

Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations

Kei-ichi Okazaki*, Nobuyasu Koga*, Shoji Takada*[†], Jose N. Onuchic[‡], and Peter G. Wolynes*^{§¶}

*Graduate School of Natural Science and Technology, Kobe University, Kobe 657-8501, Japan; [†]Core Research for Evolutionary Science and Technology, Japan Science and Technology Corp., Kobe 657-8501, Japan; and [‡]Center for Theoretical Biological Physics and [§]Department of Chemistry and Biochemistry and Department of Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093

Contributed by Peter G. Wolynes, May 26, 2006

Biomolecules often undergo large-amplitude motions when they bind or release other molecules. Unlike macroscopic machines, these biomolecular machines can partially disassemble (unfold) and then reassemble (fold) during such transitions. Here we put forward a minimal structure-based model, the “multiple-basin model,” that can directly be used for molecular dynamics simulation of even very large biomolecular systems so long as the endpoints of the conformational change are known. We investigate the model by simulating large-scale motions of four proteins: glutamine-binding protein, S100A6, dihydrofolate reductase, and HIV-1 protease. The mechanisms of conformational transition depend on the protein basin topologies and change with temperature near the folding transition. The conformational transition rate varies linearly with driving force over a fairly large range. This linearity appears to be a consequence of partial unfolding during the conformational transition.

conformational transition | cracking | partial unfolding | funnel

To function, biomolecules often undergo large-amplitude structural changes upon binding or releasing ligands. These structural changes organize the workings of biomolecular machines such as the ribosome, molecular chaperones, and molecular motors. Structural information on the conformational ensembles before and after the conformation changes is often available through x-ray crystallography or NMR. These experiments, however, provide primarily quasistatic information. They reveal directly less about the transition dynamics between two end structures. The overall dynamics of basin-hopping can be studied by pump-probe experiments or from NMR relaxation. These experiments, however, usually monitor directly only a few local structure changes in what is typically a huge system. Thus, we see that global time-dependent structural information at high resolution is rarely obtained directly by experiments. Simulations can potentially provide full time-dependent structural information on biomolecular machines. Yet conventional atomistic simulations currently only reach times up to microseconds (1). This time scale falls orders of magnitude short of the typical physiologically important time scales of milliseconds to seconds. To overcome this limitation, one approach is to coarse-grain the molecular representation (2). Reduction in complexity allows one to simulate much longer times. This so-called “minimalist approach” has been quite successful for studying protein folding (3–8). The purpose of this article is to investigate a minimal structure-based model, which we call the “multiple-basin model,” to simulate large-scale conformational changes when structures for the endpoints of the transition are available. This approach can be used for simulations of even very large biomolecular complexes.

To motivate the present model, we first note that two qualitatively different kinds of protein motions occur depending on the amplitude of motion. For small deviations from the fiducial native structure motions are well approximated by quasi-harmonic dynamics. Indeed, very simple quasi-harmonic topology-based models reproduce the size of fluctuations at single-residue resolution as

evidenced by experimental B-factors (9, 10). Structure changes upon ligand binding usually occur predominantly in directions that correspond to combinations of a few low-frequency modes (11–16). Ikeguchi *et al.* (17) have shown using their linear response theory that the direction of structural change can be predicted from the dominant principal components of the structural fluctuations in the unbound states multiplied vectorially by the force exerted by the ligand. Conformational transitions, however, must involve rearrangement of nonlocal contacts of amino acid pairs (see Fig. 1 *Lower*). Such notions clearly require going beyond the quasi-harmonic picture (18–21): The protein breaks some contacts specific to the initial conformation and forms new contacts that are specific to the final conformation. This second, large-amplitude regime of protein dynamics has been termed a “proteinquake” (22) and may involve “cracking” (18, 21) or local unfolding. As the protein softens and its fluctuations increase with temperature, a third regime where the protein may globally unfold begins to play a further role in function especially for the so-called “natively unfolded proteins” (23, 24).

We see that a minimal requirement for describing conformational transitions is to be able to model both the quasi-harmonic fluctuations around each basin and the transient and partial unfolding near the transition region. A natural framework to formalize such a model is provided by energy landscape theory. Energy landscape theory has established that proteins have evolved to have funnel-like energy landscapes (25–27). The bottom of the funnel corresponds to a fiducial native structure having the lowest effective solvent averaged free energy (as an individual conformation). As the protein unfolds, the number of structures multiplies while the effective energy increases. An ideal funnel-like energy landscape can be realized by the so-called Gō model (6, 7, 28–30), first introduced for lattice models (where of course there is no harmonic motion) in which a protein is represented as a chain with attractive interactions between pairs of residues that interact in the native structure and repulsive interactions for all other residue pairs. These models have been very effective for studying folding mechanisms. Moreover, the off-lattice version of perfect funnel models (6, 7) already goes far to meet our present needs because it models quite well the quasi-harmonic motion in the limit of the weak fluctuations (31).

In the standard energy landscape for folding proteins, however, only a single dominant minimum, which corresponds to the native structure, is assumed. Studying conformational changes of functional proteins requires more than one basin to be taken into account. Each basin should correspond to the structure with or without ligands. The standard structure-based models are not

Conflict of interest statement: No conflicts declared.

Abbreviations: GBP, glutamine-binding protein; DHFR, dihydrofolate reductase; MD, molecular dynamics; PDB, Protein Data Bank.

[¶]To whom correspondence should be addressed. E-mail: pwolynes@ucsd.edu.

© 2006 by The National Academy of Sciences of the USA

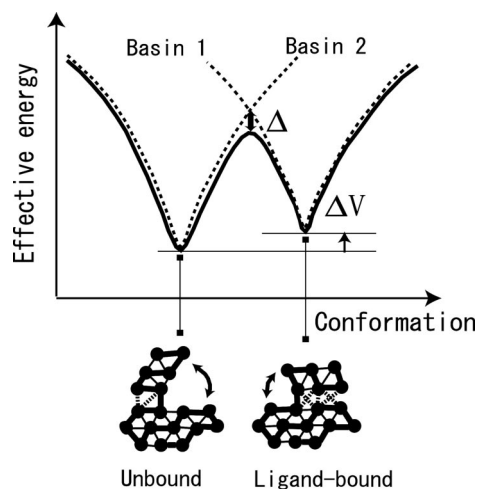


Fig. 1. Schematic view of multiple-basin energy landscape of proteins. Two single funnels used for model construction are depicted by dashed lines. Conformational change is associated with the rearrangement of some contacts. Contacts specific to the unbound conformation are broken, and new contacts are formed in bound conformation. Thick solid bonds correspond to covalent linkages.

directly applicable to this case of multiple basins. Here we explicitly build up an energy landscape encoding multiple near-degenerate basins. Given two reference structures supplied by x-ray crystallography, we first created two independent structure-based potentials (32), which were then smoothly connected to make a double-well energy landscape. Very recently, a closely related approach having the same spirit was put forward by Best *et al.* (20). We will survey several systems with the model and highlight the cracking phenomenon.

Specifically, we simulated conformational transitions for four proteins, glutamine-binding protein (GBP), S100A6 (which is a structural analog of calmodulin), dihydrofolate reductase, and HIV-1 protease. The global character of the conformational motions, the residue-specific involvement in transition pathways, and the temperature dependence of rates were then investigated and related to changes in protein topology. We also show that the linear dependence of rate on driving force arises as a consequence of local unfolding near the transition region.

Results

Multiple-Basin Model. We explicitly describe the model for the case with two major structurally characterized basins, basin 1 and 2 as in Fig. 1. (For a generalization to n basins see *Materials and Methods*.) The model's energy landscape within each individual basin is a perfect funnel. We start with two virtual funnels (dashed curves in Fig. 1). The target energy landscape (solid curve) coincides with one of the dashed curve near basin 1 and merges with the other dashed curve near basin 2. Around the transition region, the energy landscape should be smoothly connected. From a mathematical viewpoint, the way to connect smoothly two basins is not unique. Here we choose a smooth connection that is computationally convenient.

Mathematically we construct first two single-basin potentials as $V(R | R_1)$ and $V(R | R_2) + \Delta V$ (dashed lines in Fig. 1), where R collectively represents the coordinates of the protein structure and R_1 and R_2 correspond to the coordinates of the fiducial structures. For convenience, the input potential $V(R | R_v)$ is defined so that the energy at the bottom of basin v is (nearly) zero (see *Materials and Methods* for details). Then ΔV is introduced to modulate the relative stability of the two basins to correspond to empirically determined energetics; a larger ΔV makes basin 2 less stable than basin 1. We

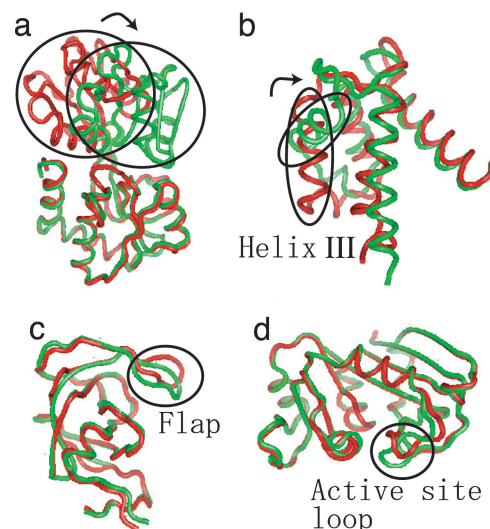


Fig. 2. Conformational change of four proteins studied. (a) GBP. (b) S100A6. (c) HIV-1 protease. (d) DHFR. The red structures are those of unbound states (occluded in the case of DHFR), and the green structures are those of bound states (41).

then introduce a coupling between the two potentials to make a smoothed double basin potential V_{MB} . We use (merely for construction) an analogy with the quantum mechanics of electron transfer to define such a smooth potential, V_{MB} , as the eigenvalue of the characteristic equation:

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

where Δ is a coupling constant and (c_1, c_2) is the eigenvector. The condition that a nontrivial solution exists leads to the secular equation:

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

We use the lower-energy solution as the multiple-basin potential:

$$V_{MB} = \frac{V(R|R_1) + V(R|R_2) + \Delta V}{2} - \sqrt{\left(\frac{V(R|R_1) - V(R|R_2) - \Delta V}{2} \right)^2 + \Delta^2}. \quad [3]$$

V_{MB} being continuous and differentiable can directly be used for molecular dynamics (MD) simulations. The corresponding eigenvector (c_1, c_2) indicates whether the system resides in basin 1 or 2; thus, we also can use $\chi = \ln(c_2/c_1)$ as a reaction coordinate for the transition. This present interpolation is particularly useful because it allows us to freely tune one of the two fundamental energy scales: The coupling constant Δ modifies directly the energy barrier, and ΔV modulates the relative stability.

Simulation of Conformational Changes. We first illustrate the results using GBP. GBP is composed of two domains connected by a hinge (see Fig. 2a). Without substrates, GBP is found in an open form (red in Fig. 2). Upon binding to glutamine, the hinge swings to make the closed form (green). Using the open form for basin 1 and basin 2 as the closed form as fiducial structures, we tuned ΔV so that the protein spends equal time in each basin. At this tuning, ligand/protein concentration coincides with the dissociation constant. The simulation temperature T was set to 0.8 times the folding transition

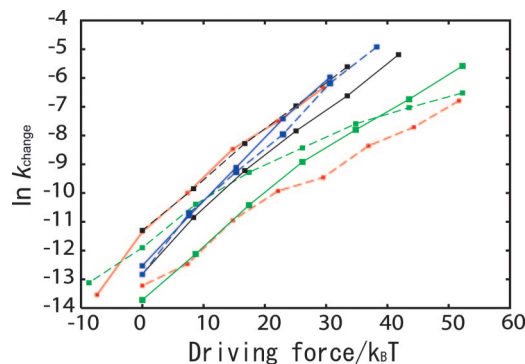


Fig. 6. The transition rate constant as a function of the driving force of the conformational transition (Tafel plot). For each of four proteins, both transitions from unbound to bound (dashed lines) and those from bound to unbound (solid lines), were investigated. The vertical axis is the logarithm of the transition rate coefficient, and the horizontal axis is the driving force. The driving force is $\pm \Delta V$, where the sign is chosen so that the increase in the driving force corresponds to stabilize the final state. GBP is in red, S100A6 is in green, DHFR is in black, and HIV-1 protease is in blue.

specific choice of the reaction coordinate (χ in Fig. 3d) reflecting recrossing effects (33). By using this estimate the dependence of the transition rate on the driving force can be determined. In the electron-transfer processes, a quadratic dependence of the barrier on stability is predicted (34). This finding has been experimentally proven to be fairly accurate (35). For the conformational changes of proteins, Miyashita *et al.* (18) also argued that a fully elastic model would also give such a curved dependence. On the other hand, they suggested that local unfolding, or cracking, will lead to a linear dependence over a large range of the driving force. We now examine this notion using multibasin model MD simulations.

The conformational transition rate coefficient k_{change} was estimated as the inverse of the first passage time as in ref. 36. The transition is considered complete when χ first reaches the value at the minimum of the final basin. For very large driving forces, the conformational transition becomes barrier-less and limited only by diffusion, as expected. In this regime the rate becomes saturated. Here we limit ourselves to thermodynamic conditions having significant barriers: The trajectories reside in the initial basin at least 100 MD steps (on average) before making the first transition.

For the four proteins studied, we calculated both the rates going from open to closed and in the other direction over as large a range of the driving force without reaching the barrier-less regime. The resulting Tafel plots (Fig. 6) are surprisingly linear (with the exception of the binding reaction for S100A6, which is curved). The linear dependence is the consequence of the local unfolding as described by Miyashita *et al.* (18, 21).

Discussion

Building on the elastic picture of Miyashita and collaborators, the multiple-basin models somewhat similar to the present one have recently been proposed. Maragakis and Karplus (19) put forward a plastic network model in which individual basins are approximated by the Tirion harmonic model and are then smoothly connected by the secular equation formulation, as we do in Eq. 3. Being locally purely harmonic local unfolding is not taken into account in this model. This plastic network model should work best for small-amplitude conformational changes. The model developed by Best *et al.* (20) is close in spirit to the present multibasin Hamiltonian because it also connects single-basin potentials in a smooth function. Their interpolation was achieved by the analogy to Boltzmann averaging instead of the secular equation formula. Their model should give similar results to ours when there are two basins. When there are more than two basins, the present model allows the

PNAS | August 8, 2006 | vol. 103 | no. 32 | 11847

possibility to modulate the barrier heights of each basin-hopping motion individually and thus is somewhat flexible.

Directly modeling the multiple-basin energy landscapes is at an early stage. There is considerable room for creativity and improvement at this low-resolution scale. An ingredient that is missing in the present model as well as in the others (18–21) is precisely how to correctly account for the interaction with ligands. In the current model the effects of ligand are implicit in the value of ΔV , which modulates the overall stability. But such a term does not account for the local nature of the interactions with the ligand. For example, S100A6 binds to calcium ions at the EF-hand loops, and the interactions with calcium ions make the EF-hand loop more rigid. This locality may play an important role in allostery. Further work in this direction is necessary.

The present multiple-basin model can directly be used for MD simulation of very large biomolecular systems, such as molecular motors. Recently, the molecular mechanisms of the rotary motor F₁-ATPase were studied by using the switching Gō model (37). In that work, the change in the nucleotide state was modeled as a “vertical excitation,” resulting in switching between single-basin models. The multiple-basin model proposed here provides a natural framework for realizing thermally activated conformational motions coupling ligand binding and release.

Materials and Methods

Proteins Studied. We studied four proteins: GBP, S100A6, DHFR, and HIV-1 protease. For each, two Protein Data Bank (PDB) structures were used to construct the multiple-basin model. GBP is composed of two domains. Without substrates, GBP is found in an open form (PDB ID code 1GGG; red in Fig. 2a), and upon glutamine binding the hinge between domains swings, forming the closed structure (PDB ID code 1WDN; green). S100A6 is a structural analog of calmodulin. Its conformational change, a shear motion (38), occurs between the apo state (PDB ID code 1K9P; red in Fig. 2b) and the calcium-bound holo state (PDB ID code 1K9K; green). DHFR changes its active-site loops via a shear motion (38) between the occluded state (PDB ID code 1RX6; red in Fig. 2c) and the closed state (PDB ID code 1RX2; green). HIV-1 protease has a β -hairpin loop (the flap) that adopts an open conformation (PDB ID code 3HVP; red in Fig. 2d) without an inhibitor but acquires a closed conformation (PDB ID code 4HVP; green) with an inhibitor. The conformational change is relatively small and is of the hinge type (38).

Multiple-Basin Model: Mathematical Expressions. The multiple-basin model V_{MB} is defined in Eq. 1 in terms of the single-basin model potentials $V(R | R_\nu)$ where R_ν stands for the reference structure, the structure at the bottom of the basin ν . For the single-basin model, we used Clementi *et al.*'s version of the off-lattice Gō model (6, 7). In this model, each amino acid is represented as a bead located at the C α position, two consecutive amino acids are connected by harmonic springs, and local and nonlocal interactions are designed to bias the surface toward the reference structure R_ν . (In case of folding, the reference structure is the native.)

The local interactions in the original version of Clementi *et al.* (6, 7) are:

$$V_{\text{local}}(R|R_\nu) = \sum_{\text{bonds}} K_b(b_i - b_{\nu,i})^2 + \sum_{\text{angles}} K_\theta(\theta_i - \theta_{\nu,i})^2 + \sum_{\text{dihedral}} \{K_\phi^{(1)}[1 - \cos(\phi_i - \phi_{\nu,i})] + K_\phi^{(3)}[1 - \cos 3(\phi_i - \phi_{\nu,i})]\}, \quad [4]$$

where b_i is the bond length between i th and $(i + 1)$ th C α atoms, θ_i is the i th bond angle between i th and $(i + 1)$ th bonds, and ϕ_i

is the dihedral angle around the $(i + 1)$ th bond. Parameters with the subscript ν indicate the values of the corresponding variables at the reference structure R_ν . Constants K are independent of residue number i , $K_b = 100.0$, $K_\theta = 20.0$, $K_\phi^{(1)} = 1.0$, 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 strong strain in limited and specific portions of the protein, such as the hinge region. Physically, the strain is relieved by breaking fragile local interactions involving this portion. Taking this into consideration, we have reduced the coefficient K_θ of this portion when the strain energy is larger than the cutoff value $\varepsilon_\theta = 1.0$. Explicitly, K_θ in Eq. 4 is replaced with the site-specific constant $K_{\theta i}$, which is defined as $K_{\theta i} = \min[K_\theta, \varepsilon_\theta/(\theta_{1i} - \theta_{2i})^2]$. In the same way, for the potential on ϕ , we have reduced the values of $K_\phi^{(1)} = 2K_\phi^{(3)}$ when the strain energy in the i th dihedral angle exceeds the cutoff value $\varepsilon_\phi = 0.5$.

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 is considered to be in “contact” when at least one nonhydrogen atom of the i th amino acid is within 6.5 Å of any nonhydrogen atom of the j th amino acid. For the multiple-basin model, the sets of fiducial contact pairs of the reference structures are not equivalent. We need to classify the amino acid pairs into three types: (i) those pairs that make contact in all reference structures, (ii) pairs that make contact in some of, 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 modify the functional form for the following reasons. In general, the repulsive part of the pair interaction is quite sharp, and thus the pair energy increases very rapidly when they come closer than the critical distance. For example, in the two-basin case, for type 1 and 2 pairs the critical distances of two reference structures are not the same. This difference induces a large energy gap between the two single-basin potential values, leading to a transition barrier from one basin to another that is unphysically large. To avoid this, 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.

Mathematically, we can express this change in the energy for type 1 and 2 pairs as follows. We divide the nonlocal interactions into $V_{\text{native-attr}}$ and V_{repl} . The former is the attraction interaction between the native contact pairs and is given by:

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

Here $\varepsilon_1 = 0.18$ was used. We note that, for convenience, we shifted the zero energy at the bottom of the curve instead of the energy at infinite distance, as in the original Clementi *et al.* version (6, 7). 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:

$$V_{\text{repl}}(R|R_\nu) = \varepsilon_1 \sum_{i>j-3}^{\text{Type1\&2}} \max \left\{ 0, 5 \left(\frac{r_{0ij}^{\min}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{0ij}^{\min}}{r_{ij}} \right)^{10} \right\} \\ + \varepsilon_2 \sum_{i>j-3}^{\text{Type3}} \left(\frac{C}{r_{ij}} \right)^{12}, \quad [6]$$

where

$$r_{0ij}^{\min} = \min_{\{v=1,2,\dots\}} \{r_{v,ij}\}, \quad [7]$$

which is the smallest characteristic distance, $\varepsilon_2 = 0.18$, and $C = 4.0 \text{ \AA}$.

Combining these terms, the total potential of the single-basin model that constitutes the input to the multiple-basin model is $V(R | R_\nu) = V_{\text{local}} + V_{\text{native-attn}} + V_{\text{repul}}$. 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.

Although the multiple-basin model was explicitly described for the two-basin case in this article, its generalization to the n -basin case is straightforward. The multiple-basin model potential energy V_{MB} is defined as the solution of the $n \times n$ secular equation:

$$D \equiv \begin{vmatrix} V(R|R_1) - V_{\text{MB}} & \Delta_{12} & & & \\ \Delta_{12} & V(R|R_2) + \Delta V_2 - V_{\text{MB}} & & & \\ \vdots & & \Delta_{23} & & \\ \vdots & & \vdots & & \\ \Delta_{1n} & & \Delta_{2n} & & \\ & \ddots & & \Delta_{1n} & \\ & \ddots & & \Delta_{2n} & \\ & \vdots & & \vdots & \\ & & \Delta_{n-1,n} & & \\ & & V(R|R_n) + \Delta V_n - V_{\text{MB}} & & \end{vmatrix} = 0. \quad [8]$$

Practically, unless more than three $V(R|R_v)$ are near degenerate, we can solve 2×2 equations with two smallest V values. In case of the degeneracy, calculating the force on each residue needs a slightly involved procedure. From $\partial D/\partial \mathbf{r}_i = 0$, we can calculate the force $\mathbf{F}_i$ acting on i th residue as:

$$\mathbf{F}_i = \sum_{v=1}^n \tilde{d}_{vv} \mathbf{F}_i^{(v)} / \sum_{v=1}^n \tilde{d}_{vv}, \quad [9]$$

where $\tilde{d}_{\nu\mu}$ is the minor determinant and $\mathbf{F}_i^{(\nu)} = -\partial V(R \mid R_\nu)/\partial \mathbf{r}_i$.

Simulation Protocol. In this work, MD simulation was carried out by using the constant-temperature Newtonian dynamics. The velocity Verlet algorithm (39) was used for time propagation, and the temperature was controlled by the simple Berendsen thermostat (40). The mass for all amino acids was set to be the same.

The simulation temperatures were determined by using a crude estimate of the folding transition temperature T_F for each reference structure of studied proteins. For S100A6 and DHFR, T_F was calculated by the protocol of ref. 36. After this, we assumed that T_F scales with the number of contacts per residue. Thus, using the data on S100A6 and DHFR, we estimated the T_F values of other proteins. Default simulation temperatures of conformational change were set as $0.8 T_F^{(\min)}$, where $T_F^{(\min)}$ is the smaller of the two T_F values associated with two reference structures. (If $T_F \sim 80^\circ\text{C}$, $0.8 T_F$ becomes 9°C) Preliminary tests indicated that at higher temperatures proteins would globally unfold, whereas at lower temperatures the conformational transitions were too slow to sample in reasonable simulation times.

It was necessary to fix the values of the two parameters introduced for the multiple-basin model, the coupling term Δ and the relative stability ΔV . The former controls the energy barrier between two states; the larger is Δ , and the smaller is the barrier. ΔV modulates the relative stability of two states. Both parameters can be determined by using experimental input. For convenience, we adjusted these parameters so that reversible transitions between two conformations were realized for each of the proteins studied. Starting with a fairly small value, Δ is gradually increased until the transitions take place within acceptable computation times. Second, ΔV was tuned so that transitions from each conformation take place with equal frequency. The resulting value is ΔV_{eq} . The parameters so obtained are as follows (in units of $k_B T$): $\Delta = 59$ and $\Delta V_{\text{eq}} = -4.4$ for GBP, $\Delta = 66$ and $\Delta V_{\text{eq}} = 11$ for S100A6, $\Delta = 6$ and $\Delta V_{\text{eq}} = -2$ for DHFR, and $\Delta = 20$ and $\Delta V_{\text{eq}} = 2$ for HIV-1 protease.

N.K. is supported by a Japan Society for the Promotion of Science Research Fellowship for Young Scientists. This work was partly supported by the Water and Biomolecules program of the Ministry of Education, Culture, Sports, Science, and Technology, Japan. Work in University of California at San Diego was supported by National Science Foundation Grants PHY0216576 and PHY0225630.

1. Daggett, V. (2003) *Protein Simulations* (Elsevier Academic, San Diego).
2. Tozzini, V. (2005) *Curr. Opin. Struct. Biol.* **15**, 144–150.
3. Ueda, Y., Taketomi, H. & Go, N. (1975) *Int. J. Pept. Res.* **7**, 445–459.
4. Socci, N. D., Onuchic, J. N. & Wolynes, P. G. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2031–2035.
5. Takada, S., Luthy-Schulten, Z. & Wolynes, P. (1999) *J. Chem. Phys.* **110**, 11616–11629.
6. Clementi, C., Jennings, P. A. & Onuchic, J. N. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 5871–5876.
7. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000) *J. Mol. Biol.* **298**, 937–953.
8. Honycutt, J. D. & Thirumalai, D. (1992) *Biopolymers* **32**, 695–709.
9. Tirion, M. M. (1996) *Phys. Rev. Lett.* **77**, 1955–1908.
10. Bahar, I. & Jernigan, R. L. (1998) *J. Mol. Biol.* **281**, 871–884.
11. McCammon, J. A., Gelin, B. R., Karplus, M. & Wolynes, P. G. (1976) *Nature* **262**, 325–326.
12. Cui, Q., Li, G., Ma, J. & Karplus, M. (2004) *J. Mol. Biol.* **340**, 345–372.
13. Krebs, W. G., Alexandrov, V., Wilson, C. A., Echols, N., Yu, H. & Gerstein, M. (2002) *Proteins* **48**, 682–695.
14. Hayward, S., Kitao, A. & Berendsen, H. J. (1997) *Proteins* **27**, 425–437.
15. Tama, F. & Sanejouand, Y. H. (2001) *Protein Eng.* **14**, 1–6.
16. Tama, F., Valle, M., Frank, J. & Brooks, C. L., III (2003) *Proc. Natl. Acad. Sci. USA* **100**, 9319–9323.
17. Ikeguchi, M., Ueno, J., Sato, M. & Kidera, A. (2005) *Phys. Rev. Lett.* **94**, 078102.
18. Miyashita, O., Onuchic, J. N. & Wolynes, P. G. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 12570–12575.
19. Maragakis, P. & Karplus, M. (2005) *J. Mol. Biol.* **352**, 807–822.
20. Best, R. B., Chen, Y. G. & Hummer, G. (2005) *Structure (London)* **13**, 1755–1763.
21. Miyashita, O., Wolynes, P. G. & Onichic J. N. (2005) *J. Phys. Chem.* **109**, 1959–1969.
22. Ansari, A., Berendzen, J., Bowne, S. F., Frauenfelder, H., Iben, I. E., Sauke, T. B., Shyamsunder, E. & Young, R. D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5000–5004.
23. Bracken, C., Iakoucheva, L. M., Romero, P. R. & Dunker, A. K. (2004) *Curr. Opin. Struct. Biol.* **14**, 570–576.
24. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 8868–8873.
25. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins* **21**, 167–195.
26. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7524–7528.
27. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8721–8725.
28. Go, N. (1983) *Annu. Rev. Biophys. Bioeng.* **12**, 183–210.
29. Shoemaker, B. A., Wang, J. & Wolynes, P. G. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 777–782.
30. Takada, S. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 11698–11700.
31. Takano, M., Higo, J., Nakamura, H. K. & Sasai, M. (2004) *Natural Comput.* **3**, 377–393.
32. Zuckerman, D. M. (2004) *J. Phys. Chem. B* **108**, 5127–5137.
33. Frauenfelder, H. & Wolynes, P. G. (1985) *Science* **229**, 337–345.
34. Marcus, R. A. & Sutin, N. (1985) *Biochim. Biophys. Acta* **811**, 265–332.
35. Miller, L. T., Calcatera, L. T. & Gloss, G. L. (1984) *J. Am. Chem. Soc.* **106**, 3047.
36. Koga, N. & Takada, S. (2001) *J. Mol. Biol.* **313**, 171–180.
37. Koga, N. & Takada, S. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 5367–5372.
38. Gerstein, M. & Krebs, W. (1998) *Nucleic Acids Res.* **26**, 4280–4290.
39. Frenkel, D. & Smit, B. (1996) *Understanding Molecular Simulations: From Algorithms to Applications* (Academic, San Diego).
40. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984) *J. Chem. Phys.* **81**, 3684–3690.
41. DeLano, W. L. (2002) *PyMOL User's Manual* (DeLano Scientific, San Carlos, CA).