On distributions of barrier crossing times as observed in single-molecule studies of biomolecules

Running title: Barrier Crossing Times in Biomolecules

Alexander M. Berezhkovskii[1], and Dmitrii E. Makarov[2]

[1]Mathematical and Statistical Computing Laboratory, Office of Intramural Research, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892

[2]Department of Chemistry and Biochemistry and Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712

Abstract

Single-molecule experiments that monitor time evolution of molecular observables in real time have expanded beyond measuring transition rates toward measuring distributions of times of various molecular events. Of particular interest is the first passage time for making a transition from one molecular configuration (a) to another (b) and conditional first passage times such as the transition path time, which is the first passage time from a to b conditional upon not leaving the transition region intervening between a and b. Another experimentally accessible (but not yet studied experimentally) observable is the conditional exit time, i.e., the time to leave the transition region through a specified boundary. The distributions of such times contain a wealth of mechanistic information about the transitions in question. Here we use the first and the second (and, if desired, higher) moments of these distributions to characterize their relative width, for the model where the experimental observable undergoes Brownian motion in a potential of mean force. We show that, while the distributions of transition path times are always narrower than exponential (in that the ratio of the standard deviation to the distribution's mean is always less than 1), distributions of first passage times and of conditional exit times can be either narrow or broad, in some cases displaying long power-law tails. The

conditional exit time studied here provides a generalization of the transition path time that also allows one to characterize the temporal scales of failed barrier crossing attempts.

Why it matters.

Single-molecule measurements directly visualize dynamics of proteins, DNA, and molecular motors as they cross activation barriers and carry out their biological functions, but, in contrast to the complexity of the molecular motion involving many atoms, single-molecule signals are inherently low-dimensional, reporting only on a few degrees of freedom. Towards solving the inverse problem of learning about the underlying dynamics from single-molecule signals, this paper discusses distributions of barrier crossing timescales expected for the important model where the time evolution of the single-molecule signal can be described as Brownian dynamics.

1. Introduction.

Learning about dynamics of molecules from experimental data is a difficult inverse problem. Molecular trajectories occur in a 3*N*-dimensional space, where *N* is the number of participating atoms (including those of the solvent in the case of biomolecular processes). Experimental measurements, in contrast, typically provide a low-dimensional signal. In single-molecule studies, this signal is the time-dependence $x(t)$ of a single variable x (or, in some of the state-of-the-art studies, two variables) such as the distance between two molecular groups, or it is a list of photon arrival times. For several decades, the common approach to this inverse problem has been to postulate a specific phenomenological model – that the trajectory $x(t)$ is described as one-dimensional diffusion in a potential of mean force – and to deduce the parameters of this model (such as the diffusivity *D*) by fitting experimental observations. But recent improvements in the spatial and temporal resolution of single-molecule experiments offer an opportunity to both refine such models (e.g., by deducing more accurate potentials or by allowing for coordinate-dependent diffusivity) and to move beyond phenomenological models toward more accurate, data-driven ones[1, 2]. Importantly, recent theoretical advances (to be discussed below) show that certain details about the underlying multidimensional dynamics are encoded in the low-dimensional signal $x(t)$, thereby suggesting that progress toward solving the above-mentioned inverse problem is possible.

A phenomenon of particular significance for the field of biomolecular dynamics is that of large conformational transitions attained via the crossing of a free energy barrier (Fig. 1). More precisely, suppose that the experimentally determined potential of mean force $U(x)$ (defined through the requirement that the equilibrium distribution $p_{eq}(x)$ of the coordinate x is the Boltzmann distribution in this potential, $p_{eq}(x) \propto \exp\left\{-\frac{U(x)}{k_B T}\right\}$) has a double-well shape with two wells separated by a barrier, each well corresponding to distinct molecular species. The experimental observable then provides a natural (albeit not necessarily optimal) choice of the reaction coordinate for the inter-well transitions, or "chemical reactions". Biomolecular folding is a common example of such a process. With a more general potential $U(x)$, this picture can also be extended to such phenomena as biomolecular binding[3, 4], or even to nonequilibrium processes such as the stroke of a molecular motor[5, 6].

Although the theory presented below is general and does not rely on any assumptions about the shape of $U(x)$, for concreteness let us assume that we are dealing with biomolecular folding. The coordinate x, then, may correspond to the molecule's extension, as in single-molecule force spectroscopy[7], or may be related to the distance between two dyes as in fluorescence resonance energy transfer (FRET) studies[8]. The potential $U(x)$ then typically has two wells corresponding to the molecule's unfolded and folded states.

There are several timescales of interest in this case. If the barrier height is much greater than the thermal energy $k_B T$ then the typical time for a transition between the two wells is much longer than the equilibration time within each well, and it is meaningful to define such equilibration timescales, τ_A and τ_B, for each well. In the field of protein folding, the latter timescales, usually referred to as the "reconfiguration" times[8-12], have attracted considerable experimental attention, particularly for the unfolded state. It can be measured, e.g., using single-molecule fluorescence correlation spectroscopy[13-18].

Reconfiguration timescale is a dynamical characteristic of thermal fluctuations around in a (meta)stable thermodynamic state. To understand the kinetics of transitions between such states, one needs information about the dynamics of barrier crossing. If we choose some point a within the folded state and a point b within the unfolded state, then the first passage time $t_{FP}(a \to b)$ to go from a to b provides a measure of the overall time required to accomplish the transition. In general, this time of course will depend on the choice of the points a and b, but if these points are separated by a barrier that is much higher than $k_B T$ then the average value of this time is insensitive to the choice of the initial and final points. Indeed, typical trajectories originating from a or b will relax into the adjacent free energy well on a timescale which is much shorter than the escape time from the well, and the much longer escape time from the well will provide a dominant contribution to the mean first passage time. In this limit, the inverse mean first passage time defines the rate (coefficient) k for the transition (see, e.g., refs.[19, 20]). Moreover, for a double-well free energy landscape the distribution of this time is close to exponential, as expected for first-order chemical kinetics.

To obtain further mechanistic insight into the dynamics of inter-well transitions, one can dissect a trajectory that takes the system from a to b into pieces where the system dwells in the initial state and where it is caught mid-way between the initial and the final state. The latter are arguably most interesting as they determine the transition "mechanism". To make this idea precise, we call the interval (a, b) the "transition region". Typically, its boundaries are chosen to contain the free energy barrier (Fig. 1). Consider now a typical trajectory that crosses a at $t = 0$ and eventually arrives, for the first time, at b (Fig. 1). In most cases it will not stay continuously within the interval (a, b) but rather escape it to the left well region $(-\infty, a)$. It may further reenter the transition region and loop back into the left well multiple times. Finally, it will enter the transition region for the last time and proceed to cross the point b thereby exiting the transition region toward the right well region $(b, +\infty)$. Thus the first-passage path from a to b generally consists of "internal loops" entering and exiting the transition region through the same boundary (red in Fig. 1), "external loops" where the system dwells inside the well region $(-\infty, a)$ (black in Fig. 1), and a "transition path" where the system traverses the transition region from one boundary to the other[21, 22] (green in Fig. 1).

The dynamics in the transition region, then, can be characterized by:

(1) The return time $t_R(a \to a)$, which is the temporal duration of failed attempts to cross the transition region (the red trajectory pieces in Fig. 1). This is the time that the system starting at the boundary a will take to return to this boundary conditional upon staying within the barrier region for $0 < t < t_R(a \to a)$.

(2) The transition path time $t_{TP}(a \to b)$, which is the temporal duration of successful attempts to cross the barrier (the green trajectory piece in Fig. 1). It is the time that the system starting at the boundary a will take to reach the other boundary b conditional upon staying within the barrier region for $0 < t < t_{TP}(a \to b)$.

The transition path time has received considerable theoretical[2, 20-37] and experimental[3, 38-44] attention in the last decade, while the return time, to our knowledge, has not yet been studied experimentally. As will be seen below, both of these times, as well as the first passage time, are limiting cases of the more general conditional exit time[45, 46], which is the time to exit the transition region through a given boundary (*a* or *b*) having started from a given point x_0 within the transition region.

The purpose of this study is two-fold. First, we propose that the more general yet experimentally accessible conditional exit time offers additional information about barrier crossing dynamics, particularly about failed barrier crossing attempts not captured by the transition path time. Second, we would like to explore the shape of the distribution of the conditional exit time (and its limiting cases such as the return and transition path time). In particular, the distribution width can be characterized by the ratio of the distribution variance to its mean (known as the coefficient of variation), which, for the case of diffusive dynamics in a potential of mean force, can be calculated analytically. Experimentally, the coefficient of variation, which only requires the first and the second moments of the distribution, is a more robust statistical measure than the distribution itself. Note that, in the context of chemical kinetics and molecular biophysics, the first-passage-time distribution, particularly for the model of 1D diffusive motion, is well studied[20, 45, 47, 48], and the expression for the first moment of the transition-path-time distribution has also been derived more recently[25, 49]. In contrast, higher moments of the distribution of the transition path time, which are required to quantify the distribution's width, have only received attention over the past year[35, 37, 50].

The importance of the distribution shape and width is exemplified by several recent developments: First, it was shown that the short-time behavior of the first passage time distribution contains information about the number of kinetic intermediates of a process[51, 52]. Second, the shape of the distribution of the transition path time measured in a FRET study of protein binding suggested the existence of an intermediate, encounter complex (i.e. potential well in the transition region) that could not be observed directly[3]. Third, it was shown[35, 37, 50] that the relative width of the distribution of the transition path time, as quantified by its coefficient of variation, cannot be too large if the system's dynamics along the coordinate x is diffusive: thus observation of broad distributions of transition path times can only be explained by a higher-dimensional free energy landscape allowing for parallel pathways. If transition-path-time distributions for diffusive dynamics are always narrow, what can we say about the distributions of other characteristic times of interest? In all of these cases, distributions of the observed times inform one about global features of the underlying free energy landscapes. Our hope, then, is that studying the distributions of various characteristic barrier crossing times is a viable step toward solving the inverse problem of learning about underlying multidimensional landscapes from low-dimensional signals.

We will further show here that these distributions display interesting (and sometimes pathological) behavior in some of the limiting cases. In particular, the mean return time $t_R(a \to a)$ is identically zero in the case of diffusive dynamics. This result has an important experimental implication: because of the finite spatial and temporal resolution of any experiment, the precise crossing of a boundary cannot be detected. As a result, an attempt to measure $t_R(a \to a)$ will lead to a finite time that depends on, e.g., experimental resolution, data sampling rate and/or the smoothing procedure used to reduce the noise in the experimental signal. We propose that such artifacts can be avoided if, instead, one considers the conditional exit time $t_E(x_0 \to a)$, where one starts from a point $x_0 \in (a, b)$ rather than from a. This resulting time has a well-defined distribution for any $x_0 > a$, and its behavior as x_0 approaches the left boundary a (i.e. when it becomes the return time) is quite interesting, as the distribution becomes "arbitrarily broad" for a sufficiently small distance $(x_0 - a)$ from the boundary in the sense that its mean becomes infinitely smaller than the standard deviation from the mean (note, however, that when inertial effects are neglected as in the Smoluchowski equation model, the absolute values of the standard deviation and the mean both approach 0 as $x_0 \to a$). In this regard, the distribution of the return time is drastically different from that of the transition path time. The physical origins of this behavior will be explained.

Likewise, we will show that distributions of first passage times can be broad, with a coefficient of variation that can be arbitrarily large. In contrast, distributions of the transition path time are always narrow, with a coefficient of variation below 1.

2. General formalism.

2.1. Distributions of conditional exit times and first passage times. To determine the statistical properties of conditional exit times we envisage a large ensemble of particles each starting, at $t = 0$, from a point x_0 that belongs to the transition region (a, b). The stochastic trajectories of each particle may be described by the usual overdamped Langevin equation[20], in which the force acting on the particle includes, in addition to the deterministic component $-U'(x)$, a friction force as well as a random noise component. We monitor each particle until it reaches either boundary a or boundary b, record the time it took to reach this boundary (and which boundary was reached), and then discard it from the ensemble. We seek the conditional probability density $\rho_{a(b)}(t|x_0)$, which describes the distribution of the time to exit the transition region for particles exiting through the boundary $a(b)$. Note that $\rho_{a(b)}(t|x_0)$ can be thought of as the probability density of the first passage time from x_0 to $a(b)$ conditional upon not crossing the point $b(a)$. Further note that

$$p_{a\to b}^{TP}(t) = p_{b\to a}^{TP}(t) = \lim_{x_0\to a} \rho_b(t|x_0) = \lim_{x_0\to b} \rho_a(t|x_0) \qquad (1)$$

is the distribution of the transition path time for the transition paths starting in a and ending in b, which is the same as the distribution of the transition path time for paths going from b to a because of the time reversal symmetry of Brownian dynamics[19, 26, 27]. Moreover, the limit

$$p_{a\to a}^{R}(t) = \lim_{x_0\to a} \rho_a(t|x_0) \ (2)$$

gives the distribution of the return time for recrossing the boundary a. We will, however, show that this limit is pathological for the model of diffusive dynamics.

The ensemble of trajectories originating from x_0 and terminating at the transition region boundaries can be described by the Smoluchowski equation. Specifically, the (conditional) probability density $G(x,t|x_0)$ of finding the particle at point $x \in (a,b)$ at time t (Green's function) obeys the equation:

$$\frac{\partial G}{\partial t} = \frac{\partial}{\partial x} D(x) e^{-\beta U(x)} \frac{\partial}{\partial x} e^{\beta U(x)} G \qquad (3)$$

with the initial condition

$$G(x,0|x_0) = \delta(x - x_0) \qquad (4)$$

and the absorbing boundary conditions at the boundaries

$$G(a,t|x_0) = G(b,t|x_0) = 0 \ (5)$$

The probability density $\rho_{a(b)}(t|x_0)$ is proportional to the flux of particles $j(a,t|x_0)$ or $j(b,t|x_0)$ exiting to the boundary a or b:

$$\rho_a(t|x_0) = -\frac{j(a,t|x_0)}{\phi(x_0\to a)} = \frac{1}{\phi(x_0\to a)} D(a) \frac{\partial G(x,t|x_0)}{\partial x}\Big|_{x=a} \quad (6)$$

$$\rho_b(t|x_0) = \frac{j(b,t|x_0)}{\phi(x_0\to b)} = -\frac{1}{\phi(x_0\to b)} D(b) \frac{\partial G(x,t|x_0)}{\partial x}\Big|_{x=b} \quad (7)$$

Here $\phi(x_0 \to a) = -\int_0^\infty dtj(a,t|x_0)$ and $\phi(x_0 \to b) = 1 - \phi(x_0 \to a) = \int_0^\infty dtj(b,t|x_0)$ are the fractions of particles in the ensemble exiting through a and b: those are known the splitting probabilities[48] (i.e. the probabilities to exit through the boundaries a and b).

The probability distribution of the first passage time from x_0 to b is calculated analogously, with the only difference that no absorbing boundary condition is imposed at $x = a$, since the trajectory $x(t)$ can cross the point a any number of times before being terminated at $x = b$ (See Fig. 1). In this case we have

$$p_{x_0 \to b}^{FP}(t) = -D(b)\frac{\partial G(x,t|x_0)}{\partial x}\Big|_{x=b} \quad (8)$$

for any starting point $x_0 < b$. We note that for a potential that has the property $U(x) \to \infty$ for $x \to -\infty$, this first passage time distribution can be obtained as the $a \to -\infty$ limit of the conditional exit time distribution $\rho_b(t|x_0)$. In other words, since the absorbing boundary at $a \to -\infty$ is never reached, the first passage times from x_0 to b are identical to the exit times. Thus first passage times can be viewed as a limiting case of conditional exit times.

2.2. Recursive equations for the distribution moments. Similarly to an approach known in the literature (see, e.g., ref. [48]), we now outline a general procedure for the calculation of the moments of the distributions of the conditional exit times introduced above. For concreteness, let us focus on the distribution $\rho_a(t|x_0)$ of the exit time to the boundary a. We are interested in its n-th moment,

$$\langle t^n(x_0 \to a\,)\rangle = \int_0^\infty dtt^n \rho_a(t|x_0) = \frac{1}{\phi(x_0 \to a)}\int_0^\infty dtt^n j(a,t|x_0) \equiv \frac{1}{\phi(x_0 \to a)}\langle \tau^n(x_0 \to a\,)\rangle \quad (9)$$

where we have introduced the moments of a *unnormalized* distribution

$$\langle \tau^n(x_0 \to a\,)\rangle = \int_0^\infty dtt^n j(a,t|x_0) = \int_0^\infty dtt^n D(a)\frac{\partial G(x,t|x_0)}{\partial x}\Big|_{x=a} \quad (10)$$

To find such moments, we start with the adjoint Smoluchowski equation (see, e.g., refs[20, 53]), which considers $G(x,t|x_0)$ as a function of the starting point x_0:

$$\frac{\partial G}{\partial t} = e^{\beta U(x_0)}\frac{\partial}{\partial x_0}D(x_0)e^{-\beta U(x_0)}\frac{\partial G}{\partial x_0}, \quad (11)$$

with the initial condition of Eq. 4 and the boundary conditions

$$G(x,t|a) = G(x,t|b) = 0 \quad (12)$$

Multiplying both sides of this equation by t^n and integrating over time, we obtain

$$e^{\beta U(x_0)} \frac{\partial}{\partial x_0} D(x_0) e^{-\beta U(x_0)} \frac{\partial F_n}{\partial x_0} = -nF_{n-1}, n > 0, \quad (13)$$

$$e^{\beta U(x_0)} \frac{\partial}{\partial x_0} D(x_0) e^{-\beta U(x_0)} \frac{\partial F_0}{\partial x_0} = -\delta(x - x_0), \quad (14)$$

where we have introduced the auxiliary functions

$$F_n(x|x_0) = \int_0^\infty dt t^n G(x, t|x_0), \quad (15)$$

which satisfy the boundary condition

$$F_n(x|a) = F_n(x|b) = 0. \quad (16)$$

From Eq. 10, the moments $\langle \tau^n(x_0 \to a)\rangle$ (Eq. 10) now can be written as

$$\langle \tau^n(x_0 \to a)\rangle = D(a) \frac{\partial F_n(x|x_0)}{\partial x}\Big|_{x=a} (17)$$

Using Eqs. 13 and 17, we find that these moments satisfy the following equation:

$$e^{\beta U(x_0)} \frac{\partial}{\partial x_0} D(x_0) e^{-\beta U(x_0)} \frac{\partial \langle \tau^n(x_0 \to a)\rangle}{\partial x_0} = -n\langle \tau^{n-1}(x_0 \to a)\rangle, \; n > 0, (18)$$

which should be supplemented with the boundary conditions

$$\langle \tau^n(x_0 \to a)\rangle|_{x=a} = \langle \tau^n(x_0 \to a)\rangle|_{x=b} = 0 \quad (19)$$

Thus if the $(n-1)$-th moment, $\langle \tau^{n-1}(x_0 \to a)\rangle$, is known then the next moment $\langle \tau^n(x_0 \to a)\rangle$ can be obtained by integrating Eq. 18 twice.

To obtain the *n*-th moment of interest, $\langle t^n(x_0 \to a)\rangle$, the moment $\langle \tau^n(x_0 \to a)\rangle$ needs to be divided by the splitting probability (Eq. 9). As this probability is the 0-th order moment of the unnormalized distribution, it can be obtained by solving Eq. 14 with the boundary conditions of Eq. 16, resulting in the known result[48]

$$\phi(x_0 \to a) = D(a) \frac{\partial}{\partial x} \left[\int_0^\infty dt G(x, t|x_0)\right]_{x=a} = D(a) \frac{\partial F_0(x|x_0)}{\partial x}\Big|_{x=a} = \frac{\int_{x_0}^b dx e^{\beta U(x)}/D(x)}{\int_a^b dx e^{\beta U(x)}/D(x)}, \quad (20)$$

Note that while the equation hierarchy, Eq. 18, is similar to that derived in the literature for first-passage times[37, 48], there is an important technical difference: because $\rho_{a(b)}(t|x_0)$ is a conditional distribution, Eq. 18 is satisfied by the moments $\langle \tau^n(x_0 \to a)\rangle$ and not by $\langle t^n(x_0 \to a)\rangle$.

3. Results

3.1. Exit times. We now give the general analytical solutions for the first and second moments of the exit times. Starting with the unnormalized distribution's "zeroth moment" (which is the splitting probability), integrating Eq. 18 twice, and dividing by the splitting probability (Eq. 9) we find the first moment

$$\langle t(x_0 \to a\,)\rangle_E = \frac{1}{\int_{x_0}^{b} dx\, e^{\beta U(x)}/D(x)} \int_a^{x_0} \frac{dx\, e^{\beta U(x)}}{D(x)} \int_a^{b} \frac{dy\, e^{\beta U(y)}}{D(y)} \int_x^{y} dz \phi(z \to a) e^{-\beta U(z)}, \qquad (21)$$

where the splitting probability $\phi(z \to a)$ is defined in Eq. 20. (Note that here we use the notation $\langle \dots \rangle_E$ to indicate averaging over the distribution of conditional exit times). Similarly, the second moment now can be expressed in terms of the first moment by integrating Eq.18 twice. This gives

$$\langle t^2(x_0 \to a\,)\rangle_E = \frac{2}{\int_{x_0}^{b} dx e^{\beta U(x)}/D(x)} \int_a^{x_0} \frac{dx\, e^{\beta U(x)}}{D(x)} \int_a^{b} \frac{dy\, e^{\beta U(y)}}{D(y)} \int_x^{y} dz \phi(z \to a) \langle t(z \to a\,)\rangle_E e^{-\beta U(z)}\,.$$

(22)

The expression for the n-th moment, $\langle t^n(x_0 \to a\,)\rangle_E$, $n = 3,4 \dots$, is obtained by replacing $2\langle t(z \to a\,)\rangle_E$ in Eq. 22 by $n\langle t^{n-1}(z \to a\,)\rangle_E$.

We are particularly interested in the distribution's relative width, which is conventionally quantified by its coefficient of variation, which is equal to the ratio of the standard deviation to the distribution's mean:

$$C = \frac{[\langle t^2\rangle - \langle t\rangle^2]^{\frac{1}{2}}}{\langle t\rangle} \qquad (23)$$

Narrower distributions have smaller values of the coefficient of variation, and the coefficient of variation for an exponential distribution is 1. The value of the coefficient of variation encodes important information about the underlying dynamics and energy landscape. For example, in the case of barrier crossing by transition paths, lower values of this coefficient correspond to higher barriers[35], while a value of 1 (corresponding to a single-exponential distribution) suggests (inasmuch as the model of diffusive dynamics is applicable) that transition paths cross a potential well (intermediate) that traps the system[3]. Even more interestingly, regardless of the underlying potential of mean force this coefficient cannot possibly exceed 1 for the model of one-dimensional diffusive dynamics model[35, 37]; thus experimental observation of C exceeding 1 automatically invalidates such a model. But what about coefficients of variation for the more general exit time?

For the double-well potential shown in Fig. 2, Fig. 3 shows how the coefficient of variation of the exit time through the left boundary depends on the starting point x_0. For x_0 approaching the right boundary, $x_0 \to b$, the exit time approaches the transition path time $t_{TP}(b \to a)$; it has been shown previously[35, 37] that the coefficient of variation for this time is always less than 1 (assuming the validity of the diffusive dynamics model).

In the opposite limit, $x_0 \to a$, the exit time becomes the return time $t_R(a \to a)$, and we observe that the coefficient of variation diverges. Mathematically, this divergence arises because both the first and the second moments grow linearly with $x_0 - a$ in the limit $x_0 - a \to 0$.

Broadening of the exit time distribution as the starting point x_0 approaches the absorbing boundary a is also observed directly in the distribution's shape shown, for three values of x_0, in Fig. 4. These distributions were obtained using Eq. 6, from a numerical solution of the Smoluchowski equation that uses the spectral expansion method – see ref.[27] for further details. Note that, because both the first and the second moments of this distribution increase with increasing x_0 (see the inset of Fig. 3, where the first moment of the distribution is plotted as a function of x_0), the time in each distribution is rescaled by its mean value. In other words, the first moments of each distribution as plotted in Fig. 4 are the same and equal to 1.

At first glance, the distribution $\rho_a(t|x_0)$ obtained for a value x_0 that is closest to the left boundary seem more "narrow" than the other two, showing a sharper rise at short times. But this is not so (inasmuch as the coefficient of variation is a good measure of width) because of the long tail exhibited by this distribution. This tail is more readily observed when the same distribution is plotted on a logarithmic scale (Fig. 4, inset).

To further understand the distribution broadening as x_0 approaches the boundary a, it is instructive to consider the case of zero potential, $U(x) = 0$, with a coordinate-independent diffusion coefficient $D(x) = D$. Setting, without loss of generality, $a = 0$, the integrals of Eq. 21 and 22 can be evaluated analytically to give

$$\langle t(x_0 \to 0\,)\rangle_E = \frac{(2b - x_0)x_0}{6D} \quad (24)$$

$$\langle t^2(x_0 \to 0\,)\rangle_E = \frac{x_0\left(8b^3 + 8b^2x_0 - 12bx_0^2 + 3x_0^3\right)}{180D^2}. \quad (25)$$

Both functions are proportional to x_0 in the limit $x_0 \to 0$, thus leading to $x_0^{-1/2}$ divergence of the coefficient of variation (Eq. 23).

Moreover, the distribution of the exit time for the case of zero potential can be calculated analytically. Although this problem of free diffusion on an interval with absorbing boundaries was studied in the literature – see, e.g., ref.[45] – we provide a derivation for $\rho_0(t|x_0)$ in Appendix A for completeness, and as a simple illustration of the approach here. Specifically, the Laplace transform of $\rho_0(t|x_0)$ is given by

$$\tilde{\rho}_0(s|x_0) = \frac{\sinh\left[(b-x_0)\sqrt{\frac{s}{D}}\right]}{(b-x_0)\sinh\left[b\sqrt{\frac{s}{D}}\right]} \qquad (26)$$

Consider now the case $x_0 \ll b, b\sqrt{\frac{s}{D}} \gg 1$. In the time domain, the second inequality implies that we are considering timescales much shorter than the characteristic diffusion time on the segment $(0, b)$, i.e., $t \ll \frac{b^2}{D}$. In this limit, we obtain $\tilde{\rho}_0(s|x_0) \approx \frac{1}{b-x_0} e^{-x_0\sqrt{\frac{s}{D}}}$, and taking the inverse Laplace transform we find

$$\rho_0(t|x_0) \approx \frac{x_0}{2\sqrt{\pi D}} t^{-\frac{3}{2}} e^{-\frac{x_0^2}{4Dt}} \qquad (27)$$

In fact, Eq. 27 provides an accurate description of the exit time distribution even in the presence of a nonzero potential $U(x)$ as long as the time t is not too long - see Figure 5. This is because, at short enough timescales, diffusion dominates over the drift, and the presence of the potential is immaterial. At longer times, however, the exit time distribution exhibits an exponential tail, in contrast to a power-law tail predicted by Eq. 27. A true power-law tail will be observed in the absence of a potential when the right boundary is removed to infinity – see Section 3.2.

Returning to the case of zero potential, when $x_0^2 \ll Dt$ then the exponential in Eq. 27 can be replaced by 1. On the other hand, when $Dt \ll b^2$, the particle originating from x_0 does not have time to reach the right boundary b. When both of these conditions are satisfied, $\frac{x_0^2}{D} \ll t \ll \frac{b^2}{D}$, Eq. 27 predicts a power lay $\rho_0(t|x_0) \propto t^{-\frac{3}{2}}$, a behavior characteristic of a “broad” distribution displaying multiple timescales. Note, however, that the absolute width of the distribution, as opposed to the relative width quantified by the coefficient of variation, decreases as $x_0 \to 0$. Indeed, at $x_0 = 0$ the Laplace transform of this distribution is constant, and the distribution is the delta function, $\rho_0(t|x_0) = \delta(t)$, indicating that a particle cannot leave the absorbing boundary $x_0 = 0$. In other words, the return time is identically zero in this case.

This delta-function distribution of the return time should be expected of diffusive dynamics (or equivalently, of the overdamped Langevin equation where the inertial term containing acceleration is omitted). It is pathological: if inertial effects are taken into account, the system crossing the left boundary a with a positive velocity will travel to the right over some finite time. Thus it will take a finite time to return to the boundary. The typical return time, then, would be comparable to the velocity correlation time, which is typically very short and thus unresolvable by current single-molecule techniques[54]. This conclusion has important experimental implications: an attempt to measure the return time will likely be confounded by experimental artifacts such as the limited spatial and temporal resolution or the smoothing of the trajectory (which introduces an artificial "velocity" that would be absent in true diffusive dynamics). These difficulties can be avoided by measuring the exit time distribution $\rho_a(t|x_0)$ instead, with the initial point x_0 being within the limits allowed by the experiment's spatial resolution.

3.2. First passage times. We now consider the distribution of the first passage time $p^{FP}_{x_0 \to b}(t)$ to arrive at the (target) boundary b starting from $x = x_0 < b$. The first moment of this distribution is given by the known expression[20, 45, 47]

$$\langle t(x_0 \to b\,)\rangle_{FP} = \int_{x_0}^{b} \frac{dx\, e^{\beta U(x)}}{D(x)} \int_{-\infty}^{x} dy e^{-\beta U(y)}, \qquad (28)$$

and the second moment can be obtained similarly to the approach of Section 3.1 by integrating Eq. 18:

$$\langle t^2(x_0 \to b\,)\rangle_{FP} = 2\int_{x_0}^{b} \frac{dx\, e^{\beta U(x)}}{D(x)} \int_{-\infty}^{x} dy e^{-\beta U(y)} \langle t(y \to b\,)\rangle_{FP} \qquad (29)$$

The dependence of the coefficient of variation on the initial position x_0 for the double-well potential shown in Fig. 2 and for a coordinate-independent $D(x)$ is illustrated in Fig. 6. For x_0

located not too far from the left potential minimum (located at $x = -1$), we observe that its value is close to 1. This is easy to understand: thermally activated escape from a potential well separated from the target boundary b by a sufficiently high barrier is governed by 1st-order kinetics, resulting in an exponential distribution of escape times, whose coefficient of variation is 1. Even though the barrier in this case is only $2k_BT$, the coefficient of variation is only slightly below 1.

When the initial position x_0 moves toward b, however, the coefficient of variation increases, and it diverges as the starting point approaches the target, $x_0 \to b$. This behavior is similar to that of the exit time discussed in Section 3.1 and illustrated in Fig. 3. The physical origins of this behavior will be further discussed below. Mathematically, the divergent behavior of the coefficient of variation follows from Eqs. 28 and 29 (and from the definition of the coefficient of variation, Eq. 23). Indeed, based on these equations, both the first and the second moments of the distribution are, to lowest order in the distance to the target $(b - x_0)$, are proportional to $(b - x_0)$, and thus the coefficient of variation scales as $C \propto (b - x_0)^{-\frac{1}{2}}$.

To gain further insight into the shapes of the first passage time distributions, consider the analytically tractable case of a linear potential,

$$U(x) = -Fx \qquad (30)$$

describing a particle subjected to a constant force $F \geq 0$. Using Eqs. 28 and 29, we find, for $x_0 < b$,

$$\langle t(x_0 \to b\,)\rangle_{FP} = \frac{b - x_0}{\beta F D} \qquad (31)$$

and

$$\langle t^2(x_0 \to b\,)\rangle_{FP} = \frac{2}{(\beta F)^3 D^2}\left[(b - x_0) + \frac{\beta F (b - x_0)^2}{2}\right], \qquad (32)$$

from which, using Eq. 23, we find

$$C = \left[\frac{2}{\beta F (b - x_0)}\right]^{1/2} \qquad (33)$$

showing divergent behavior as $x_0 \to b$.

A particularly interesting case is that of free diffusion, which is obtained by setting the force F to zero. Formally, Eq. 33 predicts that the coefficient of variation should diverge for any initial position x_0, but, of course, both the first and the second moments of the distribution diverge in this case, so the coefficient of variation is not well defined at exactly zero force (see Eq. 23). Nevertheless, the distribution of the first passage time itself is well defined for free diffusion. It is given by (see, e.g., ref. [45]; Appendix B provides a derivation for completeness)

$$p_{x_0 \to b}^{FP}(t) = \frac{b-x_0}{2\sqrt{\pi D}} t^{-\frac{3}{2}} e^{-\frac{(b-x_0)^2}{4Dt}} \qquad (34)$$

Eq. 34 is identical to Eq. 27 (provided that x_0 is replaced by $b - x_0$; both of these quantities are the distances to the absorbing boundary in each case). However, unlike Eq. 27, which only holds at intermediate timescales, Eq. 34 has a power-law tail decaying as $t^{-3/2}$ at arbitrarily long times. This difference is due to the fact that diffusion on an infinitely long segment, unlike diffusion within a transition region of finite length, does not possess a longest characteristic timescale. More specifically, as noted above, the power law in Eq. 27 holds only at times short enough that the particle does not have enough time to reach the right boundary b, and thus the exit time is indistinguishable from the first passage time. In contrast, no such boundary exists in the present case, diffusion takes place on a semi-infinite segment, and arbitrarily long excursions from the point x_0 are possible. Although the distribution of Eq. 34 is itself normalized, all of its moments diverge.

These findings show that first passage time distributions can be broad, with a coefficient of variation exceeding 1, or even with its moments being infinite. This is different from the distribution of the transition path time, which, for one-dimensional diffusive dynamics, is always below 1, as further discussed below.

3.3. Transition path times. The 1st and 2nd moments of the distribution of the transition path can be found from Eqs. 21-22 as the limiting case where the starting point x_0 coincides with the absorbing boundary b. Note that this gives moments of the transition path time from b to a, but given the time-reversal symmetry of transition paths, they are the same as those for transition paths from a to b. The result for the first moment has been derived previously[25, 49, 55]:

$$\langle t(a \to b\,)\rangle_{TP} = \langle t(b \to a\,)\rangle_{TP} = \left(\int_a^b \frac{dx\, e^{\beta U(x)}}{D(x)}\right) \int_a^b dz \phi(z \to x_0)\phi(z \to b) e^{-\beta U(z)} \qquad (35)$$

For the second moment, we arrive at

$\langle t^2(a \to b)\rangle_{TP} = \langle t^2(b \to a)\rangle_{TP} = 2(\int_a^b \frac{dx\, e^{\beta U(x)}}{D(x)}) \int_a^b dz \phi(z \to x_0)\phi(z \to b)\langle t(z \to b)\rangle_E e^{-\beta U(z)}$, (36)

where $\langle t(x_0 \to b)\rangle_E$ is the mean exit time through boundary b conditional upon not crossing the boundary a. This result was reported earlier in ref.[35] for coordinate-independent diffusivity, and in ref.[37].

As proven in refs.[35, 37], the coefficient of variation of the transition path time distribution (Eq. 23) always remains below 1 (Fig. 6). For the potential shown in Fig. 2 and for initial points $x_0 < -1$ located to the left of the potential minimum at $x = -1$, the coefficients of variation for both transition path time- and first passage time distributions are close to each other and to the value $C = 1$ expected for an exponential distribution (Fig. 6). This is not surprising: if the starting point $x_0 < -1$ is located on a steep left wall of a potential well, a trajectory originating from x_0 will likely proceed toward the right of the starting point, and thus a first passage and a transition path time should be nearly the same. Moreover, their distributions should be close to exponential, with $C \approx 1$, as these times should be close to the time of thermally activated escape from the left potential well.

As the initial point x_0 approaches the target b, the coefficient of variation for the transition path time distribution stays below 1, while the coefficient of variation for the first passage time distribution increases and diverges for $x_0 \to b$, a behavior explained in Section 3.2.

4. Discussion and concluding remarks.

Our findings show that the distribution of the transition path time is special in that, in the case of purely diffusive dynamics, its coefficient of variation cannot exceed 1, in contrast to the distributions of the first passage and conditional exit times. In other words, transition path time distributions are narrow (narrower than exponential) while first passage time and exit time distributions may be arbitrarily broad, with their coefficients of variations being arbitrarily large.

What is the physical origin of this difference? It is instructive to consider the free diffusion case discussed in Section 3.2. Consider a Brownian particle starting to the left of the target at $x_0 < b$. It may proceed directly toward the target b, in which case its trajectory will be a transition path from x_0 to b, or it may exercise an arbitrarily long detour to the left of the starting point before finally finding the target at $x = b$ (note that in 1D the particle will eventually find the target with certainty). In a sense, these two scenarios may be considered as two parallel pathways contributing to the first passage time distribution, but only with the first scenario contributing to the transition path time distribution. As shown in refs.[35, 50], it is the existence of such parallel pathways with disparate characteristic timescales that leads to broad distributions, and, indeed, Eq. 34 exhibits a power-law tail describing long-lasting events.

Exit times, Eqs. 21-22, provide a more general description of barrier crossing dynamics than the better-studied transition path time. Exit times have been discussed previously in the context of finding practical reaction coordinates[46]. It is widely believed that the splitting probability (committor) has attractive mathematical properties that makes it in a certain sense an optimal reaction coordinate[25, 53, 56-62]. The splitting probability answers the following question: starting from a certain point in phase or configuration space, what is the probability to get to the reaction product before getting to the reactant? In our one-dimensional model, the function $\phi(x_0 \to b)$ (x_0 being the starting point) gives the answer to this question, provided that the segment $(-\infty, a)$ is regarded as the reactant and $(b, +\infty)$ as the product. The conditional exit time is a complementary quantity that answers a question that is concerned with the transition timescale: starting from a certain point in phase or configuration space, what is the (mean) time to get to the product conditional upon reaching the product before getting to the reactant? In our case this question is answered by the mean exit time $\langle t(x_0 \to b)\rangle_E$. The exit time generalizes the notion of the transition path time: in the limit where the initial point x_0 coincides with the boundary a, this time becomes the transition path time $t_{TP}(a \to b)$. Another limiting case of the conditional exit time is the (mean) return time $\langle t(x_0 \to a)\rangle_E|_{x_0 \to a} = \langle t(a \to a)\rangle_R$, which is the mean time it takes to return to the boundary a starting from this boundary and conditional upon not exiting through the boundary b. For the case of diffusive dynamics considered here, however, this time is identically equal to zero.

This paper has focused on the model of diffusive dynamics along a one-dimensional reaction coordinate. Given the many successes of this model in application to biomolecular folding (see, e.g., refs.[63-65]), we view elaboration on further consequences of this model for the dynamics in the transition region a worthy pursuit, but it also gives us an opportunity to delineate potential limitations of this model in application to experimental data. Indeed, the following three properties are strictly satisfied by this model, yet they may be violated when the dynamics along the reaction coordinate are influenced by memory effects and/or when multi-dimensionality is essential[66, 67]:

(1) Locality of exit times: Distributions of exit times (and transition path times in particular) are independent of the potential $U(x)$ outside the transition region (a, b). This is obvious from the derivation outline described in Section 2.1, since the properties of the potential outside the transition interval do not enter into the picture (Of course, Markovianity of diffusion, allowing one to disregard any knowledge of the system's past prior to its arrival at the point x_0 is key here). Note that this local property is not true for first-passage times from a to b, which depend on the properties of the potential $U(x)$ for $x < a$.

(2) The distribution of the transition path time has a coefficient of variation that cannot exceed 1, regardless of the potential and the transition region boundaries. This property is a direct consequence of Eqs. 23, 35 and 36[35].

(3) Transition path times exhibit forward/backward symmetry, $p_{a\to b}^{TP}(t) = p_{b\to a}^{TP}(t)$, as already stated in Eq. 1. Note that this property is true even for systems that are not in equilibrium[68] because of the property (1).

While strictly true for diffusive dynamics, some (or all) of these properties may not be true if the dynamics along x is not diffusive. For example, property (1) cannot be generally true for non-Markovian dynamics, property (2) has already been found to be violated for the dynamics of reaction coordinates in protein folding[35], and property (3), while true for any equilibrium system, was found to be violated for systems that are simultaneously non-equilibrium and non-Markovian[68]. Violation of any of the above exact predictions would invalidate, with certainty, the diffusive picture of the dynamics along a reaction coordinate and call for a more accurate model.

We conclude this paper with comments on how experimental time resolution may affect information that can be deduced from the distributions of the various times discussed here. Generally, limited time resolution may result in missing short-time events, thereby skewing the apparent distribution toward longer times while simultaneously making it narrower. More formally, we may write the true distribution $p(t)$ of the time of interest as

$$p(t) = f p_{un}(t) + (1-f) p_{obs}(t), \qquad (37)$$

where f is the fraction of unobserved short-time events with a normalized distribution $p_{un}(t)$, and $1-f$ the fraction of observed events with a normalized distribution $p_{obs}(t)$. The latter distribution is the one observed experimentally. Thus we have

$$C^2 + 1 = \frac{\langle t^2 \rangle}{\langle t \rangle^2} = \frac{f\langle t^2 \rangle_{un} + (1-f)\langle t^2 \rangle_{obs}}{[f\langle t \rangle_{un} + (1-f)\langle t \rangle_{obs}]^2} \qquad (38)$$

where $\langle t \rangle_{un}$, $\langle t \rangle_{obs}$, $\langle t^2 \rangle_{un}$, $\langle t^2 \rangle_{obs}$ are the first and second moment of the unobserved and observed distributions, as indicated by the subscripts. Given that unobserved events are shorter than the observed ones, we have $\langle t \rangle_{un} < \langle t \rangle_{obs}$, and it is obvious that the true distribution mean $\langle t \rangle = f\langle t \rangle_{un} + (1-f)\langle t \rangle_{obs}$ is always shorter than the observed value $\langle t \rangle_{obs}$.

Let us now focus on the case of measuring transition path times. Given that higher barriers (for the same transition region boundaries) correspond to shorter mean transition path times[19, 25], this longer observed values of the transition path time could lead the experimentalist to deduce a lower transition barrier. On the other hand, as follows from Eq. 38, the true coefficient of variation C is *greater* than the observed one, $C_{obs} = \frac{[\langle t^2 \rangle_{obs} - \langle t \rangle_{obs}^2]^{\frac{1}{2}}}{\langle t \rangle_{obs}}$, and a narrower observed distribution of the transition path time, with a greater coefficient of variation, corresponds to a *higher* barrier. Thus analysis of experimental distribution widths may reveal experimental artifacts.

Appendix A. Exit time distribution for free diffusion.

Following the approach outlined in Section 2.1, we write the diffusion equation describing the time evolution of Green's function $G(x, t|x_0)$ of a free particle (i.e., Eq. 3 with constant diffusivity D and with $U(x) = 0$):

$$\frac{\partial G}{\partial t} = D\frac{\partial^2 G}{\partial x^2}, \qquad (A1)$$

with the initial condition

$G(x,0|x_0) = \delta(x - x_0)$, (A2)

and with absorbing boundary conditions (cf. Eq. 5)

$G(a = 0, t|x_0) = G(b, t|x_0) = 0$ (A3)

To solve Eq. A1, we rewrite it in Laplace space:

$s\,\hat{G} - \delta(x - x_0) = D\frac{\partial^2 \hat{G}}{\partial x^2}$, (A4)

where

$\hat{G}(x, s|x_0) = \int_0^\infty dt e^{-st} G(x, t|x_0)$ (A5)

The solution of Eq. A4 satisfying the absorbing boundary conditions can be written in the form

$$\hat{G}(x, s|x_0) = \begin{cases} A\ \sinh[x\sqrt{\frac{s}{D}}], 0 \le x < x_0 \\ B \sinh[(b - x)\sqrt{\frac{s}{D}}], x_0 < x \le b \end{cases} \quad \text{(A6)}$$

The coefficients A and B can be determined from the continuity of the Green's function at $x = x_0$,

$\hat{G}(x_0 + 0, s|x_0) = \hat{G}(x_0 - 0, s|x_0)$, (A7)

and from the condition

$D\frac{\partial \hat{G}}{\partial x}\Big|_{x_0-0}^{x_0+0} = -1$, (A8)

which is obtained by integrating Eq. A4 over an infinitesimal interval from $x_0 - 0$ to $x_0 + 0$.

The exit time distribution can now be found from Eq. 6 rewritten in Laplace space:

$\hat{\rho}_0(t|x_0) = \frac{1}{\phi(x_0 \to 0)} D\frac{\partial \hat{G}(x, t|x_0)}{\partial x}\Big|_{x=0}$, (A9)

where the splitting probability $\phi(x_0 \to 0)$ is the integral of the flux exiting through the left boundary,

$\phi(x_0 \to 0) = D\int_0^\infty dt \frac{\partial G(x,t|x_0)}{\partial x}\Big|_{x=0} = D\frac{\partial \hat{G}(x,t|x_0)}{\partial x}\Big|_{x=0}$ (A10)

The expression in Eq. 26 is then obtained using Eqs. A6-A10.

Appendix B. First passage time distribution for free diffusion

Following the approach outlined in Section 2.1, we seek the solution $G(x, t|x_0)$ of the diffusion equation

$$\frac{\partial G}{\partial t} = D\frac{\partial^2 G}{\partial x^2}, \qquad \text{(B1)}$$

with the initial condition

$$G(x,0|x_0) = \delta(x - x_0) \text{ (B2)}$$

and with absorbing boundary condition at $x = b$:

$$G(b,t|x_0) = 0 \text{ (B3)}$$

We assume that $x_0 < b$. The solution for $x < b$ is easily constructed from Green's function of the freely diffusing particle using the method of images:

$$G(x,t|x_0) = \frac{1}{\sqrt{4\pi Dt}}\left[e^{-\frac{(x-x_0)^2}{4Dt}} - e^{-\frac{(x-2b+x_0)^2}{4Dt}}\right] \qquad \text{(B4)}$$

The probability distribution of the first passage time from x_0 to the boundary b is equal to the flux at the target boundary,

$$p_{x_0\to b}^{FP}(t) = -D\frac{\partial G(x,t|x_0)}{\partial x}\bigg|_{x=b} = \frac{b-x_0}{2\sqrt{\pi D}}t^{-\frac{3}{2}}e^{-\frac{(b-x_0)^2}{4Dt}}, \text{ (B5)}$$

which is the expression in Eq. 34. Note that this distribution is automatically normalized, since the integral of the flux over time, equal to the probability of crossing $x = b$ at any time, is 1.

Acknowledgements

AMB was supported by the Intramural Research Program of the NIH, Center for Information Technology. DEM was supported by the Robert A. Welch Foundation (Grant No. F- 1514) and the National Science Foundation (Grant No. CHE 1955552).

Figure legends

Figure 1. Barrier crossing dynamics: A first-passage path from a to b consists of "internal loops" entering and exiting the transition region through the same boundary[22] (red), "external loops" where the system dwells inside the well region $(-\infty, a)$ (black), and a "transition path" where the system traverses the transition region from one boundary to the other (green). Typical one-dimensional free energy landscape of folding is shown on the right.

Figure 2. Double-well potential $U(x) = 2k_BT(x^2 - 1)^2$ used to illustrate the results here. The transition region boundaries, $a = -1, b = 1$, are shown as dashed lines.

Figure 3. Coefficient of variation for the exit time distribution $\rho_a(t|x_0)$ as a function of the starting point x_0. The transition region boundaries are shown as dashed lines. Inset shows the mean exit time as a function of x_0. This time is measured in dimensional units, with L^2/D (where $2L$ is the distance between the two potential minima and D the diffusivity) being the unit of time. The unit of distance is L.

Figure 4. Probability distribution $\rho_a(t|x_0)$ of the exit time to the left boundary (Eq.6) for different values of the initial position x_0, as indicated in the legend. The time in each case is normalized by the distribution's mean. The potential $U(x)$ and the transition region boundaries are the same as in Figures 2-3. Of the three distributions shown here, the one corresponding to the point x_0 closest to the left boundary is the broadest (i.e. has the largest coefficient of variation) because of its long tail. This tail is easy to see when the same data is plotted on a logarithmic scale (inset).

Figure 5. Eq. 27 vs. the probability distribution for the exit time to the left boundary, $\rho_a(t|x_0)$, in the double-well potential of Fig. 2. In both cases, the distance $x_0 - a$ is much shorter than the length $b - a$ of the transition region and is equal to 0.1. The time is measured in dimensional units, with L^2/D (where $2L$ is the distance between the two potential minima and D the diffusivity) being the unit of time. The distance L provides a unit of distance.

Figure 6. Coefficient of variation *C* for the first passage time from x_0 to b (solid line), and for the transition path time from x_0 to b (dashed-dotted line), as a function of the starting point x_0. Inset: same data for the transition path time.

References

1. Z. Kilic, I. Sgouralis and S. Presse, Biophys J **120** (3), 409-423 (2021).
2. D. E. Makarov, J Phys Chem B **125** (10), 2467-2476 (2021).
3. F. Sturzenegger, F. Zosel, E. D. Holmstrom, K. J. Buholzer, D. E. Makarov, D. Nettels and B. Schuler, Nat Commun **9** (1), 4708 (2018).
4. G. Schreiber, G. Haran and H. X. Zhou, Chem Rev **109** (3), 839-860 (2009).
5. A. B. Kolomeisky, *Motor Proteins and Molecular Motors*. (CRC Press, 2015).
6. A. M. Berezhkovskii and D. E. Makarov, J Phys Chem Lett **11** (5), 1682-1688 (2020).
7. N. Q. Hoffer and M. T. Woodside, Curr Opin Chem Biol **53**, 68-74 (2019).
8. B. Schuler and W. A. Eaton, Curr Opin Struct Biol **18** (1), 16-26 (2008).
9. B. Schuler, E. A. Lipman and W. A. Eaton, Nature **419** (6908), 743-747 (2002).
10. S. J. Hagen, J. Hofrichter, A. Szabo and W. A. Eaton, Proc. Natl. Acad. Sci. USA **93**, 11615 (1996).
11. O. Bieri and T. Kiefhaber, Biological chemistry **380** (7-8), 923-929 (1999).
12. O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello and T. Kiefhaber, Proceedings of the National Academy of Sciences of the United States of America **96** (17), 9597-9601 (1999).
13. Z. Wang and D. E. Makarov, J. Phys. Chem. B **107**, 5617 (2003).
14. I. V. Gopich, D. Nettels, B. Schuler and A. Szabo, J Chem Phys **131** (9), 095102 (2009).
15. A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgartel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin and B. Schuler, Proc Natl Acad Sci U S A **104** (1), 105-110 (2007).
16. D. Nettels, I. V. Gopich, A. Hoffmann and B. Schuler, Proc Natl Acad Sci U S A **104** (8), 2655-2660 (2007).
17. A. Soranno, B. Buchli, D. Nettels, R. R. Cheng, S. Muller-Spath, S. H. Pfeil, A. Hoffmann, E. A. Lipman, D. E. Makarov and B. Schuler, Proceedings of the National Academy of Sciences of the United States of America **109**, 17800-17806 (2012).
18. E. Haas and I. Z. Steinberg, Biophysical Journal **46**, 429 (1984).
19. D. E. Makarov, *Single Molecule Science: Physical Principles and Models*. (CRC Press, Taylor & Francis Group, Boca Raton, 2015).
20. R. Elber, D. E. Makarov and H. Orland, *Molecular Kinetics in Condense Phases: Theory, Simulation, and Analysis*. (Wiley and Sons, 2020).
21. A. M. Berezhkovskii, L. Dagdug and S. M. Bezrukov, J Chem Phys **147** (13), 134104 (2017).
22. A. M. Berezhkovskii, L. Dagdug and S. M. Bezrukov, J Phys Chem B **121** (21), 5455-5460 (2017).
23. B. W. Zhang, D. Jasnow and D. M. Zuckerman, J. Chem. Phys. **126**, 074504 (2007).
24. D. M. Zuckerman and T. B. Woolf, J. Chem. Phys **116** (6), 2586-2591 (2002).
25. G. Hummer, J Chem Phys **120**, 516-523 (2004).
26. A. M. Berezhkovskii, G. Hummer and S. M. Bezrukov, Phys Rev Lett **97** (2), 020601 (2006).
27. S. Chaudhury and D. E. Makarov, J. Chem. Phys. **133**, 034118 (2010).
28. E. Pollak, Phys Chem Chem Phys **18** (41), 28872-28882 (2016).
29. E. Pollak, Phys Rev Lett **118** (7), 070401 (2017).

30. E. Carlon, H. Orland, T. Sakaue and C. Vanderzande, The Journal of Physical Chemistry B **122** (49), 11186-11194 (2018).
31. M. Laleman, E. Carlon and H. Orland, J Chem Phys **147** (21), 214103 (2017).
32. P. Cossio, G. Hummer and A. Szabo, Journal of Chemical Physics **148**, 123309 (2018).
33. S. V. Malinin and V. Y. Chernyak, J Chem Phys **132** (1), 014504.
34. E. Medina, R. Satija and D. E. Makarov, J Phys Chem B **122** (49), 11400-11413 (2018).
35. R. Satija, A. M. Berezhkovskii and D. E. Makarov, Proc Natl Acad Sci U S A **117** (44), 27116-27123 (2020).
36. R. Satija, A. Das and D. E. Makarov, The Journal of Chemical Physics **147** (15), 152707 (2017).
37. D. Hartich and A. Godec, (arXiv:2011.04628 . Accepted for publication in Phys. Rev. X, 2021).
38. H. S. Chung and W. A. Eaton, Curr Opin Struct Biol **48**, 30-39 (2018).
39. H. S. Chung, J. M. Louis and W. A. Eaton, Proc Natl Acad Sci U S A **106** (29), 11837-11844 (2009).
40. H. S. Chung, J. M. Louis and W. A. Eaton, Science **335**, 981-984 (2012).
41. K. Neupane, D. A. Foster, D. R. Dee, H. Yu, F. Wang and M. T. Woodside, Science **352** (6282), 239-242 (2016).
42. K. Neupane, D. B. Ritchie, H. Yu, D. A. Foster, F. Wang and M. T. Woodside, Phys Rev Lett **109** (6), 068102 (2012).
43. H. Yu, A. N. Gupta, X. Liu, K. Neupane, A. M. Brigley, I. Sosova and M. T. Woodside, Proc Natl Acad Sci U S A **109** (36), 14452-14457 (2012).
44. J. Y. Kim and H. S. Chung, Science **368** (6496), 1253-1257 (2020).
45. S. Redner, *A Guide to Fisrt Passage Times*. (Cambridge University Press, 2001).
46. P. Ma, R. Elber and D. E. Makarov, J Chem Theory Comput **16** (10), 6077-6090 (2020).
47. A. Szabo, K. Schulten and Z. Schulten, J Chem Phys **72**, 4350 (1980).
48. C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. (Springer-Verlag, Berlin, 1983).
49. H. S. Chung and I. V. Gopich, Phys Chem Chem Phys **16** (35), 18644-18657 (2014).
50. A. M. Berezhkovskii, S. M. Bezrukov and D. E. Makarov, J Chem Phys **154** (11), 111101 (2021).
51. X. Li and A. B. Kolomeisky, J Chem Phys **139** (14), 144106 (2013).
52. A. L. Thorneywork, J. Gladrow, Y. Qing, M. Rico-Pasto, F. Ritort, H. Bayley, A. B. Kolomeisky and U. F. Keyser, Sci Adv **6** (18), eaaz4642 (2020).
53. B. Peters, *Reaction Theory and Rare Events* (Elsevier, 2017).
54. A. M. Berezhkovskii and D. E. Makarov, J Chem Phys **148** (20), 201102 (2018).
55. A. M. Berezhkovskii, M. A. Pustovoit and S. M. Bezrukov, Journal of Chemical Physics **116**, 9952-9956 (2002).
56. W. E and E. Vanden-Eijnden, J. Stat. Phys. **123** (3), 503-523 (2006).
57. J. Lu and E. Vanden-Eijnden, J Chem Phys **141** (4), 044109 (2014).
58. E. Vanden-Eijnden, in *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*, edited by M. M. Ferrario, G. Ciccotti and K. Binder (Springer, 2006).
59. B. Peters, Annu Rev Phys Chem **67**, 669-690 (2016).
60. A. M. Berezhkovskii and A. Szabo, J Phys Chem B **117** (42), 13115-13119 (2013).

61. P. V. Banushkina and S. V. Krivov, J Chem Phys **143** (18), 184108 (2015).
62. P. V. Banushkina and S. V. Krivov, Biochem Soc Trans **43** (2), 157-161 (2015).
63. N. D. Socci, J. N. Onuchic and P. G. Wolynes, J. Chem. Phys. **104**, 5860-5868 (1996).
64. D. Klimov and D. Thirumalai, Phys. Rev. Lett. **79**, 317 (1997).
65. K. Neupane, A. P. Manuel and M. Woodside, Nature Physics **12**, 700-703 (2016).
66. P. I. Zhuravlev, M. Hinczewski, S. Chakrabarti, S. Marqusee and D. Thirumalai, Proc Natl Acad Sci U S A **113** (6), E715-724 (2016).
67. S. M. Avdoshenko and D. E. Makarov, J Phys Chem B **120** (8), 1537-1545 (2015).
68. A. M. Berezhkovskii and D. E. Makarov, Journal of Chemical Physics **151**, 065102 (2019).