Transition-path properties for folding reactions in the limit of small barriers

Andrew G. T. Pyo, Noel Q. Hoffer, Krishna Neupane, and Michael T. Woodside

Citation: J. Chem. Phys. **149**, 115101 (2018); doi: 10.1063/1.5046692
View online: https://doi.org/10.1063/1.5046692
View Table of Contents: http://aip.scitation.org/toc/jcp/149/11
Published by the American Institute of Physics

Transition-path properties for folding reactions in the limit of small barriers

Andrew G. T. Pyo, Noel Q. Hoffer, Krishna Neupane, and Michael T. Woodside
Department of Physics, University of Alberta, Edmonton, Alberta T6G 2E1, Canada

(Received 29 June 2018; accepted 3 September 2018; published online 19 September 2018)

Transition paths are of great interest because they encapsulate information about the mechanisms of barrier-crossing reactions. Analysis of experiments measuring biomolecular folding reactions has relied on expressions for properties of transition paths such as transition-path times and velocities that hold in the limit of large harmonic barriers, but real molecules often have relatively small barriers. Recent theoretical work presented more general expressions for transition-path properties. Here we extend this work, deriving expressions from the general case that can be applied to small harmonic barriers. We first compared the performance of small-barrier, large-barrier, and general solutions when applied to simulated transitions, focusing on improvements in estimates of the diffusion coefficient determined from transition times and velocities. We then applied these expressions to experimental data from force spectroscopy measurements of DNA hairpins. We found that the low-barrier approximation and exact solution reduced or resolved the small but systematic inconsistencies that had arisen from assuming large harmonic barriers, demonstrating the practical utility of the new equations for analyzing experimental data. *Published by AIP Publishing.* https://doi.org/10.1063/1.5046692

I. INTRODUCTION

Transition paths are those parts of a reaction during which the barrier region separating initial and final states is crossed without returning (Fig. 1). The properties of transition paths reflect crucial information about reaction mechanisms, motivating recent work focusing on transition paths in biomolecular folding owing to the new insight they can provide into the microscopic mechanisms of a very complex process.[1] In particular, measurements of transition paths in folding reactions have provided information about fundamental parameters describing the folding dynamics including the properties of the free-energy landscape underlying the dynamics,[2,3] the diffusion coefficient governing the time scale for motions over the landscape,[3–5] internal friction within the molecule,[6,7] and the reaction coordinate used to describe the folding.[8,9]

Although transition paths contain a wealth of information, they are extremely difficult to observe experimentally owing to their brief durations, and it is only in the last few years that it has been possible to study them directly. Experimental studies of transition paths have made use of a few central results from theoretical work on transition paths obtained under the assumption of a large, harmonic barrier. Initial work[10,11] measuring the average transition-path time, τ_{TP}, for example, made use of an expression for τ_{TP} valid for harmonic barriers in the large-barrier limit,[12,13]

$$\tau_{TP} \approx \frac{\ln\left(2e^{\gamma}\beta\Delta G^{\ddagger}\right)}{\beta D\kappa_b}, \tag{1}$$

where γ is the Euler-Mascheroni constant, $\beta = 1/k_BT$ is the reciprocal of the thermal energy, $\Delta G^{\ddagger}$ is the barrier height, D is the diffusion coefficient, and κ_b is the curvature of the barrier. An expression for the distribution of transition times, $P_{TP}(t)$, was also derived in the same limit[12,14] and used to fit the distribution of times observed experimentally for individual transition paths in proteins and nucleic acids,[3,4]

$$P_{TP}(t) \approx \frac{\omega_K\sqrt{\beta\Delta G^{\ddagger}}}{1-\mathrm{erf}\sqrt{\beta\Delta G^{\ddagger}}}\frac{\exp\left[-\beta\Delta G^{\ddagger}\coth(\omega_K t/2)\right]}{\sinh(\omega_K t/2)\sqrt{2\pi\sinh(\omega_K t)}}, \tag{2}$$

where $\omega_K = \beta D\kappa_b$ is the inverse of the relaxation time. More recently, theoretical investigations into the shape of transition paths[15–17] led to an expression for the average velocity along a transition path in the limit of a large harmonic barrier,[15]

$$\langle v_{TP}(x)\rangle \approx \left\langle v(x^{\ddagger})\right\rangle[1+(x\omega_K/\langle v(x^{\ddagger})\rangle)^2]^{1/2}$$

with

$$\langle v(x^{\ddagger})\rangle = 2(e^{-\gamma}D\omega_K)^{-1/2}, \tag{3}$$

where $x^{\ddagger}$ is the location of the top of the barrier. Equation (3) was used to analyze recent measurements of the average velocity on transition paths in DNA hairpins.[5]

These theoretical descriptions of transition-path properties in the limit of large harmonic barriers have generally described the experimental data quite well. Using them to deduce the value of D for various molecules, for example, returned results that were broadly self-consistent and similar to the values obtained from analysis of kinetic rates using Kramers' theory,[2–5,11] which also assumes a harmonic barrier.[18] However, high-resolution measurements of the shape of the energy barrier[4,19,20] reveal that although approximating the barriers as harmonic is reasonable, anharmonicity is usually present to some degree. Indeed, small but systematic differences between the estimate of D obtained from Eqs. (1)–(3) were attributed in part to barrier anharmonicity.[4,5] Furthermore, many molecules have barriers that are small or even

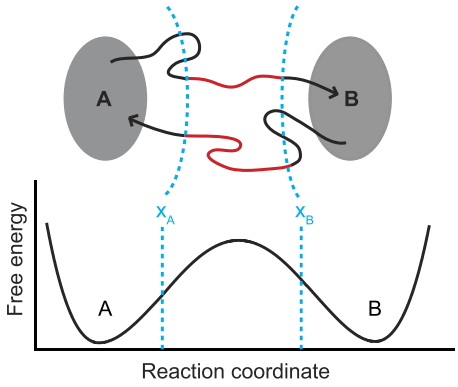

FIG. 1. **Transition paths**: Transition paths are the productive part (red) of the trajectories in a reaction from A to B or vice versa—in contrast to the non-productive fluctuations (grey)—that traverse the energy barrier separating A from B defined by the boundaries x_A and x_B.

non-existent,[21] for which Eqs. (1)–(3) would not be applicable. These considerations motivated work to find expressions for the transition-path times and velocities that are both simple enough to use for analyzing experimental data and valid under more general barrier conditions.

Berezhkovskii and Makarov[22] recently derived expressions for the average time and velocity as a function of position along the transition path without assuming large harmonic barriers. Here we extend this previous work, obtaining expressions that can be used in the limit of small harmonic barriers. We compare the effectiveness of the approximations (high- and low-barrier limits) and exact theory when applied to simulated folding trajectories over harmonic and anharmonic barriers with a range of heights, finding that the crossover to the low-barrier regime occurs at barrier heights of ~4 k_BT. Finally, we apply the theory to experimental data that was previously analyzed with Eqs. (1) and (3), re-evaluating the diffusion coefficient extracted from the data to quantify the error in the estimate of D introduced by approximating the barriers as large and harmonic.

II. METHODS

Simulations were performed as described previously.[5,23] Briefly, trajectories of transition paths were found by the numerical solution of the Itô stochastic differential equation using the Milstein scheme.[24] The time step in the simulations was set to $\Delta t \approx \tau_{TP}/1500$. The simulations were done for different barrier shapes as described in the main text, using a constant diffusion coefficient of D = 500 nm^2/s (a value chosen for convenience) and a thermal energy of k_BT = 4.1 × 10^{-21} J. At least 5000 simulated transitions were generated for each barrier shape. The simulations were analyzed to determine τ_{TP} and $\langle v_{TP}(x)\rangle$, following the methods described previously for analyzing experimental transition paths.[3,5]

III. RESULTS

We first re-derive the general expressions for $\langle v_{TP}(x)\rangle$ and $\langle t_{TP}(x)\rangle$ presented in Ref. 22 in terms of the committor function, using a slightly different approach that starts by assuming that folding is described by a 1D Smoluchowski equation,

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{\partial}{\partial x}D(x)e^{-\beta G(x)}\frac{\partial}{\partial x}e^{\beta G(x)}\rho(x,t). \tag{4}$$

Here $x(t)$ is the trajectory of the reaction coordinate, $\rho(x,t)$ is the probability density function, and $G(x)$ is the free-energy landscape governing the folding. An advantage of this approach is that it describes the time evolution of the probability density, which is in principle an experimental observable. We re-express the reaction dynamics in terms of the committor function (also known as the splitting probability), $\phi_B(x)$, the probability that the molecule will reach state B as a function of the reaction coordinate, which is given by[25]

$$\phi_B(x) = \frac{\int_{x_A}^{x} D(x')^{-1}e^{\beta G(x')}dx'}{\int_{x_A}^{x_B} D(x')^{-1}e^{\beta G(x')}dx'}, x \in (x_A, x_B). \tag{5}$$

Here x_A and x_B define the boundaries between the barrier region and, respectively, the folded and unfolded states (Fig. 1). Following previous work,[23] the Smoluchowski equation conditioned for a transition event expressed in terms of ϕ_B becomes

$$\frac{\partial}{\partial t}\rho_{TP}(x,t) = \frac{\partial}{\partial x}D(x)\phi_B{}^2(x)e^{-\beta G(x)}\frac{\partial}{\partial x}\frac{e^{\beta G(x)}}{\phi_B{}^2(x)}\rho_{TP}(x,t), \tag{6}$$

where $\rho_{TP}(x,t) \equiv \rho(x,t|x_A \to x_B)$, with an absorbing boundary at $x = x_B$ and subject to the initial condition $\rho(x, t_0|x_A \to x_B) = \lim_{\varepsilon\to 0^+}\delta(x - x_A - \varepsilon)$. Equation (6) yields the time evolution of the probability density of the transition paths, which can be readily solved numerically.[26]

The average velocity as a function of position along the transition path, $\langle v_{TP}(x)\rangle$, is then given by

$$\langle v_{TP}(x)\rangle = \frac{\int_{t_0}^{\infty} J(x,t)dt}{\int_{t_0}^{\infty}\rho_{TP}(x,t)dt}, \tag{7}$$

where $J(x,t) \equiv J(x,t|x_A \to x_B)$ is the probability current. Since the probability distribution for transition paths is given by[27]

$$\int_{t_0}^{\infty}\rho_{TP}(x,t)dt = \frac{\phi_B(x)(1-\phi_B(x))e^{-\beta G(x)}}{\int_{x_A}^{x_B}\phi_B(x)(1-\phi_B(x))e^{-\beta G(x)}dx}, \tag{8}$$

we can combine Eqs. (6) and (8) with the continuity equation for particle conservation to solve explicitly for the time integral of J, which is just the reciprocal of τ_{TP},

$$\int_{t_0}^{\infty}J(x,t)dt = \left[\int_{x_A}^{x_B}\phi_B(x)(1-\phi_B(x))e^{-\beta G(x)}dx \times \int_{x_A}^{x_B}D(x)^{-1}e^{\beta G(x)}dx\right]^{-1} = \tau_{TP}{}^{-1}, \tag{9}$$

as found previously.[22] From here, we recapitulate the results from Ref. 22. In terms of the committor function, the velocity is given by

$$\langle v_{TP}(x)\rangle = D(x)\frac{\frac{d}{dx}\phi_B(x)}{\phi_B(x)(1-\phi_B(x))}, \tag{10}$$

which for constant diffusivity yields D in terms of the velocity at the barrier top,

$$D = \frac{1}{4}\left\langle v\left(x^{\ddagger}\right)\right\rangle e^{\beta G(x^{\ddagger})}\int_{x_A}^{x_B} e^{\beta G(x)}dx. \tag{11}$$

Last, as shown previously,[22] by integrating the reciprocal of the transition velocity, the average time to reach the position x along the transition path can be expressed as

$$\langle t_{TP}(x|x_A \to x_B)\rangle = \int_{x_A}^{x} \phi_B(x')(1-\phi_B(x'))e^{-\beta G(x')}dx' \times \int_{x_A}^{x_B} D(x)^{-1}e^{\beta G(x)}dx. \tag{12}$$

In the case of constant D, the diffusion coefficient and the mean transit time are related by[27]

$$D = \int_{x_A}^{x_B} e^{-\beta G(x)}\left(\int_{x_A}^{x} e^{\beta G(x')}dx'\right)\left(\int_{x}^{x_B} e^{\beta G(x')}dx'\right)dx \times\left[\langle\tau_{TP}\rangle\int_{x_A}^{x_B} e^{\beta G(x')}dx'\right]^{-1}. \tag{13}$$

Given that $\langle v(x^{\ddagger})\rangle$ and τ_{TP} are experimental observables in transition-path measurements[3,5] and the energy-barrier profile can sometimes be measured directly, as with the single-molecule force spectroscopy methods,[28] Eqs. (11) and (13) provide a more precise way to determine D, the fundamental parameter that connects the thermodynamics of the energy landscape to the dynamics of the folding.

Because energy barriers in folding reactions are often approximated as harmonic and D is typically assumed to be constant, it is useful to re-express the equations above for this special case. For a parabolic potential $G(x) = -\frac{1}{2}\kappa x^2$, where κ is the curvature of the barrier and the transition region spans the range $x = -L$ to L, the transition-path velocity for constant D is[22]

$$\langle v_{TP}(x|-L \to L)\rangle = \frac{4D\alpha\,\mathrm{erf}(\alpha L)e^{-\alpha^2x^2}}{\sqrt{\pi}\left(\mathrm{erf}^2(\alpha L)-\mathrm{erf}^2(\alpha x)\right)}, \tag{14}$$

where $\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-x'^2}dx'$ is the error function and $\alpha = \sqrt{\frac{\beta\kappa}{2}}$ [hence $(\alpha L)^2$ is the barrier height]. D is then given by

$$D = \frac{\sqrt{\pi}}{4\alpha}\mathrm{erf}(\alpha L)\left\langle v\left(x^{\ddagger}\right)\right\rangle. \tag{15}$$

These results are valid for harmonic barriers regardless of the barrier height. Approximations have already been computed in the high-barrier limit,[12,15–17] but not yet in the low-barrier limit. For small barriers, we use the approximation $\mathrm{erf}^2(x) \approx 1 - \exp(-\pi^2x^2/8)$ for small x, leading to

$$\langle v_{TP}(x|-L \to L)\rangle = \frac{2\sqrt{2}D\alpha}{\sqrt{\pi}}\frac{\sqrt{\sinh\left(\frac{\pi^2}{16}\alpha^2L^2\right)}e^{\frac{\pi^2}{32}\alpha^2L^2}}{\sinh\left(\frac{\pi^2}{16}\alpha^2(L^2-x^2)\right)e^{\left(1-\frac{\pi^2}{16}\right)\alpha^2x^2}} \tag{16}$$

and

$$\langle t_{TP}(x|-L \to L)\rangle = \frac{\pi\left[\lambda\left(\mathrm{erf}^2(\alpha L)-1\right)(\mathrm{erfi}(\alpha x)+\mathrm{erfi}(\alpha L))+\mathrm{erf}(\lambda\alpha x)+\mathrm{erf}(\lambda\alpha L)\right]}{8D\lambda\alpha^2\mathrm{erf}(\alpha L)}, \tag{17}$$

where $\mathrm{erfi}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{x'^2}dx'$ and $\lambda = \sqrt{\frac{\pi^2}{8}-1}$. Finally, by evaluating Eq. (17) at $x = L$, the average transition-path time can be found,

$$\tau_{TP} = \frac{\pi\left[\lambda\left(\mathrm{erf}^2(\alpha L)-1\right)\mathrm{erfi}(\alpha L)+\mathrm{erf}(\lambda\alpha L)\right]}{4D\lambda\alpha^2\mathrm{erf}(\alpha L)}. \tag{18}$$

We note that these results are still valid as the barrier height goes to zero. Taking the limit $\kappa \to 0$, Eq. (16) for the average transition-path velocity simplifies to $\langle v_{TP}(x)\rangle \approx 2DL/(L^2 - x^2)$, consistent with the exact solution; similarly, Eq. (18) for the average transition-path time simplifies to $\tau_{TP} \approx 2L^2/3D$, consistent with the expression found previously by Kim and Netz.[16]

A. Application to Brownian dynamics simulations

To test these equations and discern the practical effects of using them to analyze data from molecules with different barriers, we first applied them to analyze transition paths in Brownian dynamics simulations of motion with constant diffusion over parabolic barriers of varying barrier heights. We started by examining the average velocity profile, $\langle v_{TP}(x)\rangle$, for two barrier heights: 15 k_BT (high-barrier limit) and 3 k_BT (low-barrier limit). We found that the average velocity calculated directly from the simulated trajectories (Fig. 2, black) matched the analytical predictions from Eq. (14) (Fig. 2, red) very well for both barriers. By contrast, although the high-barrier approximation, Eq. (3), matched the empirical result for the high-barrier case [Fig. 2(a), cyan], as expected, it systematically underestimated the velocity in the low-barrier case [Fig. 2(b), cyan]. Likewise, the low-barrier approximation, Eq. (16), matched the empirical result well for the low-barrier case [Fig. 2(b), blue], but systematically overestimated the velocity in the high-barrier case [Fig. 2(a), blue].

We next considered the average transition-path time, τ_{TP}. Transitions were simulated over harmonic barriers with heights ranging from 1 to 6 k_BT. The values for τ_{TP} obtained

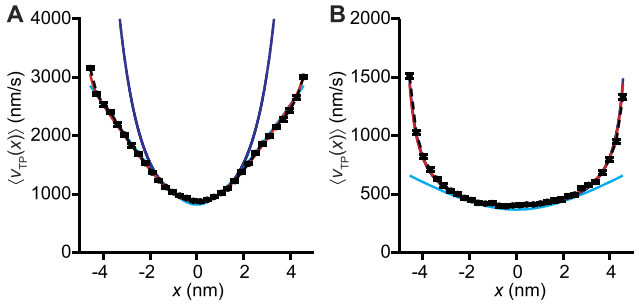

FIG. 2. **Average transition-path velocity from simulated transitions over harmonic barriers**: (a) The average transition-path velocity $\langle v_{TP}(x)\rangle$ found directly from simulations of transitions over a large, 15-k_BT harmonic barrier (black) matches very well with the result predicted by the exact solution in Eq. (14) (red) and the high-barrier approximation in Eq. (3) (cyan) but is poorly described by the low-barrier approximation in Eq. (16) (blue). (b) The average velocity found directly from simulations of transitions over a small, 3-k_BT harmonic barrier (black) agrees well with the predictions of both the exact result (red) and the low-barrier approximation (blue) but is poorly described by the high-barrier approximation (cyan).

empirically for each barrier height from the simulated transitions (Fig. 3, black) were compared to the values predicted by Eq. (1) in the high-barrier limit (Fig. 3, cyan), to the values predicted by Eq. (18) in the low-barrier limit (Fig. 3, blue), and to the values predicted by the exact solution, Eq. (12) (Fig. 3, red). Whereas the exact solution matched the empirical results very well over the whole range of barrier heights, the high-barrier approximation overestimated τ_{TP} for barriers lower than ~4 k_BT, and the low-barrier approximation underestimated τ_{TP} for barriers higher than ~4 k_BT. The crossover between high-barrier and low-barrier approximations was thus around 4 k_BT. The under- and over-estimations resulting, respectively, from the low- and high-barrier approximations were not very large in the range of barrier heights studied, however, being less than 10%–15%.

Many of the experimental studies of transition paths to date have used the properties of transition paths to investigate the diffusion coefficient D[3–7] because transition-path properties are far less sensitive to the height of the energy barrier (which is difficult to measure precisely) than are other kinetic properties like rates.[1,6,29] We therefore examined the reliability of the various estimates of D that can be obtained from transition paths using the formulae above, in the different limits. Looking first at the value of D obtained from $\langle v(x^{\ddagger})\rangle$, we used simulations over harmonic barriers with heights varying from 1 to 6 k_BT to compare the actual value of D imposed in the simulations [Fig. 4(a), black] to the value calculated from the exact solution via Eq. (15) [Fig. 4(a), red]. We found that Eq. (15) returned the expected value of D over the whole range of barrier heights. Repeating the analysis using the low-barrier approximation [Eq. (16)] yielded very similar results [Fig. 4(a), blue], indicating that this approximation is reasonable for barrier heights up to at least 6 k_BT. However, the large-barrier approximation [Eq. (3)] tended to overestimate D [Fig. 4(a), cyan], with the estimate worsening noticeably (albeit not dramatically) for barrier heights below ~3 k_BT. We note that a ~6% overestimate of D from the high-barrier approximation persists even in the high-barrier limit. Applying the same kind of analysis to τ_{TP}, we found that the values of D returned by the exact solution, Eq. (13) [Fig. 4(b), red], again agreed very well with the value imposed in the simulations over the range of barrier heights used [Fig. 4(b), black], whereas the high-barrier approximation [Eq. (1)] tended to overestimate D, with the overestimation increasing noticeably but not dramatically for barrier heights ~4 k_BT and below [Fig. 4(b), blue]. By contrast, the low-barrier approximation [Eq. (18)] tended to underestimate D somewhat for barrier heights ~5 k_BT and above [Fig. 4(b), blue].

Finally, we looked at the effects of relaxing the assumption that the barrier is harmonic by simulating transitions over two types of anharmonic barriers: one with a barrier that is sharper than in the harmonic case, described by $G(x) = -\Delta G^{\ddagger}[\tanh(x/2)/\tanh(L/2)]^2$ [Fig. 5(a), inset], where $\Delta G^{\ddagger}$ is the barrier height; and the other with a barrier that is broader than in the harmonic case, described by $G(x) = -\Delta G^{\ddagger}[\sinh(x/2)/\sinh(L/2)]^2$ [Fig. 5(c), inset]. In each case,

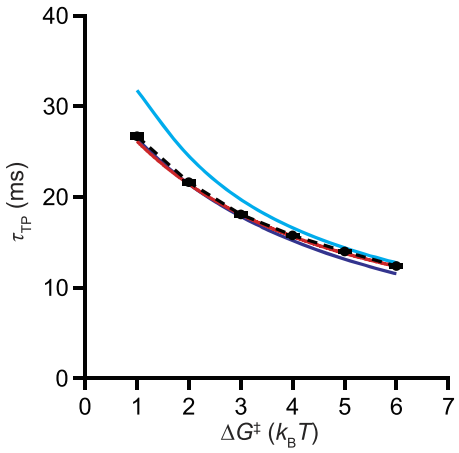

FIG. 3. **Average transition-path time from simulated transitions over harmonic barriers**: The average transition-path time determined directly from simulations over a harmonic barrier with height ranging from 1 to 6 k_BT (black) agrees well with the exact solution from Eq. (13) (red) over the whole range of barrier heights. The high-barrier approximation from Eq. (1) (cyan) agrees well at barrier heights of ~5 k_BT and higher, whereas the low-barrier approximation from Eq. (18) (blue) agrees well at barrier heights of ~4 k_BT and lower.

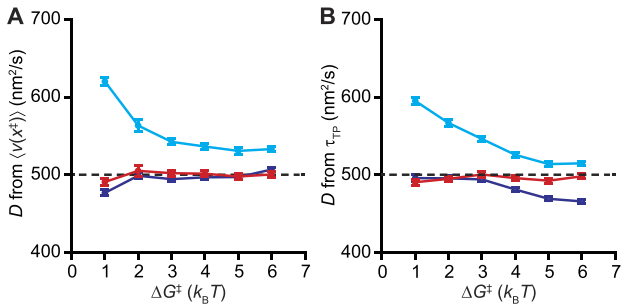

FIG. 4. **Estimates of the diffusion coefficient for simulations over harmonic barriers**: (a) The estimate for D from the exact solution for $\langle v(x^{\ddagger})\rangle$ via Eq. (15) (red) agrees with the value imposed in the simulation (black) across the whole range of barrier heights, whereas the high-barrier solution (cyan) systematically overestimates D, especially for barriers below ~4 k_BT. (b) The estimate for D from the exact solution for τ_{TP} (red), determined from the simulation results by solving Eq. (13) for D, agrees with the value imposed in the simulation (black) across the whole range of barrier heights, whereas the high-barrier solution (cyan) overestimates D for barriers below ~5 k_BT and the low-barrier solution (blue) underestimates D for barriers above ~4 k_BT.

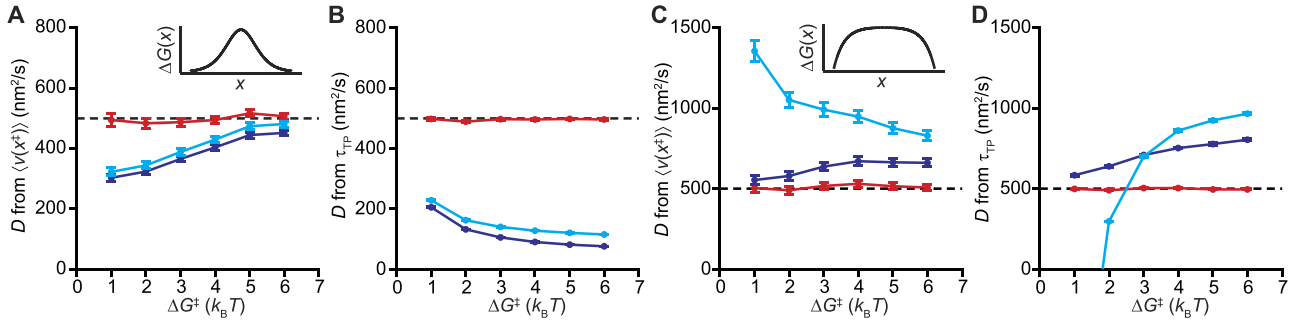

FIG. 5. **Estimates of the diffusion coefficient for anharmonic barriers**: (a) The estimate for D from the exact solution for $\langle v(x^{\ddagger})\rangle$ via Eq. (11) (red) agrees with the value imposed (black) in simulations of transitions over a barrier with higher curvature at the top than a harmonic barrier of the same width and height (inset) across the whole range of barrier heights, whereas the high-barrier (cyan) and low-barrier (blue) harmonic approximations both underestimate D for barriers below ~5–6 k_BT. (b) The estimate for D from the exact solution for τ_{TP} via Eq. (13) (red) agrees with the value imposed in the simulations (black) across the whole range of barrier heights, whereas the high-barrier (cyan) and low-barrier (blue) harmonic approximations both significantly underestimate D for all barrier heights. (c) For simulations over a barrier with lower curvature at the top than harmonic (inset), estimates of D from the exact solution for $\langle v(x^{\ddagger})\rangle$ (red) agree with the imposed value (black), whereas the high-barrier (cyan) and low-barrier (blue) harmonic approximations both overestimate D; the low-barrier approximation performs markedly better, especially for small barriers. (d) The estimate for D from the exact solution for τ_{TP} (red) agrees with the value imposed in the simulations (black) for all barrier heights, whereas the high-barrier (cyan) and low-barrier (blue) harmonic approximations overestimate D. At the smallest barriers, the high-barrier approximation yields unphysical negative values.

$\Delta G^{\ddagger}$ was varied from 1 to 6 k_BT, as for the harmonic barriers above. Considering first the sharper barrier, we found excellent agreement between the D values predicted from $\langle v(x^{\ddagger})\rangle$ by Eq. (11) [Fig. 5(a), red] and the imposed value [Fig. 5(a), black]. However, approximating the barrier as harmonic and using Eq. (3) for the high-barrier limit yielded values that were systematically too low [Fig. 5(a), cyan], diverging from the imposed value for barrier heights lower than ~5 k_BT. The same held true when using Eq. (16) for the low-barrier limit—it was no better even at low barrier heights. The value of D recovered from τ_{TP} from the exact solution, Eq. (13) [Fig. 5(b), red], agreed well with the imposed value [Fig. 5(b), black], but the harmonic approximations in the high-barrier [Fig. 5(b), cyan] and low-barrier [Fig. 5(b), blue] limits both yielded values that were several-fold lower for all barrier heights examined. Turning next to the barrier that was broader than harmonic, the value of D recovered from $\langle v(x^{\ddagger})\rangle$ by Eq. (11) [Fig. 5(c), red] agreed well with the imposed value, as above, but now the approximation as a large harmonic barrier overestimated D by 2–3-fold [Fig. 5(c), cyan], whereas the small harmonic barrier approximation overestimated D by only 10%–40% [Fig. 5(c), blue]. Looking instead at D from τ_{TP}, the exact solution [Fig. 5(d), red] recovered the imposed value [Fig. 5(d), black], and the two harmonic barrier approximations again tended to overestimate D, except in the case of the large-barrier approximation at the smallest barrier heights, where the solution yielded an unphysical negative value [Fig. 5(d), cyan].

B. Application to experimental data

The results from analysis of the simulated transitions show the kinds of errors that can arise from approximating the barrier shape. Indeed, it was previously noted that applying Eqs. (1)–(3), which assume large harmonic barriers, to transition paths measured for DNA hairpins using optical tweezers[4,5] led to small but systematic differences between the values of D recovered from different physical quantities (average transition-path times, decay of the transition-time distributions, and barrier-top velocities). A breakdown in the assumptions underlying Eqs. (1)–(3) was proposed to explain the discrepancies in the results. We therefore applied the formulae above to test the extent to which the assumption of a large harmonic barrier distorted the results of the analysis.

We used the constant-D expressions for $\langle v(x^{\ddagger})\rangle$ and τ_{tp} in Eqs. (11), (13), (15), and (18) to re-evaluate the value of D implied by the measurements of two hairpins, based on the shape of the energy barrier that had previously been measured for each hairpin.[4,19] For hairpin 20R25/T4, which has a barrier that is close to harmonic but relatively small (height measured with respect to the boundaries of the barrier region of 2.3 k_BT),[4] we found that applying Eq. (15) to estimate D from τ_{TP} yielded a new value of $D = 1.9 \pm 0.5 \times 10^5$ nm^2/s, down from the previous estimate[4] using Eq. (1) of $D = 2.6 \pm 0.3 \times 10^5$ nm^2/s. Applying instead Eq. (18) to estimate D from $\langle v(x^{\ddagger})\rangle$ yielded a new value of $D = 2.1 \pm 0.3 \times 10^5$ nm^2/s, again reduced from the previous estimate[5] using Eq. (3) of $D = 2.6 \pm 0.5 \times 10^5$ nm^2/s. These new estimates were both closer to the value obtained from Eq. (2) ($D = 1.3 \pm 0.2 \times 10^5$ nm^2/s) than the previous estimates, reducing the variance between the three estimates over 3-fold compared to the previous work.[5] Note that these values on the order of 10^4–10^6 nm^2/s are consistent with previous measurements and modeling of nucleic acid folding.[11,30,31]

We repeated this analysis for hairpin 30R50/T4, which has a barrier that is roughly twice as large as the barrier for hairpin 20R25/T4 but more anharmonic. Applying Eq. (13) to estimate D from τ_{TP} led to a new value of $D = 2.4 \pm 0.3 \times 10^5$ nm^2/s, instead of $3.5 \pm 0.3 \times 10^5$ nm^2/s using Eq. (1),[4] whereas applying Eq. (11) led to a new estimate from $\langle v(x^{\ddagger})\rangle$ of $D = 2.0 \pm 0.4 \times 10^5$ nm^2/s, instead of $2.5 \pm 0.4 \times 10^5$ nm^2/s using Eq. (3).[5] Again, both of these new estimates are more consistent with the value found from Eq. (2), $D = 1.8 \pm 0.2 \times 10^5$ nm^2/s, in this case reducing the variance between the different estimates roughly 8-fold compared to what was found previously.[4] For both hairpins, then, going beyond the large harmonic barrier approximation helped deliver improved, more consistent estimates of D.

IV. DISCUSSION

The harmonic approximation for energy barriers is almost always used to interpret folding data, not only in the context of transition paths as discussed here but also in terms of rates, such as in the celebrated expression of Kramers for diffusive barrier crossing.[18] The large harmonic barrier approximation has worked reasonably well in initial studies of transition paths in proteins and nucleic acids, for example, yielding values of D that are consistent within factors of order unity with values obtained via other means like Kramers' theory.[2,3] However, as future experiments aim to increase the precision with which fundamental quantities like D are determined from transition-path measurements, expressions for key observables like $\langle \nu_{\mathrm{TP}}(x) \rangle$ and τ_{TP} that enable analysis to go beyond the limit of large harmonic barriers and incorporate the effects of small barrier heights or anharmonicity must be applied. Our work shows that the general expressions for $\langle \nu_{\mathrm{TP}}(x) \rangle$ and τ_{TP} derived in Ref. 22 and the low-barrier limits derived herein can indeed help improve the precision of the data analysis.

The results from analyzing the simulations indicate some of the trends that can be expected from using specific approximations when analyzing experimental data. For harmonic barriers, the large-barrier assumption generated only modest errors, even down to barriers as small as 1 $k_{\mathrm{B}}T$, where the actual D, for example, was ~20%–25% smaller than estimated. Nevertheless, the small-barrier approximation provided distinct improvements at the lowest barrier heights. The presence of anharmonicity, on the other hand, led to much larger errors when applying the large and small harmonic barrier approximations. For barriers that have higher curvature at the top than harmonic barriers of the same height and width would have, the estimates of D are systematically lower than they should be because the high curvature of the harmonic approximation implies a much higher barrier than is actually the case with the anharmonic barrier, and hence a faster velocity and shorter transit time than actually occurs. This explanation also accounts for the poor performance of the low-barrier approximation: even for nominally small barriers, the implied harmonic barrier height is still large. For barriers that are flatter than harmonic, on the other hand, D is overestimated by applying the harmonic approximation because the actual barrier is higher than implied by the harmonic approximation and hence the velocity and transit time are faster than would be expected in the approximation. Again, this explanation also accounts for the improved performance of estimates using the small-barrier approximation. Taken together, these results imply that estimates of D using the harmonic approximation are considerably more sensitive to the presence of anharmonicity than they are to the height of the barrier. Note that the fact that barrier anharmonicity has a greater effect on D than barrier height could be expected based on Eq. (1), where the transition path time has a linear dependence on the barrier's curvature but a much weaker, logarithmic dependence on its height.

From the simulation results, we can also determine the crossover point at which the high-barrier approximations start to generate noticeable errors. In almost every case, the high-barrier approximations began to diverge from the low-barrier or exact results at a barrier height of 3–5 $k_{\mathrm{B}}T$, allowing the high-barrier limit to be defined fairly confidently as anything above this level. Previous measurements of transition paths have examined molecules with barrier heights close to but sometimes slightly below this boundary, explaining why the high-barrier approximation worked reasonably well but generated minor discrepancies between estimates of quantities like the diffusion coefficient that were obtained from different methods. As seen above, applying more appropriate approximations or exact formulas helped to resolve these discrepancies.

The main challenge with applying the equations for τ_{TP} and $\langle \nu_{\mathrm{TP}}(x) \rangle$ is that they require detailed knowledge of the shape of the energy barrier. Energy-barrier shapes can in some cases be measured directly, as with the single-molecule force spectroscopy methods used here,[28] or else they may be estimated from simulations and/or modeling.[32–35] Ideally, the position-dependence of the diffusion coefficient should also be known since the common approximation that D is constant is not generally correct.[36,37] This position dependence can in some cases be deduced by combining experiments and modeling,[38] but reliable methods for direct measurement remain to be established;[39] hence, the constant-D approximation may often be unavoidable in practical applications.

Although we have used nucleic acid folding to demonstrate the application of Eqs. (10)–(18), these results are, of course, equally applicable to protein folding. Theoretical considerations and experimental measurements[34,40–44] both suggest that energy barriers for protein folding are generally small compared to the total free-energy change. In many cases, they may be close to the 3–5 $k_{\mathrm{B}}T$ limit for assuming a large barrier, making the results presented here useful for obtaining better estimates of the transition-path properties. The small-barrier limit will be particularly relevant for studies of molecules with very small or even non-existent barriers, such as fast-folding or downhill-folding proteins, which continue to be the target of considerable interest because their small barriers allow transition states to be populated more easily.[45]

V. CONCLUSIONS

We showed that expressions for important transition-path properties like the average time for transitions and the average velocity along the transition path that are valid outside the limit of the large-harmonic-barrier approximation led to improved accuracy for deducing properties of the folding from simulations when applied to small and/or anharmonic energy barriers, and that they resolved discrepancies observed in the analysis of experimental transition-path data. These expressions should prove useful to obtain more precise evaluations of the microscopic properties of folding reactions as transition-path measurements continue to expand to a widening range of molecules.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council Canada and Alberta Innovates

Technology Solutions. M.T.W. was supported by a Guggenheim Fellowship.

[1]H. S. Chung, J. Mol. Biol. **430**, 409 (2018).
[2]H. Yu, D. R. Dee, X. Liu, A. M. Brigley, I. Sosova, and M. T. Woodside, Proc. Natl. Acad. Sci. U. S. A. **112**, 8308 (2015).
[3]K. Neupane, D. A. N. Foster, D. R. Dee, H. Yu, F. Wang, and M. T. Woodside, Science **352**, 239 (2016).
[4]K. Neupane, F. Wang, and M. T. Woodside, Proc. Natl. Acad. Sci. U. S. A. **114**, 1329 (2017).
[5]K. Neupane, N. Q. Hoffer, and M. T. Woodside, Phys. Rev. Lett. **121**, 018102 (2018).
[6]H. S. Chung and W. A. Eaton, Nature **502**, 685 (2013).
[7]H. S. Chung, S. Piana-Agostinetti, D. E. Shaw, and W. Eaton, Science **349**, 1504 (2015).
[8]K. Neupane, A. P. Manuel, J. Lambert, and M. T. Woodside, J. Phys. Chem. Lett. **6**, 1005 (2015).
[9]K. Neupane, A. P. Manuel, and M. T. Woodside, Nat. Phys. **12**, 700 (2016).
[10]H. S. Chung, K. McHale, J. M. Louis, and W. A. Eaton, Science **335**, 981 (2012).
[11]K. Neupane, D. B. Ritchie, H. Yu, D. A. N. Foster, F. Wang, and M. T. Woodside, Phys. Rev. Lett. **109**, 068102 (2012).
[12]S. Chaudhury and D. E. Makarov, J. Chem. Phys. **133**, 034118 (2010).
[13]H. S. Chung, J. M. Louis, and W. A. Eaton, Proc. Natl. Acad. Sci. U. S. A. **106**, 11837 (2009).
[14]B. W. Zhang, D. Jasnow, and D. M. Zuckerman, J. Chem. Phys. **126**, 074504(2007).
[15]D. E. Makarov, J. Chem. Phys. **143**, 194103 (2015).
[16]W. K. Kim and R. R. Netz, J. Chem. Phys. **143**, 224108 (2015).
[17]P. Cossio, G. Hummer, and A. Szabo, J. Chem. Phys. **148**, 123309 (2018).
[18]H. A. Kramers, Physica **7**, 284 (1940).
[19]M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block, Science **314**, 1001 (2006).
[20]A. P. Manuel, J. Lambert, and M. T. Woodside, Proc. Natl. Acad. Sci. U. S. A. **112**, 7183 (2015).
[21]H. Gelman and M. Gruebele, Q. Rev. Biophys. **47**, 95 (2014).
[22]A. M. Berezhkovskii and D. E. Makarov, J. Chem. Phys. **148**, 201102 (2018).
[23]J. Lu and J. Nolen, Probab. Theory Relat. Fields **161**, 195 (2015).
[24]P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, Berlin, Heidelberg, 1992).
[25]R. Du, V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. S. Shakhnovich, J. Chem. Phys. **108**, 334 (1998).
[26]D. J. Bicout and A. Szabo, J. Chem. Phys. **109**, 2325 (1998).
[27]G. Hummer, J. Chem. Phys. **120**, 516 (2004).
[28]M. T. Woodside and S. M. Block, Annu. Rev. Biophys. **43**, 19 (2014).
[29]M. T. Woodside, J. Lambert, and K. S. D. Beach, Biophys. J. **107**, 1647 (2014).
[30]J.-C. Lin and D. Thirumalai, J. Am. Chem. Soc. **130**, 14080 (2008).
[31]A. Ansari, S. V. Kuznetsov, and Y. Shen, Proc. Natl. Acad. Sci. U. S. A. **98**, 7771 (2001).
[32]N. D. Socci, J. N. Onuchic, and P. G. Wolynes, J. Chem. Phys. **104**, 5860 (1996).
[33]J. E. Shea and C. L. Brooks, Annu. Rev. Phys. Chem. **52**, 499 (2001).
[34]E. R. Henry and W. A. Eaton, Chem. Phys. **307**, 163 (2004).
[35]S. Piana, K. Lindorff-Larsen, and D. E. Shaw, Proc. Natl. Acad. Sci. U. S. A. **109**, 17845 (2012).
[36]J. Chahine, R. J. Oliveira, V. B. P. Leite, and J. Wang, Proc. Natl. Acad. Sci. U. S. A. **104**, 14646 (2007).
[37]R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. U. S. A. **107**, 1088 (2010).
[38]T. Cellmer, E. R. Henry, J. Hofrichter, and W. A. Eaton, Proc. Natl. Acad. Sci. U. S. A. **105**, 18320 (2008).
[39]D. A. N. Foster, R. Petrosyan, A. Pyo, A. Hoffmann, F. Wang, and M. Woodside, Biophys. J. **114**, 1657 (2018).
[40]J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins **21**, 167 (1995).
[41]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**, 14452 (2012).
[42]A. N. Naganathan, J. M. Sanchez-Ruiz, and V. Muñoz, J. Am. Chem. Soc. **127**, 17970 (2005).
[43]D. De Sancho, U. Doshi, and V. Muñoz, J. Am. Chem. Soc. **131**, 2074 (2009).
[44]D. Thirumalai, J. Phys. I **5**, 1457 (1995).
[45]V. Muñoz and M. Cerminara, Biochem. J. **473**, 2545 (2016).