Protein folding on rugged energy landscapes: Conformational diffusion on fractal networks

Gregg Lois, Jerzy Blawzdziewicz, and Corey S. O’Hern
Department of Mechanical Engineering and Department of Physics,
Yale University, New Haven, Connecticut 06520-8286

We employ simulations of model proteins to study folding on rugged energy landscapes. We construct “first-passage” networks as the system transitions from unfolded to native states. The nodes and bonds in these networks correspond to basins and transitions between them in the energy landscape. We find power-laws between the folding time and number of nodes and bonds. We show that these scalings are determined by the fractal properties of first-passage networks. Reliable folding is possible in systems with rugged energy landscapes because first passage networks have small fractal dimension.

Understanding how proteins reliably fold to their native conformations despite frustration in the form of non-native interactions between residues is an important, open question. Advances in experimental techniques, such as single-molecule fluorescence [1] and fast thermal quenching methods [2], have enabled a quantitative characterization of the dynamics that occur during folding of single proteins. For example, we now know that a large number of metastable conformations are sampled during the folding and unfolding processes, as observed in folding stability [3] and mechanical denaturation [4] studies.

How does a protein fold reliably to its native conformation even though a large number of metastable states exist? For over twenty years the answer to this question has been the principle of minimal frustration [5]. Within this framework, one recognizes that metastable states are present, but assumes that the barriers separating local energy minima are sufficiently low that there is still a large thermodynamic force driving folding to the native state [6].

We make a crucial first step in answering this question by studying the properties of a model protein that reliably folds to its native state on a rugged energy landscape with $10^2 - 10^4$ distinct basins sampled during folding. (A basin is a region of configuration space, or collection of conformations, that relaxes to a single local energy minimum when thermal fluctuations are suppressed [11].) Instead of discrete pathways through the energy landscape, we find a statistical ensemble of pathways with large fluctuations in folding times. The folding time and number of distinct basins sampled during folding scale as a power-law, which suggests that reliable folding on rugged landscapes can be described as conformational diffusion on a fractal network of basins.

**Heteropolymer model:** To study proteins with rugged energy landscapes, simulation models should possess three key features: (1) unique native state, (2) many metastable, local energy minima, and (3) large energy barriers that separate local minima so that $\delta E \sim \Delta E$. Further, we must be able to search configuration space in a reasonable amount of computer time, which excludes all-atom simulations. In these studies, we will focus on a model heteropolymer that exhibits features (1)-(3).

We model proteins as heteropolymers composed of equal-sized spherical monomers with hydrophobic and hydrophilic interactions [12]. The model includes hydrophilic monomers (white) and two types of hydrophobic monomers (red and green) as shown in Fig. 2. Green and red monomers interact via an attractive Lennard-
using Brownian dynamics, where the temperature of this heteropolymer is given by the particular set of 14 green and red monomers [15]. The native conformation of this heteropolymer is given by the particular set of 14 green-red contacts shown in Fig. 2 (c).

Thermal fluctuations of the heteropolymer are studied using Brownian dynamics, where the temperature $T$ is reported in units of the attractive energy, e.g. $T = 1/3$ corresponds to thermal energy $E_{\text{att}}/3$. To compare results for rugged and funnelled energy landscapes, we also simulated the same heteropolymer with Go-interactions [16], where attractive interactions are purely repulsive [13]. We also include a FENE potential [14] between adjacent monomers to maintain the polymer constraint. We simulate the 18-mer sequence gggwrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr!g
that the exponents $\Gamma$ and $\Lambda$ are independent of $q$ formed simulations in the range $1 < q < q_{basins}$ every with that energy more uniformly. In systems with funneled energy landscapes (model), a protein with energy $E_{i}$ for the same heteropolymer model with Go-interactions. To large activation barriers. In contrast, $\Gamma$ landscape at energy $E$ with rugged landscapes. A system with a rugged energy dependent exploration of configuration space in systems $T \approx T_{th}$ different temperatures $\propto b^{i}$ and $t \propto N_{i}^{\Lambda}$, and we find $N_{i} \propto N_{f}^{1/\kappa d_{f}}$, or results further indicate that first-passage networks are self-similar and fractal.

**Origin of power laws:** If we assume that first-passage networks are fractal, we can predict the exponent $\Gamma$ from the fractal scaling exponents of the network. This assumption will be verified a posteriori. On any network we can define the chemical distance $\Delta c$ given by the shortest path between two nodes of the network. This distance is useful because it depends only on network connectivity and is independent of the embedding space. For a fractal network, we expect

$$\Delta c \propto t^{\kappa},$$

$$N(\Delta c) \propto \Delta c^{d_{f}},$$

where $N(\Delta c)$ is the number of distinct basins sampled within chemical distance $\Delta c$ and time interval $t$, $d_{f}$ is the chemical fractal dimension, and the exponent $\kappa$ characterizes the scaling of chemical distance with time.

Given these relations, the correlation between $N_{i}$ and $N_{t}$ can be explained as follows. A single first-passage network is formed over folding time $\tau_{f} \propto N_{t}$, during which the system explores average chemical distance $\Delta c \propto N_{t}^{\kappa}$ (Eq. 2). Moreover, for a given chemical distance $\Delta c$, the number of sampled basins on the first passage network scales as $N_{t} \propto N(\Delta c) \propto \Delta c^{d_{f}}$ (Eq. 3). Thus, both $N_{i}$ and $N_{t}$ are related to $\Delta c$, and we find $N_{i} \propto N_{f}^{1/\kappa d_{f}}$, or

$$\Gamma = \frac{1}{\kappa d_{f}}.$$  

The prediction for $\Gamma$ relies on the first-passage networks being fractal. In Fig. 6 (a), we test Eq. 2 and observe that $\Delta c$ grows as a power law at large $t$ for all temperatures studied. We average $\Delta c$ over 1500 first-passage networks and only include $t < \tau_{f}$ for each realization. The exponent $\kappa$ decreases with $T$, which implies that colder systems explore chemical distance more slowly.

In Fig. 5 (b) we test Eq. 4 and find that, over the limited range of chemical distance accessible to our small heteropolymer, the chemical fractal dimension $d_{f}$ is well-defined and depends linearly on temperature. $N(\Delta c)$ is

In Fig. 4, $N_{b}, N_{i}$ and $N_{t}$ show strong fluctuations from one realization to the next; however, the fluctuations obey power-law scaling:

$$N_{b} \propto N_{t}^{\Lambda} \quad \text{and} \quad N_{t} \propto N_{i}^{\Gamma}. \quad (1)$$

This correlation is non-trivial and depends on global properties of first-passage networks. We find that distributions of local features of the network, such as single-jump activation times and distances, and the number of bonds per node, are exponential. Thus, local properties of first-passage networks cannot be responsible for the power-law scaling.

In Fig. 5 we plot the scaling exponents $\Gamma$ and $\Lambda$ at different temperatures $T$. While $\Lambda$ reaches a plateau at $\approx 1.4$ at small $T$, $\Gamma$ continues to increase with decreasing $T$. The increase of $\Gamma$ is a signature of temperature-dependent exploration of configuration space in systems with rugged landscapes. A system with a rugged energy landscape at energy $E$ only samples a small temperature-dependent fraction of conformations at that energy due to large activation barriers. In contrast, $\Gamma \approx 1.5$ at all $T$ for the same heteropolymer model with Go-interactions. In systems with funneled energy landscapes (i.e. the Go model), a protein with energy $E$ samples conformations with that energy more uniformly.

The data shown in Fig. 4 are obtained by identifying basins every $q = 1000$ time steps. We have also performed simulations in the range $1 < q < 10^{4}$ and observe that the exponents $\Gamma$ and $\Lambda$ are independent of $q$. These
Moreover, the complete network. This behavior is not peculiar to also trace out fractal networks that are independent of lying space is, proteins with rugged energy landscapes nodes, no matter how large the dimension of the under-

trace out a two-dimensional fractal network of sampled networks in configuration space. However, as far as folding is concerned, our results suggest that this network is not relevant. Just as normal diffusion will trace out two-dimensional fractal network of sampled nodes, no matter how large the dimension of the underlying space is, proteins with rugged energy landscapes also trace out fractal networks that are independent of the complete network. This behavior is not peculiar to proteins with rugged energy landscapes, but is also expected in glass-forming materials at low temperature [21]. Moreover, \( d_f \) decreases with temperature, and is always much smaller than the dimension of configuration space \( D \), which implies that \( N_f \sim (\Delta c)^{d_f} \ll (\Delta c)^D \). This provides a mechanism by which systems with rugged energy landscapes can fold reliably without kinetic pathways and offers a novel resolution to Levinthal’s paradox [22].

Financial support from NSF grant numbers CBET-0348175 (GLJB), DMS-0835742 (CSO), and DMR-0448838 (CSO), and Yale’s Institute for Nanoscience and Quantum Engineering (GL) is acknowledged.

\[ 1 \] E. A. Lipman, B. Schuler, O. Bakajin and W. A. Eaton, Science 301, 1233 (2003).
\[ 2 \] W. A. Eaton, V. Munoz, P. A. Thompson, C.-K. Chan and J. Hofrichter, Curr. Opin. Struct. Biol. 7, 10 (1997).
\[ 3 \] H. Yang, G. Luo, P. Karranchanapramon, T.-M. Louie, I. Rech, S. Cova, L. Xun and X. S. Xie, Science 302, 262 (2003); J. Breidenbeck, J. Helbing, J. R. Kumita, G. A. Woolley and P. Hamm, Proc. Natl. Acad. Sci. USA 102, 2379 (2005).
\[ 4 \] J. Bruje, R. I. Hermans Z., K. A. Walther and J. M. Fernandez, Nature Physics 2, 282 (2006).
\[ 5 \] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA 84, 7524 (1987).
\[ 6 \] J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes, Annu. Rev. Phys. Chem. 48, 545 (1997).
\[ 7 \] J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, Proteins 21, 167 (1995).
\[ 8 \] D. Baker and D. A. Agard, Biochemistry 33, 7505 (1994); J. C. Whisstock and S. P. Bottomley, Curr. Opin. Struct. Bio. 16, 761 (2006).
\[ 9 \] A. Vitalis, X. Wang and R. V. Pappu, Biophys. J. 93, 1923 (2007).
\[ 10 \] C. Soto, Nature Rev. Neuroscience 4, 49 (2003).
\[ 11 \] F. H. Stillinger and T. A. Weber, Phys. Rev. A 25, 978 (1982).
\[ 12 \] J. D. Honeycutt and D. Thirumalai, Biopolymers 32, 695 (1992).
\[ 13 \] G. Lois, J. Blawzdziewicz and C. S. O’Hern, Biophys. J. 95, 2692 (2008).
\[ 14 \] R. M. Jendrejaczek, M. D. Graham and J. J. de Pablo (2000) J. Chem. Phys. 7, 2894-2900.
\[ 15 \] A contact is defined when the distance between two monomers is less than the separation at which the concavity of the Lennard-Jones potential changes sign.
\[ 16 \] H. Taketomi, Y. Ueda and N. Go, Int. J. Pept. Protein Res. 7, 445 (1975).
\[ 17 \] S.-I. Segawa and M. Sugihara Biopolymers 23, 2473-2488 (1984).
\[ 18 \] W. H. Wang, W. Utsumi and X.-L. Wang, Europhys. Lett. 71, 611 (2005).
\[ 19 \] We use \( \Delta c \) because the Euclidean distance between basins in configuration space is not uniquely defined.
\[ 20 \] R. Orbach, Science 313, 814 (1986); T. Nakayama and K. Yakubo, Rev. Mod. Phys. 66, 381 (1994).
\[ 21 \] G. Lois, J. Blawzdziewicz and C. S. O’Hern, Phys. Rev. Lett. 102, 015702 (2009).
\[ 22 \] R. Zwanzig, A. Szabo and B. Bagchi, Proc. Natl. Acad. Sci. USA 89, 20 (1992).