Fast protein folding is governed by memory-dependent friction

Benjamin A. Dalton[a] (iD), Cihan Ayaz[a] (iD), Henrik Kiefer[a] (iD), Anton Klimek[a] (iD), Lucas Tepper[a], and Roland R. Netz[a,1] (iD)

When described by a low-dimensional reaction coordinate, the folding rates of most proteins are determined by a subtle interplay between free-energy barriers, which separate folded and unfolded states, and friction. While it is commonplace to extract free-energy profiles from molecular trajectories, a direct evaluation of friction is far more elusive and typically relies on fits of measured reaction rates to memoryless reaction-rate theories. Here, using memory-kernel extraction methods founded on a generalized Langevin equation (GLE) formalism, we directly calculate the time-dependent friction acting on the fraction of native contacts reaction coordinate Q, evaluated for eight fast-folding proteins, taken from a published set of large-scale molecular dynamics protein simulations. Our results reveal that, across the diverse range of proteins represented in this dataset, friction is more influential than free-energy barriers in determining protein folding rates. We also show that proteins fold in a regime where the finite decay time of friction significantly reduces the folding times, in some instances by as much as a factor of 10, compared to predictions based on memoryless friction.

protein folding | reaction rate theory | molecular friction | generalized Langevin equation | non-Markovian dynamics

For many proteins, functionality depends on successfully folding into a specific three-dimensional conformational state. This requires that a linear polypeptide chain is driven to explore conformation space by interactions with a solvating environment and is shaped by both solvent interactions and internal interactions between amino acids. When described by a low-dimensional reaction coordinate, folding and unfolding are typically described by a free-energy landscape (1–6), with distinct states separated by free-energy barriers. For proteins that fold in less than 100 μs, so-called "fast-folding" proteins, barrier heights determined from experiment, theory, and simulation are of the order of just a few k_BT's in height. Theories for describing protein folding are often phrased as reaction rate theories with an explicit dependence on the free energy barrier and friction. An accurate understanding of the kinetics in barrier-crossing reactions is important not only in characterizing protein folding but also in many other domains, including nucleation theory and chemical reaction kinetics. Reaction rate theory dates back to Arrhenius, who, in 1889, showed that the transition times between reactant and product scale according to $\kappa e^{U_0/k_BT}$, where U_0 represents the energy barrier height between the two states (7). In principle, the prefactor κ can depend on both the free-energy profile and friction. Building on the transition state theory of Eyring (8), Kramers was the first to include an explicit dependence of κ on solvent friction (9). The friction in Kramers' theory is frequency-independent, suggesting that the environment of the reacting system, and possibly internal processes, relax infinitely fast compared to the barrier crossing time. Despite its simplicity, Kramers' theory predicts reaction rates in both overdamped and inertia-dominated regimes, making it suitable for describing a wide range of systems. Several advancements have emerged that bridge the gap between the overdamped and inertia-dominated regimes, while also accommodating finite friction relaxation times (10–14). When considering proteins, it is understood that friction arises from a complex interplay between interactions with the surrounding solvent and internal intramolecular interactions (15–18). This ultimately leads to finite relaxation times (19). A direct evaluation of the friction acting on a protein has not been possible and approaches have rather relied on determining friction indirectly from memoryless reaction rate theory (20–23). This paper presents a method for directly extracting the friction that affects the conformational dynamics of proteins from molecular dynamics simulation trajectories, which accounts for both internal and external friction contributions, as well as finite relaxation times.

We evaluate friction for eight fast-folding proteins using previously published large-scale molecular dynamics (MD) simulation trajectories (24). These simulations, conducted by the Shaw group on the purpose-built Anton super-computer (25, 26),

Significance

To describe protein-folding dynamics, protein molecular structures are typically projected onto some low-dimensional reaction coordinate. This dimensionality reduction produces non-Markovian friction, resulting in protein dynamics that is typically overdamped and influenced by the history of the reaction coordinate. Despite this, protein-folding dynamics is usually described as a memoryless Markovian process with effective friction fitted to experimental or simulation data. Using recently developed techniques, we extract friction memory kernels from molecular dynamics simulations of 8 fast-folding proteins. We show that memory effects are significant and that the folding times depend more on the friction magnitudes than on the free-energy profiles. Quantitative theoretical predictions of protein folding times thus require full knowledge of the friction memory kernel and appropriate non-Markovian reaction rate theory.

represent a breakthrough in large-scale simulation. The resulting trajectories cover protein lengths ranging from 10 to 80 amino acid residues and simulation times of 100 μs to 3 ms. They represent a diverse protein dataset, composed of mixtures of α-helix and β-hairpin secondary structures and an assortment of different tertiary structures.

We analyze the protein folding trajectories in the framework of a non-Markovian, generalized Langevin equation (GLE). Although previous studies have examined non-Markovian effects in protein folding (27, 28) and Markov-state models have been expanded to include memory effects (29), our investigation is the first to employ such extensive simulation data and utilize recent GLE memory-kernel extraction methods (30–32). The GLE is a reduced dynamic description of a complex multidimensional system (33, 34). Here, we map the all-atom dynamics of the composite water–protein systems onto a one-dimensional fraction of native contacts reaction coordinate Q (35, 36) and hence extract friction memory kernels from $Q(t)$. Q has been previously established as a good reaction coordinate for the proteins studied in this paper (36). Free-energy profiles and friction are specific to a given reaction coordinate. The extracted friction kernels reflect the combined dissipative effects of solvent and internal relaxation on the reaction coordinate over a wide range of time scales. To simulate multiple folding and unfolding events, Lindorff-Larsen et al. selected for proteins that have fast folding times (24). Using the extracted free-energy profiles and friction kernels, we

predict folding times based on reaction rate theory on different levels of sophistication and compare to folding times from MD simulations. Our analysis shows that the variation of folding and unfolding times across the considered proteins is correlated more strongly with friction than with free energy barriers. In fact, the most accurate prediction for protein folding times is given by a previously developed multi-time-scale, non-Markovian theory (13, 14). Interestingly, we find that memory decay times are long for all proteins in the dataset, with some being as long as the folding times themselves. This indicates that Q is a poor reaction coordinate when judged based on its non-Markovianity. Overall, our results suggest that friction plays a dominant role in governing the fast conformational dynamics of proteins, and when represented by a reduced-dimensional reaction coordinate, one must account for strong non-Markovian effects.

Results & Discussion

Extracting Friction from Protein Simulations. The native state structures and the number of residues of each protein are presented in Fig. 1A. The fraction of native contacts reaction coordinate Q provides a contact-based measure for the deviation of a given state away from the native configuration, which typically represents the folded state. Q is closer to $Q = 1$ in the folded state and closer to $Q = 0$ when unfolded. In Fig. 1B, we show a 250 μs trajectory segment for the α_3D
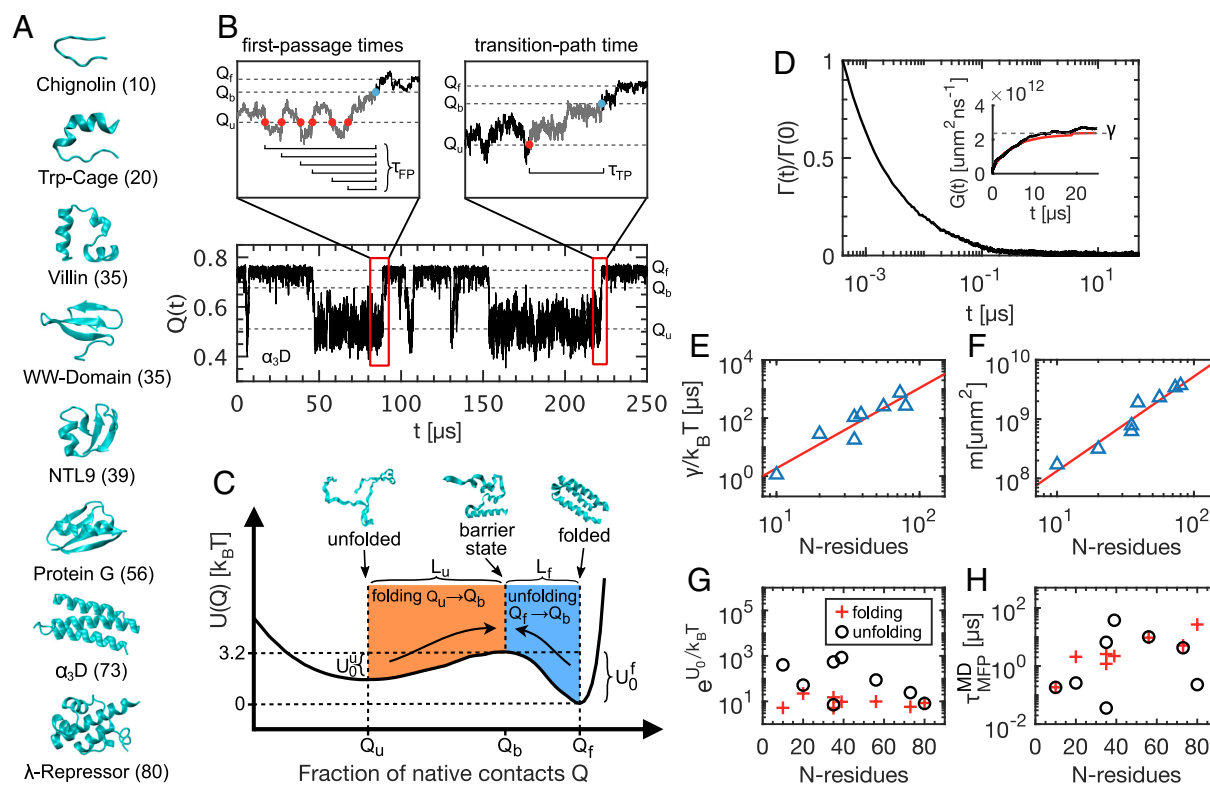

Fig. 1. The folding and unfolding of eight fast-folding proteins. (A) Native states for all proteins with the number of amino acids indicated. All protein trajectories are taken from ref. 24. (B) 250 μs trajectory segment for the $Q(t)$ reaction coordinate (α_3D protein). *Left* magnification: a sequence of folding first-passage events, from the unfolded state Q_u to the barrier *top* Q_b. *Right* magnification: an example folding transition path and corresponding transition path time τ_{TP}. (C) Free-energy profile for the α_3D protein. Configuration snapshots show example unfolded, barrier-top, and folded states. Q_u, Q_b, and Q_f are the reaction coordinate values in the unfolded, barrier-top, and folded states, respectively. For α_3D, the barrier faced by the unfolded protein, $U_0^u = U(Q_b) - U(Q_u) = 1.8\ k_BT$, is less than the barrier faced by the folded protein $U_0^f = U(Q_b) - U(Q_f) = 3.2\ k_BT$. The distance from the unfolded state to the barrier top $L_u = Q_b - Q_u$ is greater than the distance from the folded state to the barrier top $L_f = Q_f - Q_b$. (D) Normalized GLE memory kernel $\Gamma(t)$, extracted from $Q(t)$ for the α_3D protein. Inset: Running integral $G(t)$ (black line), the limiting total friction γ (dashed line), and an exponential fit (red curve). (E) Total friction γ for each protein, divided by k_BT, plotted as a function of the number of residues N, (power law with exponent ~ 2.8 red line). (F) Effective mass m, plotted as a function of N (power law with exponent ~ 1.5 red line). (G) and (H) show the Arrhenius factor e^{U_0/k_BT} and barrier crossing times τ_{MFP}^{MD} for folding and unfolding transitions individually.

protein (for a summary of simulation details, see *Methods* and *SI Appendix*, section 1). The folded and unfolded states are discernible, located at $Q_f = 0.75$ and $Q_u = 0.51$, respectively, separated by a barrier at $Q_b = 0.66$. Throughout this paper, we consider folding transitions as leading from the unfolded state minimum to the barrier top ($Q_u \to Q_b$) and unfolding transitions as leading from the folded state minimum to the barrier top ($Q_f \to Q_b$). This accounts for the asymmetries observed in the free-energy profiles for the eight proteins. In the *Left* magnification above the trajectory, we identify a sequence of folding first-passage events. The mean of all first-passage events gives the mean first-passage time τ_{MFP}^{MD} for the folding reaction. Similarly, we evaluate the unfolding times as the mean of all first-passage events from Q_f to Q_b. In the *Right* magnification, we present an example folding transition path, linking the unfolded state minimum directly to the barrier top. The mean transition path time τ_{MTP} is the mean of all transition path durations for either folding or unfolding transition. From the trajectories of $Q(t)$ for each protein, we calculate the free-energy profile $U(Q)$ acting on the reaction coordinate. In Fig. 1*C*, we show $U(Q) = -k_B T \log[\rho(Q)]$ for the $\alpha_3 D$ protein, where $\rho(Q)$ is the probability density for $Q(t)$, k_B is Boltzmann's constant, and T is the temperature, which is different for each protein. The free-energy profiles for all proteins are asymmetric, so that the barrier heights faced by the folded and unfolded states (U_0^u and U_0^f) are not equal. Similarly, the distances in reaction-coordinate space from the minima to the barrier top (L_f and L_u) are also different. The free-energy profiles for all proteins are shown in *SI Appendix*, section 2.

Central to this paper is extracting time-dependent friction kernels $\Gamma(t)$ from $Q(t)$ trajectories of each of the eight proteins. We describe the stochastic dynamics of $Q(t)$ with the one-dimensional generalized Langevin equation (GLE)

$$m\ddot{Q}(t) = -\int_0^t \Gamma(t - t')\dot{Q}(t')dt' - \nabla U[Q(t)] + F_R(t), \quad [1]$$

where $\Gamma(t)$ is the friction memory kernel, $F_R(t)$ is the random force term satisfying the fluctuation-dissipation theorem $\langle F_R(t) F_R(t') \rangle = k_B T \Gamma(t - t')$, m is the effective mass of the reaction coordinate, and $U(Q)$ is the free-energy profile. Although Eq. 1 neglects nonlinear friction and is approximate (37–39), it has been demonstrated to be accurate for polypeptides (32), pair reactions in water (40), and chemical bond vibrations (41). There are different techniques for extracting dynamic friction from discrete time-series trajectories (30, 31, 42–44). We explain the method used to extract $\Gamma(t)$ from $Q(t)$ in detail in *SI Appendix*, section 3. In short, we directly extract the running integral of the memory kernel $G(t) = \int_0^t \Gamma(t')dt'$ using a Volterra extraction scheme, which is appropriate for arbitrary, nonlinear free-energy profiles $U(Q)$ (31, 32). The memory kernel is then given by $\Gamma(t) = dG(t)/dt$, which is evaluated numerically. In Fig. 1*D*, we show the memory kernel for the $\alpha_3 D$ protein, normalized by $\Gamma(0)$. The dynamics of the $\alpha_3 D$ protein exhibit significant memory effects, which is evident from the inset that shows the running integral $G(t)$ reaching a plateau after about 10 µs. We present the memory kernels for all proteins in *SI Appendix*, section 4.

Chain-Length Dependence of Protein Dynamics. After extracting the full time-dependent friction kernel via $G(t)$, we have access to the steady-state friction $\gamma = G(t \to \infty)$, which we refer to as the total friction throughout. In Fig. 1*E*, we show γ for each protein as a function of the number of residues in each

chain N. Since each system has a unique temperature, we divide by $k_B T$ and hence display the inverse total diffusivity along Q. We fit a power law and find that $\gamma/k_B T = 3.2 \times 10^{-3} N^{2.8}$ µs (red line). For a normalized reaction coordinate that is a sum of N uncorrelated atomic distances, one expects linear scaling in N (*SI Appendix*, section 5). However, Q is a highly nonlinear reaction coordinate. Additionally, it is known that reptation dominates the dynamics of collapsed polymers (45, 46). Both of these effects contribute to the super-linearity in N-scaling for the total friction. The effective mass of Q, calculated using the equipartition theorem $m = k_B T/\langle \dot{Q} \rangle^2$, also depends on N (Fig. 1*F*). The power law scaling is given by $m = 3.4 \times 10^6 N^{1.5}$ unm² (red line). The super-linearity here is solely due to the nonlinearity of Q (37). The Arrhenius factors $e^{U_0/k_B T}$, plotted in Fig. 1*G*, show very little correlation with protein size. The folding and unfolding kinetics, however, quantified by the mean first-passage times τ_{MFP}^{MD} (Fig. 1*H*), increase with N. The correlation is weak but the results are consistent with experimental results for a different dataset that encompasses larger proteins (47) (see *SI Appendix*, section 6 for a statistical analysis). Overall, Fig. 1 *E–H* reveal that, when projected onto Q, friction plays the dominant role in governing the dynamics and kinetics of the proteins considered here. A similar conclusion was previously reported using a 1D Markovian Langevin model (48). We explore this in more detail in a subsequent section.

Memory Duration Is Significant in Protein Folding. Extracting $\Gamma(t)$ provides not only the total friction γ but also a time scale for the duration of memory effects. In addition, γ can be used predictively to assess other key dynamic time scales. The inertia time $\tau_m = m/\gamma$ is the time scale at which the system becomes diffusive (49), while the diffusion time,

$$\tau_D = \frac{\gamma L^2}{k_B T}, \quad [2]$$

is the time taken for a Brownian particle to diffuse over a characteristic distance L in the absence of free-energy gradients. L is the distance from the unfolded state minimum to the top of the folding barrier (L_u) or the distance from the folded state minimum to the barrier top (L_f). A system is considered to be in the overdamped regime when $\tau_m \ll \tau_D$. All eight proteins in this study meet this condition for both L_u and L_f (Fig. 2*A*). One borderline case is Chignolin unfolding ($\tau_m/\tau_D = 0.027$), for which both L_f and γ are small.

By comparing τ_D to the decay times of extracted memory kernels, we can assess the expected influence of memory on barrier-crossing processes (12). Memory-dependent friction is typically multimodal, with time scales spanning many orders of magnitude (32). To assign a single effective time-scale to the memory decay, we evaluate the first-moment according to $\tau_{mem} = \int_0^\infty t\Gamma(t)dt / \int_0^\infty \Gamma(t)dt$. In *SI Appendix*, section 7, we address the issue of temporal discretisation of the MD data. Fig. 2*B* shows that memory times are nonnegligible when compared to diffusion times. Memory has been shown to accelerate barrier crossing kinetics in the range $1 \times 10^{-2} < \tau_{mem}/\tau_D < 1 \times 10^1$ (12–14), as indicated in Fig. 2*B*. Thus, for the proteins considered here, memory is predicted to accelerate protein folding and unfolding, which we quantify further below. In Fig. 2*C*, we see that memory times are of the same order as the reaction times, indicating that Q is a poor reaction coordinate. The overall concurrence between τ_{mem}, τ_D, and τ_{MFP}^{MD} predicts a coupling between memory effects and folding kinetics.
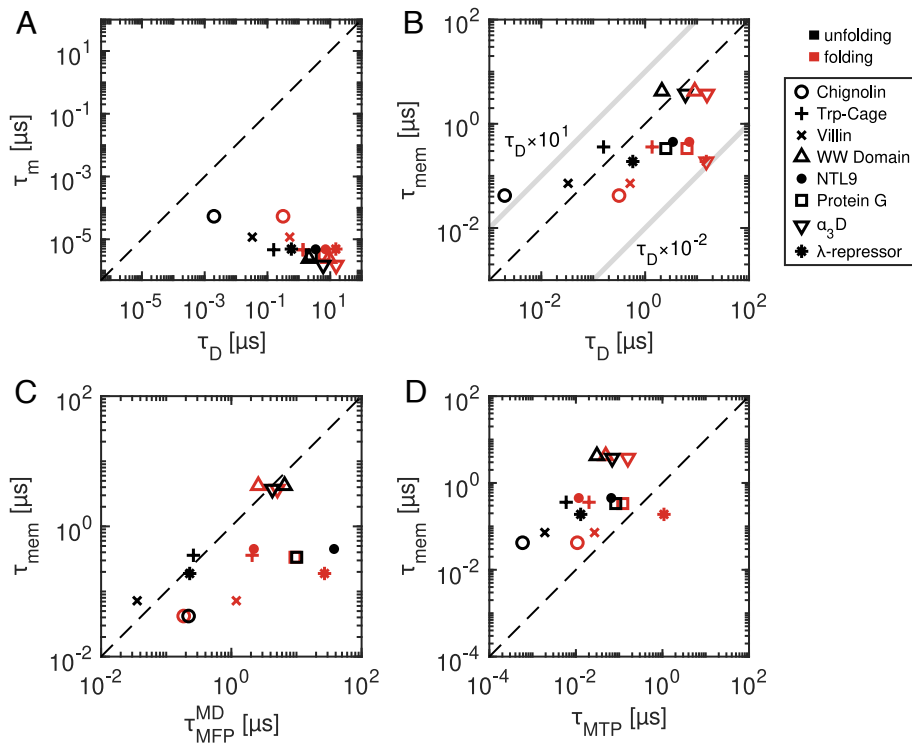
Fig. 2. Comparison of relevant time scales for protein folding and unfolding kinetics. (*A*) The inertia time scales τ_m compared to the diffusion times τ_D, showing that all systems are in the over-damped regime. (*B*) Memory decay time scales τ_{mem}, calculated from the first moment of the memory kernel $\Gamma(t)$, compared to the diffusion times τ_D. The light gray lines indicate the bounding regime for τ_{mem} between $\tau_D \times 10^{-2}$ and $\tau_D \times 10^1$, which is the domain in which memory-induced kinetic acceleration is expected. (*C*) Memory times τ_{mem} compared to the folding and unfolding times, expressed as the mean first-passage times τ_{MFP}^{MD}. (*D*) Memory times compared to the transition path times τ_{MTP} leading from the folded and unfolded state minima to the barrier tops. The broken lines in each plot indicate exact equivalence between the respective times.

In Fig. 2*D*, we show that transition path times are much shorter than memory times for all proteins, sometimes by up to two orders of magnitude. This is significant since it indicates that entire transition processes occur under the influence of memory effects that extend back relatively far into the protein's dynamic history. Transition paths, which are the segments of a trajectory where the protein undergoes a transition from one state to a target state, are of great interest since they reveal the actual folding mechanisms of a protein. They have been extensively studied in experiments (23, 50–55) and in simulations and theory (24, 36, 56–60).

Non-Markovian Rate Theory Predicts Protein Folding Times.
From the simulation trajectories, we extract the folding and unfolding times, evaluated via the mean first-passage times τ_{MFP}^{MD}. Since we also extract the friction directly from the simulation trajectories, we can compare the performance of various reaction rate theories. To quantify deviations between predicted and measured folding times, we calculate the root-mean-square logarithmic deviation (RMSLD), which is a logarithmic measure of the relative deviations given by RMSLD = $\sqrt{\sum_{i=1}^{n}(\log(\tau_{MFP,i}^{MD}) - \log(\tau_{MFP,i}^{theo}))^2/n}$. Here, $\tau_{MFP,i}^{MD}$ is the value measured from MD, $\tau_{MFP,i}^{theo}$ is the theoretical prediction for a given model, and $n = 16$ is the total number of folding and unfolding reactions (for details, see *SI Appendix*, section 8). Fig. 3*A* shows that τ_D provides a remarkably good prediction of the MD data, despite completely neglecting contributions from free-energy profiles. A Markovian prediction, which in addition to friction explicitly accounts for the extracted free energy profile (49), is given by

$$\tau_{MFP}^{Mar}(Q_s, Q_e) = \tau_D \xi_U$$
$$= \tau_D \int_{Q_s}^{Q_e} e^{U(x)/k_B T} \left[\int_{-\infty}^{x} e^{-U(y)/k_B T} \frac{dy}{L} \right] \frac{dx}{L}. \quad [3]$$

Eq. **3** assumes constant friction across Q and evaluates the mean first-passage times between a start and end-point on the free-energy profile, Q_s and Q_e. For folding reactions, for example, we calculate $\tau_{MFP}^{Mar}(Q_u, Q_b)$ (*SI Appendix*, section 9). In Fig. 3*B*, we see that the Markovian predictions are generally slow compared to the MD data. The RMSLD of 1.41 indicates that the Markovian prediction (Eq. **3**) is slightly more accurate than τ_D, which has an RMSLD of 1.51, indicating that by explicitly including the free-energy profile via the correction factor ξ_U, defined in Eq. **3**, we only slightly improve the prediction. To emphasize this, we consider the predictive power of the free energy profiles alone and hence evaluate $\tau_D' \xi_U$ (Fig. 3*C*), where τ_D' is the diffusion time using a common friction coefficient for all proteins, chosen to minimize the RMSLD. The relatively large RMSLD value of 1.98 indicates that ξ_U is not a good predictor for folding times. We conclude that friction alone describes the variation of folding and unfolding times among this set of proteins well, while the explicit inclusion of free-energy profiles does little to improve the prediction.

We now consider the influence of non-Markovian effects on the folding time predictions. A formula was recently constructed that predicts the memory-induced speed-up and slow-down regimes (12–14). Tested against extensive simulations, this formula explicitly accounts for non-Markovian dynamics correctly for multiexponential memory. Based on this formula, we introduce a non-Markovian correction factor ξ_{noMar} as

$$\xi_{noMar} = \frac{\tau_{MFP}^{H}(\tau_D, U_0, \{\gamma_i\}, \{\tau_i\}, \tau_m)}{\tau_{MFP}^{H}(\tau_D, U_0, \{\gamma_i\}, 0, 0)}. \quad [4]$$

Here, τ_{MFP}^{H} is the barrier-crossing time of a one-dimensional reaction coordinate with finite mass and a memory kernel $\Gamma(t) = \sum_{i=1}^{M}(\gamma_i/\tau_i)\exp(t/\tau_i)$, characterized by M friction amplitudes $\{\gamma_i\}$ and M memory times $\{\tau_i\}$ (13, 14) (see *Methods* section Eqs. **M2**–**M4** and *SI Appendix*, section 10 for details of ξ_{noMar} and τ_{MFP}^{H}). The multiexponential fits of our extracted memory kernels
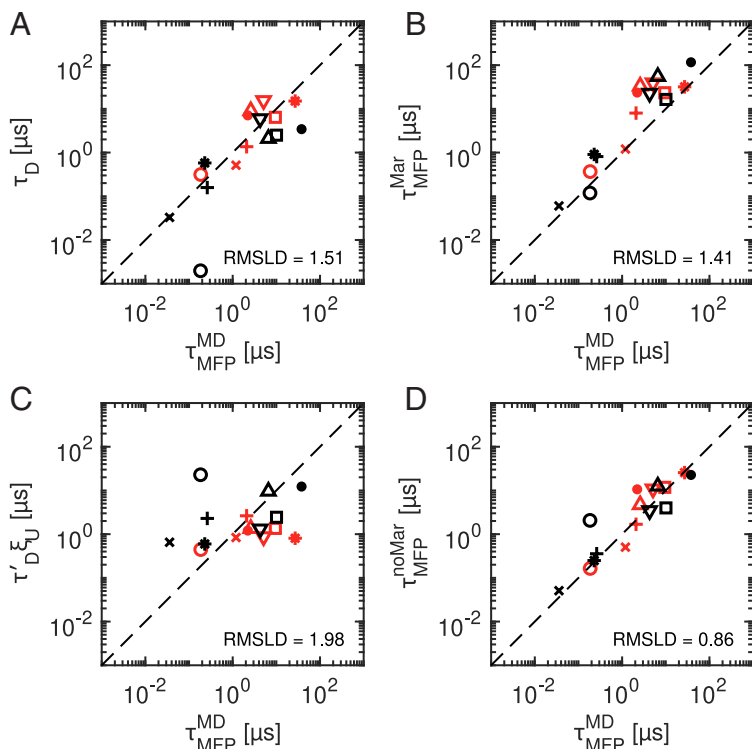
Fig. 3. Comparison of simulated protein folding and unfolding times $\tau_{\text{MFP}}^{\text{MD}}$ with predictions on different levels of theory. (*A*) Diffusion times τ_{D}, according to Eq. **2**, which do not depend on free-energy barriers. (*B*) Markovian predictions $\tau_{\text{MFP}}^{\text{Mar}}$, according to Eq. **3**, which use the extracted free-energy profile for each protein. (*C*) Free-energy-dependent factor ξ_{U} that accounts for the effect of the free energy profile but not for the friction, according to Eq. **3**. (*D*) Non-Markovian predictions $\tau_{\text{MFP}}^{\text{noMar}}$, according to Eq. **5**. Multiexponential memory kernels are explicitly accounted for. The root-mean-squared deviations of the logarithmically transformed data (RMSLD) provides a measure of relative deviations for each model prediction from the observed data $\tau_{\text{MFP}}^{\text{MD}}$ over the combined set of folding and unfolding times. A lower RMSLD indicates that a given model prediction is more accurate, where the dashed lines indicate an exact prediction. In (*C*), τ_{D}' is calculated using $\gamma = 1.1 \times 10^9 \, \text{unm}^2\mu\text{s}^{-1}$ for all proteins, chosen to minimize the RMSLD. See Fig. 2 for symbol legend.

are discussed in *SI Appendix*, section 4, for Chignolin we use $M=2$ and for all other proteins $M=3$. We note that all parameters in Eq. **4** are uniquely extracted from the MD simulations. The denominator in Eq. **4** represents the overdamped Markovian limit, where the inertial time and all memory times limit to zero. The non-Markovian prediction for the barrier-crossing time can thus be written as

$$\tau_{\text{MFP}}^{\text{noMar}} = \xi_{\text{noMar}}\tau_{\text{MFP}}^{\text{Mar}} = \xi_{\text{noMar}}\xi_{\text{U}}\tau_{\text{D}}. \qquad [5]$$

In Fig. 3*D*, we see that $\tau_{\text{MFP}}^{\text{noMar}}$ improves the overall prediction of the MD folding and unfolding times (except for Chignolin, Villin and Protein G), as confirmed by the relatively low RMSLD value of 0.86. See *SI Appendix*, section 8 for additional statistical analysis of the data in Fig. 3.

Proteins Fold in the Memory-Induced Speed-Up Regime. Memory effects can either speed-up or slow-down barrier-crossing times when compared to the memoryless limit, depending on the value of $\tau_{\text{mem}}/\tau_{\text{D}}$ (12–14). In our analysis, we use $\tau_{\text{MFP}}^{\text{Mar}}$, as given by Eq. **3**, to represent the overdamped, memoryless limit. By plotting $\tau_{\text{MFP}}^{\text{MD}}/\tau_{\text{MFP}}^{\text{Mar}}$ against the rescaled memory times $\tau_{\text{mem}}/\tau_{\text{D}}$ for both folding (red symbols) and unfolding (black symbols) of all proteins in Fig. 4*A*, we reveal that speed-up is most pronounced for intermediate values of $1\times10^{-1} < \tau_{\text{mem}}/\tau_{\text{D}} < 1\times10^1$, while in the short memory-time limit $\tau_{\text{mem}}/\tau_{\text{D}} \to 0$, the reaction times trend to Markovian behaviour, i.e., $\tau_{\text{MFP}}^{\text{MD}}/\tau_{\text{MFP}}^{\text{Mar}} \to 1$. This is in line with previous treatments of model systems (13, 14) and simulations of short homo-peptide chains (32). Fig. 4*A*, therefore, indicates that protein folding and unfolding times are significantly accelerated by memory effects, in some instances by as much as a factor of 10, which is a remarkable contribution.

This notion can be made more quantitative by using ξ_{noMar} predictively. According to Eq. **5**, ξ_{noMar} describes the ratio $\tau_{\text{MFP}}^{\text{noMar}}/\tau_{\text{MFP}}^{\text{Mar}}$ and quantifies deviations from Markovianity. In

Fig. 4*A*, we observe an excellent agreement between ξ_{noMar} (blue and yellow symbols for folding and unfolding) and the rescaled simulation data, suggesting that a multimodal, non-Markovian reaction rate theory accurately describes the accelerated barrier-crossing in protein folding simulations, even predicting a 10-fold acceleration for intermediate memory times.

To gain a deeper understanding of how memory affects protein folding times, we rescale all memory times in Eq. **4** by a common factor α and generate an α-dependent memory kernel $\Gamma^{\alpha}(t) = \sum_{i=1}^{M}(\gamma_i/\alpha\tau_i)\exp(t/\alpha\tau_i)$. Such time-rescaling can be considered analogous to varying the solvent viscosity. In Fig. 4*B*, we show $\xi_{\text{noMar}}^{\alpha}$ from Eq. **4** using the α-dependent memory kernel $\Gamma^{\alpha}(t)$ for four proteins as solid lines, plotted as a function of $\alpha\tau_{\text{mem}}/\tau_{\text{D}}$. The vertical dashed lines denote $\alpha = 1$, corresponding to the original proteins described by the MD simulations. We overlay the MD results ($\tau_{\text{MFP}}^{\text{MD}}/\tau_{\text{MFP}}^{\text{Mar}}$—the same colors and symbols as in Fig. 4*A*) at $\alpha = 1$, which show good agreement with Eq. **4**. For comparison, we also show the folding and unfolding times predicted by Grote–Hynes (GH) theory (11) as broken lines, which also explicitly accounts for memory effects (see *Methods* section Eqs. **M6–M10**). The Grote–Hynes curves, $\tau^{\text{GH},\alpha}/\tau_{\text{GH}}^{\text{HF}}$, agree with $\xi_{\text{noMar}}^{\alpha}$ for small memory times but disagree significantly for longer memory times. Here, $\tau^{\text{GH},\alpha}$ is the GH prediction (τ^{GH} - Eq. **M6**) with memory times rescaled by α, and $\tau_{\text{HF}}^{\text{GH}}$ is the high-friction, memoryless GH limit (Eq. **M10**) (see *SI Appendix*, section 10 for details). Previous work has discussed the breakdown of GH theory for long memory times for model systems (12, 40). Here, we demonstrate the relevance for protein folding kinetics. While GH theory agrees with ξ_{noMar} for some proteins (e.g., α_3D and the λ-repressor in Fig. 4*B*), it significantly fails for other proteins (e.g., Trp-Cage and WW Domain; see also *SI Appendix*, Figs. S8 and S9), depending on the value of $\tau_{\text{mem}}/\tau_{\text{D}}$. Specifically, Trp-Cage unfolding is described by a particularly high value of $\tau_{\text{mem}}/\tau_{\text{D}} = 2$, which corresponds to the onset of memory-induced slowdown, as predicted well by Eq. **4** but not by GH theory.
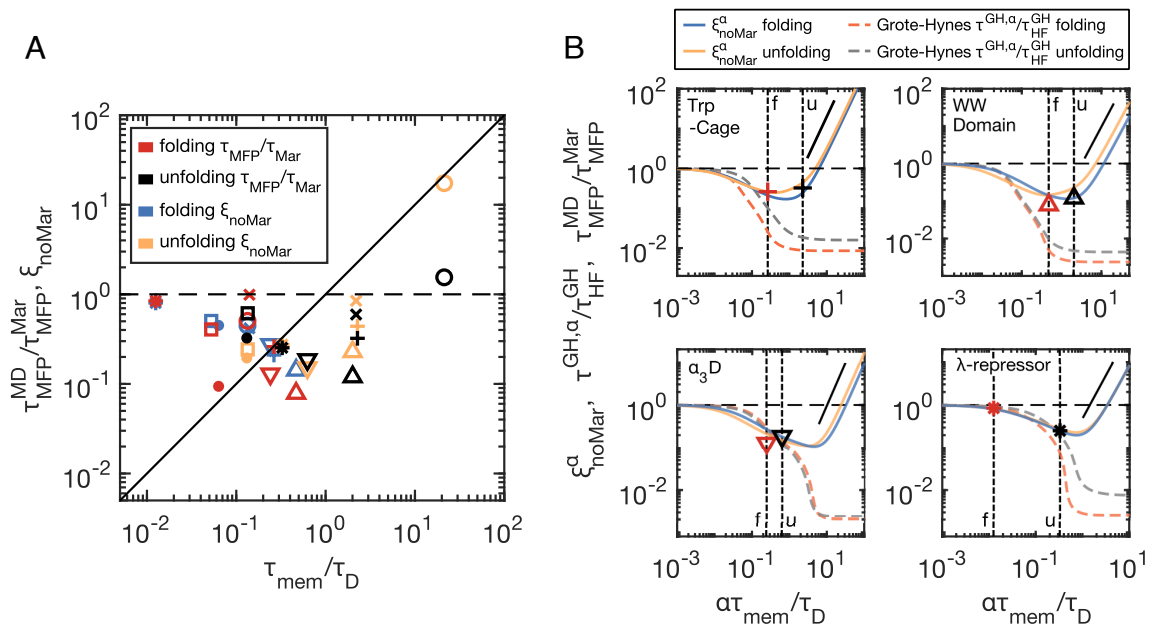
Fig. 4. Barrier crossing times indicate memory-induced speed up. (A) Deviations from Markovian barrier crossing kinetics, plotted as a function of rescaled memory times τ_{mem}/τ_D, for the folding and unfolding of all eight proteins. Rescaled MD simulation values for each protein ($\tau_{MFP}^{MD}/\tau_{MFP}^{Mar}$, red and black symbols) are compared to multimodal, non-Markovian prediction (ξ_{noMar}, Eq. **4**, blue and yellow symbols). See Fig. 2 for symbol legend. (B) Reaction-time curves generated by scaling all memory times by a common factor α. Vertical dashed lines indicate $\alpha = 1$, which are different for folding (f) and unfolding (u) due to unique τ_D values. The multimodal non-Markovian prediction (ξ_{noMar}^{α} from Eq. **4**—folding (blue) and unfolding (yellow)) is compared to the Grote–Hynes prediction ($\tau^{GH,\alpha}$ from Eq. **M6**), rescaled by the memoryless, high-friction limit (τ_{HF}^{GH} - Eq. **M10**) for folding (orange) and unfolding (grey). The blue and yellow symbols shown in (A) for the four example proteins coincide with the curve intercepts of the $\alpha = 1$ lines in (B) for ξ_{noMar}^{α}. The red and black symbols are the same as those shown in (A), corresponding to the MD results for each protein, located at $\alpha = 1$. The solid black lines show quadratic scaling, predicted for ξ_{noMar} in the long memory-time limit. Dashed horizontal black lines show unity for all plots.

Conclusion

We have extracted memory kernels from published MD simulation trajectories of eight fast-folding proteins (24) and have shown that, for the Q reaction coordinate, each protein's memory time is comparable to its folding time and typically vastly exceeds its transition path time. By extracting the memory kernels, and from that the total friction γ, we show that a simple prediction for the barrier crossing times, which only depends on friction, i.e., τ_D, represents the simulated barrier crossing times well (Fig. 3A). Explicitly accounting for the exact free-energy profiles (Figs. 3 B and C) does not particularly improve upon this prediction. A theory that explicitly accounts for multi-time-scale memory-dependent friction most accurately predicts the folding and unfolding times observed in the MD simulations (Fig. 3D), demonstrating that non-Markovian effects must be accounted for when describing protein dynamics via some reduced-dimensional reaction coordinate.

The quality of reaction coordinates has been estimated using transition-state-ensemble and commitment-probability methods (58, 61, 62), based on the idea that an ideal reaction coordinate should project all transition states onto a single point and be Markovian (28). Using such methods and the same dataset from the Shaw group as we use here (24), Best and Hummer argued that Q is a good reaction coordinate for describing protein folding dynamics (36). However, we demonstrate that Q displays strong non-Markovianity, suggesting that qualifying a reaction coordinate is subtle. In *SI Appendix*, section 11, we show memory kernels for other standard reaction coordinates, such as the end-to-end distance, radius of gyration, and the RMSD from the native state. The resultant memory times span orders of magnitudes but are typically of the order of the diffusion time scale τ_D. Therefore, none of these reaction coordinates

can be considered as a good reaction coordinate when judged based on their Markovianity. Regardless, by faithfully extracting memory friction functions from MD trajectories and using non-Markovian reaction rate theories, the folding kinetics of proteins can be analyzed and accurately predicted. Such a method may prove to be ideal in the experimental context (50–52, 63, 64) where the investigator does not have the luxury to choose optimal reaction coordinates.

Alternative models, which neglect non-Markovian effects but assume position-dependent friction (5, 58), fail to generate a unique friction profile $\gamma(Q)$ that describes both folding and unfolding kinetics (*SI Appendix*, section 12). This was already demonstrated for a simple α-helix forming homo-alanine chain (32). Similarly, we show in *SI Appendix*, section 13 that Markovian simulations with position-independent friction cannot reproduce the characteristic subdiffusive dynamics exhibited by proteins, contrary to simple GLE simulations. It is possible that position-dependent memory effects may also be important for protein folding in general (37–39), but as we demonstrate in *SI Appendix*, section 13, any such effects are minor and therefore negligible for the proteins studied here. Finally, we mention that multidimensional projection techniques may significantly improve our analysis, especially for proteins that exhibit multiple distinct folding or unfolding pathways, such as NTL9 and the WW-Domain (24). However, this is beyond the scope of our present investigation.

Methods

In the document, *SI Appendix*, section 1, we present details for the molecular dynamics simulations, including various relevant simulation parameters. Additionally, we include a range of measured time-scales and other extracted

quantities, and we compare results for our analysis of reaction times to those from Lindorff-Larsen et al. (24).

The Fraction of Native Contacts Reaction Coordinate. For each protein, we project the backbone C_α atomic positions from the all-atom trajectories taken from ref. 24 onto the fraction of native contacts reaction coordinate Q, evaluated with a soft cut-off potential (36). The evaluation of $Q(t)$ requires a reference state, which we take to be the native state for each trajectory. To evaluate the native state, we follow previous implementations and select from amongst the member states of the trajectories (as opposed to using, for example, the PDB entry). The approach is similar to that used by Lindorff-Larsen et al. (24) and Best et al. (36), which follows from ref. 65. Briefly, we sample a subset of evenly spaced states from the full trajectory. For each pair of states, we calculate the corresponding root-mean-squared deviation (RMSD) between the two states. If the RMSD between two states is less than 0.2 nm, then we place the pair into a list. We assign the state that has the most listed pairs satisfying the RMSD condition as the native state for a given protein. The native states for each protein are displayed in Fig. 1*A*. Note that for proteins with more than one independent trajectory segments, we select a single native state from amongst all trajectory segments, which we then use for all segments. In the native state, we define all C_α pairs that are separated by at least 5 residues in the primary sequence and which are separated by less than 0.9 nm in Cartesian distance, as the native contacts. Each protein will have N_{nc} native contacts. $\mathbf{s}_{ij}^0$ are the separation vectors for all native contact pairs in the native state, which have magnitudes $s_{ij}^0 = \sqrt{\mathbf{s}_{ij}^0 \cdot \mathbf{s}_{ij}^0}$. $\mathbf{s}_{ij}(t)$ are the separation vectors for all native contact pairs at each time, with magnitudes $s_{ij}(t) = \sqrt{\mathbf{s}_{ij}(t) \cdot \mathbf{s}_{ij}(t)}$. This gives the fraction of the native contacts that are deemed to be in contact at time t as

$$Q(t) = \frac{1}{N_{nc}} \sum_{i<j} \frac{1}{1 + e^{\beta(s_{ij}(t) - \gamma s_{ij}^0)}}, \qquad \text{[M1]}$$

where the summation indices i and j are only for native contact pairs. Here, we set the parameters such that $\beta = 30 \text{ nm}^{-1}$ and $\gamma = 1.6$.

Non-Markovian Correction Factor ξ_{noMar}. From Kappler et al. (13) and Lavacchi et al. (14), we describe multiexponential memory dependent barrier crossing times as a sum of contributions from M overdamped contributions $\{\tau_{OD}^i\}$, and M energy-diffusion contributions $\{\tau_{ED}^i\}$, where $i = 1, 2, ..., M$. The individual contributions are defined as follows. For the overdamped contributions, we have

$$\tau_{OD}^i = \tau_D \frac{\gamma_i}{\gamma} \frac{e^{\beta U_0}}{\beta U_0}$$
$$\times \left[\frac{\pi}{2\sqrt{2}} \frac{1}{1 + 10\beta U_0 \tau_i / \tau_D} + \sqrt{\beta U_0 \frac{\tau_m}{\tau_D}} \right], \qquad \text{[M2]}$$

and for the energy-diffusion contributions

$$\tau_{ED}^i = \tau_D \frac{\gamma}{\gamma_i} \frac{e^{\beta U_0}}{\beta U_0}$$
$$\times \left[\frac{\tau_m}{\tau_D} + 4\beta U_0 \left(\frac{\tau_i}{\tau_D} \right)^2 + \sqrt{\beta U_0 \frac{\tau_m}{\tau_D}} \right]. \qquad \text{[M3]}$$

Here, $\beta = 1/k_B T$. We combine Eqs. **M2** and **M3** such that the predicted mean passage times are given by

$$\tau_{MFP}^H(\tau_D, U_0, \{\gamma_i\}, \{\tau_i\}, \tau_m)$$
$$= \sum_{i=1}^{M} \tau_{OD}^i + \left[\sum_{i=1}^{M} \frac{1}{\tau_{ED}^i} \right]^{-1}. \qquad \text{[M4]}$$

$\{\gamma_i\}$ and $\{\tau_i\}$ are the sets of M amplitudes and time-scales that appear in Eqs. **M2** and **M3**. The total friction γ that appears in Eqs. **M2** and **M3** is accounted for since $\gamma = \sum_{i=1}^{M} \gamma_i$. The overdamped Markovian limit is achieved by setting all memory time scales and inertial times equal to 0 and is given by

$$\tau_{MFP}^H(\tau_D, U_0, \{\gamma_i\}, 0, 0) = \tau_D \pi e^{\beta U_0} / 2\beta U_0 \sqrt{2}, \qquad \text{[M5]}$$

leading to the non-Markovian correction factor ξ_{noMar} and the non-Markovian barrier crossing time $\tau_{MFP}^{\text{noMar}}$, as given by Eqs. **4** and **5**, respectively. The α-scaling $\xi_{\text{noMar}}^\alpha$, as discussed in the main text, is described in detail in *SI Appendix, section 10*.

Grote–Hynes Theory and Barrier Crossing Predictions. The Grote–Hynes (GH) prediction for barrier crossing with frequency-dependent friction is given by ref. 11

$$\tau^{GH} = \frac{2\pi \omega_{max}}{\lambda \omega_{min}} e^{\beta U_0}, \qquad \text{[M6]}$$

where $\omega_{max} = \sqrt{|U_{max}''|/m}$ and $\omega_{min} = \sqrt{U_{min}''/m}$ are the barrier frequencies at the free energy maximum and minimum, assumed to be parabolic such that U_{max}'' and U_{min}'' are the quadratic curvatures at those locations. λ is the barrier reactive frequency determined by solving the Grote–Hynes equation

$$\lambda^2 + \lambda \frac{\tilde{\Gamma}(\lambda)}{m} = \omega_{max}^2. \qquad \text{[M7]}$$

$\tilde{\Gamma}(\lambda)$ is the Laplace transform of the friction memory kernel, given by

$$\tilde{\Gamma}(\lambda) = \int_0^\infty \Gamma(t') e^{-\lambda t'} dt'. \qquad \text{[M8]}$$

In the Markovian limit, i.e., when $\Gamma(t) = \gamma \delta(t)$, it can be shown that Eq. **M6** reduces to the well-known Kramers (Kr) prediction for a massive particle in the medium-to-high friction regime

$$\tau_{Kr}^{GH} = \left[\sqrt{\frac{\gamma^2}{4m^2} + \omega_{max}^2} - \frac{\gamma}{2m} \right]^{-1} \frac{2\pi \omega_{max}}{\omega_{min}} e^{\beta U_0}. \qquad \text{[M9]}$$

In the high-friction (HF) limit, i.e., $\tau_m/\tau_D \to 0$, Eq. **M9** further reduces to

$$\tau_{HF}^{GH} = \frac{2\pi \gamma}{m \omega_{min} \omega_{max}} e^{\beta U_0}. \qquad \text{[M10]}$$

An extended discussion of the Grote–Hynes calculation, including the α-scaling discussed in the main text, is given in *SI Appendix, section 10*.

Data, Materials, and Software Availability. Previously published data were used for this work (24). Processed data is available upon reasonable request.

Author affiliations: ᵃFachbereich Physik, Freie Universität Berlin, Berlin 14195, Germany

1. J. D. Bryngelson, P. G. Wolynes, Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* **93**, 6902–6915 (1989).
2. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Struct.* **21** (1995).
3. K. A. Dill, H. S. Chan, From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
4. B. Schuler, W. A. Eaton, Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008).
5. M. Hinczewski, J. C. M. Gebhardt, M. Rief, D. Thirumalai, From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4500–4505 (2013).
6. S. S. Plotkin, P. G. Wolynes, Buffed energy landscapes: Another solution to the kinetic paradoxes of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4417–4422 (2003).
7. S. Arrhenius, Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Zeitschrift für Physikalische Chemie* **4U**, 226–248 (1889).
8. H. Eyring, The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107–115 (1935).
9. H. A. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–304 (1940).
10. V. I. Mel'nikov, S. V. Meshkov, Theory of activated rate processes: Exact solution of the Kramers problem. *J. Chem. Phys.* **85**, 1018–1027 (1986).
11. R. F. Grote, J. T. Hynes, The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *J. Chem. Phys.* **73**, 2715–2732 (1980).
12. J. Kappler, J. O. Daldrop, F. N. Brünig, M. D. Boehle, R. R. Netz, Memory-induced acceleration and slowdown of barrier crossing. *J. Chem. Phys.* **148**, 14903 (2018).
13. J. Kappler, V. B. Hinrichsen, R. R. Netz, Non-Markovian barrier crossing with two-time-scale memory is dominated by the faster memory component. *Euro. Phys. J. E* **42**, 119 (2019).
14. L. Lavacchi, J. Kappler, R. R. Netz, Barrier crossing in the presence of multi-exponential memory functions with unequal friction amplitudes and memory times. *EPL* **131** (2020).
15. A. Ansari, C. M. Jones, E. R. Henry, J. Hofrichter, W. A. Eaton, The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **256**, 1796–1798 (1992).
16. I. Echeverria, D. E. Makarov, G. A. Papoian, Concerted dihedral rotations give rise to internal friction in unfolded proteins. *J. Am. Chem. Soc.* **136**, 8708–8713 (2014).
17. J. C. F. Schulz, L. Schmidt, R. B. Best, J. Dzubiella, R. R. Netz, Peptide chain dynamics in light and heavy water: Zooming in on internal friction. *J. Am. Chem. Soc.* **134**, 6273–6279 (2012).
18. D. de Sancho, A. Sirur, R. B. Best, Molecular origins of internal friction effects on protein-folding rates. *Nat. Commun.* **5**, 4307 (2014).
19. O. F. Lange, H. Grubmüller, Collective Langevin dynamics of conformational motions in proteins. *J. Chem. Phys.* **124**, 214903 (2006).
20. R. B. Best, G. Hummer, Diffusive model of protein folding dynamics with Kramers turnover in rate. *Phys. Rev. Lett.* **96**, 228104 (2006).
21. R. B. Best, G. Hummer, Coordinate-dependent diffusion in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1088–1093 (2010).
22. M. Hinczewski, Y. von Hansen, J. Dzubiella, R. R. Netz, How the diffusivity profile reduces the arbitrariness of protein folding free energies. *J. Chem. Phys.* **132**, 245103 (2010).
23. H. S. Chung, S. Piana-Agostinetti, D. E. Shaw, W. A. Eaton, Structural origin of slow diffusion in protein folding. *Science (New York, N.Y.)* **349**, 1504–1510 (2015).
24. K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, How fast-folding proteins fold. *Science* **334**, 517–520 (2011).
25. D. E. Shaw *et al.*, Millisecond-scale molecular dynamics simulations on Anton, in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* (2009), pp. 1–11.
26. D. E. Shaw *et al.*, Atomic-level characterization of the structural dynamics of proteins. *Science (New York, N.Y.)* **330**, 341–346 (2010).
27. S. S. Plotkin, P. G. Wolynes, Non-Markovian configurational diffusion and reaction coordinates for protein folding. *Phys. Rev. Lett.* **80**, 5015–5018 (1998).
28. A. M. Berezhkovskii, D. E. Makarov, Single-molecule test for Markovianity of the dynamics along a reaction coordinate. *J. Phys. Chem. Lett.* **9**, 2190–2195 (2018).
29. S. Cao, A. Montoya-Castillo, W. Wang, T. E. Markland, X. Huang, On the advantages of exploiting memory in Markov state models for biomolecular dynamics. *J. Chem. Phys.* **153**, 14105 (2020).
30. J. O. Daldrop, J. Kappler, F. N. Brünig, R. R. Netz, Butane dihedral angle dynamics in water is dominated by internal friction. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5169–5174 (2018).
31. B. Kowalik *et al.*, Memory-kernel extraction for different molecular solutes in solvents of varying viscosity in confinement. *Phys. Rev. E* **100**, 12126 (2019).
32. C. Ayaz *et al.*, Non-Markovian modeling of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023856118 (2021).
33. R. Zwanzig, Memory effects in irreversible thermodynamics. *Phys. Rev.* **124**, 983–992 (1961).
34. H. Mori, Transport, collective motion, and brownian motion. *Progr. Theor. Phys.* **33**, 423–455 (1965).
35. E. Shakhnovich, G. Farztdinov, A. M. Gutin, M. Karplus, Protein folding bottlenecks: A lattice Monte Carlo simulation. *Phys. Rev. Lett.* **67**, 1665–1668 (1991).
36. R. B. Best, G. Hummer, W. A. Eaton, Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17874–17879 (2013).
37. C. Ayaz, L. Scalfi, B. A. Dalton, R. R. Netz, Generalized Langevin equation with a nonlinear potential of mean force and nonlinear memory friction from a hybrid projection scheme. *Phys. Rev. E* **105**, 54138 (2022).
38. H. Vroylandt, P. Monmarché, Position-dependent memory kernel in generalized Langevin equations: Theory and numerical estimation. *J. Chem. Phys.* **156**, 244105 (2022).
39. H. Vroylandt, On the derivation of the generalized Langevin equation and the fluctuation-dissipation theorem. *Euro. Lett.* **140**, 62003 (2022).
40. F. N. Brünig, J. O. Daldrop, R. R. Netz, Pair-reaction dynamics in water: Competition of memory, potential shape, and inertial effects. *J. Phys. Chem. B* **126**, 10295–10304 (2022).
41. F. N. Brünig, O. Geburtig, A. von Canal, J. Kappler, R. R. Netz, Time-dependent friction effects on vibrational infrared frequencies and line shapes of liquid water. *J. Phys. Chem. B* **126**, 1579–1589 (2022).
42. J. E. Straub, M. Borkovec, B. J. Berne, Calculation of dynamic friction on intramolecular degrees of freedom. *J. Phys. Chem.* **91**, 4995–4998 (1987).
43. J. O. Daldrop, B. G. Kowalik, R. R. Netz, External potential modifies friction of molecular solutes in water. *Phys. Rev. X* **7**, 41065 (2017).
44. E. Darve, J. Solomon, A. Kia, Computing generalized Langevin equations and generalized Fokker-Planck equations. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10884–10889 (2009).
45. P. G. de Gennes, Reptation of a polymer chain in the presence of fixed obstacles. *J. Chem. Phys.* **55**, 572–579 (1971).
46. A. Alexander-Katz, H. Wada, R. R. Netz, Internal friction and nonequilibrium unfolding of polymeric globules. *Phys. Rev. Lett.* **103**, 28102 (2009).
47. K. W. Plaxco, K. T. Simons, D. Baker, Contact order, transition state placement and the refolding rates of single domain proteins11Edited by P. E. Wright. *J. Mol. Biol.* **277**, 985–994 (1998).
48. W. Zheng, R. B. Best, Reduction of all-atom protein folding dynamics to one-dimensional diffusion. *J. Phys. Chem. B* **119**, 15247–15255 (2015).
49. R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, 2001).
50. H. S. Chung, W. A. Eaton, Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* **502**, 685–688 (2013).
51. H. S. Chung, K. McHale, J. M. Louis, W. A. Eaton, Single-molecule fluorescence experiments determine protein folding transition path times. *Science* **335**, 981–984 (2012).
52. K. Neupane *et al.*, Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* **352**, 239–242 (2016).
53. K. Neupane, A. P. Manuel, M. T. Woodside, Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *Nat. Phys.* **12**, 700–703 (2016).
54. F. Sturzenegger *et al.*, Transition path times of coupled folding and binding reveal the formation of an encounter complex. *Nat. Commun.* **9**, 4708 (2018).
55. N. Zijlstra, D. Nettels, R. Satija, D. E. Makarov, B. Schuler, Transition path dynamics of a dielectric particle in a bistable optical trap. *Phys. Rev. Lett.* **125**, 146001 (2020).
56. R. Satija, A. Das, D. E. Makarov, Transition path times reveal memory effects and anomalous diffusion in the dynamics of protein folding. *J. Chem. Phys.* **147**, 152707 (2017).
57. J. O. Daldrop, W. K. Kim, R. R. Netz, Transition paths are hot. *EPL (Europhys. Lett.)* **113**, 18004 (2016).
58. G. Hummer, From transition paths to transition states and rate coefficients. *J. Chem. Phys.* **120**, 516–523 (2004).
59. M. Laleman, E. Carlon, H. Orland, Transition path time distributions. *J. Chem. Phys.* **147**, 214103 (2017).
60. E. Carlon, H. Orland, T. Sakaue, C. Vanderzande, Effect of memory and active forces on transition path time distributions. *J. Phys. Chem. B* **122**, 11186–11194 (2018).
61. P. L. Geissler, C. Dellago, D. Chandler, Kinetic pathways of ion pair dissociation in water. *J. Phys. Chem. B* **103**, 3706–3710 (1999).
62. P. G. Bolhuis, D. Chandler, C. Dellago, P. L. Geissler, Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **53**, 291–318 (2002).
63. M. Pirchi *et al.*, Photon-by-photon hidden Markov model analysis for microsecond single-molecule FRET kinetics. *J. Phys. Chem. B* **120**, 13065–13075 (2016).
64. H. Y. Aviram *et al.*, Direct observation of ultrafast large-scale dynamics of an enzyme under turnover conditions. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3243–3248 (2018).
65. X. Daura *et al.*, Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.* **38**, 236–240 (1999).