Supporting Information

Best et al. 10.1073/pnas.0910390107

SI Text

Coarse-Grained Folding Models. In the original version of the model, non-bonded interactions are described by a pair potential in which native contacts (those amino acid pairs in contact in the native structure) are treated using a 12-10-6 potential:

$$V_{\mathrm{G\bar{o}}}(r_{ij}) = \epsilon_{ij}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad [1]$$

Non-native contacts were treated as repulsive. Simulations were run using Langevin dynamics at the folding temperature of the respective Gō model (292 K for prb_{7-53} and 310 K for protein G), with a friction coefficient of 0.1 ps^{-1}.

A number of variants of this model are considered: variant 1 is the "non-Gō" potential described in the main text while variants 2 and 3 are shown in Fig. S4.

1. A recently described model for protein–protein interactions (1) was used to add non-Gō energy, i.e, interactions specified by the identity of the residues rather than by the native structure. The non-bonded energy in the presence of non-Gō interactions was given by

$$V_{\mathrm{nb}} = \sum_{\substack{\mathrm{native}\\(i,j)}} V_{\mathrm{G\bar{o}}}(r_{ij}) + \sum_{\substack{\mathrm{non-native}\\(i,j)}} V_{\mathrm{ng}}(r_{ij}) + \sum_{\substack{\mathrm{all}\\(i,j)}} V_{\mathrm{elec}}(r_{ij}) \qquad [2]$$

   Here, the $V_{\mathrm{ng}}(r_{ij})$ and $V_{\mathrm{elec}}(r_{ij})$ are, respectively, the modified Lennard–Jones potential and Debye–Hückel type expression of Kim and Hummer (1) (model "a" in the paper).
2. A Morse potential was used to treat the attractive part of the native contact potential, instead of Eq. S1:

$$V(r_{ij}) = \begin{cases} V_{\mathrm{G\bar{o}}}(r_{ij}) & r_{ij} \le \sigma_{ij} \\ \epsilon_{ij}[(1-\exp(-\alpha(r_{ij}-\sigma_{ij})))^2 - 1] & r_{ij} > \sigma_{ij} \end{cases}. \qquad [3]$$

   The parameter α was set to 1.7 to match the curvature of the 12-10-6 potential at the minimum for a native contact distance of 6 Å.
3. The magnitude of the torsion potentials in the original model was halved.

Calculation of Diffusion Coefficients. We describe the dynamics along a given coordinate as one-dimensional diffusion. The coordinate q is divided into a set of contiguous non-overlapping bins and the dynamics is described in terms of transitions between the bins using a spatially discretized form of the Smoluchowski equation due to Bicout and Szabo (2). The time evolution of the population in bins i follows a system of rate equations:

$$p_i(t) = R_{i,i-1}p_{i-1}(t) - (R_{i-1,i} + R_{i+1,i})p_i(t) + R_{i,i+1}p_{i+1}(t). \qquad [4]$$

From a simulation (or single-molecule experimental) trajectory $q(t)$, the statistics of transitions from bin i to bin j after a lag time Δt, are accumulated in a transition matrix $N(j, \Delta t|i, 0)$. A Bayesian approach is used to find the equilibrium probabilities P_i and rate coefficients $R_{j,i}$ that are most consistent with the observed transition matrix, as described previously (3, 4).

The potential of mean force in interval i, $F(q_i)$ may be obtained from the equilibrium populations P_i in each interval via

$$F(q_i) \approx -k_{\mathrm{B}}T \ln \frac{P_i}{\Delta q} \qquad [5]$$

where Δq is the width of the interval in q. The position-dependent diffusion coefficients $D_{i+1/2}$ can be obtained from the rate coefficients from

$$D_{i+1/2} \approx \Delta q^2 R_{i+1,i}\left(\frac{P_i}{P_{i+1}}\right)^{1/2} \qquad [6]$$

where $D_{i+1/2} = D((q_i + q_{i+1})/2)$ is the diffusion coefficient centered between the bins at q_i and q_{i+1}. Uncertainties were estimated from a block error analysis of the sampled parameters.

What happens if the chosen coordinate q is not a good reaction coordinate? If q does not properly separate the transition states (with splitting/commitment probabilities (5, 6) of 1/2) from reactant and/or product states, the estimated $D(q)$ will depend strongly on the lag time Δt at which transitions are observed. For short lag times, $D(q)$ will be large (because even relatively fast fluctuations appear to carry the system between reactant and product states); for long times, it will be small (because fast fluctuations are averaged out and only the long-time folding/unfolding dynamics dominates, similar to what is seen in a number correlation function). Correspondingly, the rates of folding and unfolding predicted from the diffusion model will be large for small Δt and small for large Δt. Only in the limit of very long Δt are the predicted rates expected to approach the "real" ones. However, trajectories $q(t)$ artificially generated from the diffusion model obtained in the limit $\Delta t \to \infty$ will not resemble the $q(t)$ trajectories obtained from the molecular simulations if q is a poor reaction coordinate. The reason for that is that $D(q)$ will be artificially reduced in the transition region to compensate with a small kinetic prefactor for the low free energy barrier of the poor coordinate q.

Effect of Non-Gō Contacts on Stability. Intuitively, one would expect that non-Gō contacts would only play a role in the unfolded state, most likely stabilizing it. However, our non-Gō potential is fairly long-ranged, including 12-6 Lennard–Jones and electrostatic terms (1). Fig. S3*A* and *B* shows that the overall number of non-native contacts is similar in both states and consequently there is a similar contribution of the non-Gō terms to the energy of both the folded and unfolded states, so that the difference between the two is small. Although the non-Gō energy difference between unfolded and folded states shows that the non-Gō contacts do slightly disfavor the folded state (Table S1), it turns out that the repulsive non-native contacts in the original Gō model disfavor the folded state a little more (Fig. S3*C* and *D* and Table S1). With a difference in non-native energy between folded and unfolded of $\Delta E_{\mathrm{nonG\bar{o}}}$ for the non-Gō model and $\Delta E_{\mathrm{G\bar{o}}}$ for the Gō model, the change in folding energy upon replacing the Gō non-native interactions with non-Gō interactions, $\Delta\Delta E = \Delta E_{\mathrm{nonG\bar{o}}} - \Delta E_{\mathrm{G\bar{o}}}$, is −0.13 kcal/mol ($-0.22\ k_{\mathrm{B}}T$) for prb and −2.24 kcal/mol ($-3.64\ k_{\mathrm{B}}T$) for protein G; these values are in reasonable agreement with the stabilization observed for each protein in Fig. 1. We note that the difference in total potential energy between unfolded and folded, ≈ -21 and ≈ -40 kcal/mol for prb_{7-53} and protein G, respectively, is still dominated by native contact energies, as expected.

1. Kim YC, Hummer G (2008) Coarse-grained models for simulation of multiprotein complexes: Application to ubiquitin binding. *J Mol Biol* 375:1416–1433.
2. Bicout DJ, Szabo A (1998) Electron transfer reaction dynamics in non-debye solvents. *J Chem Phys* 109:2325–2358.
3. Hummer G (2005) Position-dependent diffusion coeficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J Phys* 7:516–523.
4. Best RB, Hummer G (2006) Diffusive model of protein folding dynamics with Kramers turnover in rate. *Phys Rev Lett* 96:228104.
5. Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich ES (1998) On the transition coordinate for protein folding. *J Chem Phys* 108:334–350.
6. Best RB, Hummer G (2005) Reaction coordinates and rates from transition paths. *Proc Natl Acad Sci USA* 102:6732–6737.
7. Karanicolas J, Brooks III CL (2002) The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci* 11:2351–2361.
8. Hu J, Ma A, Dinner AR (2006) Monte Carlo simulations of biomolecules: The MC module in CHARMM. *J Comp Chem* 27:203–216.
9. Dinner AR (2000) Local deformations of polymers with nonplanar rigid main chain internal coordinates. *J Comp Chem* 21:1132–1144.

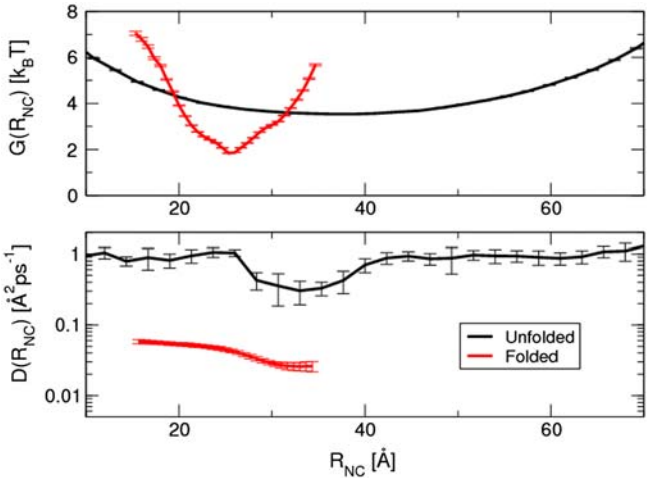

Fig. S1. Estimated $F(r_{NC})$ and $D(r_{NC})$ for a single pair distance r_{NC} (distance between N and C termini in prb). Since for the prb model system this distance is a poor folding coordinate (e.g., it cannot separate folded and unfolded conformations), the folding trajectory was divided into folded ($Q > 0.7$) and unfolded states, and transition matrices $N(r_j, t|r_i, 0)$ were estimated separately from these trajectory fragments. In this way, separate free energies and diffusion coeffi-cients were obtained for this pair distance in the folded (*Red*) and unfolded (*Black*) states of the protein.

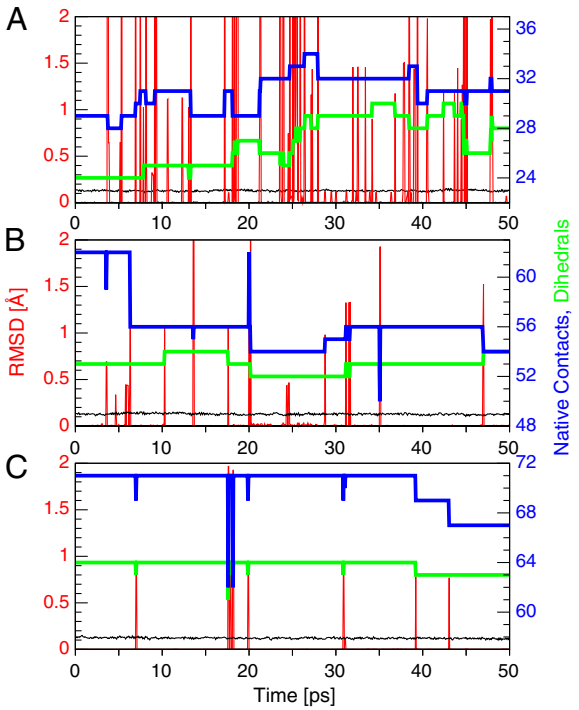

Fig. S2. Barrier crossing from quenched trajectories of prb. Trajectories were initiated from (*A*) an unfolded configuration ($Q \approx 0.4$), (*B*) a transition-state configuration ($Q \approx 0.67$) and (*C*) a folded configuration ($Q \approx 0.9$), and later "quenched" by extensive energy minimization of the saved structures. The left hand scale gives the rms distance (RMSD) to the preceding saved structure in the trajectory, with the RMSD for the original trajectory given in black and for the quenched trajectory in red. The numbers of native contacts and native torsion angles in the quenched trajectory are shown in blue and green, respectively. In (*B*) and (*C*) the number of torsion angles has been increased by adding 15 and 27, respectively, to place it on the same scale as the number of native contacts.

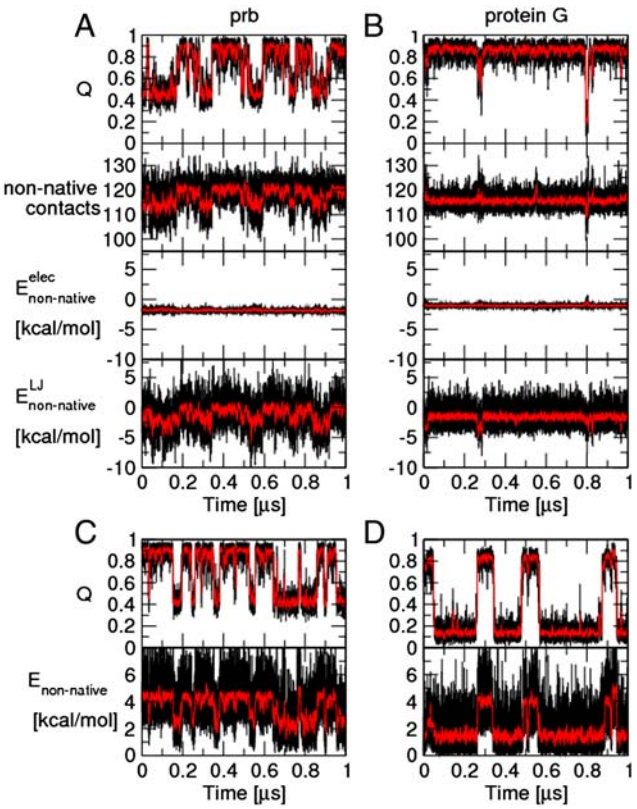

Fig. S3. Non-native contacts in folded and unfolded proteins. The fraction of native contacts, Q, total number of non-native contacts, non-native electrostatic energy and non-native Lennard–Jones energy 1 are shown for (*A*) prb and (*B*) protein G trajectories with the Gō+non-Gō potential. The fraction of native contacts and total non-native (purely repulsive) energy for bare Gō models are given in (*C*) and (*D*) for prb and protein G, respectively. Note that the folded state is slightly destabilized by both the purely repulsive non-native contacts of the original pure Gō model (*C* and *D*), and by the weak non-native attractions in the non-Gō model. To count non-native contacts, the cutoff distance was 1.2σ, with σ taken from the Lennard–Jones part of the non-native potential (1).

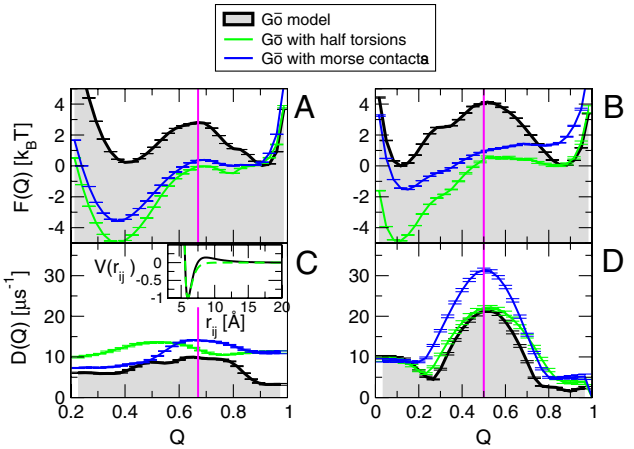

Fig. S4. Effect of changes to torsion angle and native contact potentials. The free energy $F(Q)$ (*Upper*) and position-dependent diffusion coefficients $D(Q)$ (*Lower*) are shown for prb (*Left*) and protein G (*Right*). The color code of the curves is (Black) Karanicolas–Brooks Gō model, (*Red*) non-Gō interactions added, (*Blue*) Morse contact potentials, and (*Green*) halved torsion potentials. The inset to (*E*) compares the Morse pair potential (*Broken Green Lines*) with the original Karanicolas–Brooks "desolvation" potential (7).

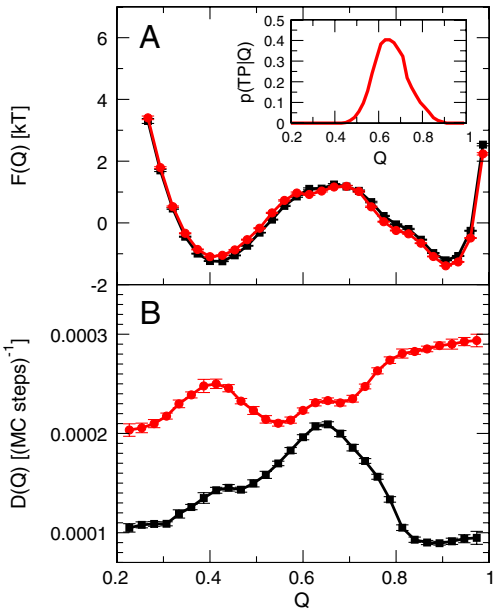

Fig. S5. Monte Carlo (MC) "dynamics" of prb. Position-dependent (*A*) free energies $F(Q)$ and (*B*) diffusion coefficients $D(Q)$ obtained by fitting a 1D diffusive model to MC trajectories; the black curve is for the simple MC with only single torsional rotation moves and the red curve includes moves with concerted rotations of up to seven torsion angles. Inset to (*A*) indicates the quality of the reaction coordinate Q for MC dynamics with $p(TP|Q)$. MC simulations were run using the MC module of CHARMM (8), with one of two move sets: ("simple") Displacement moves of up to 0.2 Å, and single torsion rotations of up to 20°; "concerted" moves available in the simple move set, with the addition of concerted rotation moves of up to seven torsion angles as described by Dinner (9), with maximum 15° displacements. Moves were chosen from a uniform distribution on the allowed intervals. Note that harmonic bonds without SHAKE constraints were used for the MC simulations.

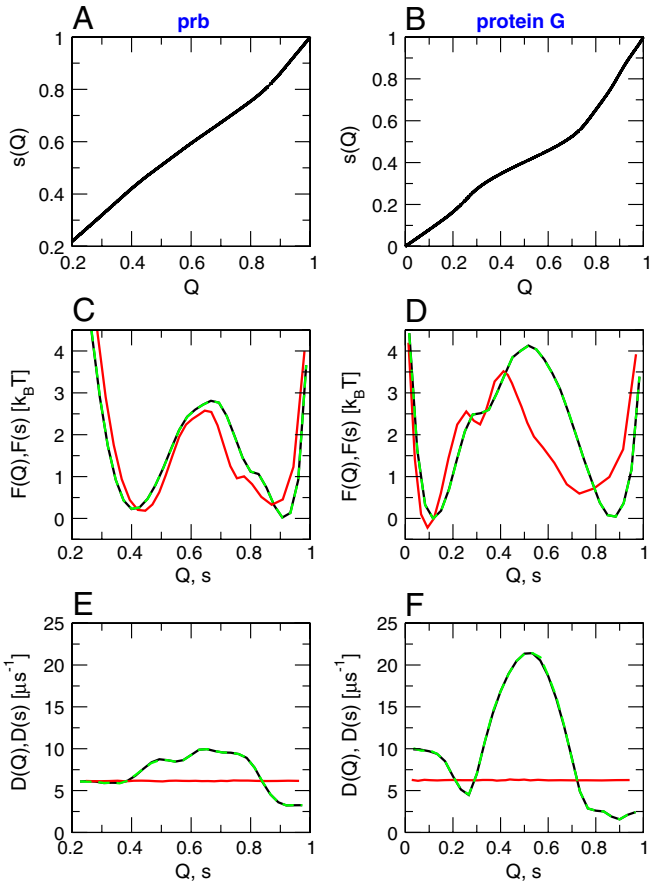

Fig. S6. Mapping the fraction of native contacts Q onto a coordinate s on which the diffusion coefficient $D(s)$ is constant for prb (*Left*) and protein G (*Right*). (*A*),(*B*) Mappings $s(Q)$ defined as described in the text to obtain a position-invariant $D(s)$; like Q, s is defined on the interval $[0, 1]$; (*C*), (*D*) Free energy surfaces $F(s)$ (*Red*) and $F(Q)$ (*Black*) for prb and protein G; (*E*), (*F*) Position-dependent diffusion coefficients $D(s)$ (*Red*) and $D(Q)$ (*Black*). The result of the reverse transform from s back to Q are shown by broken green lines in (*C-F*).

Table S1. Contributions to non-native pair energies in prb_{7-53} and protein G. For each energy term E, we present the average over the unfolded state $\langle E\rangle_U$, folded state $\langle E\rangle_F$, and their difference $\Delta E = \langle E\rangle_F - \langle E\rangle_U$. The terms considered are the Lennard–Jones ($E^{LJ}_{nonG\bar{o}}$) and electrostatic ($E^{elec}_{nonG\bar{o}}$) parts of the non-Gō potential ($E^{tot}_{nonG\bar{o}}$), averaged over simulations with the non-Gō interactions included. Corresponding averages of the non-native pair energies in simulations with the original Gō model ($E_{G\bar{o}}$) are also given. The folded state was defined by $Q > 0.7$ for prb_{7-53} and by $Q > 0.55$ for protein G.

Energy term E	$\langle E\rangle_U$ (kcal/mol)	$\langle E\rangle_F$ (kcal/mol)	$\langle E\rangle_F - \langle E\rangle_U$ (kcal/mol)
prb_{7-53}			
$E^{LJ}_{nonG\bar{o}}$	−2.21	−0.50	1.71
$E^{elec}_{nonG\bar{o}}$	−1.78	−1.85	−0.04
$E^{tot}_{nonG\bar{o}}$	−3.99	−2.35	1.67
$E_{G\bar{o}}$	2.53	4.32	1.80
Protein G			
$E^{LJ}_{nonG\bar{o}}$	−2.67	−1.87	0.80
$E^{elec}_{nonG\bar{o}}$	−0.63	−1.04	−0.41
$E^{tot}_{nonG\bar{o}}$	−3.30	−2.91	0.39
$E_{G\bar{o}}$	1.46	4.09	2.63