PHYSICAL REVIEW E **88**, 040702(R) (2013)

Energy landscape and dynamics of proteins: An exact analysis of a simplified lattice model

Marek Cieplak[1,*] and Jayanth R. Banavar[2,†]
[1]*Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland*
[2]*Department of Physics, University of Maryland, College Park, Maryland 20742, USA*
(Received 9 July 2013; published 22 October 2013)

We present the results of exact numerical studies of the energy landscape and the dynamics of a 12-monomer chain with contact interactions encoding the ground state on a square lattice. In spite of its simplicity, the model is shown to exhibit behavior at odds with the standard picture of proteins.

DOI: 10.1103/PhysRevE.88.040702 PACS number(s): 87.10.Hk, 87.15.A−, 87.15.hj

Proteins are amazing molecular machines and serve as the workhorses of living cells. In spite of their simple linear chain topology, the development of a theoretical framework for understanding proteins is a daunting challenge; this is because of the distinct chemistries and geometries of the side chains of naturally occurring amino acids and the essential role played by the solvent molecules within the cell. Tremendous progress has been made in recent years through borrowing simple concepts and techniques from physics. In particular, ideas from the spin glass field, such as the principle of minimal frustration [1] and the principle of maximum compatibility, as exemplified by the highly studied Go model [2], have resulted in a framework for describing the folding landscape of a protein [3,4]. Also, the concept of a two-state system (corresponding to folded and unfolded conformations; the simplest fast folding proteins are often found to be unfolded or folded and rarely in a partially folded state) along with a transition state between them has proven to be useful as a guide to experiments and their theoretical interpretation [5–9]. Finally, from a computational point of view, exact studies of simple lattice models of proteins [10–15] have yielded invaluable insights.

Here we report the results of exact thermodynamic and dynamical analyses of a Go model on a lattice and show, quite surprisingly, that its behavior is very different from the standard picture of proteins. The dynamics of folding follows the treatment of Cieplak *et al.* [16] (see also [17,18]) and is coupled with an exact analysis of the conformational space describing the energy landscape. Our results suggest one of two possibilities. One scenario is that our findings regarding the nature of the energy landscape and the dynamics within it change dramatically on considering more complex models, leading to the emergence of the expected folding funnel and two-state behavior. The more likely scenario is that it may be much too simplistic to interpret experimental results using the prevailing paradigms based on the concepts of the folding landscape and the transition state theory.

The system studied is a lattice polymer defined in Fig. 1 on a square lattice. It has contact interactions with energy of −1 between certain pairs of beads when they are neighbors. We consider a Go-like model in which just six pairs of beads, 1–8, 2–7, 3–6, 5–12, 6–11, and 7–10, are attractive and all other interactions corresponding to non-native contacts are absent. Such an artificial contact Hamiltonian that encodes the desired native state is meant to be a caricature of the extreme limiting case of minimal frustration, which is then supposed to lead to a folding funnel and ideal two-state behaviors.

The ground state is nondegenerate, has an energy of −6, corresponding to all native contacts being formed, and is an S-shaped 3×2 conformation denoted as NAT in Fig. 1. Following Chan and Dill [11], the possible kinetic moves involve jumps in the orientation of the terminal bonds, kink switches (diagonal moves of a bead connected by two bonds that are set at a right angle), and two-bead crankshaft turns (a mirror-like reflection of the U-shaped pieces of the chain). The attempt rate for the latter is taken to be four times as big as the attempt rate for the single-bead moves [13]. In addition to NAT, there are 15 036 other conformations. Five of these are simple energy minima. The lowest among them are denoted by A and B; they correspond to an energy of −4 and are also maximally compact. Getting out of these states by means of the allowed moves requires input of energy to break one or more of the bonds. There are also a number of extended minima. These are sets of states: the states within a set are accessible to each other without any change in energy, but getting to any other conformation necessarily requires a supply of energy. The lowest extended minima are denoted C and D.

The exact time evolution of the system is governed by the master equation [19]

$$\frac{dP_\alpha}{dt} = -\sum_\beta M_{\alpha\beta} P_\beta, \tag{1}$$

where $P_\alpha = P_\alpha(t)$ is the probability of finding the sequence in conformation α, of energy E_α, at time t. The matrix elements of M are given by

$$M_{\alpha\beta} = -w_{\alpha\beta} \leqslant 0 \quad \text{if} \quad \alpha \neq \beta, \quad M_{\alpha\alpha} = \sum_{\beta\neq\alpha} w_{\beta\alpha}, \tag{2}$$

where $w_{\alpha\beta} = w(\beta \to \alpha)$ is the transition rate from conformation β to conformation α. We take $w_{\alpha\beta} = w^{(1)}_{\alpha\beta} + w^{(2)}_{\alpha\beta}$, where

$$w^{(\sigma)}_{\alpha\beta} = \frac{1}{\tau} R_\sigma \left[1 + \exp\left(\frac{E_\alpha - E_\beta}{T}\right)\right]^{-1}, \tag{3}$$

with $R_1 = 0.2$ and $R_2 = 0.8$. Here, σ refers to the single- and double-monomer moves, and τ is a microscopic time scale. If there is no move of type σ linking β with α, then $w^{(\sigma)}_{\alpha\beta} = 0$. This form of $w_{\alpha,\beta}$ guarantees that $P^{\rm eq}_\alpha \sim e^{-E_\alpha/T}$ is a steady-state

*mc@ifpan.edu.pl
†banavar@umd.edu

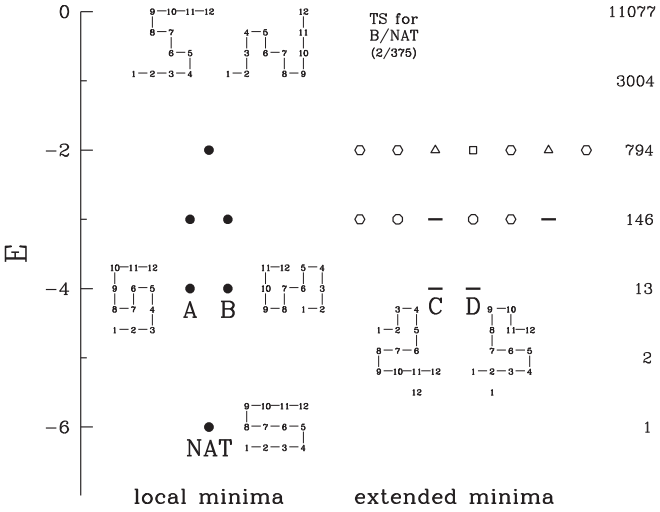

FIG. 1. The energy landscape of the system. The solid circles indicate the energies of the local energy minima. The conformations of the three lowest minima, states NAT, A, and B, are shown next to the circles. The open n-gons correspond to extended minima comprising n conformations. In particular, the horizontal bars indicate extended minima of two conformations. C and D denote two of these with the lowest energy. There are 375 transition states between NAT and state B. They have an energy of -2. Two of these, which are key transit conformations during folding, are shown at the top. The numbers of states at a given energy, $n(E)$, are indicated at the very right. They can be used to determine the free energy $F(E)$ of the system. At T_f, $F(E)$ has a local maximum at $E = -4$.

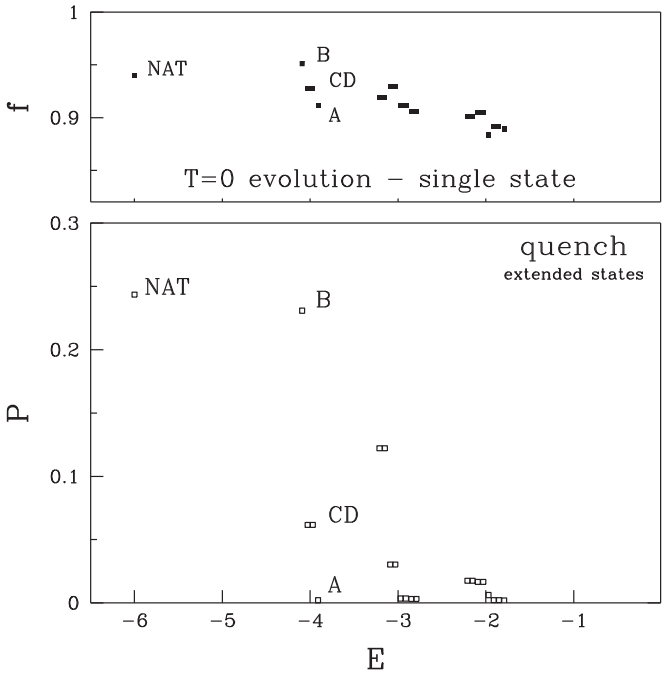

FIG. 2. The top panel shows the fraction f of all starting states that have a $T = 0$ path to each of the local or extended minima. The starting states are considered one at a time. The energies corresponding to the minima are displaced around the true value to bring out separation of various minima. The bottom panel considers a $T = 0$ steepest descent quench to the energy minima and indicates the probability P of arriving at a particular minimum. The initial starting state corresponds to all open conformations (i.e., with no native contact) being occupied evenly.

solution of the master equation corresponding to equilibrium at temperature T.

The master equation can be solved by bringing it into a matrix form by introducing $\vec{P} = (P_1, \ldots, P_{\mathcal{N}})$, where $\mathcal{N}$ denotes the total number of conformations. The time dependence of vector $\vec{P}(t)$ at time $t = n\tau$ can be obtained by applying n times the recursion $\vec{P}[(n+1)\tau] = (1 + \hat{M}\tau)\vec{P}(n\tau)$ to $\vec{P}_{in}$ describing the initial state of the system. In order to generate folding conditions we consider the native state, denoted as NAT, to be a sink for the probability distribution [19,20]. This can be accomplished by allowing for transitions that lead to state NAT and disallowing all those which lead out of state NAT, i.e., $M_{\beta,\mathrm{NAT}} = 0$ for each β. The unfolding conditions are created when all open conformations, i.e., with zero energy, act as probability sinks.

In order to characterize the connectivity of the conformational space, we first consider a $T = 0$ quench with NAT acting as a sink. We place the probability of 1 in one state at a time and ask how many of the $\mathcal{N}$ conformations have a downhill path leading to specific simple or extended minima. The top panel of Fig. 2 shows that 94% of conformations have a path to NAT. However, minimum B is slightly more accessible; 95% of all states connect to it. The corresponding numbers for states A, C, and D are 91%, 93%, and 93%, respectively. Another measure of conformational connectivity is obtained by distributing the initial probability of 1 evenly across the 11 077 open states and by monitoring which states accumulate the probability. The bottom panel of Fig. 2 indicates that NAT accumulates 24% of the probability, minimum B accumulates 23%, but minimum A accumulates only 0.2%. Thus the probability to arrive at NAT from the open conformations without a need to overcome any energy barriers is substantial. This kind of energetic landscape is consistent with the kinetic partitioning mechanism of protein folding [21–23].

We now consider kinetic processes at the folding temperature T_f, defined as one at which the probability of staying in the native conformation P_0 is equal to $1/2$ in equilibrium. T_f provides a measure of thermodynamic stability, and for our system it is equal to 0.476 in units of the contact energy. Figure 3 demonstrates that the system behaves as a two-state model: all kinetic processes are described by a simple single exponential function in time. The folding, unfolding, and relaxation times (t_f, t_u, and t_r, respectively) are shown in Fig. 3. The relaxation time is defined as the characteristic time to reach equilibrium and is determined by not introducing any probability sinks. The two-state condition that $k_r = k_f + k_u$, where the k's denote the rates or inverse times, is satisfied approximately, again in accord with the simple transition state picture.

The picture of folding that emerges is not the canonical one of a transition state separating unfolded conformations and the native state. Rather, around a quarter of the trajectories lead directly to NAT in a downhill or flat manner, whereas the remaining trajectories end up in local or extended energy minima. Getting out of these minima needs thermal activation, which is the rate determining step for folding. Figure 2 shows that minimum B constitutes the most potent trap. Indeed, when

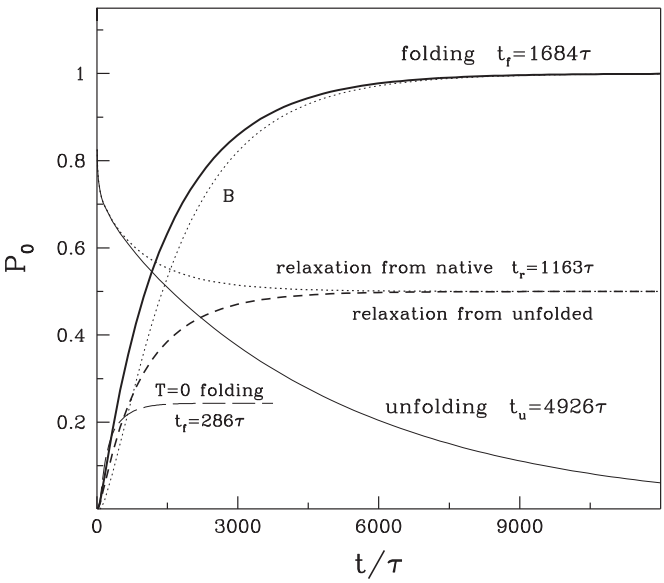

FIG. 3. Time dependence of P_0 at T_f for various processes. The thick solid line corresponds to folding when NAT acts as the probability sink. There is no significant difference in the evolution between the starting state consisting only of the open states or of all states occupied evenly. The dotted line with the symbol B next to it corresponds to an evolution which starts in conformation B. The line with the long dashes corresponds to folding at $T = 0$. In this case, only 24.36% of the full probability ends up in the native state. The thin solid line corresponds to unfolding: all open conformations act as probability sinks. The other lines correspond to relaxation to equilibrium: there are no probability sinks, but the starting state is either NAT (dotted line) or it comprises all open states (the line with the short dashes). The numbers are the characteristic times as obtained by making fits to single exponentials. Their inverses give the rates of folding, unfolding, and relaxation. The two relaxation processes have almost the same time constants.

one starts the evolution in state B, the characteristic folding time is only some 50τ longer than t_f derived by starting from the unfolded state. Note that any path from B to NAT entails first reducing the number of contacts from four and then increasing it to six, so the fraction of the native contacts present does not represent a relevant "reaction coordinate" for the process.

The transition states of interest are those that separate B from the native state. We identify 375 such states with energy of -2 as the lowest energy states from which the $T = 0$ trajectories end up in either NAT or B. Two of these 375 transition states are shown at the top of Fig. 1. These two are the most commonly traversed transition states (9.4% and 8.8%) which can be determined by starting with state B and evolving the system at T_f with all the transition states acting as sinks. Among the 375 transition states, there is a total of 61 states with contacts between pairs 3–6 and 7–10, 157 states with contacts between pairs 6–11 and 7–10, and another 157 states with contacts between pairs 2–7 and 3–6. We find that when one blocks the transition states by raising their energies to high values, the folding time from the unfolded state increases by at least an order of magnitude.

Traditionally, the properties of the transition state ensemble have been inferred experimentally through φ-value analysis [24,25]. The φ_i values for folding are given by

$$\varphi_i = \frac{\delta k_f}{k_f} \Big/ \left(\frac{\delta k_f}{k_f} - \frac{\delta k_u}{k_u} \right), \tag{4}$$

where δk_f (δk_u) denotes the change in the folding (unfolding) rate when an amino acid at site i along the sequence is mutated. The common interpretation of the φ values is that numbers close to 1 signify that the site is in a native-like environment in the transition state(s), whereas numbers close to 0 suggest a non-native local environment in the transition state.

We carried out a variant of the φ-value analysis in our model by enhancing contact energies associated with bead i by a small percentage and by determining the impact of this perturbation on k_f and k_u [20]. The values of φ depend on whether the folding starts from the open conformations or from state B, especially for $i = 1$, for which φ_i is equal either to 1 or $\frac{1}{2}$, respectively. The values are large for sites 1, 3, and 6 and $\varphi = 0$ for $i = 2$. Sites 4 and 5 have no native contacts, and the values for the other sites are determined from symmetry. Enhancing the strength of contact 1–8 is observed to make the folding time longer. This is consistent with the fact that both state B and the native state contain this contact, whereas the 375 transition states do not. Thus getting out of state B requires first breaking contact 1–8 and then reestablishing it in NAT. Unfolding from NAT is also affected, and the net result is a nonzero value of φ_1, as observed. A small enhancement of the 2–7 contact is observed not to affect the folding time, i.e., $\varphi_2 = 0$. This can be understood by noting that the enhancement affects neither state B nor the 218 transition states, including the 61 which are visited the most. Enhancing the 3–6 contact makes folding faster and unfolding slower. This is because 218 transition states incorporate it and so do NAT and B; native-like arrangements are favored. Thus φ_3 and φ_6 should

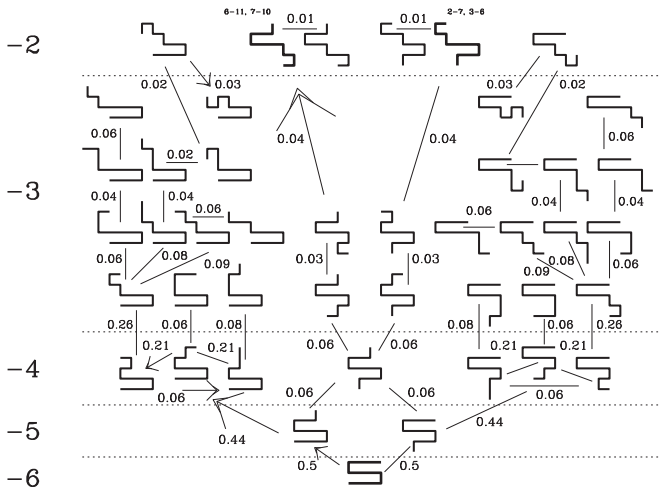

FIG. 4. A map of the larger net probability fluxes during unfolding from the native state. The figure is symmetric. The dotted horizontal lines delineate conformations corresponding to a given value of energy, as indicated on the left. The directions of the fluxes are generally upward or sideways in energy and are not indicated by arrows. Note, however, an anomaly near the top left of an arrow pointing downward in energy; this arises from the local connections between conformations and the probability of traversing them at the folding transition temperature.

be nonzero, as observed. The values of φ_i are qualitatively consistent with their expected meaning as a representation of the native-like local environment in the transition state ensemble. The important exception is site 1: among the 375 transition states there are no conformations involving contact 1–8. This casts doubt on the correctness of analyses akin to that presented in Ref. [26] in which the φ values were "interpreted in terms of native-like inter-residue contacts" and other measures of similarity to NAT.

We now consider the unfolding process from NAT to the open conformations as described by the total probability fluxes between conformations when integrated over the whole duration of the evolution (Fig. 4). The folding process in the vicinity of NAT is essentially the same but with the directions of the transitions reversed. The diagram is consistent with the idea of kinetic partitioning. The central part depicts routes, which rapidly arrive at states of energy of -2. The side parts of the diagram correspond to paths, which are less direct. The picture that emerges shows numerous nearly equivalent pathways.

We have presented a detailed analysis of the energy landscape and the dynamics of perhaps the simplest and most widely used model for understanding protein folding. Contrary to common belief, the energy landscape is not akin to a folding funnel, the model exhibits kinetic partitioning, the transition states are between a significant trap and the native state, and the φ values do not simply reflect the commonality of contacts between the transition state and the native state, but rather are a sensitive function of the trap state(s), the native state, and the ensemble of transition states. We note that while our analysis is exact, the model is necessarily exceedingly simple. It is short, two-dimensional, and on a lattice. One would expect that for more realistic models, the landscape and the dynamics would, if anything, be much more complex and harder to interpret. It is possible that going to three-dimensional models can expand the role of the kinetic partitioning linked to the native state. Using the master equation for continuum systems would require some scheme of discretization such as the one used in Ref. [27].

The computer resources were financed by the European Regional Development Fund under the Operational Programme Innovative Economy NanoFun POIG.02.02.00-00-025/09. This research has been supported by Polish National Science Centre Grant No. 2011/01/B/ST3/02190.

[1] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).
[2] N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[3] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[4] K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[5] A. R. Fersht, Curr. Opin. Struct. Biol. **5**, 79 (1995).
[6] S. Takada, Proc. Natl. Acad. Sci. USA **96**, 11698 (1999).
[7] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[8] S. B. Ozkan, K. A. Dill, and I. Bahar, Biopolymers **68**, 35 (2003).
[9] A. N. Naganathan and M. Orozco, Phys. Chem. Chem. Phys. **13**, 15166 (2011).
[10] P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992).
[11] H. S. Chan and K. A. Dill, J. Chem. Phys. **99**, 2116 (1993).
[12] C. J. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. USA **90**, 6369 (1993).
[13] A. Sali, E. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).
[14] H. Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996).
[15] D. K. Klimov and D. Thirumalai, Proteins **43**, 465 (2001).
[16] M. Cieplak, M. Henkel, J. Karbowski, and J. R. Banavar, Phys. Rev. Lett. **80**, 3654 (1998).
[17] O. Collet, Phys. Rev. E **67**, 061912 (2003).
[18] T. R. Weikl, M. Palassini, and K. A. Dill, Protein Sci. **13**, 822 (2004).
[19] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981); J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).
[20] I. Chang, M. Cieplak, J. R. Banavar, and A. Maritan, Protein Sci. **13**, 2446 (2004).
[21] J. F. Sinclair, M. M. Ziegler, and T. O. Baldwin, Struct. Biol. **1**, 320 (1994).
[22] M. J. Todd, G. H. Lorimer, and D. Thirumalai, Proc. Natl. Acad. Sci. USA **93**, 4030 (1996).
[23] D. Thirumalai, D. K. Klimov, and S. A. Woodson, Theor. Chem. Acc. **96**, 14 (1997).
[24] A. Matouschek, J. T. Kelis, Jr., L. Serrano, and A. R. Fersht, Nature (London) **340**, 122 (1989).
[25] S. E. Jackson and A. R. Fersht, Biochemistry **30**, 10428 (1991).
[26] M. Vendruscolo, E. Paci, C. M. Dobson, and M. Karplus, Nature (London) **409**, 641 (2001).
[27] D. De Sancho, J. Mittal, and R. B. Best, J. Chem. Theor. Comput. **9**, 1743 (2013).