

## Extraction of kinetics from equilibrium distributions of states using the Metropolis Monte Carlo method

Sergei F. Chekmarev <sup>\*</sup>

*Institute of Thermophysics, SB RAS, 630090 Novosibirsk, Russia  
and Department of Physics, Novosibirsk State University, 630090 Novosibirsk, Russia*



(Received 3 December 2021; accepted 25 February 2022; published 14 March 2022)

The Metropolis Monte Carlo (MMC) method is used to extract reaction kinetics from a given equilibrium distribution of states of a complex system. The approach is illustrated by the folding/unfolding reaction for two proteins: a model  $\beta$ -hairpin and  $\alpha$ -helical protein  $\alpha_3$ D. For the  $\beta$ -hairpin, the free energy surfaces (FESs) and free energy profiles (FEPs) are employed as the equilibrium distributions of states, playing a role of the potentials of mean force to determine the acceptance probabilities of new states in the MMC simulations. Based on the FESs and PESs for a set of temperatures that were simulated with the molecular dynamics (MD) method, the MMC simulations are performed to extract folding/unfolding rates. It has been found that the rate constants and first-passage time (FPT) distributions obtained in the MMC simulations change with temperature in good agreement with those from the MD simulations. For  $\alpha_3$ D protein, whose equilibrium folding/unfolding was studied with the single-molecule FRET method [Chung *et al.*, *J. Phys. Chem. A* **115**, 3642 (2011)], the FRET-efficiency histograms at different denaturant concentrations were used as the equilibrium distributions of protein states. It has been found that the rate constants for folding and unfolding obtained in the MMC simulations change with denaturant concentration in reasonable agreement with the constants that were extracted from the photon trajectories on the basis of theoretical models. The simulated FPT distributions are single-exponential, which is consistent with the assumption of two-state kinetics that was made in the theoretical models. The promising feature of the present approach is that it is based solely on the equilibrium distributions of states, without introducing any additional parameters to perform simulations, which suggests its applicability to other complex systems.

DOI: [10.1103/PhysRevE.105.034407](https://doi.org/10.1103/PhysRevE.105.034407)

### I. INTRODUCTION

The equilibrium distribution of states and the kinetics of their interconversion are two inherent characteristics of any complex system that are important to know for understanding its behavior. At the same time, the former are generally much easier to determine than the latter, both in experiment and in simulations. One representative example is the protein folding/unfolding reaction. Single-molecule experiments, such as the Förster resonance energy transfer (FRET), atomic-force microscopy (AFM), and optical tweezers (OT) techniques, make it possible to determine the equilibrium distributions of states, in the form of either the FRET-efficiency histograms (FRET [1–7], folding/unfolding) or free energy profiles (AFM [8,9] and OTT [10], unfolding). However, in order to extract kinetics from these distributions, theoretical models are needed (FRET [11–15] and AFM and OTT [16–19]). A similar difference in the possibility of obtaining the equilibrium distribution of states and determining the kinetics exists for the molecular dynamics (MD) simulations of folding of relatively large proteins (of the order of 100 residues and larger). While brute force all-atom simulations, which are able to give a detailed insight into the folding kinetics, are very time

consuming [20,21], several methods exist that allow very efficient calculation of the equilibrium distributions of protein states [22], such as the replica exchange molecular dynamics [23–25] and the weighted histogram analysis method [26] in combination with sampling from biased MD simulations [27]. Therefore, it is tempting to find out whether the knowledge of the equilibrium distribution of states of a system allows us to obtain information about its kinetics, and if so, to what extent.

This paper examines the possibility of extracting kinetics from the equilibrium distributions of protein states with the use of the Metropolis Monte Carlo (MMC) algorithm [28]. Originally designed for calculating the equation of states, the MMC method has become a powerful tool for the study of equilibrium properties of various systems [29,30]. However, the capabilities of this method are not limited to the study of the equilibrium properties; it can be used to investigate the dynamic properties as well. Using the diffusion of a Brownian particle as an example, it was shown that on the coarse level, the evolution of states in the MMC simulations can be described by the Fokker-Planck equation [31]. The method has long been successfully employed to study the protein folding reaction in the framework of lattice models [32–36]. In this *dynamic* implementation, the MMC method differs from the *kinetic* MC method, in which the probability of transition between two states is determined by the rate constant (or,

\*chekmarev@itp.nsc.ru

more generally, by the height of the barrier) rather than by the difference in populations of these states [37,38].

In the present paper, the acceptance probability of a new state in the course of the MMC run is determined by the equilibrium distribution of states known from simulations or experiment. The typical examples of such distributions are the free energy surfaces (FESs) and free energy profiles (FEPs), although, in principle, any equilibrium distribution of states along one or several collective variables can be used for this purpose. These distributions represent potentials of mean force (PMF) [27,39,40], and thus they can be assumed to govern kinetics.

A known shortcoming of the MMC method is that a generated MMC trajectory represents a Markovian sequence of states rather than the change of these states in time [29,30]. Therefore, only the relative values of the reaction rates for different reaction conditions, e.g., for temperature, are available. In order to have the absolute value of rates, a fit to the timescale of the original data is required [41–44]. At the same time, in contrast to other possible approaches which could employ the PMFs to calculate kinetics, e.g., the use of the Langevin equation [45], no additional parameters are needed to perform the MMC simulations for a given equilibrium distribution of states. It is a considerable advantage of the present approach.

The approach is illustrated for two systems. One is a model  $\beta$ -hairpin protein. In order to have baseline results, the equilibrium MD simulations were performed. The obtained FESs and FEPs were employed as the input distributions of states for subsequent MMC simulations. It has been shown that the rate constants of folding and unfolding calculated from the FESs and FEPs change with temperature in good agreement with those from the MD simulations. The other system is an  $\alpha$ -helical protein  $\alpha_3$ D, for which folding and unfolding rate constants for different denaturant concentrations were extracted from single-molecule FRET trajectories [3] using theoretical models [12,13]. In this case, the FRET-efficiency histograms (FEHs) obtained in the experiment were employed as the equilibrium distributions of states for the MMC simulations. It has been found that the folding and unfolding rate constants obtained in the MMC simulations vary with the denaturant concentration in reasonable agreement with those extracted from the FRET trajectories. For both proteins, the first-passage time (FPT) distributions in the MMC simulations have been found to be single exponential, which agree with those obtained in the MD simulations ( $\beta$ -hairpin protein) and correspond to the assumption of two-state kinetics employed in the theoretical models to extract rate constants ( $\alpha_3$ D protein).

The paper is organized as follows. Section II describes the MD (Sec. II A) and MMC (Sec. II B) simulations for the  $\beta$ -hairpin protein, Sec. III presents the results of the MMC simulations for the  $\alpha_3$ D protein, and Sec. IV briefly summarizes the results of the work.

## II. $\beta$ -HAIRPIN PROTEIN: SYSTEM AND SIMULATION

### A. System and molecular dynamics simulations

To perform MD simulations, a coarse-grained ( $C_\alpha$ -bead) protein model and a G $\ddot{o}$ -type interaction potential [46] were

used. Briefly, the approach is as follows (for details, see Ref. [47]). The  $C_\alpha$  model was constructed on the basis of the solution NMR data for 12-residue HP7 protein (2evq.pdb) [48]. Specifically, the first conformation in the NMR ensemble of the protein structures was employed for this purpose, which was then considered as a reference native structure of the protein. The G $\ddot{o}$ -type potential accounted for the rigidity of the backbone and the contributions of native and nonnative contacts [49]. Two  $C_\alpha$  beads were considered to be in native contact if they were not the nearest neighbors along the protein chain and had the interbead distance not longer than  $d_{\text{cut}} = 7.5 \text{ \AA}$ . In this case, the number of native contacts is  $N_{\text{nat}} = N_{\text{nat}}^{\text{NAT}} = 27$ . The simulations were performed with a constant-temperature MD based on the coupled set of Langevin equations [50]. The time step was  $\Delta t = 0.0125\tau$ , where  $\tau$  is the characteristic time. At the length scale  $l = 7.5 \text{ \AA}$  and the attractive energy  $\epsilon = 2.2 \text{ kcal/mol}$  [51],  $\tau = (Ml^2/\epsilon)^{1/2} \approx 2.6 \text{ ps}$ , where  $M = 110 \text{ Da}$  is the average mass of the residue. The friction constant  $\gamma$  in the Langevin equations was  $\gamma = 10M/\tau$ . In what follows, the length, in particular, the radius of gyration, is measured in angstroms. Unless otherwise noted, the other quantities (energy, temperature, and time) are dimensionless; specifically, the energy and temperature are measured in units of  $\epsilon$  (in the latter case with the Boltzmann constant set to unity), and the time is measured in units of  $\tau$ .

The equilibrium simulations were carried out for a range of temperatures at which both folding and unfolding events occurred quite frequently, specifically, at  $T = 0.35, 0.375, 0.4, 0.425, \text{ and } 0.45$ . The regime of folding varied from downhill (low temperatures) to uphill (larger temperatures) folding through a marginal two-state folding (intermediate temperatures). For each temperature, the MD trajectory was run until  $10^3$  folding/unfolding events took place. The protein was considered to be unfolded if  $N_{\text{nat}} \leq 5$  and to be folded if the root-mean-square deviation from the reference native structure was less than  $1 \text{ \AA}$  [52]. Based on the simulated trajectories, the FESs and FEPs were calculated. The FES was constructed using the number of native contacts ( $N_{\text{nat}}$ ) and the radius of gyration ( $R_g$ ) as collective variables [27,53]

$$F(N_{\text{nat}}, R_g) = -T \ln P(N_{\text{nat}}, R_g), \quad (1)$$

where  $P(N_{\text{nat}}, R_g)$  is the probability to find the protein in the point  $(N_{\text{nat}}, R_g)$ . The FEP was calculated as

$$F(N_{\text{nat}}) = -T \ln P(N_{\text{nat}}), \quad (2)$$

where  $P(N_{\text{nat}})$  is the probability that the protein has  $N_{\text{nat}}$  native contacts. The resulting FESs and FEPs for the lower, intermediate, and upper temperatures are shown in Figs. 1 and 2, respectively. In structure, they are common for folding of  $\beta$ -hairpins (experiment [54] and simulation studies in explicit [55,56] and implicit [57,58] solvents).

Both folding and unfolding reactions were considered. For this, the equilibrium MD trajectory was divided into segments of the trajectory corresponding to the transition from the unfolded to the native state (folding) and to the backward transition (unfolding) [52]. Based on these segments of the trajectory, the MD rate constants of folding and unfolding, as well as the corresponding FPT distributions, were calculated (see below).

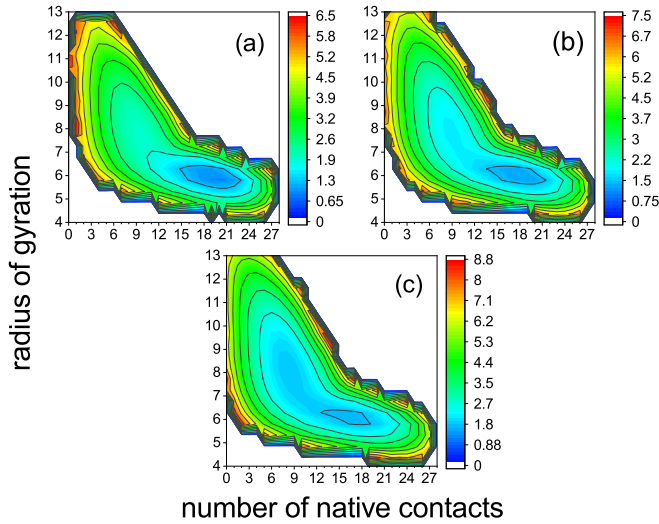


FIG. 1.  $\beta$ -hairpin, free energy surfaces: (a)  $T = 0.35$ , (b)  $T = 0.4$ , and (c)  $T = 0.45$ . The radius of gyration is given in angstroms.

### B. Monte Carlo simulations

Having the FESs and FEPs, the MMC simulations were performed. According to the Metropolis algorithm [28], the acceptance probability of the transition from state A to B is determined as

$$W(A \rightarrow B) = \begin{cases} 1 & \text{if } P(B)/P(A) \geq 1 \\ P(B)/P(A) & \text{if } P(B)/P(A) < 1 \end{cases}, \quad (3)$$

where  $P(A)$  and  $P(B)$  are the equilibrium probabilities of the old (A) and new (B) states, respectively. A trial move from A to B is accepted if  $P(B)/P(A) \geq \alpha$ , where  $\alpha$  is a random number uniformly distributed between 0 and 1. In application to the free energy landscapes of Figs. 1 and 2,  $P(B)/P(A) = \exp\{-[F(B) - F(A)]/T\}$ . For FESs (Fig. 1), A and B represent the points in the  $(N_{\text{nat}}, R_g)$  space, and for FEPs (Fig. 2), the points in the  $N_{\text{nat}}$  space. The corresponding

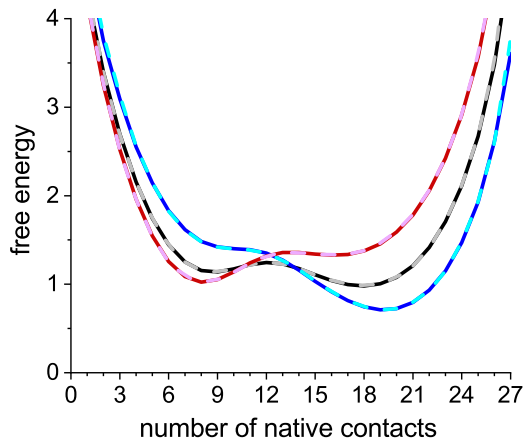


FIG. 2.  $\beta$ -hairpin, free energy profiles:  $T = 0.35$  (blue and cyan),  $T = 0.4$  (black and gray), and  $T = 0.45$  (red and magenta). The first colors in the brackets are for the MD simulations, and the second colors are for the MMC simulations based on the FESs of Fig. 1.

free energies  $F(A)$  and  $F(B)$  are determined by Eqs. (1) and (2).

In the case of FESs, the MMC simulation space represented a rectangle with  $0 \leq N_{\text{nat}} \leq 28$  and  $0 \leq R_g \leq 15 \text{ \AA}$ . It was divided into  $28 \times 30$  segments ( $dN_{\text{nat}} = 1$  and  $dR_g = 0.5$ ), forming a mesh of  $29 \times 31$  discrete states. The trial moves consisted of random shifts of the point along the  $N_{\text{nat}}$  and  $R_g$  axes by  $-1$  or  $+1$  mesh step so that the eight nearest points surrounding the current point were sampled with equal probability. To verify that detailed balance is satisfied, a long (“equilibrium”) MMC trajectory with many folding/unfolding events was run. It was found that the balance was approximately maintained for each point-to-point transition and was exactly fulfilled for the transitions within the simulation cell, i.e., the number of transitions from the central point of the cell to the eight surrounding points was equal to the number of the backward transitions. Accordingly, the equilibrium MMC simulations successfully reproduced the original equilibrium distributions on which they were based (one example is shown in Fig. 2). To simulate the process of folding, the trajectories were initiated at point (5, 11.8), representing the unfolded state, and terminated at point (27, 5.78), representing the native state (Fig. 1). For the unfolding trajectories, these points served as the terminal and initial states, respectively.

In the case of FEPs, the MMC simulations were conducted in the linear  $N_{\text{nat}}$  space with the boundaries  $N_{\text{nat}} = 0$  and  $N_{\text{nat}} = 27$ . The trial moves represented random changes of the current value of  $N_{\text{nat}}$  by  $-1$  or  $+1$ , which were taken with equal probability. In equilibrium, detailed balance was strictly maintained. The folding trajectories started at  $N_{\text{nat}} = 5$  and terminated at  $N_{\text{nat}} = 27$ , and the unfolding trajectories used these points as terminal and initial state, respectively.

In each case (the FES- or FEP-based simulations, the protein folding or unfolding, and a given temperature),  $10^3$  trajectories were run. For the FES-based simulations, the acceptance probability increased from  $\approx 0.57$  ( $T = 0.35$ ) to  $\approx 0.67$  ( $T = 0.45$ ) for folding trajectories, and from  $\approx 0.57$  ( $T = 0.35$ ) to  $\approx 0.62$  ( $T = 0.45$ ) for unfolding trajectories. For the FEP-based simulations, it increased from  $\approx 0.87$  ( $T = 0.35$ ) to  $\approx 0.9$  ( $T = 0.45$ ) for both folding and unfolding trajectories.

In order to fit the MMC steps to the MD time steps, the average ratios between the MMC and MD rate constants over the whole temperature range were calculated. In particular, for the FES-based folding trajectories, the time step was determined as  $(\Delta t)_{\text{FES}}^{\text{F}} = (1/N) \sum_{i=1}^N (k_{\text{F}})_{\text{MMC},i} / (k_{\text{F}})_{\text{MD},i}$ , where  $N = 5$  is the number of temperatures, and  $(k_{\text{F}})_{\text{MMC}}$  and  $(k_{\text{F}})_{\text{MD}}$  are the rate constants for the MMC and MD folding trajectories, respectively. The rate constants were calculated as the inverses of the corresponding mean first-passage times in units of number of steps (the MMC simulations) and  $\tau$  (the MD simulations). Similarly, the time steps for the FEP based and unfolding MMC simulations were determined. It was found that  $(\Delta t)_{\text{FES}}^{\text{F}} \approx 3.4 \times 10^{-3} \tau$ ,  $(\Delta t)_{\text{FES}}^{\text{U}} \approx 6.3 \times 10^{-3} \tau$ , and  $(\Delta t)_{\text{FEP}}^{\text{U}} \approx (\Delta t)_{\text{FEP}}^{\text{F}} \approx 0.18 \tau$ . These values of the time steps show that the MMC timescales generally depend on both the direction of the process (folding or unfolding) and the dimensionality of the state distributions for the MMC simulations (FES or FEP).

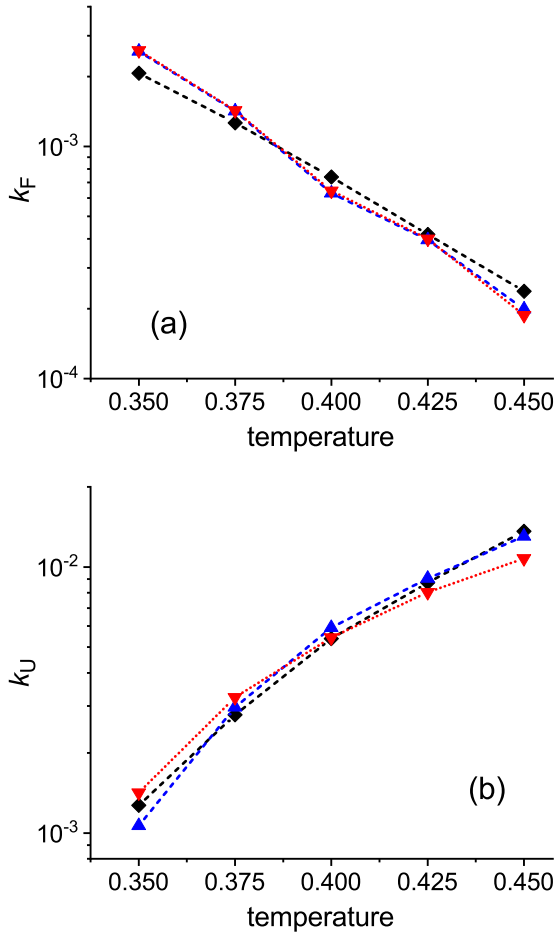


FIG. 3.  $\beta$ -hairpin, rate constants of (a) folding and (b) unfolding: the MD simulations (black diamonds and dashed line), the FES-based MMC simulations (red triangles and dotted line), and the FEP-based MMC simulations (blue triangles and dashed line). The lines are to guide the eye.

Figures 3 and 4 compare the results of the MMC and MD simulations. Figure 3 presents the rate constants obtained in the MMC simulations in comparison to those in the MD simulations. The MMC “time” scales were adjusted to the MD timescales using the above values of the MMC steps in  $\tau$  units, e.g., for the MMC folding simulations based on the FES, the adjusted rate constant at  $T = 0.35$  was obtained to be  $(k_F)_{\text{MMC}} = (\tilde{k}_F)_{\text{MMC}} / (\Delta t)_{\text{FES}}^F \approx 2.6 \times 10^{-3}$ , where  $(\tilde{k}_F)_{\text{MMC}} \approx 8.75 \times 10^{-6}$  is the original rate constant in the inverse number of the MMC steps, and  $(\Delta t)_{\text{FES}}^F \approx 3.4 \times 10^{-3} \tau$ . Figure 3 shows that the rate constants from the MMC simulations are in good agreement with those from the MD simulations through the entire temperature range. The same result would evidently be obtained if the MMC step in  $\tau$  units was determined by equating the MMC rate constant at one temperature with the corresponding MD rate constant. The above values of the time steps were also used to adjust the timescales for the MMC FPT distributions. Figure 4 presents these distributions in comparison with the distributions obtained in the MD simulations and shows that, similar to the rate constants, they are also in agreement, including a single-exponential decay. Importantly, the good agreement of

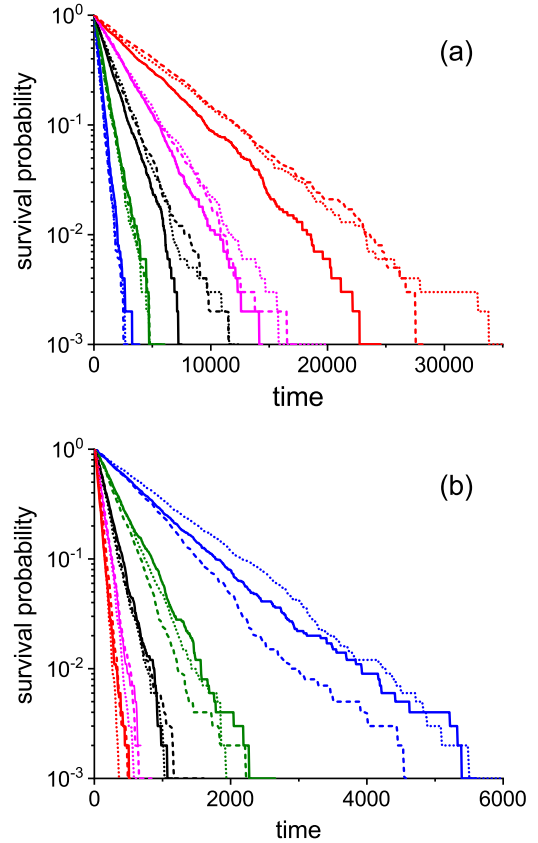


FIG. 4.  $\beta$ -hairpin, first-passage time distributions: (a) folding and (b) unfolding. Solid lines are for the MD simulations, dashes lines are for the MMC simulations based on the free energy surfaces of Fig. 1, and dotted lines are for the MMC simulations based on the free energy profiles of Fig. 2. Temperatures and lines:  $T = 0.35$  (blue),  $T = 0.375$  (olive),  $T = 0.4$  (black),  $T = 0.425$  (magenta), and  $T = 0.45$  (red).

the MMC rate constants with the MD constants through all temperature range indicates that the time step in the MMC simulations is not affected by temperature.

### III. $\alpha_3$ D PROTEIN: EXPERIMENT AND MONTE CARLO SIMULATIONS

The equilibrium folding/unfolding of this 73-residue three-helix bundle protein (2a3d.pdb) at room temperature and different denaturant (GdmCl) concentrations (1.5 M to 3 M) was studied using single-molecule FRET [3]. Two models assuming two-state kinetics were employed to analyze the photon trajectories. One is a maximum likelihood estimation (MLE), in which the photon trajectory were fitted to the sequences of transitions between unfolded and folded states [12], and the other is an approximation of the FRET-efficiency histogram (FEH) by a sum of Gaussians whose weights depended on the probabilities of surviving the folded and unfolded states for the bin time ( $3G$ ) [13]. Using these approaches, the rate constants of folding and unfolding were determined, which were found in good agreement, except for the denaturant concentrations of 1.5 M and, partially, of 3.0 M (cf. the corresponding lines of Table 1 of Ref. [3]). Figure 5

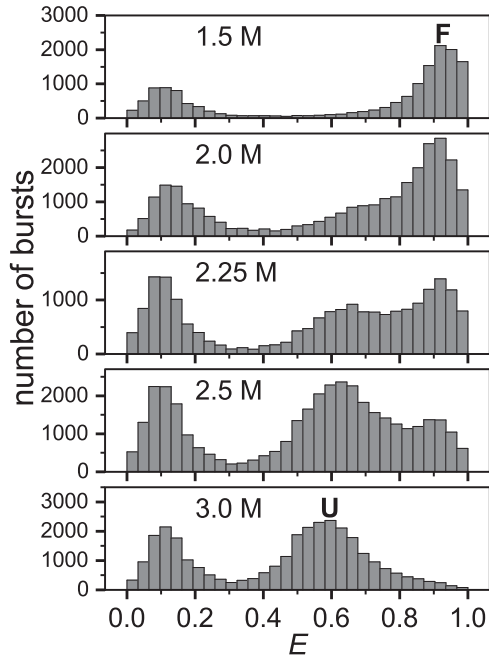


FIG. 5.  $\alpha_3\text{D}$ : FRET-efficiency histograms [3]. Figures inside the panels represent the GdmCl concentrations. Labels U and F indicate the unfolded and folded state, respectively.

shows the FEHs for freely diffusing protein molecules. The time bin to count donor and acceptor photons is 2 ms. Two peaks are significant: at intermediate and higher values of FRET efficiency ( $E$ ), which represent the unfolded (U) and folded (F) states, as indicated in Fig. 5. The peaks at  $E \sim 0$  originate from molecules with inactive acceptors and are not included in consideration. It was found that the FEHs are well approximated by a sum of three Gaussians (3G model). Two Gaussians describe the unfolded and folded states, and the third one accounts for the intermediate values of  $E$  that appear due to the transitions between the folded and unfolded states.

It should be noted that the present FEHs may not represent the true PMFs, as it would be if they obeyed the Förster “distance-efficiency” equation  $E = 1/[1 + (r/R_0)^6]$  ( $r$  is the donor-to-acceptor separation distance, and  $R_0$  is the Förster radius). First, the number of photons in time bin ( $\sim 10^2$ ) is not sufficiently large to avoid the effect of shot noise on the distribution of protein states. In this case, the broadening of the folded and unfolded peaks may mainly be a result of short noise rather than of the presence of folded- and unfolded-like conformations of the protein. Second, the binning time (2 ms) is comparable to the relaxation time ( $\sim 1$  ms), so that the signal from different protein conformations, which typically interconvert on a much shorter timescale ( $\sim 1 \mu\text{s}$  and shorter), may be placed in the same bin. In particular, the intermediate values of  $E$  between the unfolded and folded peaks may represent a mixture of photons that come from the unfolded and folded states rather than the transitional conformations. However, if detailed balance is fulfilled, it seems more important that the intermediate values of  $E$  adequately represent the probabilities of the intermediate protein conformations rather than the conformations themselves.

The MMC moves were determined by Eq. (3), in which the values of FRET efficiency represented the protein states, and the numbers of bursts in the FEHs served as (non-normalized) probabilities of these states,  $P(E)$ . According to the FEHs of Fig. 5, the whole range of FRET efficiencies, from 0 to 1, was divided into 31 segments. To take into account that the peak at  $E \sim 0$  corresponds to the inactive acceptor, the populations of states at  $0 \leq E \leq E_{\text{thr}}$ , where  $E_{\text{thr}}$  is the value of  $E$  at which  $P(E)$  has a minimum between the inactive and unfolded peaks, were set to zero,  $P(E) = 0$ . Each MMC step was taken to be an equally probable random move to a neighboring segment. Detailed balance was strictly satisfied similar to the FEP-based MMC simulations for the  $\beta$ -hairpin (Sec. II B). The unfolded and folded states were chosen to correspond to the maxima of the unfolded and folded peaks (Fig. 5), with the values of FRET efficiency  $E_U = 0.61$  and  $E_F = 0.93$ , respectively. To simulate folding, the MMC trajectories were initiated at  $E = E_U$  and terminated at  $E = E_F$ , and to simulate unfolding, the states  $E_F$  and  $E_U$  were used as the initial and terminal states, respectively. In all cases, i.e., at each denaturant concentration for folding and unfolding,  $10^3$  MMC trajectories were run. For both folding and unfolding reactions, the acceptance probability varied insignificantly, from  $\approx 0.87$  to  $\approx 0.92$ . Based on the simulated folding and unfolding trajectories, the rate constants for folding ( $k_F$ ) and unfolding ( $k_U$ ) were determined, and the corresponding FPT distributions were constructed. Similar to  $\beta$ -hairpin, the MMC “time” scale was adjusted to the experimental timescale by calculating the average ratios between simulated and experimental rate constants. Specifically, the MMC time steps for folding and unfolding were determined, respectively, as  $(\Delta t)^F = (1/N) \sum_{i=1}^{i=N} (k_F)_{\text{MMC},i} / (k_F)_{\text{exp},i}$  and  $(\Delta t)^U = (1/N) \sum_{i=1}^{i=N} (k_U)_{\text{MMC},i} / (k_U)_{\text{exp},i}$ , where  $N = 3$  is the number of denaturant concentrations ranging from 2.0 M to 2.5 M (i.e., 2.0 M, 2.25 M, and 2.5 M), at which the relative populations of both folded and unfolded states were larger than 1%. It was obtained  $(\Delta t)^F \approx 7.5 \times 10^{-3}$  ms and  $(\Delta t)^U \approx 8.7 \times 10^{-3}$  ms. Given the MMC time steps, the rate constants for folding and unfolding were recalculated as  $(\tilde{k}_F)_{\text{MMC}} = (k_F)_{\text{MMC}} / (\Delta t)^F$  and  $(\tilde{k}_U)_{\text{MMC}} = (k_U)_{\text{MMC}} / (\Delta t)^U$ , where the rate constants denoted by the tilde are the original MMC constants in the inverse number of the MMC steps.

Figure 6 compares the adjusted MMC rate constants with those extracted from the FRET trajectories with the 3G model (Table 1 of Ref. [3]). It shows that the MMC simulations based on the FRET histograms reasonably predict the change of the rate constants with denaturant concentration, for both folding and unfolding reactions. The only exception are the concentrations at which the normalized population of the target state is as small as  $\approx 1\%$ , as at 3 M GdmCl in the case of folding, where  $P(0.93) \approx 1.2 \times 10^{-2}$ , and at 1.5 M GdmCl in the case of unfolding, where  $P(0.61) \approx 9 \times 10^{-3}$ . More generally, the MMC simulations are expected to give satisfactory results when the peaks for the target states are resolved. Figure 7 also presents the simulated FPT distributions for folding and unfolding. The distributions are single exponential, which is consistent with the assumption of two-state kinetics that was made in the theoretical models for extracting the rate constants from the FRET data [3]. Since the FEHs for the GdmCl

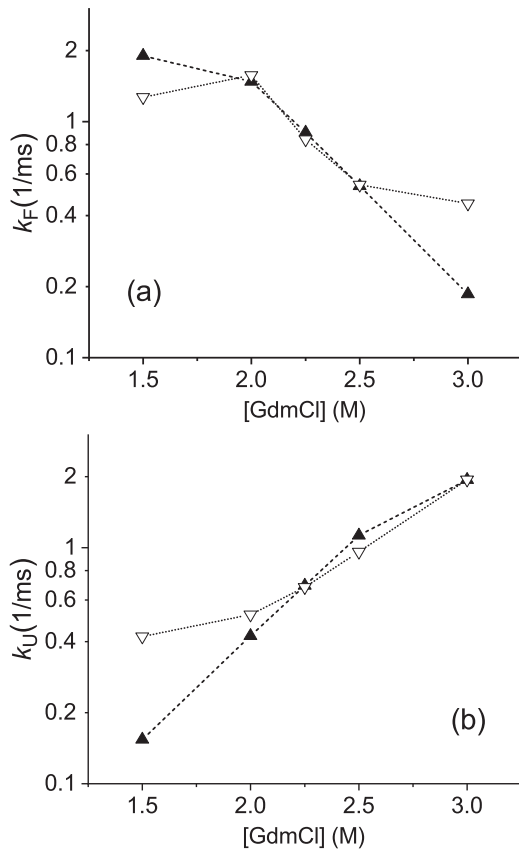


FIG. 6.  $\alpha_3D$ : Comparison of the rate constants for protein (a) folding and (b) unfolding. Open triangles show the experimental rate constants obtained with the 3G model [3], and solid triangles represent simulated rate constants.

concentrations of 1.5 M, 2.0 M, and 3.0 M do not have minima between the unfolded and native states that would represent the transition state barriers, the folding kinetics in these cases should generally be considered as downhill rather than two-state [59–61]. At the same time, it does not rule out that the distributions of folding times can remain single-exponential, as has been discussed, e.g., in Refs. [62,63], in particular, for the  $\alpha_3D$  protein [63].

#### IV. CONCLUSIONS

It has been shown that the Metropolis MC (MMC) simulations based on the equilibrium distributions of states can be successfully used to predict how characteristic reaction times in a complex system change with the determining conditions. As an example, the process of folding and unfolding for two proteins was considered. The first protein, a  $\beta$ -hairpin protein, whose coarse-grained ( $C_\alpha$ -bead) model was constructed on the basis of the solution NMR data for 12-residue HP7 protein [48], was employed to develop and test the approach. Performing MD simulations, the baseline characteristics of the folding/unfolding process were determined for a set of temperatures: the equilibrium free energy surfaces (FESs) and free energy profiles (FEPs), the rate constants of folding and unfolding, and the first-passage time (FPT) distributions for these reactions. Using the FESs and FEPs as the potentials

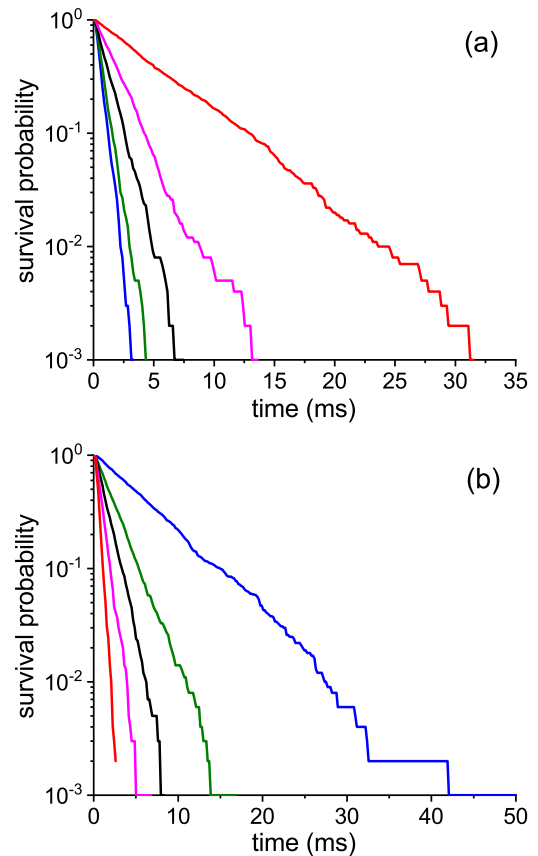


FIG. 7.  $\alpha_3D$ : Simulated first-passage time distributions for protein (a) folding and (b) unfolding.

of mean force (PMFs), the MMC simulations were carried out to obtain relative rate constants of folding and unfolding and the FPT distributions. It has been found that they change with temperature in good agreement with those from the MD simulations, and, after fitting the MMC steps to the MD time steps, they agree well in absolute value. The other protein is a three-helical  $\alpha_3D$  protein, whose folding/unfolding reaction was studied at different denaturant concentrations using single-molecule FRET method [3]. In this case, the measured FRET-efficiency histograms (FEHs) were employed as the equilibrium distributions of states. Because of the uncertainty posed by experimental limitations (a relatively small number of photons in time bins, and a long binning time in comparison to the characteristic times of interconversion of protein molecules), it was not quite clear whether these FEHs could be used as the PMFs. However, it has been found that for the denaturant concentrations at which the FEH peak for the target folded (unfolded) state is resolved, the rate constants of folding (unfolding) change with the denaturant concentration in satisfactory agreement with those extracted from photon trajectories with the use of theoretic models [12,13]. Also, it has been found the FPT distributions are single-exponential, which is consistent with the assumption of two-state kinetics in the theoretical models [12,13] for extracting the rate constants from the FRET data. Although the kinetics observed for  $\beta$ -hairpin and  $\alpha_3D$  proteins do not go beyond two-state kinetics, there seem to be no reasons why adding a new state would

reduce the ability of the present approach. To support this, the approach was tested on a two-dimensional (2D) potential energy surface [64], which was designed to mimic the process of folding of an imaginary protein three states, i.e., unfolded, folded and off-pathway intermediate states [65]. It has been shown that the approach works equally well in this case, i.e., the FPT distribution obtained with the MMC simulations based of the FES generated from the given 2D PES agrees well with the double-exponential distribution obtained in the MD simulations (see the Supplemental Material [66]). To obtain absolute values of the rate constants, a time-dependent quantity is required, which, on the one hand, would be known in the experiment (physical or computational) and, on the other, could be obtained in the MMC simulations based on the

given equilibrium distribution of states, e.g., the self-diffusion coefficient, if this distribution were in physical space [41]. The promising feature of the present approach is that it is based solely on the equilibrium distributions of states, without introducing any additional parameters to perform simulations, which suggests its applicability to other complex systems.

#### ACKNOWLEDGMENTS

I thank Irina Gopich for valuable discussions of the results of the FRET experiments for folding of  $\alpha_3D$  protein, and Hoi Sung Chung for sharing the corresponding FRET-efficiency histograms. The work was supported by a contract with the IT SB RAS.

- 
- [1] B. Schuler, E. A. Lipman, and W. A. Eaton, *Nature (London)* **419**, 743 (2002).
- [2] F. Huang, L. Ying, and A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **106**, 16239 (2009).
- [3] H. S. Chung, I. V. Gopich, K. McHale, T. Cellmer, J. M. Louis, and W. A. Eaton, *J. Phys. Chem. A* **115**, 3642 (2011).
- [4] J. Liu, L. A. Camposa, M. Cerminara, X. Wang, R. Ramanathan, D. S. Englisha, and V. Muñoz, *Proc. Natl. Acad. Sci. USA* **109**, 179 (2012).
- [5] B. Schuler and H. Hofmann, *Curr. Opin. Struct. Biol.* **23**, 36 (2013).
- [6] M. Gabba, S. Poblete, T. Rosenkranz, A. Katranidis, D. Kempe, T. Züchner, R. G. Winkler, G. Gompfer, and J. Fitter, *Biophys. J.* **107**, 1913 (2014).
- [7] W. A. Eaton, *J. Phys. Chem. B* **125**, 3452 (2021).
- [8] H. Yu, A. N. Gupta, X. Liu, K. Neupane, A. M. Brigley, I. Sosova, and M. T. Woodside, *Proc. Natl. Acad. Sci. USA* **109**, 14452 (2012).
- [9] M. Batchelor, K. Papachristos, M. Stofella, Z. T. Yew, and E. Paci, *Biochim. Biophys. Acta* **1864**, 129613 (2020).
- [10] C. Bustamante, L. Alexander, K. Maciuba, and C. M. Kaiser, *Annu. Rev. Biochem.* **89**, 443 (2020).
- [11] S. A. McKinney, C. Joo, and T. Ha, *Biophys. J.* **91**, 1941 (2006).
- [12] I. V. Gopich and A. Szabo, *J. Phys. Chem. B* **113**, 10965 (2009).
- [13] I. V. Gopich and A. Szabo, *J. Phys. Chem. B* **114**, 15221 (2010).
- [14] R. Ramanathan and V. Muñoz, *J. Phys. Chem. B* **119**, 7944 (2015).
- [15] I. S. Osadko, *J. Phys. Chem. C* **120**, 697 (2016).
- [16] G. Hummer and A. Szabo, *Proc. Natl. Acad. Sci. USA* **98**, 3658 (2001).
- [17] G. Hummer and A. Szabo, *Acc. Chem. Res.* **38**, 504 (2005).
- [18] C. Hyeon and D. Thirumalai, *J. Phys.: Condens. Matter* **19**, 113101 (2007).
- [19] O. K. Dudko, G. Hummer, and A. Szabo, *Proc. Natl. Acad. Sci. USA* **105**, 15755 (2008).
- [20] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, *Science* **330**, 341 (2010).
- [21] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, *Science* **334**, 517 (2011).
- [22] C. Camilloni and F. Pietrucci, *Adv. Phys.: X* **3**, 1477531 (2018).
- [23] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
- [24] Y. M. Rhee and V. S. Pande, *Biophys. J.* **84**, 775 (2003).
- [25] F. Rao and A. Caflisch, *J. Chem. Phys.* **119**, 4035 (2003).
- [26] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
- [27] J.-E. Shea and C. L. Brooks, III, *Annu. Rev. Phys. Chem.* **52**, 499 (2001).
- [28] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. N. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [29] D. P. Landau and K. Binder, *Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).
- [30] K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics: An Introduction*, 4th ed. (Springer, Berlin, 2002).
- [31] K. Kikuchi, M. Yoshida, T. Maekawa, and H. Watanabe, *Chem. Phys. Lett.* **185**, 335 (1991).
- [32] N. Gö and H. Taketomi, *Proc. Natl. Acad. Sci. USA* **75**, 559 (1978).
- [33] A. Rey and J. Skolnick, *Chem. Phys.* **158**, 199 (1991).
- [34] A. Šali, E. Shakhnovich, and M. Karplus, *Nature (London)* **369**, 248 (1994).
- [35] S. F. Chekmarev, S. V. Krivov, and M. Karplus, *J. Phys. Chem. B* **109**, 5312 (2005).
- [36] S. F. Chekmarev, A. Yu. Palyanov, and M. Karplus, *Phys. Rev. Lett.* **100**, 018107 (2008).
- [37] D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- [38] A. F. Voter, in *Radiation Effects in Solids*, edited by K. E. Sickafus, E. A. Kotomin, and B. P. Uberuaga (Springer, NATO Sci. Ser., Dordrecht, 2007), p. 1–23.
- [39] B. Roux, *Comput. Phys. Commun.* **91**, 275 (1995).
- [40] D. Trzesniak, A.-P. E. Kunz, and W. F. van Gunsteren, *ChemPhysChem* **8**, 162 (2007).
- [41] H. E. A. Huitema and J. P. van der Eerden, *J. Chem. Phys.* **110**, 3267 (1999).
- [42] G. Rutkai and T. Kristóf, *J. Chem. Phys.* **132**, 104107 (2010).
- [43] E. Sanz and D. Marenduzzo, *J. Chem. Phys.* **132**, 194102 (2010).
- [44] K. M. Bal and E. C. Neyts, *J. Chem. Phys.* **141**, 204104 (2014).
- [45] G. Tiana, L. Sutto, and R. A. Broglia, *Physica A: Stat. Mech. Appl.* **380**, 241 (2007).
- [46] N. Gö, *Annu. Rev. Biophys. Bioeng.* **12**, 183 (1983).
- [47] S. F. Chekmarev, *J. Chem. Phys.* **139**, 145103 (2013).

- [48] N. H. Andersen, K. A. Olsen, R. M. Fesinmeyer, X. Tan, F. M. Hudson, L. A. Eidenschink, and S. R. Farazi, *J. Am. Chem. Soc.* **128**, 6101 (2006).
- [49] T. X. Hoang and M. Cieplak, *J. Chem. Phys.* **112**, 6851 (2000).
- [50] R. Biswas and D. R. Hamann, *Phys. Rev. B* **34**, 895 (1986).
- [51] S. Miyazawa and R. L. Jernigan, *J. Mol. Biol.* **256**, 623 (1996).
- [52] S. F. Chekmarev, *Phys. Chem. Chem. Phys.* **23**, 17856 (2021).
- [53] R. B. Best, G. Hummer, and W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **110**, 17874 (2013).
- [54] V. Muñoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, *Nature (London)* **390**, 196 (1997).
- [55] R. Zhou, B. J. Berne, and R. Germain, *Proc. Natl. Acad. Sci. USA* **98**, 14931 (2001).
- [56] Y. Gao, Y. Li, L. Mou, B. Lin, J. Z. H. Zhang, and Y. Mei, *Sci. Rep.* **5**, 10359 (2015).
- [57] A. R. Dinner, T. Lazaridis, and M. Karplus, *Proc. Natl. Acad. Sci. USA* **96**, 9068 (1999).
- [58] B. Zagrovic, E. J. Sorin, and V. Pande, *J. Mol. Biol.* **313**, 151 (2001).
- [59] J. Sabelko, J. Ervin, and M. Gruebele, *Proc. Natl. Acad. Sci. USA* **96**, 6031 (1999).
- [60] S. F. Chekmarev, S. V. Krivov, and M. Karplus, *J. Phys. Chem. B* **110**, 8865 (2006).
- [61] P. T. Bui and T. X. Hoang, *Biophys. J.* **120**, 4798 (2021).
- [62] S. J. Hagen, *Proteins: Struct. Funct. Genet.* **50**, 1 (2003).
- [63] J. Kubelka, J. Hofrichter and W. A. Eaton, *Curr. Opin. Struct. Biol.* **14**, 76 (2004).
- [64] K. Müller and L. D. Brown, *Theor. Chim. Acta* **53**, 75 (1979).
- [65] S. F. Chekmarev, *PLoS ONE* **10**, e0121640 (2015).
- [66] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.105.034407> for the simulation of folding of an imaginary protein on a two-dimensional energy surface with the unfolded, folded, and off-pathway states, and the comparison of the first-passage time distributions for the resulting three-state kinetics obtained in the MD and MMC simulations.