



King's Research Portal

DOI:

[10.1063/1.5004774](https://doi.org/10.1063/1.5004774)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Leahy, C. T., Kells, A., Hummer, G., Buchete, N.-V., & Rosta, E. (2017). Peptide dimerization-dissociation rates from replica exchange molecular dynamics. *Journal of Chemical Physics*, 147(15).
<https://doi.org/10.1063/1.5004774>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Peptide dimerization-dissociation rates from replica exchange molecular dynamics

*Cathal T. Leahy,^{1,2} Adam Kells,³ Gerhard Hummer,⁴
Nicolae-Viorel Buchete,^{1,2,*,**} and Edina Rosta^{3,*,**}*

¹School of Physics, University College Dublin, Belfield, Dublin 4, Ireland

²Institute for Discovery, University College Dublin, Belfield, Dublin 4, Ireland

³Department of Chemistry, King's College London, London SE1 1DB, United Kingdom

⁴Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Max-von-Laue-
Straße 3, D-60438 Frankfurt am Main, Germany

* Corresponding Authors

** E-mail: buchete@ucd.ie

*** E-mail: edina.rosta@kcl.ac.uk

Date: September 12th, 2017

ABSTRACT

We show how accurate rates of formation and dissociation of peptide dimers can be calculated using direct transition counting (DTC) for analyzing replica-exchange molecular dynamics (REMD) simulations. First, continuous trajectories corresponding to system replicas evolving at different temperatures are used to assign conformational states. Second, we analyze the entire REMD data to calculate the corresponding rates at each temperature directly from the number of transition counts. Finally, we compare the kinetics extracted directly, using the DTC method, with indirect estimations based on trajectory likelihood maximization using short-time propagators, and on decay rates of state autocorrelation functions. For systems with relatively low-dimensional intrinsic conformational dynamics, the DTC method is simple to implement and leads to accurate temperature-dependent rates. We apply the DTC rate-extraction method to all-atom REMD simulations of dimerization of amyloid-forming NNQQ tetrapeptides in explicit water. In an assessment of the REMD sampling efficiency with respect to standard MD, we find a gain of more than a factor of two at the lowest temperature.

INTRODUCTION

High performance computing hardware capacity has continued its remarkable growth in recent years with speeds rising by and large in accordance with Moore's Law. From a software point of view, the developments have been equally dramatic. Molecular dynamics (MD) packages are now capable of reaching microsecond simulations routinely and millisecond simulations are accessible on machines specialized in MD simulations,¹ and in aggregate through distributed computing projects such as Folding@Home². Nevertheless, despite all the major advances, computational resources are still limited in what they can achieve in standard MD simulations with explicit solvent molecules on even modestly sized molecular systems, due to the complexity of their conformational dynamics. This has led to the development of more efficient ways to extract the thermodynamic properties of the system.

The use of enhanced sampling methods has become commonplace when simulating proteins and biomolecules. Replica-exchange MD (REMD)^{3, 4} and simulated tempering^{5, 6} are some of the most popular modern methods used to cross high energy barriers, and to map the free energy landscape of biomolecular systems, available in most MD simulation packages. In practice, REMD simulations provide accurate estimates of the populations of conformational states of a molecular system. However, extracting quantitative kinetic information from REMD trajectories regarding the transitions between the various conformational states is generally more challenging.⁷⁻¹⁴ Previously proposed methods rely either on *a priori* assumptions on the functional dependence of the transition rates on temperature (e.g. Arrhenius-like^{15, 16}), or on more complex statistical analysis of transition paths¹⁷ or algorithms using likelihood maximization and multi-dimensional optimization methods.^{7, 18}

Here, we use the direct transition counting (DTC) of Refs. 11 and 12 in its simplest form as a method for calculating transition rates from REMD trajectories that is easier to implement and leads to similarly accurate temperature-dependent rates as compared to the alternative, more complex and indirect methods.^{7, 18} We compare the DTC results to those of the maximum likelihood propagator based (MLPB)

method. We apply both methods (i.e., DTC and MLPB) to all-atom REMD simulations of dimerization of computationally and experimentally-relevant amyloid-forming NNQQ tetrapeptides,^{8, 19} in explicit water – one of the smallest two-state-like systems featuring peptide-peptide interactions that is, nevertheless, challenging to analyze systematically using REMD.¹⁸ We validate the rates extracted using the DTC method both by comparison to the corresponding MLPB rates and by analyzing the decay rates of the state autocorrelation functions of the system.²⁰⁻²² We assess the corresponding REMD efficiency,²³ and we obtain remarkably good agreement with the theoretically predicted errors in estimating the dimer and dissociated populations.

II. THEORETICAL METHODS

a. DTC Method

The following is a derivation of the DTC equations directly from a short time expansion of kinetic rate equations. In Ref. 12, it was shown that the resulting DTC estimators for the rate coefficients are statistically optimal (i.e., being the maximum likelihood solutions for a rate system satisfying detailed balance).

We assume that the conformational space of a system can be discretized into N distinct states that obey a master equation, $dp_m(t)/dt = \sum_{n=1}^N [k_{nm}p_n(t) - k_{mn}p_m(t)]$, where p_m is the probability of being in state m at time t and k_{nm} is the rate of transition from state m to state n .^{21, 24-34} In matrix notation, this is written as $\frac{d\mathbf{p}(t)}{dt} = \mathbf{K}(t)\mathbf{p}(t)$, where $\mathbf{K}(t)$ is the $N \times N$ rate matrix and $\mathbf{p}(t)$ is the time dependent column vector of probabilities with elements such that $p_n(t) > 0, n \in \{1, \dots, N\}$. At equilibrium, $\mathbf{K}\mathbf{p}^o = 0$ and the first right eigenvector of \mathbf{K} (corresponding to the first eigenvalue $\lambda_1 = 0$) is therefore given by \mathbf{p}^o (i.e., the vector of equilibrium populations that has positive elements, $p_n^o > 0, n \in \{1, \dots, N\}$),

and it is normalized according to the relation $\sum_{n=1}^N p_n^o = 1$).

Central to the DTC method^{11, 12} for estimating rates is the assignment of conformational states of the system. Here, the states were assigned by following each replica at different temperatures, using a transition based assignment (TBA) method described and used in previous studies.^{18, 21, 22} Our analysis of the NNQQ dimer conformational dynamics follows the approach presented in Ref. 18, and is summarized also in Fig. S2. Briefly, two specific interatomic distances are used as initial reaction coordinates, as shown in Fig. S1, as they allow a good discrimination between the different peptide-peptide interaction modes. While these reaction coordinates need to be reasonably good, the subsequent state assignment step does not depend on their absolute quality, as the TBA method adds more specific information from analyzing the transition paths (i.e., time sequence of transition events) to the state assignment process.

The method to determine the rate matrix \mathbf{K} is as follows (see Fig. S2). Consider an REMD simulation of the system of interest with M replicas, for t_{REMD} simulation time for each replica. The atomistic trajectories for each replica $i \in \{1, \dots, M\}$ are simplified by using the TBA method to project them onto states $s_i(t) \in \{1, \dots, N\}$ at times $t \in [0, t_{\text{REMD}}]$. For deriving the rates of the corresponding master equation, the intrinsic system dynamics in each of the s_i states is assumed to be Markovian, an assumption that we test subsequently.^{18, 21} Here, we also define the temperature, $\theta_i(t)$, at which the system is at time t for replica i , with values in the discrete interval $\theta_i(t) \in \{T_1, \dots, T_M\}$. We can now count the number of transitions, C_{nm}^i , from state m to state n from all replicas, for each temperature T_i with $i \in \{1, \dots, M\}$ using

$$C_{nm}^i = \sum_{j=1}^M \sum_{q=0}^{t_{\text{REMD}}/\Delta t - 1} \delta[s_j(q\Delta t), m] \delta[s_j(q\Delta t + \Delta t), n] \delta[\theta_j(q\Delta t), T_i],$$

where Δt is the frequency at which the coordinates of the system are saved and the states are assigned along the replica trajectories $s_i(t)$. $\delta(a,b)$ is one if $a=b$ and zero otherwise. We assume that Δt is small enough such that the number of additional transitions (including at different replica temperatures) occurring within the Δt time interval is negligible. This approximation corresponds to truncating the Taylor expansion of the matrix exponential at the linear term: $\tilde{\mathbf{C}}(\Delta t) = e^{\mathbf{K}\Delta t} = \mathbf{I} + \mathbf{K}\Delta t + O(\Delta t^2)$, where $\tilde{\mathbf{C}}(\Delta t)$ is the column-normalized transition probability matrix (Markov matrix) with lagtime Δt . The truncation error can thus be assessed analytically and, as shown below, for typical MD simulations it is negligible. We also assume that a replica m remains at the same temperature $T_i = \theta_i(q\Delta t) = \theta_i(q\Delta t + \Delta t)$ after a transition (i.e., a replica does not change states during an exchange event). This is a very good assumption for REMD simulations where trajectories are saved frequently, at an interval Δt that is typically smaller than the exchange attempt frequency. After assigning states (e.g., as done here by using the TBA method), the transition times cannot be resolved within the interval Δt between two data points along the trajectory.

Finally, using the number of transitions \mathbf{C} , the rate matrix $\mathbf{K}^{(i)}$ at temperature T_i is determined by first symmetrizing the unnormalized transition matrix to enforce detailed balance $C_{nm}^{sym,i} = (C_{nm}^i + C_{mn}^i)/2$, and then by using

$$[2] \quad K_{nm}^{(i)} = \frac{C_{nm}^{sym,i}}{P_m^{(i)} \Delta t} \text{ for } m \neq n, \text{ and}$$

$$[3] \quad K_{mm}^{(i)} = - \sum_{n=1(n \neq m)}^N K_{nm}^{(i)},$$

where $P_m^{(i)}$ is the unnormalized equilibrium probability of state m , calculated by simply counting the number of times the REMD trajectory is in each particular state m at temperature T_i

$$[4] \quad P_m^{(i)} = \sum_{j=1}^M \sum_{q=0}^{t_{REMD}/\Delta t} \delta[s_j(q\Delta t), m] \delta[\theta_j(q\Delta t), T_i] = \sum_{n=1}^N C_{nm}^i.$$

Note that $P_m^{(i)}$ can also be estimated using the symmetrized matrix, $P_m^{(i)} = \sum_{n=1 (n \neq m)}^N C_{nm}^{sym}$, which in practice

leads to very small differences as compared to using Eq. [4]. Eq. [2] for the rates was obtained by Stelzl and Hummer¹² using likelihood maximization. Symmetrized transition counts in Eq. [2] were obtained as a direct consequence of detailed balance. Note that Ref. 12 also considers the simultaneous estimation of equilibrium populations and rates in the kinetic model, and extends DTC to cases with transition paths that could be longer than the interval between replica exchanges by introducing fractional transition counts.

We apply the DTC method to study the association and dissociation of a dimer system of two NNQQ monomers (Fig. 1) and compare the resulting rates to the ones calculated using an alternative, more complex indirect MLPB method.^{18, 21} The MLPB method uses Green's functions to express the likelihood of a trajectory between Markov states. The conditional probability $G(n, \Delta t | m, 0)$ for being in state n after a lagtime Δt having been in state m at time $t_0 = 0$ is related to the rate matrix \mathbf{K} using

$$[5] \quad G(n, \Delta t | m, 0) = \left[e^{\mathbf{K} \cdot \Delta t} \right]_{nm}.$$

The likelihood of a Markovian trajectory, Λ is calculated using

$$[6] \quad \Lambda = \prod_{n=1}^N \prod_{m=1}^N \left[G(n, \Delta t | m, 0) \right]^{C_{nm}(\Delta t)},$$

where $C_{nm}(\Delta t)$ is the transition matrix corresponding to the lagtime Δt , as defined earlier (omitting the index i for the REMD temperatures), and N is the total number of states. In the MLPB method, the elements of the rate matrix \mathbf{K} are found by using a multi-dimensional stochastic search (i.e., simulated annealing using a Metropolis Monte Carlo algorithm as described in Ref. 21) that uses the minimization of the $-\log \Lambda$ as the optimization function.^{18, 21}

b. REMD Simulations

Our REMD simulations of NNQQ dimers were performed as described in Ref. 18, with the MD package Gromacs^{35, 36}, using Langevin dynamics with a friction coefficient of 0.1 ps^{-1} ,³⁷ an integration time step of 2 fs, Berendsen pressure coupling,³⁸ and a particle-mesh Ewald method with a switching distance for nonbonded electrostatics and van der Waals interactions at 8.5 Å and a cutoff distance of 10 Å. The simulations were in the NPT ensemble with the Amber 99sb force field³⁹ and explicit TIP3P⁴⁰ water molecules. The simulation box side was 40 Å, and contained 6525 atoms in total, including 2132 water molecules. To enhance the sampling, REMD is used with 16 replicas running at temperatures spaced according to an optimized protocol⁴¹ in the range of 310.00 K to 369.08 K.¹⁸

Coordinates were saved every 1 ps, and REMD exchanges were also attempted every 1 ps, with an average acceptance probability of ~30%. Attempting an exchange as often as possible has been found to enhance the sampling even further.^{23, 42} Five initial conditions were run, each starting from a potential microcrystal structure as suggested by X-ray micro-crystallography experiments by Sawaya et al.¹⁹ The five initial conditions were simulated for 164 ns for each replica, giving a total REMD running time of 820 ns and thus a total MD simulation time of 13.12 μs . As shown in Fig. S3 (and also in Figs. S2 and S3 from Ref. 18), this is more than twice the amount of data needed for convergence of the relevant kinetic quantities.

As an additional test for convergence we also investigated the “equal occupancy” of replicas at each temperature,⁴³ as shown in Fig. S7 of Ref. 18) which is a useful method for assessing the performance of parallel tempering simulations.^{43, 44}

The trajectories were then analyzed using the workflow indicated in Fig. S2 and the corresponding rate matrices were extracted and compared.

Measuring the decay of the autocorrelation function is another method to extract the slowest relaxation time of the system. For a two-state model with states denoted as *A* and *B*, the overall relaxation

rate of an REMD simulation with M replicas can be given as the weighted sum of each of the relaxation rates of the per temperature data and weighted by the normalized product of the probabilities²³

$$[7] \quad \kappa = \frac{1}{\tau_{relax}} = \frac{\sum_{i=1}^M \lambda_i (1 - p_A^i) p_A^i}{\sum_{i=1}^M (1 - p_A^i) p_A^i} .$$

where p_A^i is the population of state A at temperature T_i , and λ_i is the relaxation rate of the system at temperature T_i . Therefore, κ is the overall relaxation rate of the entire REMD simulation, a weighted average of the relaxation rates at each temperature. This expression was derived in the fast exchange limit building a coarse-grained kinetic network model using the local equilibrium approximation. For this model, the solution for the relaxation rate and the populations of the total number of states in A (or B) is mathematically equivalent with the one-dimensional diffusive harmonic oscillator model for large number of replicas.²³

The normalized folding state correlation function $c(t)$ is given at long times by

$$[8] \quad c_i(t) \equiv \frac{\langle \Delta s(t) \Delta s(0) \rangle}{\langle \Delta s^2 \rangle} \approx \frac{(1 - p_A^i) p_A^i e^{-\kappa t}}{\sum_{i=1}^M (1 - p_A^i) p_A^i} ,$$

in the limit of fast exchange and for large M . Therefore, the slope of the natural logarithm of the autocorrelation function allows us to estimate the relaxation rate κ .

A similar formula was derived for simulated tempering (ST) simulations.⁴⁵ A converged ST trajectory is essentially equivalent with a replica trajectory in an REMD simulation. Therefore, the following formula can be applied to the per replica data of the REMD simulations

$$[9] \quad \lambda_{replica}^{eff} = \frac{\sum_{i=1}^M k_{AB}^i p_B^i}{\sum_{i=1}^M p_B^i} + \frac{\sum_{i=1}^M k_{BA}^i p_A^i}{\sum_{i=1}^M p_A^i} ,$$

where k_{AB}^i is the rate for transitions from B to A at temperature T_i . Eq. [9] can be rewritten in a similar way to Eq. [7] to give

$$[10] \quad \lambda_{\text{replica}}^{\text{eff}} = \frac{N \sum_{i=1}^N \lambda^i k_A^i p_A^i (1 - p_A^i)}{\sum_{i=1}^N p_A^i \sum_{i=1}^N (1 - p_A^i)}.$$

Thus, the effective relaxation rate for the per replica R-trajectories of the system is given by the weighted sum of the relaxation rates of each of the per replica trajectories, weighted by the product of the folding and unfolding probabilities and normalized by the product of the sums of the folding and unfolding probabilities, and multiplied by the number N of replicas. The differences between Eqs. [7] and [10] are the multiplicative factor of N , and that the normalization is the sum of the products and the product of the sums of the probabilities, respectively.

The efficiency η_k at temperature T_k of an REMD simulation is given by

$$[11] \quad \eta_k = \frac{\sum_{i=1}^N k_{U_i} p_{B_i}}{N k_{U_k} p_{B_k}}.$$

Here indices U and B refer to unbound and bound states, respectively. This relation can be thought of as the ratio of total number of transitions per unit time averaged over the N replicas compared to the number of transitions that occur at the temperature of interest T_k (usually the lowest temperature). For all values of η_k greater than 1 it is more efficient to run a REMD simulation as opposed to a standard MD simulation at that temperature (i.e., using the same total computer time).

The variance $\sigma_{REMD}^2(t_{sim})$ in the population estimates of the (un)bound populations in an REMD simulation at temperature T_k is given by²³

$$\sigma_{REMD}^2(t_{sim}) = \frac{2}{t_{sim}} \frac{p_k^2 q_k^2}{\sum_{i=1}^N p_i q_i \lambda_i},$$

where t_{sim} is the total simulation time (for each replica), and the unbound fractions are simply $q_i = 1 - p_i$.

III. DISCUSSION

Here, following the analysis in Ref. 18, the 7×7 rate matrix extracted for the NNQQ dimeric system is coarse grained to a two-state system with the associated dimer (states 1 to 6) as one macrostate and the dissociated state (state 7) the second macrostate.¹⁸ The equilibrium populations shown in Fig. 2(a) are extracted by counting the amount of time spent in each state. We see a linear increase in the population of the dissociated state with temperature. This behavior is consistent with the expectation that the unbound state has higher entropy and enthalpy.

In the following, we analyze the rate matrices and compare them to the results found by the MLPB method. In a spectral analysis, the slowest relaxation times from the two methods agree well, as seen in Fig. 2(b). The profile of the relaxation rate is relatively flat across the temperatures. This can be understood when the individual k_{off} and k_{on} rates (i.e., for dimer dissociation and association, respectively) are plotted in Figs. 3(a) and 3(b) respectively. Again the match is remarkably good. The rate of dissociation k_{off} does not depend on the concentration of the system. A clear trend is observed showing an increase in k_{off} as the temperature increases, in keeping with an Arrhenius-like rate. By contrast, k_{on} depends only weakly on temperature, with increased diffusion at higher temperatures compensated by an increased simulation box volume. The relaxation rate is proportional to $k_{off} + k_{on}$. Because k_{on} is an order of magnitude larger than k_{off} , the relaxation rate is nearly independent of temperature.

Overall the DTC method^{11, 12} is much easier to implement than the MLPB method. Excellent agreement between the two methods is found. We thus recommend using DTC^{11, 12} as a simple and more direct way of extracting rates for an REMD system in cases where transitions between states can be

olved and are fast.

The decay of the state autocorrelation function is another method to estimate the slowest relaxation rate of a system, i.e., by estimating the inverse of the slope of the autocorrelation curve (e.g., see Fig. S4). Fig. 4(a) shows the slowest relaxation times (τ_2) obtained from fits of the autocorrelation function calculated “per temperature” (red circles, ACF T) T-trajectories (see ¹⁸), and “per replica” R-trajectories (blue crosses, ACF R). Values obtained from formulas given above (Eqs. [9] and [10])²³ for the effective relaxation rate of the REMD simulation are also shown (marked “eff remd”, and “eff rep”). The average values for ACF T and ACF R are added for reference as dashed lines. The effective relaxation rates can be seen to approximate very well the intrinsic timescales of the system. Interestingly, they are not only very close to each other, but fall definitely within the upper and lower bounds of the values extracted from the fitted autocorrelation functions (red and blue lines).

In Fig. 4(b) we show the computational efficiency (η) of running a REMD simulation with M replicas (as compared to running an equivalent, M times longer, MD simulation at the target temperature). We see that it is approximately 2.2 times more efficient to run the REMD simulation with M replicas of the system at the lowest (i.e., our “target”) temperature of 310 K (Fig. 4(a)). We see a turnover point at 340.49 K beyond which it becomes less efficient to run an REMD simulation using the current temperature range, as molecular transition rates generally increase with temperature. The variance σ^2 of the equilibrium populations p_i^o (see Fig. 2(a)), can be calculated as described in Ref. 23 and illustrated in Ref. 11.

In actual implementations of REMD calculations, one needs to choose a value for the time interval, Δt_{REMD} , at which replica exchanges are attempted. In order to gain a more quantitative insight, and to examine the convergence of the 7-state rates calculated via the DTC method, a series of replica exchange Monte Carlo (REMC) simulations^{46,47} were run for different Δt_{REMD} intervals. The Markov matrices used to propagate the system in the MC steps were analytically calculated using the (“exact”) rates obtained

From the REMD simulations at each temperature ($\tilde{\mathbf{C}}(\Delta t) = e^{\mathbf{K}\Delta t}$). From the resultant REMC simulation data, the populations were calculated from the simulation trajectories, and were compared with the exact populations using the Kullback-Leibler (KL) divergence⁴⁸ measure

$$[12] \quad KL(t) = \sum_i P_i^{MC}(t) \log\left(\frac{P_i^{MC}(t)}{P_i^{exact}}\right).$$

Figure 5(a) shows the convergence of the 7-state populations captured by the KL measure as a function of the simulation length.

The DTC method was then used to calculate the corresponding rates from the REMC trajectories. In Fig. 5(b), the REMC rates are compared to the input, exact (REMD) rates by using a root mean square deviation (RMSD) measure defined as

$$[13] \quad \epsilon(t) = \sqrt{\frac{\sum_{i,j,i \neq j} (K_{ij}^{MC}(t) - K_{ij}^{exact})^2}{N(N-1)}}.$$

The KL divergence can in theory tend to zero in the long time limit, whereas the rates are limited by our step size Δt used in constructing the Markov matrices for REMC. In the DTC method, it is assumed that the higher order terms can be neglected, thus, for REMC simulations using the normalized counts, $\tilde{\mathbf{C}}(\Delta t)$, the theoretically optimal rates can be estimated as

$$[14] \quad \mathbf{K} = (\tilde{\mathbf{C}} - \mathbf{I})/\Delta t.$$

As shown in Fig. 5, the KL divergence of the populations converges towards zero for sufficiently long simulation times (Fig. 5a), and the corresponding error in calculating the rates from REMC data converges towards its theoretical minimum (Fig. 5b). In general, this convergence is faster for more frequent replica exchange attempts, an observation that agrees with previous studies on the error of equilibrium populations,^{23, 42} and shown here as well for the calculated kinetic rates.

CONCLUSIONS

In summary, we found that the DTC method^{11, 12} for extracting rates directly from REMD simulations is simple to implement, in comparison to previously proposed methods, which either use a more complex maximum likelihood approach,^{18, 21} statistical analysis of transition paths,¹⁷ or rely on *a priori* assumptions about the functional form (e.g. Arrhenius-like,^{15, 16}) of the temperature-dependence of the underlying transition rates. REMD simulations are increasingly popular options for achieving enhanced sampling, yet extracting routinely quantitative kinetic information from REMD trajectories regarding the transitions between the various conformational states is generally challenging.^{7-10, 12-14}

In the simple implementation used here to study peptide binding-unbinding, the DTC method^{11, 12} proceeds in two major steps, requiring only the ability to assign Markovian states to REMD trajectories. In a first step (as illustrated by the workflow in Fig. S2), we use the continuous R-trajectories corresponding to system replicas evolving at the various temperatures to assign conformational states, using the trajectory-based state assignment (TBA) method introduced earlier.^{18, 21} In a second and final step, we analyze the entire REMD data to calculate the corresponding rates at each temperature, both directly, from the number of transition counts,^{11, 12} and also indirectly, from short-time propagators (using a maximum likelihood approach as in Refs. 18, 21) or from state correlation functions.^{20, 22}

Here, the DTC method^{11, 12} was applied to dimer formation of NNQQ peptides. We obtained excellent agreement between the rates extracted using the DTC method and our previous, more elaborate maximum likelihood-based method. We also tested theoretical predictions for the slowest relaxation time of the system by monitoring the decay of the autocorrelation function both in per temperature REMD space (i.e., using T-trajectories), and in the per replica space (i.e., using R-trajectories). We assessed the corresponding REMD efficiency, and we showed using REMC simulations that more frequent exchanges

increased the error in determining the kinetic rates, in addition to leading to more accurate populations of the NNQQ dimers.

The DTC method^{11, 12} should be useful in the increasingly broad range of replica exchange studies where there is a need for accurate calculations of transition rates between states, besides the more typical calculation of equilibrium populations and the associated free energy differences. For example, together with Hamiltonian replica exchange,⁴⁹⁻⁵³ simulated tempering,^{45, 54-56} lambda dynamics,⁵⁷ generalized ensemble sampling,^{58, 59} or with other recent enhanced sampling methods,⁶⁰⁻⁶⁴ DTC may be used to extract more complete and accurate kinetic and thermodynamic data (*e.g.*, possibly in conjunction with the DHAM method⁶⁵ for relating biased to unbiased transition counts). In addition, the DTC method can also be combined with analysis methods that take advantage of additional information such as the automatic identification of Markovian transition states⁶⁶ from MD trajectories.

ABBREVIATIONS

DTC = direct transition counting; MD = molecular dynamics; REMD = replica-exchange molecular dynamics; TBA = transition-based assignment; MLPB = maximum likelihood propagator-based; DHAM = dynamic histogram analysis method

SUPPLEMENTARY MATERIAL

See **supplementary material** for the four supplementary figures and their captions, Figs. S1 to S4, mentioned in the main text.

ACKNOWLEDGMENTS

We are grateful to Dr. Attila Szabo from the National Institutes of Health for helpful and stimulating discussions. G.H. thanks Dr. Lukas Stelzl for insightful discussions. We wish to thank the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC), and the Biowulf cluster at the National Institutes of Health, United States (<http://biowulf.nih.gov>) for the provision of computational facilities and support. We gratefully acknowledge financial support from the Irish Research Council (IRC) for C.T.L and N.V.B., and the EPSRC (grant number EP/N020669/1) for E.R.

FIGURE CAPTIONS

Figure 1. (a) Dimer formation of two tetrameric NNQQ amyloid peptides with association rate k_{on} and dissociation rate k_{off} . **(b)** Schematic illustration of a replica R-trajectory (black line) that visits several temperatures during an REMD simulation, while transitioning for example between two states S1 and S2. The red line is the corresponding state R-trajectory obtained using a state-assignment procedure (e.g., the TBA method, see text). An attempt to exchange temperatures is made every Δt_{REMD} and is either accepted (marked as “A”) or rejected (“R”). Note that R-trajectories are continuous, even though they visit various temperatures, while the per-temperature T-trajectories would be interrupted at times when exchange attempts are accepted.

Figure 2. (a) Equilibrium population p^o of the dissociated state as a function of temperature ($T_0 = 310.00$ K, $T_{15} = 369.08$ K). **(b)** Slowest relaxation time τ of the system, as a function of the temperature, extracted for each temperature T using the MLPB method (blue, x marks) and the direct transition counting (red circles) method. Error bars report the standard error of the mean.

Figure 3. Temperature dependence of the NNQQ dimer rates of **(a)** dissociation k_{off} , and **(b)** association k_{on} estimated from the REMD T-trajectories using MLPB method (blue, x marks) and the direct transition counting method (red circles). Note that no a priori assumption on the functional form of the T-dependence of the rates (e.g., Arrhenius-like or not) was necessary.

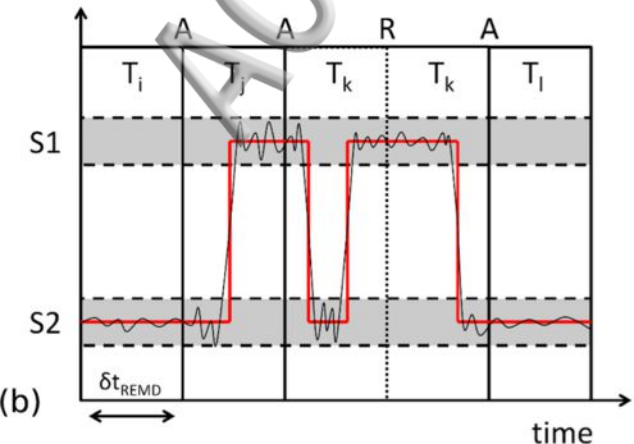
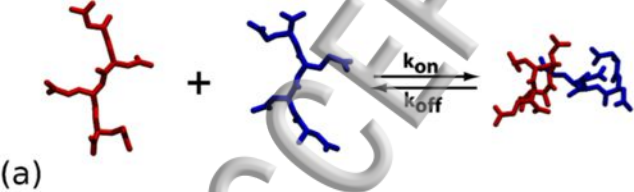
Figure 4. (a) Slowest relaxation time τ , calculated by several different methods from the NNQQ REMD simulation data using the state autocorrelation function for each temperature (red circles, “ACF T”), and for each replica (blue x marks, “ACF R”). The effective relaxation rates of the system (“eff. remd” and “eff. rep”) are calculated using analytical formulas (see text). The average values of “ACF T” and “ACF R” (denoted as “avg. ACF T”, red dashes, and “avg. ACF R”, blue dashes, respectively) are also shown for comparison. The rates extracted directly from the “per replica” R-trajectory rate matrices are also shown (marked as “rates R”, green line). **(b)** Relative computational efficiency, η , of the REMD simulation at each temperature. Note that $\eta > 1$ at a certain temperature T_i implies that REMD is more efficient than the corresponding standard MD (i.e., for N -times longer simulations, where N is the number of replicas) at that temperature.

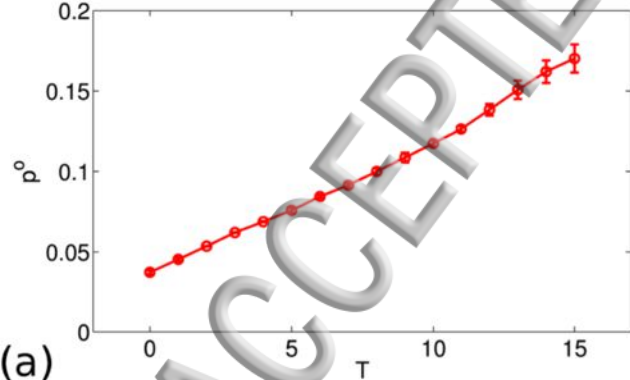
Figure 5. (a) The KL divergence of the REMC calculated populations as a function of simulation time, at replica exchange attempt time intervals of 10 ps, 100 ps, 1 ns, and 10 ns. **(b)** The error estimated using Eq. [14] for the REMC simulation rates with respect to the exact rates as a function of simulation time, for varying replica exchange attempt intervals. The inset shows the same data on a log-log scale.

REFERENCES

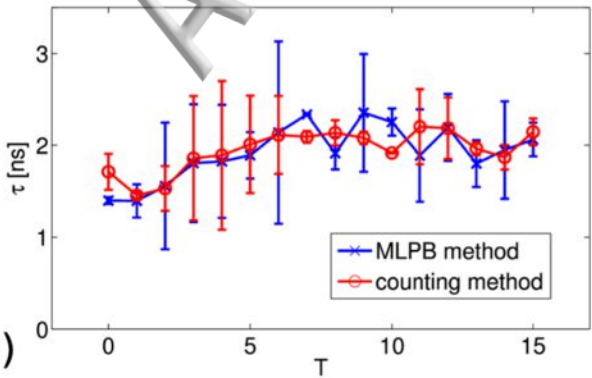
- ¹ K. Lindorff-Larsen *et al.*, *Science* **334** (2011) 517.
- ² M. Shirts, and V. S. Pande, *Science* **290** (2000) 1903.
- ³ Y. Sugita, and Y. Okamoto, *Chem. Phys. Lett.* **314** (1999) 141.
- ⁴ A. E. Garcia, and K. Y. Sanbonmatsu, *Proc. Natl. Acad. Sci. U. S. A.* **99** (2002) 2782.
- ⁵ A. P. Lyubartsev *et al.*, *J. Chem. Phys.* **96** (1992) 1776.
- ⁶ E. Marinari, and G. Parisi, *Europhys. Lett.* **19** (1992) 451.
- ⁷ N. V. Buchete, and G. Hummer, *Phys. Rev. E* **77** (2008) 4.
- ⁸ B. Strodel, C. S. Whittleston, and D. J. Wales, *J. Am. Chem. Soc.* **129** (2007) 16005.
- ⁹ N. V. Buchete, *Biophysical journal* **103** (2012) 1411.
- ¹⁰ A. T. Frank, and I. Andricioaei, *The Journal of Physical Chemistry B* **120** (2016) 8600.
- ¹¹ C. T. Leahy, in *PhD Thesis* (University College Dublin, School of Physics, Belfield, Dublin 4, Ireland, 2013), p. 107.
- ¹² L. S. Stelzl, and G. Hummer, *Journal of Chemical Theory and Computation* **13** (2017) 3927.
- ¹³ M. Carballo-Pacheco, and B. Strodel, *Protein Science* **26** (2017) 174.
- ¹⁴ D. De Sancho, and R. B. Best, *J. Am. Chem. Soc.* **133** (2011) 6809.
- ¹⁵ D. van der Spoel, and M. M. Seibert, *Physical Review Letters* **96** (2006) 238102.
- ¹⁶ S. Piana, K. Lindorff-Larsen, and D. E. Shaw, *Proc Natl Acad Sci U S A* **109** (2012) 17845.
- ¹⁷ F. Zhu, *The Journal of Chemical Physics* **146** (2017) 124128.
- ¹⁸ C. T. Leahy *et al.*, *The Journal of Physical Chemistry Letters* **7** (2016) 2676.
- ¹⁹ M. R. Sawaya *et al.*, *Nature* **447** (2007) 453.
- ²⁰ D. J. Bicout, and A. Szabo, *Protein Science : A Publication of the Protein Society* **9** (2000) 452.
- ²¹ N. V. Buchete, and G. Hummer, *J. Phys. Chem. B* **112** (2008) 6057.
- ²² G. S. Buchner *et al.*, *BBA-Proteins Proteomics* **1814** (2011) 1001.
- ²³ E. Rosta, and G. Hummer, *J. Chem. Phys.* **131** (2009) 12.
- ²⁴ R. Zwanzig, *Journal of Statistical Physics* **30** (1983) 255.
- ²⁵ C. Schütte *et al.*, *Journal of Computational Physics* **151** (1999) 146.
- ²⁶ B. L. De Groot *et al.*, *Journal of Molecular Biology* **309** (2001) 299.
- ²⁷ Y. Levy, J. Jortner, and R. S. Berry, *Physical Chemistry Chemical Physics* **4** (2002) 5052.
- ²⁸ W. C. Swope *et al.*, *J. Phys. Chem. B* **108** (2004) 6582.
- ²⁹ D. S. Chekmarev, T. Ishida, and R. M. Levy, *J. Phys. Chem. B* **108** (2004) 19487.
- ³⁰ S. Sriraman, I. G. Kevrekidis, and G. Hummer, *The journal of physical chemistry. B* **109** (2005) 6479.
- ³¹ J. D. Chodera *et al.*, *Multiscale Model. Simul.* **5** (2006) 1214.
- ³² W. Zheng *et al.*, *J. Phys. Chem. B* **113** (2009) 11702.
- ³³ A. M. Berezhkovskii, F. Tofoleanu, and N.-V. Buchete, *Journal of Chemical Theory and Computation* **7** (2011) 2370.
- ³⁴ A. M. Berezhkovskii, R. D. Murphy, and N.-V. Buchete, *The Journal of Chemical Physics* **138** (2013) 036101.
- ³⁵ D. Van der Spoel *et al.*, *J. Comput. Chem.* **26** (2005) 1701.
- ³⁶ B. Hess *et al.*, *Journal of Chemical Theory and Computation* **4** (2008) 435.
- ³⁷ R. W. Pastor, B. R. Brooks, and A. Szabo, *Molecular Physics* **65** (1988) 1409.
- ³⁸ H. J. C. Berendsen *et al.*, *J. Chem. Phys.* **81** (1984) 3684.
- ³⁹ V. Hornak *et al.*, *Proteins* **65** (2006) 712.
- ⁴⁰ W. L. Jorgensen *et al.*, *J. Chem. Phys.* **79** (1983) 926.
- ⁴¹ A. Patriksson, and D. van der Spoel, *Physical Chemistry Chemical Physics* **10** (2008) 2073.

- ⁴²g. J. Sindhikara, D. J. Emerson, and A. E. Roitberg, *Journal of Chemical Theory and Computation* **6** (2010) 2804.
- ⁴³ J. D. Doll, and P. Dupuis, *The Journal of Chemical Physics* **142** (2015) 024111.
- ⁴⁴ J. D. Doll *et al.*, *The Journal of Chemical Physics* **137** (2012) 204112.
- ⁴⁵ E. Rosta, and G. Hummer, *J. Chem. Phys.* **132** (2010) 9.
- ⁴⁶ D. J. Earl, and M. W. Deem, *Physical Chemistry Chemical Physics* **7** (2005) 3910.
- ⁴⁷ R. H. Swendsen, and J.-S. Wang, *Physical Review Letters* **57** (1986) 2607.
- ⁴⁸ S. Kullback, and R. A. Leibler, *Ann. Math. Statist.* **22** (1951) 79.
- ⁴⁹ Y. Sugita, A. Kitao, and Y. Okamoto, *The Journal of Chemical Physics* **113** (2000) 6042.
- ⁵⁰ H. Fukunishi, O. Watanabe, and S. Takada, *The Journal of Chemical Physics* **116** (2002) 9058.
- ⁵¹ Y. Okamoto, *Journal of Molecular Graphics and Modelling* **22** (2004) 425.
- ⁵² J. D. Faraldo-Gómez, and B. Roux, *J. Comput. Chem.* **28** (2007) 1634.
- ⁵³ E. Rosta *et al.*, *J. Am. Chem. Soc.* **133** (2011) 8934.
- ⁵⁴ A. P. Lyubartsev *et al.*, *The Journal of Chemical Physics* **96** (1992) 1776.
- ⁵⁵ E. Marinari, and G. Parisi, *EPL (Europhysics Letters)* **19** (1992) 451.
- ⁵⁶ T. Nagai *et al.*, *J. Comput. Chem.* **37** (2016) 2017.
- ⁵⁷ S. Banba, Z. Guo, and C. L. Brooks, *The Journal of Physical Chemistry B* **104** (2000) 6903.
- ⁵⁸ C. Lu *et al.*, *Journal of Chemical Theory and Computation* **12** (2016) 41.
- ⁵⁹ D. Wu *et al.*, *Methods in Enzymology* **577** (2016) 57.
- ⁶⁰ X. Wu, and B. R. Brooks, *Chem. Phys. Lett.* **381** (2003) 512.
- ⁶¹ Q. Lu *et al.*, *The Journal of Chemical Physics* **141** (2014) 18C525.
- ⁶² A. Dickson, L. S. Ahlstrom, and C. L. Brooks, *J. Comput. Chem.* **37** (2016) 587.
- ⁶³ J. Xia *et al.*, *J. Comput. Chem.* **37** (2016) n/a.
- ⁶⁴ H. Jung, K.-i. Okazaki, and G. Hummer, *The Journal of Chemical Physics* **147** (2017) 152716.
- ⁶⁵ E. Rosta, and G. Hummer, *Journal of Chemical Theory and Computation* **11** (2015) 276.
- ⁶⁶ L. Martini *et al.*, *Physical Review X* (2017)

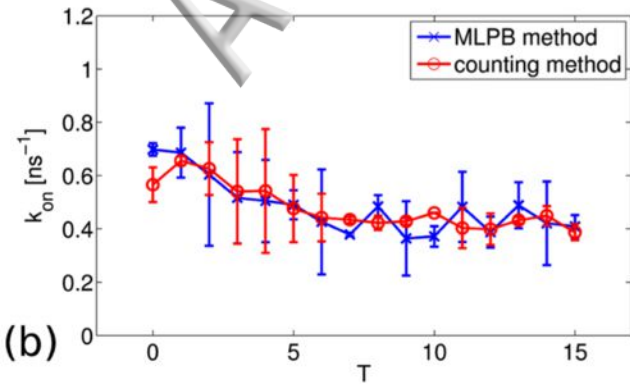
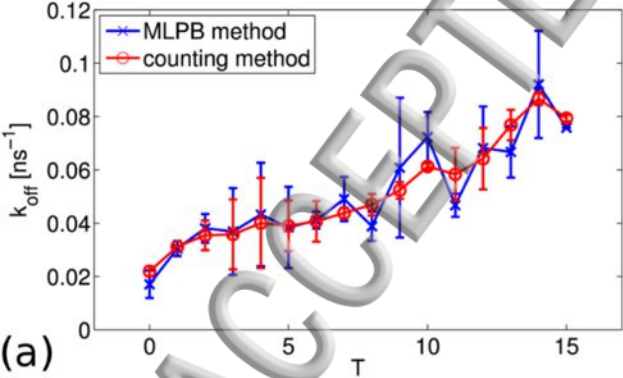


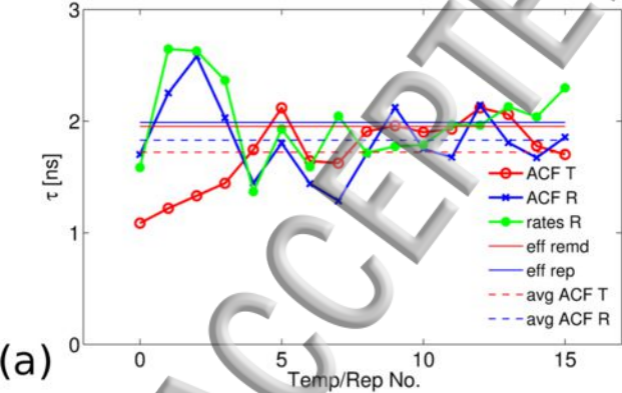


(a)

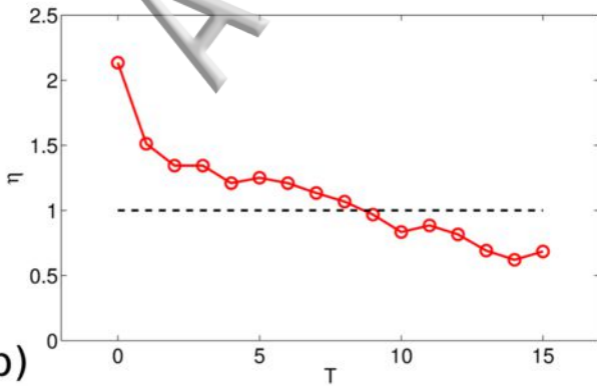


(b)

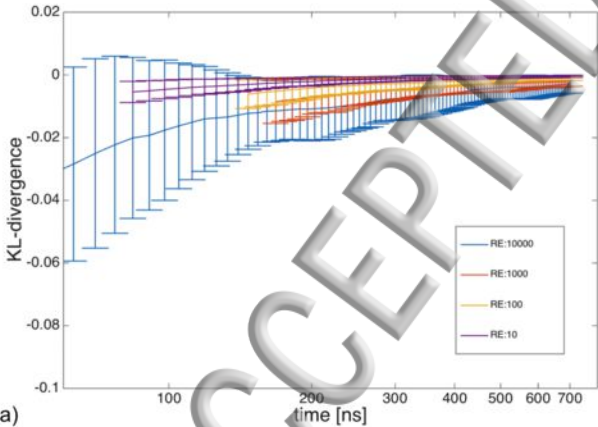




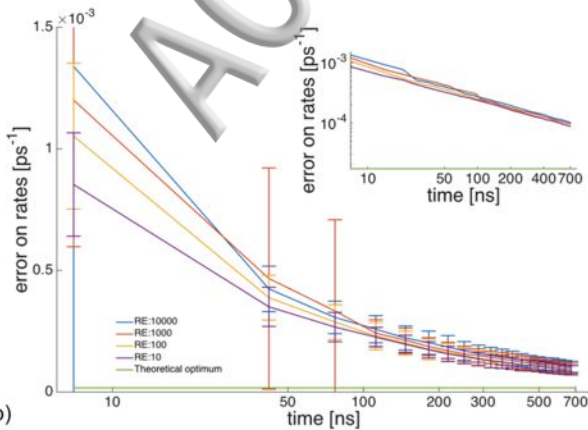
(a)



(b)



(a)



(b)