Rare-event trajectory ensemble analysis reveals metastable dynamical phases in lattice proteins

Antonia S. J. S. Mey,[1,*] Phillip L. Geissler,[2,3,†] and Juan P. Garrahan[1,‡]

[1]*School of Physics & Astronomy, University of Nottingham, Nottingham NG7 2RD, United Kingdom*
[2]*Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA*
[3]*Chemical Sciences and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

We explore the dynamical large deviations of a lattice heteropolymer model of a protein by means of path sampling of trajectories. We uncover the existence of nonequilibrium dynamical phase transitions in ensembles of trajectories between active and inactive dynamical phases, whose nature depends on the properties of the interaction potential. We consider three potentials: two heterogeneous interaction potentials and a homogeneous Gō potential. When preserving the full heterogeneity of interactions due to a given amino acid sequence, either in a fully interacting model or in a native contacts interacting model (heterogeneous Gō model), the observed dynamic transitions occur between equilibrium highly native states and highly native but kinetically trapped states. A native activity is defined that allows us to distinguish these dynamic phases. In contrast, for the homogeneous Gō model, where all native interaction energies are uniform and the amino acid sequence plays no role, the dynamical transition is a direct consequence of the static bistability between the unfolded and the native state. In the two heterogeneous interaction models the native-active and native-inactive states, despite their thermodynamic similarity, have widely varying dynamical properties, and the transition between them occurs even in lattice proteins whose sequences are designed to make them optimal folders.

PACS number(s): 05.70.Fh, 64.60.De

I. INTRODUCTION

In statistical mechanics, when trying to uncover the physical mechanisms behind complex emergent behavior of materials or natural systems, we very often consider highly simplified model versions of such systems in the hope that these models are simple enough to allow thorough investigation while, at the same time, retaining the basic physical ingredients of the problem of interest [1–4]. In the case of protein folding, this approach has helped shape our current understanding through theoretical and computational studies of models that discard fine details of molecular structure and/or make simplifying assumptions about the interaction energies of amino acid residues [5–7]. One such idealized model is the representation of a heteropolymer by a self-avoiding walk on a cubic lattice, as originally proposed by Gō, where each occupied lattice site represents an amino acid and each edge represents an unbreakable backbone bond [8,9]. This model has been widely studied [9–11], has been shown to mimic the elementary aspects of protein folding, and is the system we consider here.

The question we address is that of the existence of highly metastable, or "glassy," states [12] in lattice protein models. A well-"designed" sequence (one that makes the protein an efficient folder) should minimize kinetic bottlenecks en route to the native state—sometimes referred to as the "principle of minimal frustration" in the context of natural proteins [13]. In contrast, a heteropolymer with a random or a poorly designed sequence is plagued by kinetic traps and, at low temperatures, becomes arrested in an amorphous compact state, analogous to a glass in this context [14–16]. On the

surface at least, this would appear to indicate that in well-designed protein sequences glassy states are absent. Here we show, however, that even well-designed lattice proteins, i.e., those with sequences that allow them to reach the desired native state efficiently, possess glass-like arrested states. These states are thermodynamically unlikely (thus folding events are successful on average and occur fast) yet highly kinetically metastable. We show that, in fact, dynamics takes place close to the first-order coexistence between an equilibrium and "active" dynamical phase and a nonequilibrium and "inactive" (or glassy) phase. We also show that glass-like states can be highly native: just as in the glass problem, while active and inactive states differ markedly in their dynamics, they cannot be distinguished by simple structural measures (here, degree of nativeness). We obtain these results by studying the large-deviation (LD) properties of ensembles of dynamical trajectories via the so-called "*s*-ensemble" method [17–21], recently used to uncover dynamical phase behavior and transitions in glasses and other systems with complex dynamics. While our results are for a highly idealized system, it is not far-fetched to speculate that more detailed models of proteins will have dynamical phase behavior as rich as (if not richer than) that we report here.

The paper is organized as follows. In Sec. II we introduce the lattice models we study. In Sec. III we give a brief overview of the *s*-ensemble approach for uncovering dynamical phase behavior, both in a general context and for the more specific case of the lattice protein model we study here, including how to access the *s* ensemble with biased simulations. Section IV presents the results for the heterogeneous and full interaction models. In Sec. V the results are collected into a schematic phase diagram, and in Sec. VI we provide a discussion of our observations in a more biological context and also compare them to other recent results in the literature. The Appendixes present further model and simulation details.

---

*antonia.mey@fu-berlin.de
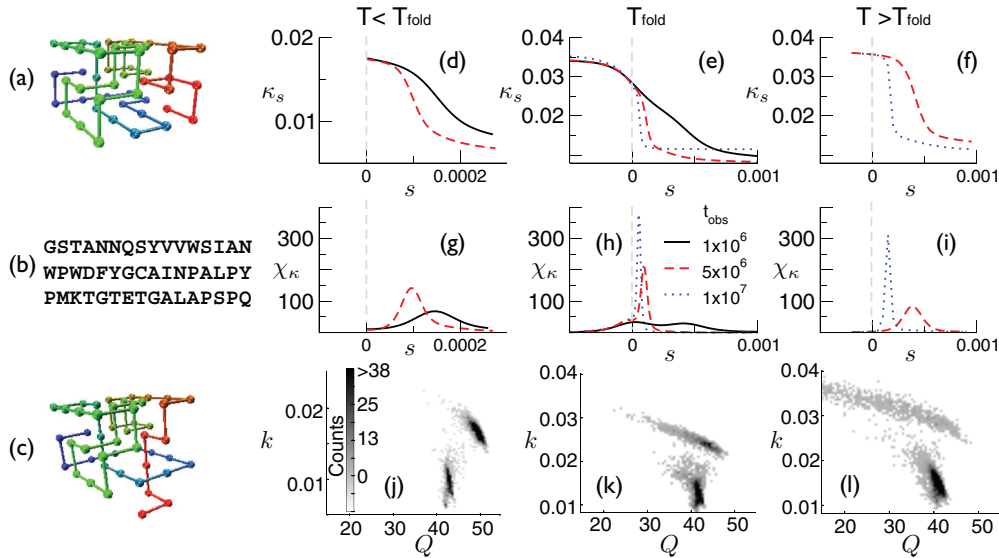†geissler@berkeley.edu
‡juan.garrahan@nottingham.ac.uk

FIG. 1. (Color online) Results of HeGō s-ensemble simulations of 30 000 trajectories for different T and t_{obs}: (a) native state; (b) sequence; (c) representative trapping state. κ_s as a function of s for temperatures of (d) $T_{fold} > T = 0.175$ (e) $T_{fold} = T = 0.19$ (f) $T_{fold} < T = 0.205$. Different curves correspond to different trajectory lengths ($t = t_{obs}$) in units of Monte Carlo steps: $t = 1 \times 10^6$ [solid (black) line], $t = 5 \times 10^5$ [dashed (red) line] and $t = 1 \times 10^7$ [dotted (blue) line]. The equilibrium mean first folding time (MFFT) at $T = T_{fold}$ is $t = 1 \times 10^6$. (g–i) corresponding dynamical susceptibilities $\chi_K(s)$. (j–l) Joint distribution of activity (per unit time and residue) k and nativeness Q from subsamples of 5000 trajectories.

II. MODELS

We study the standard lattice protein model [14–16] of a self-avoiding walk on a three-dimensional cubic lattice, where each site on the walk represents an amino acid or residue of the protein, and each bond that connects these sites a backbone bond. The chain is unbreakable, self-avoiding, and ergodic, and allowed Monte Carlo moves maintain these properties. More details on the model and simulations are provided in the Appendixes. With appropriately chosen interactions between residues a lattice heteropolymer displays the characteristic two-state kinetics and thermodynamics of simple proteins [7]: at high temperatures the stable thermodynamic state is that of an extended and mobile chain—the "unfolded" state—while at low temperatures the stable state is compact and less mobile—the "native" state—the change of state being "first-order-like," i.e., a first-order crossover (due to the finite extent of the system), whose formation is often initiated with the nucleation of a set of key contacts between residues [22–24]. As in real proteins, the folding into a specific native state is encoded in the amino acid sequence, which in turn determines the interactions.

Figure 1(a) shows one of the native structures we consider: a chain of length $L = 48$, with $N = 57$ native contacts in its fully folded state. We analyze three possible energy functions. In all cases interactions between pairs of noncontiguous residues (i.e., two residues which are not direct neighbors along the backbone) are considered only when they lie on two lattice sites which are one lattice spacing a away from each other. The first energy function we consider includes all interactions between such pairs of residues. The interaction energy between them depends on the two residues involved and is parameterized by the Miyazawa-Jernigan interaction matrix [25]: we call this the Full model. The second energy function corresponds to

considering the same interactions as in the Full model only between residues which form a *native contact*, i.e., between residues which should be a lattice spacing away in the native state: we call this the heterogeneous Gō (HeGō) model. The third energy function considers only native interactions with a uniform interaction energy between native contacts: this is the homogeneous Gō (HoGō) model. Only for the Full and HeGō model is the sequence relevant. The one shown in Fig. 1(b) was specifically designed to be a fast folder to the native structure in Fig. 1(a) [22]. Results of the HoGō model are mostly left for the Appendixes.

The system's dynamics is that of the standard Metropolis Monte Carlo, using a previously proposed move set consisting of single- and two-monomer moves [23]. The three models we consider—Full, HeGō, and HoGō—display two-state thermodynamics and folding kinetics [22,26,27], which are reviewed in the Appendixes for completeness.

III. DYNAMICAL PHASE BEHAVIOR AND s ENSEMBLE OF TRAJECTORIES

We are primarily interested in understanding the space-time dynamics, which is achieved by looking at rare events in the equilibrium path ensembles [17,18,20]. We denote by X_t a trajectory, of time extension $t = t_{obs}$, from such an ensemble, and denote its corresponding path probability $P[X_t]$. In order to investigate the dynamical phase structure, we define a dynamical order parameter termed the activity $K[X_t]$, extensive in both system size and observation time, which we use to classify trajectories. A convenient choice is given by the "native activity," that is, the total number of events in which a native contact is made or broken in a trajectory. As in the case of glasses, activity is the natural order parameter to

explore metastability in systems displaying complex collective dynamics [18,28].

For an equilibrated system at a given temperature, the path ensemble $P[X_t]$ of trajectories X_t can be sampled straightforwardly by generating dynamical trajectories starting from an equilibrated initial state. This is not very efficient for exploring rare events. To explore the tails of $P[X_t]$ we can formally define a modified ensemble of trajectories biased by activity

$$P_s[X_t] \equiv \frac{P[X_t]e^{-s\,K[X_t]}}{Z_t(s)}. \tag{1}$$

The parameter s is a biasing "counting" field conjugate to the activity $K[X_t]$ [18]. The exponential factor in Eq. (1) biases the probability of trajectories towards those which are less (more) active when $s > 0$ ($s < 0$) compared to the unbiased ensemble. The normalization factor,

$$Z_t(s) \equiv \sum_{X_t} P[X_t]e^{-s\,K[X_t]}, \tag{2}$$

is the moment-generating function for K [that is, $\langle K^n \rangle = (-)^n \partial_s^n Z_t(s)|_{s=0}$] and can be thought of as a dynamical partition function associated with the ensemble of trajectories biased by s.

In analogy with an equilibrium statistical mechanics problem, the dynamical partition sum $Z_t(s)$ is the object of interest. At long times it acquires an LD form [18,29]:

$$Z_t(s) \sim e^{t\psi(s)}. \tag{3}$$

The LD function $\psi(s)$ can be thought of as a dynamical free energy. It is directly related to the order parameter distribution at long times. That is, the probability $P_t(K)$ of observing an activity K over time t also has an LD form for long times, $P_t(K) \sim e^{-t\phi(K/t)}$, and the functions ϕ and ψ are connected by a Legendre transform [29].

Just like the free energy in an equilibrium problem, the analytic structure of $\psi(s)$ as a function of s tells us about *dynamical* phases and possible phase transitions (or crossovers, in the case of systems of finite extent) between them. The average native activity, as a function of s, is given by

$$\kappa_s \equiv \frac{\langle K \rangle_s}{Nt} = \frac{\sum_{X_t} P_s[X_t]K[X_t]}{Nt} \sim -\frac{\partial \psi(s)}{\partial s}, \tag{4}$$

where we have defined κ_s as activity per unit time and unit native contact (where N is the number of native contacts in the fully folded state). When $s = 0$, Eq. (4) gives the average rate at which native contacts are formed and broken in the normal dynamics of the system. When $s \neq 0$, Eq. (4) gives the average rate associated with atypical trajectories biased to be more or less active by the field s. A singular change in $\psi(s)$ with s translates into a singular change in κ_s, which will therefore serve as the order parameter, which allows us to distinguish between dynamical phases in the lattice heteropolymer. Alternative choices for the activity, such as the total number of formed/broken contacts, irrespective of whether or not they are native, give equivalent results, as discussed in the Appendixes.

The average $\langle \cdot \rangle_s$ in Eq. (4) is over the ensemble of trajectories biased by s as in Eq. (1), which we call the s ensemble. The s ensemble can be probed numerically by a variation of transition path sampling (TPS) as used in [20] and

[30]: in effect, a Monte Carlo scheme in trajectory space that samples the distribution $P_s[X_t]$. A trajectory X_t can be sliced into n segments of time extent $\tau = t/n$. Each segment of the trajectory can serve as a shooting point from which part of the original trajectory is regenerated. For the reversible trajectories we are sampling here, the shooting direction can be forwards or backwards. The probability of accepting a new trajectory $X_t^{(\text{new})}$ thus generated is given by a Metropolis criterion dependent on the change in activity, $\min[1, e^{-s(K_{\text{new}} - K_{\text{old}})}]$. This procedure guarantees eventual convergence to the s ensemble $P_s[X_t]$ at a given temperature of the system.

IV. RESULTS

In Fig. 1 we show results for the s ensemble for the HeGō model with the native structure and sequence in Figs. 1(a) and 1(b); Figs. 1(d)–1(f) show the average native activity κ_s, Eq. (4), as a function of s for temperatures, below, at, and above the folding temperature $T_{\text{fold}} = 0.19$ for this system (T_{fold} is defined as the temperature at which 50% of the native contacts are formed, on average, in equilibrium). For $s < 0$ the native activity is larger than the typical one, and for $s > 0$ it is smaller, as expected from Eq. (1). But what is notable is that the change from more active to less active as a function of s becomes sharp with increasing the observation time of the trajectory. This is also seen in the increase in the peak of the dynamical susceptibility,

$$\chi_K(s) \equiv \frac{\langle K^2 \rangle_s - \langle K \rangle_s^2}{Nt}, \tag{5}$$

shown in Figs. 1(g)–1(i). Such behavior is indicative of a first-order transition between an active/equilibrium dynamical phase and an inactive/metastable dynamical phase. Even if time can become infinitely large, the transition is rounded since the protein is a system of finite extent in space. The first-order dynamical transition in trajectories shown in Fig. 1 is highly reminiscent of what is observed in models of glasses, where the inactive phase is associated with dynamical metastability [20,31].

The natural question to ask is whether there is a structural signature of the active-inactive transition we observe in the HeGō model. An obvious structural order parameter is the "nativeness," i.e., the overall time average of formed native contacts $Q \equiv t^{-1} \int_0^t \sum_j n_j(t')dt'$, where j runs over all native contacts ($j = 1, \ldots, 57$ for the specific case in Fig. 1), and $n_j(t) = 1, 0$ indicates whether native contact j is made/broken at time t. Figures 1(j)–1(l) show the joint probability of activity (per unit time and residue) $k \equiv (Nt)^{-1}K$ and nativeness Q of all trajectories in an s ensemble near the critical value $s_c(T)$. The bimodality of the distribution is evident, as expected from the first-order nature of the transition, but what is notable is that the active/inactive basins, which differ greatly in dynamical activity, are difficult to distinguish in terms of nativeness. Inactive trajectories are a consequence of a highly metastable state which is also as *native* as equilibrium states; a characteristic conformation associated with this inactive near-native state is shown in Fig. 1(c). This, again, is reminiscent of glasses, where the inactive glassy state is difficult to distinguish structurally from the relaxing liquid [21].
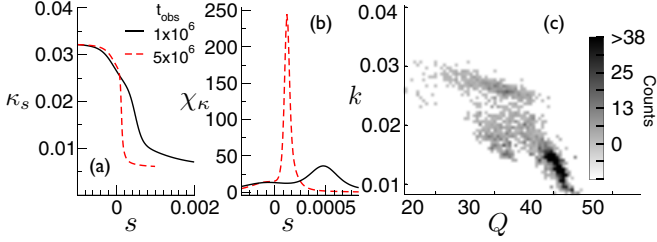
FIG. 2. (Color online) Results of the Full model where the equilibrium ($s = 0$) MFFT to the folded state is $t = 2 \times 10^6$ at T_{fold}. (a) κ_s with respect to s at $T_{\text{fold}} = 0.19$ averaged over 20 000 trajectories. (b) Fluctuations χ_K in terms of s. (c) Joint probability density of k and Q from a subsample of 5000 trajectories.

The existence of inactive yet highly native states seems to require the heterogeneity of interactions associated with the amino acid sequence. In the Appendixes we present a similar s-ensemble analysis for a HoGō model for the same structure as in Fig. 1(a): for this model, where the sequence plays no role in the interactions, there is a straightforward linear relation between dynamical activity k and structural nativeness Q, indicating that kinetic trapping is determined by the same structural states—unfolded and native—that determine the thermodynamics. In contrast, the Full model for the sequence in Fig. 1(b) shows active-inactive transitions of similar richness as the HeGō model (see Fig. 2). This is indicative of the fact that for qualitative folding models sequence heterogeneity is essential, as for the case of the Full and HeGō models [22].

In Fig. 3 we further explore the nature of the active and inactive dynamical phases. Figure 3(a) shows the joint density profile of the activity and nativeness k and Q from a subsample of 5000 trajectories at $T = T_{\text{fold}}$, for two values of s: $s = 0$ (typical dynamics) and $s = s_c \approx 10^{-4}$ [an enlarged version of Fig. 1(k)]. In the former case there is a clear
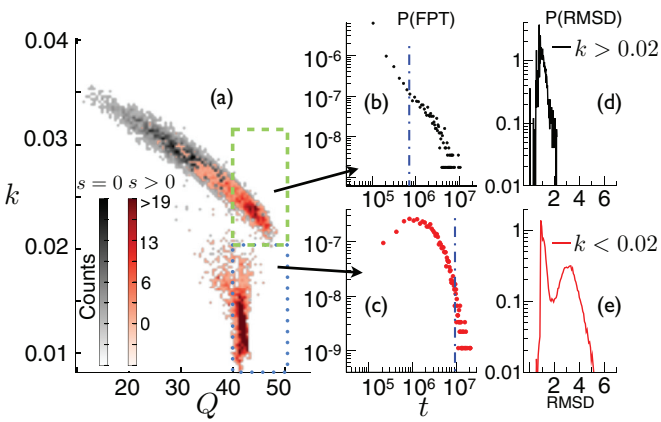
correlation between k and Q, while at "space-time" coexistence between the two dynamical phases two clear basins of very distinct activity but similar structural nativeness are found, as previously discussed. Representative configurations taken from trajectories fulfilling the requirements of $Q > 40$ and $k < 0.02$ or $k > 0.02$, respectively, are used in order to investigate the dynamic nature of the highly native active or inactive trajectories. This is indicated in Fig. 3(a) by the respective boxes [dashed (green) box, $k > 0.02$; dotted (blue) box, $k < 0.02$]. Figures 3(b) and 3(c) look at the distributions of first passage times (FPTs) to the fully native state starting from a typical conformation of the active or inactive phases, respectively, rapidly quenched to equilibrium ($s = 0$). Not only do the inactive states have a much larger mean first folding time [MFFT; indicated by the dash-dotted (blue) lines], but also the distribution is exponential, rather than stretched (as a power law, as is characteristic of near-native equilibrium states in this model). Figures 3(d) and 3(e) show the distribution of the root mean square distance (RMSD) in the FPT trajectories. Active trajectories are sharply peaked around RMSD values of 1 lattice unit, corresponding to the RMSD value of the initial structures, which is already "en route" to the native state, as shown in Fig. 3(d). Low-activity initial structures escape trapped states by an unfolding event (RMSD peak of 3 lattice units) followed by the folding event [see Fig. 3(e)]. This supports the idea of a long-lived and metastable trapping state, prolonging folding times significantly.

V. DYNAMICAL PHASE DIAGRAM

From our simulation results, a schematic phase diagram in the parameter space $\{T, s\}$ can be constructed as shown in Fig. 4. For all values of T there is a dynamical phase corresponding to equilibrium trajectories that is smoothly connected to $s = 0$. The phase boundary of this equilibrium dynamical phase is at $s_c(T)$, indicated by the solid (black) curve in Fig. 4. At temperatures $T > T_{\text{fold}}$, equilibrium trajectories are very active as the typical states are unfolded
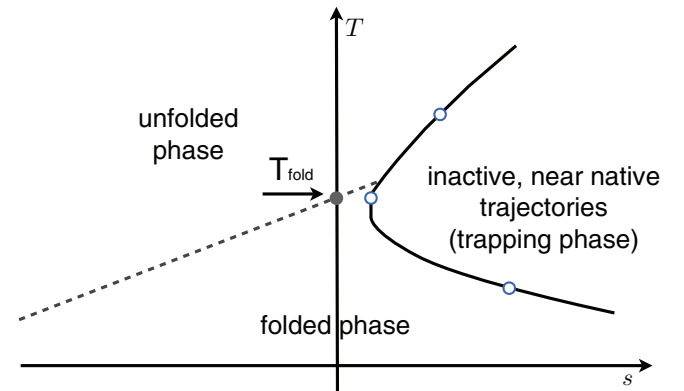

FIG. 3. (Color online) Dynamic phase classification. (a) Densities of the joint distribution of k and Q for $s = 0$ (black; lower distribution) and $s = 1 \times 10^{-4}$ (red; upper distribution) for $t_{\text{obs}} = 5 \times 10^6$ and $T = T_{\text{fold}}$ of 5000 trajectories. (b, c) FPT distribution of 5000 FPT trajectories for initial configurations selected from critical trajectories with $Q > 40$ and $k > 0.02$ (b) and $Q > 40$ and $k < 0.02$ (c). Dashed-dotted lines indicate MFFT of the distributions. (d, e) RMSD distribution for initial states with $Q > 40$ and $k > 0.02$ (d) and $Q > 40$ and $k < 0.02$ (e).

FIG. 4. (Color online) Schematic dynamic phase diagram of the HeGō and Full models in the parameter space $\{s, T\}$. First-order dynamic transitions, giving rise to inactive yet highly native trajectories, are indicated by the solid black line. Open circles along the line of the first-order transitions are taken from the results in Fig. 1. The dashed (gray) line is an extension of the thermodynamic native-to-non-native transition into the s space.

(unfolded phase), while at $T < T_{\text{fold}}$ equilibrium trajectories display a lower activity since typical states are native (folded phase). The transition between these two regimes, indicated by the dashed (gray) line in Fig. 4, is a direct extension of the thermodynamic crossover at T_{fold} [filled (gray) circle]. As shown in the Appendixes, this straightforward connection between static and dynamic phases is all that is present in the HoGō model.

In the HeGō and Full models, that is, the models with the interaction heterogeneity of the amino acid sequence, a third dynamical phase exists. At $s_c(T)$ there is a transition to the nonequilibrium phase of metastable, or trapped, trajectories of very low activity (trapping phase; three open circles, corresponding to the three temperatures, in Fig. 1). From the structural point of view, the trapped phase appears to be as native as the equilibrium phase despite having a very low activity, showing that in the more complex HeGō and Full models dynamic behavior is not a simple consequence of static behavior.

We surmise the phase diagram in Fig. 4 from the analysis of the results presented in Fig. 1. As we show in the Appendixes, such behavior is also valid for different sequences optimized for the same native state in Fig. 1(a). Depending on the sequence and native state, the phase of inactive trapping trajectories may contain an even richer phase structure, but we expect it to be fundamentally similar to that in Fig. 4.

VI. DISCUSSION

We have shown that very simplified models of proteins, as long as interaction variability is maintained, display a complex dynamical phase structure analogous to that of glassy systems. Just as in ordinary many-body systems, the mere existence of a phase transition to a highly inactive phase implies that rare fluctuations will play a role in the observed dynamics. It is not far-fetched to assume that such behavior can also be observed in nature, especially as there are certain biological systems that seem to be good candidates for such dynamics. As an example one might think of fibril formation, where certain kinetically rare trapping events predispose protein molecules to aggregation [32]. Once such an aggregate is formed it is very stable yet not in its native configuration. A first approach to study arrested protein dynamics in all atom models was attempted in very recent work [12], where the s ensemble was applied to Markov state models of a selection of all-atom protein models, obtained by means of molecular dynamics simulations. While a crossover between active and inactive phases was observed in all cases studied, no sharp transition to a glass-like state, like the one we presented here in lattice proteins, was found. This may be due to the fact that the biased distribution itself was not sampled, as only the equilibrium Markov state model was biased according to the s ensemble with observation times no larger than the folding time scale. In contrast, our observation times were up to 10 times the MFFT of the system and explicitly sample from the biased distribution, thus showing much sharper transitions to non-native, metastable states. Observing such dynamic transitions in all-atom protein models by direct sampling from the biased distribution may still be problematic due to the sampling problem that plagues the field of all-atom protein molecular dynamics simulations.

APPENDIX A: MODEL DETAILS

As discussed above, a heteropolymer of length L and with N native contacts (mostly $L = 48$ and $N = 57$), represented by a self-avoiding walk on a lattice, is studied. The length L represents the number of beads that sit on the vertices of the cubic lattice. All edges have a uniform length a and represent the lattice spacing. The heteropolymer's native state is defined by a maximally compact structure as depicted in the schematic in Fig. 1(a). The choice of the native state was motivated by a series of previous studies [22,26,27]. We look at three variants for the interaction potential between monomers: the Gō interaction potential (HoGō), which considers only contacts (i.e., monomers which are not direct neighbors along the backbone and which are a single lattice spacing a away from each other) that occur in the native state to which it assigns a uniform interaction energy. The heterogeneous Gō potential (HeGō) has nonuniform interaction energies between these native contacts; and in the full interaction potential (Full), the energy of all contacts (native and non-native) is considered.

The energy function for the three models can be written generically [6],

$$E = \sum_{i=1}^{L-1} \sum_{j>i}^{L} U(r_{ij}) + \sum_{i=1}^{L-3} \sum_{j=i+3}^{L} N_{ij} B_{ij} \Delta(r_{ij} - a), \quad \text{(A1)}$$

where $\vec{r}_i$ indicates the position of the ith monomer and $r_{ij} = |\vec{r}_i - \vec{r}_j|$. The potential $U(r_{ij})$ restricts the walker to be self-avoiding, as it takes a value of ∞ for $r = 0$ and of 0 for any value of $r > 0$. The term B_{ij} is an energy interaction matrix, which is determined by the sequence of the amino acids chosen [for example, that in Fig. 1(b)]. Interaction values for different amino acids are drawn from the model of Miyazawa and Jernigan [25] and set the base value ϵ_0 for the energy scale. In this model, only residues in contact, i.e., one lattice spacing away from each other, are of interest, and the matrix Δ holds the information of the list of contacts in a given configuration of the heteropolymer. In the HoGō and HeGō models N_{ij} holds the information about the set of native contacts, so that $N_{ij} = 1$ if the contact is present in the native state and 0 otherwise. This restriction is lifted for the Full interaction model, with all entries given by 1. The energy function now allows us to compute the instantaneous energy of the system at each simulation step. The simulation is a Monte Carlo simulation using the standard Metropolis acceptance criterion, $P_{\text{accept}} = \min[1, \exp(-\beta \Delta E)]$. The set of moves consists of single-monomer and double-monomer moves defined in [23]. The single moves are flips of the terminal bead or corner bead, and the double moves are "crank-shaft" moves [23]. We found that attempting a ratio of 80% monomer to 20% crank-shaft moves gave a good sampling. The time observable t given earlier corresponds to the incremental count of each attempt to move one of the beads, thus generating trajectories of any given
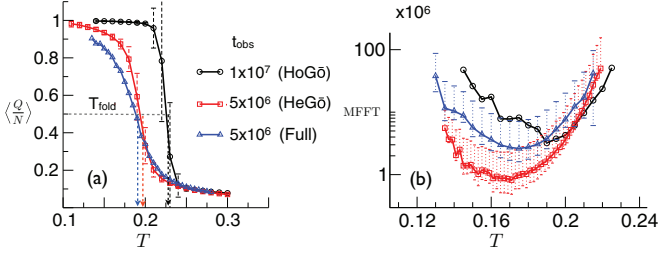
FIG. 5. (Color online) (a) Two-state thermodynamics: average number of native contacts as a function of temperature for the three models: HoGō [(black) circles], HeGō [(red) squares], and Full [(blue) triangles]. (b) MFFT from random high-temperature states as a function of temperature. Averages were taken from 1000 independent simulation runs.

observational time $t = t_{obs}$. Initial configurations were drawn from short equilibrium trajectories at a high temperature, $T = 100\epsilon_0/k_B$. The Boltzmann constant was set to $k_B = 1$ in all simulations. All other temperatures are unitless, defined by the Miyazawa and Jernigan interaction scale.

APPENDIX B: EQUILIBRIUM AND FOLDING BEHAVIOR

We recomputed relevant known [16,22,23,26,27] equilibrium results of the lattice polymer models considered above. The two-state thermodynamics of the three models considered can be observed in the behavior of the average number of native contacts in equilibrium [Fig. 5(a)], which crosses over from being low in the unfolded state at high T to high in the native state at low T [where the choice for the native state was that of the structure shown in Fig. 1(a)]. The sigmoidal behavior mimics that of real protein denaturation. The T_{fold} temperature, indicated by dashed arrows, is defined as the temperature at which 50% of all native contacts are formed. The second observable is the mean first passage time to the folded state, or the mean first folding time (MFFT), shown in Fig. 5(b).

We have defined the native activity K as the total number of formed and broken native bonds in a trajectory of time extent t. The average native activity per unit time and native bond is $\kappa_{s=0}$, Eq. (4). It is shown as a function of T for the HoGō model in Fig. 6(a) and for the HeGō and Full models in Fig. 6(c). In all cases it shows a two-step behavior similar to that of other thermodynamic quantities. In fact, in equilibrium trajectories, there is a close correlation between (intensive) native activity, $k = K/Nt$, and nativeness Q, as shown in Figs. 6(b) and 6(d) for the case of $T = T_{fold}$. The correlation is particularly sharp for the HoGō case.

APPENDIX C: s-ENSEMBLE SIMULATION DETAILS

Dynamical LDs and the s-ensemble approach for studying statistical properties of trajectories is discussed at length in Refs. [18–20,33]. The key aim is to obtain the partition sum, Eq. (2), in its LD form, Eq. (3). How the LD function $\psi(s)$ relates to the (physically observable) probability distribution of the activity $P_t(K)$ is sketched in Fig. 7. The panels on the left in Fig. 7 describe the case of a single dynamical phase: $\psi(s)$ is an analytic function for all s [Fig. 7(a)]; in
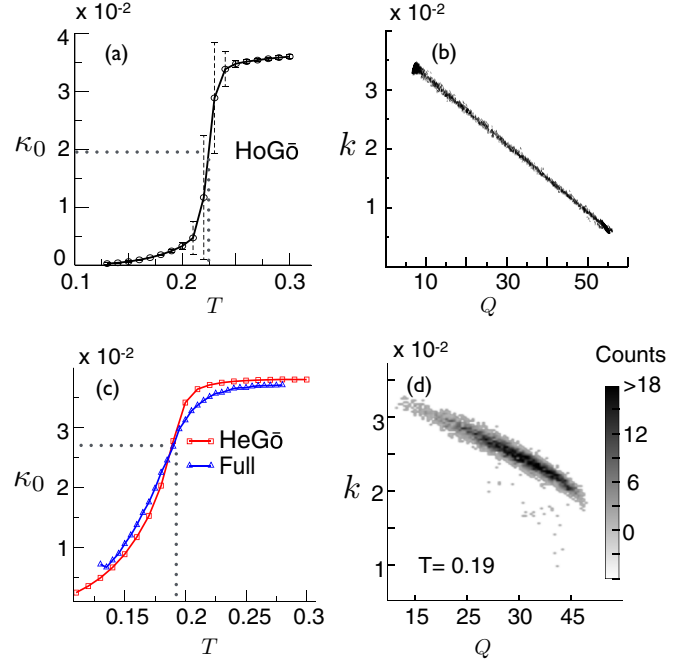

FIG. 6. (Color online) Equilibrium results (a,b) for the HoGō model and (c,d) for the HeGō and Full model. (a,c) Average native activity $\kappa_{s=0}$ in equilibrium as a function of temperature T. Folding temperature T_{fold} indicated by the dashed line. (b,d) Native activity k vs nativeness Q in equilibrium trajectories. Trajectory length was 1×10^7 Monte Carlo steps for the HoGō model [(black) circles] and 5×10^6 Monte Carlo steps for the Full [(blue) triangles] and the HeGō [(red) squares] models.

this case the distribution $P_t(K)$ is unimodal [Fig. 7(c)]; and the average activity κ_s changes smoothly with s [Fig. 7(e)]. The panels on the right of Fig. 7, in contrast, describe the case with two dynamical phases, with a first-order transition between them: $\psi(s)$ is has a singularity at s_c [Fig. 7(b)]; in this case the distribution $P_t(K)$ is bimodal [Fig. 7(d)]; and the average activity κ_s is discontinuous at s_c [Fig. 7(f)]. Note that while $P_t(K)$ is the probability distribution observed in normal dynamics, i.e., when $s = 0$, its shape is strongly affected by the presence of the singularity in the LD function $\psi(s)$ at s_c.

The s ensemble is composed of trajectories of the original dynamics but where their probabilities are biased by s according to Eq. (1). For accessing this ensemble a modified TPS strategy was employed [20,30]. An initial trajectory of a predefined length of time steps was generated. Along this trajectory with equal probability a shooting point was chosen from which either the first or the second part of the trajectory was regenerated, again with equal probability. In this way a second trajectory similar in activity K to that of the first was obtained. In order to bias according to the s ensemble the new trajectory was accepted according to the Metropolis acceptance criterion $P^s_{accept} = \min[1, e^{-s(K^{new} - K^{old})}]$. This acceptance uses the extensive native activities. As we are trying to approach the limit of $t \to \infty$, we choose computationally feasible trajectory lengths. By this we mean trajectories that are long enough, ideally $t \geqslant$ MFFT, so that even large values of the bias still allow fast equilibration into the s ensemble. For all simulation results shown here an initial
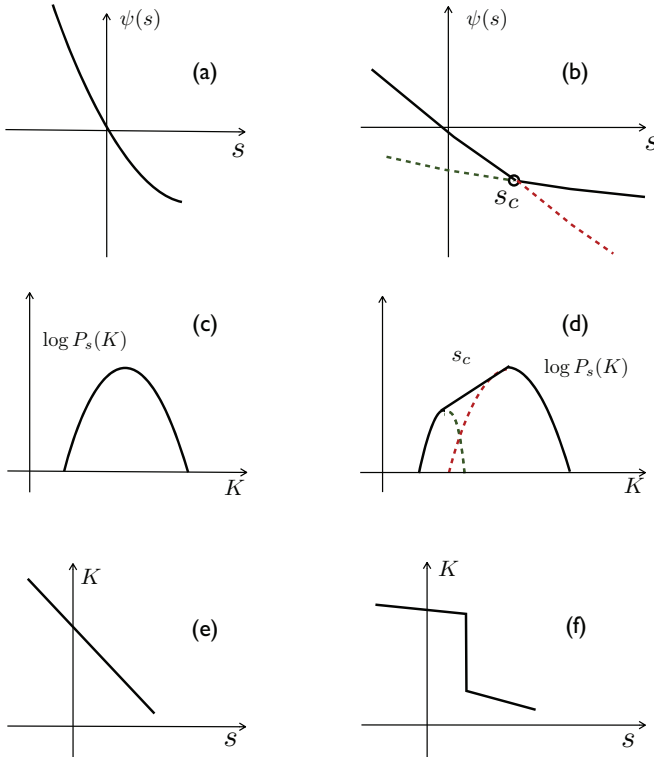
FIG. 7. (Color online) Sketch of the relation between the LD function ψ and the order parameter distribution $P(K)$. (a, c, e) The case with a single phase and a unimodal distribution. (b, d, f) The case with two phases and a bimodal distribution, resulting in a first-order transition between the two phases.

set of 2000 trajectories was discarded towards equilibration into the s ensemble. A choice of $t \geqslant$ MFFT will ensure that trapping states even longer-lived than the folded state can be uncovered and are allowed to dominate the dynamics in the biased setting. In order to reduce the required computations, after an initial scan through different values of s, the simulation time was concentrated to values close to a critical s. The set of trajectories was then reweighted with a histogram reweighting method in order to compute the dependence of s and κ_s, as well as the susceptibilities.

APPENDIX D: s ENSEMBLE OF THE HoGō MODEL

Figure 8 shows results for the s ensemble in the HoGō model obtained from TPS simulations of 10^4 accepted trajectories. Figure 8(a) is for temperatures $T < T_{\text{fold}}$; Fig. 8(b), for temperatures $T > T_{\text{fold}}$. Both panels show a clear transition from high to low activity as s is changed. For the case $T < T_{\text{fold}}$ the transition occurs at $s_c < 0$ and moves progressively to more negative values of s as T decreases from T_{fold}. Conversely, for $T > T_{\text{fold}}$, the transition occurs at $s_c > 0$ and moves to larger values as T increases. These results suggest the phase diagram in Fig. 8(c): the transition between the dynamical active and inactive phases is a direct extension of the thermodynamic transition between the unfolded and the native state.
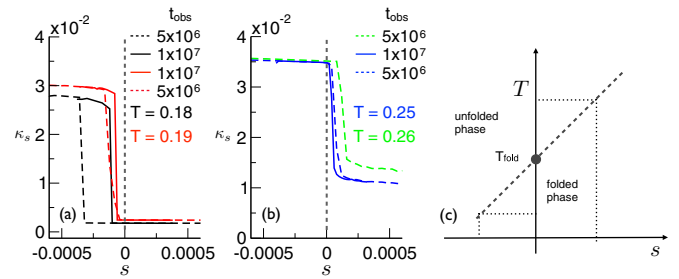

FIG. 8. (Color online) The s ensemble of the HoGō model. (a) Average native activity κ_s as a function of s for temperatures below T_{fold} ($T = 0.18$ and $T = 0.19$ as indicated). Dashed lines correspond to observation times of $t = 5 \times 10^5$; solid lines, to $t = 1 \times 10^6$. (b) Same as (a), but for $T > T_{\text{fold}}$ ($T = 0.25$ and $T = 0.26$). (c) Schematic dynamical phase diagram: the active phase is the dynamical phase associated with the unfolded thermodynamic phase, and the inactive phase is associated with the native thermodynamic phase. The dashed line indicates the first-order transition between the dynamical phases (dotted intersects are the observed data at high and low temperatures). The dynamical transition intersects the temperature axis at T_{fold}.

APPENDIX E: DIFFERENT DYNAMICAL OBSERVABLES

The emergence of complex dynamic behavior using the native activity (K) as an order parameter has been presented. An obvious question arises: Is this complex dynamic behavior also observed if the dynamic observable is changed? Another possible choice of observable is the general activity, that is, the count of all contacts broken or formed (G), irrespective of whether they are native, over the whole trajectory. The intensive general activity is then defined by $g = \frac{G}{Nt}$, and the averaged general activity as

$$g_s = \frac{\langle G \rangle_s}{Nt} = \frac{\sum_{X_t} P_s[X_t] G[X_t]}{Nt}. \quad (E1)$$

Figure 9 shows the results for a TPS simulation of 30 000 biased trajectories for different values of s at a temperature of $T_{\text{fold}} = 0.19$ in the HeGō model. The observation time t was again chosen in such a way that $t \geqslant$ MFFT at T_{fold}. The general activity is now higher than the native one (cf. Fig. 1), but again, a sharp crossover is observed at $s \approx 0$. Taking a set of critical trajectories using, again, the histogram reweighting method, the dependence of g_s with respect to s was obtained for two
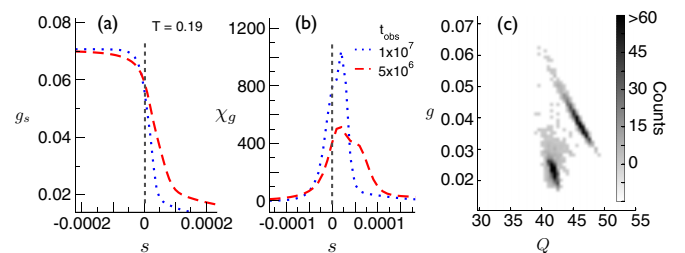

FIG. 9. (Color online) General activity in the HeGō model from averages of 30 000 trajectories at $T = 0.19$. (a) g_s with respect to s for $t = 5 \times 10^6$ [dashed (red) line] and $t = 1 \times 10^7$ [dotted (blue) line] and (b) the corresponding susceptibility χ_g. (c) The joint distributions of g and Q form of a set of critical trajectories for $t = 5 \times 10^6$.
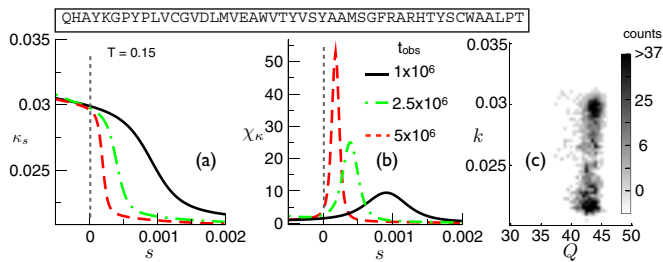
FIG. 10. (Color online) s-ensemble results for a fast folder with the sequence shown using the HeGō model. (a) κ_s with respect to s at $T = 0.15$ for trajectory lengths $t = 1 \times 10^6$ [solid (black) line], $t = 2.5 \times 10^6$ [dashed (red) line], and $t = 5 \times 10^6$ [dash-dotted (green) line]. (b) Susceptibility and (c) joint distribution of k and Q. The active-to-inactive trapped transition becomes very pronounced for longer observation times, and the two phases, while dynamically very distinct, are virtually indistinguishable in terms of nativeness Q.

observation times, $t = 5 \times 10^6$ [dashed (red) line] and $t = 1 \times 10^7$ [dotted (blue) line]. Again, the transition gets sharper for longer times. The joint distribution of g and Q for a set of critical trajectories shows a pattern similar to that in Fig. 1.

APPENDIX F: DIFFERENT SEQUENCES

In order to highlight the fact that the existence of a dynamically trapped phase is not the consequence of a specific sequence (or a specific native structure), we show here the case of a second sequence, different from that in Fig. 1(b), which also folds to the native structure in Fig. 1(a) in the HeGō model. For this sequence (Fig. 10), $T_{\text{fold}} \approx 0.155$. The folding dynamics follows a different pathway to the sequence in Fig. 1(a). The MFFT at a temperature near T_{fold} is about five times larger than that in Fig. 1(a) yet still counts as a "fast folder" [22]. Figures 10(a)–10(c) show the results of s-ensemble simulations at $T = 0.15$ for trajectory observation times $t = 1 \times 10^6$, 2.5×10^6, and 5×10^6. The longest observation time is close to the MFFT of the system at the observed temperature. Figures 10(a) and 10(b) show that the dynamical transition becomes increasingly pronounced with longer observation times, and the transition point is close to $s = 0$, indicating that the inactive state is highly metastable. Figure 10(c) shows the joint distribution of k and Q from 10^4 trajectories consisting of $t = 10^6$ Monte Carlo steps each at s_c. For this case the active and trapped phases are virtually indistinguishable in terms of their structural nativeness.

[1] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).

[2] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley, Reading, MA, 1992).

[3] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 2000).

[4] L. Peliti, *Statistical Mechanics in a Nutshell* (Princeton University Press, Princeton, NJ, 2011).

[5] E. I. Shakhnovich and A. M. Gutin, Nature **346**, 773 (1990).

[6] E. I. Shakhnovich, Chem. Rev. **106**, 1559 (2006).

[7] G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, Proc. Natl. Acad. Sci. USA **103**, 16623 (2006).

[8] H. Abe and N. Go, Biopolymers **20**, 1013 (1981).

[9] N. Gō, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).

[10] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101**, 1519 (1994).

[11] E. Paci, M. Vendruscolo, and M. Karplus, Proteins **47**, 379 (2002).

[12] J. K. Weber, R. L. Jack, and V. S. Pande, J. Am. Chem. Soc. **135**, 5501 (2013).

[13] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins **21**, 167 (1995).

[14] I. E. T. Iben, D. Braunstein, W. Doster, H. Frauenfelder, M. K. Hong, J. B. Johnson, S. Luck, P. Ormos, A. Schulte, P. J. Steinbach, A. H. Xie, and R. D. Young, Phys. Rev. Lett. **62**, 1916 (1989).

[15] J. Chahine, H. Nymeyer, V. B. P. Leite, N. D. Socci, and J. N. Onuchic, Phys. Rev. Lett. **88**, 168101 (2002).

[16] A. Gutin, A. Sali, V. Abkevich, M. Karplus, and E. I. Shakhnovich, J. Chem. Phys. **108**, 6466 (1998).

[17] M. Merolle, J. P. Garrahan, and D. Chandler, Proc. Natl. Acad. Sci. USA **102**, 10837 (2005).

[18] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, Phys. Rev. Lett. **98**, 195702 (2007).

[19] V. Lecomte, C. Appert-Rolland, and F. van Wijland, J. Stat. Phys. **127**, 51 (2007).

[20] L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, Science **323**, 1309 (2009).

[21] R. L. Jack, L. O. Hedges, J. P. Garrahan, and D. Chandler, Phys. Rev. Lett. **107**, 275702 (2011).

[22] B. Gin, J. P. Garrahan, and P. L. Geissler, J. Mol. Biol. **392**, 1303 (2009).

[23] A. Sali, E. Shakhnovich, and M. Karplus, J. Mol. Biol. **235**, 1614 (1994).

[24] L. G. Garcia and A. F. Pereira de Araújo, Proteins **62**, 46 (2006).

[25] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).

[26] E. I. Shakhnovich, Phys. Rev. Lett. **72**, 3907 (1994).

[27] L. A. Mirny, V. I. Abkevich, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. USA **95**, 4976 (1998).

[28] M. Baiesi, C. Maes, and B. Wynants, Phys. Rev. Lett. **103**, 010602 (2009).

[29] H. Touchette, Phys. Rep. **478**, 1 (2009).

[30] E. S. Loscar, A. S. J. S. Mey, and J. P. Garrahan, J. Stat. Mech. (2011) P12011.

[31] T. Speck, A. Malins, and C. P. Royall, Phys. Rev. Lett. **109**, 195703 (2012).

[32] F. Chiti and C. Dobson, Nat. Chem. Biol. **5**, 15 (2009).

[33] J. P. Garrahan, R. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, J. Phys. A **42**, 075007 (2009).