

Article

Accelerating Kinetics with Time-Reversal Path Sampling

Zhirong Liu 

Beijing National Laboratory for Molecular Sciences (BNLMS), College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China; liuzhirong@pku.edu.cn

Abstract: In comparison to numerous enhanced sampling methods for equilibrium thermodynamics, accelerating simulations for kinetics and nonequilibrium statistics are relatively rare and less effective. Here, we derive a time-reversal path sampling (tRPS) method based on time reversibility to accelerate simulations for determining the transition rates between free-energy basins. It converts the difficult uphill path sampling into an easy downhill problem. This method is easy to implement, i.e., forward and backward shooting simulations with opposite initial velocities are conducted from random initial conformations within a transition-state region until they reach the basin minima, which are then assembled to give the distribution of transition paths efficiently. The effects of tRPS are demonstrated using a comparison with direct simulations of protein folding and unfolding, where tRPS is shown to give results consistent with direct simulations and increase the efficiency by up to five orders of magnitude. This approach is generally applicable to stochastic processes with microscopic reversibility, regardless of whether the variables are continuous or discrete.

Keywords: accelerated kinetics; nonequilibrium statistics; time reversibility; enhanced sampling; protein folding

1. Introduction

Thermodynamics and kinetics are two fundamental aspects of physical chemistry. In general, thermodynamics and equilibrium statistics are relatively easy to handle compared to kinetics and nonequilibrium statistics. Based on the ubiquitous Boltzmann distribution, various efficient enhanced sampling methods have been developed to greatly accelerate the molecular simulations of equilibrium properties [1], e.g., umbrella sampling [2,3], the histogram method [4,5], temperature–replica exchange [6–8], integrated tempering sampling [9], and metadynamics [10,11]. In recent years, machine learning techniques were also combined with conventional enhanced sampling to explore the vast conformational space of molecules [12–16].

In contrast, there are less known profound principles for kinetics and nonequilibrium statistics. The Onsager reciprocal relations keenly caught the time-reversal symmetry in the underlying microscopic dynamics to express the equality of certain ratios between different pairs of forces and flows [17–19]. The Jarzynski equality [20], which revealed an unexpected connection between irreversible work and free-energy difference, actually utilized the in

is inherently related to the maximum entropy principle [34], and thus, machine learning has potential applications in path statistics and kinetics computation. Other approaches, such as hyperdynamics (using a bias potential to upshift the free-energy basins) [35–37], transition path theory [38], reactive flux (Bennett–Chandler estimation of the transmission coefficient to correct the transition-state theory approximation) [39,40], and the aimless shooting algorithm [41], were also explored. Overall, accelerating kinetic simulations often rely on more complex assumptions than enhanced samplings of thermodynamics and are usually less effective.

In this paper, we present a related approach to accelerate kinetic simulation for systems with a free-energy barrier. The approach is established based on the time-reversal symmetry of the microscopic dynamics and the existence of equilibrium distribution and is thus expected to be generally applicable. The obtained formula to calculate the transition rate is bias-free and easy to implement.

2. Theoretical Results

The microscopic state of a system is described by a point in the phase space. Imagine that a phase space has many states (similar to the ensemble picture) that evolve and bifurcate (due to stochastics) over time, forming some infinitely long trajectories. To make practical statistics on trajectories, one should use some methods to cut infinitely long trajectories into paths (segments of trajectories) with a length (duration) that is finite. In the literature, there are two main schemes for cutting trajectories. One is to cut trajectories at a fixed time so that the resulting paths have the same duration. This scheme was adopted by TPS [24–26] and S-shooting [42]. Another is to cut trajectories with some fixed planes in the phase space with specified conformational features. For example, for a system with a double-well free-energy profile, the basin minima can be chosen as the cutting point, and the trajectory is cut whenever it crosses the cutting points (Figure 1a). This scheme was adopted in methods such as FFS [29,30], TIS [27,28], and the Bayesian relation method [31–33] and was also adopted in this study.

The thermodynamic properties are determined by the state statistics, while the kinetic properties are embedded in the path statistics, i.e., the distribution of paths. To make the analysis more clear, the trajectories and paths are assumed to be composed of a sequence of conformations (plus velocities if necessary) sampled with a very small time step dt (see Figure 1a). Then, a path- i ($\mathcal{P}_i$) is described in

$$\mathcal{P}_i = \{\mathcal{C}_{i,j}\}, \quad j = 1, \dots, L_i, \quad (1)$$

where $\mathcal{C}_{i,j}</$

ahead (evolve) to occupy the position of the preceding one, regardless of the path duration and speeds, which is convenient for calculation. For the transition/reaction



the kinetic equation is

$$\frac{dN_A}{dt} = -kN_A, \quad (4)$$

where N_A is the number of molecules in basin A. Therefore, the transition rate coefficient k is calculated as

$$k = -\frac{dN_A}{N_A dt} = \frac{1}{dt} \frac{\sum_i H_{A-B}(\mathcal{P}_i)}{\sum_{i,j} H_A(\mathcal{C}_{i,j})}, \quad (5)$$

where the characteristic function $H_{A-B}(\mathcal{P}_i)$ indicates whether the path is classified into the type A–B, and $H_A(\mathcal{C}_{i,j})$ indicates whether the conformation falls in the basin A. The denominator in the above equation includes the contributions of paths A–A, A–B, and B–A, but generally, A–A is dominant, and the others can be ignored (unless in the conditions of high temperatures or the shallow wells, where direct simulations can be easily conducted) to yield

$$k = \frac{1}{dt} \frac{\sum_i H_{A-B}(\mathcal{P}_i)}{\sum_i L_i H_{A-A}(\mathcal{P}_i)}. \quad (6)$$

Therefore, the kinetics can be readily obtained from path statistics. Equation (6) is theoretically exact, so it is inherently equivalent to other path-based approaches with fixed-plane cuts [27–33]. But, it is not very useful on its own since direct sampling is usually inefficient, and thus, further ideation is needed as explained below.

In thermodynamics, sampling can be accelerated by applying bias or other means to enhance the sampling probability at certain regions of phase space. Similarly, methods can be developed to enhance the sampling of certain paths in kinetics simulations. Obviously, a direct simulation to generate trajectories/paths to calculate Equation (6) is very inefficient since most produced paths belong to types A–A and B–B but not transition paths A–B and B–A, and the resulting error in the numerator in Equation (6) is large. Then, comes the main idea of this study (Figure 1b), i.e., to enhance the sampling of

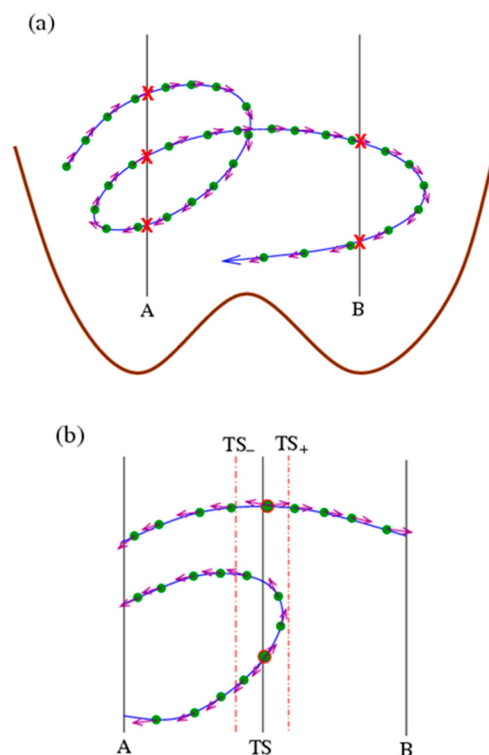


Figure 1. Schematics on accelerating kinetics with time reversibility. **(a)** The construction of the path ensemble. A very long trajectory (blue line) is cut into some short paths (segments) using two cutting planes A and B located at the basin minima of the free-energy profile (brown line). The cutting points are marked with red crossings. Filled green circles represent the conformations sampled on the trajectory with a fixed time step (dt) and violet arrows indicate their velocities. **(b)** A collection of paths crossing the transition state (TS). Starting from an initial conformation (bigger circles with red edge) (obeying equilibrium distribution) within the TS range $[TS_-, TS_+]$ (indicated by dashed-dotted lines), forward and backward simulations with opposite initial velocities are conducted until they reach any cutting planes (A or B), and, with the time-reversal symmetry, they can be assembled to give a path passing the TS range.

In actual simulations, the initial conformations are randomly chosen from an equilibrium distribution within a TS region $[TS_-, TS_+]$, which are determined via umbrella sampling in the examples below, and forward and backward shooting simulations with opposite initial velocities are conducted until they reach basin minima A or B. Then, the backward half path is reversed and assembled with the forward one to give a path passing the TS range, which may be A–A (i.e., A–TS–A), A–B, B–A, and B–B (i.e., B–TS–B) types. It is noted that such a path may contain more than one conformation in the TS region $[TS_-, TS_+]$ (the number of which is denoted as n_{TS}), which have the same probability to be chosen as the initial state to generate the path. To avoid any double counting of the same path, the sampled paths using shooting simulations should be corrected by a weight of $1/n_{TS}$ to be consistent with those in Equation (6). In addition, it is noted that the denominator of Equations (5) and (6) is actually the equilibrium conformation number in basin A. Taking all these together, it yields

$$k = \frac{1}{dt} \frac{N_{TS}}{N_A} \frac{\sum_i H_{A-B}(\mathcal{P}_i)}{N_{TS}} = \frac{1}{dt} \frac{N_{$$

from an equilibrium result N_{TS}/N_A and a shooting simulation, both of which are easy to implement. It is applicable to both continuous and discrete variables. We term the method time-reversal path sampling (tRPS).

It is noted that the forward/backward shooting moves with time reversibility were widely employed in previous path samplings [24–28,31–33,42], e.g., TPS [24–26], TIS [27,28], and the old-fashioned Bennett–Chandler approach [39]. In most cases, they were used as a means to perturb the old paths in order to provide trial paths for Monte-Carlo-like algorithm in constructing the path ensemble slice by slice, and the purpose is to calculate the correlation function or the conditional probability between adjacent slices used in rate formula. In comparison, here we utilized the time reversibility to directly convert the difficult-to-calculate quantity (A–B paths) into an easy-to-calculate quantity (TS–A and TS–B half paths). We did not cast any sampled paths as was carried out in the rejection/acceptance step in MC. We do not need to consider any other intermediate slices except the TS one. Visually speaking, in order to determine the height of the top of the steps, TIS and many closely related methods jump upstairs one by one, while tRPS directly jumps down from the top of the steps to the floor. See Supporting Information for more details.

3. Numerical Results

We test the method on the protein-folding problem [43]. Although protein modeling has advanced rapidly over the past 50 years, a direct approach to simulate protein-folding kinetics is still challenging [44]. We consider a well-known model protein, chymotrypsin inhibitor 2 (CI2) with 64 residues (PDB ID: 2CI2), which folds and unfolds in a simple two-state manner [45]. A coarse-grained Gō-like model was adopted to describe the conformational energetics of protein in the folding and unfolding processes [46,47], where the protein conformation is represented by the C α coordinates of the amino acid residues (see Supporting Information). Molecular dynamic (MD) simulations were conducted to obtain the equilibrium conformation distribution and folding/unfolding rates.

The obtained equilibrium free-energy profiles of CI2 using the umbrella sampling technique are plotted in Figure 2 as a function of the number of native contacts (Q). The free-energy profiles exhibit a typical double-well form and are quite smooth, with one basin minimum at $Q \approx 20$ for unfolded states and another at $Q \approx 110$ for folded states. The folded states become more stable with decreasing the temperature. The midpoint temperature at which folded and unfolded states exhibit equal stability was determined to be $T = 0.855$, where the free-energy barrier is about $6.4 k_B T$

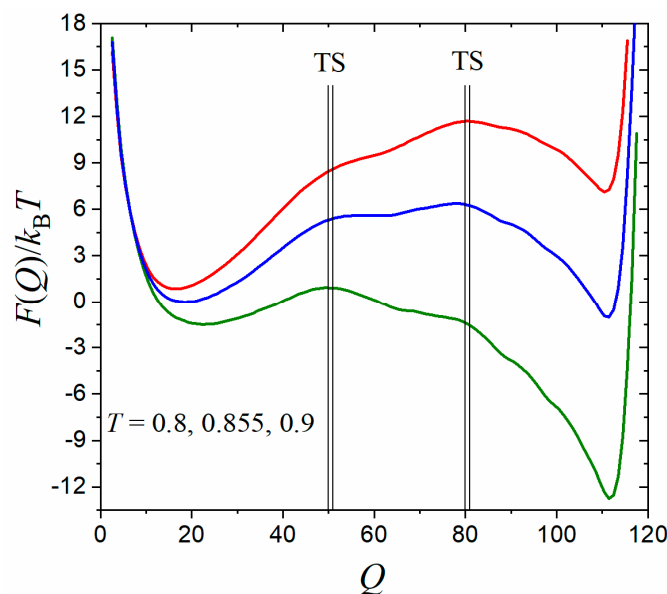


Figure 2. The free-energy profiles of the protein CI2 as a function of the number of native contacts (Q) at different reduced temperatures (from top to bottom): $T = 0.9, 0.855, 0.8 \epsilon/k_B$ (where ϵ is the native contact energy strength). Two choices of transition-state (TS) regions at around $Q = 50$ and $Q = 80$ are indicated by thin black lines.

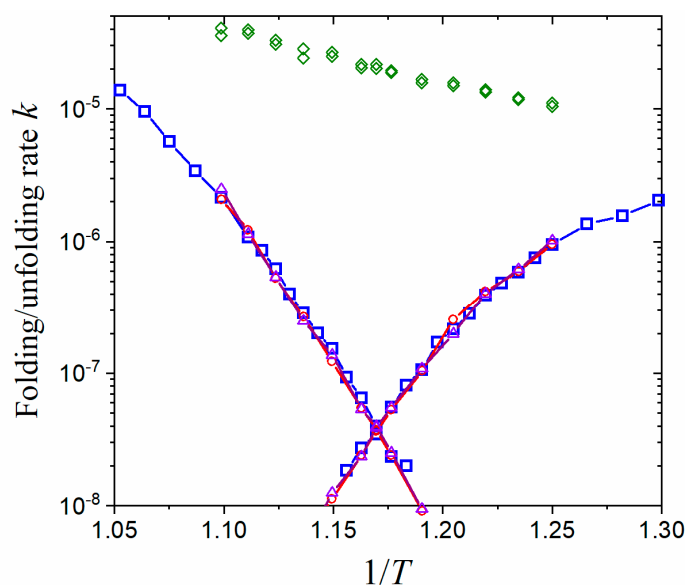
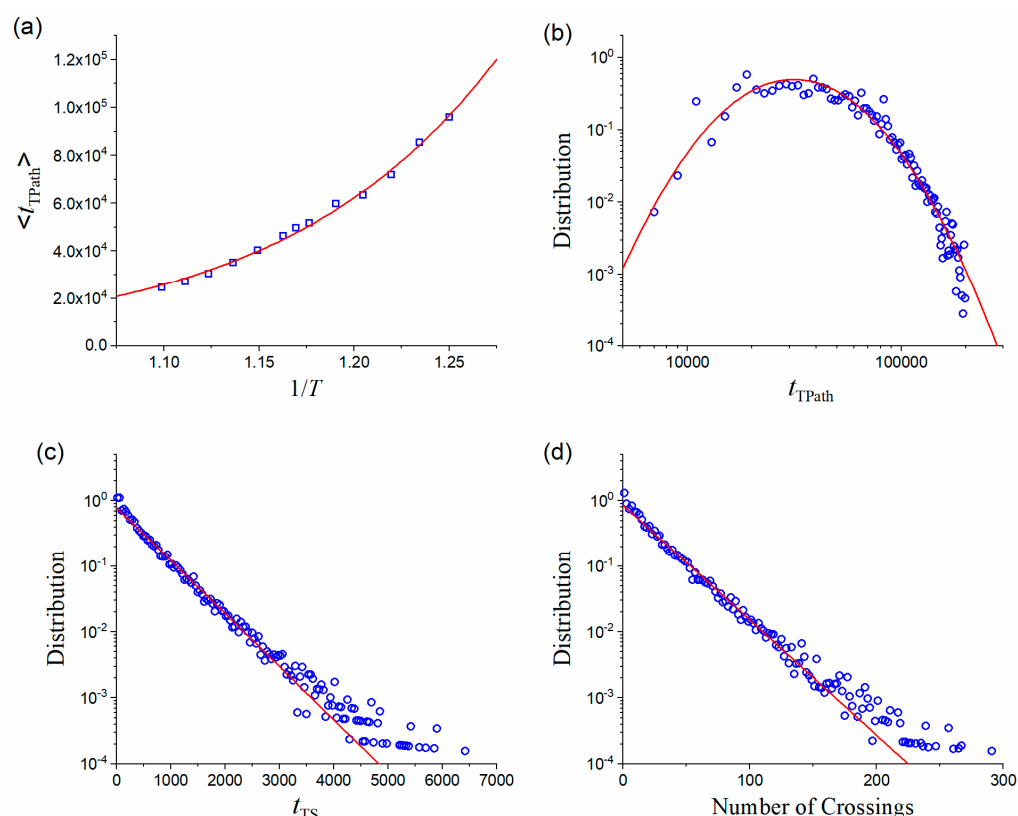


Figure 3. Accelerating folding/unfolding kinetics of CI2 with the tRPS method. Folding (right branches) and unfolding (left branches) rates were measured in a unit of $(\Delta t)^{-1}$, where Δt is the MD time step. The temperature T was measured in a unit of ϵ/k_B . The folding/unfolding rates of direct simulations were plotted in blue squares, while those obtained using tRPS (Equation (7)) were shown in red circles (with $Q_{TS} \in [80, 81]$) and violet triangles (with $Q_{TS} \in [50, 51]$). The sampling rates of transition paths were plotted in scattered green diamonds to demonstrate the acceleration effect. Each datapoint of direct simulations was averaged from about 400 folding/unfolding runs, while those of tRPS were each averaged from about 4000 paths.

Usual models of kinetics suggest that the system has to oscillate around the bottom of a basin many times before it finally crosses the barrier to successfully transit into another basin. This is in line with the protein-folding example here: the average duration of A–A and B–B paths remains roughly unchanged with increasing temperature (Figure S2), similar

to the characteristics of a simple pendulum. To have one successful A–B or B–A transition, it has to oscillate $10^3 \sim 10^6$ times. The acceleration of tRPS originates from the fact that it does not need to spend a lot of time on the massive oscillations but can directly sample the transition paths.

With tRPS, transition paths can readily be obtained for analyses (Figure 4). The average duration of transition paths ($\langle t_{\text{TPath}} \rangle$) is in an order of magnitude of 10^4 steps and increases exponentially with decreasing temperature (Figure 4a), much larger than that for A–A and B–B oscillation paths (in the order of magnitude of 10^2 steps, see Figure S2). The logarithmic t_{TPath} roughly obeys a Gaussian distribution (Figure 4b), similar to previous studies [22,33]. Remarkably, a transition path usually has multiple opportunities to cross the TS region, the number of which well obeys an exponential distribution, i.e., a memoryless distribution (Figure 4d). Consequently, the duration of a path spent within the TS region, t_{TS} , also obeys the exponential distribution (Figure 4c). At the midpoint temperature $T = 0.855$, the average number of crossing times is about 25, and there is an average t_{TS} of about 540 steps. The large number of crossing times indicates that the conventional transition-state theory would inevitably overestimate the transition rates since it assumes that a transition path crosses the TS region only once. Another discrepancy with the transition-state theory is that the activation enthalpy of folding/unfolding (the minus slope of the unfolding curve in Figure 3 is about 62ε at midpoint temperature) is not equal to the enthalpy difference between the TS region and the unfolded/folded state (contributed by the N_{TS}/N_A term in Equation (7), which is about 72ε for unfolding) due to the contribution of the last term in Equation (7).



The tRPS method can be combined with direct simulations to provide much more comprehensive results. For example, the minimal/maximal Q value of a path can be adopted to measure how far it can go, and the resulting path population decreases exponentially with the distance between $Q_{\min/\max}$ and the cutting point (Figure 5). This makes the paths with distant $Q_{\min/\max}$ hard to sample in direct simulations. The tRPS method, on the other hand, samples only the paths that cross the TS region and thus possesses the capability to probe distant $Q_{\min/\max}$. The patches they provided can be combined to give a smooth and complete distribution (Figure 5).

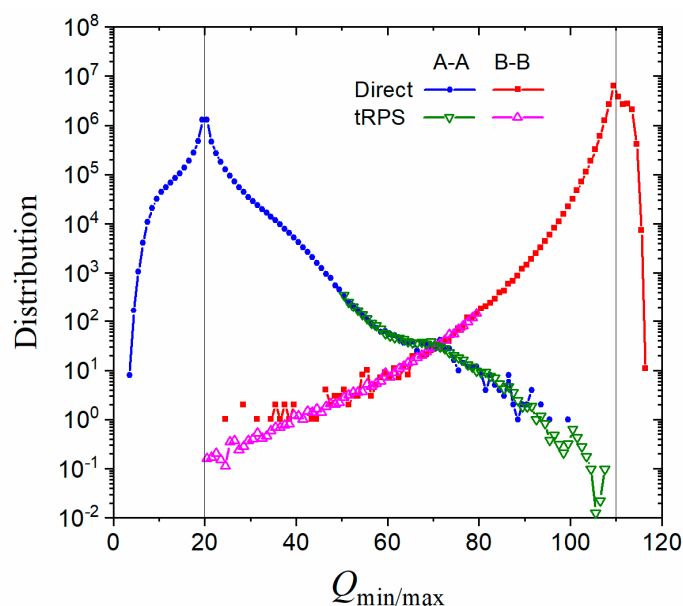


Figure 5. Distribution of minimal/maximal Q for A-A and B-B paths at $T = 0.855$. $Q_A = 20$ and $Q_B = 110$ were used in cutting paths. Filled symbols represent datapoints from direct simulations, while open symbols for those from tRPS with $Q_{TS} \in [50, 51]$ (down triangles) and $Q_{TS} \in [80, 81]$ (up triangles).

Transition rates decay exponentially with the barrier height, but the duration of transition paths usually depends on the barrier in a much weaker logarithmic law [22,48]. This makes tRPS even more powerful when the barrier is high. As proof, we apply tRPS on another protein, acylphosphatase (PDB ID: 1APS) with 98 residues, which was listed as a slow-folding protein in a previous study [49]. The obtained free-energy profile is smooth, possessing a high barrier of about $15 k_B T$ at a midpoint temperature of $T = 0.913$ (Figure 6a). The folding/unfolding is slow and extremely difficult to determine with direct simulations. Therefore, we conducted direct simulations only in some feasible temperature ranges (filled circles/squares in Figure 6b) and applied tRPS to complete the gaps (open circles/squares). The results of direct simulation and tRPS in the overlapping area are well consistent with each other. The data combine to give a nice chevron plot. As a main expense of tRPS, the average duration (open diamonds

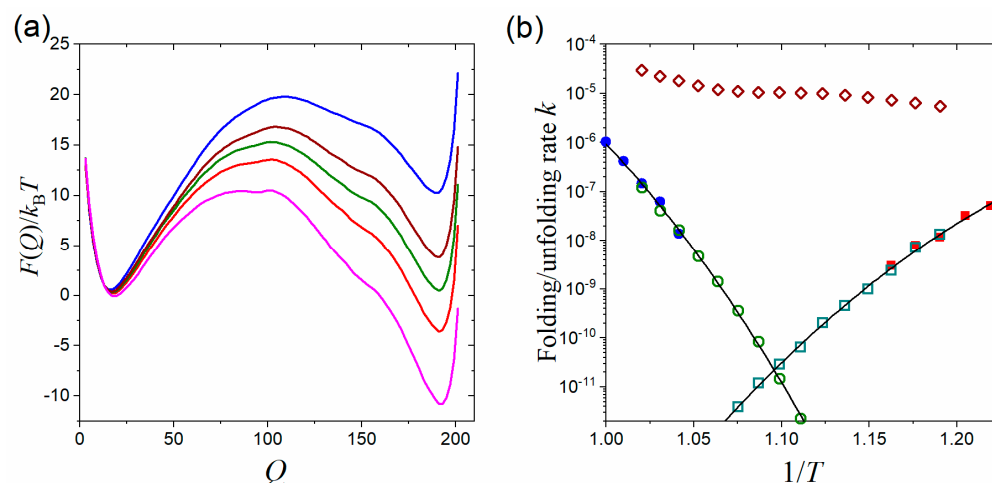


Figure 6. Equilibrium thermodynamics and kinetics of protein folding/unfolding for acylphosphatase. (a) The free-energy profiles at different reduced temperatures (from top to bottom): $T = 0.95, 0.925, 0.913, 0.9, 0.875 \epsilon/k_B$. (b) Accelerating kinetics with tRPS. The temperature T was measured in a unit of ϵ/k_B . The folding/unfolding rates of direct simulations were plotted in filled squares/circles, while those obtained using tRPS were shown in open squares/circles (with $Q_{TS} \in [100, 102]$). The sampling rates of transition paths were plotted in scattered diamonds to demonstrate the acceleration effect.

4. Discussion

The validity of tRPS relies on the time reversibility. Although it cannot be applied to irreversible systems as methods like FFS [29], it possesses the benefits of “shooting from the top” [33] to avoid the possible inadequate choice of initial states in the basin that are capable/incapable of crossing the barrier. Although the TS region was not particularly optimized in our examples (Figure S3), the resulting rates from tRPS are satisfactory (Figures 3 and 6). Another underlying assumption of tRPS

consistent with direct simulations. In areas where direct simulations are inaccessible, the accelerating method provided reasonable results at little cost, with an increase in efficiency up to five orders of magnitude. The technique is easily applied to other kinds of calculations, such as quantum dynamics and chemical reactions.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/molecules28248147/s1>: the native-centric Gō-like model; molecular dynamics simulations; a theoretical comparison with TIS; Figure S1: the temperature dependence of free-energy difference; Figure S2: some path properties determined from direct simulations; Figure S3: the path number sampled in tRPS; and Figure S4: the duration of shooting paths from the TS region for acylphosphatase. References [50–52] are cited in Supplementary Materials.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 22273003. Part of the simulations was performed on the High-Performance Computing Platform of the Center for Life Science (Peking University).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article and supplementary materials.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Yang, Y.I.; Shao, Q.; Zhang, J.; Yang, L.J.; Gao, Y.Q. Enhanced sampling in molecular dynamics. *J. Chem. Phys.* **2019**, *151*, 070902. [CrossRef] [PubMed]
2. Torrie, G.M.; Valleau, J.P. Non-physical sampling distributions in Monte-Carlo free-energy estimation—Umbrella sampling. *J. Comp. Phys.* **1977**, *23*, 187–199. [CrossRef]
3. Kumar, S.; Rosenberg, J.M.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021. [CrossRef]
4. Ferrenberg, A.M

21. Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **1935**, *3*, 107–115. [[CrossRef](#)]
22. Zhang, Z.; Chan, H.S. Transition paths, diffusive processes, and preequilibria of protein folding. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 20919–20924. [[CrossRef](#)] [[PubMed](#)]
23. Bolhuis, P.G.; Swenson, D.W.H. Transition path sampling as Markov chain Monte Carlo of trajectories: Recent algorithms, software, applications, and future outlook. *Adv. Theory Simul.* **2021**, *4*, 2000237. [[CrossRef](#)]
24. Dellago, C.; Bolhuis, P.G.; Csajka, F.S.; Chandler, D. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* **1998**, *108*, 1964–1977. [[CrossRef](#)]
25. Bolhuis, P.G.; Chandler, D.; Dellago, C.; Geissler, P.L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318. [[CrossRef](#)] [[PubMed](#)]
26. Dellago, C.; Bolhuis, P.G.; Geissler, P.L. Transition path sampling. In *Advances in Chemical Physics*; Prigogine, I., Rice, S.A., Eds.; LIDSEN Publishing Inc.: Auburn, OH, USA, 2002; Volume 123, pp. 1–78.
27. van Erp, T.S.; Moroni, D.; Bolhuis, P.G. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* **2003**, *118*, 7762–7774. [[CrossRef](#)]
28. Cabriolu, R.; Refsnes, K.M.S.; Bolhuis, P.G.; van Erp, T.S. Foundations and latest advances in replica exchange transition interface sampling. *J. Chem. Phys.* **2017**, *147*, 152722. [[CrossRef](#)] [[PubMed](#)]
29. Allen, R.J.; Warren, P.B.; ten Wolde, P.R. Sampling rare switching events in biochemical networks. *Phys. Rev. Lett.* **2005**, *94*, 018104. [[CrossRef](#)]
30. Allen, R.J.; Frenkel, D.; ten Wolde, P.R. Simulating rare events in equilibrium or nonequilibrium stochastic systems. *J*

51. Guo, Z.Y.; Thirumalai, D. Kinetics of protein-folding—Nucleation mechanism, time scales, and pathways. *Biopolymers* **1995**, *36*, 83–102. [[CrossRef](#)]
52. van Erp, T.S. Reaction rate calculation by parallel path swapping. *Phys. Rev. Lett.* **2007**, *98*, 268301. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.