Published in final edited form as:

J Phys Chem Lett. 2018 May 03; 9(9): 2190–2195. doi:10.1021/acs.jpclett.8b00956.

Single-Molecule Test for Markovianity of the Dynamics along a Reaction Coordinate

Alexander M. Berezhkovskii[†], Dmitrii E. Makarov^{*,‡,§}

[†]Mathematical and Statistical Computing Laboratory, Office of Intramural Research, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892, United States

[‡]Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States

[§]Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712, United States

Abstract

In an effort to answer the much-debated question of whether the time evolution of common experimental observables can be described as one-dimensional diffusion in the potential of mean force, we propose a simple criterion that allows one to test whether the Markov assumption is applicable to a single-molecule trajectory $x(t)$. This test does not involve fitting of the data to any presupposed model and can be applied to experimental data with relatively low temporal resolution.

Graphical Abstract

Kinetic measurements of protein folding and other complex biomolecular process are often interpreted using the model that assumes simple diffusion along some reaction coordinate $x^{1,2}$ with a (generally coordinate-dependent) diffusion coefficient $D(x)$. This is a Markov

*Corresponding Author: makarov@cm.utexas.edu.

The authors declare no competing financial interest.

process that lacks memory. Although this model and its low-dimensional extensions successfully account for numerous experimental^{3–6} and simulational^{7–9} observations, it remains a phenomenological description. On theoretical grounds, however, memory effects are expected to be ubiquitous^{10,11} and to have important consequences on the observed dynamics and rates.^{12–21} Moreover, they are commonly observed in computer simulations, and, less commonly, in experimental studies of biomolecules.^{22–32}

Single-molecule measurements that directly probe the time evolution $x(t)$ of an experimentally measurable quantity (such as a protein's extension in single-molecule pulling studies) offer an opportunity to estimate the magnitude of memory effects,^{18,32} but no consensus yet exists regarding their importance. On one hand, several studies report that the dynamics of protein and DNA extension, as measured by single-molecule force spectroscopy, is consistent with the one-dimensional (1D) diffusion model with no memory.^{5,33} On the other hand, this model appears inadequate as a global fit of both rates and transition path times, while describing each of these data sets separately;³⁴ this finding was interpreted as evidence that a one-dimensional model without memory is insufficient.^{24,35}

A key difficulty in estimating memory effects from single-molecule trajectories is their relatively low time resolution. For this reason, previous comparisons of the 1D Markov diffusion model with experimental data have been indirect and employed fitting certain experimental observables (e.g., transition rates, distributions of transition path times, etc.). This is in contrast to trajectories obtained from molecular simulations, for which, for example, the memory kernel in the generalized Langevin equation can be directly computed,^{36–38} or the assumption of diffusive motion can be tested by examining short-time evolution of mean-square displacements.²³ In what follows we propose a simple and direct test of the Markov assumption that can be applied to any sufficiently long single-molecule trajectory measured experimentally or simulated on a computer. The test does not make any assumptions about the specific equations of motion governing the system and, importantly, does not require high temporal resolution of the observed trajectory $x(t)$. We expect that this test would be straightforward to apply to existing single-molecule data and, in anticipation of such applications, illustrate how it works by analyzing (i) a trajectory obtained by numerically integrating a generalized Langevin equation with memory and (ii) a trajectory of a protein obtained from molecular simulations.³⁹

We consider a long trajectory $x(t)$ that is bounded in space, as is usually the case for common reaction coordinates considered in protein folding and other intramolecular processes. We would like to establish whether $x(t)$ is a Markov process. To this end we pick an interval (a, b) (where $a < b$) that is sufficiently well-sampled by the trajectory. Suppose at some time a point x lying within this interval is visited by the trajectory. This event could be part of four possible scenarios (see Figure 1): The trajectory could arrive through the boundary $a(b)$, visit x , and then exit the interval through the same boundary. Or the trajectory could also arrive through the boundary $a(b)$, pass the point x , and exit through the other boundary $b(a)$. In the latter case, we say that x belongs to a transition path from a to b (b to a). Thus, all trajectories passing through x are either transition paths (from a to b or b to a) or loops that escape the interval through the same boundary through which they entered.

For each of the above events, we can define corresponding conditional probabilities $P(a \rightarrow ax)$, $P(b \rightarrow bx)$, $P(a \rightarrow bx)$, and $P(b \rightarrow ax)$. For example, $P(a \rightarrow bx)$ is the probability that a chosen point x belongs to a transition path from a to b , $P(b \rightarrow bx)$ is the probability that the trajectory passing through x is a loop entering and exiting the interval (a, b) through the boundary b , etc. In what follows, we will show that when the dynamics along x is Markovian, we have

$$\max_{a < x < b} P(a \rightarrow bx) = \max_{a < x < b} P(b \rightarrow ax) = 1/4 \quad (1)$$

$$\min_{a < x < b} [P(a \rightarrow ax) + P(b \rightarrow bx)] = 1/2 \quad (2)$$

In contrast, for non-Markov dynamics in the overdamped regime we have

$$\max_{a < x < b} P(a \rightarrow bx) = \max_{a < x < b} P(b \rightarrow ax) < 1/4 \quad (3)$$

$$\min_{a < x < b} [P(a \rightarrow ax) + P(b \rightarrow bx)] > 1/2 \quad (4)$$

Violation of the equalities of eqs 1 and 2 is, therefore, a signature of non-Markovian dynamics allowing one to assess the validity of the one-dimensional Markov diffusion model without attempting to use this model to fit the data.

Note that eq 2 follows from eq 1 and eq 4 follows from eq 3 because the sum of all four probabilities is unity, $P(a \rightarrow ax) + P(b \rightarrow bx) + P(a \rightarrow bx) + P(b \rightarrow ax) = 1$. Also note that the probabilities $P(a \rightarrow bx)$ and $P(b \rightarrow ax)$ are identical because of the time reversal symmetry assumed here; indeed, upon time reversal, any transition path from a to b becomes a transition path from b to a and vice versa. Because the direct and the time-reversed trajectories are statistically indistinguishable,⁴⁰ the above two probabilities must be the same. For these reasons, throughout the rest of this Letter, we focus on the behavior of the probability $P(a \rightarrow bx)$.

To explain the physical basis of the inequalities of eqs 3 and 4, consider first the simplest example of Markovian dynamics described by the overdamped Langevin equation:

$$0 = -U'(x) - \gamma_x \dot{x} + \zeta_x(t) \quad (5)$$

where $U(x)$ is the potential of mean force, γ_x a friction coefficient, and ζ_x a Gaussian random force with zero mean, which satisfies the fluctuation–dissipation theorem (FDT), $\langle \zeta_x(t) \zeta_x(t') \rangle = 2\gamma_x k_B T \delta(t - t')$, with the angular brackets denoting averaging over realizations of the random force. In this case, the probability $P(a \rightarrow bx)$ is

$$P(a \rightarrow b|x) = \Phi_a(x)\Phi_b(x) = \Phi_b(x)[1 - \Phi_b(x)] \quad (6)$$

Here $\Phi_{a(b)}(x)$ is the splitting probability⁴¹ (also known as committor^{42–46} or PFOLD⁴⁷) that the system starting in x will reach the boundary a (b) before b (a). Equation 6 takes advantage of the fact that, in a Markov case, $P(a \rightarrow b|x)$ is the product of two probabilities: (i) the probability $\Phi_b(x)$ that the trajectory exits the interval (a, b) through the right boundary and (ii) the probability $\Phi_a(x)$ that the trajectory enters through the left boundary, where time reversal symmetry of the Langevin dynamics has been taken into account. In Figure 1, the exiting and entering segments of the transition path are colored red and green.

The splitting probability $\Phi_b(x)$ is a monotonically increasing function with $\Phi_b(a) = 0$ and $\Phi_b(b) = 1$; therefore, $P(a \rightarrow b|x)$, eq 6, is a nonmonotonic function of x that vanishes at the boundaries ($P(a \rightarrow ba) = P(a \rightarrow bb) = 0$) and has a maximum at $x = x^*$ where $\Phi_a(x^*) = \Phi_b(x^*) = 1/2$:

$$\max_{a < x < b} P(a \rightarrow b|x) = P(a \rightarrow b|x^*) = 1/4 \quad (7)$$

As the simplest example of non-Markovian dynamics, consider now the overdamped generalized Langevin equation (GLE) of the form

$$0 = -U'(x) - \gamma_x \dot{x} - k \int_0^t e^{-(t-t')/\tau} \dot{x}(t') dt' + \zeta(t) \quad (8)$$

where $\zeta(t)$ is a Gaussian random force with zero mean, and where the FDT now takes the form $\langle \zeta(t)\zeta(t') \rangle = 2\gamma_x k_B T \delta(t-t') + k_B T k e^{-|t-t'|/\tau}$. A key observation is that the dynamics described by the GLE with a single-exponential memory kernel is equivalent to two-dimensional memoryless Langevin dynamics with one additional degree of freedom y .^{48,49}

$$\begin{aligned} 0 &= -\partial U_{2D}(x, y)/\partial x - \gamma_x \dot{x} + \zeta_x(t) \\ 0 &= -\partial U_{2D}(x, y)/\partial y - \gamma_y \dot{y} + \zeta_y(t) \end{aligned} \quad (9)$$

where $U_{2D}(x, y) = U(x) + k(y-x)^2/2$. The two components of the random force, $\zeta_x(t)$ and $\zeta_y(t)$, are uncorrelated Gaussian forces with zero means; $\zeta_x(t)$ satisfies the FDT given below eq 5, while $\zeta_y(t)$ satisfies $\langle \zeta_y(t)\zeta_y(t') \rangle = 2\gamma_y k_B T \delta(t-t')$, where $\gamma_y = k\tau$ is the friction coefficient along y .

Because the dynamics described by eq 9 is Markovian, an equivalent of eq 6 holds in the extended space, where the splitting probability is viewed as a function of x and y .

$$\begin{aligned}\tilde{P}(a \rightarrow b|x, y) &= \tilde{\Phi}_a(x, y)\tilde{\Phi}_b(x, y) \quad (10) \\ &= \tilde{\Phi}_b(x, y)[1 - \tilde{\Phi}_b(x, y)]\end{aligned}$$

Here, the tilde above a quantity indicates that it is a function of the coordinates in the extended space. Whether a fragment of a trajectory in the xy -space is a transition path is determined exclusively by the dynamics of its x -component, which is the experimental observable. We can then obtain the probability $P(a \rightarrow b|x)$ from $\tilde{P}(a \rightarrow b|x, y)$ by averaging over y . More precisely, consider the joint probability density $\tilde{\rho}(x, y, a \rightarrow b)$ that the system is found at (x, y) while on a transition path (from a to b) and similarly the joint probability

$$\rho(x, a \rightarrow b) = \int dy \tilde{\rho}(x, y, a \rightarrow b) \quad (11)$$

of finding the system at x while on a transition path. By definition of the conditional probability, we have

$$\tilde{\rho}(x, y, a \rightarrow b) = \tilde{P}(a \rightarrow b|x, y)\tilde{p}_{\text{eq}}(x, y) \quad (12)$$

where $\tilde{p}_{\text{eq}}(x, y)$ is the equilibrium distribution of x and y .

Similarly, we have

$$\rho(x, a \rightarrow b) = P(a \rightarrow b|x)p_{\text{eq}}(x) \quad (13)$$

for the corresponding quantities defined in x -space only. Combining eqs 11–13, we obtain

$$\begin{aligned}P(a \rightarrow b|x) &= \int dy \tilde{P}(a \rightarrow b|x, y)p_{\text{eq}}(y|x) \quad (14) \\ &= \langle \tilde{P}(a \rightarrow b|x, y) \rangle_x\end{aligned}$$

Here, angular brackets indicate averaging over y at fixed x

$$\langle \dots \rangle_x = \int dy (\dots)p_{\text{eq}}(y|x) \quad (15)$$

weighted with the local in x equilibrium density of y

$$p_{\text{eq}}(y|x) = \frac{\tilde{p}_{\text{eq}}(x, y)}{p_{\text{eq}}(x)} = \frac{\tilde{p}_{\text{eq}}(x, y)}{\int dy \tilde{p}_{\text{eq}}(x, y)} \quad (16)$$

Using the same definition of the average, we also find

$$\Phi_b(x) = \langle \tilde{\Phi}_b(x, y) \rangle_x \quad (17)$$

Finally, using eqs 10 and 14 and the fact that $\langle \tilde{\Phi}_b^2(x, y) \rangle_x \geq \langle \tilde{\Phi}_b(x, y) \rangle_x^2$ we obtain

$$\begin{aligned} P(a \rightarrow b|x) &= \langle \tilde{\Phi}_b(x, y) \rangle_x - \langle \tilde{\Phi}_b^2(x, y) \rangle_x \\ &\leq \langle \tilde{\Phi}_b(x, y) \rangle_x - \langle \tilde{\Phi}_b(x, y) \rangle_x^2 = \Phi_b(x) - \Phi_b^2(x) \end{aligned} \quad (18)$$

The maximum value of $\Phi_b(x) - \Phi_b^2(x)$ achieved on the interval (a, b) is $1/4$ (cf. eq 7). The left-hand side (lhs) of eq 18, therefore, cannot exceed $1/4$, but it can be less. In fact, the only way the lhs of eq 18 can attain its maximum value of $1/4$ at some $x = x^*$ is if $\tilde{\Phi}_b(x^*, y) = 1/2$ for any y . This implies that the dynamics along x and y are decoupled and, therefore, the dynamics along x is Markovian. The inequality stated by eq 3 is thus a fingerprint of non-Markovian dynamics.

Figure 2 illustrates the inequality of eq 3 for the system obeying eq 8. The probability $P(a \rightarrow b|x)$ was computed from a long simulated trajectory $x(t)$ using the identity⁴⁵

$$P(a \rightarrow b|x) = \frac{p(x|a \rightarrow b)P(a \rightarrow b)}{p_{\text{eq}}(x)} \quad (19)$$

where the equilibrium distribution $p_{\text{eq}}(x)$ is the distribution of all the points belonging to the trajectory $x(t)$, $p(x|a \rightarrow b)$ the distribution of all the points that belong to transition paths from a to b only, and $P(a \rightarrow b)$ the fraction of time the system is found on a transition path from a to b . For short memory times τ , we expect the dynamics described by eq 8 to be close to the Markov limit. Indeed, for $\tau = 0.05$, the function $P(a \rightarrow b|x)$ reaches a maximum value that is close to $1/4$. As the memory time increases, however, the non-Markovian character of the dynamics is readily detected by observing that the maximum of $P(a \rightarrow b|x)$ becomes less than $1/4$.

Although eq 3 was derived for the GLE with a single-exponential memory kernel, eq 8, it holds for an overdamped GLE with a memory kernel of arbitrary form. Indeed, introducing a set of auxiliary degrees of freedom y_n coupled to x via a potential $k_n(y_n - x)^2/2$ results in a GLE of the form similar to eq 8, but with memory kernel equal to the sum of exponentials

$$0 = -U'(x) - \gamma_x \dot{x} - \int_0^t \sum_n k_n e^{-(t-t')/\tau_n} \dot{x}(t') dt' + \zeta(t) \quad (20)$$

The derivation of eq 3 can be generalized to such a multidimensional system. By allowing a continuum of auxiliary degrees of freedom, we can now extend the proof to a GLE with arbitrary memory kernel. Moreover, as seen from its proof, eq 3 does not require that the time evolution of x obeys a GLE as long as it can be viewed as a projection of some higher-dimensional Markov process in an extended configuration space. We further emphasize that our results are valid regardless of how the boundaries a and b are chosen. In particular, it is not assumed that the transition region $a < x < b$ separates two basins of attraction and contains a barrier; indeed, no such basins of attraction exist in the potential chosen in the example of Figure 2.

As a more realistic case, we now apply our test of the Markov assumption to the folding of a designed double norleucine mutant of 35-residue villin headpiece C-terminal fragment (HP-35) protein, for which a long trajectory displaying many folding–unfolding transitions has been simulated by Piana, Lindorff-Larsen, and Shaw.³⁹ Following their work, the root-mean-square deviation from the native structure is used here as a dynamic variable $x(t)$. Based on the decay of its autocorrelation function and on the distributions of transition-path times, an earlier study²⁴ concluded that the dynamics of this coordinate is strongly non-Markovian. Consistent with these conclusions, $P(a \rightarrow b|x)$ peaks at a value of ~ 0.11 , much lower than the 0.25 expected for a Markov process (Figure 3). Without even attempting to model the data as a one-dimensional diffusion process, with position-dependent or position-independent diffusion coefficient, one establishes that this would not be an adequate model for the folding dynamics along x .

In the preceding discussion, we considered dynamics in configuration space. In particular, eqs 3 and 4 were derived by assuming that the dynamics of x is a projection of Markovian dynamics occurring in some extended configuration space. This is a common assumption for many biophysical phenomena that take place in solution, where overdamped dynamics is a good approximation for most relevant degrees of freedom. When inertial effects are not negligible, however, dynamics in phase space must be considered. Here we briefly discuss how inertial effects are reflected in $P(a \rightarrow b|x)$, focusing on the case where the dynamics along x alone is not Markovian, but the dynamics in the phase space (x, v) is Markovian, as exemplified by the (nonoverdamped) Langevin equation

$$m\ddot{x} = m\dot{v} = -U'(x) - \gamma_x \dot{x} + \zeta_x(t) \quad (21)$$

where m is a mass. The equivalent of eq 10 written in phase space is then

$$\begin{aligned}\tilde{P}(a \rightarrow b|x, v) &= \tilde{\Phi}_a(x, -v)\tilde{\Phi}_b(x, v) \quad (22) \\ &= \tilde{\Phi}_b(x, v)[1 - \tilde{\Phi}_b(x, -v)]\end{aligned}$$

There is an essential difference between eqs 10 and 22: to form a transition path from two trajectory segments originating from x , the velocity of one of them must be reversed (cf. green and red trajectory segments in Figure 1). This renders our previous analysis, which led to the conclusion that the maximum value of $P(a \rightarrow b|x)$ cannot exceed 1/4, inapplicable. For a symmetric potential, $U(x) = U(-x)$, and for a symmetric transition region, $a = -b$, one can, in fact, show that $P(a \rightarrow b|0)$ always exceeds 1/4! Indeed, because of symmetry, we have $\tilde{\Phi}_a(0, -v) = 1 - \tilde{\Phi}_b(0, -v) = \tilde{\Phi}_b(0, v)$, and therefore

$$\tilde{P}(a \rightarrow b|0, v) = \tilde{\Phi}_b^2(0, v) \quad (23)$$

By averaging $\tilde{P}(a \rightarrow b|0, v)$ over the velocity (analogous to averaging over y in eq 14), we obtain

$$\begin{aligned}P(a \rightarrow b|0) &= \langle \tilde{P}(a \rightarrow b|0, v) \rangle_v = \langle \tilde{\Phi}_b^2(0, v) \rangle_v \quad (24) \\ &\geq \langle \tilde{\Phi}_b(0, v) \rangle_v^2 = \Phi_b^2(0) = \frac{1}{4}\end{aligned}$$

because symmetry dictates that $\Phi_b(0) = 1/2$. Finding values of $P(a \rightarrow b|x)$ that exceed 1/4, therefore, would indicate importance of inertial effects (Figure 4).

In conclusion, we showed that the probability $P(a \rightarrow b|x) = P(b \rightarrow a|x)$ that a given point x belongs to a transition path between arbitrarily chosen boundaries a and b and the probability $P(a \rightarrow ax) + P(b \rightarrow bx)$ that x belongs to a loop (i.e., not a transition path) obey inequalities (eqs 3 and 4), which turn into equalities (eqs 1 and 2) only in the Markov case. Testing whether these equalities are satisfied for an experimentally observable reaction coordinate $x(t)$ directly shows whether its underlying dynamics is Markovian. Because the above probabilities do not depend on time explicitly, this Markovianity test can be applied to data that have limited time resolution, as is often the case with experimental signals.

ACKNOWLEDGMENTS

Discussions with Ron Elber and Mauro Mugnai are gratefully acknowledged. We thank D.E. Shaw Research for providing access to their molecular dynamics trajectories, Stefano Piana and Kresten Lindorff-Larsen for clarifications regarding their data analysis, and Atanu Das for help with data analysis. A.M.B. was supported by the Intramural Research Program of the NIH, Center for Information Technology. D.E.M. was supported by the Robert A. Welch Foundation (Grant No. F-1514) and the National Science Foundation (Grant No. CHE 1566001).

REFERENCES

- (1). Szabo A; Schulten K; Schulten Z First passage time approach to diffusion controlled reactions. J. Chem. Phys 1980, 72, 4350.

- (2). Socci ND; Onuchic JN; Wolynes PG Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys* 1996, 104, 5860–5868.
- (3). Gopich IV; Nettels D; Schuler B; Szabo A Protein dynamics from single-molecule fluorescence intensity correlation functions. *J. Chem. Phys* 2009, 131, 095102. [PubMed: 19739874]
- (4). Schuler B; Eaton WA Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol* 2008, 18, 16–26. [PubMed: 18221865]
- (5). Neupane K; Manuel AP; Woodside M Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *Nat. Phys* 2016, 12, 700–703.
- (6). Dudko OK; Hummer G; Szabo A Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U. S. A* 2008, 105, 15755–60. [PubMed: 18852468]
- (7). Mugnai ML; Elber R Extracting the diffusion tensor from molecular dynamics simulation with Milestoning. *J. Chem. Phys* 2015, 142, 014105. [PubMed: 25573551]
- (8). Best RB; Hummer G Diffusion models of protein folding. *Phys. Chem. Chem. Phys* 2011, 13, 16902–11. [PubMed: 21842082]
- (9). Best RB; Hummer G Coordinate-dependent diffusion in protein folding. *Proc. Natl. Acad. Sci. U. S. A* 2010, 107, 1088–93. [PubMed: 20080558]
- (10). Zwanzig R Nonequilibrium Statistical Mechanics; Oxford University Press: 2001.
- (11). van Kampen NG Remarks on Non-Markov Processes. *Braz. J. Phys* 1998, 28, 90–96.
- (12). Grote RF; Hynes JT The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *J. Chem. Phys* 1980, 73, 2715–2732.
- (13). Grote RF; Hynes JT Reactive modes in condensed phase reactions. *J. Chem. Phys* 1981, 74, 4465–4475.
- (14). Portman JJ; Takada S; Wolynes PG Microscopic theory of protein folding rates. II. Local reaction coordinates. *J. Chem. Phys* 2001, 114, 5082–5096.
- (15). Makarov DE Interplay of non-Markov and internal friction effects in the barrier crossing kinetics of unfolded proteins. *J. Chem. Phys* 2013, 138, 014102. [PubMed: 23298023]
- (16). Plotkin SS; Wolynes PG Non-Markovian Configurational Diffusion and Reaction Coordinates for Protein Folding. *Phys. Rev. Lett* 1998, 80, 5015.
- (17). Panja D Generalized Langevin equation formulation for anomalous polymer dynamics. *J. Stat. Mech.: Theory Exp* 2010, 2010, L02001.
- (18). Presse S; Lee J; Dill KA Extracting conformational memory from single-molecule kinetic data. *J. Phys. Chem. B* 2013, 117, 495–502. [PubMed: 23259771]
- (19). Pollak E; Talkner P Reaction rate theory: what it was, where is it today, and where is it going? *Chaos* 2005, 15, 026116.
- (20). Hanggi P; Talkner P; Borkovec M 50 years fter Kramers. *Rev. Mod. Phys* 1990, 62, 251.
- (21). Pollak E Theory of activated rate-processes - a new derivation of Kramers expression. *J. Chem. Phys* 1986, 85, 865–867.
- (22). Sangha AK; Keyes T Proteins fold by subdiffusion of the order parameter. *J. Phys. Chem. B* 2009, 113, 15886–94. [PubMed: 19902909]
- (23). Avdoshenko SM; Das A; Satija R; Papoian GA; Makarov DE Theoretical and computational validation of the Kuhn barrier friction mechanism in unfolded proteins. *Sci. Rep* 2017, 7, 269. [PubMed: 28325911]
- (24). Satija R; Das A; Makarov DE Transition path times reveal memory effects and anomalous diffusion in the dynamics of protein folding. *J. Chem. Phys* 2017, 147, 152707. [PubMed: 29055292]
- (25). Luo G; Andricioaei I; Xie XS; Karplus M Dynamic distance disorder in proteins is caused by trapping. *J. Phys. Chem. B* 2006, 110, 9363–7. [PubMed: 16686476]
- (26). Cote Y; Senet P; Delarue P; Maisuradze GG; Scheraga HA Anomalous diffusion and dynamical correlation between the side chains and the main chain of proteins in their native state. *Proc. Natl. Acad. Sci. U. S. A* 2012, 109, 10346–51. [PubMed: 22689963]

- (27). Volk M; Milanesi L; Waltho JP; Hunter CA; Beddard GS The roughness of the protein energy landscape results in anomalous diffusion of the polypeptide backbone. *Phys. Chem. Chem. Phys* 2015, 17, 762–82. [PubMed: 25412176]
- (28). Krivov SV Is protein folding sub-diffusive? *PLoS Comput. Biol* 2010, 6, e1000921. [PubMed: 20862361]
- (29). Elber R Perspective: Computer simulations of long time dynamics. *J. Chem. Phys* 2016, 144, 060901. [PubMed: 26874473]
- (30). Grossman-Haham I; Rosenblum G; Namani T; Hofmann H Slow domain reconfiguration causes power-law kinetics in a two-state enzyme. *Proc. Natl. Acad. Sci. U. S. A* 2018, 115, 513–518. [PubMed: 29298911]
- (31). Gowdy J; Batchelor M; Neelov I; Paci E Nonexponential kinetics of loop formation in proteins and peptides: a signature of rugged free energy landscapes? *J. Phys. Chem. B* 2017, 121, 9518–9525. [PubMed: 28950699]
- (32). Presse S; Peterson J; Lee J; Elms P; MacCallum JL; Marqusee S; Bustamante C; Dill K Single molecule conformational memory extraction: p5ab RNA hairpin. *J. Phys. Chem. B* 2014, 118, 6597–603. [PubMed: 24898871]
- (33). Neupane K; Manuel AP; Lambert J; Woodside MT Transition-path probability as a test of reaction-coordinate quality reveals DNA hairpin folding is a one-dimensional diffusive process. *J. Phys. Chem. Lett* 2015, 6, 1005–1010. [PubMed: 26262860]
- (34). Neupane K; Foster DA; Dee DR; Yu H; Wang F; Woodside MT Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* 2016, 352, 239–42. [PubMed: 27124461]
- (35). Makarov DE Reconciling transition path time and rate measurements in reactions with large entropic barriers. *J. Chem. Phys* 2017, 146, 071101. [PubMed: 28228023]
- (36). Straub JE; Borkovec M; Berne BJ Calculation of dynamic friction on intramolecular degrees of freedom. *J. Phys. Chem* 1987, 91, 4995–4998.
- (37). Daldrop JO; Kowalik BG; Netz RR External potential modifies friction of molecular solutes in water. *Phys. Rev. X* 2017, 7, 041065.
- (38). Berkowitz M; Morgan JD; Kouri DJ; McCammon JA Memory kernels from molecular dynamics. *J. Chem. Phys* 1981, 75, 2462.
- (39). Piana S; Lindorff-Larsen K; Shaw DE Protein folding kinetics and thermodynamics from atomistic simulation. *Proc. Natl. Acad. Sci. U. S. A* 2012, 109, 17845–50. [PubMed: 22822217]
- (40). Makarov DE Single Molecule Science: Physical Principles and Models; CRC Press, Taylor & Francis Group: Boca Raton, FL, 2015; p 214.
- (41). Gardiner CW Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences; Springer-Verlag: Berlin, 1983.
- (42). E W; Vanden-Eijnden E Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem* 2010, 61, 391–420. [PubMed: 18999998]
- (43). E. W; Vanden-Eijnden E Towards a Theory of Transition Paths. *J. Stat. Phys* 2006, 123, 503–523.
- (44). Vanden-Eijnden E Transition path theory In *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*; Ferrario M, M., Ciccotti G, Binder K, Eds.; Springer: 2006.
- (45). Hummer G From transition paths to transition states and rate coefficients. *J. Chem. Phys* 2004, 120, 516–523. [PubMed: 15267886]
- (46). Peters B Reaction Theory and Rare Events; Elsevier: 2017; p 634.
- (47). Du R; Pande VS; Grosberg AY; Tanaka T; Shakhnovich ES On the transition coordinate for protein folding. *J. Chem. Phys* 1998, 108, 334–350.
- (48). Straub JE; Berne BJ Energy diffusion in many-dimensional Markovian systems: The consequences of competition between inter- and intramolecular vibrational energy transfer. *J. Chem. Phys* 1986, 85, 2999–3006.
- (49). Straub JE; Borkovec M; Berne BJ Non-Markovian activated rate processes: Comparison of current theories with numerical simulation data. *J. Chem. Phys* 1986, 84, 1788–1794.

Figure 1.

Examples of trajectories crossing some intermediate point x that are transition paths from a to b (the combined trajectory composed of the green and red lines) and loops (green and red lines). A transition path from a to b can be thought of as a combination of a forward trajectory originating from some intermediate point x and proceeding to b and a time-reversed backward trajectory originating at x and going to a .

Figure 2.

Non-Markovian character of GLE dynamics, eq 8, readily observed in the behavior of $P(a \rightarrow b|x)$ (right). Here, the potential (measured in units of thermal energy) is given by $U(x) = x^8 - x + 0.5 \sin(7x)$ and is depicted on the left, with the boundaries a and b shown as vertical dashed lines. Other parameters are $\gamma_x = 1$; $k = 10$; and $\tau = \gamma_y/k = 0.05$ (red), 0.5 (green), and 5.0 (blue).

Figure 3. Probability $P(a \rightarrow b|x)$, where x is the root-mean-square deviation from the native structure for a designed HP-35 protein, peaking at value that is much less than the $1/4$ expected in the Markov case. This immediately rules out one-dimensional diffusion, with a position-dependent or position-independent diffusion coefficient, as an adequate model for the dynamics along x . In this analysis, the interval (a, b) was chosen exactly as the transition region in previous studies.^{24,39}

Figure 4.

Peak value of $P(a \rightarrow b|x)$ (right panel) exceeding $1/4$ for Langevin dynamics (eq 21) in a symmetric potential $\beta U(x) = 2(x^2 - 1)^2$ with a mass $m = 1$. The potential is depicted in the left panel, with the boundaries a and b shown as vertical dashed lines. For low friction ($\gamma_x = 0.01$, red), the maximum of $P(a \rightarrow b|x)$ approaches $1/2$ as expected for ballistic dynamics (at x coinciding with the barrier top, $P(a \rightarrow b|x)$ equals the probability to have a positive velocity, which is $1/2$). As the friction coefficient is increased ($\gamma_x = 1$ green, $\gamma_x = 10$, blue), the overdamped Markov limit is approached, where $\max_{a < x < b} P(a \rightarrow b|x) = 1/4$.