Coordinate-dependent diffusion in protein folding

Robert B. Best[a,1], **and Gerhard Hummer**[b,1]

[a]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom [b]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

Edited by Peter G. Wolynes, University of California, San Diego, La Jolla, CA, and approved November 24, 2009 (received for review September 10, 2009)

Diffusion on a low-dimensional free-energy surface is a remarkably successful model for the folding dynamics of small single-domain proteins. Complicating the interpretation of both simulations and experiments is the expectation that the effective diffusion coefficient *D* will in general depend on the position along the folding coordinate, and this dependence may vary for different coordinates. Here we explore the position dependence of *D*, its connection to protein internal friction, and the consequences for the interpretation of single-molecule experiments. We find a large decrease in *D* from unfolded to folded, for reaction coordinates that directly measure fluctuations in Cartesian configuration space, including those probed in single-molecule experiments. In contrast, *D* is almost independent of *Q*, the fraction of native amino acid contacts: Near the folded state, small fluctuations in position cause large fluctuations in *Q*, and vice versa for the unfolded state. In general, position-dependent free energies and diffusion coefficients for any two good reaction coordinates that separate reactant, product, and transition states, are related by a simple transformation, as we demonstrate. With this transformation, we obtain reaction coordinates with position-invariant *D*. The corresponding free-energy surfaces allow us to justify the assumptions used in estimating the speed limit for protein folding from experimental measurements of the reconfiguration time in the unfolded state, and also reveal intermediates hidden in the original free-energy projection. Lastly, we comment on the design of future single-molecule experiments that probe the position dependence of *D* directly.

coordinate transformation | internal friction | kinetic prefactor | projected dynamics | reaction coordinate

One of the remarkable consequences of the energy landscape theory of protein folding is that folding can be effectively modeled as diffusion along a one-dimensional reaction coordinate (1–5). In such a diffusion model, the rate of folding is determined by the height of the free-energy barrier and a kinetic prefactor that depends on the diffusion coefficient along the reaction coordinate. For proteins with high free-energy barriers separating the folded and unfolded states, only the diffusion coefficient at the barrier top contributes to the rate. For ultrafast folding proteins, where the barrier heights can be quite small or vanish completely, the diffusion coefficient along the entire reaction coordinate contributes to the rate (6–11). The prefactor therefore also approximates the "speed limit" for protein folding, analogous to the diffusion-limited rate for bimolecular reactions (7, 8, 12, 13). A number of groups have investigated the prefactor experimentally by varying solvent viscosity (14–18). However, viscogens tend to affect the rate not only by increasing the solvent friction, but also by changing the free-energy barrier height and curvature, requiring measurements at viscogen concentrations that alter neither the stability nor the viscosity-corrected activation energy (11, 19).

A more direct method of obtaining diffusion coefficients for folding is to probe the dynamics of polypeptides or unfolded proteins, for example by using ultrafast spectroscopy to measure rates of contact formation between specific residues (12, 20–26), or FRET cross-correlation functions to probe the reconfiguration time in the unfolded state (27, 28). Determination of the diffusion coefficient in this class of experiments requires that the end-to-end distance distribution be known. Moreover, to be relevant to folding, it must be assumed that the diffusion coefficient in the denatured well and at the barrier top are the same (29). Interestingly, some of these studies (24, 27, 28) have shown that dilution of denaturant results in a slowing of diffusion, presumably as a result of increased internal friction.

Even if a projection of the true high-dimensional dynamics of folding onto a single coordinate is successful in the sense that the projected dynamics is diffusive without memory effects, the projection will in general lead to a dependence of the diffusion coefficient on position along that coordinate (1). In particular, any continuous one-to-one transformation that locally stretches or compresses the reaction coordinate results in an equally good coordinate, but with a different local diffusion coefficient and free-energy surface. This position dependence of diffusion coefficients is not just a theoretical curiosity: The interpretation of experiments is often based on a position-invariant diffusion coefficient. Although the position dependence was neglected for simplicity in early calculations of the one-dimensional diffusion coefficient from simulation (3), position-dependent diffusion coefficients have been estimated from simulation for model systems (30, 31) and more recently for folding (10, 32–34). One of the difficulties in such calculations is the separation of the effects arising from the local diffusion and from curvature in the free-energy surface (31, 33). For diffusive dynamics initiated at position Q_i on the reaction coordinate Q, the diffusion coefficient $D(Q_i)$ is related to the variance $\sigma^2(\Delta t)$ of Q after a short time Δt through $\sigma^2(\Delta t) \approx 2D(Q_i)\Delta t$. However, this approximation is only valid at short times Δt because it ignores the effects of curvature in the free-energy surface, a difficulty that becomes particularly pronounced if the dynamics becomes diffusive only at long times Δt (32).

The position dependence of D reported in different simulation studies has been quite varied, with some finding a maximum in the diffusion coefficient at intermediate values of the reaction coordinate near the barrier top (32–34), and others finding a monotonic decrease towards the folded state (10). The origin of the differences is not immediately clear, and one should probably not expect any general trend in the position dependence of the diffusion coefficient. Whereas the global dynamics of the chain in Cartesian space is expected to slow down for increasingly compact states, one often has little intuition for projections of the high-dimensional protein configuration space onto reaction coordinates that are nonlinear functions of the atomic positions.

Here, we use coarse-grained molecular simulations of a pair of experimentally well-characterized two-state proteins, prb_{7-53} (35) and protein G (36), to investigate the position dependence of the diffusion coefficient along different reaction coordinates. We find

Author contributions: R.B.B. and G.H. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: rbb24@cam.ac.uk or hummer@helix.nih.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0910390107/DCSupplemental.

that diffusion coefficients for coordinates that directly probe fluctuations in Cartesian space, such as pair distances, are markedly position dependent. In contrast, we find that a projection of the dynamics onto the commonly used folding coordinate Q results in only a small position dependence in $D(Q)$. This lack of position dependence can be explained by the compression of large fluctuations in Euclidian distance in the unfolded state into small fluctuations in Q. Moreover, we show that if the dynamics is well-described by alternative one-dimensional reaction coordinates, the respective position-dependent diffusion coefficients and free energies can be mapped onto each other by a simple transformation. We use this same transformation to construct a coordinate s along which the diffusion coefficient $D(s)$ is position invariant. We find that the resulting free-energy surface $F(s)$ justifies the assumptions used to calculate the prefactor for protein folding rates from the reconfiguration time in the unfolded state measured in single-molecule experiments. Intriguingly, we find that $F(s)$ also reveals intermediates in the local free-energy profile that are hidden in the projection onto Q. We dissect the contributions to the "internal friction" for our model and find that the largest contribution from internal friction to $D(Q)$ near the barrier comes from the formation and breaking of amino acid contacts, rather than the crossing of torsional barriers. Nonnative interactions reduce the diffusion coefficient near the barrier. This effect compensates the reduction in barrier height due to these interactions, such that the folding rate is almost unchanged from that of the Gō model.

Results

Dependence of D on Q. If the projection of the dynamics onto a one-dimensional reaction coordinate can be described as diffusive, the effective diffusion coefficient for motion along that coordinate will in general be position dependent. As a reaction coordinate for folding, we initially consider the fraction of native contacts, Q (37), which has a low value when the protein is unfolded and a high value (near unity) for the fully folded state. Q has been shown to be a good reaction coordinate for folding in several cases (32, 38, 39). We describe the projected dynamics on a one-dimensional reaction coordinate (here, Q) as diffusive, using the Smoluchowski equation, with free-energy surface $F(Q)$ and position-dependent diffusion coefficient $D(Q)$, $\partial_t p(Q,t) = \partial_Q\{D(Q)\exp(-\beta F(Q))\partial_Q[\exp(\beta F(Q))p(Q,t)]\}$ where $p(Q,t)$ is the time-dependent probability density along Q.

Fig. 1*A* and *B* show typical two-state folding trajectories $Q(t)$ for the two proteins we have studied, the three-helix bundle, prb_{7-53} (35) (hereafter referred to as prb) and protein G (36). These proteins were chosen to be representative of two different structural classes of single-domain proteins (α and α/β; the structures are shown at the top of Fig. 1). We have used a standard type of "Gō-like" polymer model to study the folding dynamics (40). As expected from the Q trajectories, the free-energy surface $F(Q)$ (Fig. 1*C* and *D*) has two minima, located at low and high Q (corresponding respectively to the unfolded and folded protein), separated by a single free-energy barrier.

Q turns out to be a good reaction coordinate for the folding of both proteins. To assess the ability of Q to capture the ensemble of folding transition states, we use a Bayesian criterion (41) and calculate the probability $p(\text{TP}|Q)$ of being on a transition path, given a particular value of Q (38). $p(\text{TP}|Q)$ reaches approximately 0.4 and approximately 0.45 for prb and protein G, respectively, close to the theoretical maximum of 0.5 for diffusive dynamics, indicating that the value of Q is predictive of transition states (38, 41).

The position-dependent diffusion coefficients $D(Q)$ are shown in Fig. 1*E* and *F*. A number of features are worth noting. First, $D(Q)$ varies by only a factor 2–3 over the range of the coordinate, counter to our expectation of a strong slowdown with increasing nativeness (10). This effect is small, given that the folding rates depend only linearly on the diffusion coefficient, but exponentially on the barrier height. Second, although $D(Q)$ is slightly lower for the folded protein than the unfolded one, in agreement with our expectations, the diffusion coefficient also has a maximum near the barrier.

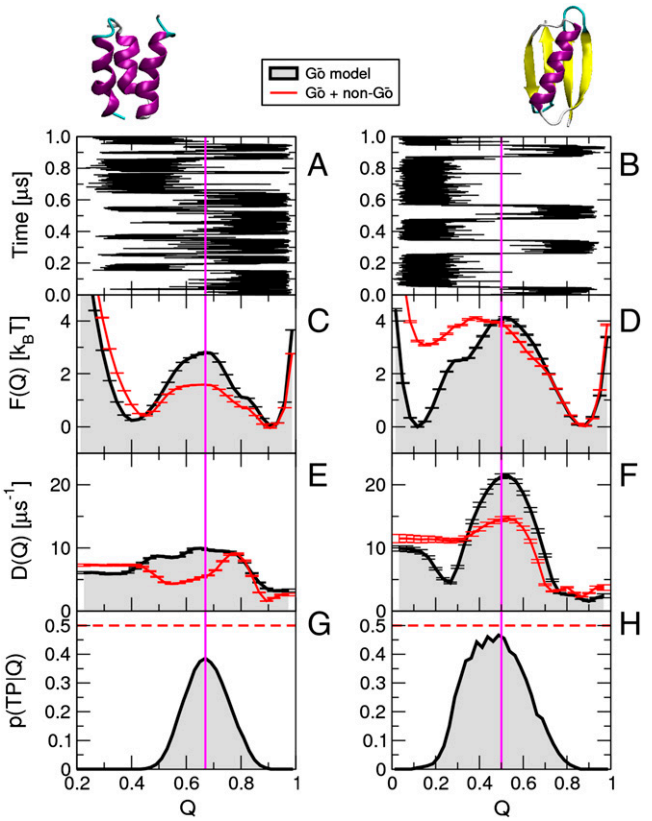

Fig. 1. Fraction of native contacts, Q, as reaction coordinate for folding of prb_{7-53} (*Left*) and protein G (*Right*). (*A* and *B*) Trajectory segments $Q(t)$ for Gō models. (*C* and *D*) Free-energy surfaces $F(Q)$ for Gō models (*Black*) and with nonnative attractions (*Red*). (*E* and *F*) Corresponding position-dependent diffusion coefficients $D(Q)$. (*G* and *H*) Conditional probability $p(\text{TP}|Q)$ of being on a transition path, with the dashed line indicating the theoretical maximum. The folded structures of each protein are shown at the top of the figure.

Alternative Reaction Coordinates. Experimental measurements of diffusion coefficients for folding usually probe a coordinate in Cartesian space, such as the distance between a pair of residues (e.g., in a FRET experiment). Using such a coordinate, Nettels et al. (27) recently found a fivefold change in diffusion coefficient for unfolded proteins on diluting denaturant from 8 M to 1 M GdmCl and Möglich et al. (24) obtained a fourfold change in diffusion coefficient between 8 M and 0 M GdmCl for $(\text{Gly-Ser})_{16}$. We find a similarly large decrease (>10-fold) in diffusion coefficient going from unfolded to folded, if only pair distances are considered (Fig. S1). In fact, this result explains why diffusion coefficients obtained from kinetic models of protein unfolding in pulling experiments (42), which are more sensitive to the native and transition state regions of the free-energy surface, are much lower than those inferred from measurements on unfolded proteins (12, 20–27).

However, any single pair distance is extremely unlikely to be a good folding coordinate. In particular, it would be difficult to even resolve the unfolded and folded states using the instantaneous value of such a coordinate (43, 44). Even though for any coordinate r the decay of the correlation function $\langle r(t)r(0)\rangle$ at long times t reflects the slowest relaxing modes, for poor coordinates the amplitude of the slow decay will be small. To avoid

these difficulties with single pair distances, we use d_{RMS} (45) as an alternative coordinate, defined here as the rms difference of distances r_{ij} between the N native residue pairs (i,j) and the reference distances in the native structure r_{ij}^0, $d_{RMS} = [N^{-1}\sum_{(i,j)}(r_{ij} - r_{ij}^0)^2]^{1/2}$. This coordinate better preserves the magnitude of fluctuations in Cartesian space and is approximately related to the radius of gyration R_g (far from the native structure $d_{RMS} \approx 2R_g$). However, unlike R_g, it is also a reasonably good reaction coordinate near the native state (particularly for protein G), in that it can identify transition states, as illustrated by the distribution of $p(\mathrm{TP}|d_{RMS})$ (Fig. 2, insets) (41). We note, however, that the maximum of $p(\mathrm{TP}|d_{RMS})$ is at $d_{RMS} \approx 3$ Å, whereas the free-energy barrier is at approximately 5 Å. This shift in the transition state relative to the barrier is a consequence of the strong position dependence of the diffusion coefficient $D(d_{RMS})$ in the barrier region (Fig. 2*C*) (10).

The potential of mean force on d_{RMS} looks quite different from the familiar two-state free-energy surfaces on Q, with a narrow folded minimum and broad unfolded one. This structure reflects the larger fluctuations in the unfolded state when measured in Cartesian space. The trend in the position-dependent diffusion coefficient is also more in line with expectations from experiment (27) and intuition. There is a monotonic decrease in the diffusion coefficient $D(d_{RMS})$ toward the folded state and a variation of around two orders of magnitude, in contrast to the nearly constant $D(Q)$.

The apparent differences in position dependence between Q and d_{RMS} can be understood from the relationship between the two coordinates. In Fig. 3, d_{RMS} is plotted against Q for each protein. The unfolded state spans a wide range of d_{RMS}, yet this is compressed into only small fluctuations in Q, whereas for the folded state the reverse is true. For the unfolded protein, large changes in conformation can occur with a change of only a few contacts, but even a small conformational change near the folded state will cause a large shift in Q.

As a second alternative coordinate, we consider the number of native torsion angles, $n = N_\phi^{-1}\sum_i \exp(-(\phi_i - \phi_i^0)^2/\sigma^2)$, where the sum runs over the N_ϕ torsion angles ϕ_i with native values ϕ_i^0 and $\sigma = 60°$. This coordinate resembles the coordinate used in Ising-like folding models, namely the number of "correct" or ordered residues (46–49). Using the Bayesian criterion $p(\mathrm{TP}|n)$, n is found to be a good reaction coordinate for protein G and a reasonable one for prb (Fig. 4, insets). Like Q, n compresses the fluctuations in the unfolded state because single torsional transitions will generally be associated with larger Cartesian displacements for the unfolded protein. But in contrast to Q, n also compresses the transition region (Fig. 3*C* and *D*). As a consequence, the diffusion coefficient $D(n)$ at the barrier is substantially lower than in the unfolded minimum.

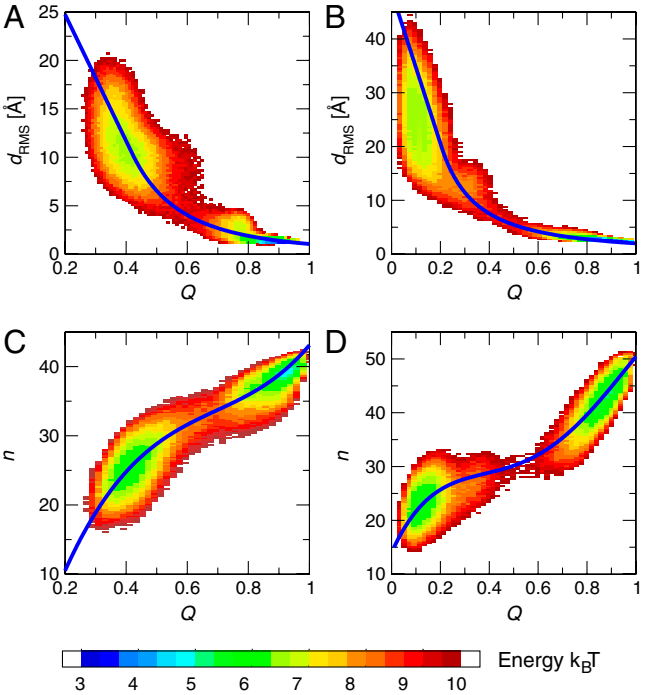

Fig. 3. Relation between different reaction coordinates. Two-dimensional free-energy surfaces from long equilibrium simulations are shown as functions of d_{RMS} and Q for (*A*) prb_{7-53} and (*B*) protein G, and for the number of native torsion angles against Q for (*C*) prb_{7-53} and (*D*) protein G. Continuous curves represent the coordinate transformations $s = s(r)$ discussed in the text.

General Relation Between Diffusion on Different Coordinates. The results of the previous section suggest that the appearance of the diffusion coefficient along different coordinates is altered simply by "stretching" or "shrinking" one coordinate relative to another (see Fig. 3), and that there is nothing special about the trend in position-dependent free energies and diffusion coefficients observed along a particular coordinate. In fact, given

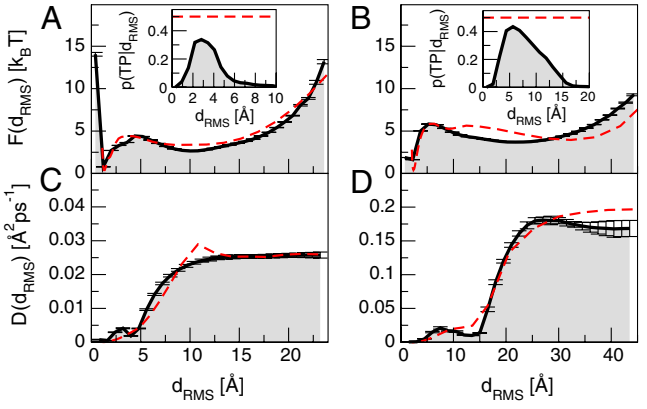

Fig. 2. Distance-matrix rms deviation d_{RMS} as reaction coordinate. (*A* and *B*) free energy $F(d_{RMS})$; (*C* and *D*) position-dependent diffusion coefficients $D(d_{RMS})$; and (*Insets*) the conditional probability of being on a transition path, $p(\mathrm{TP}|d_{RMS})$. The left and right columns refer to prb_{7-53} and protein G, respectively. Broken red lines are obtained by mapping the free energies and diffusion coefficients along Q onto d_{RMS}.

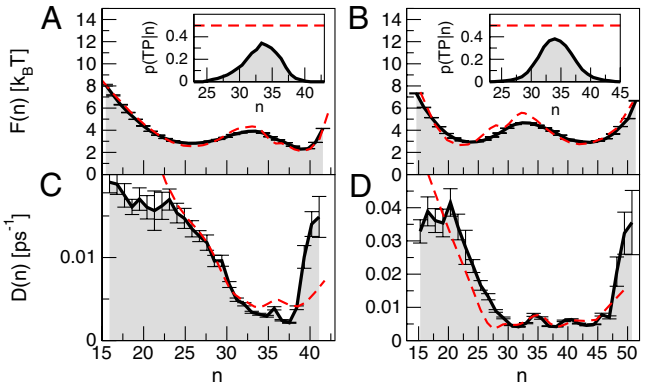

Fig. 4. Number of native torsion angles n as reaction coordinate. (*A* and *B*) free energy $F(n)$; (*C* and *D*) position-dependent diffusion coefficients $D(n)$; and (*Insets*) conditional probability of being on a transition path, $p(\mathrm{TP}|n)$. The left and right columns refer to prb_{7-53} and protein G, respectively. Broken red lines are obtained by mapping the free energies and diffusion coefficients along Q onto n.

two coordinates on which the projected dynamics is well-described as one-dimensional diffusion, it should be possible to define a monotonic map from one coordinate to the other. For two coordinates r and $s = s(r)$ (for which the inverse $r = r(s)$ exists), the position-dependent free energies $F(s)$ and diffusion coefficients $D(s)$ for s can be obtained from those for r via the transformations $\beta F(s) = \beta F(r) + \ln|ds/dr|$ and $D(s) = D(r)(ds/dr)^2$, respectively, with $|ds/dr|$ the Jacobian of the variable change. We note that the map $s(r) = \int^r [D_0/D(r)]^{1/2} dr$, used by Rhee and Pande (50), results in a constant diffusion coefficient $D(s) = D_0$; an alternative approach that results in position-invariant D has been described by Krivov and Karplus (51). Note that the accuracy of the transformation to $s = s(r)$ depends on the dynamics being well-described as one-dimensional diffusion along both coordinates. We have used these transformations to map the free energies $F(Q)$ and diffusion coefficients $D(Q)$ from Q onto the other two coordinates. Smooth, monotonic transformations $d_{\mathrm{RMS}}(Q)$ and $n(Q)$ were defined by fits to the data in Fig. 3. The functional forms used were a quartic polynomial for $n(Q)$, and for $d_{\mathrm{RMS}}(Q)$, a continuous spline consisting of a power law fit at high Q and a straight line at low Q. The resulting free energies and diffusion coefficients, shown by broken red lines in Figs. 2 and 4 for $D(d_{\mathrm{RMS}})$ and $D(n)$, respectively, are remarkably close to the direct estimates, given that the individual coordinates are imperfect.

Origin of the Position Dependence of $D(Q)$. The main contribution to the diffusion coefficient is the crossing of local barriers on the energy landscape (32). We have analyzed more closely the barriers traversed at each position on the energy landscape for the three-helix bundle prb by initiating short trajectories from structures at $Q = 0.4$, 0.9, and 0.7 corresponding to the unfolded, folded, and transition states respectively. Barrier crossing is probed by "quenching" each step of the trajectory by extensive local minimization, after which traversals of microscopic barriers on the energy landscape are identified with a nonzero rms deviation between adjacent frames in the trajectory (Fig. S2) (52). The changes associated with each transition can be clearly identified by calculating the number of native contacts and native dihedrals for the quenched trajectories. While in the unfolded state, there is frequent barrier crossing and most transitions are only associated with the formation/breaking of single contacts, or no contacts (average change of 0.69 contacts per transition). In the folded state and transition state, barrier crossings generally cause multiple contacts to be formed or broken, with an average change in total contacts of 3.27 and 2.19 per transition, respectively (Fig. S2). Thus the slowdown in the rate of local barrier crossings as the protein becomes more compact and native-like is compensated by an increase in the number of contacts formed or lost in each event. The imperfect cancellation of these effects gives rise to the peak in $D(Q)$ near the barrier.

Nonnative Contacts, Roughness, and Internal Friction. Nonnative attractions are expected to affect the roughness of the energy landscape (53–56). To explore the effects of nonnative attractive interactions on the folding dynamics, we add recently published amino acid pair potentials developed for protein–protein interactions (57) on top of the original Gō-like model for native interactions (see *SI Text*). These non-Gō interactions slightly stabilize the folded structures of both proteins (Fig. 1). Whereas this stabilization seems counterintuitive (non-Gō contacts might be expected to favor the unfolded state), it can be explained by a combination of the gains from the relatively long-ranged non-Gō potential and a partial elimination of slightly destabilizing nonnative interactions in the original Gō model (see Fig. S3 and *SI Text*).

Nonnative interactions affect the diffusion coefficient predominantly in the barrier region where $D(Q)$ is reduced (Fig. 1); they have less effect beyond the barrier near the folded state, where most residues are already native-like and it would be harder to form nonnative interactions. This slowdown of the dynamics along Q on the barrier top may arise from the increased "frustration" involved in the correct selection of native contacts. However, the folding rate is affected by both $D(Q)$ and the barrier height. For both proteins, we find that nonnative attractions lower the barrier height, in agreement with expectations from theory (58) and previous simulations (59). Remarkably, for prb, the reduced $D(Q)$ at the barrier is sufficient to approximately cancel the rate enhancement that would result from the reduction in barrier height alone. The folding rate for the plain Gō-like model is 25.8(2.0) μs^{-1}. The barrier is lowered by approximately $1.7 k_B T$, which would be expected to give an approximately fivefold increase in rate if the diffusion coefficient were unchanged. However, the decrease in diffusion coefficient is sufficient that the rate increases only by a factor of approximately 1.3 to 33.1(2.8) μs^{-1}. The strong effect of nonnative interactions on D is thus masked in the apparent rate.

We have explored other factors that contribute to the roughness of the energy landscape and affect the dynamics at the folding barrier (53). Specifically, we lowered the energy barrier for torsional transitions and found an increase in the diffusion coefficient in the folded minimum, but only a slight enhancement at the barrier and in the unfolded state (Fig. S4). In addition, we explored the role of contact formation and breaking. Elimination of the desolvation barrier in the attractive contact potentials increases the diffusion coefficient both in the folded state and at the barrier (Fig. S4), suggesting that for our model contact formation is more important for the dynamics of folding than the crossing of torsional barriers.

If we use $D(Q)$ as a measure of the internal friction along the folding coordinate (11), we find that nonnative attractive interactions increase the internal friction in the relevant barrier region as well as barriers for native contact formation and breaking, with only a modest effect from torsion barriers.

Concluding Remarks

Several other studies have estimated position-dependent diffusion coefficients from simulations of folding. Using a Gō model, Yang et al. (33) found a maximum in $D(Q)$ for intermediate values of Q and similar results were obtained using a related coordinate for the folding of a model of a β-hairpin in explicit water (34). This qualitative agreement suggests that the form for $D(Q)$ found here is not limited to our specific model. We explain the weak dependence of $D(Q)$ on Q and the maximum near the barrier as resulting from the trade-off between the slowdown of the dynamics with compaction and an increase in the magnitude of Q fluctuations for more folded species. However, another study found a strong position dependence of $D(Q)$, with a monotonic decrease by an order of magnitude going from unfolded to folded (10). Possible reasons for this form of $D(Q)$ are the choice of Monte Carlo move set (60) (see Fig. S5), or the inclusion of non-Gō interactions. Another possible reason is that for a lattice model, the curvature of the folded basin in the free-energy profile is substantially greater than that of the unfolded basin (10). A variable transformation stretching Q near the folded basin to match the curvatures of the two free-energy minima would increase the diffusion coefficient for compact configurations. The resulting position dependence of both the local free energies and diffusion coefficients would then be more akin to those obtained with off-lattice Gō models (32, 33).

The position dependence of the diffusion coefficient is also important to experiments. Inspired by the results from energy landscape theory, experimental folding kinetics are often interpreted using the one-dimensional diffusive model in (overdamped) Kramers theory (61). This model gives the rate for escape from a one-dimensional potential well as $k = k_0 \exp(-\Delta/k_B T)$, where

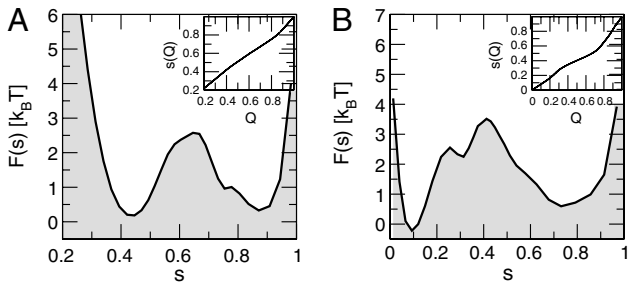

Fig. 5. Free-energy surfaces $F(s)$ for the coordinate s on which the diffusion coefficient $D(s)$ is position-independent: (*A*) prb_{7-53}, and (*B*) protein G. By identifying ω_u^2 and ω_b^2 with $|F''(s)|$ in the unfolded minimum and at the barrier top respectively, we obtain for prb_{7-53} $\omega_u = 17.2$, $\omega_b = 15.8$ $(k_BT)^{1/2}$, and for protein G $\omega_u = 25.1$, $\omega_b = 16.4$ $(k_BT)^{1/2}$.

$k_0 = \omega_u\omega_b D_b/(2\pi k_B T)$ is the kinetic prefactor with ω_u^2 and ω_b^2 the curvatures of the free-energy surface in the unfolded well and barrier regions, respectively, and Δ the barrier height. The theory does not explicitly consider a position-dependent diffusion coefficient, but the value of D_b in the rate expression is that near the barrier top. The kinetic prefactor k_0 is important because it sets the speed limit for folding (6–8, 12) and allows one to calculate free-energy barrier heights from folding rates. One way to estimate k_0 from experiment is by rewriting the rate expression in terms of the reconfiguration time in the unfolded state, $\tau_u = k_BT/\omega_u^2 D_u$, i.e., $k_0 = \omega_b/2\pi\omega_u\tau_u$ (29), which is correct if the diffusion coefficients in the unfolded state and at the folding barrier agree, $D_u = D_b$. Provided that for the same reaction coordinate $\omega_u \approx \omega_b$, then the prefactor becomes $k_0 \approx 1/2\pi\tau_u$. The reconfiguration time τ_u in the unfolded state can be estimated from FRET correlation functions and the width of FRET efficiency distributions (28, 29) by monitoring coordinates that are not necessarily good folding coordinates (e.g., the end-to-end distance of the protein as in Fig. S1).

Here we can explicitly test the assumptions $D_u \approx D_b$ and $\omega_u \approx \omega_b$, by introducing a new coordinate $s = s(Q)$ for which $D(s)$ is constant such that $D_u = D_b$ by construction $s(Q) = \int_0^Q dQ/\sqrt{D(Q)}/\int_0^1 dQ/\sqrt{D(Q)}$. The similarity of the curvatures of the corresponding free-energy surfaces $F(s)$ (Fig. 5) at the barrier (ω_b^2) and in the unfolded minimum (ω_u^2) suggests that the kinetic prefactor and the height of the free-energy barrier for folding can indeed be obtained consistently from a combination of experiments that probe the unfolded state and the folding rate (see Fig. S6 for a complete comparison between s and Q). It will be important to test this approximation further for larger proteins and more complex models including all-atom detail (62) and explicit solvent (63). An interesting feature of the surfaces $F(s)$ for both prb_{7-53} and protein G in Fig. 5 is the appearance of subsidiary barriers at $Q \approx 0.75$ on the folded side of the main barrier and $Q \approx 0.25$ on the unfolded side, respectively. Because these barriers are concealed in $F(Q)$, they are compensated by a local reduction of $D(Q)$; stretching Q by the transformation $s(Q)$ reveals the presence of the additional barriers in $F(s)$, and of corresponding low-population intermediates.

Experimentally, the most direct method of deriving position-dependent diffusion coefficients for folding would be to monitor some sufficiently collective coordinate r in a single-molecule experiment, e.g., a well-chosen amino acid pair distance monitored by FRET or the quantum yield of the fluorophore which is dependent on solvent exposure. One could either follow the time dependence of such a coordinate in an unrestrained system, e.g., by monitoring FRET efficiencies; or, if feasible, one could harmonically restrain the system at a series of r_i values, e.g., using optical tweezers or an atomic force microscope, while at the same time monitoring its fluctuations. In the former case, $D(r)$ could be obtained from the procedure used here; in the latter case, one can estimate the local diffusion coefficient along r from the ratio of the local variance and correlation time, $D(r_i) \approx \mathrm{Var}_i(r)/\tau_i$ (31). Both procedures would require a relatively high time resolution and low experimental noise.

Materials and Methods

Coarse-grained models of prb and protein G were built from their protein data bank structures [prb: 1prb (64); protein G: 2gb1 (65) using the Karanicolas and Brooks Gō model (40)]. The protein is represented by beads located at the alpha carbon positions, with fixed Cα-Cα "bond" lengths, harmonic angle (Cα-Cα-Cα) potentials, and a knowledge-based Fourier series for the torsional potential. Long Langevin dynamics simulations were run at the folding temperature with a friction coefficient of 0.1 ps^{-1} with Chemistry at HARvard Molecular Mechanics (CHARMM) (66); MC simulations were also run with CHARMM (60). From the trajectory data, diffusion coefficients were determined as described previously (31, 32). Simulation details are provided in *SI Text*.

ACKNOWLEDGMENTS. We thank Dr. William A. Eaton and Prof. Ben Schuler for helpful comments on the manuscript. R.B. is supported by a Royal Society University Research Fellowship. G.H. is supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

1. Bryngelson JD, Wolynes PG (1989) Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J Phys Chem*, 93:6902–6915.
2. Camacho CJ, Thirumalai D (1993) Kinetics and thermodynamics of folding in model proteins. *Proc Natl Acad Sci USA*, 90:6369–6372.
3. Socci ND, Onuchic JN, Wolynes PG (1996) Diffusive dynamics of the reaction coordinate for protein folding funnels. *J Chem Phys*, 104:5860–5868.
4. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol*, 4:10–19.
5. Plotkin SS, Wolynes PG (1998) Non-Markovian congurational diffusion and reaction coordinates for protein folding. *Phys Rev Lett*, 80:5015–5018.
6. Sabelko J, Ervin J, Gruebele M (1999) Observation of strange kinetics in protein folding. *Proc Natl Acad Sci USA*, 96:6031–6036.
7. Yang WY, Gruebele M (2003) Folding at the speed limit. *Nature*, 423:193–197.
8. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding "speed limit". *Curr Opin Struct Biol*, 14:76–88.
9. Sadqi M, Fushman D, Muñoz V (2006) Atom-by-atom analysis of global downhill protein folding. *Nature*, 442:317–321.
10. 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*, 104:14646–14651.
11. Cellmer T, Henry ER, Hofrichter J, Eaton WA (2008) Measuring internal friction in ultra-fast protein folding kinetics. *Proc Natl Acad Sci USA*, 105:18320–18325.
12. Hagen SJ, Hofrichter J, Szabo A, Eaton WA (1996) Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding. *Proc Natl Acad Sci USA*, 93:11615–11617.
13. Ghosh K, Ozkan SB, Dill KA (2007) The ultimate speed limit to protein folding is conformational searching. *J Am Chem Soc*, 129:11920–11927.
14. Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA (1992) The role of solvent viscosity in the dynamics of protein conformational changes. *Science*, 256:1796–1798.
15. Jacob M, Geeves M, Holtermann G, Schmid FX (1999) Diffusional barrier crossing in a two-state protein folding reaction. *Nat Struct Biol*, 6:923–926.
16. Plaxco KW, Baker D (1998) Limited internal friction in the rate-limiting step of a two-state protein folding reaction. *Proc Natl Acad Sci USA*, 95:13591–13596.
17. Qiu L, Hagen SJ (2004) A limiting speed for protein folding at low solvent viscosity. *J Am Chem Soc*, 126:3398–3399.
18. Ladurner AG, Fersht AR (1999) Upper limit of the time scale for diffusion and chain collapse in chymotrypsin inhibitor 2. *Nat Struct Biol*, 6:28–31.
19. Jas GS, Eaton WA, Hofrichter J (2001) Effect of viscosity on the kinetics of α-helix and β-hairpin formation. *J Phys Chem B*, 105:261–272.
20. Bieri O, et al. (1999) The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc Natl Acad Sci USA*, 96:9597–9601.
21. Lapidus LJ, Eaton WA, Hofrichter J (2000) Measuring the rate of intramolecular contact formation in polypeptides. *Proc Natl Acad Sci USA*, 97:7220–7225.
22. Neuweiler H, Schultz A, Böhmer M, Enderlein J, Sauer M (2003) Measurement of submicrosecond intramolecular contact formation in peptides at the single molecule level. *J Am Chem Soc*, 125:5234–5330.
23. Laurence TA, Kong X, Jäger M, Weiss S (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc Natl Acad Sci USA*, 102:17348–17353.
24. Möglich A, Joder K, Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc Natl Acad Sci USA*, 103:12394–12399.
25. Buscaglia M, Lapidus LJ, Eaton WA, Hofrichter J (2006) Effects of denaturants on the dynamics of loop formation in polypeptides. *Biophys J*, 91:276–288.

26. Fierz B, et al. (2007) Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. *Proc Natl Acad Sci USA*, 104:2163–2168.
27. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single molecule photon statistics. *Proc Natl Acad Sci USA*, 104:2655–2660.
28. Nettels D, Hoffmann A, Schuler B (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J Phys Chem B*, 112:6137–6146.
29. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419:743–747.
30. Straub JE, Borkovec M, Berne BJ (1987) Calculation of dynamic friction on intramolecular degrees of freedom. *J Phys Chem*, 91:5722–5729.
31. Hummer G (2005) Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J Phys*, 7:516–523.
32. Best RB, Hummer G (2006) Diffusive model of protein folding dynamics with Kramers turnover in rate. *Phys Rev Lett*, 96:228104.
33. Yang S, Onuchic JN, Levine H (2006) Effective stochastic dynamics on a protein folding energy landscape. *J Chem Phys*, 125:054910.
34. Yang S, Onuchic JN, Garcia AE, Levine H (2007) Folding time predictions from all-atom replica exchange simulations. *J Mol Biol*, 372:756–763.
35. Wang T, Zhu Y, Gai F (2004) Folding of a three-helix bundle at the folding speed limit. *J Phys Chem B*, 108:3694–3697.
36. McCallister EL, Alm E, Baker D (2000) Critical role of β-hairpin formation in protein G folding. *Nat Struct Biol*, 7:669–673.
37. Nymeyer H, García AE, Onuchic JN (1998) Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc Natl Acad Sci USA*, 95:5921–5928.
38. Best RB, Hummer G (2005) Reaction coordinates and rates from transition paths. *Proc Natl Acad Sci USA*, 102:6732–6737.
39. Cho SS, Levy Y, Wolynes PG (2006) *P* versus *Q*: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc Natl Acad Sci USA*, 103:586–591.
40. Karanicolas J, Brooks CL III (2002) The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci*, 11:2351–2361.
41. Hummer G (2004) From transition paths to transition states and rate coefficients. *J Chem Phys*, 120:516–523.
42. Hummer G, Szabo A (2003) Kinetics from non-equilibrium single-molecule pulling experiments. *Biophys J*, 85:5–15.
43. Best RB, Paci E, Hummer G, Dudko OK (2008) Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J Phys Chem B*, 112:5968–5976.
44. Best RB, Hummer G (2008) Protein folding kinetics under force from molecular simulation. *J Am Chem Soc*, 130:3706–3707.
45. Zagrovic B, Snow CD, Khaliq S, Shirts MR, Pande VS (2002) Native-like mean structure in the unfolded ensemble of small proteins. *J Mol Biol*, 323:153–164.
46. Zwanzig R, Szabo A, Bagchi B (1992) Levinthal's paradox. *Proc Natl Acad Sci USA*, 89:20–22.
47. Muñoz V, Eaton WA (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA*, 96:11311–11316.
48. Bruscolini P, Pelizzola A (2002) Exact solution of the Muñoz-Eaton model for protein folding. *Phys Rev Lett*, 88:258101.
49. Kubelka J, Henry ER, Hofrichter J, Eaton WA (2008) Chemical, physical and theoretical kinetics of an ultrafast folding protein. *Proc Natl Acad Sci USA*, 105:18655–18662.
50. Rhee YM, Pande VS (2005) One-dimensional reaction coordinate and the corresponding potential of mean force from commitment probability distribution. *J Phys Chem B*, 109:6780–6786.
51. Krivov SV, Karplus M (2008) Diffusive reaction dynamics on invariant free energy profiles. *Proc Natl Acad Sci USA*, 105:13841–13846.
52. Elber R, Karplus M (1987) Multiple conformational states of proteins: A molecular dynamics analysis of myoglobin. *Science*, 235:318–321.
53. Portman JJ, Takada S, Wolynes PG (2001) Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics. *J Chem Phys*, 114:5082–5096.
54. Hyeon C, Thirumalai D (2003) Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments. *Proc Natl Acad Sci USA*, 100:10249–10253.
55. Chavez LL, Onuchic JN, Clementi C (2004) Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates. *J Am Chem Soc*, 126:8426–8432.
56. Wolynes PG (2005) Recent successes of the energy landscape theory of protein folding and function. *Q Rev Biophys*, 38:405–410.
57. Kim YC, Hummer G (2008) Coarse-grained models for simulation of multiprotein complexes: Application to ubiquitin binding. *J Mol Biol*, 375:1416–1433.
58. Plotkin SS (2001) Speeding protein folding beyond the Gō model: How a little frustration sometimes helps. *Proteins*, 45:337–345.
59. Clementi C, Plotkin SS (2004) The effect of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci*, 13:1750–1766.
60. Hu J, Ma A, Dinner AR (2006) Monte Carlo simulations of biomolecules: The MC module in CHARMM. *J Comp Chem*, 27:203–216.
61. Kramers HA (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–303.
62. Yang JS, Wallin S, Shakhnovich EI (2008) Universality and diversity of folding mechanics for three-helix bundle proteins. *Proc Natl Acad Sci USA*, 105:895–900.
63. Rhee YM, Sorin EJ, Jayachandran G, Lindahl E, Pande VS (2004) Simulations of the role of water in the protein folding mechanism. *Proc Natl Acad Sci USA*, 101:6456–6461.
64. Johansson MU, et al. (1997) Solution structure of the albumin-binding GA module: A versatile bacterial protein domain. *J Mol Biol*, 266:859 865.
65. Gronenborn AM, et al. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science*, 253:657–661.
66. Brooks BR, et al. (2009) CHARMM: The Biomolecular Simulation Program. *J Comput Chem*, 30:1545–1614.