Supporting Information

Zhang and Chan 10.1073/pnas.1209891109

SI Text

Computational Details of Our Coarse-Grained Explicit-Chain Model. Our explicit-chain results were simulated using a C_α coarse-grained model with desolvation barriers in its native-centric potential (1, 2) and an improved nonnative repulsive term (3). The definition of native contacts and the Langevin dynamics parameters are identical to the definition and parameters in ref. 2. Time is measured in a number of simulation time steps. For each of the eight proteins in Table S1, a long simulation covering >50 folding–unfolding cycles was used to determine the normalized equilibrium conformational distribution $P_{\rm eq}(Q)$. All results reported in this paper were simulated around each model protein's transition midpoint, i.e., when $P_{\rm eq}(Q_{\rm D}) \approx P_{\rm eq}(Q_{\rm N})$. Kinetic properties of each protein were deduced from ≥1,200 folding trajectories. We have assessed the reliability of our simulations by considering the deviations in results from independent runs of >1,300 trajectories (in addition to runs listed in Table S1) for 2CI2 and 1SHF. For these proteins' data in Fig. 3, the largest deviation in $\langle({\rm CO})_{\rm TP}\rangle/\langle{\rm CO}_{\rm FP}\rangle$ was <2%, and there was essentially no deviation in $\langle[\Delta^{\ddagger}P_{\rm c}(Q)]^2\rangle$. The contact maps for $P_{{\rm FP},ij}(Q^{\ddagger})$ of 1SHF from two independent runs were also practically identical.

Nonexplicit-Chain Monte Carlo Diffusion Model: Metropolis and Kawasaki Algorithms. As discussed in the text, the application of a diffusion perspective to understand experimental protein folding (refs. 4–9 and references therein) has benefitted significantly from recent advances in general theory of diffusive dynamics (refs. 10–12 and references therein). Although often only idealized potential functions (11) rather than more realistic biophysics-based potential functions were used in analytical treatments of diffusion, such developments are important. They established universal principles (10), pointed to potential limitations of using a 1D diffusion picture to describe configurational changes in 3D (12), and provided conceptual insights (e.g., the general trend of cooperative transitions and transition paths exhibited in Fig. 1 for our 2CI2 model is quite similar to figure 1 in ref. 11 for an idealized double-well potential). For protein folding, much theoretical progress and physical understanding has been gained recently from an extensive body of work from Best and Hummer (refs. 4–6 and references therein) and Wang and coworkers (refs. 7–9 and references therein). Among many advances, one of these efforts demonstrated how the parameters governing an approximate mapping between explicit-chain folding/unfolding dynamics and a 1D diffusive process with a coordinate-dependent diffusion coefficient may be optimized (4). While recognizing the importance of these recent achievements, the aim of the present study is not directed primarily to an accurate reproduction of explicit-chain data through the introduction of a coordinate-dependent diffusion coefficient. Instead, as a complement to the explicit-chain, structure-based simulations in this study, we endeavored to uncover basic concepts with regard to the relationship between diffusion and the deviations from preequilibrium (as applied to stopped-flow folding trajectories) and the properties of folding transition paths revealed by our explicit-chain dynamics. For this purpose, we asked how far a simple picture of diffusion with a constant coordinate-independent diffusion coefficient along a free energy profile defined by Q can account for the newly uncovered explicit-chain behaviors. Mathematically, 1D diffusive processes with different coordinate-dependent diffusion coefficients can be mapped onto one another provided that the coordinate (progress) variable is also transformed (6). However, the process that offers arguably the most physical appeal is the one with a coordinate-independent diffusion coefficient $D = D_0$, because it coincides with our intuitive notion of diffusion and spatial homogeneity. Because the coordinate-dependent $D(Q)$ values for model protein free energy profiles defined on the common progress variable Q (which is also a simple intuitive construct) have been found to be "near constant" (6)—thus, no large errors are expected to be incurred by using D_0 instead of $D(Q)$—we stipulated that useful insights would be gained by applying coordinate-independent diffusion on Q-defined free energy profiles to address the issues of preequilibrium and transition paths at hand.

Here, we used nonexplicit-chain MC simulations to model various diffusion processes with an effective coordinate (Q)-independent D_0. All simulations were based on the free energy profile $\beta\Delta G(Q) = -\ln P_{\rm eq}(Q)$ obtained from our explicit-chain model. We found it instructive to apply and compare the common Metropolis criterion (13) and a version of Kawasaki criterion (14, 15) for move acceptance. The former is simpler, whereas the latter is more accurate for our purpose (see below). As described in *Methods*, the probability for an attempted move to be accepted is equal to $\min[1,\exp(-\beta\Delta G_Q)]$, where $\Delta G_Q = G(Q \pm \delta Q) - G(Q)$, in the Metropolis algorithm; whereas the probability is equal to $A^{-1}\exp(-\beta\Delta G_Q/2)$, for some constant A, in the Kawasaki algorithm. As will be shown below, the dynamics that follows from the Metropolis algorithm are a good approximation of—although not equivalent to—a diffusive process governed by the Smoluchowski equation. In contrast, the dynamics that follows from our Kawasaki algorithm are a discretized version of a Smoluchowski process. Nevertheless, the results that we obtained using the two different move acceptance algorithms were extremely similar (Fig. S2).

The Kawasaki algorithm that we used for the present work corresponds to the $\gamma = 0$ case of equation 13 in ref. 15 (relevant discussions in pp. 29–32 of ref. 15). Before considering the detailed derivation below to show that the Kawasaki algorithm is a discretized solution to the Smoluchowski equation with a constant D, it is instructive to inspect the expression for discretized coordinate-dependent diffusion coefficient $D_{i+1/2} \approx \Delta q^2\, R_{i+1,i}\, (P_i/P_{i+1})^{1/2}$ given by equation 6 in *SI Text* of ref. 6, where Δq is the interval between two consecutive discretized coordinate values q_i and q_{i+1}, P_i and P_{i+1} are equilibrium populations, and $R_{i+1,i}$ is the rate coefficient from i to $i + 1$ (6). If we now substitute our Kawasaki expression for $R_{i+1,i} = A^{-1}p_{\pm}\exp[-\beta(G_{i+1} - G_i)/2] = A^{-1}p_{\pm}(P_{i+1}/P_i)^{1/2}$ into the above relationship for $D_{i+1/2}$ (G_i is free energy at q_i), we arrive at a diffusion coefficient $D_{i+1/2} \approx \Delta q^2\, A^{-1}\, p_{\pm}$ that is coordinate-independent. This correspondence is consistent with the relation $D_0 \approx A^{-1}\, p_{\pm}\, (\delta Q)^2$ to be shown below for our Kawasaki Monte Carlo (MC) dynamics. It is noteworthy that a Kawasaki-like half-exponential (square-root Boltzmann factor) form was also adopted for the elements of the transition matrix in the "theoretical model" for ultrafast protein folding in the work by Kubelka et al. (ref. 16, equations S9 and S10) to describe hopping between adjacent reaction coordinates.

Here, we have chosen $A = \exp(1/2) = 1.649$ and accordingly, excluded small regions of the model free energy profiles with $\beta\Delta G_Q > 1$ from Kawasaki MC simulations. Because these regions (all with $Q < Q_{\rm D}$) are energetically highly unfavorable to begin with, adoption of a larger A to further limit the extent of these forbidden regions is not expected to lead to appreciable changes in the MC simulation results.

A total of 2,500–200,000 MC trajectories were simulated for each protein using both the Metropolis and Kawasaki algorithms. In all cases, we verified that the original free energy profile $G(Q)$ was reproduced when simulation was performed without an absorber at Q_{N}.

Metropolis and Kawasaki Algorithms as Discretized Approximate Solutions to the Smoluchowski Equation. Consider the Smoluchowski equation (Eq. **1**) for constant D_0 (Eq. **S1**):

$$\begin{aligned}\partial_t p(Q,t)/D_0 &= \partial_Q\big(\exp[-\beta G(Q)]\,\partial_Q\{\exp[\beta G(Q)]p(Q,t)\}\big)\\ &= \beta\big[\partial^2 G(Q)/\partial Q^2\big]p(Q,t)\\ &\quad+\beta[\partial G(Q)/\partial Q]\,[\partial p(Q,t)/\partial Q]\\ &\quad+\big[\partial^2 p(Q,t)/\partial Q^2\big].\end{aligned}\qquad\textbf{[S1]}$$

For MC simulations using discrete time steps on a free energy profile defined by discrete values of Q with increment $\delta Q = 1/\tilde{Q}_{\mathrm{n}}$ as specified in *Methods*, the differential quantities in the above equation may be approximated by the following discretized expressions:

$$\partial_t p(Q,t) \approx p(Q,t+1) - p(Q,t),\qquad\textbf{[S2]}$$

$$\partial G(Q)/\partial Q \approx \begin{cases}[G(Q+\delta Q)-G(Q)]/(\delta Q) & \textbf{[S3A]}\\ [G(Q)-G(Q-\delta Q)]/(\delta Q)\,, & \textbf{[S3B]}\end{cases}$$

$$\partial p(Q,t)/\partial Q \approx \begin{cases}[p(Q+\delta Q,t)-p(Q,t)]/(\delta Q) & \textbf{[S4A]}\\ [p(Q,t)-p(Q-\delta Q,t)]/(\delta Q)\,, & \textbf{[S4B]}\end{cases}$$

$$\partial^2 G(Q)/\partial Q^2 \approx [G(Q+\delta Q)+G(Q-\delta Q)-2G(Q)]/(\delta Q)^2,\qquad\textbf{[S5]}$$

$$\partial^2 p(Q,t)/\partial Q^2 \approx [p(Q+\delta Q,t)+p(Q-\delta Q,t)-2p(Q,t)]/(\delta Q)^2.\qquad\textbf{[S6]}$$

Note that two equally reasonable discretized expressions are provided by **S3A** and **S3B** for $\partial G(Q)/\partial Q$ and by **S4A** and **S4B** for $\partial p(Q, t)/\partial Q$. We will exploit these alternative forms in the formal development below.

Similarities and Differences Between Metropolis and Smoluchowski Dynamics. It follows from the Metropolis criterion of min[1,exp($-\beta\Delta G_Q$)] that, depending on the relative values of $G(Q-\delta Q)$, $G(Q)$, and $G(Q+\delta Q)$, four cases need to be considered separately:

Case I: $G(Q+\delta Q) > G(Q)$ and $G(Q-\delta Q) > G(Q)$.
Case II: $G(Q+\delta Q) > G(Q)$ and $G(Q-\delta Q) \le G(Q)$.
Case III: $G(Q+\delta Q) \le G(Q)$ and $G(Q-\delta Q) > G(Q)$.
Case IV: $G(Q+\delta Q) \le G(Q)$ and $G(Q-\delta Q) \le G(Q)$.

For case I [i.e., $G(Q+\delta Q) > G(Q)$ and $G(Q-\delta Q) > G(Q)$],

$$\begin{aligned}{[p(Q,t+1)-p(Q,t)]/p_\pm} &= p(Q-\delta Q,t)+p(Q+\delta Q,t)\\ &\quad -p(Q,t)\{\exp(-\beta[G(Q+\delta Q)-G(Q)])\\ &\quad +\exp(-\beta[G(Q-\delta Q)-G(Q)])\}\\ &\approx p(Q-\delta Q,t)+p(Q+\delta Q,t)-2p(Q,t)\\ &\quad +\beta p(Q,t)[G(Q+\delta Q)-2G(Q)+G(Q-\delta Q)]\\ &\approx\{[\partial^2 p(Q,t)/\partial Q^2]+\beta p(Q,t)[\partial^2 G(Q)/\partial Q^2]\}(\delta Q)^2,\end{aligned}$$

where the first equality follows from our Metropolis MC algorithm (the rate of change of $p(Q, t)$ with respect to time t is a product of the overall left/right transition probability $p_\pm$ and the min[1,exp($-\beta\Delta G_Q$)] criterion, hence, the $1/p_\pm$ factor on the left-hand side). The first $\approx$ entails approximating exp(x) by $1 + x$, which is valid for small x (the same applies to cases II–IV and the Kawasaki MC case below), whereas the last step (second $\approx$ sign) was reached by using expressions **S5** and **S6**. Thus, for this case, in which $G(Q)$ is a local minimum, the Metropolis MC algorithm accounts for two of three terms in the Smoluchowski equation in Eq. **S1** if D_0 is identified with $p_\pm\,(\delta Q)^2$. However, the Metropolis MC algorithm fails to account for the $\beta[\partial G(Q)/\partial Q][\partial p(Q, t)/\partial Q]$ term in the Smoluchowski equation.

For case II [i.e., $G(Q+\delta Q) > G(Q)$ and $G(Q-\delta Q) \le G(Q)$],

$$\begin{aligned}{[p(Q,t+1)-p(Q,t)]/p_\pm} &= p(Q+\delta Q,t)-p(Q,t)\\ &\quad -p(Q,t)\exp(-\beta[G(Q+\delta Q)-G(Q)])\\ &\quad +p(Q-\delta Q,t)\exp(-\beta[G(Q)-G(Q-\delta Q)])\\ &\approx p(Q+\delta Q,t)-2p(Q,t)+p(Q-\delta Q,t)\\ &\quad +\beta p(Q,t)\,[G(Q+\delta Q)-G(Q)]\\ &\quad -\beta p(Q-\delta Q,t)[G(Q)-G(Q-\delta Q)]\\ &= p(Q+\delta Q,t)-2p(Q,t)+p(Q-\delta Q,t)\\ &\quad +\beta p(Q,t)[G(Q+\delta Q)-2G(Q)+G(Q-\delta Q)]\\ &\quad +\beta[p(Q,t)-p(Q-\delta Q,t)][G(Q)-G(Q-\delta Q)]\\ &\approx\{[\partial^2 p(Q,t)/\partial Q^2]+\beta p(Q,t)[\partial^2 G(Q)/\partial Q^2]\\ &\quad +\beta[\partial G(Q)/\partial Q][\partial p(Q,t)/\partial Q]\}(\delta Q)^2,\end{aligned}$$

where the last step was reached by using expressions **S3B**, **S4B**, **S5**, and **S6**. The second equality in the above equation amounts only to a rearrangement of terms. Thus, by comparing Eq. **S1** with the approximate equality between the left-hand side and the expression after the second $\approx$ sign in the above formulation, it is clear that the Metropolis MC algorithm is a discretized version of the Smoluchowski equation, with $D_0 = p_\pm\,(\delta Q)^2$ for this case.

For case III [i.e., $G(Q+\delta Q) \le G(Q)$ and $G(Q-\delta Q) > G(Q)$],

$$\begin{aligned}{[p(Q,t+1)-p(Q,t)]/p_\pm} &= p(Q-\delta Q,t)-p(Q,t)\\ &\quad -p(Q,t)\exp(-\beta[G(Q-\delta Q)-G(Q)])\\ &\quad +p(Q+\delta Q,t)\exp(-\beta[G(Q)-G(Q+\delta Q)])\\ &\approx p(Q-\delta Q,t)-2p(Q,t)+p(Q+\delta Q,t)\\ &\quad +\beta p(Q,t)[G(Q-\delta Q)-G(Q)]\\ &\quad +\beta p(Q+\delta Q,t)[G(Q+\delta Q)-G(Q)]\\ &= p(Q-\delta Q,t)-2p(Q,t)+p(Q+\delta Q,t)\\ &\quad +\beta p(Q,t)[G(Q-\delta Q)-2G(Q)+G(Q+\delta Q)]\\ &\quad +\beta[p(Q+\delta Q,t)-p(Q,t)][G(Q+\delta Q)-G(Q)]\\ &\approx\{[\partial^2 p(Q,t)/\partial Q^2]+\beta p(Q,t)[\partial^2 G(Q)/\partial Q^2]\\ &\quad +\beta[\partial G(Q)/\partial Q][\partial p(Q,t)/\partial Q]\}(\delta Q)^2,\end{aligned}$$

where the last step was reached by using expressions **S3A**, **S4A**, **S5**, and **S6**. The second equality amounts only to a rearrangement of terms. Thus, similar to case II, the Metropolis MC algorithm is seen as a discretized version of the Smoluchowski equation, with $D_0 = p_\pm\,(\delta Q)^2$ for this case as well.

For case IV [i.e., $G(Q+\delta Q) \le G(Q)$ and $G(Q-\delta Q) \le G(Q)$],

$$\begin{aligned}{[p(Q,t+1)-p(Q,t)]/p_\pm} &= -2p(Q,t)\\ &\quad +p(Q-\delta Q,t)\exp(-\beta[G(Q)-G(Q-\delta Q)])\\ &\quad +p(Q+\delta Q,t)\exp(-\beta[G(Q)-G(Q+\delta Q)])\\ &\approx -2p(Q,t)+p(Q-\delta Q,t)+p(Q+\delta Q,t)\\ &\quad -\beta p(Q-\delta Q,t)[G(Q)-G(Q-\delta Q)]\\ &\quad +\beta p(Q+\delta Q,t)[G(Q+\delta Q)-G(Q)]\\ &= -2p(Q,t)+p(Q+\delta Q,t)+p(Q-\delta Q,t)\\ &\quad +\beta p(Q,t)[G(Q+\delta Q)-2G(Q)+G(Q-\delta Q)]\\ &\quad +\beta[p(Q,t)-p(Q-\delta Q,t)][G(Q)-G(Q-\delta Q)]\\ &\quad +\beta[p(Q+\delta Q,t)-p(Q,t)][G(Q+\delta Q)-G(Q)]\\ &\approx\{[\partial^2 p(Q,t)/\partial Q^2]+\beta p(Q,t)[\partial^2 G(Q)/\partial Q^2]\\ &\quad +2\beta[\partial G(Q)/\partial Q][\partial p(Q,t)/\partial Q]\}(\delta Q)^2,\end{aligned}$$

where the last step was reached by using expressions **S3–S6**. Again, the second equality amounts only to a rearrangement of terms. Thus, for this case, in which $G(Q)$ is a local maximum, if D_0 is identified with $p_\pm\,(\delta Q)^2$, the Metropolis MC algorithm correctly accounts for two of three terms in Eq. **S1** but affords a coefficient to the $\beta[\partial G(Q)/\partial Q][\partial p(Q, t)/\partial Q]$ term that is double the coefficient in the Smoluchowski equation.

Taken together, Metropolis MC describes a diffusion process that is a discretized version of the process governed by the Smoluchowski equation with $D_0 = p_\pm (\delta Q)^2$ when $G(Q)$ is not a local minimum or a local maximum. Interestingly, the average coefficient for the $\beta[\partial G(Q)/\partial Q]$ $[\partial p(Q,-t)/\partial Q]$ term in Metropolis MC (zero for local minima and two for local maxima) is equal to the coefficient in the Smoluchowski equation. Because an overwhelming majority of the $G(Q)$ values in the free energy profiles that we considered is neither local minimum nor local maximum, we expect Metropolis MC to provide a good approximation to diffusive dynamics. Moreover, when $G(Q)$ is a local minimum or local maximum (in a discrete sense), the $\partial G(Q)/\partial Q$ value is expected to be ≈0, and thus, the errors incurred by using Metropolis MC to model diffusive dynamics should not be significant, because the error terms are proportional to $\partial G(Q)/\partial Q$. Consistent with this expectation, for all except one of our model proteins, the Metropolis MC-simulated mean first passage times (MFPTs) with $p_\pm = 0.45$ entail effective $D_0 \approx 0.44$–0.48 [in units of $(\delta Q)^2$] when compared against the analytical formula for MFPT derived from Eq. **1** (Fig. S2). All results from Metropolis MC simulations shown in this work were obtained using $p_\pm = 0.45$.

Correspondence Between Kawasaki and Smoluchowski Dynamics. According to our Kawasaki algorithm,

$$\begin{aligned}
[p(Q,t+1)-p(Q,t)]/A^{-1}p_\pm &= -p(Q,t)\{\exp(-\beta[G(Q+\delta Q)-G(Q)]/2)+\exp(-\beta[G(Q-\delta Q)-G(Q)]/2)\}\\
&\quad +p(Q+\delta Q,t)\exp(-\beta[G(Q)-G(Q+\delta Q)]/2)+p(Q-\delta Q,t)\exp(-\beta[G(Q)-G(Q-\delta Q)]/2)\\
&\approx -2p(Q,t)+p(Q+\delta Q,t)+p(Q-\delta Q,t)+\beta p(Q,t)[G(Q+\delta Q)-2G(Q)+G(Q-\delta Q)]/2\\
&\quad -\beta p(Q+\delta Q,t)[G(Q)-G(Q+\delta Q)]/2-\beta p(Q-\delta Q,t)[G(Q)-G(Q-\delta Q)]/2\\
&= -2p(Q,t)+p(Q+\delta Q,t)+p(Q-\delta Q,t)+\beta p(Q,t)[G(Q+\delta Q)-2G(Q)+G(Q-\delta Q)]\\
&\quad +\beta[p(Q+\delta Q,t)-p(Q,t)][G(Q+\delta Q)-G(Q)]/2+\beta[p(Q,t)-p(Q-\delta Q,t)][G(Q)-G(Q-\delta Q)]/2\\
&\approx \{[\partial^2 p(Q,t)/\partial Q^2]+\beta p(Q,t)[\partial^2 G(Q)/\partial Q^2]+\beta[\partial G(Q)/\partial Q][\partial p(Q,t)/\partial Q]\}(\delta Q)^2.
\end{aligned}$$

Again, the first equality follows from the Kawasaki criterion itself, whereas the last step was reached by using expressions **S3–S6**. Thus, within the approximations represented by the ≈ signs described above, Kawasaki MC dynamics are a discretized version of a diffusion process described by the Smoluchowski equation with $D_0 = A^{-1}p_\pm (\delta Q)^2$. We used $p_\pm = 0.5$ for our Kawasaki MC simulations.

Similarities and Differences Between Conformations Sampled by the Explicit-Chain Folding and Unfolding Trajectories in the Barrier Region. To assess the applicability of the common transition state theory (TST)-inspired notion that the transition states encountered in stopped-flow folding and unfolding are identical for two-state proteins (17, 18), we have compared the contact pattern of the conformations in the barrier region sampled along folding paths (FPs) with those patterns sampled in equilibrium simulations that include both folding and unfolding trajectories (Fig. S7*A*). The results show that FP and equilibrium contact patterns at $Q^\ddagger$ can be significantly different (except for models 1BDD and 1HYW in Fig. S7*A*), indicating that conformations traversed during stopped-flow folding and unfolding are not statistically identical for these model proteins. Interestingly, folding and unfolding kinetics seem to be more symmetric for the two fastest folding model proteins 1BDD and 1HYW, because they exhibit close to zero differences at $Q^\ddagger$.

It is important to note that the observation in Fig. S7*A* does not violate time reversal symmetry, which is a fundamental tenet in classical physics and applies to our model Langevin dynamics (ref. 19 directly verifies this principle in a Langevin dynamics simulation of translocation times for particles traversing membrane channels). Although every forward path is matched to an equally likely reverse path (e.g., the paths A-B-C and C-B-A in Fig. S7*B*) as required by time reversal symmetry, part of a stopped-flow folding path (blue path from A to B in Fig. S7*B*) may not be counted (measured) as part of a stopped-flow unfolding path. In the example in Fig. S7*B*, the dashed red path from B to A is not part of a stopped-flow unfolding path. (Schematic drawings similar to those drawings in Fig. S7 have been used in ref. 20.) This is because both A-B-C and C-B-A in Fig. S7*B* recross Q_D, which serves as the starting point for folding paths and the finish line for unfolding paths. Aiming to capture the essential feature of stopped-flow measurements, our definition of folding and unfolding paths allows the paths to recross their respective starting points at Q_D and Q_N, but the path is considered to be completed when it first crosses the finish line at Q_N and Q_D, respectively (i.e., the folding and unfolding paths do not recross their respective finish line). If part of the folding (or unfolding) path that is not counted to the reverse unfolding (or folding) path also recrosses $Q^\ddagger$ in the transition region (as for the A-B part of the folding path A-B-C in Fig. S7*B*), it is possible that the transition-state conformations sampled by folding and unfolding paths are different, which is exemplified by our explicit-chain simulation data in Fig. S7*A*.

Another manifestation of the asymmetry between folding and unfolding is the ensemble of conformations in the transition region last visited by a folding transition path vs. those conformations last visited by an unfolding transition path. These conformations are of intuitive interest, because they may play a comparatively more important role in taking the protein over the mountain top to facilitate folding or unfolding than other conformations in the transition region. Because all folding transition paths start at Q_D and end at Q_N without recrossing Q_D and all unfolding transition paths start at Q_N and end at Q_D without recrossing Q_N, time reversal symmetry requires that each folding transition path (e.g., blue path from D to E in Fig. S7*B*) is matched to an equally likely reverse unfolding transition path (e.g., red path from E to D in Fig. S7*B*). If these transition paths recross $Q^\ddagger$, however, the $Q^\ddagger$ configuration last visited by the folding transition path (e.g., *v* in Fig. S7*B*) can be different from the $Q^\ddagger$ configuration last visited by the unfolding transition path (e.g., *iv* in Fig. S7*B*). For all eight model proteins in the present study, we found appreciable differences between the contact patterns of folding and unfolding trajectories when they last visited $Q^\ddagger$ (Fig. S7*C*), indicating that there is significant recrossing of the barrier region.

Since the early experiments on ribonuclease A (21), it has long been known that conformations sampled along stopped-flow folding and unfolding pathways can be different. Although it has been argued that reversibility of stopped-flow folding and unfolding trajectories applies to some proteins, such as chymotrypsin inhibitor 2, within experimental and computational uncertainties (18, 22), such behavior is not universal. A more recent example is single-chain monellin. Experiment showed that its kinetics of unfolding is less complex than the kinetics of folding (23). One possible origin of such asymmetry between folding and unfolding is kinetic trapping caused by nonnative interactions during folding (15, 24, 25). In this regard, it is noteworthy that Fig. S7 *A* and *C* indicates that asymmetry between stopped-flow folding and unfolding kinetics can also arise in native-centric models in the absence of favorable nonnative interactions. These results show that kinetic subtleties in protein folding beyond common notions inspired by 1D TST can readily emerge in explicit-chain dynamics.

1. Cheung MS, García AE, Onuchic JN (2002) Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc Natl Acad Sci USA* 99(2):685–690.
2. Ferguson A, Liu Z, Chan HS (2009) Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins. *J Mol Biol* 389(3): 619–636. Corrigendum: 401:153(2010).
3. Zhang Z, Chan HS (2010) Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. *Proc Natl Acad Sci USA* 107(7):2920–2925.
4. Hummer G (2005) Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J Phys* 7:34.
5. Best RB, Hummer G (2006) Diffusive model of protein folding dynamics with Kramers turnover in rate. *Phys Rev Lett* 96(22):228104.
6. Best RB, Hummer G (2010) Coordinate-dependent diffusion in protein folding. *Proc Natl Acad Sci USA* 107(3):1088–1093.
7. Chahine J, Oliveira RJ, Leite VBP, Wang J (2007) Configuration-dependent diffusion can shift the kinetic transition state and barrier height of protein folding. *Proc Natl Acad Sci USA* 104(37):14646–14651.
8. Oliveira RJ, Whitford PC, Chahine J, Leite VBP, Wang J (2010) Coordinate and time-dependent diffusion dynamics in protein folding. *Methods* 52(1):91–98.
9. Xu W, Lai Z, Oliveira RJ, Leite VBP, Wang J (2012) Configuration-dependent diffusion dynamics of downhill and two-state protein folding. *J Phys Chem B* 116(17): 5152–5159.
10. Zuckerman DM, Woolf TB (2002) Transition events in butane simulations: Similarities across models. *J Chem Phys* 116:2586–2591.
11. Zhang BW, Jasnow D, Zuckerman DM (2007) Transition-event durations in one-dimensional activated processes. *J Chem Phys* 126(7):074504.
12. Cheng RR, Makarov DE (2011) Failure of one-dimensional Smoluchowski diffusion models to describe the duration of conformational rearrangements in floppy, diffusive molecular systems: A case study of polymer cyclization. *J Chem Phys* 134(8): 085104.
13. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092.
14. Kawasaki K (1966) Diffusion constants near the critical point for time-dependent Ising models. I. *Phys Rev* 145:224–230.
15. Chan HS, Dill KA (1998) Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins* 30(1):2–33.
16. Kubelka J, Henry ER, Cellmer T, Hofrichter J, Eaton WA (2008) Chemical, physical, and theoretical kinetics of an ultrafast folding protein. *Proc Natl Acad Sci USA* 105(48): 18655–18662.
17. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224(3):771–782.
18. Jackson SE, elMasry N, Fersht AR (1993) Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: A critical test of the protein engineering method of analysis. *Biochemistry* 32(42):11270–11278.
19. Berezhkovskii AM, Hummer G, Bezrukov SM (2006) Identity of distributions of direct uphill and downhill translocation times for particles traversing membrane channels. *Phys Rev Lett* 97(2):020601.
20. Chaudhury S, Makarov DE (2010) A harmonic transition state approximation for the duration of reactive events in complex molecular arrangements. *J Chem Phys* 113: 034118.
21. Tsong TY, Baldwin RL, Elson EL (1971) The sequential unfolding of ribonuclease A: Detection of a fast initial phase in the kinetics of unfolding. *Proc Natl Acad Sci USA* 68 (11):2712–2715.
22. McCully ME, Beck DAC, Daggett V (2008) Microscopic reversibility of protein folding in molecular dynamics simulations of the engrailed homeodomain. *Biochemistry* 47 (27):7079–7089.
23. Patra AK, Udgaonkar JB (2007) Characterization of the folding and unfolding reactions of single-chain monellin: Evidence for multiple intermediates and competing pathways. *Biochemistry* 46(42):11727–11743.
24. Chan HS, Dill KA (1994) Transition states and folding dynamics of proteins and heteropolymers. *J Chem Phys* 100:9238–9257.
25. Guo Z, Thirumalai D (1995) Kinetics of protein folding – Nucleation mechanism, time scales and pathways. *Biopolymers* 36:83–102.

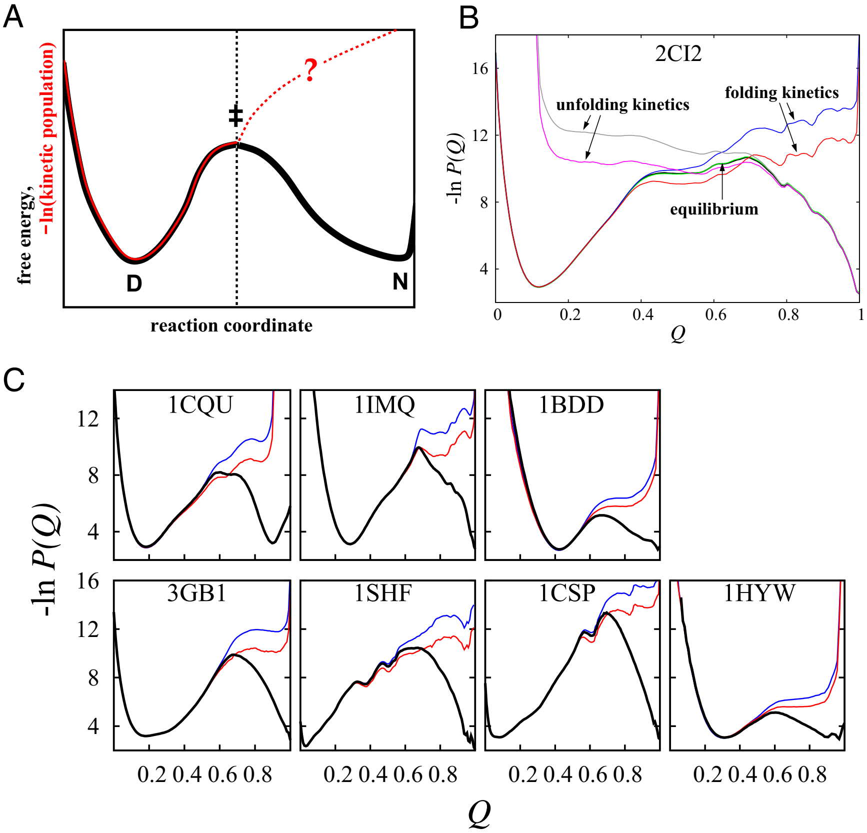

Fig. S1. Evaluating the preequilibrium idea. (*A*) The preequilibrium assumption is illustrated by a hypothetical free energy profile along a reaction coordinate for folding; here D, ‡, and N denote, respectively, the denatured (unfolded), transition, and native (folded) states of the protein. In the common TST-inspired interpretation of stopped-flow folding kinetics data, a complete overlap between the thermodynamic free energy profile (black curve) and a profile of the nonequilibrium population during the transient folding process ($-\ln P_{\mathrm{FP}}$; red curve) is assumed for the entire unfolded (D) regime up to the peak of the free energy profile at ‡. The dotted red curve underscores that TST for folding provides no prediction for $-\ln P_{\mathrm{FP}}$ between ‡ and N. (*B*) Simulated thermodynamic and kinetic profiles of model 2CI2. The black thermodynamic profile (marked as "equilibrium" and seen here as overlapping with other profiles plotted in other colors) and the thin blue and red curves (marked as "folding kinetics") are equivalent, respectively, to the $-\ln P_{\mathrm{eq}}(Q) + c$, $-\ln P_{\mathrm{FP}}(Q)$, and $-\ln P_{\mathrm{FP}}|_{\mathrm{s}}(Q)$ explicit-chain simulated profiles in Fig. 2*A* (plotted in the same color). The only difference is that the shift ($c = -0.30$) to highlight conformity to a folding preequilibrium is now applied to the kinetic profiles, and therefore, the black, blue, and red profiles here are, respectively, $-\ln P_{\mathrm{eq}}(Q)$, $-\ln P_{\mathrm{FP}}(Q) - c$, and $-\ln P_{\mathrm{FP}}|_{\mathrm{s}}(Q) - c$. In the present plot, the magenta curve is the ensemble unfolding path profile $-\ln P_{\mathrm{UFP}}(Q) - c'$, and the gray curve is the single-molecule unfolding path profile $-\ln P_{\mathrm{UFP}}|_{\mathrm{s}}(Q) - c'$ of the same model 2CI2 ($c' = -1.38$). These profiles, marked as "unfolding kinetics", are analogous, respectively, to the $-\ln P_{\mathrm{FP}}(Q) - c$ and $-\ln P_{\mathrm{FP}}|_{\mathrm{s}}(Q) - c$ profiles; the only difference is that the $-\ln P_{\mathrm{UFP}}(Q) - c'$ and $-\ln P_{\mathrm{UFP}}|_{\mathrm{s}}(Q) - c'$ profiles are defined by unfolding trajectories instead of folding trajectories, with c' introduced to highlight conformity to an unfolding preequilibrium. The green profile is the combination, $-\ln[P_{\mathrm{FP}}(Q) + P_{\mathrm{UFP}}(Q)]$, of the ensemble folding path and unfolding path kinetic profiles (no shift). The green profile overlaps almost perfectly, which it should, with the $-\ln P_{\mathrm{eq}}(Q)$ profile (no shift) obtained from thermodynamic simulations. (*C*) Thermodynamic and kinetic FP profiles of the other seven model proteins that we studied. Results are equivalent to the results in Fig. 2*A* but without the nonexplicit-chain MC and analytically derived data. The simulated equilibrium free energy profile $-\ln P_{\mathrm{eq}}(Q) + c$ and the kinetic $-\ln P_{\mathrm{FP}}(Q)$ and $-\ln P_{\mathrm{FP}}|_{\mathrm{s}}(Q)$ profiles are shown, respectively, by the black curve and the thin blue and red curves. A value for c is chosen for each model protein to allow the thermodynamic and kinetic profiles to coincide around Q_{D} so as to highlight the extent of preequilibration. Here, $c = -0.30$, -0.24, -0.45, -0.18, -0.48, -0.19, and -0.6, respectively, for 1CQU, 1IMQ, 1BDD, 3GB1, 1SHF, 1CSP, and 1HYW.

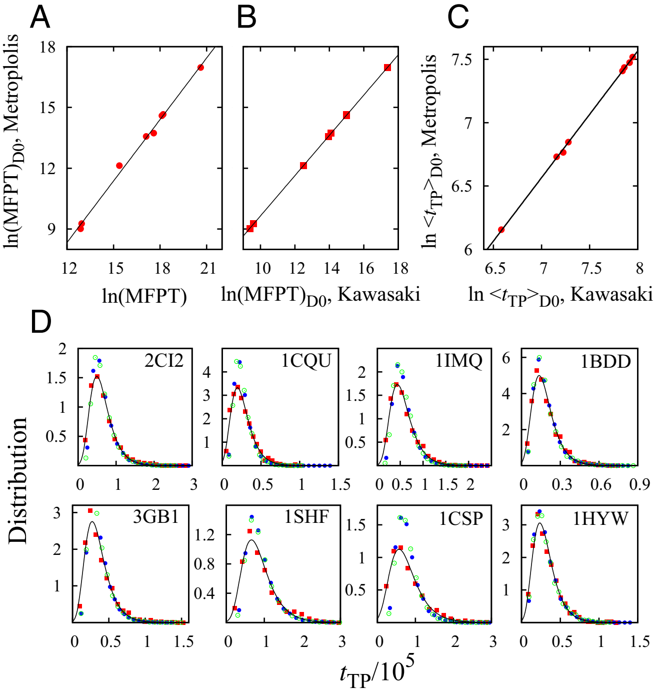

Fig. S2. Nonexplicit-chain constant-D_0 diffusion provides a good rationalization for general trends in the explicit-chain simulation data. (*A*) An excellent correlation is observed between ln(MFPT) from explicit-chain simulations (horizontal axis) and ln(MFPT)$_{D0}$ from nonexplicit-chain MC simulations using the Metropolis algorithm (vertical axis). The MFPT values are provided in Table S2. The line through the data points is the least-squares fit with slope = 1.01 and *r* = 0.997. Each (MFPT)$_{D0}$ value was averaged from 2,500–200,000 MC trajectories using an equal probability $p_\pm = 0.45$ for attempting either a $Q \to Q + \delta Q$ or a $Q \to Q - \delta Q$ transition at every MC time step. A near-perfect agreement that covers ~3.3 orders of magnitude is seen between our MC simulation and explicit-chain dynamics with respect to the folding rates. From the MC-simulated (MFPT)$_{D0}$ values, we define an effective diffusion coefficient $D_{\rm eff} = \sum_{Q=Q_D}^{Q_N} P_{\rm eq}(Q)^{-1} \sum_{Q'=0}^{Q} P_{\rm eq}(Q')/(\mathrm{MFPT})_{D0}$ by using the theoretical expression in Eq. **5** for (MFPT)$_{D0}$. Note that the above expression for $D_{\rm eff}$ is in units of $(\delta Q)^2$. Consistent with expectation (in the text), we found that $D_{\rm eff} \approx p_\pm = 0.45$. For the eight proteins studied (in the same order as listed in Table S1), $D_{\rm eff}$ = 0.45, 0.45, 0.43, 0.44, 0.44, 0.44, 0.48, and 0.39, respectively. (*B* and *C*) Excellent correlation is seen between the values computed using Metropolis and Kawasaki MC algorithms for the logarithmic MFPT (*B*; *r* = 0.99998) and the logarithmic ‹$t_{\rm TP}$› (*C*; *r* = 0.99975). (*D*) Distribution of $t_{\rm TP}$ for the eight model proteins that we studied. In each panel, the continuous distribution curve was fitted to the red data points from explicit-chain Langevin dynamics simulations, which is shown in Fig. 2*C*. Also included here are blue and green data points obtained from nonexplicit-chain (MC) simulations using the Metropolis and Kawasaki algorithms, respectively. To facilitate comparison, the MC results are shown here in time units that were set by requiring the average MC $t_{\rm TP}$ to be equal to the ‹$t_{\rm TP}$› from explicit-chain simulations [e.g., for 2CI2, the plotted Metropolis MC-simulated $t_{\rm TP}$ value is ‹$t_{\rm TP}$›/‹$t_{\rm TP}$›$_{D0}$ = 39.2 times the number of Metropolis MC steps; the corresponding factors for other model proteins are given by their respective ‹$t_{\rm TP}$›/‹$t_{\rm TP}$›$_{D0}$ values in Table S2]. The results show that the three $t_{\rm TP}$ distributions are very similar. It should be noted that the general form of the continuous distribution curves $P(t_{\rm TP}) = Bf\exp(-f\, t_{\rm TP})\exp[-B\exp(-f\, t_{\rm TP})]/[1 - \exp(-B)]$ that we used to fit simulation data here and in Fig. 2*C* was derived in the work by Malinin and Chernyak (1) using idealized potential functions that are similar but not identical to the free energy profiles in our explicit-chain protein models. In this formula, $B = \beta G$, where *G* is the quantity defined in the work by Malinin and Chernyak (1) and should not be confused with the free energy used elsewhere in the present paper and *f* is the quantity *k* in the same reference. The fitting parameters for the eight model proteins are (listed in the order as in Table S1) *B* = 7.7, 5.9, 7.9, 6.4, 7.6, 7.6, 5.7, and 7.1 and *f* = 4.1, 9.0, 4.7, 13.6, 7.5, 3.1, 3.1, and 8.3. We found no clear correspondence between the fitting parameters for the continuous curve and features (such as curvature) in our simulated free energy profiles.

1. Malinin SV, Chernyak VY (2010) Transition times in the low-noise limit of stochastic dynamics. *J Chem Phys* 132(1):014504.

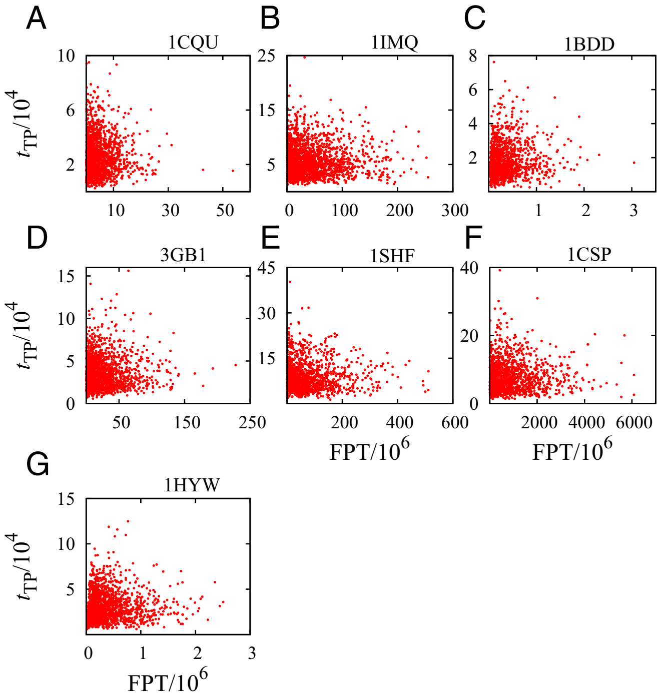

Fig. S3. Relationship between explicit-chain simulated transition path time (t_{TP}) and FPT. Same as Fig. 2*D* but for the other seven proteins studied in this work. The scatters are essentially random, with $r = -0.025$, -0.008, 0.057, 0.041, -0.105, 0.003, and 0.015, respectively, for *A*–*G*.

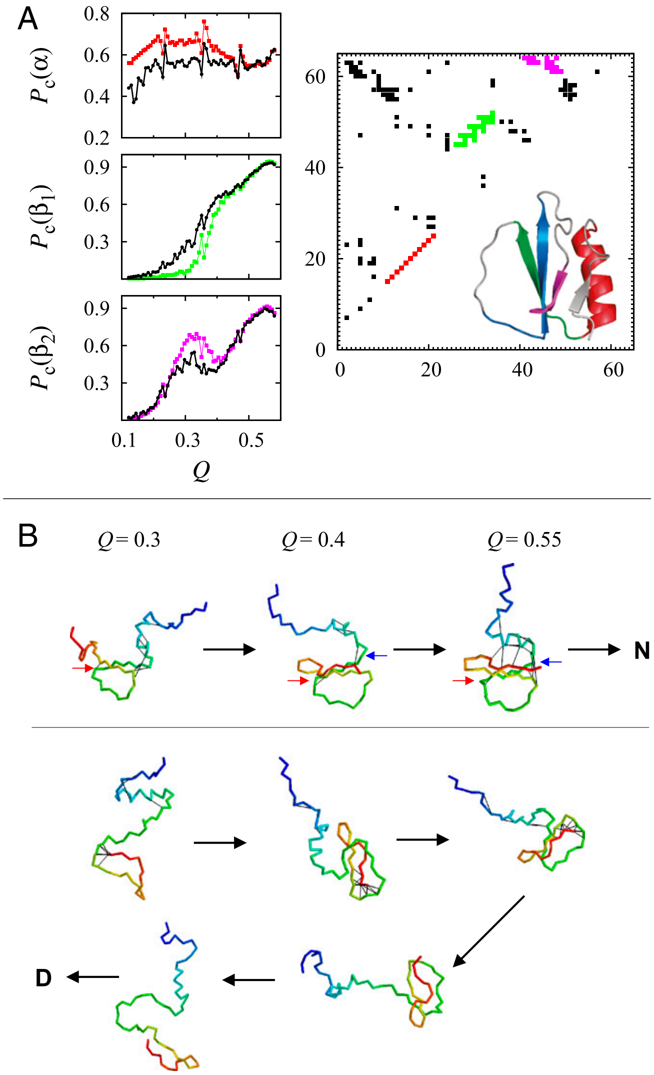

Fig. S4. Additional illustrations of the atypical nature of transition paths (TPs) among the folding trajectories in the quasi-preequilibrium. Results shown are for model 2CI2. (*A*, *Left*) Contact probabilities (P_c) of selected secondary structure elements along TPs (black data points) are compared with the corresponding probabilities along FPs (color data points). Each data point in *Left* was averaged from more than 400 sampled conformations. (*A*, *Right*) Residue numbers *i,j* are represented by the horizontal and vertical axes of the contact map. The native contacts of 2CI2 considered in *Left* are shown in color; the other native contacts are shown in black. The ribbon diagram (*Right*) is the 2CI2 native structure. In the contact map, contacts marked by red, green, and magenta denote, respectively, the contacts in an α-helix (α; shown in red in the ribbon diagram), a parallel β-sheet (β_1; formed by the green and blue strands in the ribbon diagram), and an antiparallel β-sheet (β_2; formed by the magenta and blue strands in the ribbon diagram). From *Left*, the TPs (black data points) are seen to have a smoother increases in P_c as Q increases (and a better approximation to a monotonic increase) vis-à-vis the corresponding variations of P_c with Q along FPs (color data points). (*B*) Comparing a folding transition path with a nontransition trajectory in the unfolded (denatured) basin. Selected 2CI2 conformations with the given Q values (0.3, 0.4, and 0.55) (compare with Fig. 4) along a TP (*Upper*) and a nontransition trajectory (non-TP; *Lower*) are each depicted as a C_α trace colored from blue (N terminus) to red (C terminus). The black arrows between conformations indicate time progression. Every TP in our 2CI2 model starts from the unfolded (denatured) minimum of the free energy profile at $Q = Q_D \approx 0.12$ and ends at the native state at $Q = Q_N = 1$ without revisiting Q_D. In contrast, a non-TP starting at Q_D cannot reach Q_N unless it revisits Q_D at least one time. To contrast the different patterns of contact development between this particular pair of TP and non-TP trajectories, all 14 contacts in the $Q = 0.55$ TP conformation that are not in the $Q = 0.55$ non-TP conformation are marked by connecting lines between C_α positions in the $Q = 0.55$ TP conformation. (Each of the $Q = 0.55$ conformations has a total of 72 native contacts.) Some of these contacts already exist in the earlier $Q = 0.3$ and $Q = 0.4$ TP conformations along the TP trajectory; those contacts are marked in the $Q = 0.3$ and $Q = 0.4$ TP conformations as well. Likewise, all 14 contacts in the $Q = 0.55$ non-TP conformation that are not in the $Q = 0.55$ TP conformation are marked in the $Q = 0.55$ non-TP conformation. Accordingly, subsets of these contacts that exist in the first $Q = 0.3$ and the first $Q = 0.4$ non-TP conformations (forward branch of the non-TP) are marked. The contacts indicated by the small red and blue arrows along the TP are, respectively, between residues 34 and 50 and between residues 24 and 63. The contact between 34 and 50 is in the β_1 parallel β-sheet structure (contact map in *A*). The early formation and persistence of this contact is in line with the higher average population of β_1 in TP than in non-TP (*A*, *Middle Left*). The contact between 24 and 63 has a high contact order (= 63–24 = 39). Bringing the helix and the last β-strand at the C terminus into proximity, this contact helps restrict conformation freedom and facilitate folding. In contrast, the highest contact order among those contacts in the shown $Q = 0.55$ non-TP conformation that do not exist in the TP conformation with the same Q is only equal to 23 (between residues 40 and 63). These observations may point to a rationalization for the preference for structures with higher contact order at early stages of TP relative to non-TP structures with the same Q. For our model 2CI2, the average contact order values for TPs and non-TPs at $Q = 0.3$, 0.4, and 0.55 are $\langle CO_{TP} \rangle$ = 0.0470, 0.0733, and 0.116, respectively, and $\langle CO_{FP} \rangle$ = 0.0423, 0.0707, and 0.116, respectively. The TP conformations shown here may be considered to be representative in that their CO_{TP} values are 0.0470, 0.0774, and 0.118, respectively, which are very similar to the corresponding $\langle CO_{TP} \rangle$ values. Likewise, the CO_{FP} values for the non-TP conformations shown here in the forward (backward) branch are 0.0421 (0.0383), 0.0701 (0.0686), and 0.115, respectively, which are also

Legend continued on following page

quite similar to the corresponding ‹$\mathrm{CO_{FP}}$› values for Q = 0.3, 0.4, and 0.55, respectively. Nonetheless, it should be emphasized that the TP and non-TP shown here are only two examples chosen among many trajectories that we simulated. Our simulation data showed that the development of contact pattern along individual TPs can be very different. The development of contact pattern along individual non-TPs can differ a lot as well. For instance, although on average, the $P_c(\alpha)$ of TPs is lower than the $P_c(\alpha)$ of non-TPs at Q = 0.3 (*A Upper Left*), for the Q = 0.3 TP and non-TP conformations shown here, the former has a higher helical content than the latter (0.636 and 0.545 of native, respectively). Thus, the different behavioral trends of TPs and non-TPs observed in Figs. 3 and 4 and *A* are the results of averaging over ensembles of highly diverse trajectories, a comprehensive analysis of which is beyond the scope of this work. Future effort is needed to gain deeper insight into how TPs differ from non-TPs.

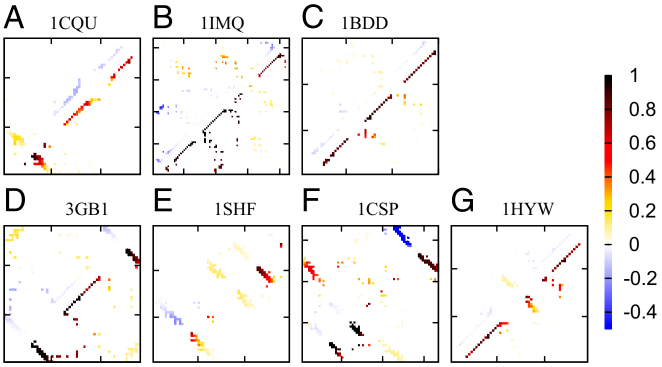

Fig. S5. Comparing TP and FP contact patterns. For each protein, the deviations $P_{\mathrm{TP},ij}(Q) - P_{\mathrm{FP},ij}(Q)$ of the TP native contact probabilities from the FP native contact probabilities are provided by the upper left contact map, whereas the $P_{\mathrm{FP},ij}(Q)$ values are provided by the lower right contact map. As for Fig.4, the contact probability or difference in contact probabilities for residue pair i,j in this figure and Fig. S7 is depicted by a small square at position i,j that is color-coded in accordance with the color scale on the right. Results shown in this figure for each protein are for the Q value at which the largest differences between TPs and FPs for the given protein are exhibited in Fig. 3. The Q values for *A*–*G* are, respectively, 0.3, 0.65, 0.52, 0.62, 0.20, 0.51, and 0.40. Similar to the 2CI2 in Fig. 4, the model proteins studied in this figure show a disfavoring of early formation of local contacts in TPs relative to a typical FP trajectory in the pre-equilibrium. All helical contact patterns are disfavored along TPs. Moreover, the low-CO β-structure near the C terminus of 1CSP involving residues 46–65 is clearly disfavored along the TPs of this model proteins. Nonetheless, more subtle behaviors are also observed with some low-CO β-structures disfavored while other low-CO β-structures are favored by TPs (e.g., 1SHF).

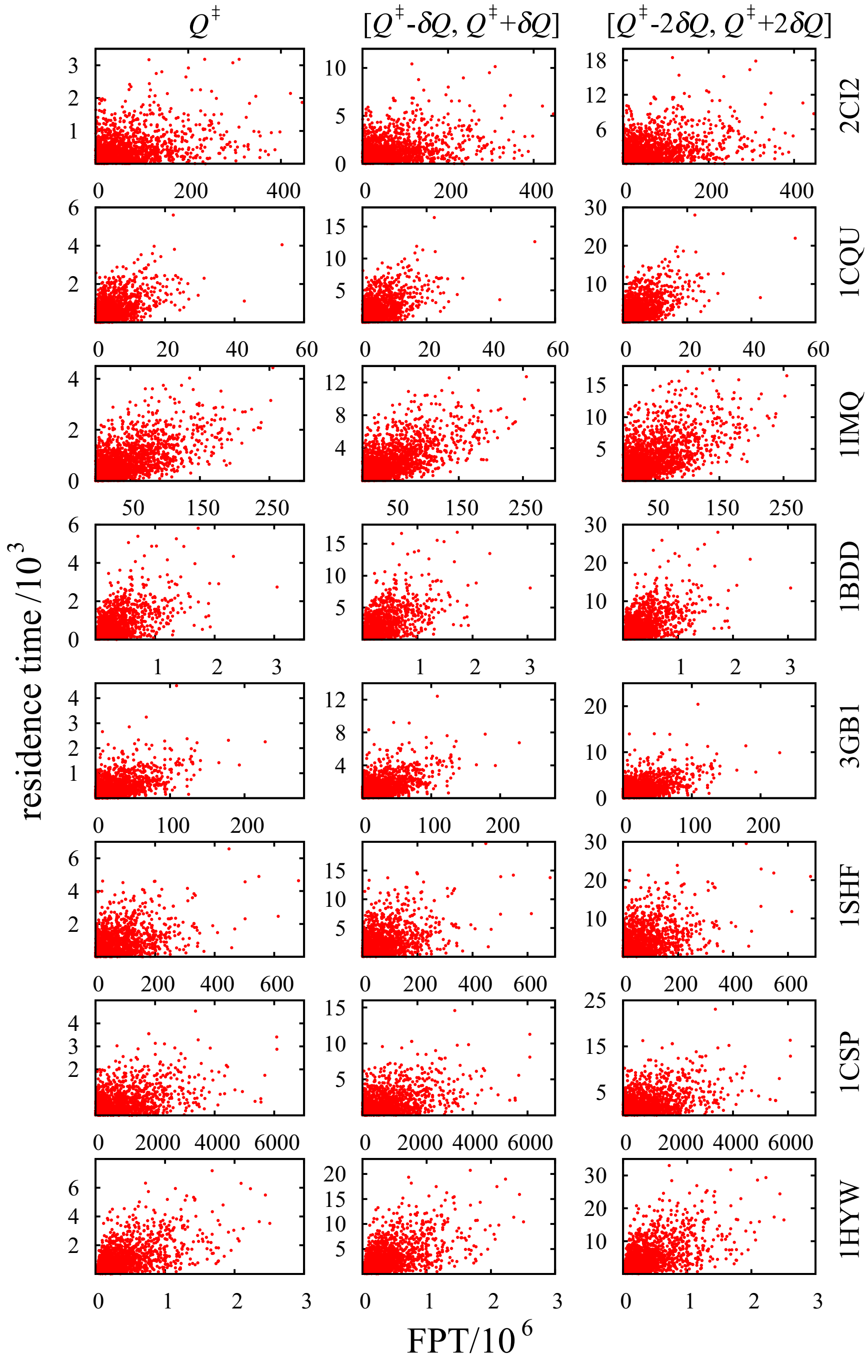

Fig. S6. Scatter plot of FPT with residence time $t(Q^{\ddagger})$ of folding trajectories (FPs) at $Q^{\ddagger}$ (*Left*), residence time at $Q^{\ddagger}$ and its two neighboring Q values [i.e., $t(Q^{\ddagger} - \delta Q) + t(Q^{\ddagger}) + t(Q^{\ddagger} + \delta Q)$; *Center*], and residence time at $Q^{\ddagger}$ and its four neighboring Q values [i.e., $t(Q^{\ddagger} - 2\delta Q) + t(Q^{\ddagger} - \delta Q) + t(Q^{\ddagger}) + t(Q^{\ddagger} + \delta Q) + t(Q^{\ddagger} + 2\delta Q)$; *Right*]. The distributions are essentially random. For each of the model proteins studied, the SD of $t(Q^{\ddagger})$ is approximately equal to the mean value ‹$t(Q^{\ddagger})$›.

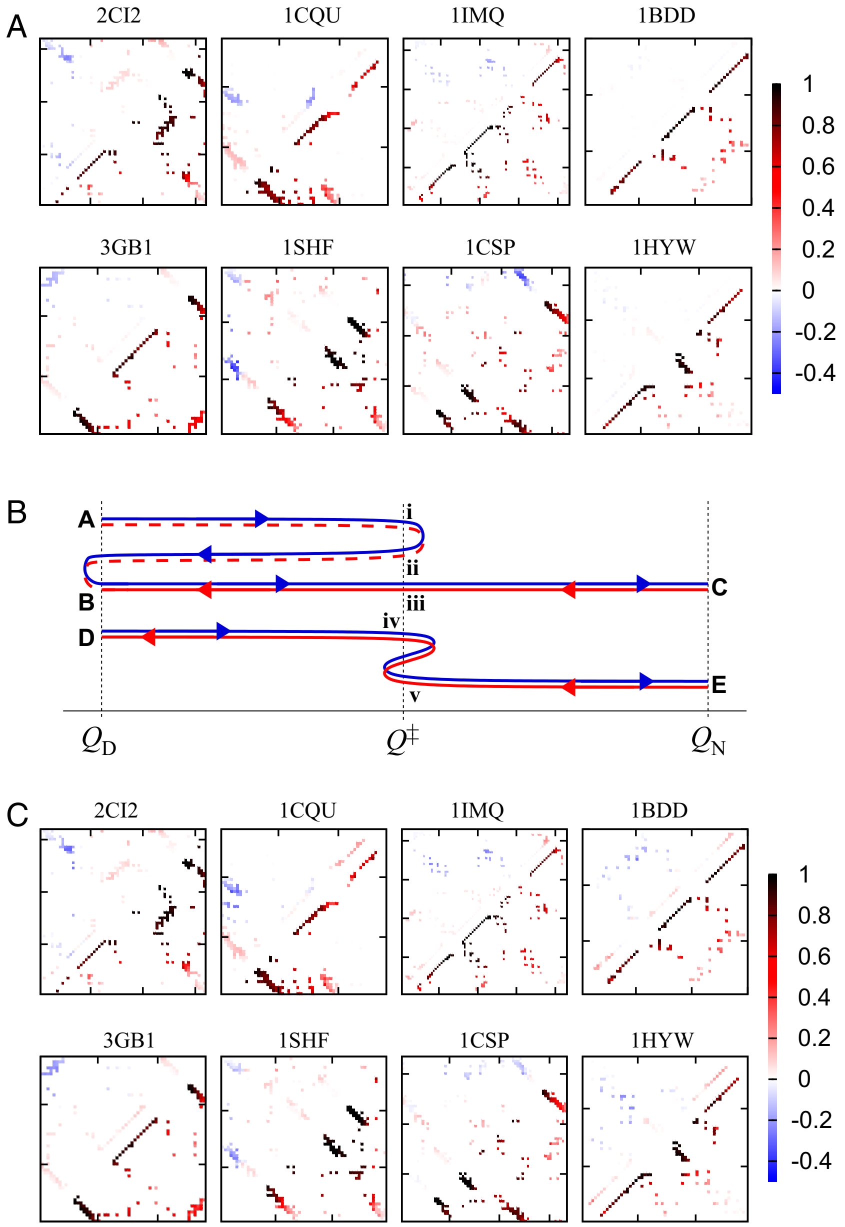

Fig. S7. Symmetry and asymmetry between folding and unfolding kinetics. (*A*) FP and thermodynamic contact patterns can differ at the folding barrier. For each protein, the upper left contact map shows the differences, $P_{\mathrm{FP},ij}(Q^\ddagger) - P_{\mathrm{eq},ij}(Q^\ddagger)$, in native contact probabilities sampled by folding trajectories and equilibrium folding/unfolding trajectories at the thermodynamic free energy barrier $Q^\ddagger$. The lower right map shows the $P_{\mathrm{eq},ij}(Q^\ddagger)$ values obtained by equilibrium sampling of >1,000 conformations at $Q^\ddagger$ for each protein. The mean square deviation $\sum_{i,j}[P_{\mathrm{FP},ij}(Q^\ddagger)-P_{\mathrm{eq},ij}(Q^\ddagger)]^2/\tilde{Q}_n = 0.0109,\ 0.0115,\ 0.0058,$ $0.0001,\ 0.0054,\ 0.0236,\ 0.0209,$ and 0.0003, respectively, for 2CI2, 1CQU, 1IMQ, 1BDD, 3GB1, 1SHF, 1CSP, and 1HYW. (*B*) Schematics of folding and unfolding trajectories. The folding trajectories are shown in blue, and their exact reverse trajectories are shown in red, with a small offset for clarity. A-B-C (blue) is an example folding path that starts at Q_D and ends at Q_N. It passes through $Q^\ddagger$ three times (*i*, *ii*, and *iii*). Time reversal symmetry implies that the reverse trajectory C-B-A exists with equal probability. However, only the C-B part of this trajectory (solid red curve) that passes through $Q^\ddagger$ one time (at *iii*) contributes to an unfolding path, because Q_D is reached by the red path at B. In other words, the dotted red curve from B to A does not contribute to an unfolding path. It follows that the conformations sampled by folding and unfolding paths at $Q^\ddagger$ can be different. D-E is an example folding transition path. Its exact reverse, E-D, is an unfolding transition path. This example shows that the conformations sampled by the folding and unfolding transition paths at their respective last visit of $Q^\ddagger$ can be different (*v* for folding and *iv* for unfolding). This possibility is illustrated in *C*. For each model protein in *C*, the upper left map shows the difference in contact probabilities between simulated folding and unfolding trajectories when the model protein last visited $Q^\ddagger$ (peak of the equilibrium free energy profile) before it is fully folded (for folding trajectories) or fully unfolded (for unfolding trajectories). The plotted differences in the upper left maps are such contact probabilities of the folding trajectories minus the corresponding contact probabilities of the unfolding trajectories, whereas the sums of these folding and unfolding contact probabilities are provided by the lower right maps.

Table S1. Proteins modeled in the present study

Protein	Protein Data Bank ID code (residues range)	n	$\tilde{Q}_n$	No. of folding trajectories	$\langle t(Q^{\ddagger})\rangle$
Chymotrypsin inhibitor 2	2CI2 (20–83)	64	131	1,451	498.8
N-terminal domain of the ribosomal protein L9	1CQU (1–56)	56	107	1,600	662.0
Colicin E9 immunity protein IM9	1IMQ (1–86)	86	164	2,020	755.3
B domain of protein A	1BDD (1–60)	60	84	1,200	889.1
B1 domain of protein G	3GB1 (1–56)	56	103	1,600	444.4
Fyn SH3 domain	1SHF (84–142)	59	129	1,440	893.7
Cold shock protein	1CSP (1–67)	67	141	1,404	562.4
Bacteriophage λ-protein W	1HYW (1–58)	58	96	1,500	1,125.1

For each protein, the number of residues n, the number of native contacts $\tilde{Q}_n$, the number of folding trajectories (FPs) simulated using our explicit-chain model, and the average residence time $\langle t(Q^{\ddagger})\rangle$ among the FPs at the putative transition state are tabulated [$Q^{\ddagger}$ is the peak of $-\ln P_{eq}(Q)$].

Table S2. Barrier heights and folding times in the explicit-chain Langevin dynamics model and the nonexplicit-chain Metropolis MC model with an effective coordinate-independent diffusion coefficient D_0

Protein Data Bank ID code	$\Delta G^{\ddagger}/k_BT$	MFPT	$(\mathrm{MFPT})_{D0}$	$\langle t_{TP}\rangle$	$\langle t_{TP}\rangle_{D0}$	$\langle t(Q^{\ddagger})\rangle_{D0}$	$\mathrm{MFPT}/(\mathrm{MFPT})_{D0}$	$\langle t_{TP}\rangle/\langle t_{TP}\rangle_{D0}$	$\langle t(Q^{\ddagger})\rangle/\langle t(Q^{\ddagger})\rangle_{D0}$
2CI2	7.75	7.4048e7	2.172e6	6.47e4	1,649.4	18.1	34.0921	39.2264	27.558
1CQU	5.25	4.7539e6	1.855e5	2.62e4	837.9	30.9	25.6275	31.2686	21.4239
1IMQ	6.82	4.3563e7	9.194e5	5.68e4	1,761.3	22.2	47.382	32.2489	34.0225
1BDD	2.40	3.8971e5	8.261e3	1.83e4	471.3	25.1	47.1747	38.8288	35.4223
3GB1	6.70	2.6970e7	7.829e5	3.57e4	940.2	21.0	34.4488	37.9706	21.1619
1SHF	8.10	8.0551e7	2.323e6	8.79e4	1,840.9	21.7	34.6754	47.7484	41.1843
1CSP	10.27	9.0674e8	2.346e7	7.84e4	1,695.6	17.0	38.6505	46.2373	33.0824
1HYW	2.07	4.1964e5	1.059e4	3.10e4	867.1	30.8	39.6261	35.7514	36.5292

Times are given, respectively, in Langevin time steps and number of attempted MC moves. Note that the average residence time $\langle t(Q^{\ddagger})\rangle$ among the FPs at $Q^{\ddagger}$ in the explicit-chain model is provided in Table S1; $\langle t(Q^{\ddagger})\rangle_{D0}$ is the corresponding average residence time in the nonexplicit-chain Metropolis MC model. Otherwise, the notation is the same as the notation in the text. The ratios listed in columns 8–10 compare the time scales in the two classes of models. The ratios are quite stable (ranging from 21.2 to 47.7) in that they do not exhibit significant variations among different model proteins. The $\mathrm{MFPT}/(\mathrm{MFPT})_{D0}$, $\langle t_{TP}\rangle/\langle t_{TP}\rangle_{D0}$, and $\langle t(Q^{\ddagger})\rangle/\langle t(Q^{\ddagger})\rangle_{D0}$ ratios averaged over the eight model proteins are, respectively, 37.7, 38.7, and 31.3. These average ratios may be used as conversion factors between the Langevin and Metropolis MC time units in the present study. To facilitate comparison between the average transition path times obtained from the two classes of models, the $\langle t_{TP}\rangle_{D0}$ values plotted in Fig. 2*B* are scaled by a factor that is equal to the average value of $\langle t_{TP}\rangle/\langle t_{TP}\rangle_{D0} \times 10^{-4}$. In other words, each of the $\langle t_{TP}\rangle_{D0}$ values plotted in Fig. 2*B* is 0.00387 times the corresponding $\langle t_{TP}\rangle_{D0}$ value in this table.