Percolation-like phase transitions in network models of protein dynamics

Jeffrey K. Weber and Vijay S. Pande^{a)}

Department of Chemistry, Stanford University, Stanford, California 94305, USA

(Received 1 December 2014; accepted 19 May 2015; published online 4 June 2015)

In broad terms, percolation theory describes the conditions under which clusters of nodes are fully connected in a random network. A percolation phase transition occurs when, as edges are added to a network, its largest connected cluster abruptly jumps from insignificance to complete dominance. In this article, we apply percolation theory to meticulously constructed networks of protein folding dynamics called Markov state models. As rare fluctuations are systematically repressed (or reintroduced), we observe percolation-like phase transitions in protein folding networks: whole sets of conformational states switch from nearly complete isolation to complete connectivity in a rapid fashion. We analyze the general and critical properties of these phase transitions in seven protein systems and discuss how closely dynamics on protein folding landscapes relate to percolation on random lattices. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4921989>]

INTRODUCTION

As the study of protein folding has progressed since the days of Anfinsen,¹ pragmatic marriages have emerged between the theory that describes protein dynamics and existing frameworks within physics. For example, simple polymer theories have provided a reference point for understanding the configurational diversity in the more complicated protein heteropolymer systems.² Frustrated equilibration kinetics within protein conformational ensembles exhibit natural connections to relaxation in molecular and spin glasses, which have been studied for decades.³ Theories of evolution, design, and optimization have been leveraged to explain how protein native states emerge from a melange of structurally non-specific conformations.⁴ More recently, network theory has been used in concert with machine learning techniques to yield quantitative descriptions of microscopic protein dynamics.⁵

Though omitted from this cadre of methods, percolation theory presents an intriguing paradigm for thinking about protein folding phenomenology. Percolation classically refers to the circuitous passage of a liquid through vacancies in a mesoporous solid.⁶ As vacancies are added at random to a lattice, the network of empty space that results evolves in an interesting fashion. In particular, the extant set of lattice vacancies switches in a rapid manner from a state of full isolation (in which all defects are disconnected) to a state of full connection (yielding a single, cavernous vacancy). This near-instantaneous event is referred to as the “percolation phase transition,” and the universal properties of this crossover have been well characterized using the theory of critical phenomena.^{7–9}

In generalized terms, the percolation problem can be constructed on a graph, with edges between nodes (added at random based on a probabilistic threshold) representing vacancies. The critical behavior near the percolation phase

transition in such general networks has been catalogued in all dimensions.^{10–13} As one approaches the thermodynamic limit in graph cardinality, the percolation transition becomes “infinitely” sharp; interesting scaling questions, including whether the crossover is truly first order or second order in nature, remain open.^{14,15}

Recently, detailed network representations for protein dynamics (built from atomistic molecular dynamics (MD) simulations) have been constructed in the form of Markov state models (MSMs).^{16–18} One wonders, given the similarities between protein folding dynamics and frustrated, glassy relaxation kinetics (both featuring potentially slow equilibration among many free energetic minima), if protein folding networks exhibit percolation-like characteristics. Can proteins, in some sense, be thought to “percolate” through their free energy landscapes in transit between unfolded and folded states, in a manner similar to liquid navigating vacancies in a mesoporous solid? Notably, percolation-like properties have been identified in folded protein structures in terms of C_α contact networks, side chain proximity networks, and fractions of hydrophobic residues;^{19–21} percolation thresholds in protein-protein interaction networks have also been described.²² To our knowledge, however, the configurational dynamics of proteins have yet to be characterized from the perspective of percolation theory. In this article, we apply a probability threshold to systematically eliminate dynamical connections in MSM networks below a set transition rate, and we analyze the critical and structural properties of the connected clusters of nodes that result. We find, as more and more prominent connections are removed, a “percolation-like” phase transition indeed occurs: the network of conformational states becomes completely disconnected over a narrow range of transition rates. More generally, we aim to interrogate the properties of protein dynamics from a fresh theoretical perspective. Do the structures of protein folding networks conform to those of random percolation networks generated on lattices? We discuss these considerations, in depth, after a basis for MSM and percolation theory has been established.

^{a)} Author to whom correspondence should be addressed. Electronic mail: pande@stanford.edu

THEORY, MODELS, AND METHODS

Markov state models of protein dynamics

In recent years, specialized computational architectures have enabled researchers to probe atomistic protein dynamics at the millisecond time scale, bringing detailed simulations of protein folding, conformational change, and protein-protein interactions within reach. The sheer volume of MD data such large-scale computations generate has created novel challenges for data analysis and interpretation. Collecting parallel trajectories on distributed platforms like Folding@home²³ offers huge advantages for sampling extensivity; however, one must derive statistical means to synthesize the disconnected, weakly dependent trajectories that comprise resulting data sets. Even ultra-long simulations conducted in serial fashion (e.g., on optimized hardware like Anton²⁴) are difficult to interpret without a rigorous framework that facilitates the identification of dominant dynamical processes.⁵

MSMs of protein dynamics have been developed to explicitly address the aforementioned analytical challenges. By treating trajectories as observables of an underlying Markov chain, one can integrate isolated conformational changes into a parametric network that describes the protein dynamical landscape as a whole. For our purposes, this practice involves parameterizing a transition probability matrix, $\mathbb{T}$, within which transitions among discrete, mesoscopic conformational states satisfy the Markov property. Indexing arbitrary states with Latin characters, the Markov constraint implies that the probability of moving from state i to j over some discrete time step, Δt , beyond the present time, t , is independent of the system's history,⁵

$$\begin{aligned} (\mathbb{T})_{ij} &= P[X(t + \Delta t) = x_j | X(t) = x_i, \\ &\quad X(t - \Delta t) = x_k, \dots, X(0) = x_l] \\ &= P[X(t + \Delta t) = x_j | X(t) = x_i], \end{aligned}$$

where each lower-case x represents a state of an upper-case random variable, X , in the Markov chain. The time step at which the memoryless assumption yields a good representation of dynamics is highly system-dependent. For proteins, this Markov time is often set at tens of nanoseconds, a time scale well-correlated with the relaxation rates of individual amino acid side chains.²⁵ The specific strategies leveraged in constructing a state space and estimating transition counts are discussed, in depth, elsewhere.⁵ Ideally, one defines a state space in which equilibration within states is fast compared to the Markov time, whereas interstate transitions are comparatively slow. In this sense, conformations within individual states should be kinetically similar; in practice, structurally similar states are assumed to be kinetically close, to varying levels of accuracy.

MSMs are thus built in discrete space and time via some schematic like that which follows: (1) MD simulations of a protein of interest are carried out to the researcher's satisfaction, (2) the configuration space explored by MD trajectories is partitioned into a set of discrete conformational states, (3) transitions among these states are counted according to a chosen Markov time, Δt , and finally, (4) the rows of the resulting transition count matrix are normalized to yield the

FIG. 1. Cartoon schematic of MSM construction process. Once an MD trajectory has been collected (or multiple trajectories), counts are estimated over a discretized state space; statistical estimators are then employed to generate a probabilistic network model of dynamics in the discrete space. Color-coded and letter-indexed configurations shown in the cartoon at bottom are representative of state structures within the NTL9 MSM analyzed in this work.

matrix $\mathbb{T}$. Figure 1 provides a cartoon illustration of the model construction process. The eigenvalues and eigenvectors of the MSM transition describe characteristic folding time scales and their defining dynamical properties; eigen- or state spaces can be readily be projected onto specific observables to facilitate comparison with standard simulation analyses and experimental results.

In effect, a protein MSM provides a framework from which new trajectories consistent with the underlying ensemble of MD simulations can be generated *ad infinitum*. Connections in the network exist in space-time, providing estimates of the probabilities (or, equivalently, rates) by which small-scale fluctuations occur between different ensembles of thermodynamic microstates. By contrast, canonical percolation networks feature connections only in space; a simple protocol for the space-time generalization of percolation analysis is described below.

Connection to percolation systems

As noted above, MSM networks differ from the graphs studied in percolation theory in one significant attribute: MSM transition matrices represent edge-weighted graphs that define transition probabilities (or transition rates) among conformational states defined in $3N$ -dimensional space. To provide a notion of binary connectivity characteristic to percolation

theory, we convert transition probability matrices into adjacency matrices as follows. Based on a fixed probability threshold, two nodes are deemed connected if an edge weight in either direction exceeds the cutoff and are considered to be disconnected otherwise. Percolation properties for this binary matrix are then evaluated based on its largest connected cluster, where connectivity is determined by a standard Tarjan algorithm.²⁶ The full distribution of cluster sizes (also determined by Tarjan's approach) is further exploited in the study of critical behavior near the phase boundary.

In percolation theory, the critical properties of the percolation threshold (i.e., phase boundary) are manifested in a set of dimension-dependent critical exponents.^{9,11} Here, we focus our analysis on the simplest of such exponents: the Fisher exponent, τ , which describes the decay of cluster sizes at the phase boundary.¹³ Close to this phase boundary, the distribution of clusters, n_s , of size s , scales characteristically with this Fisher exponent in classical percolation systems,

$$n_s \sim s^{-\tau}.$$

A convenient hyperscaling relation allows us to relate the Fisher exponent to the fractal dimension, d_f of the nascent connected cluster at the percolation threshold.^{8–10,13} Denoting the spatial (embedding) dimension for the percolation system by d , this hyperscaling relation is given by

$$d_f = \frac{d}{\tau - 1}.$$

The fractal dimension provides a notion of the relative “mass” of the critical cluster—normalizing by the cluster size, the d_f describes the effective density of edges in the rapidly forming connected component.⁹ Clusters with higher fractal dimensions appear “thick” with edges, while clusters at lower fractal dimension have an “airy” or “spindly” appearance. Since MSMs are real parametric networks (composed of hundreds or thousands of states), any properties related to phase transitions in such networks will be subject to fluctuations related to finite size effects. Accordingly, for purposes of comparison, we designate the MSM percolation-like threshold to be the probability threshold at which the system's largest connected cluster contains 50% of the model's states. Fisher exponents are computed via simple power-law fitting of the cluster size distribution at this threshold, and related fractal dimensions are determined by the hyperscaling relation above. Since MSM graphs describe dynamics in a discretized $3N$ -dimensional configuration space (where $N \gg 2$), we compare critical properties in MSMs with those at the upper percolation critical dimension ($d = 6$), beyond which such properties are invariant to hyperscaling.^{9,13}

To provide structural insight into these critical properties, we analyze secondary structural characteristics within the critical cluster and the complementary ensemble of disconnected states using the DSSP algorithm.²⁷ In order to ameliorate the effects of model construction details on our results, we analyze a variety of models at various levels of coarse-graining for each system of interest. Systematic coarse-graining of the MSMs is achieved using the BACE algorithm.²⁸ Here, we analyze percolation-like phase transitions and their critical properties in seven protein folding systems: the helix bundles α 3D, BBL,

and a λ -repressor mutant, along with the generic globular proteins BBA (beta-beta-alpha fold), NuG2 (a variant of Protein G), NTL9 (taken from Protein L9), and the Fip35 WW Domain.¹⁶ Notably, these proteins exhibit folding time scales ranging from tens to many hundreds of microseconds.^{16,29}

We should again emphasize that the connections in these protein folding MSMs are purely dynamical in nature, describing either transition rates or (more directly) discrete-time transition probabilities among sub-ensembles of protein structures. Typical percolation studies are conducted on networks in which edges convey spatial relationships between nodes; the space-time generalization of percolation theory we apply here allows one to modulate connections in trajectories that carry the protein system among unfolded, intermediate, folded, and misfolded states. The implications of such dynamical perturbations are discussed below.

RESULTS AND DISCUSSION

Percolation-like phase transitions

To formalize the concept of percolation-like phase transitions in MSMs, we first analyze the largest connected cluster as a function of a sweeping probability threshold for each the models in question. Figure 2 showcases the maximal cluster size curves that result from interrogating the most fine-grained model available for each protein. The probability cutoff is rigorously defined in the space of transition probabilities; roughly speaking, however, corresponding microscopic

FIG. 2. Curves representing the size, $s_{\max}$, of largest connected cluster (normalized to graph cardinality) as a function of the probability threshold for seven protein folding MSMs. Rough microscopic rates associated with the dynamical transition probability thresholds are shown at top. The sharp crossovers between fully connected and fully disconnected regimes represent percolation-like phase transitions. The curves suggest that a clear rate threshold exists in protein folding dynamics that delineates which microscopic processes are strictly necessary for traversal of the free energy landscape. For the models studied here, this threshold falls in the range of 5–10 transitions/ μ s.

transition rates can be estimated by dividing probabilities by the MSM lag time. Such estimates, reported for simplicity, at a fixed Markov time of 10 ns, are included in the figure.

As Figure 2 indicates, we observe relatively sharp cross-overs between fully connected and disconnected regimes of the networks; these abrupt jumps typify the percolation-like phase transitions we observe on protein folding landscapes. Though some variance obviously exists among the seven models, the percolation thresholds tend to fall between a probability cutoff of 0.05 and 0.1—a threshold that extricates microscopic processes that occur fewer than 5–10 times/ μ s. In the region just below the 0.05 threshold (highlighted in blue), all seven networks become completely connected. Connections related to microscopic fluctuations that occur fewer than about 5 times/ μ s (at the extreme, fewer than about 2 times/ μ s), thus, are not strictly necessary for traversing the entire folding landscape. This observation suggests that while relatively rare fluctuations occur in protein dynamics (and are observed constantly in molecular dynamics data), only the strong edges are required to complete the full protein folding process. Presumably, such heavily weighted edges, corresponding to fast microscopic rates, have been highly optimized by evolution. While rarer fluctuations can occasionally be productive to folding (and serve to speed up the assembly process), these slower dynamical events—cutoff at a very characteristic time scale across systems—do not seem requisite to proper self-assembly.

The observed threshold for full connectivity—coincident among seven protein folding landscapes—falls close in time scale to the theorized “speed limit” for protein folding, which is set, by physical constraints, to near 500 ns for a 50-residue, single-domain system.³⁰ While the proteins studied here fold, in total, at time scales much slower than the supposed speed limit, the notion that all required microscopic “hops” along folding pathways have been optimized to be at or faster than the limit is intriguing.

The definition of “microscopic process” and the parameterization of related edge weights, of course, depend on the methods used in MSM construction. The curves shown in Figure 3 provide a simple control over potential biases that might arise from specific choices in model construction (like the clustering method, the number and size of states, and the lag time). Using a Bayesian approach, we have systematically coarse-grained each model to yield series of networks containing fewer and larger states.²⁸ The data in Figure 3 correspond only to BBA (for MSMs ranging from 200 to 600 states), but the plot suggests that maximal cluster curves are largely consistent over this range of models. To emphasize the shape of these curves, the data are aligned such that the percolation thresholds are matched. Based on finite-size scaling arguments, one would expect fewer-state networks to exhibit broader transition regions; we indeed see broader crossovers in the coarser models, and (concomitantly) we observe that percolation thresholds are shifted to somewhat lower probabilities. Still, all such crossovers in BBA occur between probability cutoffs of 0.04 and 0.08—and dominant clusters become fully connected with rate thresholds above $1/\mu$ s. The prevailing features of percolation-like transitions in proteins, therefore, are largely conserved upon coarse-graining.

FIG. 3. Overlay of dominant connected cluster curves for various coarse-grained versions of the BBA MSM. Curves are shifted such that percolation thresholds (where the dominant cluster size, $s_{\max}$ is 0.5) are aligned. Models constructed with the BACE algorithm range from 200 to 600 states in size at an interval of 20 states.²⁸ The plot illustrates that dynamical thresholds for landscape traversal are similarly sharp over a wide range of MSM resolution.

Structural characteristics of percolation clusters

Another topic of interest concerns how proteins are structured within and outside their respective (largest) critical clusters. As a null hypothesis, one might predict that conformations would be more structured within the dominant cluster; after all, one might reasonably expect less ordered, far-from-native states to be poorly connected on the folding landscape. Figure 4 provides a direct test of this hypothesis: mean aggregate α -helix, β -sheet, and overall secondary structural assignments are plotted for states within and outside the dominant critical clusters of the seven protein systems. The bars on the data points represent ± 1 standard deviation from the mean for each of the structural ensembles.

One sees immediately that states inside and excluded from the dominant clusters have remarkably similar structural characteristics, on average. If anything, member states have more β -related content and disconnected states contain more helices—but these observations are both subject to the notable exception of NTL9. These observations speak to the marked heterogeneity of rates among structured states on these protein folding landscapes. The native states of all seven proteins are in fact contained within their respective dominant critical clusters, but the conformations excluded from this strongly connected network still exhibit significant collapsed structural characteristics of their own.

“Sub-critical” percolation dynamics

Noting the interesting properties of the dominant critical cluster, we now proceed to scrutinize the properties of entire cluster size distributions near the percolation threshold. In Figure 5, we present Fisher exponents for all seven systems over a wide range of coarse-grained models; in

FIG. 4. Comparisons of secondary structural characteristics within and outside dominant critical clusters, detailing (left:) average α -helical content, (center:) average β -sheet content, and (right:) average total secondary structural content. Helices were defined by the “H” designation in the DSSP nomenclature, while residues in “B” or “E” designations were counted as sheets. All DSSP designations were included in the total accounting of secondary structure.²⁷ Data for critical cluster states are indicated in blue, while data for excluded states are shown in red. The bars on each data point represent ± 1 standard deviation within the structural ensemble.

a complementary fashion, we display corresponding fractal dimensions calculated from the hyperscaling relation at the upper critical dimension for percolation. Notably, the method for determination of Fisher exponents is agnostic to spatial dimension (as such exponents are determined by power-law fitting of the cluster size distribution). Fisher exponents for percolation at lower dimensions are smaller than the upper dimension limit presented ($\tau = 5/2$).¹³

One observes that, for the vast majority of models, Fisher exponents exceed the percolation limit by a wide margin, meaning that cluster size distributions decay much more quickly than one would expect on a standard percolation lattice. This result implies that, from the standpoint of classical percolation systems, MSMs of protein folding dynamics are sub-critical: the decay of cluster sizes at the phase boundary is accelerated and closer to exponential in nature. In other words, the states excluded from the dominant critical cluster are themselves exceptionally disconnected from other states in protein folding MSMs. Again, several models do offer exceptions to this rule: the cluster sizes for λ -repressor, in particular, fall close to the canonical distribution, and several of

the most coarse-grained models fall below the upper dimension limit. The behavior of the Fisher exponent as a function of model size is very system dependent: some models approach the percolation limit as the number of states is reduced, while others exhibit the opposite (or oscillating) behavior. Also, little to no correlation among Fisher exponents and simple characteristics like chain length, folding time, and contact order is observed. Overwhelmingly, however, the Fisher exponents land in the sub-critical regime.

The complementary values for the fractal dimension, of course, also describe largely sub-critical behavior.^{7,31} As mentioned above, small values of the fractal dimension imply an anomalously low density of connections in the dominant critical cluster. If visualized, therefore, the largest critical clusters in MSMs would appear “airy” or “spindly” compared to those for comparable percolation systems, as such MSM clusters are themselves more sparsely connected.

Considered in concert, these observations suggest that a representation of protein folding networks as random graphs (with edges laid down with fixed probability) is often not an appropriate one. Perhaps unsurprisingly, strong edges between

FIG. 5. Fisher exponents (left) and related fractal dimensions (right) for seven protein folding systems, calculated over ranges of coarse-grained MSMs. Models for each protein were coarse-grained until fits of critical properties could no longer be reliably computed. Dashed lines indicate the corresponding values for each property in classical percolation systems, at the upper critical dimension for percolation.

conformational states seem to be laid down more judiciously on protein free energy landscapes; one might suppose that some type of evolutionary design procedure supersedes pure stochasticity to limit the number of probable microscopic pathways in protein folding networks. A number of systems are closer to the percolation limit than others. Particularly in the case of the λ -repressor mutant, one could expect to construct similar critical behavior using a random percolation lattice. This observation signifies that λ -repressor dynamics are in some sense more heterogeneous than dynamics in other proteins studied. Exactly how specific protein sequences encode for more or less heterogeneous (i.e., percolation-like) behavior remains an interesting question. Simple native structural features alone seem to be incapable of providing an answer: helix bundles, for instance, appear both close to and far from the percolation limit. The dominant sub-critical behavior we do see, however, demonstrates why connectivity-based phase transitions in protein dynamics should be described as “percolation-like.” We do see sharp crossovers in connectivity that define characteristic rates for microscopic dynamics on protein folding landscapes. The distribution of connections around these crossovers though is much sparser than randomly placed edges would dictate.

CONCLUSION

In this article, we have demonstrated that abrupt jumps in connectivity do appear on protein folding landscapes when rare dynamical fluctuations are systematically suppressed. In some sense, therefore, protein dynamics can be likened to percolation on a lattice. Our analysis of critical properties, however, suggests that vacancies on the protein folding “lattice” do not appear at random: evolution is no doubt selective in opening and closing avenues through which protein dynamics can flow. In this vein, it is important to note that, while fluctuations excluded by the percolation-like threshold do not enhance connectivity on protein folding networks, all such fluctuations cannot be ignored. At equilibrium, these rare dynamical events might occasionally help or hinder productive processes to little beneficial or deleterious effect. Living systems have evolved to operate not in stasis, though, but out of equilibrium; it is out of equilibrium (under glassy or driven conditions) that seemingly rare fluctuations can take center stage.

ACKNOWLEDGMENTS

We thank the NSF (Grant No. MCB-0954714) and NIH (Grant No. R01-GM062868) for their support of this work. J.K.W. was supported by the Fannie and John Hertz Foundation on the endowed Professor Yaser S. Abu-Mostafa Fellowship.

- ¹C. B. Anfinsen *et al.*, *Science* **181**, 223 (1973).
- ²P. Flory, *J. Chem. Phys.* **66**, 5720 (1977).
- ³J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7524 (1987).
- ⁴E. Shakhnovich and A. Gutin, *J. Phys. A: Math. Gen.* **22**, 1647 (1989).
- ⁵V. S. Pande, K. Beauchamp, and G. R. Bowman, *Methods* **52**, 99 (2010).
- ⁶R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- ⁷M. Aizenman and D. J. Barsky, *Commun. Math. Phys.* **108**, 489 (1987).
- ⁸C. L. Henley, *Phys. Rev. Lett.* **71**, 2741 (1993).
- ⁹A. Kammerer, F. Höfling, and T. Franosch, *EPL* **84**, 66002 (2008).
- ¹⁰J. Essam, *J. Phys. A: Math. Gen.* **22**, 4927 (1989).
- ¹¹H. Ballesteros, L. Fernández, V. Martin-Mayor, A. M. Sudupe, G. Parisi, and J. Ruiz-Lorenzo, *J. Phys. A: Math. Gen.* **32**, 1 (1999).
- ¹²N. Jan and D. Stauffer, *Int. J. Mod. Phys. C* **9**, 341 (1998).
- ¹³G. Paul, R. M. Ziff, and H. E. Stanley, *Phys. Rev. E* **64**, 026115 (2001).
- ¹⁴D. Achlioptas, R. M. D’Souza, and J. Spencer, *Science* **323**, 1453 (2009).
- ¹⁵W. Chen and R. M. D’Souza, *Phys. Rev. Lett.* **106**, 115701 (2011).
- ¹⁶K. A. Beauchamp, R. McGibbon, Y.-S. Lin, and V. S. Pande, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17807 (2012).
- ¹⁷G. R. Bowman, V. A. Voelz, and V. S. Pande, *J. Am. Chem. Soc.* **133**, 664 (2010).
- ¹⁸V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, *J. Am. Chem. Soc.* **132**, 1526 (2010).
- ¹⁹D. Deb, S. Vishveshwara, and S. Vishveshwara, *Biophys. J.* **97**, 1787 (2009).
- ²⁰K. V. Brinda, S. Vishveshwara, and S. Vishveshwara, *Mol. Biosyst.* **6**, 391 (2010).
- ²¹J. Miao, J. Klein-Seetharaman, and H. Meirovitch, *J. Mol. Biol.* **344**, 797 (2004).
- ²²C.-S. Chin and M. P. Samanta, *Bioinformatics* **19**, 2413 (2003).
- ²³M. Shirts and V. S. Pande, *Science* **290**, 1903 (2000).
- ²⁴D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan *et al.*, *Science* **330**, 341 (2010).
- ²⁵K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.* **78**, 1950 (2010).
- ²⁶R. Tarjan, *SIAM J. Comput.* **1**, 146 (1972).
- ²⁷W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
- ²⁸G. R. Bowman, *J. Chem. Phys.* **137**, 134111 (2012).
- ²⁹K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, *Science* **334**, 517 (2011).
- ³⁰J. Kubelka, J. Hofrichter, and W. A. Eaton, *Curr. Opin. Struct. Biol.* **14**, 76 (2004).
- ³¹M. V. Menshikov, *Dokl. Akad. Nauk SSSR* **288**, 1308 (1986).