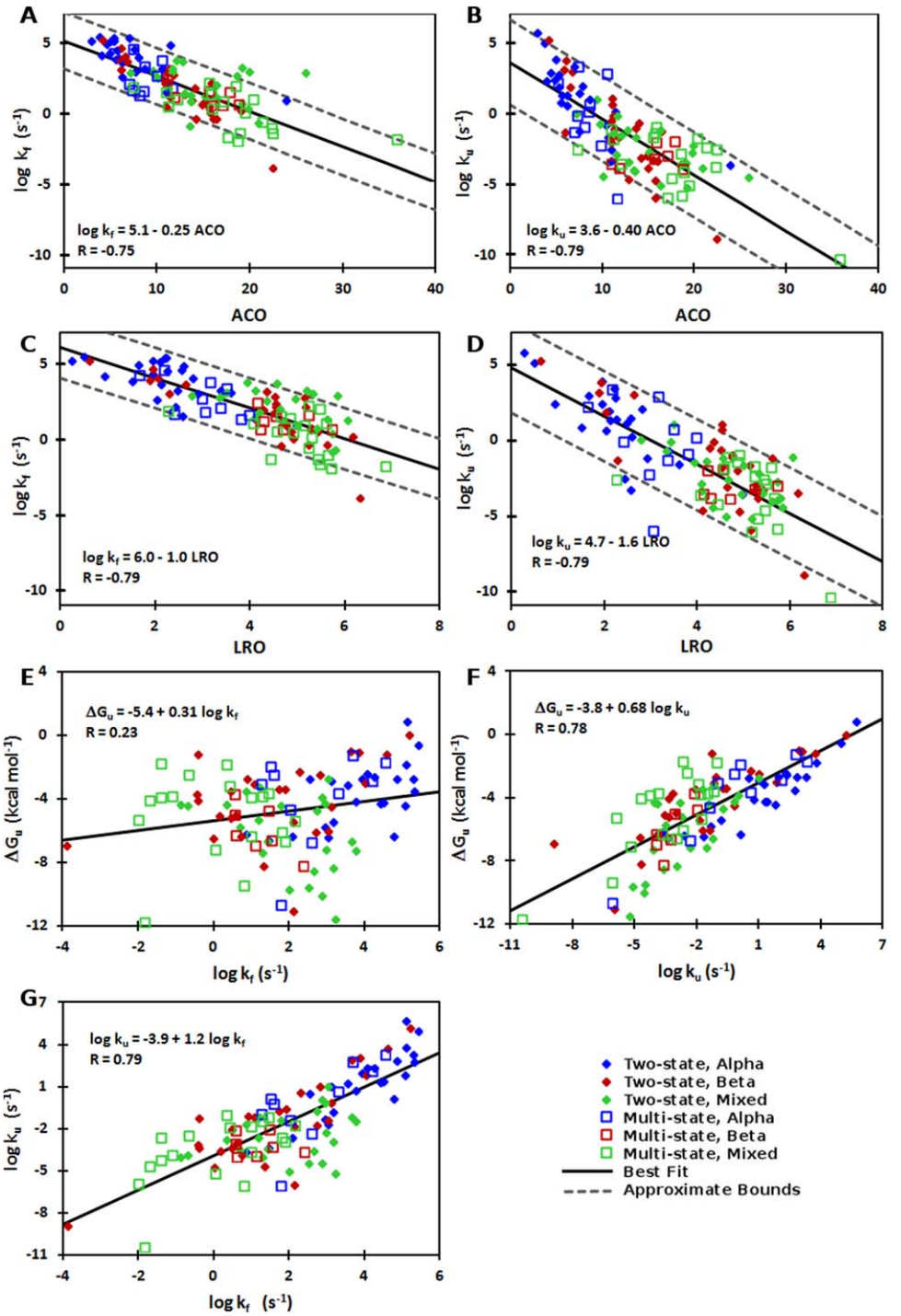

Figure 1. Correlations between structural complexity, folding and unfolding rates, and thermodynamic stability. Correlations are shown between (A) folding rates and ACO, (B) unfolding rates and ACO, (C) folding rates and LRO, (D) unfolding rates and LRO, (E) folding rates and thermodynamic stability, (F) unfolding rates and thermodynamic stability, and (G) unfolding and folding rates. The lines of best fit (solid black) and corresponding equations and correlation values are given for the whole dataset, values for subsets of data are given in Table I for two-state (filled diamonds), multistate (open squares), alpha (blue), beta (red), and mixed (green) proteins. Dotted lines for panels A–D denote ±10-fold and ±100-fold variation in k_f and k_u, respectively.

these measures of structural complexity are similarly predictive of both rates (small variations in correlation are expected owing to differing experimental conditions and the necessary extrapolation of the rates from kinetic data). Thus, both unfolding and folding rates decrease with increasing structural complexity. These results contrast with those of previous studies using smaller datasets of 22 and 25 proteins which concluded that one parameter was not equally well suited for predicting both

Table I. *Correlations and Linear Fits of Unfolding and Folding Rate Constants, Measures of Native Structure, and Thermodynamic Stability*

			Linear fit: $y = b + m^* x$		Pearson correlation	
Parameter x	Parameter y	Dataset (size)	m	b	R	P^a
ACO	$\log k_f$	Full (108)	−0.25	5.1	−0.75	1.4×10^{-20}
		Two-state (73)	−0.25	5.4	−0.73	3.6×10^{-13}
		Multistate (35)	−0.20	4.0	−0.75	2.7×10^{-7}
		Alpha (33)	−0.22	5.3	−0.62	1.1×10^{-4}
		Beta (34)	−0.35	6.1	−0.86	4.9×10^{-11}
		Mixed (41)	−0.16	3.7	−0.52	5.1×10^{-4}
		Maxwell (28)	−0.14	4.1	−0.48	9.5×10^{-3}
ACO	$\log k_u$	Full (108)	−0.40	3.6	−0.79	3.0×10^{-24}
		Two-state (73)	−0.46	4.4	−0.82	3.9×10^{-19}
		Multistate (35)	−0.30	1.9	−0.70	2.8×10^{-6}
		Alpha (33)	−0.48	4.5	−0.70	6.4×10^{-6}
		Beta (34)	−0.55	5.4	−0.82	3.5×10^{-9}
		Mixed (41)	−0.25	1.0	−0.67	2.0×10^{-6}
		Maxwell (28)	−0.30	2.1	−0.71	2.7×10^{-5}
LRO	$\log k_f$	Full (108)	−1.0	6.0	−0.79	2.9×10^{-24}
		Two-state (73)	−0.94	6.0	−0.80	2.0×10^{-17}
		Multistate (35)	−1.0	5.6	−0.75	2.7×10^{-7}
		Alpha (33)	−0.96	5.9	−0.69	1.0×10^{-5}
		Beta (34)	−1.1	6.3	−0.82	4.3×10^{-9}
		Mixed (41)	−0.95	5.8	−0.54	2.6×10^{-4}
		Maxwell (28)	−0.93	6.2	−0.68	7.6×10^{-5}
LRO	$\log k_u$	Full (108)	−1.6	4.7	−0.79	6.9×10^{-24}
		Two-state (73)	−1.6	4.9	−0.82	3.4×10^{-19}
		Multistate (35)	−1.4	3.7	−0.64	3.5×10^{-5}
		Alpha (33)	−1.8	5.2	−0.68	1.3×10^{-5}
		Beta (34)	−1.7	5.9	−0.79	3.6×10^{-8}
		Mixed (41)	−1.2	3.0	−0.56	1.4×10^{-4}
		Maxwell (28)	−1.5	4.7	−0.74	5.6×10^{-6}
$\log k_f$	$\Delta G_{F\text{-}U}$	Full (108)	0.31	−5.4	0.23	1.8×10^{-2}
		Two-state (73)	0.37	−5.6	0.27	2.3×10^{-2}
		Multistate (35)	0.23	−5.1	0.15	4.0×10^{-1} NS
		Alpha (33)	0.78	−6.6	0.47	5.2×10^{-3}
		Beta (34)	0.49	−5.3	0.36	3.8×10^{-2}
		Mixed (41)	−0.52	−5.1	−0.33	3.6×01^{-2}
		Maxwell (28)	−0.49	−4.4	−0.26	1.9×10^{-1} NS
$\log k_u$	$\Delta G_{F\text{-}U}$	Full (108)	0.68	−3.8	0.78	2.4×10^{-23}
		Two-state (73)	0.69	−4.0	0.80	2.2×10^{-17}
		Multistate (35)	0.76	−3.0	0.79	1.4×10^{-8}
		Alpha (33)	0.77	−4.4	0.89	7.0×10^{-12}
		Beta (34)	0.69	−3.3	0.83	1.7×10^{-9}
		Mixed (41)	0.82	−3.2	0.66	3.0×10^{-6}
		Maxwell (28)	0.98	−3.8	0.77	1.9×10^{-6}
$\log k_f$	$\log k_u$	Full (108)	1.2	−3.9	0.79	6.4×10^{-24}
		Two-state (73)	1.3	−4.1	0.79	7.4×10^{-17}
		Multistate (35)	1.2	−3.8	0.72	1.1×10^{-6}
		Alpha (33)	1.6	−4.8	0.83	2.6×10^{-9}
		Beta (34)	1.4	−3.9	0.82	2.6×10^{-9}
		Mixed (41)	0.62	−3.7	0.50	9.8×10^{-4}
		Maxwell (28)	0.64	−3.3	0.42	2.5×10^{-2}

Individual correlations and linear fits are shown for subsets of the data as in Figure 1. Additionally, values for the commonly used dataset of Maxwell *et al*. are shown for comparison.

[a] Two-tailed probability value.

NS Correlation is not significant at the 0.05 level (5.0×10^{-2}).

$\Delta G_{F\text{-}U} = G_F - G_U = -RT \ln (k_f/k_u)$, where R is the gas constant and T is the absolute temperature in Kelvin, gives the Gibbs free energy of the folded state relative to the unfolded state.

unfolding and folding.[27,28] Further, we find that the correlation for unfolding holds well across different structural classes (Table I), whereas another analysis[29] suggested all-beta proteins have a much weaker correlation with LRO that is opposite in sign to that for all-alpha and mixed structural classes. These apparent discrepancies are likely caused by the small dataset sizes used in the earlier study

where the alpha and beta classes had 5 and 7, compared here with 33 and 34 proteins, respectively.

We also note that unfolding rates are strongly correlated with folding rates [Fig. 1(E) and Table I], and that ACO and LRO are strongly correlated with rates at the transition midpoints (i.e., under conditions of equal thermostability; Supporting Information Fig. S1, Table SI). Thus, ACO and LRO may report on the structural complexity and relative energy of the transition state.[35] There is also a weaker correlation between protein length and the unfolding and folding rates (Supporting Information Table SI). Together these trends indicate that protein topological complexity and size affect both folding and unfolding rates.

Lastly, there is a strong correlation between unfolding rate and thermodynamic stability [Fig. 1(F), Table I]. In contrast, the correlation of folding rate with thermodynamic stability is weak [Fig. 1(G), Table I], as has been found previously.[6,36] The larger contribution of unfolding rate to thermodynamic stability has been noted before,[27,28,32,33,37] and is also apparent in the differences between the upper and lower limits of the dataset, where the upper limit on fast folding and unfolding is similar, while the lower bound for unfolding is substantially slower than that for folding (Supporting Information Table SII). These results suggest that variations in thermodynamic stability, which are determined by the ratios of folding to unfolding rates, are dominated by unfolding rather than folding rates (Table I). Why is this so? Folding may have a biologically imposed lower limit *in vivo*, such that it is sufficiently fast to avoid degradation or aggregation,[38] and an upper limit imposed by physical constraints even for the most topologically simple folds.[34] Conversely, while there may be a similar physical limit for fast unfolding, the biological limit for slow unfolding, which may be related to the need for eventual protein turnover,[39] may be more malleable due to the greatly differing roles and lifetimes of natural proteins.

Experimental and theoretical support for correlation between unfolding rates and native structure

Multiple lines of evidence suggest that unfolding rates should correlate with native structural complexity. First, while the relationships between unfolding rates and structure observed here may appear to be at odds with prior studies,[27–29] this is likely spurious due to trends being obscured previously when analyzing smaller datasets with substantial noise. Specifically, for a given value of a structural parameter (ACO or LRO), the variation in the observed unfolding rates is $\pm\sim$10-fold larger than that for the folding rates [Fig. 1(A–D)]. Thus, compared to folding rates, to detect significant correlations between unfolding rates and structural parameters, the absolute range of unfolding rates needs to be larger. The smaller datasets used in previous unfolding analyses,[27–29] which were based on the more curated set of Maxwell *et al.*,[30] had a range of $\sim$ eight orders of magnitude for the unfolding rates ($6 \times 10^{-6} - 1 \times 10^{2}$ s^{-1}). The larger dataset used here spans $\sim$16 orders of magnitude ($4 \times 10^{-11} - 5\times10^{5}$ s^{-1}), and as such the correlation between unfolding rate and structural complexity can be observed more clearly. Second, as the rates at the transition midpoint (where k_u is equal to k_f, and ΔG is 0) report on the transition state energy, the correlation of these rates with measures of structural complexity suggests that both the folding and unfolding rates (under conditions of different ΔG) should also be correlated with those same measures of structural complexity. Third, a recently developed method based on physical principles and protein structural class and size was able to predict both unfolding and folding rates for a set of 52 two-state folding proteins.[40] Finally, an analysis of 53 two-state and 19 multistate folders using a complex fractal parameter found comparable correlations with unfolding and folding rates, although the strength of the correlation was weaker than reported here.[24] The above considerations provide support for our observation of the significant correlations of structural complexity with both folding and unfolding rates.

Implications for design

The correlations reported herein indicate that the same measures of structural complexity predict both folding and unfolding rates equally well, and consequently, it may be difficult to modulate one aspect of the structure to alter (e.g., gain) folding speed, while leaving unfolding speed unaffected. Thus, it may seem a daunting task to achieve the desirable outcome of both fast folding and slow unfolding simultaneously. However, while the correlations of structural complexity and folding/unfolding rates have high statistical significance (Table I), there is nevertheless considerable variation around the lines of best fit, which we roughly estimate to be in the range of $\pm$ two orders of magnitude for folding and $\pm$ three orders of magnitude for unfolding rates [Fig. 1(A–D)]. Although some of this variation may be caused by more complex topological features such as nested structures,[25,35,41] it has been noted previously[14,42] and well documented experimentally, for example, by comparison of homologous proteins,[31–33,36,42] that while the native structure may place upper and lower boundaries on folding and unfolding rates, sequence-specific effects can be substantial. This is also illustrated by the effects, sometimes quite large, of point mutations on kinetics.[43,44] In addition, single mutations tend to have a larger

absolute effect on unfolding rather than folding rates based on our analysis of a dataset collected by Naganathan and Muñoz[45] where the change in unfolding rate is ~15-fold greater on average than for folding rates (Supporting Information Table SIII). Together, the above points suggest that while the scaffold may define broad ranges for folding/unfolding rates, sequence-specific engineering can provide substantial scope to modulate these rates in order to achieve to some extent, fast folding and slow unfolding.

Fortunately, much work has been done on the sequence-specific determinants of folding and unfolding rates, and some lessons may be learned from this. First, the nature of functional sites in proteins may modulate topological complexity and alter kinetics. This was studied for two beta-trefoil proteins: the functional myristoyl binding site of Hisactophilin is a cavity within the protein core which reduces structural complexity and so may speed folding and unfolding, whereas the binding site of Interleukin-1 beta is formed by two long loops which increase structural complexity and may slow the kinetics.[46] Second, both residual structure in the denatured state[47,48] and nonnative interactions in the transition state[49–51] can increase the folding rate independent of the overall topology. Third, unfolding rates can be slowed by introducing hydrophobic residues on the surface of the native structure, which may increase local rigidity and the barrier to local hydration,[52] or by surface electrostatic interactions, which may act as clips.[53] Lastly, it may be possible through computational simulation to identify particular weak points in the structure which, if strengthened, could increase unfolding cooperativity and therefore increase the height of the unfolding barrier.[54] Thus, multiple approaches may be used to modulate sequence-specific interactions in order to alter folding/unfolding rates.

Conclusions

We have shown, using a large dataset with highly significant correlations, that the measures of structural complexity that have emerged as strong predictors of folding rates, ACO and LRO, are equally predictive of unfolding rates, contrary to what has been reported previously when using smaller datasets.[27–29] In addition, the correlations are fairly robust to kinetic mechanism, whether two-state or multistate, and structural class, whether alpha, beta, or mixed. From a fundamental protein folding point of view, this suggests that the structural complexity reported on by ACO and LRO is a key determinant of both folding and unfolding processes.

These results have important implications for protein engineering and design. Specifically, a topologically simple scaffold may fold quickly, and so attain *in vivo* activity[34,38]; but, it will also unfold quickly, reducing kinetic stability and resistance to thermal or chemical denaturation and degradation by proteases.[55,56] Conversely, a topologically complex scaffold may possess the high kinetic stability that would be ideal for harsh industrial conditions or crowded and proteolytic biological environments,[56] but it may have difficulty folding fast enough to be biologically viable.[8,34] These conflicting kinetic constraints may limit the prospects when both fast folding and kinetic stability are required, or when improving a particular scaffold that is desirable for other reasons (such as function). However, these difficulties can be overcome as in existing proteins where the large variation in folding/unfolding rates and significant effects of point mutations[44] (Supporting Information Table SIII) demonstrate that appropriately designed sequences can ease constraints placed on the folding and unfolding rates by the structural complexity of the protein scaffold.

Methods

Our dataset is largely that of Garbuzynskiy *et al.*[34] (which includes the smaller dataset of Maxwell *et al.*[30]). We added data from the Kinetic DataBase[57] and other published kinetic data including our own[32,58–60] as well as data on our engineered Three-Foil protein[61] (for which the kinetic experiments will be published separately). In adding data, we followed the general criterion used by Garbuzynskiy *et al.*,[34] using only monomeric, single domain proteins, which lack disulfide bonds and prosthetic groups. In addition, the experimental temperature was in the range, or could be reliably extrapolated, to ~25°C, and the folding and unfolding rates were measured at, or could be extrapolated to, 0 *M* denaturant. The specific details of the sources for each member of the 108 protein dataset are given in Supporting Information Table SII.

ACO, the average sequence separation between contacting heavy atoms, was calculated as described by Ivankov *et al.*[7]

$$\mathrm{ACO} = \frac{1}{N_c}\sum_{i,j}^{N_c} |i-j|$$

where N_c is the total number of contacts between heavy atoms, and $|i-j|$ is the sequence separation in residues for a given contacting pair of atoms. Contacts are considered between heavy atoms less than 6 Å apart.

The formula for calculating LRO is that of Gromiha and Selvaraj[19]

$$\mathrm{LRO} = \frac{1}{L}\sum_{i,j}^{R_c} \boldsymbol{n}_{i,j}$$

where L is the protein length, R_c is the total number of contacting residues and $\boldsymbol{n}_{i,j}$ is 1 when $|i-j| \geq 12$ and 0 otherwise. We have modified the definition of a contact between residues to be the same as in

ACO rather than the original criterion of a Cα separation less than 8 Å. This modified form yields slightly improved correlations for both folding and unfolding; using 6 Å may correct for underestimation of long range contacts between large residues in cores.

Acknowledgment

The authors declare no conflicts of interest.

References

1. Daggett V, Fersht A (2003) The present view of the mechanism of protein folding. Nat Rev Mol Cell Biol 4: 497–502.
2. Sosnick TR, Barrick D (2011) The folding of single domain proteins–have we reached a consensus? Curr Opin Struct Biol 21:12–24.
3. Chan HS, Zhang Z, Wallin S, Liu Z (2011) Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. Annu Rev Phys Chem 62:301–326.
4. Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. Annu Rev Biophys 37:289–316.
5. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 48:545–600.
6. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 277:985–994.
7. Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV (2003) Contact order revisited: influence of protein size on the folding rate. Protein Sci 12:2057–2062.
8. Lane TJ, Pande VS (2013) Inferring the rate-length law of protein folding. PLoS One 8:e78606.
9. Gutin A, Abkevich V, Shakhnovich E (1996) Chain length scaling of protein folding time. Phys Rev Lett 77:5433–5436.
10. Zwanzig R, Szabo A, Bagchi B (1992) Levinthal's paradox. Proc Natl Acad Sci USA 89:20–22.
11. Li MS, Klimov DK, Thirumalai D (2002) Dependence of folding rates on protein length. J Phys Chem B 106: 8302–8305.
12. Wolynes PG (1997) Folding funnels and energy landscapes of larger proteins within the capillarity approximation. Proc Natl Acad Sci USA 94:6170–6175.
13. Naganathan AN, Muñoz V (2005) Scaling of folding times with protein size. J Am Chem Soc 127:480–481.
14. Istomin AY, Jacobs DJ, Livesay DR (2007) On the role of structural class of a protein with two-state folding kinetics in determining correlations between its size, topology, and folding rate. Protein Sci 16:2564–2569.
15. Rollins GC, Dill KA (2014) A general mechanism of 2-state protein-folding kinetics. J Am Chem Soc 136: 11420–11427.
16. Paci E, Lindorff-Larsen K, Dobson CM, Karplus M, Vendruscolo M (2005) Transition state contact orders correlate with protein folding rates. J Mol Biol 352: 495–500.
17. Faísca PFN, Travasso RDM, Parisi A, Rey A (2012) Why do protein folding rates correlate with metrics of native topology? PLoS One 7:e35599.
18. Sosnick TR (2008) Kinetic barriers and the role of topology in protein and RNA folding. Protein Sci 17: 1308–1318.
19. Gromiha MM, Selvaraj S (2001) Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. J Mol Biol 310:27–32.
20. Zou T, Ozkan SB (2011) Local and non-local native topologies reveal the underlying folding landscape of proteins. Phys Biol 8:066011.
21. Zhou H, Zhou Y (2002) Folding rate prediction using total contact distance. Biophys J 82:458–463.
22. Ouyang Z, Liang J (2008) Predicting protein folding rates from geometric contact and amino acid sequence. Protein Sci 17:1256–1263.
23. Micheletti C (2003) Prediction of folding rates and transition-state placement from native-state geometry. Proteins 51:74–84.
24. Tejera E, Machado A, Rebelo I, Nieto-Villar J (2009) Fractal protein structure revisited: topological, kinetic and thermodynamic relationships. Phys A Stat Mech Appl 388:4600–4608.
25. Rustad M, Ghosh K (2012) Why and how does native topology dictate the folding speed of a protein? J Chem Phys 137:205104.
26. Su JG, Li CH, Hao R, Chen WZ, Wang CX (2008) Protein unfolding behavior studied by elastic network model. Biophys J 94:4586–4596.
27. Jung J, Buglass AJ, Lee E-K (2010) Topological quantities determining the folding/unfolding rate of two-state folding proteins. J Solut Chem 39:943–958.
28. Jung J, Lee J, Moon H-T (2005) Topological determinants of protein unfolding rates. Proteins 58:389–395.
29. Harihar B, Selvaraj S (2011) Application of long-range order to predict unfolding rates of two-state proteins. Proteins 79:880–887.
30. Maxwell KL, Wildes D, Zarrine-Afsar A, De Los Rois MA, Brown AG, Friel CT, Hedberg L, Horng J-C, Bona D, Miller EJ, Vallée-Bélisle A, Main ERG, Bemporad F, Qiu L, Teilum K, Vu N-D, Edwards AM, Ruczinski I, Poulsen FM, Kragelund BB, Michnick SW, Chiti F, Bai Y, Hagen SJ, Serrano L, Oliveberg M, Raleigh DP, Wittung-Stafshede P, Radford SE, Jackson SE, Sosnick TR, Marqusee S, Davidson AR, Plaxco KW (2005) Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. Protein Sci 14:602–616.
31. Wensley BG, Gärtner M, Choo WX, Batey S, Clarke J (2009) Different members of a simple three-helix bundle protein family have very different folding rate constants and fold by different mechanisms. J Mol Biol 390:1074–1085.
32. Ferguson N, Sharpe TD, Schartau PJ, Sato S, Allen MD, Johnson CM, Rutherford TJ, Fersht AR (2005) Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. J Mol Biol 353:427–446.
33. Perl D, Welker C, Schindler T, Schröder K, Marahiel MA, Jaenicke R, Schmid FX (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. Nat Struct Biol 5:229–235.
34. Garbuzynskiy SO, Ivankov DN, Bogatyreva NS, Finkelstein AV (2013) Golden triangle for folding rates of globular proteins. Proc Natl Acad Sci USA 110:147–150.
35. Chavez LL, Onuchic JN, Clementi C (2004) Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. J Am Chem Soc 126:8426–8432.

36. Plaxco KW, Simons KT, Ruczinski I, Baker D (2000) Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. Biochemistry 39:11177–11183.
37. Gráczer E, Varga A, Hajdú I, Melnik B, Szilágyi A, Semisotnov G, Závodszky P, Vas M (2007) Rates of unfolding, rather than refolding, determine thermal stabilities of thermophilic, mesophilic, and psychrotrophic 3-isopropylmalate dehydrogenases. Biochemistry 46:11536–11549.
38. Dobson CM (2003) Protein folding and misfolding. Nature 426:884–890.
39. Matouschek A (2003) Protein unfolding–an important process in vivo? Curr Opin Struct Biol 13:98–109.
40. De Sancho D, Muñoz V (2011) Integrated prediction of protein folding and unfolding rates from only size and structural class. Phys Chem Chem Phys 13:17030–17043.
41. Fiebig KM, Dill KA (1993) Protein core assembly processes. J Chem Phys 98:3475.
42. Nickson AA, Clarke J (2010) What lessons can be learned from studying the folding of homologous proteins? Methods 52:38–50.
43. Jackson SE (1998) How do small single-domain proteins fold? Fold Des 3:R81–R91.
44. Lawrence C, Kuge J, Ahmad K, Plaxco KW (2010) Investigation of an anomalously accelerating substitution in the folding of a prototypical two-state protein. J Mol Biol 403:446–458.
45. Naganathan AN, Muñoz V (2010) Insights into protein folding mechanisms from large scale analysis of mutational effects. Proc Natl Acad Sci USA 107:8611–8616.
46. Gosavi S (2013) Understanding the folding-function tradeoff in proteins. PLoS One 8:e61222.
47. Yi Q, Scalley-Kim ML, Alm EJ, Baker D (2000) NMR characterization of residual structure in the denatured state of protein L. J Mol Biol 299:1341–1351.
48. Ratcliff K, Marqusee S (2010) Identification of residual structure in the unfolded state of ribonuclease H1 from the moderately thermophilic Chlorobium tepidum: comparison with thermophilic and mesophilic homologues. Biochemistry 49:5167–5175.
49. Zarrine-Afsar A, Wallin S, Neculai AM, Neudecker P, Howell PL, Davidson AR, Chan HS (2008) Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. Proc Natl Acad Sci USA 105:9999–10004.
50. Shental-Bechor D, Smith MTJ, Mackenzie D, Broom A, Marcovitz A, Ghashut F, Go C, Bralha F, Meiering EM, Levy Y (2012) Nonnative interactions regulate folding and switching of myristoylated protein. Proc Natl Acad Sci USA 109:17839–17844.
51. Clementi C, Plotkin SS (2004) The effects of nonnative interactions on protein folding rates: theory and simulation. Protein Sci 13:1750–1766.
52. Machius M, Declerck N, Huber R, Wiegand G (2003) Kinetic stabilization of Bacillus licheniformis alpha-amylase through introduction of hydrophobic residues at the surface. J Biol Chem 278:11546–11553.
53. Cavagnero S, Debe DA, Zhou ZH, Adams MW, Chan SI (1998) Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. Biochemistry 37:3369–3376.
54. Rodriguez-Larrea D, Minning S, Borchert TV, Sanchez-Ruiz JM (2006) Role of solvation barriers in protein kinetic stability. J Mol Biol 360:715–724.
55. Manning M, Colón W (2004) Structural basis of protein kinetic stability: resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward beta-sheet structure. Biochemistry 43:11248–11254.
56. Sanchez-Ruiz JM (2010) Protein kinetic stability. Biophys Chem 148:1–15.
57. Bogatyreva NS, Osypov AA, Ivankov DN (2009) KineticDB: a database of protein folding kinetics. Nucleic Acids Res 37:D342–D346.
58. Smith MTJ, Meissner J, Esmonde S, Wong HJ, Meiering EM (2010) Energetics and mechanisms of folding and flipping the myristoyl switch in the {beta}-trefoil protein, hisactophilin. Proc Natl Acad Sci USA 107:20952–20957.
59. Lee J, Blaber SI, Dubey VK, Blaber M (2011) A polypeptide "building block" for the β-trefoil fold identified by "top-down symmetric deconstruction." J Mol Biol 407:744–763.
60. Spector S, Raleigh DP (1999) Submillisecond folding of the peripheral subunit-binding domain. J Mol Biol 293: 763–768.
61. Broom A, Doxey AC, Lobsanov YD, Berthin LG, Rose DR, Howell PL, McConkey BJ, Meiering EM (2012) Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. Structure 20:161–171.