SI Text

Muñoz–Eaton (ME) Model. The free energy level of a protein conformation is described by an effective free energy function of ME model; $\Delta G(\{S_k\}) = J \sum_{<i,j>} \varepsilon_{ij} \prod_{k=i}^{j} S_k - T \sum_{k=1}^{N} \Delta S_{conf} \cdot S_k$, where T is a temperature and N is the number of amino acids in a protein. The conformational state of a protein with N amino acids is represented by the N-spin binary variables $\{S_k\}$ for amino acids k = 1, 2, …., N. When the dihedral angles (ϕ_{k-1}, ψ_k) of kth amino acid are native (non-native)-like, S_k takes the value 1 (0). The entropic cost of forming a native amino acid is $\Delta S_{conf} < 0$ with respect to its non-native conformation. The pairwise-contact energy between ith and jth amino acid is ε_{ij} when the distance between two C^{α}s of ith and jth amino acid is less than a threshold value, for example 6.5 Å, in its native structure. The first term is the sum of pairwise-contact energy ε_{ij} over native contacts $<i, j>$ within the cut-off distance, which is defined to be established when all amino acids from ith to jth amino acid are in the native states; namely when the product $\prod_{k=i}^{j} S_k = S_i S_{i+1} \cdots S_{j-1} S_j = 1$. In reality, ΔS_{conf} depends on the secondary structure of each amino acid, but ΔS_{conf} was chosen as a constant value (–3.8 cal/mol·K) for the sake of simplicity as was in the literature (1–4). Using different values of ΔS_{conf} sets the new value of the rescale factor J, since J was determined such that the folding mid-temperature T_m becomes a reference value, for example 300 K, for a given ε_{ij}. Thus, the value $J/\Delta S_{\text{conf}}$ sets the energy scale and T/T_m sets the temperature scale of a protein under our consideration.

Folding Kinetics by Master Equation. Provided with the exact one-dimensional free energy landscape (Fig. 1) using ME model, each point on there corresponds to a set of protein conformations having the same fraction of native residue. The reversible kinetic hopping between two adjacent points on this free energy landscape effectively describes the relaxational time evolution of the conformational probability vector $\vec{P}(t) = (P_l) = (P_0, P_1, P_2 ... P_N)$ satisfying a master

equation $\frac{d\vec{P}(t)}{dt} = -M\vec{P}(t)$, where $l = 0,1,2,...,N-1,N$ denotes $(N+1)$ set of protein conformations each having the same fraction $M = l/N$ of native residue such that $0(N)$ denotes a fully unfolded (native) state. $M = (M_{mn})$ is a relaxation matrix constructed by the transition probability from a state *n* to *m* based on Metropolis algorithm, where $M_{nn} = -\frac{1}{\tau_0}\exp(-(\Delta G_m - \Delta G_n)/RT)$ if $G_m > G_n$ and $-\frac{1}{\tau_0}$ otherwise, and $m,n = 0,1,2,...,N-1,N$. Here, ΔG_m and ΔG_n are read from Fig. 1, and τ_0 is a molecular time scale. The diagonal elements of *M* are set by $M_{mm} = -\sum_{m,(m\neq n)} M_{mn}$ such that the sum of column matrix elements is zero and M_{mn} satisfies the detailed balance condition. The time evolution of $\vec{P}(t)$ toward its equilibrium one is governed by the eigenvalues of *M*. The eigenvector elements for the zero eigenvalue is the equilibrium population of a protein conformation, and the non-zero smallest eigenvalue λ_1 governs the dominant relaxation behavior with the longest relaxation time and its eigenvector elements dictate the population flow among conformational states. The second non-zero smallest eigenvalue λ_2 and its eigenvector give a correction to the dominant relaxation behavior. A chevron plot as a function of temperature is constructed from λ_1, which becomes the folding (unfolding) rate at the low (high) temperature compared to T_m. A big separation between λ_1, with the temperature dependence of Arrhenius type, and λ_2 signifies the existence of the two-state folding behavior.

Out-of-Equilibrium Monte Carlo Simulation. Starting from an initial conformation of a protein, which can be either a fully stretched or a native conformation, the relaxation of a protein conformation to its equilibrium one for the given temperature is simulated by Monte Carlo simulation using the Metropolis algorithm. Since the protein conformation is represented by the *N*-spin variables $\{S_k\}$ for amino acids $k = 1, 2, \ldots, N$ within the context of ME model, the conformational move of a protein in three dimensional space is described by a single-spin flip kinetics using the Metropolis algorithm in the phase space of 2^N conformations. Given the effective free energy of a protein conformation from ME model, the new protein conformation *p*, after flipping a randomly chosen spin in the old conformation *q*, is generated with the probability

$\exp(-(G_p - G_q)/RT)$ if $G_p > G_q$ and 1 otherwise with the detailed balance condition satisfied. The dynamic trajectory starting from an initial protein conformation and ending at the equilibrated one for a given temperature can be constructed by the successive generation of the new accepted conformation from the out-of-equilibrium Monte Carlo simulation. The time-dependent properties of a protein is averaged over 1,000 different dynamic trajectories in our simulation. We confirmed that the dynamic-free energy-like quantity $F(M, t)$ calculated after the equilibration time for all temperatures converges well to the exact free energy landscape of ME model, implying that our Monte Carlo simulation is indeed reliable.

1. Muñoz V, Eaton WA (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 96:11311–11316.
2. Henry ER, Eaton WA (2004) Combinatorial modeling of protein folding kinetics: Free energy profiles and rates. *Chem Phys* 307:163–185.
3. Bruscolini P, Pelizzola A (2002) Exact solution of the Muñoz-Eaton model for protein folding. *Phys Rev Lett* 88:258101.
4. Chang I, Cieplak M, Banavar JR, Maritan A (2004) What can one learn from experiments about the elusive transition state? *Protein Sci* 13:2446–2457.