Reliable Protein Folding on Complex Energy Landscapes: The Free Energy Reaction Path

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

ABSTRACT A theoretical framework is developed to study the dynamics of protein folding. The key insight is that the search for the native protein conformation is influenced by the rate r at which external parameters, such as temperature, chemical denaturant, or pH, are adjusted to induce folding. A theory based on this insight predicts that 1), proteins with complex energy landscapes can fold reliably to their native state; 2), reliable folding can occur as an equilibrium or out-of-equilibrium process; and 3), reliable folding only occurs when the rate r is below a limiting value, which can be calculated from measurements of the free energy. We test these predictions against numerical simulations of model proteins with a single energy scale.

INTRODUCTION

Under appropriate conditions, proteins spontaneously fold from a one-dimensional chain of amino acids to a unique three-dimensional native conformation. How this occurs on timescales accessible to experiment—and relevant to biological function—is a question that has intrigued scientists for the past forty years. Levinthal (1) was the first to recognize the importance of timescales and point out that, assuming a random search of conformation space, proteins would not fold in a person's lifetime. This argument has come to be known as Levinthal's Paradox since proteins must fold for human life to exist in the first place.

Of course, conformation space is not sampled randomly and Levinthal's paradox has been resolved by applying statistical mechanics to the protein folding problem (2–4). Each protein conformation has a free energy that determines its probability to be sampled at temperature T. While the free energy F generally comprises a sum of many enthalpic and entropic terms, it is convenient to express it as $F = E - TS_{\mathrm{conf}}$, where S_{conf} is the conformational entropy of only the protein degrees of freedom and E is the internal energy that includes all other contributions to the free energy (from both protein and solvent). The functional dependence of E on all protein degrees of freedom is called the energy landscape (5,6), which, in general, contains many minima. For low temperatures, only the energy landscape is relevant and the protein resides in a local (or global) minimum, corresponding to a compact conformation. As temperature increases, the conformational entropy smoothes out the minima in the energy landscape and the protein adopts more extended states with larger S_{conf}. In the ''new view'' of protein folding (3,7), statistical fluctuations on an energy landscape give rise to an ensemble of folding pathways.

Often associated with the new view is the hypothesis that energy landscapes have the shape of a multidimensional funnel (4,8–10). Proponents argue that to fold reliably (transition to the native state with probability one) the energy landscape must contain a single low-lying minimum to which all conformations are channeled. If multiple funnels exist, separated by large enough energy barriers, then at low temperature or denaturant concentration a protein can become trapped in a local minimum of energy that does not correspond to its native conformation. While the existence of a single funnel is a sufficient condition for reliable protein folding, the number of proteins with a single funnel is expected to be small and the observation of kinetic traps (11–15) and glassy behavior (16,17) in biologically relevant proteins indicates that not all proteins fold on smooth funneled landscapes.

Here we address the open question: is a funneled energy landscape necessary for reliable folding? By formulating a statistical theory that includes the dynamics of folding, we find that a funneled landscape is not necessary for reliable folding. The important insight is that the rate r at which temperature or chemical denaturant concentration is decreased to induce folding affects the final conformation of the protein. For sufficiently small r, the protein always folds to its native conformation, whereas for larger r it can become trapped in a metastable state. This leads to new predictions that can be tested in experiments and simulations: First, proteins with arbitrary energy landscapes—funneled or not—can fold reliably to their native state if the rate r is below a limiting value. Second, reliable folding can occur as an equilibrium-quasistatic or nonequilibrium process. Third, in a nonequilibrium folding process, a protein can reliably fold to a local (instead of global) minimum of free energy. We conduct off-lattice simulations of model proteins and verify these predictions.

MATERIALS AND METHODS

Here we provide the details of the simulation and numerical methods used to obtain the results discussed in Simulations of Model Proteins. Simulations are performed on polymer chains of spherical monomers, each with diameter

Submitted March 10, 2008, and accepted for publication May 20, 2008.

Address reprint requests to Gregg Lois, Tel.: 203-436-1318; E-mail: gregg.lois@yale.edu.

Editor: Gregory A. Voth.

0006-3495/08/09/2692/10 $2.00

doi: 10.1529/biophysj.108.133132

σ. We include two types of monomers—attractive (green) and nonattractive (white). Interactions depend on the separation r_{ij} between monomers i and j, and it is convenient to define the normalized distance $\bar{r}_{ij} \equiv r_{ij}/\sigma$. Interactions between adjacent monomers are chosen to prevent the polymer chain from breaking, while interactions between nonadjacent monomers are either purely repulsive (for green-white or white-white interactions) or attractive (for green-green interactions). More specifically, monomers that are adjacent on the polymer chain experience a piecewise continuous potential $\Phi_{cc}(\bar{r})$ that is comprised of a purely repulsive Lennard-Jones (LJ) potential (18) for separations $\bar{r}_{ij} \le 1$ and a FENE potential (19) for separations $\bar{r}_{ij} \ge 1$,

$$\Phi_{cc}(\bar{r}_{ij}) = \begin{cases} \epsilon(\bar{r}_{ij}^{-12} - 2\bar{r}_{ij}^{-6} + 1) & \bar{r}_{ij} \le 1 \\ -\epsilon\log(1 - q^{-2}(\bar{r}_{ij} - 1)^2) & \bar{r}_{ij} > 1 \end{cases}, \quad (1)$$

where ϵ sets the energy scale and $q = 0.1$. This potential has a minimum of zero at $\bar{r}_{ij} = 1$ and diverges at $\bar{r}_{ij} = 1+q$ to prevent adjacent monomers from unbinding. Green-green interactions are described by an LJ potential

$$\Phi_{att}(\bar{r}_{ij}) = -\epsilon E_c(\bar{r}_{ij}^{-12} - 2\bar{r}_{ij}^{-6}), \quad (2)$$

with energy depth $E_c < 0$ at $\bar{r}_{ij} = 1$, whereas green-white and white-white interactions obey a repulsive LJ potential

$$\Phi_{rep}(\bar{r}_{ij}) = \begin{cases} \epsilon(\bar{r}_{ij}^{-12} - 2\bar{r}_{ij}^{-6} + 1) & \bar{r}_{ij} \le 1 \\ 0 & \bar{r}_{ij} > 1 \end{cases} \quad (3)$$

that provides a repulsive force only when particles overlap.

Thermal fluctuations are included using off-lattice Brownian dynamics simulations (18). The vector position $\vec{r}_i$ of each monomer i is determined at each time-step by the attractive and repulsive forces arising from the potentials in Eqs. 1–3 and random forces arising from thermal fluctuations. The equation of motion for monomer i is

$$m_i\frac{d^2\vec{r}_i}{d^2t} = \vec{F}_i(t) - \gamma\vec{v}_i - \frac{d}{d\vec{r}_i}\sum_{j\neq i}\left[(\Phi_{cc}(\bar{r}_{ij}) + \Phi_{att}(\bar{r}_{ij}) + \Phi_{rep}(\bar{r}_{ij})\right], \quad (4)$$

where $\vec{F}_i(t)$ is a Gaussian random force, $-\gamma\vec{v}_i$ a damping force, $-\gamma\vec{v}_i$ is a damping force, $\vec{v}_i$ is the velocity of monomer i, $\gamma = \eta\sigma^{d-1}$, η is the solvent viscosity, and d is the spatial dimension. The Gaussian random force has zero mean and a standard deviation proportional to T/η. We solve Eq. 4 using standard numerical integration techniques (18) in the limit that monomer mass $m_i = 0$.

Folding simulations are conducted by starting with $E_c = 0$ and decreasing E_c linearly in time with rate r at constant $T = 1$. Supplementary Material (Data S1) is included online of two movies that show the folding of a two-dimensional polymer chain with an ordered sequence of green and white monomers at $r\eta\sigma^2/T = 10^{-7}$ where folding occurs reliably (Movie S1, ''slowrate.mov'') and at $r\eta\sigma^2/T = 10^{-5}$ where a misfold occurs (Movie S2, ''fastrate.mov'').

We use the simulations to construct energy and free energy landscapes for model proteins. The energy landscape (see Fig. 2) is obtained by running 20 folding simulations at each of five rates $r\eta\sigma^2/T = 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}$, and 10^{-4}. Each simulation explores the range $0 < c < 0.4$ and the energy landscape is averaged over all observed states. We believe that the landscape is sufficiently sampled since we observe no difference at small D and R_g between the energy landscape pictured later in Fig. 2 and ones measured using only data from the smallest r. The free energies (see Fig. 3) are measured by ramping to the desired c-value using $r\eta\sigma^2/T = 5 \times 10^{-9}$, and then calculating a histogram of the probability $P(E, D)$ to have energy E and end-to-end distance D over 10^8 time-steps for each c-value reported. The free energy $F(E, D)$ is determined from the probability via the relation $F(E, D) = -T \log P(E, D)$.

RESULTS

We consider proteins with complex energy landscapes—not necessarily funneled—and derive the conditions under which folding occurs reliably. Generally, energy landscapes contain multiple minima, possibly separated by large energy barriers. Thus, folding is not necessarily an equilibrium process and misfolds can occur. Below we consider the dynamics of the folding process and its effect on reliable folding.

A kinetic mechanism for folding

Multiple minima in the energy landscape lead to multiple minima in the free energy. In this case, we argue that there is a basic kinetic mechanism that determines whether folding is reliable. We illustrate the kinetic mechanism by considering a transition from state **A** to state **B** on a complex energy landscape. Although we will assume that the transition is driven by a reduction of temperature, the same arguments can be applied when a change of denaturant concentration or another parameter induces folding.

In Fig. 1, schematic illustrations of the free energy are plotted at four temperatures $T_1 > T_2 > T_3 > T_4$. We will assume that a transition from **A** $\rightarrow$ **B** is induced by decreasing the temperature at a constant rate r, such that $T(t) = T_1(1 - rt)$ as a function of time t. Initially at T_1 the protein resides in state **A**. As temperature is reduced to T_2, an equilibrium transition to state **B** can occur with folding time equal to $\exp(\Delta F/T_2)/r^*$, where r^*, which depends on temperature and other physical parameters, is the microscopic rate at which conformations are sampled and ΔF is the free energy barrier between states **A** and **B**. At T_3, a third state **M** has free energy equal to that of **A**. As temperature is further reduced to T_4, the minimum corresponding to state **A** no longer exists and the activation barrier $\Delta F'$ between states **M** and **B** grows.

Dynamics are important in determining transitions from **A** $\rightarrow$ **B**. If the time that it takes for the temperature to decrease from T_2 to T_3 is less than the folding time, the protein can fall into the metastable state **M**. This sets a bound on r: if

$$r > r^f \equiv \frac{(T_2 - T_3)r^*}{T_1}\exp\left(\frac{-\Delta F}{T_2}\right), \quad (5)$$

then the protein is likely to populate the state **M**. The limiting rate r^f is inversely proportional to the equilibrium folding time $\exp(\Delta F/T_2)/r^*$ and proportional to $(T_2 - T_3)$, where T_2 (T_3) is the temperature at which states **B** (**M**) and **A** have the same free energy. Note that we use units where Boltzmann's constant $k_B = 1$.

For a misfold to occur, the escape probability from the metastable state must be sufficiently small. If the protein populates state **M** at time t_3, the probability that it has not escaped at time τ is given by

$$P(\tau) = \exp\left(-\int_{t_3}^{\tau} dt\, r^*\exp(-\Delta F'(T)/T)\right) \equiv \exp(-g(\tau)). \quad (6)$$

For a maximum waiting time τ, the protein always escapes the metastable state for $g(\tau) \gg 1$ and rarely escapes for

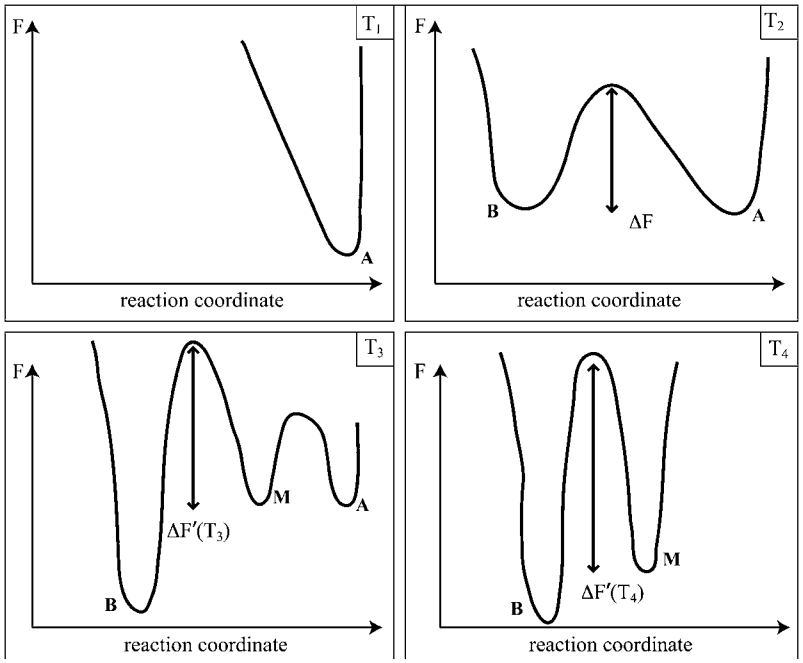

FIGURE 1 Schematic plots of the free energy versus an arbitrary reaction coordinate at four temperatures where $T_1 > T_2 > T_3 > T_4$. At T_1 only the state **A** is accessible. At T_2, transitions to state **B** occur with activation barrier ΔF. T_3 is defined as the largest temperature at which a new state **M** exists with free energy equal to that of state **A**. If the protein has not transitioned to state **B** by T_3, misfolds can occur. At T_4 the free energy barrier $\Delta F'(T)$ separating **M** and **B** becomes larger than it was at T_3.

$g(\tau) \ll 1$. The crossover between frequently escaping from and being trapped in state **M** occurs when $g(\tau) \approx 1$. Using $T(t) = T_1(1 - rt)$, we find that when the rate is

$$r > r^{s} \equiv \int_0^{T_3} r^{*} \exp\left(\frac{-\Delta F'(T)}{T}\right)\frac{dT}{T_1}, \tag{7}$$

the probability to become trapped in the metastable state **M** is significant and misfolds occur. (Since we use a waiting time τ that satisfies $T(\tau) = 0$, the lower limit of the integral in Eq. 3 is zero. Note that the limiting rates r^{f} and r^{s} can be determined for any functional form $T(t)$ and maximum waiting time τ. In the case that the final temperature is nonzero, r^{s} includes a term that grows linearly with waiting time and reliable folding at time τ only occurs if $r < r^{f}$ or $r < r^{s}$.)

From these basic considerations, it is apparent that protein folding transitions are influenced by multiple minima in the free energy and the rate r at which external parameters are varied to induce folding. To determine whether reliable folding occurs, we must address two important questions:

1. Can the protein conformation reside in a metastable local minimum?
2. Is it likely that the protein conformation becomes trapped in that local minimum?

The answers to these questions define the limiting rates r^{f} and r^{s}. The transition $\mathbf{A} \rightarrow \mathbf{B}$ occurs reliably if r obeys one of the inequalities, $r < r^{f}$ or $r < r^{s}$. In the case that $r < r^{s}$, the protein is given sufficient time to sample all states and the transition $\mathbf{A} \rightarrow \mathbf{B}$ occurs reliably as an equilibrium process. If $r^{s} < r < r^{f}$, the protein conformation becomes trapped in the state **B** without fully exploring phase space and the transition occurs reliably, but out of equilibrium. If $r > r^{f}$ and $r > r^{s}$, then the protein does not transition between **A** and **B** reliably.

The free energy reaction path

In the previous section we identified a kinetic mechanism that influences conformational transitions on complex energy landscapes. In this section we use this mechanism to formulate a general framework for understanding folding. We begin by partitioning the energy landscape into basins associated with particular protein topologies, proceed to define the free energy reaction path that describes how the protein transitions from one topology to another, and then use the kinetic mechanism described above to determine whether folding is reliable.

As a way to understand complex folding dynamics, the energy landscape of an arbitrary protein can be partitioned into basins surrounding each local minimum, analogous to the inherent structure formalism for liquids and glasses (20). In particular, the infinite number of protein conformations can be uniquely associated with a finite number of topologies, defined as protein conformations that are local minima of the internal energy. We denote a topology as $\mathbf{t}^{n}$, where n is an index that contains sufficient information to fully describe the conformation (e.g., number, type, and arrangement of bonds). The set of conformations $\mathcal{B}(\mathbf{t}^{n})$ associated with each topology $\mathbf{t}^{n}$ is the basin of attraction for that topology. The basin of attraction is defined such that all conformations that belong to $\mathcal{B}(\mathbf{t}^{n})$ relax to the topology $\mathbf{t}^{n}$ when thermal fluctuations of the protein are suppressed. Thus the infinite number of possible protein conformations is represented by a finite number of topologies and a free energy $F(\mathbf{t}^{n})$ can be

defined for the set of protein conformations $\mathcal{B}(\mathbf{t}^n)$. Formally the partition function $Z(\mathbf{t}^n)$ for conformations constrained to lie in $\mathcal{B}(\mathbf{t}^n)$ is given by

$$Z(\mathbf{t}^n) = \int_{\mathcal{B}(\mathbf{t}^n)} \exp(-E/T)\, d\Gamma, \quad (8)$$

where integration is over all coordinates Γ in the basin $\mathcal{B}(\mathbf{t}^n)$ and E is the internal energy as a function of Γ. The free energy for a protein constrained to $\mathcal{B}(\mathbf{t}^n)$ can then be written in terms of the topology $\mathbf{t}^n$ as

$$F(\mathbf{t}^n, T) = E(\mathbf{t}^n, T) - TS_{\text{conf}}(\mathbf{t}^n, T), \quad (9)$$

where $E(\mathbf{t}^n, T)$ is the internal energy of topology $\mathbf{t}^n$ and $S_{\text{conf}}(\mathbf{t}^n, T)$ is its associated entropy (20), given by

$$S_{\text{conf}}(\mathbf{t}^n, T) = \log \int_{\mathcal{B}(\mathbf{t}^n)} \exp(-[E - E(\mathbf{t}^n, T)]/T)\, d\Gamma. \quad (10)$$

The random coil state $\mathbf{t}^0$ with zero internal energy has the largest entropy and is therefore the global minimum of free energy at sufficiently large temperature.

Given a protein energy landscape that has been partitioned into a finite number of basins of attraction, master-equation approaches (21) can be used to predict the probabilities at which all topologies are sampled at temperature T. Heterogeneity in folding, i.e., multiple folding pathways for a single protein, occurs because statistical fluctuations determine the sampling probabilities (22). However, at each T there is a topology that is sampled most frequently, which is the topology with the lowest free energy. While master-equation approaches treat constant T, reliable folding depends on how the protein samples the energy landscape, which changes as external parameters are varied. Using the kinetic mechanism introduced in the previous section, we focus here on how the most-likely topology changes as T is reduced and make no assumptions about the transition pathways between the most-likely topologies (we find that reliable folding can be predicted by only including information about the most-likely topologies).

We define the free energy reaction path as the ordered sequence of most-likely topologies that the protein adopts as temperature is reduced in the equilibrium limit. That is, if the rate r is sufficiently small, the protein will come to equilibrium at all temperatures and proceed through the basins of attraction for a reproducible set of most-likely topologies $\mathbf{t}^0 \rightarrow \mathbf{t}^{n_1} \rightarrow \mathbf{t}^{n_2} \rightarrow \cdots \rightarrow \mathbf{t}^{n_N}$. Each transition occurs at the temperature where the free energy of two topologies is equal, e.g., the transition $\mathbf{t}^0 \rightarrow \mathbf{t}^{n_1}$ occurs at the temperature T^* where $F(\mathbf{t}^0, T^*) = F(\mathbf{t}^{n_1}, T^*)$. Note that free energy barriers between topologies are not relevant in the equilibrium limit since the protein explores its conformation space ergodically. Thus, for any energy landscape, the free energy reaction path is defined as the path taken through conformation space when folding occurs as an equilibrium-quasistatic process.

To determine whether folding is reliable, we apply the analysis introduced in the previous section to each transition in the free energy reaction path. If we label the transitions by $i = 1, 2, \ldots, N$, then limiting rates r_i^f and r_i^s can be determined for each transition by measuring properties of the free energy. There are then three distinct folding scenarios:

1. If $r < r_i^s$ for all i, then the protein does not become trapped in metastable conformations and folding occurs reliably in equilibrium.
2. If $r_i^s < r < r_i^f$ for a single transition i, then the protein falls out of equilibrium at transition i, but reliably folds to the topology $\mathbf{t}^{n_i}$ (since the condition $r < r_i^f$ guarantees that the protein does not fall into a different metastable state).

Note that if there exist multiple transitions with $r_i^s < r < r_i^f$, then the protein will reliably fold to the topology with the smallest value of i for which this condition holds. Finally,

3. If $r > r_i^s$ and $r > r_i^f$ for any i, and condition 2 does not hold for a smaller value of i, then the protein will not fold reliably.

From our analysis, we deduce that there are two types of reliable folding: equilibrium and nonequilibrium. While reliable equilibrium folding brings the protein to the global minimum of free energy, reliable nonequilibrium folding can target local minima. The free energy reaction path provides a useful framework to classify the relevant transitions since, depending on the rate r, a protein will do one of three things: pass through all topologies on the free energy reaction path and arrive at the topology with the smallest free energy; target an intermediate topology along the free energy reaction path and reliably fold to a local minimum of free energy; or misfold and deviate from the free energy reaction path.

Simulations of model proteins

To test the predictions of the previous section we perform off-lattice Brownian dynamics simulations of model proteins with a single attractive energy scale. We model a protein as a polymer chain containing both attractive (green) and non-attractive (white) spherical monomers of size σ. Interactions between nonadjacent green monomers are attractive with energy depth $E_c < 0$, while interactions between nonadjacent pairs of green-white or white-white monomers are purely repulsive. This model is a variant of the HP model (23). Thermal fluctuations of the protein at temperature T are included using Brownian dynamics simulations with solvent viscosity η. We observe that as the parameter $c = |E_c|/T$ increases from zero, the polymer chain transitions from a random coil to a folded conformation.

Simulations in two dimensions

To test the predictions of the theory we begin with a two-dimensional protein to simplify identification of the multiple topologies that are adopted. We consider a three-dimensional protein in the following section. In two dimensions, we

simulate several sequences of green and white monomers, both random and ordered, but focus the discussion on the specific sequence pictured in Fig. 2.

In Fig. 2, we plot the energy landscape of the polymer chain as a function of two reaction coordinates: the radius of gyration R_g and the end-to-end distance D, each normalized by the monomer diameter σ. In terms of these two reaction coordinates, three energy minima exist and are separated by energy barriers. The minima correspond to three distinct topologies that are pictured in Fig. 2. We find a total of four relevant topologies for this simple system, containing either zero $\mathbf{t}^0$, three $\mathbf{t}^3$, four $\mathbf{t}^4$, or five $\mathbf{t}^5$ bonds between attractive green monomers. Energy barriers exist among $\mathbf{t}^3$, $\mathbf{t}^4$, and $\mathbf{t}^5$ because, to transition between the topologies, it is necessary to first break a bond and then rearrange the chain conformation. Note that four green particles is the minimum number needed to ensure multiple energy minima in two dimensions, while seven are required in three dimensions. Including additional green particles introduces additional minima and more complex energy landscapes—we treat only the simplest case here.

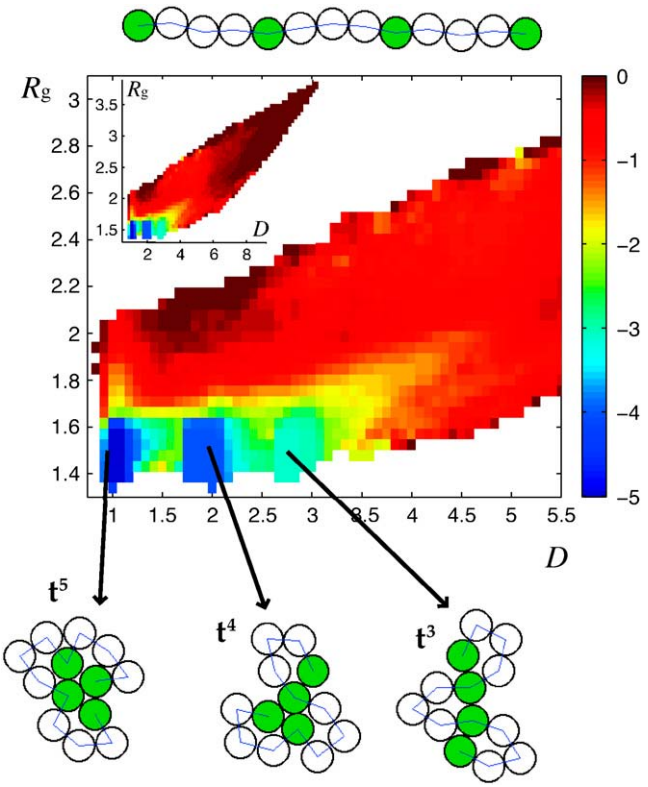

FIGURE 2 Contour plot of the energy landscape and pictures of the relevant topologies for a model protein in two dimensions. The fully extended conformation is shown at the top of the figure. The inset displays the full energy landscape and the main figure contains a magnified view of the compact states. The landscape is plotted as a function of the radius of gyration R_g and end-to-end distance D, each normalized by the monomer diameter. The color bar gives the total internal energy of the protein divided by the attraction strength $|E_c|$. There are three distinct energy minima separated by barriers and the associated topologies are pictured. Open regions correspond to protein conformations that are never sampled in the simulations.

The energy landscape of the simulated protein contains multiple low-lying minima separated by energy barriers, as is the case for many realistic proteins. We now determine the associated free energy reaction path. Measurements of free energy F/T, normalized by temperature, as a function of $E/|E_c|$ and end-to-end distance D are shown in Fig. 3 for a sequence of c-values that corresponds to the sequence of schematic plots in Fig. 1. In Fig. 3 *a*, we plot F/T for a small value of $c = 0.0040$ and observe that the random coil state $\mathbf{t}^0$ is the only free energy minimum. In Fig. 3 *b*, the value c is increased to $c_2 = 0.0085$ and there are multiple local minima in the free energy, including the topologies $\mathbf{t}^0$, $\mathbf{t}^1$, $\mathbf{t}^3$, and $\mathbf{t}^5$. The free energies of $\mathbf{t}^0$ and $\mathbf{t}^5$ are equal in Fig. 3 *b*. At a slightly higher value $c = c_3 = 0.0100$, Fig. 3 *c* exhibits three minima and the free energy of $\mathbf{t}^0$ and $\mathbf{t}^3$ are equal. Finally at $c = 0.0145$, the free energy plotted in Fig. 3 *d* exhibits a deep minimum at topology $\mathbf{t}^5$.

From the plots in Fig. 3, we conclude that the first and only transition in the free energy reaction path is $\mathbf{t}^0 \rightarrow \mathbf{t}^5$. Since there are only two topologies in the free energy reaction path, the protein will either fold reliably to $\mathbf{t}^5$ or fold unreliably to one of the three free energy minima $\mathbf{t}^3$, $\mathbf{t}^4$, and $\mathbf{t}^5$. Reliable folding to one of the local free energy minima $\mathbf{t}^3$ or $\mathbf{t}^4$ is not possible in this case since they are not a part of the free energy reaction path. In the Appendix, we calculate the limiting rates $r^f\eta\sigma^2/T = 1.8 \times 10^{-7}$ and $r^s\eta\sigma^2/T = 3.0 \times 10^{-8}$ for the single transition on the free energy reaction path, where $\eta\sigma^2/T$ is the simulation time-unit.

Now that we have determined the free energy reaction path and calculated the limiting rates, we conduct simulations in which external parameters are varied in time to induce folding. We increase the energy scale c linearly in time at rate r ($c = rt$), starting from the topology $\mathbf{t}^0$ at $c = 0$. In Fig. 4 *a*, the energy of the polymer chain is plotted as a function of c for three different values of r, with the final state labeled by its topology. From this figure we clearly see that small r targets the native state $\mathbf{t}^5$ whereas larger r leads to misfolding. In Fig. 4 *b*, we plot the probability to fold to the native state $\mathbf{t}^5$ as a function of $r\eta\sigma^2/T$, averaged over many folding trajectories studied for each r. The protein folds reliably for small rates. We have also conducted simulations on model proteins with random sequences of the same number of green and white monomers. Reliable folding also occurs at low rates for these random sequences, although the critical rates vary with sequence.

The modern theory of protein folding requires funneled energy landscapes for reliable folding (4,8–10). The simple protein model we consider here provides a contradiction to this viewpoint since there are multiple minima, none of which is especially deep, and it nevertheless folds reliably at small r. The free energy reaction path theory predicts that reliable folding can occur on arbitrary energy landscapes and

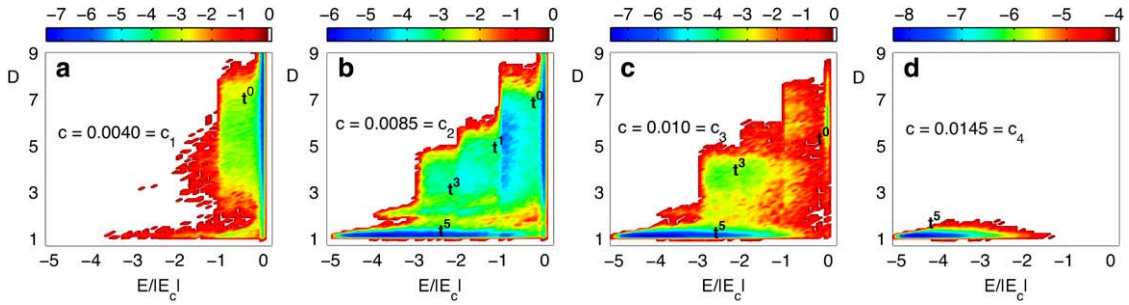

FIGURE 3 Contour plots of the free energy F/T normalized by temperature for the two-dimensional protein pictured in Fig. 2 as a function of $E/|E_c|$ (*horizontal axis*) and end-to-end distance D (*vertical axis*) for a sequence of c-values. The free energy is calculated from the probability for the protein to be in a conformation with given $E/|E_c|$ and D. Open regions correspond to protein conformations that are never sampled in the simulations.

provides a means to quantitatively determine the limiting rate below which folding is reliable. Given the values of r^f and r^s quoted above, the free energy reaction path theory predicts reliable folding for $r\eta\sigma^2/T < 1.8 \times 10^{-7}$. In Fig. 4 *b*, we have measured that reliable folding occurs for normalized rates $\lesssim 10^{-7}$. The theory therefore makes a correct quantitative prediction of the simulation results. Additionally, the values of r^f and r^s indicate that there is a range of rates $r^s < r < r^f$ where reliable folding to $\mathbf{t}^5$ occurs out of equilibrium. We test this prediction by measuring energy fluctuations for rates at which folding is reliable, as plotted in Fig. 4 *c*. For $r \leq r^s$ fluctuations are large at the transition point $c = 0.0085$, because the protein is sampling both folded and unfolded conformations as it remains in equilibrium. For $r > r^s$, fluctuations remain small near the transition point since the protein becomes trapped in the folded state and reliable folding is a nonequilibrium process.

Simulations in three dimensions

We have also tested the predictions of the free energy reaction path theory in three dimensions and find similar results. We study the model protein with the ordered sequence pictured in Fig. 5 that consists of 25 monomers, seven of which are attractive. In Fig. 5, we plot the protein energy landscape as a function of the radius of gyration R_g and end-to-end distance D, each normalized by the monomer diameter σ. There are two minima at small R_g and D, corresponding to the topologies $\mathbf{t}^{15}$ and $\mathbf{t}^{16}$ pictured in the figure.

The limiting rate below which folding is reliable can be predicted by measurements of free energy. In Fig. 6, we plot the free energy as a function of end-to-end distance D and normalized energy $E/|E_c|$ for many different values of c. In Fig. 6 *a*, the random coil state $\mathbf{t}^0$ is the only minimum in the free energy. For $c = 0.0067$, Fig. 6 *b* demonstrates that $\mathbf{t}^{16}$ and $\mathbf{t}^0$ have equal free energies. In Fig. 6 *c*, the random coil $\mathbf{t}^0$, native state $\mathbf{t}^{16}$, and metastable state $\mathbf{t}^{15}$ basins of attraction are present. At this value of $c = 0.0072$, topology $\mathbf{t}^{15}$ has a free energy equal to that of $\mathbf{t}^0$. For larger c, Fig. 6 *d* demonstrates that the protein has an increasing probability to populate the basin of attraction for $\mathbf{t}^{16}$, although the basin of attraction for $\mathbf{t}^{15}$ is still visible. From this series of free energy plots, it is apparent that the simulated protein possesses a single equilibrium transition at $c = c_2$ from $\mathbf{t}^0$ to $\mathbf{t}^{16}$, and misfolds to $\mathbf{t}^{15}$ are possible for $c > c_3$.

Given the data in Fig. 6, we conclude that the first and only transition in the free energy reaction path is $\mathbf{t}^0 \rightarrow \mathbf{t}^{16}$, where the protein folds to its native conformation. In the Appendix, we calculate the limiting rates $r^f\eta\sigma^3/T = 2.7 \times 10^{-7}$ and $r^s\eta\sigma^3/T = 2.3 \times 10^{-6}$ for the single transition on the free energy reaction path, where $\eta\sigma^3/T$ is the simulation time-unit in three dimensions.

Given the values of r^f and r^s, we expect this protein to fold reliably for $r\eta\sigma^3/T < 2.3 \times 10^{-6}$, which is consistent with the data in Fig. 7 *b*. In contrast to the two-dimensional simulations, we find $r^f < r^s$ and thus this particular protein can only fold in equilibrium. Generally we believe that the ordering of r^f and r^s can depend on the length, sequence, and energy scales of the protein.

In both two and three dimensions we have demonstrated that the folding of model proteins (with both ordered and random sequences) is dependent on the rate that external parameters are adjusted to induce folding. The free energy reaction path theory allows us to calculate the limiting rate below which folding is reliable, and we find quantitative agreement with the results of simulations. Since rate dependence is important for the simple model proteins we consider here, we expect that it will also play an important role in proteins of biological importance.

DISCUSSION

Levinthal was the first to realize that the exponential number of collapsed conformations preclude a protein from finding its native state via random sampling. The experimental observation that proteins fold reliably to a reproducible native state therefore requires an explanation. The modern view is that protein sequences have evolved to favor energy landscapes with a single funnel and can therefore fold reliably. We have demonstrated that proteins with complex energy landscapes can also fold reliably, as long as the external

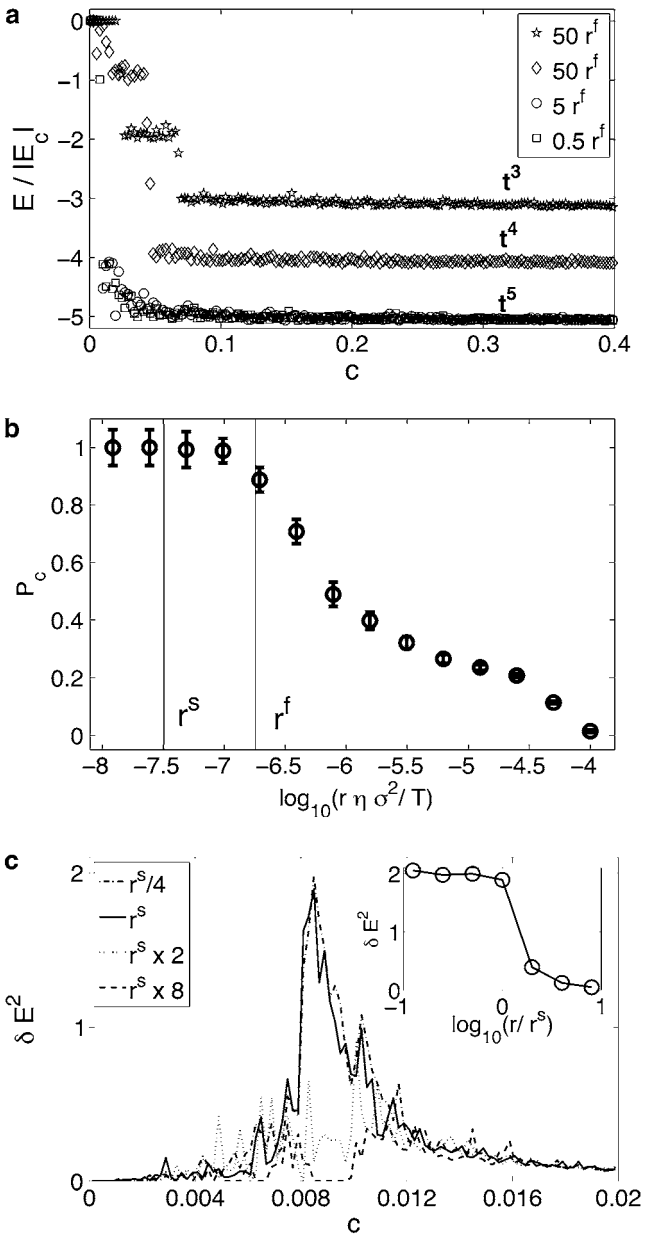

FIGURE 4 Results from folding simulations of the two-dimensional protein pictured in Fig. 2. (*a*) Folding trajectories from simulations with identical initial conditions at three different rates. The normalized energy $E/|E_c|$ is plotted as a function of c and the final state is labeled by its topology. Slow rates $r \leq 5r^f$ lead to the native state t^5 whereas fast rates lead to unreliable folding. (*b*) The probability of folding to the native state P_c as a function of rate r. Error bars are from sampling statistics. For $r\eta\sigma^2/T \lesssim 10^{-7}$, the protein folds reliably to the topology t^5. Vertical lines indicate the values of r^f and r^s calculated in the text. (*c*) Energy fluctuations $\delta E^2 = (\langle E^2\rangle - \langle E\rangle^2)/E_c^2$ as a function of c for folding simulations at different rates r. For $r \leq r^s$, the fluctuation curves appear to collapse and reliable folding occurs in equilibrium. For $r^s < r < r^f$, fluctuations depend on r and reliable folding occurs out of equilibrium. (*Inset*) Energy fluctuations at the equilibrium transition point $c = c_2 = 0.0085$ as a function of r/r^s.

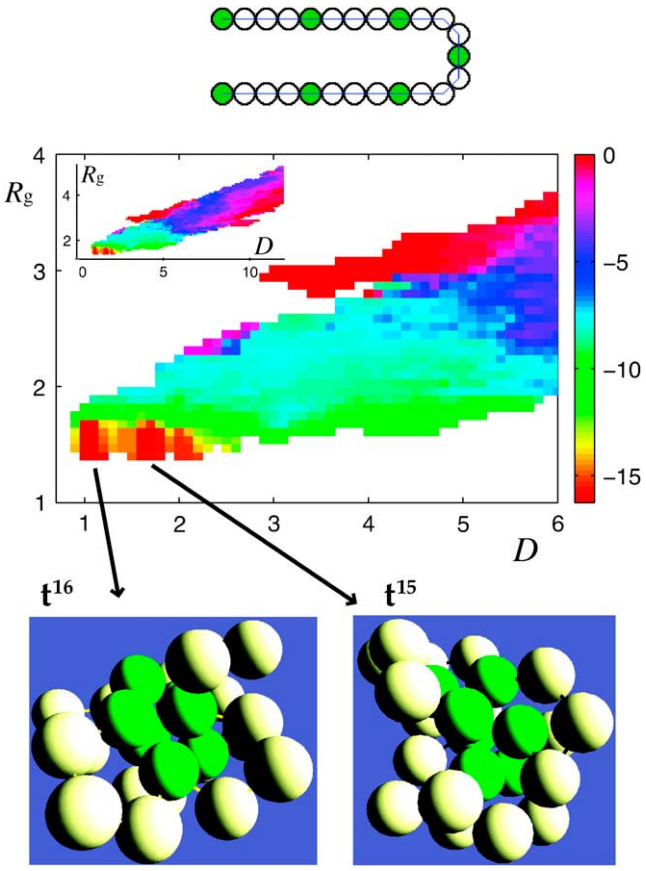

FIGURE 5 Energy landscape and relevant topologies for a three-dimensional model protein, pictured in an extended state with no bonds at the top of the figure. The inset is the full energy landscape, and the main figure contains a magnified view of the compact states. The color bar gives the total energy of the system normalized by the magnitude of the attraction strength $|E_c|$. There are two distinct energy minima separated by barriers and the topologies of each minima are pictured and labeled. Open regions correspond to protein conformations that are never sampled in the simulations.

parameters that induce folding are adjusted slowly enough. Thus the properties of the energy landscape are not sufficient to determine whether a protein will fold reliably. Instead, one must consider both dynamical effects and properties of the landscape to predict whether folding is reliable. In the limit that $r^f \rightarrow \infty$, reliable folding is ensured for all rates at which external parameters are adjusted to induce folding. This limit provides a natural definition for a funneled energy landscape since it is the only case where reliable folding is independent of rate.

Our predictions can be tested in experiments by studying folding over a range of rates, using methods such as ultrafast mixing or laser pulsing (24,25). Since the critical rates r^f and r^s depend on the underlying energy landscape, measuring their values provides a relatively simple means to extract information about the landscapes of proteins. This information can be used as a tool to further characterize folding processes in different proteins. Some progress has been made

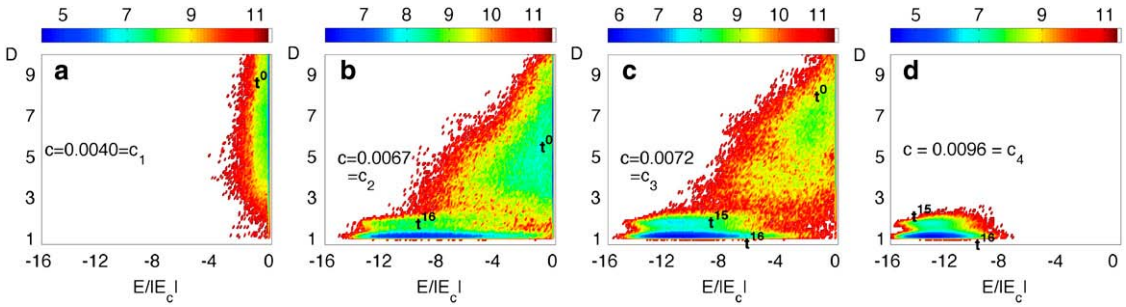

FIGURE 6 Contour plots of the free energy F/T normalized by the temperature for the three-dimensional protein pictured in Fig. 5, as a function of the normalized energy $E/|E_c|$ (*horizontal axis*) and end-to-end distance D (*vertical axis*) for four values of c. Open regions correspond to protein conformations that are never sampled in the simulations.

in this direction (26–28), and the observation of non-exponential relaxation (29) after rapid temperature jumps is consistent with our predictions. In three dimensions, the limiting rates are proportional to $r^* \sim T/\eta R_H^3$, where R_H is the hydrodynamic radius and η is the viscosity. This implies that investigations of folding in solvents with varying viscosities can greatly increase the range of experimentally accessible rates. Even in water, we estimate $r^f \lesssim 10^4\ s^{-1}$, which is easily accessible in current laser pulsing experiments (24). Moreover, due to the dependence on T, folding by changing temperature will give different limiting rates than folding by reducing denaturant concentration.

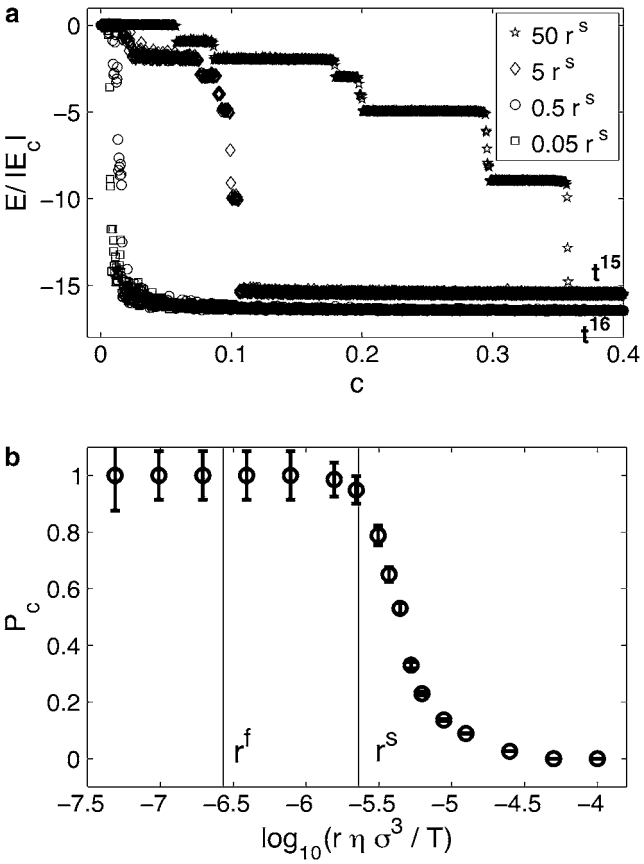

FIGURE 7 Results from folding simulations of the three-dimensional protein pictured in Fig. 5. (*a*) Folding trajectories for the three-dimensional protein in simulations with identical initial conditions at four different rates. The normalized energy $E/|E_c|$ is plotted as a function of c and the final state is labeled by its topology. Slow rates $r \leq 0.5r^s$ find the native state $\mathbf{t}^{16}$ reliably whereas fast rates give rise to unreliable folding. (*b*) The probability P_c of folding to the native state $\mathbf{t}^{16}$ as a function of rate r. Error bars are from sampling statistics. For $r\eta\sigma^3/T \lesssim 2 \times 10^{-6}$, the system folds reliably. Vertical lines indicate the values of r^f and r^s calculated in the text.

We have identified two distinct types of reliable folding: equilibrium and nonequilibrium. Even if the rate at which thermodynamic parameters are varied to induce folding is too large to access the equilibrium limit in some biological settings, reliable folding can occur out of equilibrium. If this is the case, the native state should be regarded as a reliably targeted local minimum on the free energy reaction path that remains metastable over timescales sufficient for biological function.

The importance of the free energy reaction path and the necessity of using small rates to vary external parameters presents challenges for protein folding simulations. Reliable protein folding is especially difficult to study in all-atom simulations where, due to the long timescales and large number of atoms, extremely rapid rates are used to induce folding (30,31). From our results, reliable folding depends on rate; thus, simulation studies that argue that funneled energy landscapes are necessary for reliable folding (32,33) must be carefully interpreted if only large rates are considered.

Finally, it is intriguing to speculate about folding in vivo. Given that the folded state of a protein is dependent on rate at which external parameters are varied to induce folding, and that local minima in free energy can be targeted by adjusting this rate, it is possible that protein sequence has evolved along with the biological environment in which it folds. Since the folding process is determined by protein sequence and rate, both are likely used in nature to ensure robust folding.

APPENDIX

In this Appendix we calculate the values r^f and r^s quoted in Simulations of Model Proteins, above. The limiting rates can be determined using equations similar to those in Eqs. 5 and 7,

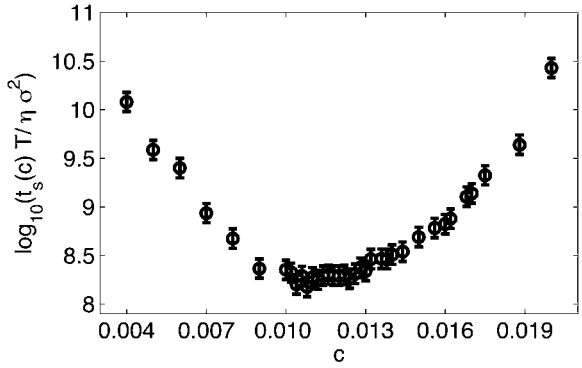

FIGURE 8 Average time for the two-dimensional protein pictured in Fig. 2 to transition from $\mathbf{t}^3$ to $\mathbf{t}^5$ as a function of c.

$$r^{\mathrm{f}} = (c_3 - c_2) r^* \exp\left(\frac{-\Delta F}{T}\right), \tag{11}$$

$$r^{\mathrm{s}} = \int_{c_3}^{\infty} r^* \exp\left(\frac{-\Delta F'(c)}{T}\right) dc. \tag{12}$$

These equations are derived for the simulation protocol where $|E_{\mathrm{c}}| = cT$ increases linearly in time to induce folding, with T constant. The maximum waiting time is taken to infinity.

Two-dimensional proteins

Here we calculate the critical rates for the ordered sequence pictured in Fig. 2. We first calculate r^{f}. Data in Fig. 3 gives $c_2 = 0.0085$ and $c_3 = 0.01$. The free energy barrier $\Delta F/T$ is measured by preparing the protein in topology $\mathbf{t}^5$ at $c = c_2$ and measuring the amount of time t_{f} required to transition to $\mathbf{t}^0$, averaged over 100 trials. The free energy barrier is related to the transition time by $t_{\mathrm{f}} = \exp(\Delta F/T)/r^*$. We measure $t_{\mathrm{f}}T/\eta\sigma^2 = 8400$, where $\eta\sigma^2/T$ is the simulation time-unit. Inserting these numbers into Eq. 11 yields $r^{\mathrm{f}}\eta\sigma^2/T = 1.8 \times 10^{-7}$.

The rate r^{s} is determined by preparing the protein in topology $\mathbf{t}^3$ and measuring the average time $t_{\mathrm{s}}(c)$ required to transition to the native topology $\mathbf{t}^5$. We average $t_{\mathrm{s}}(c)$ over 100 trials for each c-value and it is plotted in Fig. 8. Since $t_{\mathrm{s}}(c) = \exp(\Delta F'(c)/T)/r^*$, we calculate $r^{\mathrm{s}}\eta\sigma^2/T = 3.0 \times 10^{-8}$ by direct integration of $t_{\mathrm{s}}(c)^{-1}$, according to Eq. 12. Contributions to the value of r^{s} from $c > 0.02$ are negligible.

Three-dimensional proteins

Here we calculate the critical rates for the ordered sequence pictured in Fig. 5. The rate r^{f} is calculated using the values $c_2 = 0.0067$ and $c_3 = 0.0072$ determined from Fig. 6, along with the transition time t_{f} from $\mathbf{t}^{16}$ to $\mathbf{t}^0$ at $c = 0.0067$. We measure $t_{\mathrm{f}}T/\eta\sigma^3 = 1850$, averaged over 100 trials. Given these values, we calculate $r^{\mathrm{f}}\eta\sigma^3/T = 2.7 \times 10^{-7}$.

The rate r^{s} is calculated by measuring the transition time $t_{\mathrm{s}}(c)$ between topologies $\mathbf{t}^{16}$ and $\mathbf{t}^{15}$, which is shown in Fig. 9. Directly integrating this data for $c > c_3$ yields $r^{\mathrm{s}}\eta\sigma^3/T = 2.3 \times 10^{-6}$.

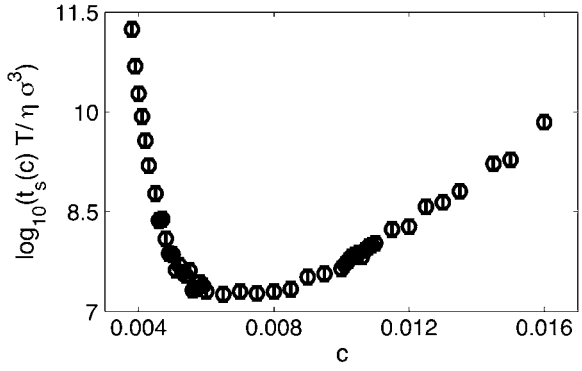

FIGURE 9 Average time for the three-dimensional protein pictured in Fig. 5 to transition from $\mathbf{t}^{15}$ to $\mathbf{t}^{16}$ as a function of c.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

Financial support from National Science Foundation grants No. CBET-0348175 (to G.L. and J.B.) and DMR-0448838 (to G.L. and C.S.O.), and Yale's Institute for Nanoscience and Quantum Engineering (G.L.), is gratefully acknowledged. We also thank Yale's High Performance Computing Center for computing time.

REFERENCES

1. Levinthal, C. 1969. Mossbauer Spectroscopy in Biological Systems. Proceedings of a meeting held at Allerton House. P. Debrunner, J. C. M. Tsibris, and E. Munck, editors. University of Illinois Press, Urbana, IL.
2. Zwanzig, R., A. Szabo, and B. Bagchi. 1992. Levinthal's paradox. *Proc. Natl. Acad. Sci. USA*. 89:20–22.
3. Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* 4:10–19.
4. Plotkin, S. S., and J. N. Onuchic. 2002. Understanding protein folding with energy landscape theory part I: basic concepts. *Q. Rev. Biophys.* 35:111–167.
5. Wales, D. J. 2003. Energy Landscapes. Cambridge University Press, Cambridge, UK.
6. Chan, H. S., and K. A. Dill. 1998. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins*. 30:2–33.
7. Baldwin, R. L. 1995. The nature of protein folding pathways: the classical versus the new view. *J. Biomol. NMR*. 5:103–109.
8. Bryngelson, J. D., and P. G. Wolynes. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA*. 84:7524–7528.
9. Leopold, P. E., M. Montal, and J. N. Onuchic. 1992. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*. 89:8721–8725.
10. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
11. Kiefhaber, T. 1995. Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. USA*. 92:9029−9090.
12. Hua, Q.-X., S. H. Gozani, R. E. Chance, J. A. Hoffmann, B. H. Frank, and M. A. Weiss. 1995. Structure of a protein in a kinetic trap. *Nat. Struct. Biol.* 2:129–138.
13. Pan, T., and T. R. Sosnick. 1997. Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity. *Nat. Struct. Biol.* 4:931–938.
14. Chang, J.-Y., L. Li, and P.-H. Lai. 2001. A major kinetic trip for the oxidative folding of human epidermal growth factor. *J. Biol. Chem.* 276:4845–4852.
15. Im, H., M.-S. Woo, K. Y. Hwang, and M.-H. Yu. 2002. Interactions causing the kinetic trap in serpin protein folding. *J. Biol. Chem.* 277:46347–46354.
16. Iben, I. E. T., 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. 1989. Glassy behavior of a protein. *Phys. Rev. Lett.* 62:1916–1919.

17. Young, R. D., H. Frauenfelder, J. B. Johnson, D. C. Lamb, G. U. Nienhaus, R. Philipp, and R. Scholl. 1991. Time- and temperature dependence of large-scale conformational transitions in myoglobin. *Chem. Phys.* 158:315–327.

18. Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation of Liquids. Oxford University Press, Oxford, UK.

19. Grest, G. S., and K. Kremer. 1986. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A.* 33:3628–3631.

20. Stillinger, F. H., and T. A. Weber. 1982. Hidden structure in liquids. *Phys. Rev. A.* 25:978–989.

21. Cieplak, M., M. Henkel, J. Karbowski, and J. R. Banavar. 1998. Master equation approach to protein folding and kinetic traps. *Phys. Rev. Lett.* 80:3654–3657.

22. Carr, J. M., and D. J. Wales. 2005. Global optimization and folding pathways of selected α-helical proteins. *J. Chem. Phys.* 123:234901.

23. Stillinger, F. H., T. Head-Gordon, and C. L. Hirshfeld. 1993. Toy model for protein folding. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 48:1469–1477.

24. Eaton, W. A., V. Munoz, P. A. Thompson, C.-K. Chan, and J. Hofrichter. 1997. Submillisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.* 7:10–14.

25. Eaton, W. A., V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry, and J. Hofrichter. 2000. Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 29:327–359.

26. Ballew, R. M., J. Sabelko, and M. Gruebele. 1996. Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. USA.* 93:5759–5764.

27. Takahashi, S., S.-R. Yeh, T. K. Das, C.-K. Chan, D. S. Gottfried, and D. L. Rousseau. 1997. Folding of cytochrome *c* initiated by submillisecond mixing. *Nat. Struct. Biol.* 4:44–50.

28. Causgrove, T. P., and R. B. Dyer. 2006. Nonequilibrium protein folding dynamics: laser-induced pH-jump studies of the helix-coil transition. *Chem. Phys.* 323:2–10.

29. Sabelko, J., J. Ervin, and M. Gruebele. 1999. Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci. USA.* 96:6031–6036.

30. Shea, J.-E., and C. L. Brooks III. 2001. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* 52: 499–535.

31. Snow, C. D., E. J. Sorin, Y. M. Rhee, and V. S. Pande. 2005. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* 34:43–69.

32. Sali, A., E. Shakhnovich, and M. Karplus. 1994. How does a protein fold? *Nature.* 369:248–251.

33. Sali, A., E. Shakhnovich, and M. Karplus. 1994. Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *J. Mol. Biol.* 235:1614–1636.