Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding

Jin Wang[a,b,c,1], Ronaldo J. Oliveira[d,e,2], Xiakun Chu[a,b,2], Paul C. Whitford[f], Jorge Chahine[d], Wei Han[b], Erkang Wang[a,1], José N. Onuchic[f,1], and Vitor B.P. Leite[d,1]

[a]State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences Changchun, Jilin 130012 China; [b]College of Physics and State Key Laboratory of Superhard Materials, Jilin University, Changchun, Jilin 130021, China; [c]Department of Chemistry, Physics and Applied Mathematics, State University of New York at Stony Brook, Stony Brook, NY 11794-3400; [d]Departamento de Física—Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, 15054-000 São José do Rio Preto, Brazil; [e]Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais,13083-970 Campinas, SP, Brazil; and [f]Center for Theoretical Biological Physics, Rice University, 6100 Main, Houston, TX 77005-1827

Contributed by José N. Onuchic, August 1, 2012 (sent for review April 26, 2012)

The energy landscape approach has played a fundamental role in advancing our understanding of protein folding. Here, we quantify protein folding energy landscapes by exploring the underlying density of states. We identify three quantities essential for characterizing landscape topography: the stabilizing energy gap between the native and nonnative ensembles δE, the energetic roughness ΔE, and the scale of landscape measured by the entropy S. We show that the dimensionless ratio between the gap, roughness, and entropy of the system $\Lambda = \delta E/(\Delta E\sqrt{2S})$ accurately predicts the thermodynamics, as well as the kinetics of folding. Large Λ implies that the energy gap (or landscape slope towards the native state) is dominant, leading to more funneled landscapes. We investigate the role of topological and energetic roughness for proteins of different sizes and for proteins of the same size, but with different structural topologies. The landscape topography ratio Λ is shown to be monotonically correlated with the thermodynamic stability against trapping, as characterized by the ratio of folding temperature versus trapping temperature. Furthermore, Λ also monotonically correlates with the folding kinetic rates. These results provide the quantitative bridge between the landscape topography and experimental folding measurements.

energy landscape theory | biomolecular dynamics

Understanding how the amino acid sequence (i.e., primary structure) of each protein enables the native three-dimensional structure to be reached is one of the major challenges in molecular biophysics. In 1969, the arguments of Levinthal led to the suggestion of an apparent kinetic paradox (1). That is, if proteins were to randomly explore all possible states, cosmological timescales would be required for each protein to find the folded configuration. However, naturally occurring proteins fold in milliseconds to seconds. Protein folding theory has resolved this paradox by demonstrating that the underlying energy landscape is "funneled" towards the native state (2–9), however, local traps may be encountered during folding. To ensure that the folding occurs on biologically relevant timescales, the steepness of the protein folding funnel should be large, compared with the roughness due to local traps. Although this theory has provided the conceptual framework for interpreting folding experiments, both qualitatively and quantitatively (2, 5–18), it has yet to be explicitly demonstrated how the shape of the underlying landscape governs the thermodynamic stability and speed of folding, as measured experimentally (19). Here, we meet this challenge by quantifying the landscape topography and establishing the connection between the thermodynamics and kinetics of protein folding.

Naturally selected proteins differ from random sequences in that they fold into unique three-dimensional functional configurations. This indicates that the information necessary to fold is embedded in the amino acid sequence. This intrinsic information manifests in the form of energetic interactions and entropic contributions, which together define the temperature-dependent free-energy landscape. For describing folding, we are interested in the solvent-averaged effective energy, rather than the direct interactions between atoms (20, 21). Whereas detailed solvent–solute interactions contribute to the overall energetics, hydrophobic and hydrophilic interactions are described as solvent-averaged effects that drive folding. Similarly, the entropy of interest is the solvent-averaged configurational entropy of the polymer. Accordingly, the energy funnel describes the funnel-like character of the effective energy and configurational entropy.

We quantify the landscape by calculating the density of states in configuration and energy space. The density of states gives the probability of each energy/configuration and it does not depend explicitly on temperature. Accordingly, it describes the intrinsic energy landscape. The density of states is obtained from the microcanonical ensemble, whereas simulations of protein folding are usually performed for the canonical ensemble, which yields information on the free energy. In order to obtain the density of states, we transform the simulated canonical ensemble to the microcanonical ensemble.

To identify a robust metric that measures the "degree of funneledness" and overall topography of the landscape, we have used structure-based models with and without explicit energetic roughness (details in *Materials and Methods*). For both, funneled and rough landscapes, we explicitly demonstrate that three quantities are essential for characterizing the effective folding landscape topography: the energy gap (bias or slope of the funnel) (22, 23) between the native state and the average nonnative states $\delta E = |E_n - \bar{E}_{\mathrm{non-native}}|$, the roughness in the energies ΔE, and the size of the funnel measured by the configurational entropy of the system $S = k_{\mathrm{B}} \ln \Omega$ (where Ω is the number of states). We show that the dimensionless ratio between the gap, roughness and configurational entropy of the system $\Lambda = \delta E/(\Delta E\sqrt{2S})$ measures the landscape topography, and quantifies the degree of the funnel. Consistent with previous work (24), Λ is strongly correlated with the thermodynamic stability against traps (characterized by the ratio of the folding temperature versus the trapping temperature) and the folding kinetic rates. These results provide the quantitative bridge between the landscape topography and experimental folding measurements.

Author contributions: J.W., R.J.O., P.C.W., J.C., W.H., E.W., J.N.O., and V.B.P.L. designed research; R.J.O., X.C., and P.C.W. performed research; J.W., R.J.O., X.C., P.C.W., and V.B.P.L. analyzed data; and J.W., R.J.O., X.C., P.C.W., J.C., W.H., E.W., J.N.O., and V.B.P.L. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: jin.wang.1@stonybrook.edu or ekwang@ns.ciac.jl.cn or jonuchic@rice.edu or vleite@sjrp.unesp.br.

[2]R.J.O. and X.C. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1212842109/-/DCSupplemental.

Results and Discussions

Density of States and the Energy Landscape of Protein Folding. We explored the effective energy landscape and quantified the underlying topography to establish the link between the thermodynamics and kinetics of protein folding. We first performed molecular dynamics simulations, using a structure-based model, with and without energetic frustration included. The canonical ensemble can be efficiently sampled with replica (temperature) exchange methods (25–32). This provides a temperature-dependent distribution of the energies, which we then transform into the microcanonical ensemble through use of the weighted histogram analysis method (WHAM) algorithm (33). Because size effects can have a large influence on dynamics, it is crucial that we partition size-scaling and structural (or topological) effects. To do so, we studied two sets of proteins: The first group is formed by 13 proteins that vary in size, and the second is composed of nine proteins with similar size and different structural motifs (see *SI Appendix* for full analysis of each protein).

To visualize the energy landscape of folding, we projected the density of states onto specific order parameters, or reaction coordinates. The simplest representation of the landscape is the energy spectrum (Fig. 1) (34–36). From this, one may extract the stability gap δE and roughness ΔE (indicated by vertical arrows for each protein). The energy gap between the native and nonnative ensembles measures the bias, or slope, towards the native state. In contrast, the dispersion in energies describes the energetic roughness of the landscape. As shown in Fig. 1, there is a distinct gap between the native energy and the average energy of the nonnative states. With these models, all of the fast folding proteins that we studied have significant gaps, or biases, towards the native state. However, for each protein, the degree of roughness and scale of the gap are unique. As we will show, these differences allow one to classify proteins according to their thermodynamic and kinetic properties, and thereby provide a quantitative bridge between the funnel diagram and the physical–chemical properties of protein folding.

Size, Steepness, and Roughness of Folding Funnels. For an energetically unfrustrated model, the number of states (i.e., the configurational entropy $S = \ln[n(E)]$) decreases monotonically with decreasing energy. In other words, the states are more sparsely distributed as the energy is lowered. This is different from the canonical ensemble distribution $n(E, T)$, where T is the temperature and low energies have higher populations $[n(E, T) \sim n(E) \exp[E/k_B T]]$. The distribution of energy states, or density of states $n(E)$, is intrinsic to the system (i.e., not temperature dependent), and thus reflects the underlying landscape of the protein folding funnel. The configurational entropy $S(E)$ measures how many states are accessible and, therefore, describes the size of the state space. Initially, the size of the funnel is large for the unfolded states. As expected, the size of the funnel decreases as the native ensemble is approached. Additionally, the rate of decrease, or the slope, of the configurational entropy towards the native state in energy is different for different proteins (Fig. 2).

To estimate the glassy-like trapping temperature T_g, which describes the temperature at which the configurational entropy associated with nonnative states vanishes, the density of states was partitioned into native and nonnative ensembles. This was achieved by classifying the simulated configurations by their structural proximity to the native state (*SI Appendix*, Fig. S5). Using the microcanonical ensemble, T_g was determined according to the thermodynamic Maxwell relationship $\partial S/\partial E = 1/T$, and $\partial S/\partial E$ was calculated for the unfolded ensemble. T_g corresponds to the temperature below which the system becomes trapped or "frozen." Consistent with these measures of T_g, in our simulations, as T approaches T_g, the kinetics significantly slows down, has increased dispersion in the folding time, and becomes nonself-averaging (*SI Appendix*, Figs. S14 and S20). From T_g, the roughness of the underlying energy landscape of folding ΔE is determined from the relation $T_g = \sqrt{\Delta E^2/2S}$, where S is the configurational entropy of nonnative states (22). In contrast to T_g, the folding temperature T_f is obtained directly from the heat capacity curve, which is experimentally accessible.

Because experiments do not report on the configurational entropy and enthalpy of each configuration directly, it is important that we understand the connection between energetic and structural metrics of protein folding. Fig. 3 *A* and *B* show the density of states as a function of the order parameter Q, which is defined here as the fraction of native contacts formed (37). The configurational entropy, or density of states, decreases monotonically with Q. In other words, as folding proceeds, the size of the funnel [measured by $S(Q)$] decreases, where the the slope is distinct for each protein. The average energy also monotonically decreases as the system becomes more folded (Fig. 3 *C* and *D*). This depicts the funneled nature of the simulated landscapes. For an analytical mean field theory of protein folding (22), the energy gap is closely linked to the slope of $E(Q)$. Here, there is also a strong correlation between the average slope of the energy landscape $E(Q)$ and the energy gap (*SI Appendix*, Figs. S12 and S18), explicitly confirming that the models describe the landscape as being funneled.

There are many qualitative folding funnel illustrations available, though none have been quantitatively derived from experimental or theoretical measurements. In Fig. 4, we show three quantified folding funnels of villin headpiece, CI2 and P13, obtained directly from simulations. These funnels encapsulate the data described in Figs. 1–3. The horizontal axis is $\sqrt{(1-Q)S_r}$, where the area in the ellipsoid is equal to the configurational entropy $S(E)$, which provides the size of the funnel. The picture illustrates that as folding proceeds towards the native state, $S(E)$ and the effective energy $E(Q)$ (vertical axis) decrease. The steepness of the funnel is determined by the slope of the energy (with respect to Q), as well as the slope of the configurational entropy. The depicted roughness in the funnel is due to the

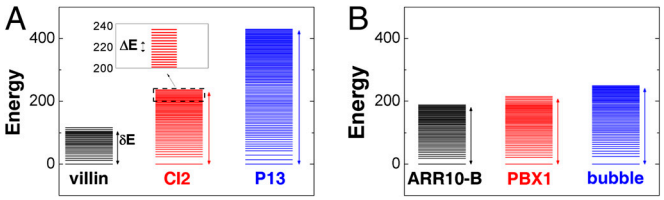

Fig. 1. Folding energy landscapes in zero dimensions. The distribution of the energy levels (spectrum) for (*A*) three (of 13 studied. See *SI Appendix*.) proteins with different sizes and (*B*) three (of nine) proteins with the same size. The lowest (native) energy E_n is set to 0 for visualization purposes. The stability gap δE between the native state and the average nonnative states is indicated by vertical arrows. Each energy level of the distribution represents the sum of a cluster of states, except for the native band. The inset is a magnification of the energy levels of CI2, with the scale of the energetic roughness ΔE indicated by a vertical arrow.

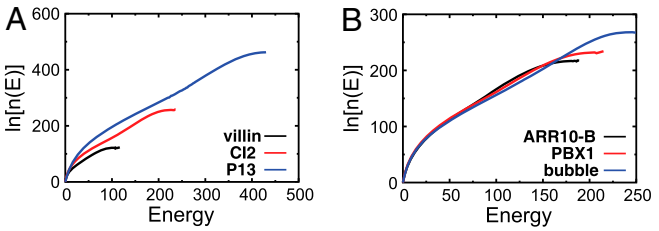

Fig. 2. Folding energy landscapes in one dimension. Logarithm of the density of states as a function of energy for (*A*) three (of 13) proteins with different sizes and (*B*) three (of nine) proteins of the same size. The lowest (native) energy E_n is set to 0 for visualization purposes.

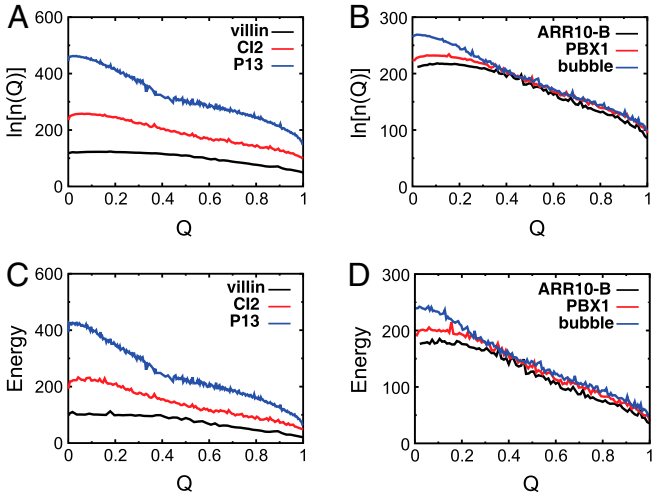

Fig. 3. Folding energy landscapes in one dimension. Logarithm of the density of states as a function of the fraction of native contacts Q for (*A*) three (of 13) proteins of different sizes and (*B*) three of the nine proteins of same size. Average energy as a function of Q for (*C*) three of the 13 different-sized proteins and (*D*) three of the nine proteins with same size.

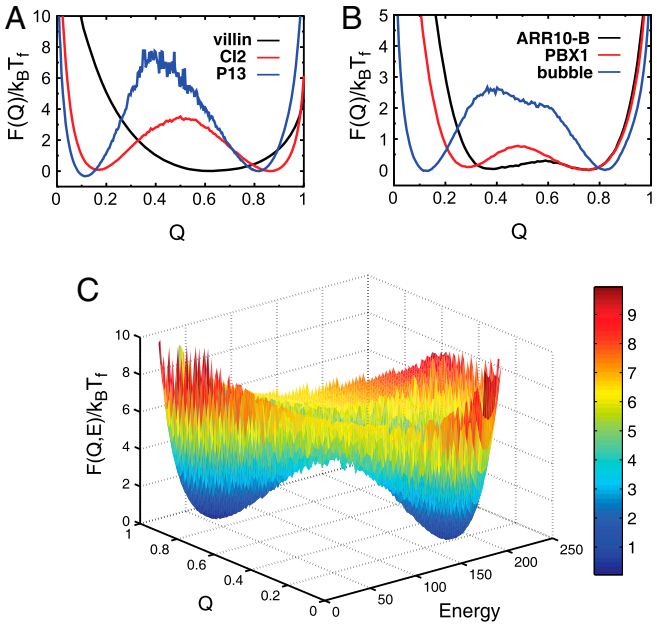

Fig. 5. The free energy at folding temperature in one dimension as a function of fraction of native contacts Q and in two dimension as a function of Q and energy. Free-energy profile, at folding temperature T_f, as a function of Q for (*A*) three (of 13) proteins with different sizes and (*B*) three (of nine) proteins with the same size. (*C*) Two-dimensional free-energy profile for CI2 as a function of Q and energy at T_f normalized by the corresponding k_BT_f.

imperfect correlation between Q and E (details in *SI Appendix*). As the protein folds, the radius of the funnel [related to $(1-Q)$], the steepness of landscape, and the size of the funnel all decrease. The overall shape of the funnel is then determined by the steepness over the roughness and the size (entropy) of the unfolded ensemble.

Free-Energy Profile of Protein Folding. In contrast to the density of states, the free-energy landscape is dependent on temperature, which can give the impression that a particular landscape is not funneled, even when the underlying density of states is. Fig. 5 *A* and *B* show the free energy as a function of the folding coordinate $F(Q)$, at the folding temperature T_f. The profiles were calculated from a C_α structure-based model for downhill folders (villin headpiece), two-state folders (CI2 and P13) with different sizes, and proteins with the same size but different structural topologies. The shape of the free-energy landscape results from a balance between energy and configurational entropy, leading to its explicit dependence on temperature. For two-state folders, at T_f the free energies of the folded and unfolded ensembles are equal. Below T_f, the free energy is dominated by the effective enthalpy and the system is biased towards the native ensemble. Above T_f, the free energy is dominated by the configurational entropy and is heavily weighted towards the nonnative ensemble. For comparison to other common measures of folding, we also calculated the free energy at T_χ (38), which is the temperature at which the native ensemble accounts for 80% of the population (*SI Appendix*, Fig. S28 *A–C*). At T_0, the folding kinetics is the fastest (*SI Appendix*, Fig. S28 *D–F*). Because T_χ is lower than T_f, the free-energy profiles are biased towards the native state. As temperature approaches T_0, the folding of each protein becomes downhill-like. Two-dimensional free-energy profiles at T_f, T_χ, T_0 are shown in Fig. 5*C* and *SI Appendix*, Fig. S28 *G* and *H*. The relative free energy of the native ensemble and nonnative ensemble provides a measure of thermodynamics (or equilibrium constants), linking to thermodynamic folding experiments.

Some of the studied proteins are strong folders, with one basin of attraction in the free-energy landscape. The slope of the free-energy landscape for the one-state folders measures how easy it is to fold. For other proteins, two basins of attraction emerge, leading to the two-state behavior of the proteins, as is often seen in protein folding experiments. The barrier heights of the two-state folders measure how difficult it is to fold. The often nonfunnel-like folding free-energy landscape is the result of a competition and imperfect cancelation between energy and configurational entropy. This is the view from the canonical ensemble. For the microcanonical ensemble, whereas the underlying energy landscape is also determined by the competition between energy, roughness, and configurational entropy, there is only one basin of attraction that shapes the energy landscape.

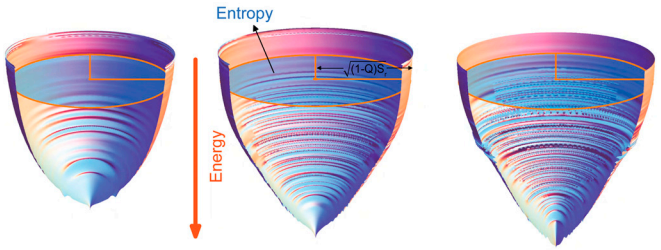

Fig. 4. Quantified funneled landscapes for villin headpiece (*Left*), CI2 (*Center*) and P13 (*Right*) obtained from simulation. The depth of the funnel is the energy and the cross-section perpendicular to the energy axis is an ellipsoid with an area equal to the configurational entropy $S(E)$. The semimajor axis of the ellipsoid at each stratum is $\sqrt{(1-Q(E))S_r}$, which is a function of the fraction of native contacts $Q(E)$, S_r is a constant for each protein. The other semiaxis is given by $S(E)/(\pi\sqrt{(1-Q(E))S_r})$. Both energy and configurational entropy are normalized by the protein size.

Energy Landscape Topography Determines the Thermodynamics of Folding. The landscape topographic measure Λ correlates with the thermodynamic folding stability against trapping for several classes of proteins and energetic models (Fig. 6): (a) proteins of different sizes with different structural topologies that are modeled without energetic roughness [i.e., completely funneled landscape (39)]; (b) proteins of different sizes, modeled with energetic roughnesses; (c) proteins of the same size, but different topologies, without energetic roughness included; (d) proteins of the same size with different structural topologies and energetic

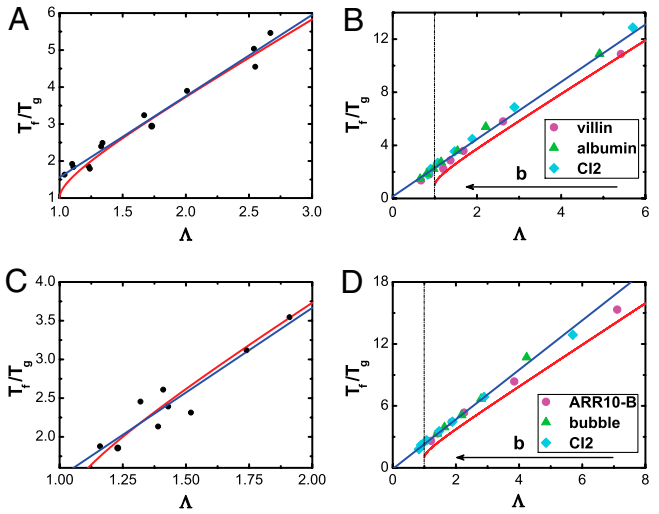

Fig. 6. The correlation between folding against trapping T_f/T_g and landscape topography Λ. The red line is the analytical mean field theory prediction of the relationship between T_f/T_g and Λ: $T_f/T_g = \Lambda + (\Lambda^2 - 1)^{1/2}$. The blue line is a linear fit of T_f/T_g versus Λ. For all cases, the correlation coefficient (c.c) is greater than 0.9. (*A*) Folding of proteins of different sizes and different structural topologies, but no energetic roughness. (*B*) Folding of different-sized proteins with energetic roughness. (*C*) Folding of proteins of similar size with different structural topologies but no energetic roughness included. (*D*) Folding of same-size proteins with energetic roughness included. In *B* and *D*, b sets the energy variance of the non-native interactions (details in *SI Appendix*). Λ monotonically correlates with b.

roughness included. For each group, we find a strong positive correlation between thermodynamic stability [characterized by the dimensionless ratio of folding transition temperature to the trapping temperature T_f/T_g (40–42)] and Λ, which is consistent with expectations from mean field theory (2, 7, 22, 43). That is, the steeper and less rugged (or the smaller the size of) the funnel, the more thermodynamically stable the protein is against energetic traps. This demonstrates that the thermodynamic stability of proteins that is measured from experiments will reflect the the underlying topography of the landscape. Specifically, these results demonstrate that T_f and T_g (44) measurements will allow us to infer the characteristics of the underlying landscape.

For each class of proteins, the origin of the correlation between Λ and T_f/T_g may be attributed to different features. Because the slope, roughness, and configurational entropy of the landscape, as well as the folding and trapping temperatures, all depend on protein size, the correlation across different-sized proteins (Fig. 6*A*) is expected from scaling considerations (*SI Appendix*, Figs. S11 and S13). This is inconsistent with mean field approximations because finite sizes and the capillarity of protein folding are accounted for here (45). When energetic roughness is included (Fig. 6*B*) the trapping temperature and Λ are altered, and the correlation is due to the energetic roughness dependence of both. For proteins with different structural topologies (Fig. 6*C*), size-scaling arguments cannot account for the correlation, which implicates the residue–residue connectivity as leading to changes in the underlying slope, roughness, and size of the landscape. When introducing energetic roughness to proteins of the same size (Fig. 6*D*) the origin of the correlation is again due to concomitant changes in landscape topography and thermodynamic stability. Thus, these results demonstrate that Λ is a measure of landscape topography that can accurately predict the thermodynamic stability against trapping, whether or not they are size-scaling or energetic in origin.

Energy Landscape Topography Determines the Kinetics of Folding. To explore the relationship between the landscape topography and kinetics, we have simulated folding and calculated the kinetic rates. We first investigated proteins of different sizes, using an energetically unfrustrated model that has been shown to produce folding kinetics that are correlated with experimental measurements (11, 45–58). Kinetic simulations were performed at the temperature T_χ, where 80% of the population occupies the native ensemble (38). We find that the rates are strongly negatively correlated with the landscape topographic measure Λ (Fig. 7). This illustrates that the steeper, less rugged, or smaller size (i.e., less entropy) of the folding funnel, the faster folding occurs (59, 60). This strong correlation indicates that experimental measures of the folding speed can be used to obtain quantitative measures of the underlying energy landscape.

Because the kinetic rate and landscape topography measure Λ depend on size, the correlation for proteins of different sizes (in the energetically unfrustrated model) may result from size-dependent effects. To address this potential limitation in our description of kinetics and topography, we calculated the folding kinetics for proteins of different sizes, with a variable degree of energetic roughness included (Fig. 7*B*). We find a comparable correlation when using models with energetic roughness. Because the landscape topography is directly related to the energetic roughness, and the kinetic rates also depend on the energetic roughness (rougher landscape slows down kinetics), the correlation between the landscape topography and folding speed reflects the influence of energetic roughness. It is worth noting that, when the nonnative interaction strength is increased from zero, the correlation between folding rate and Λ is negligible. This is consistent with theoretical considerations and simulations that have shown that a small degree of nonnative interactions (relative to a completely unfrustrated landscape) can decrease the free-energy barrier and accelerate folding rates (39, 61). In the present study, the model is structure based, where large values of Λ correspond to a very small degree of energetic roughness, whereas proteins must have a finite degree of roughness in their landscapes. Thus, for proteins in solution, one does not expect to observe such accelerating effects because the landscapes have a finite degree of roughness already. Therefore, the correlation

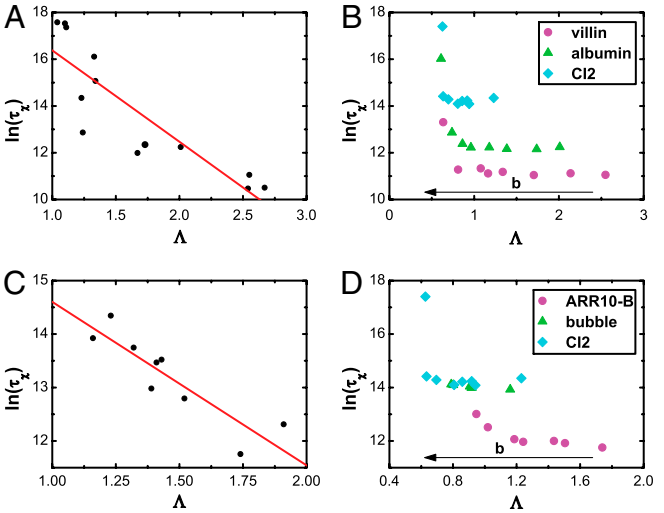

Fig. 7. The correlation between the kinetic rates of folding and landscape topography Λ, shown for: (*A*) proteins of different sizes and topologies, without energetic roughness included (c.c. = −0.87); (*B*) proteins of different sizes with varying degrees of energetic roughness included; (*C*) proteins of similar size and different structural topologies but no energetic roughness included (c.c. = −0.89); (*D*) proteins of the same size with energetic roughness included. The landscape topography Λ is strongly correlated with folding time. The folding time is in reduced units.

between folding rates and Λ is likely to be monotonic in real proteins, where larger Λ leads to larger folding rate.

To remove size-scaling effects and identify the structural contributions to the folding kinetics, we performed simulations of nine proteins that are comparable in size (Fig. 7*C*). We first explored the folding kinetics for proteins of the same size and different structural topologies, with a model that lacks energetic roughness. For those cases, the structural topologies determine the configurational entropy, landscape slope, and connectivity of substates. The correlation between the landscape topography and folding speed, therefore, demonstrates the structural-topology dependence of the landscape. To ensure that these results are not model dependent, we repeated these calculations with a model that included stabilizing nonnative interactions (Fig. 7*D*). We again find that the energetic roughness has a considerable impact on the landscape topography and leads to a correlation between the underlying topography and folding speed.

Because other metrics, such as contact order [CO, an entropic measure based on the native structure of proteins (53, 61–64)], are found to be correlated with the kinetics of folding, Λ should be correlated with those quantities. The correlation between the relative contact order (RCO) and Λ for proteins of the same size with the energetically unfrustrated model implies that Λ contains the structural topological information of each protein (Fig. 8*A*). In addition, our results show that RCO is correlated with the folding rate for the nine same-sized energetically unfrustrated proteins (Fig. 8*B*). Compared with RCO, the energy landscape topography measure Λ shows a stronger correlation with the folding rates. Although such observations reinforce the notion that native-state topology influences the kinetics, the presented landscape topography measure Λ captures the topological and energetic contributions to the landscape and is capable of predicting the thermodynamics and kinetics of folding.

Conclusions

We have quantified the energy landscape of protein folding by calculating the underlying density of states. Based on the density of states, we have shown that the dimensionless ratio between the gap, roughness, and configurational entropy of the system $[\Lambda = \delta E/(\Delta E\sqrt{2S})]$ quantifies the topography of the underlying landscape and measures the degree of its funneledness. Λ represents a universal metric by which the overall landscape topography may be extracted from thermodynamic or kinetic measurements. Further, we have demonstrated that Λ is not only appropriate for particular subsets of proteins (such as fast folding proteins) but also for proteins of different sizes and energetic character.

Each protein studied has a folding funnel associated with a unique slope, roughness, and size. The topography measure of the underlying folding landscape, Λ, is shown to be correlated strongly (monotonically) with the thermodynamic stability against traps, characterized by the ratio of thermodynamic stability temperature versus trapping temperature. In addition, the landscape topography measure Λ also monotonically correlates with the folding kinetic rates. These demonstrate how the energy landscape topography determines the thermodynamics and kinetics of protein folding.

Energy landscape theory has guided our understanding of protein folding and has led to progress in the study of folding thermodynamics and kinetics. In our studies, because the landscape topography, thermodynamics, and kinetics are so strongly correlated, Λ represents the meeting point for theoretical quantification of the landscape topography and experimental evaluation of stability and kinetics. In addition, this quantitative energy landscape framework can be extended to multidomain proteins and folding in vivo. It is worth noting that our studies have used a coarse-grained structure-based model to explore the energy landscape. However, as computing hardware and software continue to advance, the presented framework may be utilized in conjunction with all-atom molecular dynamic simulations. Such efforts will allow one to partition the effects of solvent interactions on protein folding and identify how they contribute to the global topology of the energy landscape. In closing, the current study bridges the gap between the quantification of the underlying folding energy landscape topography and experiments (or simulations) on thermodynamics and kinetics of protein folding.

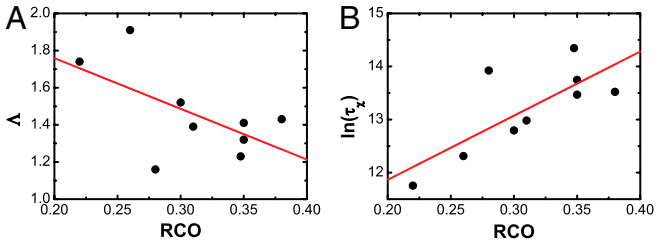

Fig. 8. The correlation between landscape topography Λ, kinetic rates of folding and RCO. (*A*) Λ versus RCO for nine proteins of the same size but different structural topologies (c.c. = −0.59). (*B*) Folding time $\ln(\tau_\chi)$ as a function of RCO (c.c. = 0.76).

Materials and Methods

A structure-based model (13), coarse grained at the residue level, was used to explore the density of states. To ensure efficient sampling, the density of states was calculated from replica exchanged molecular dynamics simulations (25), performed at 48 temperatures for each protein. To obtain a single measure of the density of states, WHAM (33) was used. The deviation of each configuration from the native structure was used to divide sampled configurations into native and nonnative ensembles. With the thermodynamic relation $1/T = \partial S/\partial E$, we calculated the slope of the density of states for the nonnative ensemble to estimate the temperature at which the configurational entropy of the nonnative states vanishes ($1/T_g$). From the trapping temperature T_g, we directly obtained the landscape topological roughness: $\Delta E_{\text{Top}} = \sqrt{2S}T_g$. S is the configurational entropy of the nonnative states. The energy gap δE is defined as the difference between the average energy of the nonnative states and the energy of native state. S and δE were calculated from the density of states.

To explore the role of the energetic roughness on the landscape topographical measure, we introduced energetic frustration into the structure-based model by including nonnative interactions (39, 65). The weight of each nonnative interaction was assigned according to a Gaussian distribution with a variance of b^2. We modulated the scale of the energetic roughness by changing b^2. $b^2 = 0$ corresponds to a structure-based model. The energetic roughness ΔE_{Ene} introduced by nonnative interactions is expressed as $\sqrt{A_{\text{max}}b^2}$, where A_{max} is the maximum number of nonnative contacts the protein can form.

The total landscape roughness of proteins ΔE_{Total} is due to both topological and energetic roughness contributions. In purely structure-based models (13, 66), there are no nonnative interactions, and the energetic contribution to the roughness is $\Delta E_{\text{Ene}} = 0$. When nonnative interactions are included, the roughness of the energy landscape is the sum of the topological roughness and energetic roughness. To calculate ΔE_{Total}, we make the assumption that modulating the variance of nonnative interactions does not significantly alter the topological contribution to the roughness. This assumption is valid because the topological roughness arises from constraints on the geometry and shape of the native structure, which is not significantly perturbed when the energetic roughness is sufficiently low (6768). In other words, the nonnative interactions in real proteins can be approximated as energetic heterogeneity perturbations to the background roughness introduced by the topology. Accordingly, we use the expression: $\Delta E^2_{\text{Total}} = \Delta E^2_{\text{Top}} + \Delta E^2_{\text{Ene}}$ to calculate the total roughness of each protein. The trapping transition temperature (T_g) will be: $T_g = \sqrt{\Delta E^2_{\text{Total}}/2S}$. The folding landscape measure is $\Lambda = \delta E/(\Delta E_{\text{Total}}\sqrt{2S})$.

ACKNOWLEDGMENTS. X.C. and E.W. are supported by the National Natural Science Foundation (NSF) of China (Grants 21190040 and 11174105) and the 973 project 2009CB930100 and 2010CB933600. J.W. is supported by the National Science Foundation. R.J.O., J.C., and V.B.P.L. were supported by Brazilian Agencies Fundação de Amparo à Pesquisa do Estado de São Paulo, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior and Conselho Nacional de Desenvolvimento Científico e Tecnológico. This

research was supported by resources supplied by the Núcleo de Computação Científica/Grid Universidade Estadual Paulista—Universidade Estadual Paulista. This work was supported by the Center for Theoretical Biological Physics sponsored by the NSF (Grant PHY-0822283) and by NSF-MCB-1214457. J.N.O. is a CPRIT Scholar in Cancer Research sponsored by the Cancer Prevention and Research Institute of Texas.

1. Levinthal C (1969) *Proceedings in Mossbauer Spectroscopy in Biological Systems*, eds P Debrunner, J Tsibris, and E Munck (University of Illinois Press, Urbana, IL), pp 22–24.
2. Bryngelson JD, Wolynes PG (1987) Spin-glasses and the statistical-mechanics of protein folding. *Proc Natl Acad Sci USA* 84:7524–7528.
3. Bryngelson JD, Wolynes PG (1989) Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J Phys Chem* 93:6902–6915.
4. Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels—A kinetic approach to the sequence structure relationship. *Proc Natl Acad Sci USA* 89:8721–8725.
5. Chan HS, Dill KA (1994) Transition-states and folding dynamics of proteins and heteropolymers. *J Chem Phys* 100:9238–9257.
6. Abkevich VI, Gutin AM, Shakhnovich EI (1994) Free-energy landscape for protein-folding kinetics—Intermediates, traps, and multiple pathways in theory and lattice model simulations. *J Chem Phys* 101:6052–6062.
7. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein-folding—A synthesis. *Proteins* 21:167–195.
8. Wang J, Onuchic J, Wolynes P (1996) Statistics of kinetic pathways on biased rough energy landscapes with applications to protein folding. *Phys Rev Lett* 76:4861–4864.
9. Brooks CL, Onuchic JN, Wales DJ (2001) Statistical thermodynamics—Taking a walk on a landscape. *Science* 293:612–613.
10. Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition-state for folding of chymotrypsin inhibitor-2 analyzed by protein engineering methods—Evidence for a nucleation-condensation mechanism for protein-folding. *J Mol Biol* 254:260–288.
11. Klimov DK, Thirumalai D (1998) Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J Chem Phys* 109:4119–4125.
12. Gruebele M, Sabelko J, Ervin J (1999) Observation of strange kinetics in protein folding. *Proc Natl Acad Sci USA* 96:6031–6036.
13. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298:937–953.
14. Nymeyer H, Garcia AE, Onuchic JN (1998) Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc Natl Acad Sci USA* 95:5921–5928.
15. Onuchic JN, Nymeyer H, Garcia 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.
16. Eaton WA, Schuler B, Lipman EA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
17. Stell G, Lee CL, Wang J (2003) First-passage time distribution and non-Markovian diffusion dynamics of protein folding. *J Chem Phys* 118:959–968.
18. Zhou YQ, Zhang C, Stell G, Wang J (2003) Temperature dependence of the distribution of the first passage time: Results from discontinuous molecular dynamics simulations of an all-atom model of the second beta-hairpin fragment of protein g. *J Am Chem Soc* 125:6300–6305.
19. Englander SW, Milne JS, Xu YJ, Mayne LC (1999) Experimental study of the protein folding landscape: Unfolding reactions in cytochrome *c*. *J Mol Biol* 290:811–822.
20. Wolynes PG (1995) Biomolecular folding in vacuo. *Proc Natl Acad Sci USA* 92:2426–2427.
21. Wolynes PG, Onuchic JN, Thirumalai D (1995) Navigating the folding routes. *Science* 267:1619–1620.
22. Plotkin SS, Wang J, Wolynes PG (1997) Statistical mechanics of a correlated energy landscape model for protein folding funnels. *J Chem Phys* 106:2932–2948.
23. Shakhnovich EI (1994) Proteins with selected sequences fold into unique native conformation. *Phys Rev Lett* 72:3907–3910.
24. Goldstein R, Luthey-Schulten Z, Wolynes PG (1992) Optimal protein-folding codes from spin-glass theory. *Proc Natl Acad Sci USA* 89:4918–4922.
25. Okamoto Y, Sugita Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151.
26. Garcia AE, Onuchic JN (2003) Folding a protein in a computer: An atomic description of the folding/unfolding of protein A. *Proc Natl Acad Sci USA* 100:13898–13903.
27. Rodinger T, Pomes R (2005) Enhancing the accuracy, the efficiency and the scope of free energy simulations. *Curr Opin Struct Biol* 15:164–170.
28. Rodinger T, Howell PL, Pomes R (2006) Distributed replica sampling. *J Chem Theor Comput* 2:725–731.
29. Zhang C, Ma J (2008) Comparison of sampling efficiency between simulated tempering and replica exchange. *J Chem Phys* 129:134112.
30. Zhang C, Ma J (2009) Enhanced sampling in generalized ensemble with large gap of sampling parameter: Case study in temperature space random walk. *J Chem Phys* 130:194112.
31. Zhang C, Ma J (2010) Enhanced sampling and applications in protein folding in explicit solvent. *J Chem Phys* 132:244101.
32. Rauscher S, Neale C, Pomes R (2010) Simulated tempering distributed replica sampling, virtual replica exchange, and other generalized-ensemble methods for conformational sampling. *J Chem Theor Comput* 5:2640–2662.
33. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) The weighted histogram analysis method for free-energy calculations on biomolecules 1. The method. *J Comput Chem* 13:1011–1021.
34. Honeycutt JD, Thirumalai D (1990) Metastability of the folded states of globular proteins. *Proc Natl Acad Sci USA* 87:3526–3529.
35. Sali A, Shakhnovich E, Karplus M (1994) How does a protein fold. *Nature* 369:248–251.
36. Shakhnovich EI, Gutin AM (1990) Implications of thermodynamics of protein folding for evolution of primary sequences. *Nature* 346:773–775.
37. Cho SS, Levy Y, Wolynes PG (2006) P versus q: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc Natl Acad Sci USA* 103:586–591.
38. Klimov DK, Thirumalai D (1996) Criterion that determines the foldability of proteins. *Phys Rev Lett* 76:4070–4073.
39. Clementi C, Plotkin SS (2004) The effects of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci* 13:1750–1766.
40. Eastwood M, Hardin C, Luthey-Schulten Z, Wolynes PG (2001) Evaluating protein structure prediction schemes using energy landscape theory. *IBM Res Dev* 45:475–497.
41. Hardin C, Eastwood MP, Prentiss M, Luthey-Schulten Z, Wolynes PG (2002) Folding funnels the key to robust protein structure prediction. *J Comput Chem* 23:138–146.
42. Fujitsuka Y, Takada S, Luthey-Schulten Z, Wolynes PG (2004) Optimizing physical energy functions for protein folding. *Prot Struct Func Gen* 54:88–103.
43. Wang J, Verkhivker GM (2003) Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Phys Rev Lett* 90:188101.
44. Iben IET, et al. (1989) Glassy behavior of a protein. *Phys Rev Lett* 62:1916–1919.
45. Wolynes PG (1997) Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc Natl Acad Sci USA* 94:6170–6175.
46. Cieplak M, Hoang TX, Li MS (1999) Scaling of folding properties in simple models of proteins. *Phys Rev Lett* 83:1684–1687.
47. Thirumalai D (1995) From minimal models to real proteins—Time scales for protein-folding kinetics. *J Phys I* 5:1457–1467.
48. Finkelstein AV, Badretdinov AY (1997) Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. *Fold Des* 2:115–121.
49. Cieplak M, Hoang TX (2000) Scaling of folding properties in go models of proteins. *J Biol Phys* 26:273–294.
50. Takada S, Koga N (2001) Roles of native topology and chain-length scaling in protein folding: A simulation study with a go-like model. *J Mol Biol* 313:171–180.
51. Li MS, Klimov DK, Thirumalai D (2002) Dependence of folding rates on protein length. *J Phys Chem B* 106:8302–8305.
52. Finkelstein AV, et al. (2003) Contact order revisited: Influence of protein size on the folding rate. *Protein Sci* 12:2057–2062.
53. Galzitskaya OV, Garbuzynskiy SO, Ivankov DN, Finkelstein AV (2003) Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics. *Proteins* 51:162–166.
54. Li MS, Klimov DK, Thirumalai D (2004) Thermal denaturation and folding rates of single domain proteins: Size matters. *Polymer* 45:573–579.
55. Munoz V, Naganathan AN, Sanchez-Ruiz JM (2005) Direct measurement of barrier heights in protein folding. *J Am Chem Soc* 127:17970–17971.
56. Naganathan AN, Munoz V (2005) Scaling of folding times with protein size. *J Am Chem Soc* 127:480–481.
57. Kouza M, Li MS, O'Brien EP, Hu CK, Thirumalai D (2006) Effect of finite size on cooperativity and rates of protein folding. *J Phys Chem A* 110:671–676.
58. Klimov DK, Thirumalai D (1998) Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J Chem Phys* 109:4119–4125.
59. Saven JG, Wang J, Wolynes PG (1994) Kinetics of protein folding: The dynamics of globally connected rough energy landscapes with biases. *J Chem Phys* 101:11037–11043.
60. Socci ND, Onuchic JN, Wolynes PG (1996) Diffusive dynamics of the reaction coordinate for protein folding funnels. *J Chem Phys* 104:5860–5868.
61. Plotkin SS (2001) Speeding protein folding beyond the go model: How a little frustration sometimes helps. *Proteins: Struct Funct Bioinform* 45:337–345.
62. Baker D, Plaxco KW, Simons KT (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277:985–994.
63. Raleigh DP, Kuhlman B, Luisi DL, Evans PA (1998) Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the n-terminal domain of the protein l9. *J Mol Biol* 284:1661–1670.
64. Dinner AR, Karplus M (2001) The roles of stability and contact order in determining protein folding rates. *Nat Struct Biol* 8:21–22.
65. Clementi C, Das P, Matysiak S (2005) Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes. *Proc Natl Acad Sci USA* 102:10141–10146.
66. Whitford PC, et al. (2009) An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins: Struct Funct Bioinform* 75:430–41.
67. Thirumalai D, Klimov DK, Woodson SA (1997) Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theor Chem Acc* 96:14–22.
68. Thirumalai D, Klimov DK (1999) Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr Opin Struct Biol* 9:197–207.