Supporting Information

Okazaki and Takada 10.1073/pnas.0802524105

SI Text

Multiple-Basin Model (1). We have previously developed a structure-based multiple-basin energy landscape model that is constructed from multiple fiducial structures. For the case of two basins, we started with two Go-type energy functions $V(R|R_i)$ with $i = 1$, and 2, where R_i is the fiducial structure of basin i. Using these, we constructed a smooth double-basin energy function for a protein, V_{protein} as the eigenvalue of the characteristic equation,

$$\begin{pmatrix} V(R|R_1) & \Delta \\ \Delta & V(R|R_2) + \Delta V \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = V_{\text{protein}} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \qquad [1]$$

where the coupling Δ controls the barrier height for the transition, ΔV is the parameter for modulating the relative stability of the two basins, and (c_1, c_2) is the eigenvector that characterizes the state. The condition that a nontrivial solution exists leads to the secular equation,

$$\begin{vmatrix} V(R|R_1) - V_{\text{protein}} & \Delta \\ \Delta & V(R|R_2) + \Delta V - V_{\text{protein}} \end{vmatrix} = 0. \qquad [2]$$

We use the lower-energy solution as the multiple-basin potential. Eigenvector (c_1, c_2) indicates whether the system resides in basin 1 or 2, thus $\chi = \ln(c_2/c_1)$ can be a useful reaction coordinate for the transition.

The single-basin energy function $V(R|R_\nu)$ was constructed based on the Clementi–Nymeyer–Onuchic model (2). The local interactions in it are

$$\begin{aligned} V_{\text{local}}(R|R_\nu) = & \sum_{\text{bonds}} K_b(b_i - b_{\nu,j})^2 + \sum_{\text{angles}} K_\theta(\theta_i - \theta_{\nu i})^2 \\ & + \sum_{\text{dihedra/}} \{K_\phi^{(1)}[1 - \cos(\phi_i - \phi_{\nu i})] \\ & + K_\phi^{(3)}[1 - \cos 3(\phi_i - \phi_{\nu i})]\} \end{aligned} \qquad [3]$$

where b_i is the bond length between ith and $(i+1)$th Cα atoms, θ_i is the ith bond angle between ith and $(i+1)$th bonds and ϕ_i is the dihedral angle around the $(i+1)$th bond. Parameters with the subscript "ν" take the values of the corresponding variables at the fiducial structure ν. $K_b = 100.0$, $K_\theta = 20.0$, $K_\phi^{(1)} = 1.0$, and $K_\phi^{(3)} = 0.5$. When a protein's landscape possesses two basins, 1 and 2, we define the local strain energies as the local interaction energies at R_1 for the single-basin model defined with the reference structure R_2. For example, the local strain energy for θ_i is $K_\theta(\theta_{1i} - \theta_{2i})^2$, which monitors the degree of structural change in this portion. Often, a large-amplitude conformational change between states 1 and 2 induces a too-strong strain in limited and specific portions of the protein, such as the hinge region. Physically, the strain is relieved by local unfolding. Taking this into consideration, we have reduced the coefficient K_θ of this portion if the strain energy is larger than the cutoff value $\varepsilon_\theta = 1.0$. Namely, explicitly, K_θ in Eq. **3** is replaced with $K_{\theta i} = \min[K_\theta, \varepsilon_\theta/(\theta_{1i} - \theta_{2i})^2]$. In the same way, for the potential on ϕ, we have reduced the values of $K_{\phi i}^{(1)} = 2K_{\phi i}^{(3)}$ when the strain energy in the ith dihedral angle exceeds the cutoff value $\varepsilon_\varphi = 0.5$.

The nonlocal interactions in the single-basin model have specific attractive and repulsive interactions for the amino acid pairs that make contacts in the fiducial structure and generic repulsive interactions for the rest of the pairs. Here, an ij pair was considered to be in "contact" when at least one nonhydrogen atom of the ith amino acid is within 6.5 Å of any nonhydrogen atom of the jth amino acid. For the multiple-basin model, the sets of fiducial contact pairs of the reference structures are not equivalent. We classified the amino acid pairs into three types: (*i*) those pairs that make contact in all reference structures, (*ii*) pairs that make contact in some, but not all, of the reference structures, and (*iii*) pairs that do not make contact in any reference structures.

For the type 1 and 2 pairs, we changed the energy function so that the repulsive interactions are identical for the two basins; of the two critical distances, the smaller one is used. Conversely, the attractive part, if any, depends on the basin. To do this, we divided the nonlocal interactions into $V_{\text{native-attr}}$ and V_{repul}. The former is the attraction interaction between the native contact pairs, and is given by,

$$\begin{aligned} V_{\text{native-attr}}(R|R_\nu) = \varepsilon_1 \sum_{i>j-3}^{\text{native-contact}} \min\Bigg\{ 1,\ 5\left(\frac{r_{\nu,ij}}{r_{ij}}\right)^{12} \\ - 6\left(\frac{r_{\nu,ij}}{r_{ij}}\right)^{10} + 1 \Bigg\}. \end{aligned}$$

Here, $\varepsilon_1 = 0.18$ was used. Note that, for convenience, we shifted the zero energy at the bottom of the curve. For the type 3 pairs, we can just use the same functional form as the original one because the repulsive force has a generic form. The repulsive part for all types can be written as

$$\begin{aligned} V_{\text{repul}}(R|R_\nu) = & \varepsilon_1 \sum_{i>j-3}^{\text{Type1\&2}} \max\left\{ 0,\ 5\left(\frac{r_{0ij}^{\min}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}^{\min}}{r_{ij}}\right)^{10} \right\} \\ & + \varepsilon_2 \sum_{i>j-3}^{\text{Type 3}} \left(\frac{C}{r_{ij}}\right)^{12}, \end{aligned}$$

where $r_{0ij}^{\min} = \min_{\{\nu=A,B,..\}} \{r_{\nu,ij}\}$, which is the smallest characteristic distance and $\varepsilon_2 = 0.18$, $C = 4.0$ Å. Combining these terms, the total potential of the single-basin model is $V(R|R_\nu) = V_{\text{local}} + V_{\text{native-attr}} + V_{\text{repul}}$. We note that, with this definition, the single-basin potential is always positive and its value at the bottom of the basin $V(R_\nu|R_\nu)$ vanishes.

Molecular Dynamics. MD simulation was carried out by using the constant-temperature Newtonian dynamics, where the mass of all residues was set to identical. The velocity Verlet algorithm was used for time propagation with a simple Berendsen thermostat. We could equivalently use the Langevin dynamics. The choice of the Newtonian dynamics with Berendsen thermostat is not essential for the current results.

Definition of States and Transition Rate Analysis. First, we defined four states. Ligand-binding state is either unbound (U) or bound (B). The conformation is classified as open (O) if $\chi \leq 0.0$ and as closed (C) if $\chi > 0.0$. So, we have the four states: UO, UC, BO, and BC.

For the transition rates, under the condition that the U and B are equally probable, we performed long-time simulations. In the trajectory, we observed reversible and multiple transitions

among the four states, and counted the number of transitions between neighboring states n_{ij}. Here, transitions in both directions should be equal, in principle, and so we take an average of numbers in two-way transitions. To avoid overcounting the recrossing, we defined the completion of one transition when χ changed its sign *and* reached its absolute value larger than unity. Under the Markovian approximation, using $k_{i\to j}\tau_i = k_{j\to i}\tau_j = n_{ij}$, we obtained the transition rates between states, where $k_{i\to j}$ ($k_{j\to i}$) is the rate constant from state i (j) to state j (i) and $\tau_i(\tau_j)$ is the residential time of state i (j) along the trajectory.

Steady-State Analysis. Within the four-state representation, we performed a simple steady-state analysis. For example, when we calculate the rate of induced-fit pathway, that is, from UO via BO to BC, the reaction scheme is UO⇄BO→BC. For the steady state, we assumed that the molecular source is at UO state and the molecular sink is at BC. Thus, the reverse transition from BC to BO is zero. We further assume that the concentration of the intermediate BO is in steady state, $\frac{d[BO]}{dt} = k_1[UO] - (k_{-1} + k_2)[BO] = 0$ where k_1 is apparent first-order rate from UO to BO at the fixed ligand concentration, k_{-1} is the dissociation rate from BO to UO, and k_2 is the rate from BO to BC. From this, we directly get the rate of induced-fit k_{if}, $\frac{d[BC]}{dt} = \frac{k_1 k_2}{k_{-1} + k_2}[UO] = k_{if}[UO]$. In the case of short-ranged interaction, this gave us $k_{if} = 0.054$. The same sort of analysis about the population-shift pathway leads to the rate of population-shift k_{ps} as $k_{ps} = 0.11$. In the case of long-ranged interaction, $k_{if} = 0.37$ and $k_{ps} = 0.12$.

1. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci USA* 103:11844–11849.

2. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and ''en-route'' intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298:937–953.