Evidence for the principle of minimal frustration in the evolution of protein folding landscapes

Franco O. Tzul[a,b], Daniel Vasilchuk[b,c], and George I. Makhatadze[a,b,c,1]

[a]Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180; [b]Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180; and [c]Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, Troy, NY 12180

Edited by José N. Onuchic, Rice University, Houston, TX, and approved January 24, 2017 (received for review August 19, 2016)

Theoretical and experimental studies have firmly established that protein folding can be described by a funneled energy landscape. This funneled energy landscape is the result of foldable protein sequences evolving following the principle of minimal frustration, which allows proteins to rapidly fold to their native biologically functional conformations. For a protein family with a given functional fold, the principle of minimal frustration suggests that, independent of sequence, all proteins within this family should fold with similar rates. However, depending on the optimal living temperature of the organism, proteins also need to modulate their thermodynamic stability. Consequently, the difference in thermodynamic stability should be primarily caused by differences in the unfolding rates. To test this hypothesis experimentally, we performed comprehensive thermodynamic and kinetic analyses of 15 different proteins from the thioredoxin family. Eight of these thioredoxins were extant proteins from psychrophilic, mesophilic, or thermophilic organisms. The other seven protein sequences were obtained using ancestral sequence reconstruction and can be dated back over 4 billion years. We found that all studied proteins fold with very similar rates but unfold with rates that differ up to three orders of magnitude. The unfolding rates correlate well with the thermodynamic stability of the proteins. Moreover, proteins that unfold slower are more resistant to proteolysis. These results provide direct experimental support to the principle of minimal frustration hypothesis.

protein folding | protein stability | protein evolution

The energy landscape theory provides a conceptual physicochemical framework for understanding protein folding. This theory is based on the principle of minimal frustration that "...quantifies the dominance of interactions stabilizing the specific native structure over other interactions that would favor nonnative, topologically distinct traps" (1). A consequence of this is that the folding energy landscape of naturally occurring proteins is funnel-shaped (1–22). The shape of this funnel depends on two main factors that can introduce frustration and roughness: topology and the extent of nonnative interactions. Topological frustration can occur when certain native interactions are formed too early and need to be undone to allow for other interactions to form first, leading to backtracking and/or cracking (23–28). Weak nonnative interactions can have complex effects on the folding landscape (29–31): small amounts of weak nonnative interactions can assist folding, whereas larger amounts can create internal friction that will slow folding (32–35). For a given protein fold, both topological and energetic frustrations will depend on the amino acid sequence. Theoretical studies have suggested that naturally occurring proteins have selected sequences that are compatible with the principle of minimal frustration (36).

The goal of this study is to experimentally test the evolutionary validity of the principle of minimal frustration by characterizing protein sequences that span a wide range of stability and evolutionary time. The principle of minimal frustration states that all naturally evolved proteins have optimized folding energy landscapes, with topological or energetic barriers that are often a consequence of forming the biologically functional folded state. We can assume that all proteins within a given functional fold evolved from a common ancestor. This common ancestor was presumably a well-folded protein and had robust folding rates. To acquire such properties, the energy landscape had to follow the principle of minimal frustration. Numerous lines of evidence suggest that the last universal common ancestor (LUCA) was a thermophilic organism (37). Correspondingly, the proteins in this organism were more thermostable than those from modern mesophilic organisms (38–43). During early evolution, different organisms had to colonize different environments that required their proteins to be more or less stable than those of LUCA. If they evolved in such a way that their folding became slower (less robust), this would suggest that the folding energy landscape is not minimally frustrated anymore, and it would violate the principle of minimal frustration. If the proteins evolved to have significantly faster folding rates, this would suggest that the LUCA was in violation of the principle of minimal frustration. Thus, proteins that belong to the same functional protein family are expected to have similar folding energy landscapes, which macroscopically should be manifested via similar folding rates. However, protein stability also needs to evolve to be compatible with the growth temperature of their host organisms. Thus, for a simple two-state system, the modulation of stability within a given family of proteins should occur via modulation of unfolding rates (i.e., more stable proteins should have slower unfolding rates, whereas folding rates remain largely independent of stability). Computational modeling, both on lattice (3) and off lattice (44–46), provides additional support for this hypothesis. Here, we experimentally test this hypothesis using functional homologs from the

Significance

A detailed understanding of how different sequences that fold into the same structure affect the folding energy landscape presents one of the current challenges of the protein folding field. The principle of minimal frustration suggests that naturally evolved proteins with the same structure should have similar folding rates and that modulation of thermodynamic stability should occur via unfolding rates. We experimentally tested this hypothesis using 15 different thioredoxins, with sequences either obtained from the extant organisms or resurrected using ancestral sequence reconstruction. We show that all of these proteins fold with similar rates and that their dramatic differences in stability are because of the differences in the unfolding rates.

Author contributions: F.O.T. and G.I.M. designed research; F.O.T. and D.V. performed research; D.V. contributed new reagents/analytic tools; F.O.T. and D.V. analyzed data; and F.O.T. and G.I.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: makhag@rpi.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613892114/-/DCSupplemental.

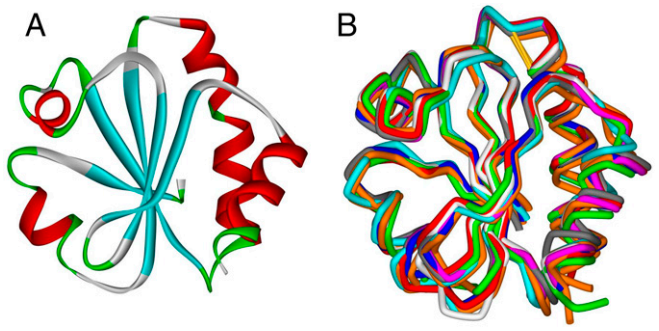

Fig. 1. Cartoon representation of the Trx protein fold. (*A*) Cartoon of the secondary structure topology of Trx fold. (*B*) Backbone alignment of 10 different structures shows conservation of the Trx fold, despite sequence identity as low as 20% (Table S4). Protein Data Bank ID codes of structures used in the alignment are 2CVK (TrxTT), 2TRX (TrxEC), 2E0Q (TrxST), 2YJ7 (TrxLPBCA), 2YN1 (TrxLGPCA), 2YNX (TrxLACA), 2YOI (TrxLECA), 2YPM (TrxLAFCA), 3ZIV (TrxAECA), and 4BA7 (TrxLBCA).

thioredoxin (Trx) protein family from extant organisms or those resurrected using ancestral sequence reconstruction (ASR) as our model system.

Results and Discussion

Model Protein Family. Trx is a class of essential small (12 kDa) redox proteins found in nearly all known organisms (47) and believed to be present in early life (37). They function as antioxidants by facilitating the reduction of other proteins through cysteine thiol-disulfide exchange. The α/β protein fold of Trx consists of five antiparallel β-sheets sandwiched by four α-helices and is highly conserved (Fig. 1). We studied extant Trx from eight different organisms, including those living under extreme environmental conditions: psychrophilic (optimal growth temperature from 5 to 10 °C) *Shewanella benthica* (TrxSB) and *Colwellia piezophilia* (TrxCP); mesophilic (optimal growth temperature of ∼37 °C) *Escherichia coli* (TrxEC) and *Homo sapiens* (TrxHS); and thermophilic (optimal growth temperature of >70 °C) *Thermus thermophilus* (TrxTT), *Pyrococcus yayanosii* (TrxPY), *Methanococcus jannaschii* (TrxMJ), and *Sulfolobus tokodaii* (TrxST).

Extant Trx Stabilities and Folding/Unfolding Kinetics. Previous studies of stability of Trxs indicated that these proteins are highly stable (41, 48). Indeed, even at pH 2, the transition temperatures range between 40 and 95 °C (Fig. 2*A*). This variation in the transition temperatures seems to correlate with the optimal growth temperatures of the corresponding organisms. For example, Trxs from thermophilic organisms, such as TrxTT, TrxPY, TrxMJ, and TrxST, have much higher Tm values than their mesophilic homologs TrxEC and TrxHS. The mesophilic proteins, in turn, have higher melting temperatures that the proteins from psychrophilic organisms TrxSB and TrxCP. This correlation of the growth temperature and protein stability is also evident from the denaturant-induced unfolding studies. Fig. 2*B* shows urea-induced unfolding profiles for the same set of proteins. The midpoint for the urea-induced unfolding transitions ranges from 1.5 to 7 M. For one protein, thermophilic TrxST, the midpoint is probably even higher, because no indication of unfolding was observed at 10 M urea.

How are the differences in thermodynamic stabilities manifested in the kinetic stabilities? Changes in thermodynamic stability can be caused by the differences in the folding rates, unfolding rates, or both. We thus performed characterization of the folding and unfolding of extant Trx variants using standard stopped flow methods. Fig. 2*C* shows the experimentally measured folding and unfolding rate, k_{obs}, in the form of chevrons for the extant Trx variants. Each data point on the chevron is an average of at least five independently measured traces, and their SDs are reported as the error bars. It is evident from these chevron plots that there is minimal variation in the rates of Trx folding. They all fold in the absence of denaturant in 80 ± 60 ms, and the difference between the fastest (20 ms) and the slowest (140 ms) is only sevenfold.

More notably, inspection of the chevron plots shown in Fig. 2*D* reveals that the variation in the unfolding rates is dramatic. Extrapolation to a zero denaturant concentration gives unfolding rate constants in the range between 7 and 22,000 s [6 d; i.e., ∼3,000-fold difference between the fastest (TrxCP) and the slowest (TrxPY) unfolding Trx proteins]. Importantly, the unfolding rates correlate with stability, whereas folding rates show little correlation with stability within the extant Trx protein family (Fig. 3). This observation suggests that the differences in stabilities of extant Trx

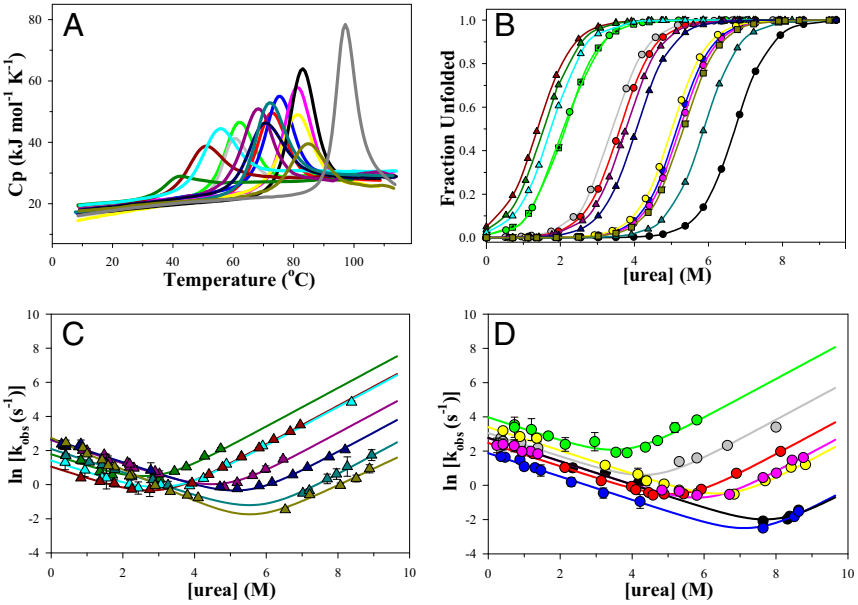

Fig. 2. Trx protein family has diverse thermodynamic and kinetic properties. (*A*) Comparison of the thermal stabilities of Trx proteins at pH 2.0 as monitored by DSC. Cp, partial molar heat capacity. (*B*) Comparison of the stabilities of Trx proteins against urea-induced unfolding at pH 2.0. Experimental points are shown as symbols, whereas the fit to a linear extrapolation model is shown by solid lines. (*C*) Chevron plots of lnk_{obs} vs. urea concentration of extant Trx variants. (*D*) Chevron plots of lnk_{obs} vs. urea concentration of ancestral Trx variants. Solid lines in *C* and *D* are the results of the global fit to Eq. **2**. Symbols and colors are as follows: brown triangle, TrxSB; green triangle (only in *A*), TrxCP; turquoise-blue triangle, TrxEC; purple triangle, TrxHS; dark-blue triangle, TrxTT; teal-green triangle, TrxPY; greenish-gray triangle, TrxMJ; gray triangle, TrxST; light-green circle, TrxLAFCA; gray circle, TrxLECA; red circle, TrxLBCA; blue circle, TrxLGPCA; pink circle, TrxAECA; yellow circle, TrxLPBCA; black circle, TrxLACA.

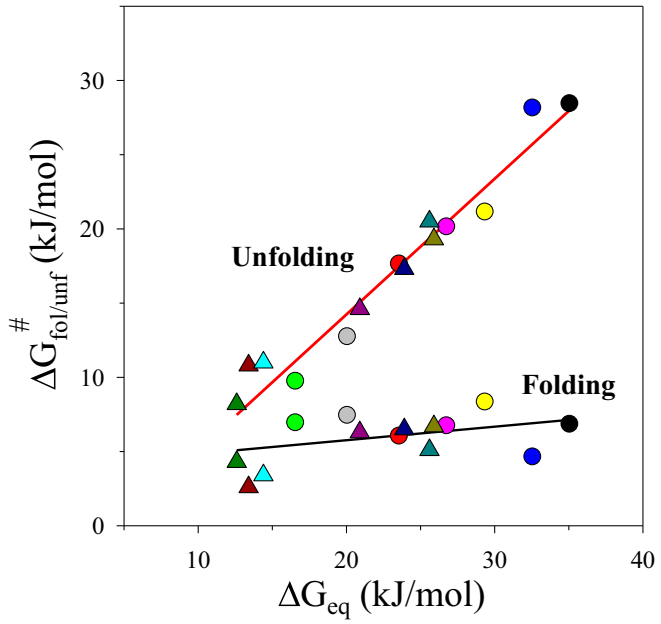

Fig. 3. Thermodynamic stability of Trx variants correlates with the unfolding rates and does not correlate with the folding rates. Equilibrium stabilities, ΔG_{eq}, are plotted vs. folding ($\Delta G^{\#}_{fol}$; black regression line) and unfolding ($\Delta G^{\#}_{unf}$; red regression line) barriers. The folding/unfolding rates are extrapolated to zero denaturant concentration using data for extant and ancestral Trx variants shown in Fig. 2 *C* and *D*. Symbol/color coding for different proteins is the same as in Fig. 2.

are largely caused by the differences in the unfolding rates of these proteins.

Ancestral Trx Stabilities and Folding/Unfolding Kinetics. Similarity of the folding rates among extant Trx proteins is not an unexpected finding, because we are assessing the kinetics of a specific fold, and there are several reports correlating various topology metrics to folding kinetic rates (49–52). Moreover, the observed robust folding rates suggest that the Trx energy landscapes are unfrustrated (3, 5, 29, 36, 53, 54). This observation, in turn, implies that the folding of extant proteins of the Trx family obeys the principle of minimal frustration and that variations in stability are caused by the unfolding rates. Was this part of the evolutionary considerations?

To answer this question, we turned our attention to the analysis of the thermodynamics and kinetics of the Trx proteins for which sequences were obtained from ASR (41, 55). ASR is a bioinformatics technique that allows sequence extrapolation from extant protein sequences back to their common ancestors. It has been used to resurrect ancestral nodes for a number of proteins and shown to produce catalytically and/or functionally active proteins (40–42, 56–58). The sequences for the ancestral Trx proteins were obtained using maximum likelihood sequence reconstruction, targeting several Precambrian nodes during Trx evolution (41). The seven sequences that were studied here are last animal and fungi common ancestor (TrxLAFCA), last eukaryotic common ancestor (TrxLECA), last bacterial common ancestor (TrxLBCA), last γ-proteobacteria common ancestor (TrxLGPCA), archaea/eukaryota common ancestor (TrxAECA), last cyanobacterial, deinococcus and thermus common ancestor (TrxLPBCA), and last archaeal common ancestor (TrxLACA). These sequences were dated back to over 4 billion years, thus enabling us to probe the thermodynamics and kinetics of folding/unfolding for proteins with sequences that arguably represent a wide evolutionary timespan (41). Furthermore, the crystal structures of these proteins have been determined and support the high conservation of the Trx protein fold (55).

Thermodynamic stabilities of the ancestral Trx proteins at pH 2 are compared in Fig. 2*A*. As with the extant Trx proteins, ancestral Trx variants showed significant variation in their transition temperatures. Proteins that are dated earlier on the evolutionary tree (e.g., LACA and AECA) have higher T_m values than the more recent proteins, such as LECA and LAFCA, which in turn, are more stable than TrxEC. Equilibrium urea-induced unfolding profiles also support this correlation (Fig. 2*B*). The differences in transition temperatures at pH 2 agree well with the T_m values obtained at pH 7 (Fig. S1), indicating a generality of the stability trend (41).

The results of folding/unfolding experiments for ancestral Trx variants measured using standard stopped flow methods are presented in the form of chevron plots in Fig. 2*D*. Again, as in the case of the extant Trx proteins, the folding rates of ancestral Trx are very similar, whereas the unfolding rates vary dramatically. This observation might suggest that the principle of minimal frustration played a role in the sequence selection early on in evolution.

Both ancestral and extant Trx proteins maintain strong correlation of increasing apparent unfolding barrier (i.e., decrease in the unfolding rate) with increase in thermodynamic stability, whereas the apparent folding barrier (as defined by the folding rates) remains independent of the thermodynamic stability. The thermal and kinetic stabilization conferred to the oldest ancestral variants is comparable with the most stable extant Trx proteins (e.g., TrxPY and TrxMJ).

The observation that both extant and ancestral Trx proteins fold with similar rates directly supports the assumption that

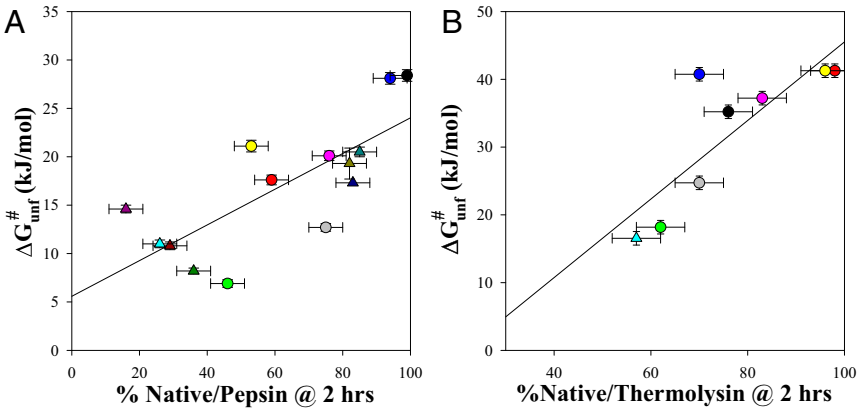

Fig. 4. Unfolding rates of Trx variants correlate with the resistance to proteolysis to pepsin (pH 2.0) (*A*) and thermolysin (pH 7.0) (*B*). Symbol/color coding for different proteins is the same as in Fig. 2.

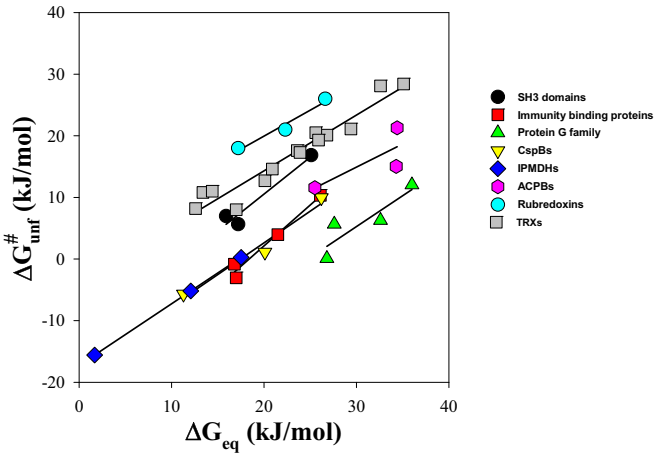

Fig. 5. Correlation between thermodynamic stabilities and unfolding rates in eight different protein families: black circle, SH3 domains (78–80); red square, immunity binding domains (81, 82); light-green triangle, protein G (83, 84); yellow inverted diamond, cold shock proteins CspB (85); blue diamond, isopropyl malate dehydrogenases IPMDH (86); pink hexagon, acyl-CoA binding proteins ACBP (87); light-blue circle, rubredoxins (88); gray square, Trxs (this work). In all cases, there is a linear dependence (average slope of 1.00 ± 0.19; maximum of 1.33; minimum of 0.75) with statistically significant correlations (R^2 = 0.95 ± 0.08). Table S5 has details.

protein sequences have evolved with minimal energetic frustration so as to rapidly fold into their stable 3D states (36). It is also clear that the differences in thermodynamic stabilities are primarily caused by differences in unfolding rates for this protein family.

Correlation of Kinetic Stability with Resistance to Proteolysis. The variation in the unfolding rate defines kinetic (unfolding rate) stability of proteins, and it was argued that one of the important consequences of the kinetic stability is to protect proteins from proteolytic degradation (59–61). To probe the correlation of the kinetic stability and resistance to proteolysis, we treated Trx samples with pepsin and thermolysin (62). Both proteases have very broad sequence specificity but very different pH activity profiles: pepsin has maximal activity at acidic pH (pH 2.0), whereas thermolysin is maximally active at neutral pH (pH 7.0). We find that both pepsin and thermolysin resistance directly correlate with the unfolding rates at pH 2.0 (Fig. 4*A*). Furthermore, the thermolysin resistance directly correlates with the unfolding rates measured previously for a subset of Trx variants at pH 7.0 (Fig. 4*B*). This correlation holds for both extant and ancestral Trx variants, supporting the notion that, independent of the pH, high kinetic stability indeed confers proteins with greater resistance to proteolytic digestion.

Concluding Remarks

We have characterized the thermodynamics and kinetics of the Trx functional fold using extant proteins from organisms inhabiting a variety of living temperatures. We also compared the thermodynamic and kinetic properties of eight extant proteins with seven Trx proteins obtained through ASR with an evolutionary timespan of 4 billion years. The main goal was to use this model protein family for studying the evolution of protein folding and unfolding rates. We have shown that, despite a dramatic difference in the thermodynamic stabilities, Trx proteins fold with a very similar rate, and the variation in thermodynamic stability is largely defined by the unfolding rates (i.e., the more thermodynamically stable proteins are kinetically more stable as well). The variation in kinetic stability correlates with the resistance against proteolytic degradation. The observation that differences in thermodynamic stability are largely defined by the differences in unfolding rates based on 15 different proteins from the Trx family seems to be specific not only to this protein family. Carstensen et al. (43) compared stabilities and folding kinetics of modern (βα)8-barrel protein HisF with its designed ancestral homolog Sym1. The authors found that Sym1 shows the same folding mechanism as HisF and that the higher thermodynamic stability of Sym1 is caused by major changes in the unfolding rates (43). Furthermore, compilation of literature data for other protein families, although not as extensive as this study and limited to extant proteins, shows very similar trends (Fig. 5) (i.e., the stability within a given protein structural family is largely modulated by the unfolding rates). This finding can be mapped to the funneled landscape, in which the transition state for a given protein fold remains minimally frustrated and modulation of the thermodynamic stability is achieved primarily through the optimization of the interactions in the native state. The illustration of the likely scenario of such effect, mapped onto a hypothetical folding energy landscape, is shown in Fig. 6.

Finally, it remains to be seen if other ancestrally reconstructed proteins maintain similar kinetic properties. If they do, ASR may be an efficient way to engineer kinetically stable enzymes.

Methods

Protein Expression, Purification, and Characterization. The gene sequences of extant Trx proteins were codon-optimized and synthesized (Blue Heron Biotechnology Inc.) for expression in *E. coli*. The DNA sequences of ancestral Trx proteins were optimized in a similar manner as described previously (41). Sequences for the extant and ancestral Trxs are given in Table S1. His_6 tags were engineered at the N terminus of all protein sequences.

Plasmids containing genes for the extant and ancestral proteins were transformed into *E. coli* BL21 (DE3), and cells were grown at 37 °C. The expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside when the cell density reached ∼0.8 optical units at 600 nm. Cells were harvested after 6 h of induction. All proteins were purified to homogeneity under native conditions using nickel- nitrilotriacetic acid (Novagen; Merck KGaA) affinity resin followed by size exclusion column chromatography according to previously published protocols (63–65).

Purities and identities of the recombinant proteins were confirmed by SDS gels and MALDI-TOF MS as previously described (63). In all cases, a single major peak was observed with a mass within 2–5 Da of that expected on the basis of the amino acid sequence (Table S2). Protein concentrations were determined spectrophotometrically using the molar extinction coefficients listed in Table S2.

The oligomerization state of all Trxs was characterized via analytical ultracentrifugation (AUC). Sedimentation equilibrium experiments were performed on a Beckman XLA analytical ultracentrifuge at pH values 3.0 (30 mM glycine-HCl), 5.5, and 7.0 (30 mM sodium cacodylate). Absorbance was monitored at 280 nm using either short- or long-column cells, and samples were allowed to equilibrate at three different rotor speeds at 20 °C. Global

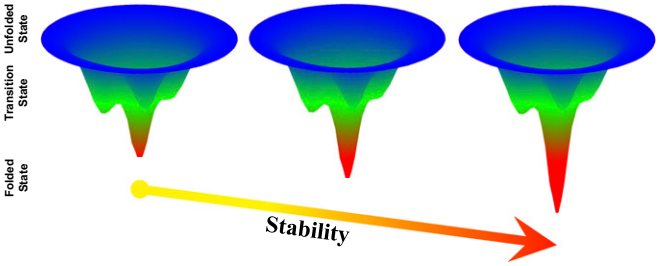

Fig. 6. Hypothetical folding energy landscape illustrating increase in stability (from left to right) because of a decrease of the unfolding rates, while folding rates remain the same. These changes in the energy landscape are consistent with the observed experimental data for the Trx protein fold (i.e., protein stability increase is caused by the increase in the energy barrier between transition-state and folded ensembles, whereas the energy barrier between unfolded and transition-state ensembles remained largely unchanged).

analysis of the centrifugation profiles was done as previously described (66). Table S3 shows molecular masses obtained from AUC experiments.

Differential Scanning Calorimetry Measurements. Differential scanning calorimetry (DSC) experiments were performed on a VP-DSC instrument (Microcal/GE-Healthcare) at a scan rate of 1.5°/min using protein concentrations of about 1.0 mg/mL. Buffers used were 30 mM glycine-HCl for pH values 2–3.5 experiments and 30 mM sodium-cacodylate for pH 7.0 experiments (67). Proteins were extensively dialyzed against the corresponding buffer. Raw DSC data were analyzed using the Origin-DSC (OriginLab) software package. Global fitting of the heat capacity profiles was done using in-house scripts using the nonlinear regression (NLREG) routine (68, 69).

Equilibrium Stability Measurements. Initially, all Trx protein stocks were extensively dialyzed in acidified water and diluted 20-fold into 30 mM glycine-HCl or buffered urea solutions adjusted to pH 2. Dilutions were done to final protein concentrations of 4–10 μM depending on their fluorophore quantum yields. Stock protein samples were diluted into buffered urea with concentrations incrementing up to 9.5 M and incubated at room temperature for 24 h. Tryptophan emission spectra were collected as a function of urea concentration using a fluorescence plate reader (TECAN Infinite M1000 Pro) in a 384-well black plate. Plate and samples were thermostated in the instrument at room temperature (23 °C ± 1 °C) for 5 min before data collection. Excitation was done at 295 nm with a gain of 150. Fluorescence emission spectra were recorded from 305 to 450 nm in 1-nm increments. TrxMJ does not have a tryptophan; therefore, tyrosine emission fluorescence was used instead. Trx proteins that did not produce a significant intensity change were additionally characterized via far-UV CD.

Urea unfolding profiles (intensity at 355 nm vs. urea concentration) were fitted globally for all Trx variants according to the linear extrapolation model (70, 71) using in-house written NLREG scripts as previously described (72).

Kinetic Stopped Flow Experiments. All Trx protein samples were prepared as described above. Data for chevron plots were collected by standard stopped flow methods on a JASCO J-815 spectropolarimeter operating in fluorescence mode that was equipped with an SFM 300 mixing module (BioLogic Science Instruments) containing a high density (HDS) mixer and a 30-μL FC-15 observation cuvette (73). Buffer containing 30 mM glycine-HCl adjusted to pH 2 was used for all kinetic experiments. Folding and unfolding reactions were initiated by a 20-fold dilution of protein-buffered stock solutions (80–200 μM) with the desired buffered urea concentration. Because of slow kinetics of unfolding, equilibration in buffered urea was required before refolding experiments could be successfully initiated. The reagent syringes, mixing chamber, and observation cuvette were thermostated at 20 °C using a circulating water bath. Fluorescence emission intensity from an N-WG 320-nm (N-WG 305 nm for TrxMJ) cutoff filter (BioLogic Science Instruments) was collected after excitation at 295 nm (280 nm for TrxMJ) through a 3-nm slit from a mercury lamp source. Voltages applied to the photomultiplier tube were set constant based on the fluorescence signal intensity at maximum amplitude (~920 V).

The Trx model system required multiexponential fits with a slow phase related to proline isomerization (74–76). Thus, the pertinent fastest folding and unfolding phases are reported. Rate constants, k_{obs}, were obtained using stretched exponential as previously described (73):

$$I(t) = pb \cdot t + y_o + A \cdot \exp(\pm k_{obs} \cdot t), \quad [1]$$

where $I(t)$ is fluorescence intensity as a function of time, y_o is the initial fluorescence intensity, A is the amplitude of the change between initial and final fluorescence intensities, k_{obs} is the observed kinetic rate constant associated with the fluorescence intensity relaxation, and pb is the sloping baseline correction for the photobleaching effect used in the Bio-Kine32 software. All traces were corrected for instrumental dead time (6 ms) before fitting, because some Trx protein variants showed significantly fast k_{obs}.

The natural logarithms of the k_{obs} are plotted as a function of urea concentration in the form of chevron plots. Each data point on the chevron plot is an average of five individual traces, and errors are taken as the SDs. Extrapolated $k_f(H_2O)$ and $k_u(H_2O)$ values were obtained by globally fitting the chevrons of all Trx variants to Eq. **2** below (77) using in-house scripts for the NLREG data-fitting package:

$$\ln(k_{obs}) = \ln\left[\exp\left(\ln(k_f(H_2O)) + \frac{m_f \cdot [urea]}{RT}\right) + \exp\left(\ln(k_u(H_2O)) + \frac{m_u \cdot [urea]}{RT}\right)\right], \quad [2]$$

where $k_f(H_2O)$ and $k_u(H_2O)$ are the folding and unfolding rates in the absence of denaturant, respectively, and m_f and m_u are the kinetic folding and unfolding m values (measured in kilojoules per mole per molar), respectively (77). The apparent folding/unfolding barriers have been calculated as $\Delta G^{\#}_{fol/unf} = RTln(k_{f/u})$, where R is the universal gas constant (8.184 J/mol K), and T is the temperature (Kelvin).

Proteolytic Digestion. Extant and resurrected Trx variants were subjected to proteolytic digestion to assess their kinetic stabilities at two pH values using appropriate proteases. For the acidic range, pepsin (Sigma-Aldrich) digestion was done in 50 mM sodium phosphate, pH 2.0. Proteolysis were carried out at 37 ± 1 °C, with a protein/enzyme mass ratio of 10:1 and a final protein concentration of 0.5 mg/mL. At the desired digestion lapse time, an aliquot was removed and quenched with 50% (wt/vol) ammonium hydroxide solution to a final concentration of 1%.

For the neutral range, thermolysin (Sigma-Aldrich) digestion was done in 50 mM sodium phosphate and 0.5 mM calcium chloride, pH 7 buffer. Proteolysis was carried out at 50 ± 1 °C, with a protein/enzyme mass ratio of 10:1 and a final protein concentration of 0.5 mg/mL. At the desired digestion lapse time, an aliquot was removed and quenched with 10% (vol/vol) formic acid to a final concentration of 0.5%.

Protein aliquots were immediately flash frozen in liquid nitrogen and stored at –20 °C until SDS/PAGE gels were ran. Samples were diluted in a 1:1 ratio with 2× bromophenol dye containing β-mercaptoethanol and resolved on precast 4–20% (wt/vol) Precise Protein SDS/PAGE gels (Thermo Scientific). Gels were stained with Coomassie Blue, and subsequently, they were destained and scanned at 1,200-dots per inch resolution using a flatbed scanner. Band intensities were analyzed with ImageJ software. Because some of the proteins were highly resistant to proteolysis, even at elevated temperatures, the $t_{1/2}$ could not be determined. Thus, we based our analysis on the fraction of the native band remaining in solution after 2 h.

ACKNOWLEDGMENTS. We thank Jose M. Sanchez-Ruiz for providing the plasmids for ancestral Trxs and Catherine Royer for the usage of the high-throughput fluorescence plate reader. This work was supported by US National Science Foundation Grant 1506468/1330249.

1. Wolynes PG, Eaton WA, Fersht AR (2012) Chemical physics of protein folding. *Proc Natl Acad Sci USA* 109(44):17770–17771.
2. Kosuri P, et al. (2012) Protein folding drives disulfide formation. *Cell* 151(4): 794–806.
3. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21(3):167–195.
4. 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(5):937–953.
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. Portman JJ, Takada S, Wolynes PG (1998) Variational theory for site resolved protein folding free energy surfaces. *Phys Rev Lett* 81(23):5237–5240.
7. Kouza M, Li MS, O'brien EP, Jr, Hu CK, Thirumalai D (2006) Effect of finite size on cooperativity and rates of protein folding. *J Phys Chem A* 110(2):671–676.
8. Campos LA, et al. (2011) A photoprotection strategy for microsecond-resolution single-molecule fluorescence spectroscopy. *Nat Methods* 8(2):143–146.
9. O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D (2008) Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc Natl Acad Sci USA* 105(36):13403–13408.
10. Cheung MS, Thirumalai D (2007) Effects of crowding and confinement on the structures of the transition state ensemble in proteins. *J Phys Chem B* 111(28): 8250–8257.
11. Cheung MS, Klimov D, Thirumalai D (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci USA* 102(13):4753–4758.
12. Klimov DK, Thirumalai D (1996) Factors governing the foldability of proteins. *Proteins* 26(4):411–441.
13. Thirumalai D, Klimov DK (1999) Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr Opin Struct Biol* 9(2):197–207.
14. Shea JE, Onuchic JN, Brooks CL, 3rd (2002) Probing the folding free energy landscape of the Src-SH3 protein domain. *Proc Natl Acad Sci USA* 99(25):16064–16068.
15. Shea JE, Brooks CL, 3rd (2001) From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem* 52:499–535.
16. Shea JE, Onuchic JN, Brooks CL, 3rd (1999) Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A. *Proc Natl Acad Sci USA* 96(22):12512–12517.
17. Shimada J, Kussell EL, Shakhnovich EI (2001) The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *J Mol Biol* 308(1):79–95.
18. Shimada J, Shakhnovich EI (2002) The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc Natl Acad Sci USA* 99(17):11175–11180.

19. Veitshans T, Klimov D, Thirumalai D (1997) Protein folding kinetics: Timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold Des* 2(1):1–22.
20. Clementi C, García AE, Onuchic JN (2003) Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: All-atom representation study of protein L. *J Mol Biol* 326(3):933–954.
21. Noel JK, Whitford PC, Onuchic JN (2012) The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *J Phys Chem B* 116(29):8692–8702.
22. Whitford PC, et al. (2009) An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins* 75(2):430–441.
23. Tripathi S, Makhatadze GI, Garcia AE (2013) Backtracking due to residual structure in the unfolded state changes the folding of the third fibronectin type III domain from tenascin-C. *J Phys Chem B* 117(3):800–810.
24. Capraro DT, Roy M, Onuchic JN, Jennings PA (2008) Backtracking on the folding landscape of the beta-trefoil protein interleukin-1beta? *Proc Natl Acad Sci USA* 105(39):14844–14848.
25. Chavez LL, Gosavi S, Jennings PA, Onuchic JN (2006) Multiple routes lead to the native state in the energy landscape of the beta-trefoil family. *Proc Natl Acad Sci USA* 103(27):10254–10258.
26. Hills RD, Jr, Brooks CL, 3rd (2008) Subdomain competition, cooperativity, and topological frustration in the folding of CheY. *J Mol Biol* 382(2):485–495.
27. Sułkowska JI, Sułkowski P, Onuchic JN (2009) Jamming proteins with slipknots and their free energy landscape. *Phys Rev Lett* 103(26):268103.
28. Whitford PC, Miyashita O, Levy Y, Onuchic JN (2007) Conformational transitions of adenylate kinase: Switching by cracking. *J Mol Biol* 366(5):1661–1671.
29. 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(27):8426–8432.
30. Zarrine-Afsar A, et al. (2008) Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc Natl Acad Sci USA* 105(29):9999–10004.
31. Clementi C, Plotkin SS (2004) The effects of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci* 13(7):1750–1766.
32. Cellmer T, Henry ER, Hofrichter J, Eaton WA (2008) Measuring internal friction of an ultrafast-folding protein. *Proc Natl Acad Sci USA* 105(47):18320–18325.
33. Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA (1992) The role of solvent viscosity in the dynamics of protein conformational changes. *Science* 256(5065):1796–1798.
34. Borgia A, et al. (2012) Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nat Commun* 3:1195.
35. Buscaglia M, Lapidus LJ, Eaton WA, Hofrichter J (2006) Effects of denaturants on the dynamics of loop formation in polypeptides. *Biophys J* 91(1):276–288.
36. Wolynes PG (2015) Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie* 119:218–230.
37. Weiss MC, et al. (2016) The physiology and habitat of the last universal common ancestor. *Nat Microbiol* 1(9):16116.
38. Gaucher EA, Govindarajan S, Ganesh OK (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451(7179):704–707.
39. Akanuma S, et al. (2013) Experimental evidence for the thermophilicity of ancestral life. *Proc Natl Acad Sci USA* 110(27):11067–11072.
40. Hobbs JK, et al. (2012) On the origin and evolution of thermophily: Reconstruction of functional precambrian enzymes from ancestors of Bacillus. *Mol Biol Evol* 29(2):825–835.
41. Perez-Jimenez R, et al. (2011) Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat Struct Mol Biol* 18(5):592–596.
42. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. *J Am Chem Soc* 135(8):2899–2902.
43. Carstensen L, et al. (2012) Conservation of the folding mechanism between designed primordial (βα)8-barrel proteins and their modern descendant. *J Am Chem Soc* 134(30):12786–12791.
44. Cieplak M, Xuan Hoang T (2000) Scaling of folding properties in go models of proteins. *J Biol Phys* 26(4):273–294.
45. Cieplak M, Hoang TX (2002) The range of the contact interactions and the kinetics of the Go models of proteins. *Int J Mod Phys C* 13(9):1231–1242.
46. Cieplak M, Hoang TX (2003) Universality classes in folding times of proteins. *Biophys J* 84(1):475–488.
47. Holmgren A (1985) Thioredoxin. *Annu Rev Biochem* 54:237–271.
48. Ladbury JE, Kishore N, Hellinga HW, Wynn R, Sturtevant JM (1994) Thermodynamic effects of reduction of the active-site disulfide of *Escherichia coli* thioredoxin explored by differential scanning calorimetry. *Biochemistry* 33(12):3688–3692.
49. 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(37):11177–11183.
50. Ivankov DN, et al. (2003) Contact order revisited: Influence of protein size on the folding rate. *Protein Sci* 12(9):2057–2062.
51. Jung J, Buglass AJ, Lee EK (2010) Topological quantities determining the folding/unfolding rate of two-state folding proteins. *J Solution Chem* 39(7):943–958.
52. Harihar B, Selvaraj S (2011) Application of long-range order to predict unfolding rates of two-state proteins. *Proteins* 79(3):880–887.
53. Onuchic JN, Nymeyer H, García AE, Chahine J, Socci ND (2000) The energy landscape theory of protein folding: Insights into folding mechanisms and scenarios. *Adv Protein Chem* 53:87–152.
54. Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14(1):70–75.
55. Ingles-Prieto A, et al. (2013) Conservation of protein structure over four billion years. *Structure* 21(9):1690–1697.
56. Wilson C, et al. (2015) Kinase dynamics. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* 347(6224):882–886.
57. Wheeler LC, Lim SA, Marqusee S, Harms MJ (2016) The thermostability and specificity of ancient proteins. *Curr Opin Struct Biol* 38:37–43.
58. Gaucher EA, Thomson JM, Burgan MF, Benner SA (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* 425(6955):285–288.
59. Romero-Romero ML, et al. (2016) Selection for protein kinetic stability connects denaturation temperatures to organismal temperatures and provides clues to archaean life. *PLoS One* 11(6):e0156657.
60. Park C, Zhou S, Gilmore J, Marqusee S (2007) Energetics-based protein profiling on a proteomic scale: Identification of proteins resistant to proteolysis. *J Mol Biol* 368(5):1426–1437.
61. Broom A, et al. (2015) Designed protein reveals structural determinants of extreme kinetic stability. *Proc Natl Acad Sci USA* 112(47):14605–14610.
62. Park C, Marqusee S (2005) Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. *Nat Methods* 2(3):207–212.
63. Strickler SS, et al. (2006) Protein stability and surface electrostatics: A charged relationship. *Biochemistry* 45(9):2761–2766.
64. Gvritishvili AG, Gribenko AV, Makhatadze GI (2008) Cooperativity of complex salt bridges. *Protein Sci* 17(7):1285–1290.
65. Gribenko AV, et al. (2009) Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proc Natl Acad Sci USA* 106(8):2601–2606.
66. Gribenko AV, Hopper JE, Makhatadze GI (2001) Molecular characterization and tissue distribution of a novel member of the S100 family of EF-hand proteins. *Biochemistry* 40(51):15538–15548.
67. Vasilchuk D, Pandharipande PP, Suladze S, Sanchez-Ruiz JM, Makhatadze GI (2014) Molecular determinants of expansivity of native globular proteins: A pressure perturbation calorimetry study. *J Phys Chem B* 118(23):6117–6122.
68. Lopez MM, Makhatadze GI (2002) Differential scanning calorimetry. *Methods Mol Biol* 173:113–119.
69. Sherrod PH (2005) Non Linear Regression Analysis Program (Nashville, TN). Available at www.nlreg.com/NLREG.pdf. Accessed February 8, 2017.
70. Pace CN (1990) Measuring and increasing protein stability. *Trends Biotechnol* 8(4):93–98.
71. Santoro MM, Bolen DW (1992) A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry* 31(20):4901–4907.
72. Loladze VV, Ibarra-Molero B, Sanchez-Ruiz JM, Makhatadze GI (1999) Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry* 38(50):16419–16423.
73. Wafer LN, Tzul FO, Pandharipande PP, Makhatadze GI (2013) Novel interactions of the TRTK12 peptide with S100 protein family members: Specificity and thermodynamic characterization. *Biochemistry* 52(34):5844–5856.
74. Wilson J, Kelley RF, Shalongo W, Lowery D, Stellwagen E (1986) Equilibrium and kinetic measurements of the conformational transition of thioredoxin in urea. *Biochemistry* 25(23):7560–7566.
75. Georgescu RE, Li JH, Goldberg ME, Tasayco ML, Chaffotte AF (1998) Proline isomerization-independent accumulation of an early intermediate and heterogeneity of the folding pathways of a mixed alpha/beta protein, *Escherichia coli* thioredoxin. *Biochemistry* 37(28):10286–10297.
76. Godoy-Ruiz R, et al. (2006) Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencies of amino acid occurrences in sequence alignments. *J Mol Biol* 362(5):966–978.
77. Maxwell KL, et al. (2005) Protein folding: Defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* 14(3):602–616.
78. Grantcharova VP, Baker D (1997) Folding dynamics of the src SH3 domain. *Biochemistry* 36(50):15685–15692.
79. Martinez JC, Pisabarro MT, Serrano L (1998) Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol* 5(8):721–729.
80. Plaxco KW, et al. (1998) The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry* 37(8):2529–2537.
81. Ferguson N, Capaldi AP, James R, Kleanthous C, Radford SE (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* 286(5):1597–1608.
82. Ferguson N, Li W, Capaldi AP, Kleanthous C, Radford SE (2001) Using chimeric immunity proteins to explore the energy landscape for alpha-helical protein folding. *J Mol Biol* 307(1):393–405.
83. McCallister EL, Alm E, Baker D (2000) Critical role of beta-hairpin formation in protein G folding. *Nat Struct Biol* 7(8):669–673.
84. Skinner JJ, et al. (2014) Benchmarking all-atom simulations using hydrogen exchange. *Proc Natl Acad Sci USA* 111(45):15975–15980.
85. Perl D, et al. (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Biol* 5(3):229–235.
86. Gráczer E, et al. (2007) Rates of unfolding, rather than refolding, determine thermal stabilities of thermophilic, mesophilic, and psychrotrophic 3-isopropylmalate dehydrogenases. *Biochemistry* 46(41):11536–11549.
87. Kragelund BB, et al. (1996) Fast and one-step folding of closely and distantly related homologous proteins of a four-helix bundle family. *J Mol Biol* 256(1):187–200.
88. 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(10):3369–3376.