PHYSICAL REVIEW E **94**, 050402(R) (2016)

Characterizing folding funnels with replica exchange Wang-Landau simulation of lattice proteins

Guangjie Shi,[1] Thomas Wüst,[2] and David P. Landau[1]
[1]*Center for Simulational Physics, The University of Georgia, Athens, Georgia 30602, USA*
[2]*Scientific IT Services, ETH Zürich, 8092 Zürich, Switzerland*
(Received 30 June 2016; published 15 November 2016)

We have studied the folding of ribonuclease A by mapping it onto coarse-grained lattice protein models. With replica exchange Wang-Landau sampling, we calculated the free energy vs end-to-end distance as a function of temperature. A mapping to the famous hydrophobic-polar (HP) model shows a relatively shallow folding funnel and flat free energy minimum, reflecting the high degeneracy of the ground state. In contrast, extending the HP model with an additional "neutral" monomer type (i.e., a mapping to the three-letter H0P model) has a well developed, rough free energy funnel with a low degeneracy ground state. In both cases, folding funnels are asymmetric with temperature dependent shape.

DOI: 10.1103/PhysRevE.94.050402

Understanding protein folding remains a grand challenge problem of modern science [1]. The resolution of Levinthal's paradox concerning the ability of proteins to fold rapidly postulates the existence of a rough, "folding funnel" in free energy space that "guides" the protein to its lowest free energy, native state [2–4]. The funnel is always portrayed schematically as a relatively symmetric function of some unknown reaction coordinate about a unique minimum (the native state), as shown in Fig. 1. However, apart from a few conceptual studies with very short chain lengths [5,6], the folding funnel has never actually been observed for a realistically large model of a protein due to the difficulty of (sufficiently) effective conformational-space sampling.

Coarse-grained protein models have played an important role in the study of protein folding via computer simulations (e.g., [7–14]) and have also been used to examine the competition between folding, adsorption on surfaces, and fibril formation [15–18]. In this work, we use Monte Carlo simulations of hydrophobic-polar (HP) and H0P lattice protein models to reveal clear, folding funnel pictures with a simple structural quantity serving as the relevant reaction coordinate.

The classic, HP lattice protein model [19,20] classifies amino acids into only two groups: hydrophobic (H) and polar (P) based on the properties of their side chains. It captures the hydrophobic interactions through an attractive interaction ϵ_{HH} between nonbonded neighboring H-mers, and there are no interactions between other nonbonded neighbor pairs. The HP model greatly simplifies the protein folding problem and allows special algorithms to be employed that take advantage of the discrete degrees of freedom; however, the ground states (native states) of HP model proteins are generally highly degenerate, unlike real proteins. (While sacrificing full atomic resolution, the HP model reduces the effective number of monomers by roughly a factor of 10, eliminates uncertainty about force fields, and allows acceleration in the sampling of phase space by many orders of magnitude.) To reduce the degree of coarseness, an extension of the HP model has been introduced, the semiflexible H0P model [21,22], that includes not only H and P monomers but also a type "0" monomer that is neutral in terms of hydrophobicity as well as the stiffness of the bond angle, ϵ_θ, as observed in real proteins. The general Hamiltonian is

$$\mathcal{H} = -\epsilon_{\mathrm{HH}} n_{\mathrm{HH}} - \epsilon_{\mathrm{H0}} n_{\mathrm{H0}} - \epsilon_{00} n_{00} - \epsilon_\theta n_\theta, \tag{1}$$

where the ϵ's are the energies corresponding to the number, n, of each kind of "bond." These modifications retain the essential simplicity of the HP model but reduce the degeneracy of the ground state by orders of magnitude. (Of course, other coarse-grained models have also expanded the number of amino acid types; e.g., [18,23].)

In order to represent the free energy landscape, i.e., folding funnel, the reaction coordinate has to be chosen carefully. After initial exploration, we found a simple structural quantity, the end-to-end distance, to be an effective choice. It can be calculated as

$$r_{ee} = |\vec{r}_1 - \vec{r}_{\mathrm{N}}|, \tag{2}$$

where $\vec{r}_1$ and $\vec{r}_{\mathrm{N}}$ are the coordinate vectors of the first and last monomers, respectively, in the chain of length N. The free energy can then be formulated as a function of the end-to-end distance and the temperature:

$$F\,(T,r_{ee}) = -k_B T\ \ln Z(T,r_{ee}), \tag{3}$$

where $Z(T,r_{ee})$ is the partition function based on both variables:

$$Z(T,r_{ee}) = \sum_E g(E,r_{ee}) e^{-E/k_B T}, \tag{4}$$

and $g(E,r_{ee})$ is the two-dimensional density of states which can be determined by Monte Carlo simulations.

The protein free energy landscape is complex and becomes effectively impossible to sample at low temperature in the canonical ensemble. In addition, entropic entanglements further complicate sampling efforts, and the study of even relatively short HP proteins is an extraordinarily challenging problem in statistical physics, computer science, statistics, and biochemistry (see Wüst and Landau [13], and references therein). To efficiently characterize the free energy of these lattice protein models we adopted the replica exchange Wang-Landau (REWL) sampling method [24,25]. This algorithm is a parallel extension of the serial Wang-Landau (WL) method [26,27] and is an iterative, efficient, and robust way to estimate the density of states. Based on splitting the

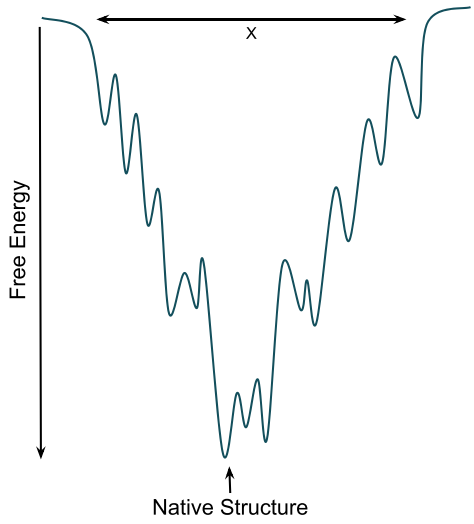

FIG. 1. Schematic view of a rough protein folding funnel vs some undefined reaction coordinate "x."

energy into overlapping windows, REWL simulations have the proven ability to reach previously inaccessible domains, and great scalability with the number of computing cores. With pull moves and bond-rebridging moves implemented, the REWL and traditional WL sampling algorithms have proven to be highly efficient for investigating lattice protein models [12,13,25].

In this work, we investigated a real protein, ribonuclease A, by first mapping it onto a 124 monomer, coarse-grained, three-dimensional (simple cubic) lattice HP model (HP124) [28] based on the hydrophobic index of each amino acid [29]; there are 47 H-mers and 77 P-mers. For HP124, the coupling constant $\epsilon_{\mathrm{HH}} = 1$ and is zero for the rest. We employed REWL

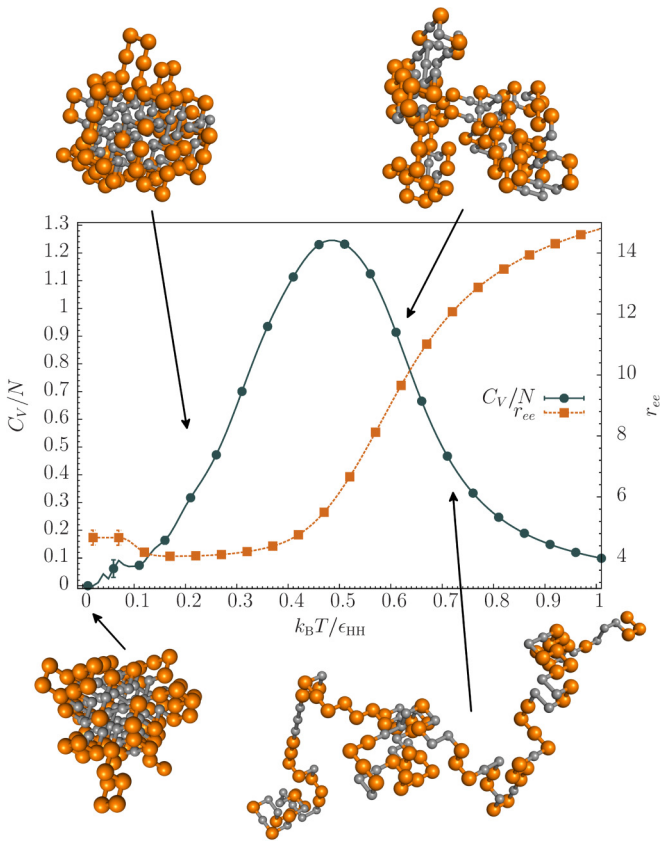

FIG. 2. The specific heat and end-to-end distance vs temperature for the HP124 lattice protein described in the text. Typical configurations are shown at the indicated temperatures: Hydrophobic monomers are colored dark gray, while polar monomers are colored orange. Error bars smaller than the data points are not shown.

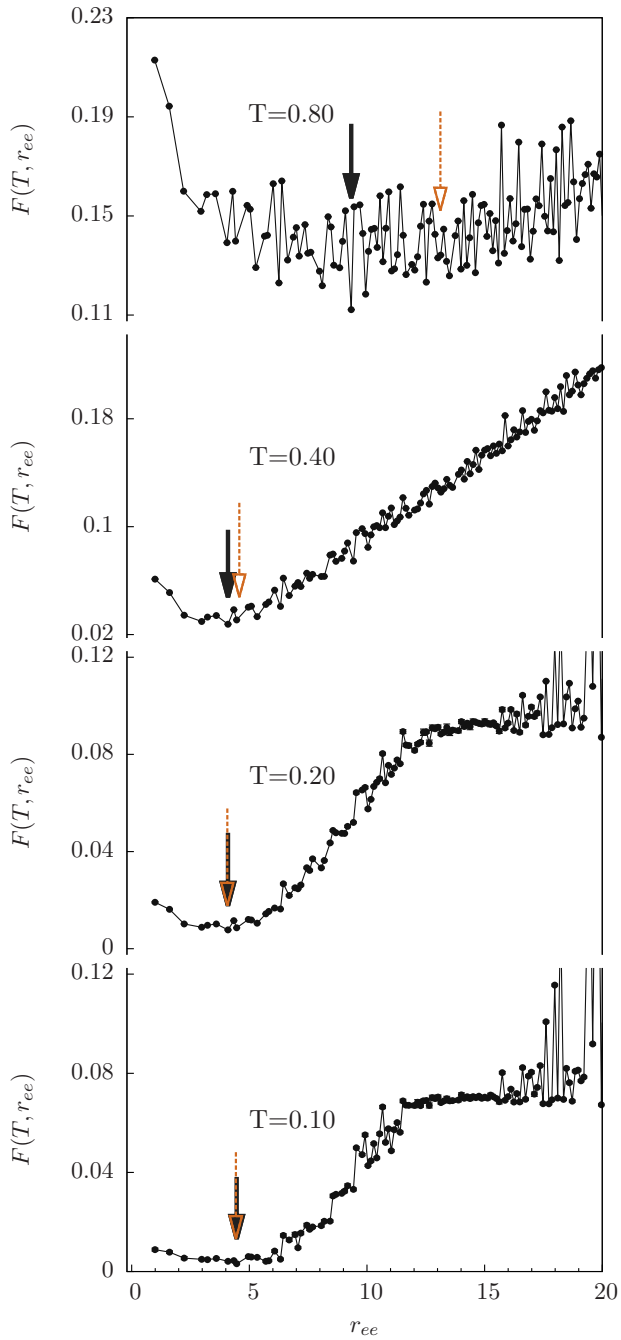

FIG. 3. Normalized free energy vs end-to-end distance at four different temperatures for the HP124 lattice protein. Black, filled arrows indicate the lowest free energy at each temperature; orange arrows show the mean end-to-end distance at that temperature. Error bars are smaller than the data points.

for determining the density of states, $\tilde{g}(E)$, of HP124 to high precision. These were followed by production runs, with $\tilde{g}(E)$ held fixed, and only the two-dimensional histogram $H(T,r_{ee})$ is then updated throughout the simulations. By reweighting the two-dimensional histogram, we could obtain two-dimensional density of states $\tilde{g}(E,r_{ee}) = \tilde{g}(E)H(T,r_{ee})$, which is the key for calculating the free energy as in Eqs. (3) and (4).

As seen in Fig. 2 the specific heat and end-to-end distance for HP124 both show a clear protein collapse "transition" near $T \approx 0.5$ followed by a very slight "bump" at quite low T. Typical protein configurations in Fig. 2 show this folding process including one of the degenerate ground states.

The free energy vs end-to-end distance at various temperatures is calculated according to Eqs. (3) and (4) and results are shown in Fig. 3. The free energy curves contain many local maxima and minima at all temperatures. These variations in free energy are significant since statistical errors in the results are smaller than the size of the symbols. The lowest free energy state is indicated by a filled, black arrow, while the mean end-to-end distance is marked by an orange arrow. At high temperature the behavior shows a shallow, "symmetric" but quite rough landscape. Upon lowering the temperature, we found that the free energy forms a clear, funnel-like structure that is skewed toward the region with low end-to-end distance values. Schematic portrayals of the protein folding funnel always present a static structure that simply guides the protein towards a fixed minimum as the temperature is lowered. Instead, we find that the lowest free energy position shifts with the change of temperature, indicating a dynamic, instead of static, nature of the folding funnel. At lower temperatures, the free energy landscape becomes relatively flat near the minimum and oddly shaped for large end-to-end distance. The relative smoothness means that the system can easily move between states, i.e., small changes in end-to-end distance do not result in significant differences in the free energy. When $T < 0.2$, the point where the free energy is

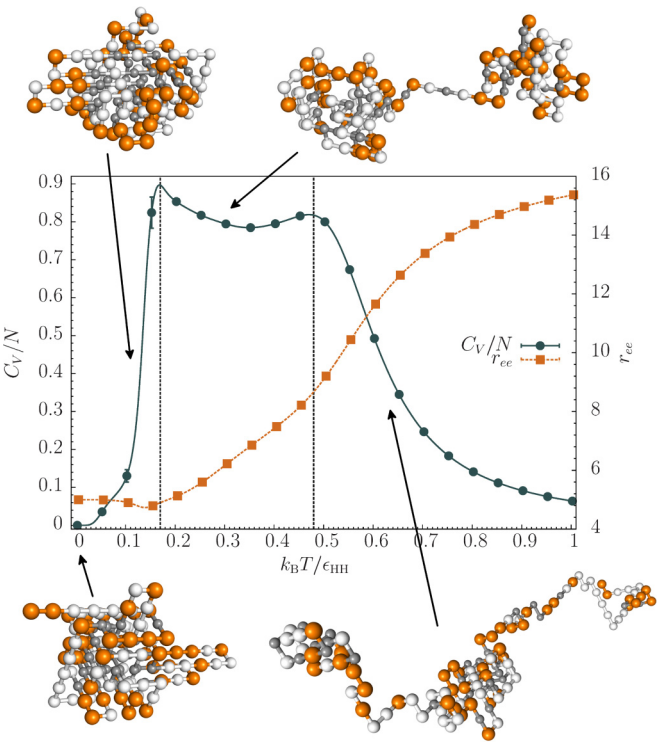

FIG. 4. The specific heat and end-to-end distance vs temperature for the H0P124 lattice protein described in the text. For structures shown, H- and "0"-mers are colored dark-gray and white, respectively, while P-mers are colored orange. Error bars smaller than the data points are not shown.

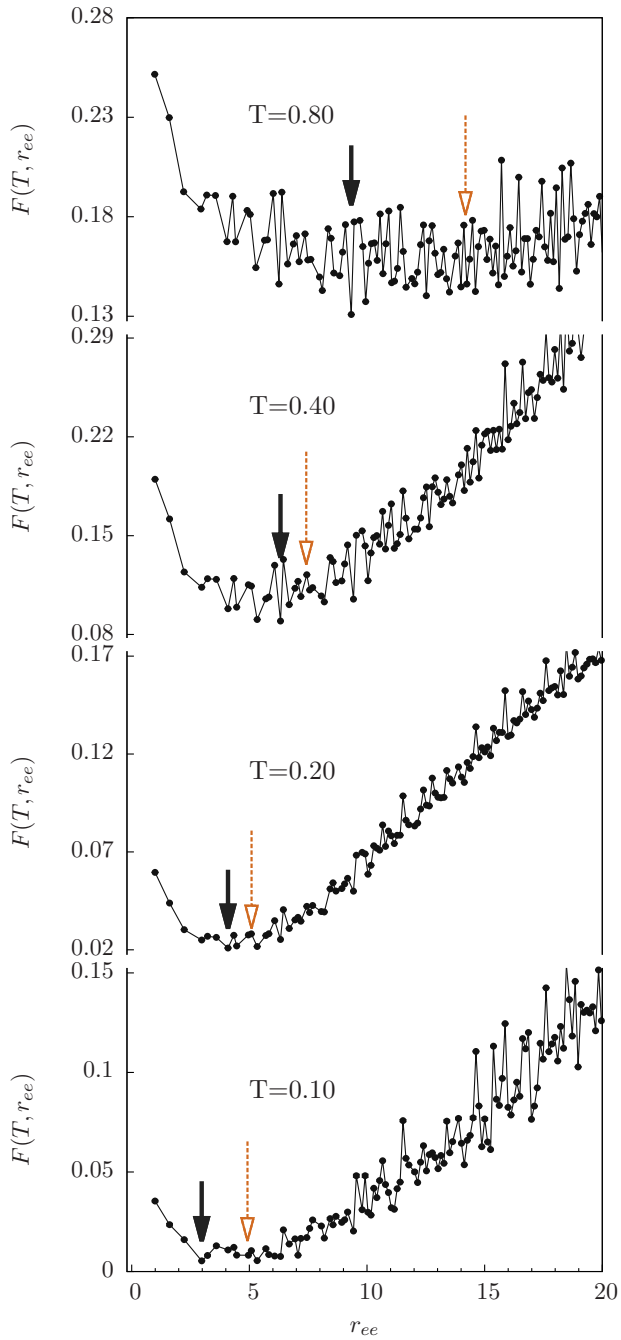

FIG. 5. Normalized free energy vs end-to-end distance at four different temperatures for the H0P124 lattice protein. Black, filled arrows indicate the lowest free energy at each temperature; orange arrows point to the mean end-to-end distance at that temperature. Error bars are smaller than the data points.

lowest coincides with the mean end-to-end distance. Lastly, through the method described in Ref. [30], we determined the ground-state degeneracy for HP124 to be $\sim 1.4 \times 10^6$, i.e., far above the "desired" unique native state.

In the "improved" model (H0P124) [21,22], the number of monomers for H, "0," and P types are, respectively, 29, 50, and 45, and the coupling constants used were $\epsilon_{\mathrm{HH}} = 1$, $\epsilon_{\mathrm{H0}} = 0.5$, $\epsilon_\theta = -0.25$. All other interactions were zero. (The exact choice of interactions is arbitrary, but ϵ_{HH} should dominate to emphasize the formation of a hydrophobic core in the folded state.) The lattice protein collapses from a random coil at a temperature $T \approx 0.5$; however, when T is lowered to 0.2, H0P124, in contrast to HP124, exhibits a second peak in the specific heat. Typical configurations show that the first "transition" is only partial and the protein collapses completely only at the lower temperature. At very low temperature there is only a slight shoulder in the specific heat but the end-to-end distance still increases slightly (Fig. 4).

The free energy vs end-to-end distance at various temperatures for H0P124 is shown in Fig. 5. The state with lowest free energy is indicated by a filled black arrow, while the mean end-to-end distance is marked by an orange arrow. At $T = 0.8$, we found a fairly shallow but rugged free energy landscape, which is similar to that for HP124. With decreasing temperature the free energy skews clearly toward the region with low end-to-end distance values, but at low T the funnel remains rough, even near the bottom. Using the method described in Ref. [30], we determined the ground-state degeneracy for H0P124 was greatly reduced and only 425 inequivalent ground states were found, i.e., a reduction of more than *four orders of magnitude* as compared to the HP model. This characteristic is much closer to what is expected for a real protein. However, the shift of the lowest free energy position with temperature, indicates a dynamic, instead of the usually depicted static, rugged folding funnel.

Although both lattice protein models possess complex, funnel-like free energy landscapes, clear differences exist between them. For HP124 (Fig. 3) at lower temperatures, the free energy curve is oddly shaped and relatively flat near the bottom, whereas the entire funnel remains rugged for H0P124. When $T < 0.2$, the position of the free energy minimum for HP124 coincides with the mean end-to-end distance, but for H0P124, even at low T, the free energy landscape remains rough near the minimum. For example, the free energy barrier preventing escape from the second lowest state (r_{ee} between 5 and 6) is approximately $7k_BT$, and this state is not even immediately adjacent to the lowest free energy state. Moreover, the lowest free energy is clearly separated from the averaged end-to-end distance and the protein can easily become trapped in a local minimum. Whereas the specific heat shows only two major events, the mean end-to-end distance changes often with temperature. This indicates that folding occurs through a series of small rearrangements that gives rise to two major configurational changes. For both models the density of states, $g(E)$, is smooth, even as the energy approaches its minimum. As a consequence, schematic representations of the funnel with a width given by a multivalued function of the entropy (see, e.g., Wolynes *et al.* [31]) are inconsistent with the actual behavior of the lattice proteins.

In summary, we uncovered folding funnels for two lattice protein models that are mapped from the protein ribonuclease A. The HP model has a relatively shallow free energy minimum, reflecting the high ground state degeneracy, while the H0P model develops a clear, rough free energy funnel with a relatively low degeneracy ground state. Unlike the schematic figures often seen in the literature, our extensive simulations of realistically-sized protein models reveal asymmetric folding funnels that change shape substantially as the temperature decreases. The high precision of our results reveals that even the locations of the free energy minima shift with temperature. This dynamic nature of the folding funnel with respect to temperature alters our perception of this fundamental concept. While the HP and H0P models are simplified descriptions of a real protein, neither the mapping nor the interactions were tuned to produce a special free energy structure. We thus believe that the general characteristics of the folding funnels found in our study (particularly for the H0P model) will persist in a more realistic description of protein folding.

This work was supported in part by NSF under Grant No. OCI-0904685. Computing resources were partially provided by the Georgia Advanced Computing Resource Center and the Texas Advanced Computing Center under XSEDE Grant No. PHY130014.

[1] K. A. Dill and J. L. MacCallum, Science **338**, 1042 (2012).
[2] K. A. Dill and H. S. Chan, Nat. Struct. Mol. Biol. **4**, 10 (1997).
[3] J. N. Onuchic and P. G. Wolynes, Curr. Opin. Struct. Biol. **14**, 70 (2004).
[4] K. A. Dill, S. B. Ozkan, T. R. Weikl, J. D. Chodera, and V. A. Voelz, Curr. Opin. Struct. Biol. **17**, 342 (2007).
[5] H. S. Chan and K. A. Dill, J. Chem. Phys. **100**, 9238 (1994).
[6] A. B. Oliveira, Jr., F. M. Fatore, F. V. Paulovich, O. N. Oliveira, Jr., and V. B. P. Leite, PLoS One **9**, 1 (2014).
[7] J. L. Zhang and J. S. Liu, J. Chem. Phys. **117**, 3492 (2002).
[8] M. Bachmann and W. Janke, Phys. Rev. Lett. **91**, 208105 (2003).
[9] S. Schnabel, M. Bachmann, and W. Janke, Phys. Rev. Lett. **98**, 048103 (2007).
[10] S. Schnabel, M. Bachmann, and W. Janke, J. Chem. Phys. **126**, 105102 (2007).
[11] J. Zhang, S. C. Kou, and J. S. Liu, J. Chem. Phys. **126**, 225101 (2007).
[12] T. Wüst and D. P. Landau, Phys. Rev. Lett. **102**, 178101 (2009).
[13] T. Wüst and D. P. Landau, J. Chem. Phys. **137**, 064903 (2012).
[14] J. Liu, B. Song, Y. Yao, Y. Xue, W. Liu, and Z. Liu, Phys. Rev. E **90**, 042715 (2014).
[15] A. Swetnam and M. P. Allen, Phys. Rev. E **85**, 062901 (2012).
[16] B. Pattanasiri, Y. W. Li, D. P. Landau, T. Wüst, and W. Triampo, J. Phys.: Conf. Ser. **402**, 012048 (2012).
[17] Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. E **87**, 012706 (2013).

[18] R. Ni, J. M. Kleijn, S. Abeln, M. A. Cohen Stuart, and P. G. Bolhuis, Phys. Rev. E **91**, 022711 (2015).
[19] K. A. Dill, Biochemistry **24**, 1501 (1985).
[20] K. F. Lau and K. A. Dill, Macromolecules **22**, 3986 (1989).
[21] G. Shi, T. Wüst, Y. W. Li, and D. P. Landau, J. Phys.: Conf. Ser. **640**, 012017 (2015).
[22] G. Shi, A. K. C. Farris, T. Wüst, and D. P. Landau, J. Phys.: Conf. Ser. **686**, 012001 (2016).
[23] D. Thirumulai and Z. Guo, Biopolymers **35**, 137 (1995).
[24] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013).
[25] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. E **90**, 023302 (2014).
[26] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[27] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).
[28] E. E. Lattman, K. M. Fiebig, and K. A. Dill, Biochemistry **33**, 6158 (1994).
[29] J. Kyte and R. F. Doolittle, J. Mol. Biol. **157**, 105 (1982).
[30] G. Shi, T. Vogel, T. Wüst, Y. W. Li, and D. P. Landau, Phys. Rev. E **90**, 033307 (2014).
[31] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).