

Simple Continuous and Discrete Models for Simulating Replica Exchange Simulations of Protein Folding[†]

Weihua Zheng,[‡] Michael Andrec,[§] Emilio Gallicchio,[§] and Ronald M. Levy^{*,§}

Department of Physics and Astronomy, Rutgers, the State University of New Jersey, 136 Frelinghuysen Road, Piscataway, New Jersey 08854, and Department of Chemistry and Chemical Biology and BioMaPS Institute for Quantitative Biology, Rutgers, the State University of New Jersey, 610 Taylor Road, Piscataway, New Jersey 08854

Received: August 8, 2007; In Final Form: November 8, 2007

The efficiency of temperature replica exchange (RE) simulations hinge on their ability to enhance conformational sampling at physiological temperatures by taking advantage of more rapid conformational interconversions at higher temperatures. While temperature RE is a parallel simulation technique that is relatively straightforward to implement, kinetics in the RE ensemble is complicated, and there is much to learn about how best to employ RE simulations in computational biophysics. Protein folding rates often slow down above a certain temperature due to entropic bottlenecks. This “anti-Arrhenius” behavior represents a challenge for RE. However, it is far from straightforward to systematically explore the impact of this on RE by brute force molecular simulations, since RE simulations of protein folding are very difficult to converge. To understand some of the basic mechanisms that determine the efficiency of RE, it is useful to study simplified low dimensionality systems that share some of the key characteristics of molecular systems. Results are presented concerning the efficiency of temperature RE on a continuous two-dimensional potential that contains an entropic bottleneck. Optimal efficiency was obtained when the temperatures of the replicas did not exceed the temperature at which the harmonic mean of the folding and unfolding rates is maximized. This confirms a result we previously obtained using a discrete network model of RE. Comparison of the efficiencies obtained using the continuous and discrete models makes it possible to identify non-Markovian effects, which slow down equilibration of the RE ensemble on the more complex continuous potential. In particular, the rate of temperature diffusion and also the efficiency of RE is limited by the time scale of conformational rearrangements within free energy basins.

1. Introduction

One of the key challenges in the computer simulation of proteins at the atomic level is the sampling of conformational space. The efficiency of many common sampling protocols, such as Monte Carlo (MC) and molecular dynamics (MD), is limited by the lack of apparent ergodicity caused by high free energy barriers between conformational states and rugged energy landscapes. Replica exchange (RE) methods^{1–5} are widely employed to enhance the conformational sampling efficiency of biomolecular simulations for the study of protein biophysics, including peptide and protein folding^{6,7} and aggregation,^{8–10} and protein–ligand interactions.^{11,12} To accomplish barrier crossings, RE methods simulate a series of replicas over a range of potential parameters^{13–17} or temperatures.⁵ In the latter, replicas exchange temperatures following a Metropolis criterion designed to preserve canonical distributions. This scheme allows conformations at physiological temperatures, where conformational interconversions are rare, to switch to higher temperatures where transitions to other conformations are more likely. In a sense, therefore, the enhancement of conformational sampling at low

temperatures is achieved by “borrowing” the faster kinetics at higher temperatures.

The popularity of RE methods is due to their ease of implementation and their ability to enhance conformational sampling while preserving canonical distributions at the thermodynamic conditions of each replica. The properties of the RE algorithm and how it can be utilized most effectively for the study of protein folding and binding has received attention recently.^{18–20} The determination of the temperature assignment and number of replicas to achieve optimal temperature mixing has been the subject of a variety of studies.^{3,21–27} Recent work has also recognized the importance of conformational relaxation as a key limiting factor that can affect the efficiency of the RE algorithm.^{18,19,26,28} While temperature RE is relatively straightforward to implement, kinetics in the RE ensemble is complicated and does not correspond in any simple way to the molecular kinetics (necessitating additional methods for the reconstruction of molecular kinetics from RE samples^{29–32}). Molecular kinetics, however, can have a strong effect on RE, especially when the kinetics has complex temperature dependence. The anti-Arrhenius behavior typical of protein folding kinetics, where the folding rate above a critical threshold temperature decreases with increasing temperature,^{33–36} is understood to occur when the transition state is energetically favored but entropically disfavored with respect to the reactants. Anti-Arrhenius behavior represents a challenge for temperature

[†] Part of the “Attila Szabo Festschrift”.

* Corresponding author. E-mail: ronlevy@lutece.rutgers.edu. Phone: 732-445-3947. Fax: 732-445-5958.

[‡] Department of Physics and Astronomy.

[§] Department of Chemistry and Chemical Biology and BioMaPS Institute for Quantitative Biology.

RE because, when folding exhibits anti-Arrhenius behavior, there exists a temperature (generally unknown) at which the folding and unfolding rates are optimal. If even higher temperatures beyond the optimal are included in the RE ensemble, this may degrade performance.²⁰

Although some comparative studies aimed at determining the benefits of RE over conventional MD for peptide folding have been conducted,^{19,37,38} it is far from straightforward to systematically explore the convergence properties of RE by brute force molecular simulations, since RE simulations of protein folding are very difficult to converge. To understand some of the basic mechanisms that determine the efficiency of RE, it is useful to study simplified low dimensionality systems that share some of the key characteristics of molecular systems. We recently investigated a discrete two-state network model for replica exchange (NRE), containing two conformational states (folded and unfolded) at each of several temperatures.²⁰ We found that the efficiency of RE for this system varies non-monotonically with respect to the temperature distribution of the replicas when the folding rate displays anti-Arrhenius behavior. The model showed that the rate of folding/unfolding events in RE is maximal when high-temperature replicas are placed near the temperature at which the harmonic mean of the folding and unfolding rates for the uncoupled system (k_f and k_u) is maximal. This result suggested that, in molecular simulations, adding high-temperature replicas does not necessarily lead to increased efficiency of exploration of conformational space, and that, instead, optimal efficiency could be obtained by placing replicas at specific temperatures determined by the temperature dependence of key kinetic rates of the system.

In this paper we extend this analysis by studying a continuous two-dimensional system designed to reproduce the anti-Arrhenius kinetics of a conformational equilibrium, such as a protein folding equilibrium, mediated by an entropic bottleneck. The two-dimensional system studied here is an extension of the potential model we originally used to study the convergence of the weighted histogram analysis method,³⁹ and is very similar in spirit to the funnel-like golf course model for protein folding studied by Szabo and co-workers.⁴⁰ This two-dimensional system is sufficiently simple to be amenable to accurate analytical and numerical solution, while including some characteristics of molecular systems that were absent from the discrete NRE model. The present model is self-contained in that the kinetic rates are determined by the potential and the move set rather than being imposed, as in the NRE model of reference 20. Furthermore, and most importantly, the unfolded and folded macrostates have, like real molecular systems, microscopic internal structure. The new model makes it possible to follow the joint microscopic evolution of the system in conformational and temperature space. It incorporates the same discrete temperature exchange scheme commonly adopted in RE molecular simulations, and it allows us to study the effects of non-Markovian processes likely present in RE simulations of molecular systems.

In the next section we present the potential model and the kinetic scheme we have employed. We review the RE method and the NRE model we previously developed. We then summarize the thermodynamic and kinetic properties of the two-dimensional system and present results showing how these determine the efficiency of the RE method. The paper is then concluded by discussing the implications of these findings for RE simulations of molecular systems.

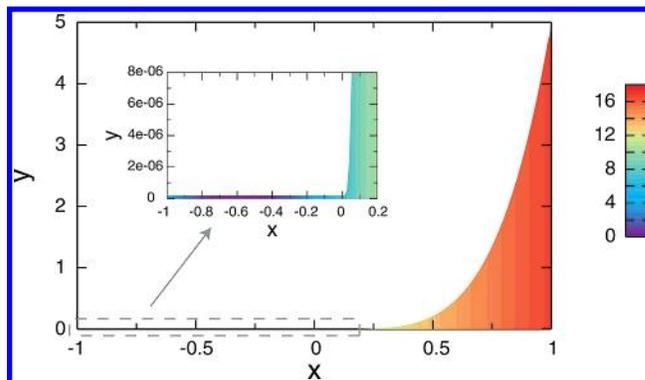


Figure 1. A schematic representation of the two-dimensional potential function used in this work. The colored area corresponds to the accessible region of the (x,y) plane, with the colors representing the magnitude of the potential energy at that (x,y) point (scale bar in kcal/mol). The potential energy is infinite in the non-colored region and for $y < 0$, $x < -1$, and $x > 1$. The inset is an enlarged view of the folded macrostate and transition region.

2. Methods

2.1. The Two-Dimensional Continuous Potential. A two-dimensional potential was constructed to mimic the anti-Arrhenius temperature dependence of the folding rate seen in proteins. We designed this potential to have an energetic barrier when going from the “folded” to the “unfolded” region, and an entropic barrier in the reverse direction. The entropic barrier is achieved by imposing a hard wall constraint that limits the space accessible to the folded region. Specifically, the particle can only move in the region $-1 \leq x \leq 1$, $0 \leq y \leq B(x)$, where the boundary function $B(x)$ is a small constant for $x \leq 0$ and an increasing function of x for $x > 0$ (Figure 1):

$$B(x) = \begin{cases} \delta & -1 \leq x \leq 0 \\ bx^{n_1} + \delta & 0 < x \leq 1 \end{cases} \quad (1)$$

The use of a boundary of this form is based on a two-dimensional potential first used in our laboratory to study the convergence of the weighted histogram analysis method,³⁹ and is very similar in spirit to simplified models for protein folding studied by Bicout and Szabo⁴⁰ and the model of an entropic barrier by Zhou and Zwanzig.⁴¹ The specific parameters δ , b , and n_1 were chosen together with the parameters of the potential function discussed below by trial and error to achieve a sufficiently strong temperature dependence to illustrate some of the possible consequences of anti-Arrhenius behavior on RE simulations. It is natural to choose the x axis to be the reaction coordinate, with $-1 \leq x \leq 0$ corresponding to the folded macrostate and $0 < x \leq 1$ corresponding to the unfolded macrostate. The move set was chosen to be compatible with this reaction coordinate (see below). In order for folding and unfolding to be activated processes, however, it is necessary to add a potential energy function that has an energetic well as a function of x in the folded region, and increases with x in the unfolded region. Specifically, we use the potential function

$$U(x,y) = \begin{cases} a_1(x+x_0)^2 & -1 \leq x < -x_1 & 0 \leq y \leq B(x) \\ -a_2x^2 + c_0 & -x_1 \leq x \leq 0 & 0 \leq y \leq B(x) \\ a_3x^{n_2} + c_0 & 0 < x < x_2 & 0 \leq y \leq B(x) \\ a_4x^{n_3} + c_1 & x_2 \leq x \leq 1 & 0 \leq y \leq B(x) \\ \infty & \text{otherwise} & \end{cases} \quad (2)$$

where $a_1 = 23.53$ kcal/mol, $a_2 = 235.3$ kcal/mol, $a_3 = 376.5$ kcal/mol, $a_4 = 11.29$ kcal/mol, $c_0 = 7.059$ kcal/mol, $b = 5$, $n_1 = 4.55$, $n_2 = 2$, $n_3 = 0.5$, and $\delta = 2 \times 10^{-7}$. The constants $x_0 = \sqrt{c_0(a_1+a_2)/a_1a_2}$, $x_1 = a_1x_0/(a_1+a_2)$, $x_2 = (a_4n_3/a_3n_2)^{1/(n_2-n_3)}$, and $c_1 = c_0 - (a_4x_2^{n_3} - a_3x_2^{n_2})$ were chosen so that the first derivative of $U(x,y)$ is continuous. A graphical representation of the two-dimensional system studied here is shown in Figure 1.

2.2. Kinetics on the Two-Dimensional Continuous Potential. We use Metropolis MC sampling to simulate the movement of a particle in this two-dimensional potential. Kinetic MC has a long history in the study of protein folding using simplified models.^{42–44} To ensure rapid equilibration along the y coordinate consistent with the choice of x as the reaction coordinate and because of the large size difference of the accessible region in the y direction between the folded and unfolded regions, we adopted an asymmetric MC proposal scheme,^{39,45} in which the step size in the y direction is proportional to $B(x)$, i.e., a proposed move (x',y') is generated uniformly in the region $x - \Delta < x' < x + \Delta$, $y - B(x)\Delta < y' < y + B(x)\Delta$. The displacement parameter Δ was chosen such that the barrier crossing is slow but not prohibitively expensive and follows a linear regime (i.e., doubling Δ causes an approximate doubling in the number of barrier crossings). To correct for the asymmetric MC proposal distribution, the factor $\theta(|y' - y|/B(x)\Delta)$ was included to satisfy detailed balance, where $\theta(z)$ equals 1 if $z < 1$ and 0 otherwise.

Rate constants in units of MC steps were obtained via MC simulation by calculating the mean first passage times (MFPTs) between the two macrostates. The same displacement parameter $\Delta = 0.05$ was used for all temperatures. A “buffer region” $-0.1 < x < 0.0437$ was defined as not belonging to either the folded or unfolded state to reduce artifactual rapid recrossings of the barrier.^{46,47} For comparison, the temperature dependence of the folding and unfolding rate constants were also estimated from the potential of mean force (PMF) using the Arrhenius equation $k = A \exp(-\Delta G^\ddagger/k_B T)$, where ΔG^\ddagger is the free energy difference between the transition state and the appropriate macrostate. Free energies were extracted from the PMF along the x axis by averaging the PMF over the macrostates and transition region using numerical integration.

2.3. RE Simulation on the Two-Dimensional Continuous Potential. RE simulations were performed by running N MC simulations at N inverse temperatures $\beta_i = (k_B T_i)^{-1}$ ($\beta_1 > \beta_2 > \dots > \beta_N$) in parallel. The state of the extended ensemble is specified by a joint configuration of N replicas $X = \{q_1, q_2, \dots, q_N\}$, where q_i is the configuration of replica i . Exchanges of configurations were attempted every N_X MC steps between pairs of replicas adjacent in temperature, and the attempted exchange $X = \{\dots, q_i, q_j, \dots\} \rightarrow X' = \{\dots, q_j, q_i, \dots\}$ was accepted with probability $w(X \rightarrow X')$. Given the potential energy function $U(q)$, the transition probability that satisfies detailed balance and reproduces the canonical ensemble is given by $w(X \rightarrow X') = \min\{1, \exp[-(\beta_j - \beta_i)(U(q_i) - U(q_j))]\}$.⁵

The efficiency of RE conformational sampling was monitored by measuring $N_{TE}(\tau|T_0)$, the number of round-trip transitions in the conformational state of a replica, conditional on the temperature of interest T_0 , that occur in a given observation time τ . A transition event is a transit of a given replica from one conformation at T_0 to the other conformation at T_0 and back again regardless of route, i.e., whether it was the result of a direct barrier crossing at T_0 or indirectly via a barrier crossing at some other temperature combined with temperature exchanges. Conceptually, this measure reflects the potential of RE to achieve rapid equilibration at the temperature of interest by

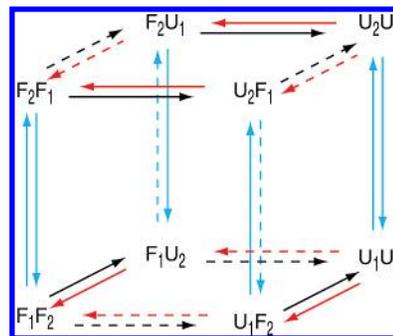


Figure 2. The kinetic network model for the discrete NRE model used by Zheng, et al.²⁰ The state labels represent the conformation (letter) and temperature (subscript) for each replica. For example, F_2U_1 represents the state in which replica 1 is folded and at temperature T_2 , while replica 2 is unfolded and at temperature T_1 . Red and black arrows correspond to folding and unfolding transitions, respectively, while the temperature at which the transition occurs is indicated by the solid and dashed lines (for T_2 and T_1 , respectively). The cyan arrows correspond to temperature exchange transitions, with the solid and dashed lines denoting transitions with rate parameters α and $w\alpha$, respectively.

means of conformational transitions at temperatures other than the temperature of interest. The transition events as defined correspond to the “reversible folding” events studied in all-atom simulations of peptide systems.^{48,49} We will use the symbol N_{TE} as a shorthand notation for $N_{TE}(\tau|T_0)$, where T_0 will generally be the lowest temperature in the simulation. For an uncoupled simulation, the number of transition events is simply the number of round trips between macrostates.

2.4. Discrete NRE. We review here the discrete kinetic network model which we devised in our recent study of RE efficiency.²⁰ In this model (unlike the continuous potential model above), the macrostates F and U (for “folded” and “unfolded”) do not possess any internal structure. Instead, it is assumed that the system evolves in time as a Poisson process, in which instantaneous transitions between F and U occur after waiting periods given by exponentially distributed random variables with means equal to the reciprocals of the folding or unfolding rates. The result (for a single replica) is an example of a “random telegraph” Markov process.⁵⁰

If the transition events are Markovian, then the simultaneous behavior of two uncoupled non-interacting replicas can be represented by the four composite states $\{F_1F_2, F_1U_2, U_1F_2, U_1U_2\}$. In each symbol, the first letter represents the configuration of replica 1, the second letter represents the configuration of replica 2, and the subscripts denote the temperature of each replica. Only transitions corresponding to a single conformational change (e.g., $F_1F_2 \rightarrow U_1F_2$) are allowed, assuming that the probability of two simultaneous changes (e.g., $F_1U_2 \rightarrow U_1F_2$) in an infinitesimal interval dt can be neglected.⁵⁰ The four-state composite system for two non-interacting replicas can be extended to create a network model of RE by introducing temperature exchanges between replicas, i.e., by allowing transitions such as $F_1U_2 \rightarrow F_2U_1$. This leads to a system with eight states arranged in a cubic network with “horizontal” folding and unfolding transitions and “vertical” temperature exchange transitions (Figure 2). For canonical equilibrium probabilities to be preserved under temperature exchanges, it is sufficient that detailed balance is satisfied, e.g., the transition probabilities $w(F_1U_2 \rightarrow F_2U_1)$ and $w(F_2U_1 \rightarrow F_1U_2)$ satisfy $P_{eq}(F_1U_2)w(F_1U_2 \rightarrow F_2U_1) = P_{eq}(F_2U_1)w(F_2U_1 \rightarrow F_1U_2)$. The ratios of forward and reverse transition probabilities for $F_1F_2 \rightleftharpoons F_2F_1$ and $U_1U_2 \rightleftharpoons U_2U_1$ are equal to 1, as interchange of temperatures does not change the equilibrium populations.

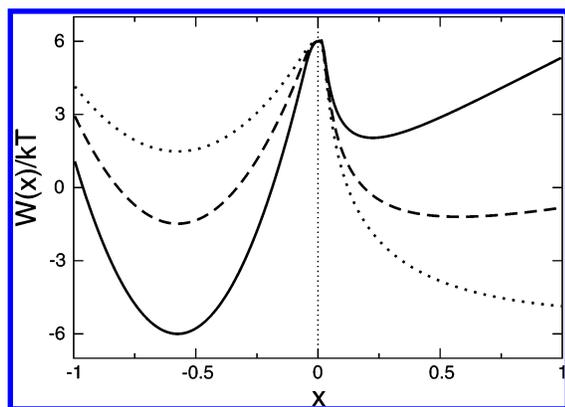


Figure 3. The PMF at three different temperatures: 296 K (solid line), 474 K (dashed line) and 789 K (dotted line). The PMF was calculated using numerical integration. To more clearly illustrate the change in the barrier height as a function of temperature, the three curves have been superimposed to coincide at $x = 0$.

The effect of the rate of temperature exchanges is included by introducing the rate parameter α , which controls the overall scaling of the temperature exchange rate relative to the folding and unfolding rates. The forward and reverse rates of the $F_1F_2 \rightleftharpoons F_2F_1$ and $U_1U_2 \rightleftharpoons U_2U_1$ transitions are set equal to α , while the other rates are set to α or $w\alpha$ as required by detailed balance, where, in this case, $w = P_{\text{eq}}(F_2U_1)/P_{\text{eq}}(F_1U_2)$ or its reciprocal such that $w < 1$ (see Figure 2). The overall average rate at which temperature exchanges occur (k_{ex}) is the probability of jumping in any instant dt from the upper to the lower face (or vice versa) of the cubic network, and is given by the equilibrium population weighted sum of the temperature exchange rates over all states:

$$k_{\text{ex}} = \frac{k_{f1}k_{f2} + 2k_{u1}k_{f2} + k_{u1}k_{u2}\alpha}{(k_{f1} + k_{u1})(k_{f2} + k_{u2})} \alpha \quad (3)$$

The NRE model was simulated using a standard method for continuous time Markov processes with discrete states,⁵⁰ also known as the ‘‘Gillespie algorithm’’. Given a current state X_0 , we identify its m neighboring states X_1, X_2, \dots, X_m and the transition rates k_1, k_2, \dots, k_m from X_0 to each of the neighboring states. We generate a waiting time in state X_0 by drawing a random number from an exponential distribution with mean $(k_1 + k_2 + \dots + k_m)^{-1}$, and select a destination state X_i from among X_1, X_2, \dots, X_m with probability $k_i/(k_1 + k_2 + \dots + k_m)$. This procedure is then repeated with the new state as the current state.

3. Results and Discussion

3.1. Thermodynamics and Kinetics of the Continuous Model System. *3.1.1. Thermodynamics.* In Figure 3 we show the PMFs corresponding to the two-dimensional potential along the x coordinate at several temperatures. PMFs calculated by MC sampling and numerical integration of the canonical distribution function agree to within statistical accuracy. The PMFs show two free energy minima corresponding to the folded ($F, x \leq 0$) and unfolded ($U, x > 0$) conformational states, separated by a free energy barrier near $x = 0$. The free energy minimum of the unfolded state and the free energy barrier have no counterparts in the potential, which is monotonically varying in both of these regions (Figure 1). These features of the PMF originate from the interplay between opposing entropic and

enthalpic driving forces. The free energy minimum of the unfolded state corresponds to the optimal balance between entropy, which drives the system toward large values of x (where the accessible space along the y coordinate is greatest), and enthalpy, which drives the system toward small values of x (where the potential energy is smallest). The free energy barrier that separates the unfolded and folded state is entropic in origin. For x near 0, the entropy is significantly reduced compared to the unfolded state, and assumes a value similar to that of the folded state (compare in Figure 1 the size of the accessible space along y at $x = 0$ and for $x > 0$ and $x < 0$). In contrast, the potential energy at $x = 0$, although smaller than in the unfolded state, is still substantially larger than in the folded state. This imbalance between entropy and potential energy causes the free energy maximum at $x = 0$.

From the point of view of folding, the free energy maximum constitutes an entropic bottleneck. In order to make a transition to the folded state, the system needs to cross the free energy barrier region at $x = 0$, where the system has lost all of the entropy required for folding without having gained all of the folding enthalpy. Similar transition bottlenecks have been described in simplified models for protein folding.^{34,40,51} After crossing this barrier, the system enters the folded state by going downhill in potential energy without further reduction in conformational entropy, since the accessible space along the y direction is the same for all points x in the folded space. Because the conformational entropy is constant for $x < 0$, the PMF in this region coincides with the potential energy. From the point of view of unfolding, the free energy maximum at $x = 0$ constitutes an enthalpic barrier. Relative to the folded state, points in the region near $x = 0$ have similar conformational entropy but larger potential energy. To reach the barrier region from the folded state, therefore, the system needs to gain potential energy (enthalpy) without the help of a concomitant increase in conformational entropy. Beyond the barrier region there is a free energy gain for moving toward the unfolded state since the gain in conformational entropy outweighs the increase in potential energy.

As shown below, the barrier region close to $x = 0$ constitutes the transition state for the folding/unfolding equilibrium. The free energy difference between the unfolded and folded states and the transition state corresponds to the free energies of activation, which determine the rate of folding and unfolding, respectively. Because of their different thermodynamic origins (entropic vs enthalpic), the free energies of activation for folding and unfolding display the opposite dependence on temperature. As Figure 3 shows, the free energy of activation for folding increases with increasing temperature relative to thermal energy (kT), where the free energy of activation for unfolding decreases with increasing temperature. This anti-Arrhenius behavior is the signature of an entropically activated process. The conformational entropy difference between the unfolded state and the transition state increases as the temperature is increased, leading to an increase in the height of the free energy barrier for folding with increasing temperature.

Figure 4 shows the temperature dependence of the population, $P_F(T)$, of the folded state, often referred to as the melting curve. The shape of the melting curve is typical of two-state protein thermal denaturation experiments. At 300 K, the system is nearly completely folded, and the fraction folded decreases with increasing temperature in favor of the unfolded state which is entropically favored. The melting temperature T_M (corresponding to equal populations of the folded and unfolded state) is approximately 460 K. At this temperature, the folded and

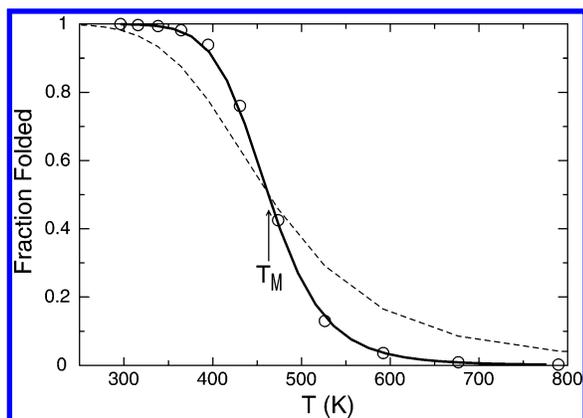


Figure 4. The temperature dependence of the fractional population folded (solid line) calculated by numerical integration of the PMF. The temperature dependence of the fraction folded corresponding to a system with a smaller average potential energy difference between the folded and unfolded states (see Appendix) is shown for comparison (dashed line). The fraction folded derived from the folding and unfolding rates obtained by MC simulation (Figure 6) is shown as circles. The melting temperature $T_M = 463$ K (corresponding to 50% folded population) is indicated.

unfolded states have equal free energy. The slope of the melting curve at the melting temperature is

$$\left(\frac{dP_F}{dT}\right)_{T=T_M} = \frac{1}{4} \frac{\bar{U}_F - \bar{U}_U}{kT^2}$$

which is proportional to the difference of the average potential energies, U_F and U_U , of the folded and unfolded states. Thus, a decrease of the average potential energy difference between the two states leads to a less steep melting curve. To illustrate this, we show in Figure 4 the melting curve corresponding to an alternative parametrization of the potential for which the average potential energy difference between the folded, unfolded, and transition states was decreased, while approximately preserving the same value of the melting temperature (see Appendix). As expected, the alternative parametrization leads to a more gradual conversion from the folded state to the unfolded state with increasing temperature (Figure 4, dashed line). The heat capacity as a function of temperature is approximately Gaussian and is peaked near T_M .

3.1.2. Kinetics. With the MC move set described in the Methods section above, the kinetics of folding/unfolding is two-state as measured by the distribution of first passage times, which is exponential (Figure 5). The Arrhenius plots of the folding and unfolding reaction rates are shown in Figure 6. The temperature dependence of the reaction rates using the Arrhenius equation with activation free energies extracted from the PMFs (Figure 3) agree well with the simulation results, and is a further indication that the kinetics is two-state and that the reaction coordinate is well represented by the x coordinate. This is a consequence of choosing a move set for which equilibration along the y coordinate is faster than that along the x coordinate. The alternative potential parametrization in the Appendix, which is characterized by a smaller average potential energy of the unfolded state relative to the folded and the transition states, leads to a weaker temperature dependence of the folding rate (Figure 6, dashed lines). Since the slope of the Arrhenius curve is proportional to the activation energy, this difference of the rates is consistent with the smaller energy of activation obtained with the alternative parametrization.

The folding rates decrease with increasing temperature, a phenomenon that has been observed in the kinetics of protein

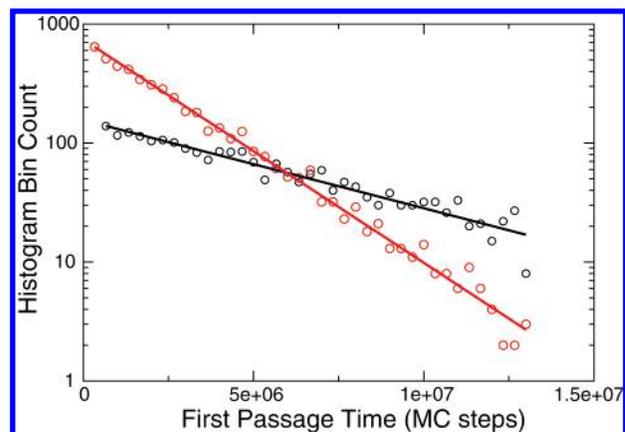


Figure 5. The distributions of first passage times for folding (black) and unfolding (red) observed during a 2.7×10^{10} -step kinetic MC at 475 K. Approximately 4700 folding and unfolding events were observed. A folding first passage time is defined as the time elapsed from when the particle enters the unfolded region from the buffer region (having previously been in the folded region), until it re-enters the folded region. The unfolding first passage time is defined similarly. The semilog plot of the histograms of the first passage times is shown as circles, while the lines represent the best-fit exponential curve.

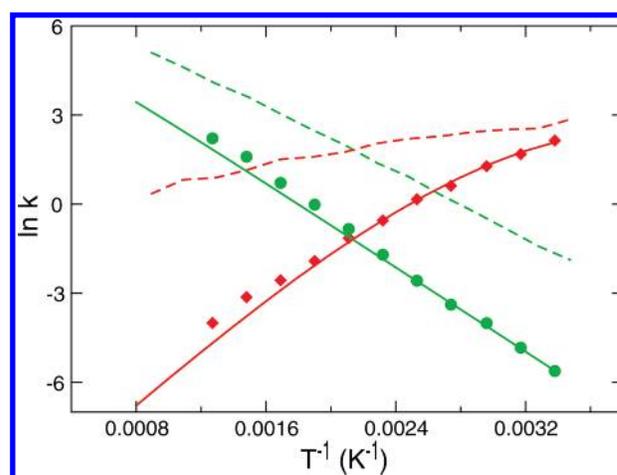


Figure 6. The temperature dependence of the folding and unfolding rate constants (solid lines and symbols). Folding and unfolding rates are indicated by red and green color, respectively. The folding and unfolding rates corresponding to a system with a smaller activation energy for folding (Appendix) are shown for comparison (dashed lines). The rate constants plotted in symbols were derived from kinetic MC simulations run at different temperatures. The solid lines represent the rates calculated using the Arrhenius equation based on activation energies derived from the PMF along x (Figure 3). Rate constants are expressed in units of 10^{-6} per MC step.

folding.^{33–36,42} Processes displaying anti-Arrhenius behavior are said to be characterized by a negative effective activation energy, whereby the enthalpy of the unfolded state is larger than that of the transition state. The free energy of activation of these processes, however, remains positive as a result of the activation entropy favoring the unfolded state. The negative activation entropy is associated with the smaller number of accessible conformations at the transition state relative to the unfolded state; that is, the transition state constitutes an entropic “bottleneck” that needs to be traversed for the transition to the folded state to occur. These elements clearly exist in the simplified two-dimensional system under investigation. Since the potential energy decreases monotonically from the unfolded state to the folded state, the average potential energy at the transition state ($x = 0$) is smaller than the average potential energy of the unfolded state, leading to the observed anti-Arrhenius behavior

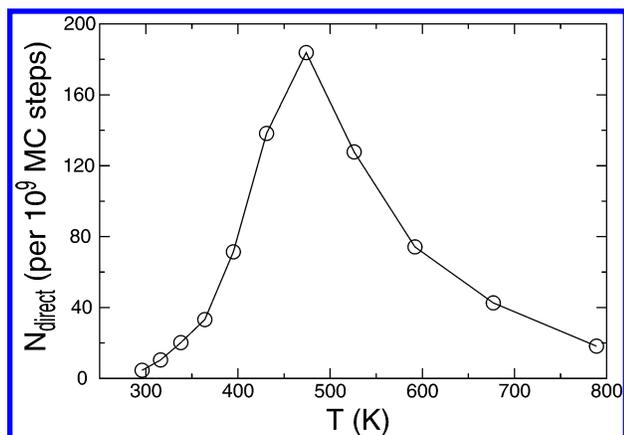


Figure 7. Number of direct round-trip transition events N_{direct} in single-temperature uncoupled simulations over the temperature range 296–789 K in 5×10^9 MC steps. The curve plotted as a solid line was calculated from the harmonic mean of the folding and unfolding rates estimated from the mean of the folding and unfolding first passage time distribution (Figure 5) obtained by MC simulations at each temperature, while the number of events counted directly from the MC simulations at individual temperatures is plotted as circles. The high level of agreement indicates that the system is very well approximated as a two-state activated process.

of the rate of folding. Despite the enthalpic driving force favoring the transition state, the free energy of activation for folding remains positive at all temperatures examined (as the calculated PMF along the x coordinate shows). This is because the entropy of the transition state is smaller than the entropy of the unfolded state because of the larger accessible configuration space along the y coordinate (Figure 1). The entropic destabilization of the transition state, which (as in protein folding) can be described as acting as a “bottleneck”, more than offsets the enthalpic stabilization, leading to the observed positive activation free energy for folding.

Often the observed folding rates of proteins show non-monotonic behavior with respect to the temperature; the folding rate increases with temperature at low temperatures as in normal Arrhenius behavior, switching to anti-Arrhenius behavior at higher temperatures, when the folding rate decreases with increasing temperature. This phenomenon is rationalized in terms of a negative activation heat capacity. The activation heat capacity is defined as the temperature derivative of the activation energy, and a negative value of the activation heat capacity indicates that the unfolded state has a larger heat capacity than the transition state. The observed negative heat capacity of activation of protein folding has been variously interpreted as being due to the hydrophobic effect^{33,42} or to the difference of the distribution of energies of the molecular conformations experienced as a function of temperature.^{34,52} The curvature of the Arrhenius plot is related to the activation heat capacity. The present simplified two-dimensional system does not have a large enough heat capacity of activation to reproduce this turnover from Arrhenius to anti-Arrhenius behavior within the temperature range we have investigated. Thus, the results extracted from this model are applicable only to the anti-Arrhenius temperature regime of the protein folding process.

Figure 7 shows the number of direct round-trip transition events N_{direct} observed during MC simulations of $N_{\text{MC}} = 5 \times 10^9$ steps as a function of temperature. We use the number of transitions as a measure of the efficiency of conformational sampling, which determines the rate of convergence of thermodynamic quantities extracted from the simulations. The results of Figure 7 show that the conformational sampling

efficiency of the uncoupled simulation varies non-monotonically with the temperature. There is a 40-fold increase in transitions from 300 to 474 K, the temperature at which the maximum is observed. This decreases for temperatures higher than 474 K, reaching a 10-fold reduction at 800 K (relative to the maximum). As the results in Figure 7 show, this behavior mirrors almost exactly the behavior of the harmonic mean $(k_f^{-1} + k_u^{-1})^{-1}$ of the folding and unfolding rates (from Figure 6) as a function of temperature (we note that our use of the term “harmonic mean” differs from standard usage by a factor of 2, which is natural given that we are considering a round trip, i.e., a single “transition event” involves two conformational transitions). The agreement between the harmonic mean of the rates and the number of direct round trip transitions is expected for a two-state activated equilibrium, since the average time of a round-trip excursion from the folded to the unfolded state and back is the sum of the average folding and unfolding times $\tau_f = k_f^{-1}$ and $\tau_u = k_u^{-1}$, respectively: $N_{\text{direct}} = N_{\text{MC}}/(\tau_f + \tau_u)$.

3.2. RE Simulations Using MC on the Continuous Potential. In a recent paper,²⁰ we analyzed the convergence and efficiency of RE using a discrete model for folding and unfolding. We found that when the physical kinetics shows anti-Arrhenius temperature dependence, there exists an optimal maximal temperature beyond which the efficiency of the RE method is degraded. Similar behavior is expected from RE simulations using the continuous two-dimensional potential, with possible differences arising from the more complex nature of the present model, where the folded and unfolded states have internal structure. We performed RE simulations on the continuous two-dimensional potential with MC as the dynamic propagator and RE proposals made periodically between adjacent temperatures every N_X MC steps. The efficiency of conformational sampling was monitored by counting the number of temperature-conditional transition events N_{TE} defined in section 2.3 above.

In order to directly compare with the results obtained previously, we first performed RE using two replicas. Although such a simulation would not be realistic in general for a protein system due to poor energy overlap and very inefficient temperature exchange, it is feasible in the two-dimensional potential. The result for a 2×10^9 -step simulation where the lower temperature is held fixed at 296 K and the upper temperature varies from 296 to 789 K is shown in Figure 8 (green, red, and blue dots). We see behavior similar to that seen for the discrete model studied previously: the number of temperature-conditional transitions N_{TE} has non-monotonic behavior and exhibits a maximum at an optimal high temperature given by the maximal harmonic mean of the folding and unfolding rates (474 K). This maximum point is approximately independent of the rate at which attempted temperature exchanges occur. While the location of the maximum is in agreement with our previous results,²⁰ the magnitude of the number of transition events is not. We have shown that, for NRE simulations employing a two-state model (folded and unfolded states), the number of transition events is given by the average over all temperatures of the harmonic means of the folding and unfolding rates, provided that the rate of temperature exchanges is sufficiently fast.²⁰ In the continuous model, we find that the number of transitions is significantly lower than that predicted from the average of the harmonic means of the rates (Figure 8, black dashed line). This may be due to the finite rate of temperature exchanges, deviations from the pure Markovian kinetics of the two-state discrete model, or a combination of these effects.

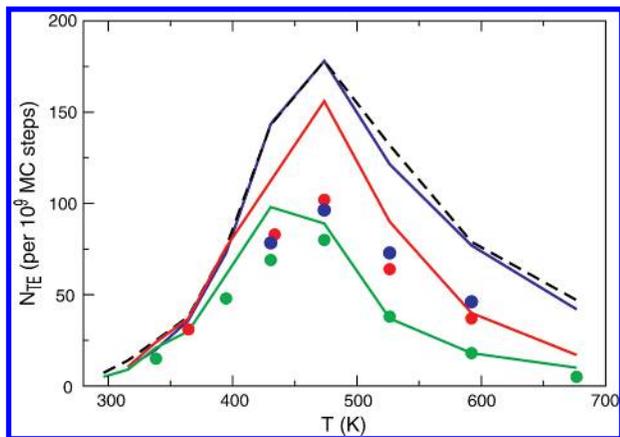


Figure 8. The dependence of the number of temperature-conditional transition events N_{TE} (section 2.3) on the temperature of the high-temperature replica for a two-replica simulation on the continuous potential (circles), and comparison with predicted transition events using the discrete NRE model (Figure 2) (curves). For all simulations, one replica was fixed at 296 K and the other replica was scanned from 296 to 789 K. The black dashed line corresponds to the discrete model prediction in the large- α limit. The solid curves are the predicted N_{TE} using the NRE model with temperature-dependent folding and unfolding rates taken from the kinetic MC simulations on the continuous potential (shown in Figure 6) and using an α corresponding to an attempted temperature exchange. The circles are the N_{TE} values observed in 2×10^9 MC step RE simulations on the continuous potential. The green, red, and blue data correspond to N_X values of 1 000, 200, and 20, respectively.

TABLE 1: Number of Temperature-Conditional Transition Events in 2×10^9 MC Steps for Two Replicas^a as a Function of the Number of MC Steps between Attempted Temperature Exchanges (N_X), and Observed Temperature-Conditional MFPTs^b

N_X	N_{TE} per replica		temperature-conditional MFPTs	
	observed (continuous)	predicted (NRE)	$F_1 \rightarrow U_1$	$U_1 \rightarrow F_1$
10 000	22	24	91.8	5.6
2 000	52	73	31.6	5.7
1 000	80	105	23.1	5.7
500	93	134	19.1	5.8
200	102	162	16.0	5.7
100	99	168	14.3	5.9
80	98	172	14.6	5.8
50	98	176	14.8	5.7
20	96	177	14.9	6.1
0 ^c		178		

^a With temperatures of 296 and 474 K. ^b In units of 10^6 MC steps; see text for details. ^c Predicted N_{TE} based on the harmonic mean relationship for the $\alpha \rightarrow \infty$ limit.

To test whether this reduced number of transitions is due to insufficiently fast temperature exchange attempts, we performed several simulations in which we varied N_X (the number of MC steps between attempted temperature exchanges). We see in Table 1 that N_{TE} is approximately constant provided that the attempted exchange rate is faster than a critical value of $N_X \approx 500$. For less frequent exchange attempts, we see a substantial decrease in the number of transitions. Thus, the number of unfolding and refolding transitions cannot be increased simply by increasing the rate of attempted exchanges.

3.3. Non-Markovian Effects Revealed by Comparison of Continuous and Discrete RE Simulations. To explore causes for the observed transition deficit, we performed simulations using the discrete NRE model (Figure 2) using kinetic parameters derived from the two-dimensional continuous potential (Figure 6). To map the rates determined using the continuous

TABLE 2: Empirical “Reverse-Engineered” Rates at Temperatures $T_1 = 296$ K and $T_2 = 474$ K (in Units of 10^{-6} MC Step) from Continuous Potential Simulation Data Assuming the Network Topology of Figure 2

	uncoupled rates	reverse-engineered rates			
		$N_X = 10\ 000$	$N_X = 2\ 000$	$N_X = 200$	$N_X = 100$
k_{f1}	6.08	5.66	6.10	5.27	6.33
k_{u1}	0.0036	0.0038	0.0036	0.0037	0.0037
k_{f2}	0.297	0.288	0.299	0.290	0.306
k_{u2}	0.420	0.420	0.419	0.427	0.425

potential to the discrete model, we used the folding and unfolding rates directly, expressed in units of 10^{-6} per MC step. Different values of α were used for the $F_1F_2 \rightleftharpoons F_2F_1$ and $U_1U_2 \rightleftharpoons U_2U_1$, and were set to $10^6/N_X$ multiplied by the empirical acceptance rate when both replicas are in the folded or unfolded state (0.853 and 0.395, respectively).

If we compare the observed number of transitions seen in the continuous model with the number predicted by the NRE model with the same rate parameters (Table 1), we see that there is good agreement when the attempted exchange rate is small, but substantial disagreement when it becomes larger. In particular, while the number of transitions using the continuous model reaches a plateau value at $N_X \approx 1000$, the predicted number of transitions in the NRE model continues to increase, asymptotically approaching the value predicted by the average of harmonic means. Similarly, comparison of the predicted and observed number of transitions as a function of temperature (Figure 8) shows a significant overestimation of the transition rate by the NRE model, and shows that this overestimation is much more severe when the rate of attempted temperature exchanges is fast. For example, while the N_{TE} predicted from the NRE model has essentially reached the asymptotic limit when $N_X = 20$ (blue curve), the observed N_{TE} values are essentially unchanged relative to those obtained when $N_X = 200$ (compare blue and red circles). The continuous two-dimensional model thus appears to contain an inherent “speed limit”, which prevents it from achieving the transition rates expected for a fully Markovian system, even if the temperature exchanges are attempted frequently.

One possible origin of this speed limit is that the average effective rates are different in the coupled and uncoupled systems. To test this, we analyzed the kinetics of the continuous RE simulation by using the NRE model to “reverse-engineer” the apparent rates by estimating the mean residence times and branching ratios for various RE macrostates. If the system is Markovian, then the rate k_{tot} given by the inverse of the mean residence time is the sum of the rates exiting that state. The rate corresponding to a given edge can then be estimated by multiplying k_{tot} by the fraction of residences that exit via that edge (the branching ratio). The results are shown in Table 2. The reverse-engineered rates generally agree with the uncoupled folding and unfolding rates estimated from kinetic MC, and this is true for both rapid and slow attempted temperature exchange rates. Therefore, the temperature exchanges do not perturb the average kinetics of the system, and cannot be a cause of the limit on the transition rates at rapid temperature exchange rates.

In order to further investigate the origin of the observed speed limit, we calculated the MFPTs for temperature conditional folding and unfolding, i.e., the average time for a replica unfolded at low temperature to become folded at low temperature (regardless of path), or vice versa. The resulting MFPTs for the continuous potential are shown in Table 1. We see there that the N_{TE} speed limit arises exclusively from a limitation in the fastest achievable unfolding rate, since the folding process

is independent of N_X and is not rate limiting. This can be understood by noting that the values of α corresponding to the N_X values used are at least 2 orders of magnitude larger than the folding and unfolding rates. To unfold, the system need only make use of temperature exchange transitions that correspond to α (i.e., the solid cyan arrows of Figure 2). Since α is already much larger than the other rates, changes to it due to changes in N_X will not significantly change the MFPT for folding.

On the other hand, the unfolding process (if it occurs via an indirect route, which is likely given the very small value of k_{u1}) requires the system to use a “ $w\alpha$ edge” (i.e., a dashed cyan arrow in Figure 2). Since $w \approx 10^{-4}$ for the temperatures used here, $w\alpha$ is now slower than or comparable to the folding and unfolding rates, and therefore changes in N_X can make a substantial impact on the unfolding MFPT. Thus, the N_{TE} speed limit can be traced to the kinetics of temperature conditional unfolding, and must arise from some difference between the unfolding kinetics in the continuous potential and the fully Markovian NRE model.

One obvious way in which the continuous and NRE models differ is that the macrostates in the continuous potential have spatial extent, unlike the NRE states which lack internal structure. This means that a finite time is required for the particle to transit the nonequivalent microstates that make up the two wells. In fact, we observe that the correlation time for diffusion in the x direction in the *unfolded well* at 474 K is approximately 1 400 MC steps. This time scale is of the same magnitude as the N_X value at which the speed limit effect of Table 1 begins to occur, suggesting that there may in fact be a connection between the observed N_{TE} speed limit and conformational diffusion within the free energy wells. Such dependence of the kinetics on the internal structure of the macrostate can lead to non-Markovian behavior.

Formally, a process is Markovian if and only if the observed propagators (Green’s functions) do not depend on the history of the trajectory prior to the current state, i.e.,

$$P(x_3, t_3 | x_1, t_1; x_2, t_2) = P(x_3, t_3 | x_2, t_2) \quad (4)$$

for all states x_1, x_2, x_3 and all times $t_1 < t_2 < t_3$. Although eq 4 could be used to directly detect deviations from Markovian behavior, previous work has typically used other analysis methods to detect such deviations.^{29,53,54} For example, in a Markovian process, the rate matrix \mathbf{K} determines the propagators via the master equation

$$\dot{\mathbf{p}}(t) = \mathbf{K}\mathbf{p}(t) \quad (5)$$

where $\mathbf{p}(t)$ is the vector of propagators at time t . The formal solution of eq 5 is given by $\mathbf{p}(t) = e^{\mathbf{K}t}\mathbf{p}(0)$, and therefore $e^{\mathbf{K}\tau}$ can be thought of as a transition matrix $\mathbf{T}(\tau)$, i.e., the matrix of probabilities of being in state x_j at time τ given that the system was in state x_i at time 0. If we denote the eigenvalues of \mathbf{K} by $\lambda_1 > \lambda_2 > \dots$ and the eigenvalues of $\mathbf{T}(\tau)$ by $\mu_1(\tau) > \mu_2(\tau) > \dots$, then $\mu_i(\tau) = e^{\lambda_i\tau}$. This can be used as a test of Markovian behavior, since $\mathbf{T}(\tau)$ can be empirically estimated from a trajectory. Different values of the lag time τ will yield different values of $\mu_i(\tau)$; however, $\tau/\ln \mu_i(\tau)$ should be independent of τ if the kinetics is Markovian.^{29,54} Alternatively, the Markov property can be tested by analyzing the transition probabilities as a function of lag time using an information theoretic measure based on Shannon’s entropy.⁵³

We have chosen to detect deviations from Markovian kinetics by examining the observed residence time distributions and branching ratios, which provides insights into the physical origin

TABLE 3: History Dependent and Independent Branching Probabilities from State U_2F_1

conditional probability	maximum likelihood estimate ^a (and 95% credible interval ^b)	
	$N_X = 200$	$N_X = 10\,000$
$P(U_1F_2 U_1F_2, U_2F_1)$	0.906 (0.904, 0.908)	0.168 (0.144, 0.195)
$P(U_1F_2 F_1F_2, U_2F_1)$	0.521 (0.153, 0.530)	0.094 (0.088, 0.101)
$P(U_1F_2 U_2F_1)$	0.849 (0.846, 0.851)	0.103 (0.096, 0.110)
$P(F_2F_1 F_2F_1, U_2F_1)$	0.477 (0.469, 0.486)	0.895 (0.888, 0.902)
$P(F_2F_1 U_1F_2, U_2F_1)$	0.092 (0.090, 0.094)	0.816 (0.788, 0.841)
$P(F_2F_1 U_2F_1)$	0.150 (0.147, 0.152)	0.886 (0.878, 0.893)

^a Maximum likelihood estimates determined using $P(a_i|b,c) = \#(b,c,a_i)/\sum_i \#(b,c,a_i)$ and $P(a_i|b) = \sum_i \#(c_i,b,a_i)/\sum_{jk} \#(c_j,b,a_i)$, where $\#(i,j,k)$ is the number of occurrences of the ordered triple (i,j,k) .
^b Bayesian credible intervals under a uniform prior given by the 0.025 and 0.975 quantiles of the distribution $P(p) \propto p^n(1-p)^{N-n}$, where n and N are the numerator and denominator, respectively, of the fraction used to calculate the maximum likelihood estimate.

and the mechanism by which the non-Markovian effects enter into the stochastic process. In our simulations on the continuous potential, we have found that the residence time distributions in the macrostates are exponential to within statistical uncertainty (data not shown), and thus by themselves are consistent with Markovian kinetics. The branching probabilities, however, are significantly dependent on the preceding macrostate. We focused on transitions entering and leaving the thermodynamically favored U_2F_1 macrostate (or its symmetry-related state F_1U_2). We ran several trajectories using different rates of attempted temperature exchange and tallied the number of times each macrostate sequence (X, U_2F_1, Y) was observed in each (where $X, Y \in \{F_2F_1, U_2U_1, U_1F_2\}$). These counts were transformed into normalized branching probabilities, where $P(X|Y)$ denotes the history-independent branching probability of next visiting macrostate X given that the system is currently in macrostate Y , and $P(X|Z, Y)$ denotes the history-dependent branching probability of next visiting macrostate X given that the system is currently in macrostate Y and had been in macrostate Z immediately prior (Table 3).

If the kinetics is Markovian, then the history-dependent and corresponding history-independent branching probabilities will be equal:

$$P(X|Z, Y) = P(X|Y)$$

from which it follows that history-dependent branching probabilities that differ only in the history condition will also be equal:

$$P(X|Z, Y) = P(X|W, Y)$$

This is clearly not the case for the data in Table 3. For example, we see that the history-dependent branching probabilities $P(U_1F_2|F_2F_1, U_2F_1)$ and $P(F_2F_1|F_2F_1, U_2F_1)$ differ significantly from their corresponding history-independent branching probabilities $P(U_1F_2|U_2F_1)$ and $P(F_2F_1|U_2F_1)$, and the branching probability $P(U_1F_2|F_2F_1, U_2F_1)$ is significantly smaller than $P(U_1F_2|U_1F_2, U_2F_1)$. This is most pronounced when the rate of attempted temperature exchanges is fast.

Examination of the kinetic scheme of Figure 2 indicates that the deviations from Markovian behavior seen in Table 3 are consistent with a reduction in the number of temperature-conditional round-trip conformational transition events. If the unfolding rate at low temperature is negligible, then a low-temperature folded conformation unfolds predominantly via indirect paths of the form $F_1F_2 \rightarrow F_2F_1 \rightarrow U_2F_1 \rightarrow U_1F_2$ or $F_1U_2 \rightarrow F_2U_1 \rightarrow U_2U_1 \rightarrow U_1U_2$. In the former case, the F_2F_1

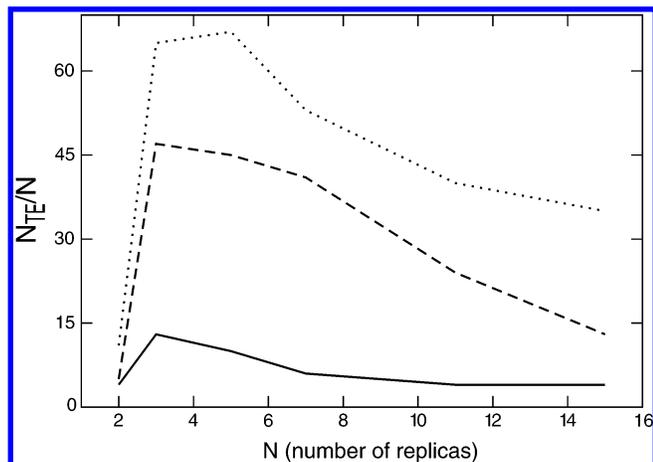


Figure 9. Number of transition events N_{TE} (section 2.3) normalized by the number of replicas in 2×10^9 MC steps for 2 to 15 replicas exponentially distributed in temperature from 296 to 789 K. Temperature exchanges were attempted every 10 000 (solid line), 1 000 (dashed line), and 200 (dotted line) MC steps.

→ U_2F_1 step is more likely to be reversed when the temperature exchange rate is rapid (Table 3), as is the $F_1U_2 \rightarrow F_2U_1$ step in the latter case (which follows by symmetry from the $U_2F_1 \rightarrow U_1F_2$ results of Table 3). Thus, increasing the rate of attempted temperature exchanges increases the probability of counterproductive backtracking relative to the Markovian case, resulting in a decrease in the rate of temperature-conditional unfolding events, and therefore a corresponding decrease in N_{TE} (since temperature-conditional unfolding was shown above to be rate-limiting).

Although the results presented here do not identify the physical origin of the non-Markovian kinetics, we hypothesize that it is due to the finite time required for diffusion of the particle within the macrostates. This effect does not arise in the NRE model, since, in there, the macrostates have no internal structure, and the probability of making a transition to a given macrostate at any instant dt is the same, regardless of which macrostate the system was in previously or how long it has been in the current macrostate. The behavior of the continuous system within the wells is not Markovian, since the system has memory that is mediated by conformational diffusion within the macrostate. This correlation in time of the particle's position (and energy) implies that there is a maximal effective value of the rate of statistically independent temperature exchanges, which is limited by the time required for conformational relaxation *within* the folded and unfolded macrostates.

3.4. Dependence of RE Efficiency on the Number of Replicas. The above results were obtained with two replicas, which is not typical for RE simulations that would be carried out for peptides and proteins. To investigate the effect of adding additional replicas, we performed a series of simulations of 2×10^9 MC steps with 2 to 15 replicas distributed uniformly in T^{-1} from 296 to 789 K. The results are shown in Figure 9. One important issue that arises when considering such a set of results is the appropriate measure of conformational sampling efficiency of RE. If we consider the total number of transition events N_{TE} (direct and indirect) in all replicas, then we would see for the most part a monotonic increase of efficiency as a function of the number of replicas N simply because the number of indirect "channels" for transitions is linearly increasing. This measure of efficiency, however, implicitly assumes that computer power is inexpensive and that the convergence rate of the simulation is the important limiting factor. If both computer resources and the convergence rate are limiting factors, a more appropriate

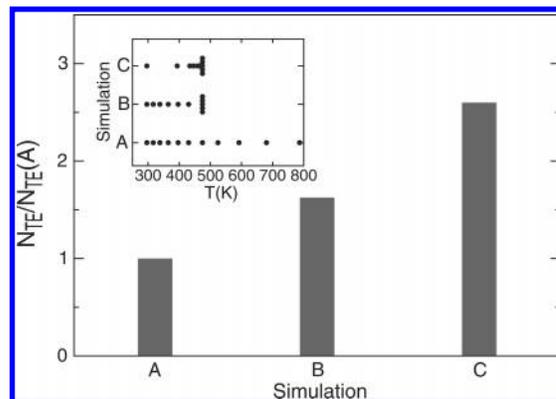


Figure 10. Number of transition events N_{TE} (section 2.3) observed in 2×10^9 MC steps for three different 11-replica RE simulations performed using the continuous potential with $N_X = 200$. The temperature distributions for the three simulations are shown in the inset. Transition counts have been normalized by the N_{TE} of simulation A.

measure is the computational efficiency calculated as the number of transition events per replica (N_{TE}/N). According to this measure, a RE simulation with $N + 1$ replicas is considered more efficient than one with N replicas only if the introduction of the additional replica provides more than a proportional increase in the number of transition events at the temperature of interest.

We find that the efficiency increases strongly as a function of N when N is small, reaches a maximum, and then decreases with N for larger N (Figure 9). This pattern is unchanged as a function of the rate of attempted temperature exchanges, showing a scaling approximately consistent with the results in Table 1. The trends seen here are qualitatively similar to that seen previously in the NRE two-state discrete model²⁰ with finite α . In that work, we attributed the decrease with increasing number of replicas beyond an optimum value in part to a combinatoric effect that decreases the relative size of the "target" space of configurations in which a replica is at the temperature of interest relative to the total temperature/configuration space. It is reasonable to assume that a similar effect is occurring here as well. We will address this in a future communication.

The results in Figure 9 were obtained with a relatively uniform distribution of temperatures. It is of interest to consider the effect on efficiency of changing that temperature distribution. In our previous work,²⁰ we concluded that, in the context of the discrete network model in the "large α " limit, the optimal temperature distribution is one replica at the temperature of interest, and the rest at the temperature that maximizes the harmonic mean of the folding and unfolding rates. That model, however, was limited in its realism in that it did not have explicit energy distribution functions. Furthermore, it is clear from the results presented in the previous section that very large effective values of α may not be achievable in real systems. The continuous two-dimensional potential studied here provides a better test system for studying these questions.

In Figure 10 we show the relative number of temperature-conditional transition events in 2×10^9 MC steps for three different temperature distributions of 11 replicas: (A) uniformly distributed in T^{-1} from 296 to 789 K, (B) 6 replicas uniformly distributed in T^{-1} from 296 to 474 K (the optimal temperature) and the remaining 5 "bunched up" at the optimal temperature, and (C) 5 replicas bunched up at the optimal temperature with the remaining distributed in the 296 to 474 K range but strongly skewed toward the optimal temperature. Temperature distribution B provides more than a 50% increase in efficiency relative

to the uniform distribution over the large temperature range. This is consistent with our discrete model results, and indicates that it is possible to include temperatures that are “too high” when the system exhibits anti-Arrhenius kinetics. However, we can increase the efficiency even further (to more than a factor of 2.5 over the baseline result) by skewing the temperature distribution to increase the number of replicas in the vicinity of the transition temperature (distribution C). Previous work by Hansmann et al. has suggested that such concentration of the temperatures near a bottleneck can improve temperature mixing.²⁶ However, the improved efficiency may simply be due to the increased number of replicas near the optimal temperature. The clarification of the relative contributions from these two effects will also be addressed in a future communication.

4. Conclusions

One of the challenges of studying the computational efficiency of RE has been the difficulty in running molecular simulations sufficiently long to obtain full convergence and meaningful statistics. This is particularly daunting if such simulations must be run multiple times to assess the effect of differences in simulation protocols and parameters. The use of simplified model systems allows for thorough theoretical, conceptual, and computational analysis of the problem that can provide insights into the factors that limit the efficiency of RE in more realistic molecular systems.

Our previous work made use of a highly simplified discrete model for protein folding with two conformational states at several temperatures.²⁰ While this system did provide useful insights, it was limited in a number of ways, and, in particular, was fully Markovian. Here we have described a two-dimensional continuous potential function and an associated move set that allows us to perform MC and RE MC simulations in a system that is small enough to quickly converge but yet is rich in a complexity that is reminiscent of molecular systems. While many of the results are consistent with those observed previously, novel effects are also seen. In particular, we have confirmed that the efficiency of RE in more complex systems is fundamentally limited by the time scale of conformational diffusion within basins, as we had anticipated.²⁰ We expect that such behavior will also be present (perhaps even more strongly) in molecular systems.

There are many unresolved questions raised by this work. One question for which our two-dimensional system would be a good model is for studying the relationship between conformational and thermal diffusion. Optimization of the diffusion of replicas in temperature space has been a major focus of recent theoretical and computational study of the RE method.^{3,21–27} However, the convergence of thermodynamic quantities is not limited by thermal diffusion *per se*, but by the exploration of the conformational space of the system. While very poor thermal diffusion obviously defeats the purpose of RE by effectively reducing it to a set of parallel uncoupled simulations, it is not clear that further optimization of thermal diffusion that is already “reasonably good” will automatically improve convergence. Some recent work has begun to address the role of basin-to-basin transitions.^{18,28} Similarly, some work on the optimization of thermal diffusion has emphasized the role of temperature bottlenecks,²⁶ which may turn out to be fundamentally conformational in nature. The exact relationship between thermal and conformational diffusion remains to be fully clarified, and we look forward to studying this and other questions using simplified continuous and discrete models of RE.

Acknowledgment. We thank Attila Szabo for discussions concerning this work. R.M.L. wishes to express his pleasure at

being fortunate enough to have known Attila for 30 years and his admiration for Attila’s intellectual honesty and style of doing science. We are very glad to participate in this special issue of the *Journal of Physical Chemistry* which honors Attila Szabo on his 60th birthday. This work has been supported by a grant from the National Institutes of Health (GM30580).

5. Appendix

The alternative potential with decreased average potential energy differences between folded, unfolded, and transition states is of the same general form as the primary potential described in the Methods section and Figure 1, but with the boundary function parameters $\delta = 10^{-5}$, $b = 1$, and $n_1 = 3.5$ and potential energy

$$U(x,y) = \begin{cases} a_1(x + x_0)^2 + b_1y^2 & -1 \leq x < -x_1 & 0 \leq y \leq B(x) \\ -a_2x^2 + b_1y^2 + c_0 & -x_1 \leq x \leq 0 & 0 \leq y \leq B(x) \\ a_3x^2 + b_1y^2 + c_0 & 0 < x \leq 1, & 0 \leq y \leq B(x) \\ \infty & \text{otherwise} \end{cases}$$

with $a_1 = 25$ kcal/mol, $a_2 = 250$ kcal/mol, $a_3 = 10$ kcal/mol, $b_1 = 1000$ kcal/mol, and $c_0 = 6$ kcal/mol. The constants x_0 and x_1 were the same as for the primary potential.

References and Notes

- (1) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.
- (2) Geyer, C. J.; Thompson, E. A. *J. Am. Stat. Assoc.* **1995**, *90*, 909–920.
- (3) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.
- (4) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
- (5) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (6) Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *84*, 775–786.
- (7) Nymeyer, H.; Gnanakaran, S.; García, A. E. *Meth. Enzymol.* **2004**, *383*, 119–149.
- (8) Cecchini, M.; Rao, F.; Seeber, M.; Caffisch, A. *J. Chem. Phys.* **2004**, *121*, 10748.
- (9) Tsai, H.-H. G.; Reches, M.; Tsai, C.-J.; Gunasekaran, K.; Gazit, E.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8174–8179.
- (10) Baumketner, A.; Shea, J.-E. *Biophys. J.* **2005**, *89*, 1493–1503.
- (11) Verkhivker, G. M.; Rejto, P. A.; Bouzida, D.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Gehlhaar, D. K.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. *Chem. Phys. Lett.* **2001**, *337*, 181–189.
- (12) Ravindranathan, K. P.; Gallicchio, E.; Friesner, R. A.; McDermott, A. E.; Levy, R. M. *J. Am. Chem. Soc.* **2006**, *128*, 5786–5791.
- (13) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
- (14) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058–9067.
- (15) Kwak, W.; Hansmann, U. H. E. *Phys. Rev. Lett.* **2005**, *95*, 138102.
- (16) Liu, P.; Huang, X.; Zhou, R.; Berne, B. J. *J. Phys. Chem. B* **2006**, *110*, 19018–19022.
- (17) Min, D.; Li, H.; Li, G.; Bitetti-Putzer, R.; Yang, W. Dual-topology Hamiltonian-replica-exchange overlap histogramming method to calculate relative free energy difference in rough energy landscape. <http://arxiv.org/abs/physics/0605005v1>, accessed 10/07.
- (18) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1200–1202.
- (19) Beck, D. A. C.; White, G. W. N.; Daggett, V. *J. Struct. Biol.* **2007**, *157*, 514–523.
- (20) Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15340–15345.
- (21) Kofke, D. A. *J. Chem. Phys.* **2002**, *117*, 6911–6914.
- (22) Kone, A.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 206101.
- (23) Predescu, C.; Predescu, M.; Ciobanu, C. V. *J. Chem. Phys.* **2004**, *120*, 4119–4128.
- (24) Predescu, C.; Predescu, M.; Ciobanu, C. V. *J. Phys. Chem. B* **2005**, *109*, 4189–4196.
- (25) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 024111.
- (26) Trebst, S.; Troyer, M.; Hansmann, U. H. E. *J. Chem. Phys.* **2006**, *124*, 174903.

- (27) Nadler, W.; Hansmann, U. H. E. On dynamics and optical number of replicas in parallel tempering simulations. <http://arxiv.org/abs/0709.3289v1>, accessed 10/07.
- (28) Zuckerman, D. M. *J. Chem. Theory Comput.* **2006**, *2*, 1693.
- (29) Swope, W. C.; Pitera, J. W.; Suits, F. *J. Phys. Chem. B* **2004**, *108*, 6571–6581.
- (30) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. *J. Phys. Chem. B* **2004**, *108*, 6582–6594.
- (31) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801–6806.
- (32) van der Spoel, D.; Seibert, M. M. *Phys. Rev. Lett.* **2006**, *96*, 238102.
- (33) Oliveberg, M.; Tan, Y.-J.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8926–8929.
- (34) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11–27.
- (35) Ferrara, P.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **2000**, *104*, 5000–5010.
- (36) Yang, W. Y.; Gruebele, M. *Biochemistry* **2004**, *43*, 13018–13025.
- (37) Zhang, W.; Wu, C.; Duan, Y. *J. Chem. Phys.* **2005**, *123*, 154105.
- (38) Periole, X.; Mark, A. E. *J. Chem. Phys.* **2007**, *126*, 014903.
- (39) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, *109*, 6722–6731.
- (40) Bicout, D.; Szabo, A. *Protein Sci.* **2000**, *9*, 452–465.
- (41) Zhou, H.-X.; Zwanzig, R. *J. Chem. Phys.* **1991**, *94*, 6147–6151.
- (42) Chan, H. S.; Dill, K. A. *Proteins* **1998**, *30*, 2–33.
- (43) Kolinski, A.; Skolnick, J. *Polymer* **2004**, *45*, 511–524.
- (44) Tiana, G.; Sutto, L.; Broglia, R. A. *Physica A* **2007**, *380*, 241–249.
- (45) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987.
- (46) Chandler, D. *J. Chem. Phys.* **1978**, *68*, 2959–2970.
- (47) Levy, R. M.; Karplus, M.; McCammon, J. A. *Chem. Phys. Lett.* **1979**, *65*, 4–11.
- (48) Rao, F.; Caflisch, A. *J. Chem. Phys.* **2003**, *119*, 4035–4042.
- (49) Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. *J. Mol. Biol.* **2005**, *354*, 173–183.
- (50) Gillespie, D. T. *Markov Processes: An Introduction for Physical Scientists*; Academic Press: Boston, 1992.
- (51) Borreguero, J.; Dokholyan, N.; Buldyrev, S.; Shakhnovich, E.; Stanley, H. *J. Mol. Biol.* **2002**, *318*, 863–876.
- (52) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93*, 6902–6915.
- (53) Park, S.; Pande, V. S. *J. Chem. Phys.* **2006**, *124*, 054118.
- (54) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. *J. Chem. Phys.* **2007**, *126*, 155102.