Optimal use of data in parallel tempering simulations for the construction of discrete-state Markov models of biomolecular dynamics

Jan-Hendrik Prinz,[1,*] John D. Chodera,[2,†] Vijay S. Pande,[3,‡]
William C. Swope,[4,§] Jeremy C. Smith,[5,¶] and Frank Noé[6,**]

[1]*Institute for Scientific Computing (IWR), University of Heidelberg, Im Neuenheimer Feld 368, 69126 Heidelberg, Germany and DFG Research Center Matheon, FU Berlin, Arnimallee 6, 14195 Berlin, Germany*

[2]*Research Fellow, California Institute of Quantitative Biosciences (QB3), University of California, Berkeley, 260J Stanley Hall, Berkeley, California 94720, USA*

[3]*Department of Chemistry, Stanford University, Stanford, CA 94305*

[4]*IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120*

[5]*UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory P.O.Box 2008 Oak Ridge TN 37831-6164, USA*

[6]*DFG Research Center Matheon, FU Berlin, Arnimallee 6, 14195 Berlin, Germany*

(Dated: October 8, 2010)

Abstract

Parallel tempering (PT) molecular dynamics (MD) simulations have been extensively investigated as a means of efficient sampling of the configurations of biomolecular systems. Recent work has demonstrated how the short physical trajectories generated in PT simulations of biomolecules can be used to construct Markov models describing biomolecular dynamics at each simulated temperature. While this approach describes the temperature-dependent kinetics, it does not make optimal use of all available PT data, instead estimating the rates at a given temperature using only data from that temperature. This can be problematic, as some relevant transitions or states may not be sufficiently sampled at the temperature of interest, but might be readily sampled at nearby temperatures. Further, the comparison of temperature-dependent properties can suffer from the false assumption that data collected from different temperatures are uncorrelated. We propose here a strategy in which, by a simple modification of the parallel tempering protocol, the harvested trajectories can be reweighted, permitting data from all temperatures to contribute to the estimated kinetic model. The method reduces the statistical uncertainty in the kinetic model relative to the single temperature approach and provides estimates of transition probabilities even for transitions not observed at the temperature of interest. Further, the method allows the kinetics to be estimated at temperatures other than those at which simulations were run. We illustrate this method by applying it to generating a Markov model of the conformational dynamics of the solvated terminally-blocked alanine peptide.

*jan-hendrik.prinz@fu-berlin.de

†Corresponding author; jchodera@berkeley.edu

‡pande@stanford.edu

§swope@almaden.ibm.com

¶smithjc@ornl.gov

**noe@math.fu-berlin.de

I. INTRODUCTION

Biological macromolecules are not static structures, but are driven by thermal motion and interactions with their molecular environment, undergoing conformational fluctuations and changing conformational states. The characterization of the statistical conformational dynamics of biomolecules is essential to understanding how these molecules work as molecular machines.

Often, a separation of timescales of characteristic dynamical relaxation times gives rise to the existence of metastable conformational states, such that the biomolecule remains in any one of these states for a long time before making a rapid transition to another state. A wealth of experimental data now supports the existence of such states, including NMR [1–3], fluorescence emission [4, 5], energy transfer [6, 7], correlation spectroscopy [8, 9], and non-equilibrium perturbation experiments [5]. Developing a quantitative understanding of what gives rise to these conformational states and the interactions that govern their transitions will have a significant impact on our understanding of many biological processes, such as signaling events, enzyme regulation, allostery, and drug design for conformationally flexible molecules.

Sampling the underlying phase space by straightforward molecular dynamics simulation often suffers from the problem that the timescales of conformational changes are usually orders of magnitude larger than simulation times accessible using current computational resources. Parallel tempering molecular dynamics (PT) simulations has been an effective and thus popular approach to overcoming the issue of convergence in molecular simluations, by allowing replicas to heat up and overcome enthalpic barriers as the simulation proceeds while still sampling from an appropriate equilibrium distribution [10–13]. At the same time this approach permits an analysis of temperature dependence of properties of interest, which is especially important for comparisons with certain experimental results [14] (e.g. melting curves, heat capacities). Although parallel tempering molecular dynamics produces unphysical replica trajectories, the short physical trajectories in between the exchanges can provide useful dynamical information. If the parallel tempering simulation is well-equilibrated, these initial configurations of the short trajectory segments will be sampled from the equilibrium at their corresponding temperatures.

Markov models provide a way of modeling the slow conformational dynamics of biomolecules based on short simulations [15–23]. In these models conformational states are envisioned as disjoint but connected regions of configurational space. The biomolecule spends long times within individual regions before undergoing rapid stochastic transitions between regions. If a separation of timescales exists between fast relaxation times within and slow equilibration between regions, the inter-state dynamics can be well described by a Markov model in discrete timesteps τ, where a coarse graining in time is required as the discretization in space prohibits the characterization of relaxation processes faster than τ. If the system is partitioned into its metastable states, τ is related to the time required to overcome internal barriers within each conformational state. However, it has been recently shown that, even in the absence of many metastable states, a Markovian model can well approximate the dynamics at long times, with this approximation error decreasing with increasing number of states [24].

Recently, Buchete and Hummer have shown that both thermodynamic and kinetic properties can be estimated over the range of temperatures by constructing Markov models using the short physical trajectories generated from parallel tempering simulations[25]. However, if a complete description of dynamics across the entire configurational space at a given temperature is desired, one quickly runs into problems if use is made of trajectories only from the temperature of interest, as some states that are sampled at other temperatures may not be well sampled at the single temperature [25]. One would like to make use of the data collected at all temperatures to characterize the kinetic behavior in all regions sampled over the full range of temperatures spanned by the parallel tempering simulation in a manner similar to equilibrium reweighting [26–31].

Here, we propose a method for integrating MD data from all temperatures by making use of *dynamical reweighting* [32], allowing a smooth, continuous and differentiable estimate of the transition probabilities at any temperature without requiring the assumption of any kinetic model (such as Arrhenius kinetics [18]) and taking advantage of the increased transition rates at higher (or, for transitions with entropic barriers, lower) temperatures. Reweighting methods (such as histogram-based [26–29] or histogram-free [30, 31]) allow the use of samples collected from multiple distributions to provide an improved estimate of the expectation value of some static property at the distribution of interest, and have been used extensively in the analysis of equilibrium thermodynamic

properties in replica-exchange simulations [33].

Dynamical reweighting has recently been proposed as a way of estimating dynamical properties (such as correlation functions) using an asymptotically optimal estimator, and also provides an estimate of the statistical error [32]. Here, we show how dynamical reweighting can be used to estimate transition probabilities (and their statistical uncertainties) for the construction of a Markov model as a smooth function of temperature, making use of data from all temperatures. This has the advantage of producing a useful Markov model at *any* temperature containing the dependence of kinetic properties on temperature, and providing an assessment of the error in the model.

We illustrate this approach for the standard test case of the terminally-blocked alanine peptide in explicit solvent. A Markov model constructed from short (6 ps) trajectories from each state has been previously shown to accurately describe the kinetics of this system at 302 K [17]. This peptide system presents a challenge for estimators based on individual temperatures, due to the presence of highly metastable states with very high free energies relative to the most populated states. These states are poorly sampled at temperatures near 300 K, even though their temporal behavior can dominate the nonequilibrium relaxation kinetics at this temperature. Finally, we determine whether using all the data using reweighting produces substantially improved kinetic models at this particular temperature and across the full range of temperatures in the parallel tempering simulation.

This paper is organized as follows: In Section II we review the theory behind Markov models of multistate conformational dynamics. We then show in Section II D how dynamical reweighting can be used to estimate temperature-dependent transition probabilities and rates for a given state decomposition. Finally, we illustrate the method in Section III by applying it to a six-state decomposition of an MD simulation of the terminally-blocked alanine peptide, and compare the results to the approach of Buchete and Hummer [25], in which Bayesian estimates of transition probabilities are obtained from a single temperature alone.

II. THEORY

A. Markov Models

Consider a system that evolves according to some stationary dynamical process. Let Ω be the configuration space with a complete decomposition [22] into M disjoint sets $\Gamma_i \subset \Omega$ such that

$$\bigcup_{i=1}^{M} \Gamma_i = \Omega;\ \Gamma_i \cap \Gamma_j = \emptyset,\ \forall i, j \in \{1, \ldots, M\},\ i \neq j. \tag{1}$$

For convenience we also define indicator functions $\chi_i(\boldsymbol{q}) \in \{0, 1\}$ for points in configuration space $\boldsymbol{q} \in \Omega$ by setting

$$\chi_i(\boldsymbol{q}) = \begin{cases} 1 & \text{if } \boldsymbol{q} \in \Gamma_i \\ 0 & \text{else} \end{cases} \tag{2}$$

i.e., the function assumes the value of unity if $\boldsymbol{q}$ belongs to set Γ_i, and zero otherwise. Based on this discretization of state space, we can define a row-stochastic transition matrix $\mathbf{T}(\tau)$ with conditional probabilities of finding the system in state j at time τ after it was originally in state i:

$$T_{ij}(\tau) = \mathbb{P}\left(\boldsymbol{q}(\tau) \in \Omega_j \mid \boldsymbol{q}(0) \in \Omega_i\right) \tag{3}$$

$$= \frac{\langle \chi_i(0)\chi_j(\tau)\rangle}{\langle \chi_i \rangle}, \tag{4}$$

and introduce $\chi_i(t) \equiv \chi_i(\boldsymbol{q}(t))$, where the dynamics is assumed to be governed by a stationary (time-independent) process. We aim here to construct a discrete-time, discrete-space Markov model that approximates the long-time dynamics of the system by virtue of

$$\boldsymbol{p}(t + k\tau) \approx \boldsymbol{p}(t)\mathbf{T}^k(\tau) \tag{5}$$

with $\boldsymbol{p}$ being the projection of some continuous distribution $\rho(\boldsymbol{q})$ onto the discrete subsets Γ_i. Eq. 5 is only an approximation to the real dynamics due to the introduction of the coarse-graining $\Gamma \equiv \{\Gamma_1, \ldots, \Gamma_M\}$ [22, 23]. It has been shown [24] that the approximation error introduced by the discretization Γ can be made arbitrarily small by either choosing more states M or increasing the underlying lag time τ. Thus, with an appropriate choice of states, we can ensure an approximation of the true dynamics by a discrete-time and discrete-space Markov model to the desired precision. Although this is a crucial step, the

process of finding an optimal decomposition of state space and appropriate lag time τ is beyond the scope of this paper, and has been discussed at length elsewhere [16, 19].

B. Estimating transition probabilities vs estimating transition rates

In many studies coarse-grained dynamics is equivalently described by the continuous-time *master equation*

$$\dot{\boldsymbol{p}}(t) = \boldsymbol{p}(t)\ \mathbf{K} \tag{6}$$

where $\mathbf{K} \in \mathbb{R}^{m\times m}$ denotes the rate matrix. K_{ij} is the rate associated with the transition $i \to j$, with $K_{ij} > 0$ for $i \neq j$ and $K_{ii} = -\sum_{j\neq i} K_{ij}$ [25, 34, 35]. While $\mathbf{T}(\tau)$ can be straightforwardly estimated from a trajectory using Eq. 4, $\mathbf{K}$ cannot because inversion of the equation

$$\mathbf{T}(\tau) = \exp(\tau\mathbf{K}) \tag{7}$$

is not unique for stochastic matrices unless $\mathbf{T}(\tau)$ is positive definite and reversible. The dynamics is only represented correctly for times larger than some Markov time τ^* and suffers from numerical issues as well. In the following we focus on the estimation of time-discrete transition probabilities. One potential solution, therefore, is to use Bayesian inference to estimate the likely rate matrix given data [15] - we discuss this issue further in Section II G 2.

C. Estimation from time-discrete trajectories

Considering a trajectory $\boldsymbol{q}(t)$ sampled at time intervals Δt we introduce $\hat{\tau} \equiv \tau/\Delta t \in \mathbb{N}$ and $\hat{L} \equiv L/\Delta t \in \mathbb{N}$ and define a temporally discrete trajectory $\boldsymbol{q}$ by

$$q_i = \boldsymbol{q}(i\Delta t) \in \Omega,\ i \in \{0, \ldots, \hat{L}\} \tag{8}$$

and the fractional count matrix $\mathbf{B}(\tau)$ of independent observed transitions by

$$B_{ij}(\tau) = \frac{1}{\hat{\tau}} \sum_{n=0}^{\hat{L}-\hat{\tau}} \chi_i\left(q_n\right) \chi_j\left(q_{n+\hat{\tau}}\right) \tag{9}$$

If the lagtime τ is long enough, the statistical dynamics over times τ and longer can be well-approximated by a Markov chain [23, 24, 36], and the likelihood that a given

transition matrix produces the observations stored in the count matrix $\mathbf{B}$ is given by the multinomial distribution

$$\mathcal{L}(\mathbf{T}) = \mathbb{P}(\mathbf{B}|\mathbf{T}) \propto \prod_{i,j=1}^{M} T_{ij}^{B_{ij}}. \tag{10}$$

As a representative, we choose the unique transition matrix $\hat{\mathbf{T}}(\tau)$ which maximizes this likelihood:

$$\hat{T}_{ij}(\tau) = [\text{argmax } \mathcal{L}(\mathbf{T})]_{ij} = \frac{B_{ij}(\tau)}{\sum\limits_{k} B_{ik}(\tau)}. \tag{11}$$

Alternatively, we can use the state-to-state time-correlation function $C_{ij}(\tau)$ [22, 23] given by

$$C_{ij}(\tau) \equiv \langle \chi_i(0)\, \chi_j(\tau) \rangle \tag{12}$$

which can be estimated in a similar fashion

$$\hat{C}_{ij}(\tau) = \frac{1}{\hat{L}-\hat{\tau}} \sum_{n=0}^{\hat{L}-\hat{\tau}} \chi_i(q_n)\chi_j(q_{n+\hat{\tau}}) = \frac{\hat{\tau}}{\hat{L}-\hat{\tau}} B_{ij}. \tag{13}$$

Although dynamical reweighting can be formulated for different dynamical models [32], the present approach is based on Hamiltonian dynamics in the canonical ensemble [32], which is time-reversible, and thus equilibrium molecular dynamics fulfills detailed balance in state space. Consequently, for trajectories sampled from equilibrium, the correlation matrix will have a symmetric form $C_{ij}(\tau) = C_{ji}(\tau)$. In this case, we can use the estimator

$$\hat{C}_{ij} = \frac{\hat{\tau}}{2(\hat{L}-\hat{\tau})}(B_{ij} + B_{ji}) = \hat{C}_{ji} \tag{14}$$

and write the transition matrix estimate $\hat{\mathbf{T}}(\tau)$ in terms of the correlation matrix estimate $\hat{\mathbf{C}}(\tau)$ as

$$\hat{T}_{ij}(\tau) = \frac{\hat{C}_{ij}(\tau)}{\sum\limits_{k} \hat{C}_{ik}(\tau)} \tag{15}$$

which will also fulfill detailed balance.

D. Transition probabilities from dynamical reweighting

We now demonstrate how transitions observed at all temperatures can be used to infer transition probabilities at any temperature of interest. For a canonical ensemble at

inverse temperature $\beta \equiv (k_B T)^{-1}$ the state-to-state correlation functions can be expressed as Boltzmann-weighted expectation functions

$$C_{ij}(\tau;\beta) = \frac{1}{Z(\beta)} \int d\boldsymbol{q}_0 d\boldsymbol{p}_0 \exp\left(-\beta \mathcal{H}(\boldsymbol{q}_0, \boldsymbol{p}_0)\right) \chi_i(\boldsymbol{q}_0) \chi_j(\boldsymbol{q}_\tau) \quad (16)$$

where $Z(\beta)$ is the complete partition function of both kinetic and potential energies.

Suppose we have a set of N_k Hamiltonian trajectories $\boldsymbol{z}_{kn}(t)$, $n = 1, \ldots, N_k$, $t \in [0, T]$, in which the initial phase space points $\boldsymbol{z}_{kn}(0)$ are sampled from canonical (NVT) distributions at corresponding inverse temperatures β_k, $k = 1, \ldots, K$. By the application of *dynamical reweighting* [31] , a correlation function $C_{ij}(\tau;\beta)$ can be estimated using the entire set of trajectories at all temperatures. For the procedure the association of a trajectory with the temperature it was sampled at is no longer relevant [31], which allows for convenient indexing of all trajectories $\boldsymbol{z}_n$ by one single index $n = 1 \ldots N$ with $N \equiv \sum_{k=1}^{K} N_k$ obtaining

$$\hat{C}_{ij}(\tau;\beta) \approx \sum_{n=1}^{N} w_n(\beta) \cdot \hat{C}_{ij}^{(n)}(\tau) \quad (17)$$

where the individual trajectory contributions to the correlation functions $\hat{C}_{ij}(\tau,\beta)$ are given by Eq. (14)

$$\hat{C}_{ij}^{(n)} = \frac{\hat{\tau}}{2(\hat{L} - \hat{\tau})} \left(B_{ij}^{(n)} + B_{ji}^{(n)} \right) \quad (18)$$

with $\mathbf{B}^{(n)}$ being the (possibly fractional) count matrix and $\hat{\mathbf{C}}^{(n)}$ the correlation matrix computed from trajectory n sampled at the respective inverse temperature β_{k_n} [22, 23]. For later analysis we keep the information of which trajectory n was sampled at which temperature β_k in a vector $\boldsymbol{e}$ with $e_n = k$.

The row-stochastic transition matrix estimate $\hat{\mathbf{T}}(\tau;\beta)$ is then computed from (15)

$$\hat{T}_{ij}(\tau;\beta) = \frac{\hat{C}_{ij}(\tau;\beta)}{\sum_k \hat{C}_{ik}(\tau;\beta)} \quad (19)$$

As $\hat{\mathbf{C}}(\tau;\beta)$ is symmetric by construction, $\hat{\mathbf{T}}(\tau;\beta)$ will be reversible (i.e., will satisfy detailed balance).

The normalized trajectory weights $w_n(\beta)$ can be computed by

$$w_n(\beta) = \hat{Z}(\beta)^{-1} \left[\sum_{k=1}^{K} N_k \hat{Z}_k^{-1} \exp\left(-(\beta_k - \beta) E_n\right) \right]^{-1} \quad (20)$$

with normalization constants $Z_k \equiv Z(\beta_k)$, and

$$\hat{Z}(\beta) = \sum_{n=1}^{N} \left[\sum_{k=1}^{K} N_k \hat{Z}_k^{-1} \exp\left(-(\beta_k - \beta) E_n\right) \right]^{-1} \tag{21}$$

where $E_{kn} \equiv H(\boldsymbol{z}_n(0))$ denotes the total energy of the trajectory, which is constant over trajectories for Hamiltonian dynamics [32]. The normalization constants Z_k are determined by the solution of a set of self-consistent equations

$$\hat{Z}_k = \sum_{n=1}^{N} w_n(\beta_k), \ \forall\, i \in \{1, \ldots, K\} \tag{22}$$

which can be obtained efficiently in a number of ways (see Appendix A), although it is often necessary to work with logarithmic representations to avoid numerical instability. A detailed exposition is presented in Ref. [32].

E. Estimation of uncertainties in transition probabilities

For a given temperature β, the statistical uncertainty in $\hat{C}_{ab} \equiv \hat{C}_{ab}(\tau; \beta)$ can be estimated in a straightforward manner [31]. We start with the $N \times K$ weight matrix $\mathbf{W}$ the elements of which are given by

$$w_{nk} \equiv w_n(\beta_k). \tag{23}$$

augmenting it by three additional columns, indexed by x, X, and Y, consisting of

$$w_{nx} = w_n(\beta)$$

$$w_{nX} = \frac{\hat{C}_{ab}^{(n)}}{\hat{C}_{ab}(\tau; \beta)} w_{nx} \ ; \ w_{nY} = \frac{\hat{C}_{a'b'}^{(n)}}{\hat{C}_{a'b'}(\tau; \beta)} w_{nx} \tag{24}$$

The uncertainty in the transition probabilities $T_{ij}(\hat{\mathbf{C}})$ can then be estimated from the uncertainty in $\hat{\mathbf{C}}$ by a first-order Taylor expansion

$$\begin{aligned} \delta^2 \hat{T}_{ij} &\equiv \left\langle (\hat{T}_{ij} - \left\langle \hat{T}_{ij} \right\rangle_\beta)^2 \right\rangle_\beta \\ &\approx \sum_{a,a',b,b'=1}^{M} \left[\frac{\partial \hat{T}_{ij}}{\partial \hat{C}_{ab}} \right] \left[\frac{\partial \hat{T}_{ij}}{\partial \hat{C}_{a'b'}} \right] \delta \hat{C}_{ab} \delta \hat{C}_{a'b'} \end{aligned} \tag{25}$$

where the covariance of the estimates $\hat{C}_{ab}$ and $\hat{C}_{a'b'}$ can be estimated as [32]

$$\delta\hat{C}_{ab}\delta\hat{C}_{a'b'} \approx \hat{C}_{ab}\,\hat{C}_{a'b'}\left[\hat{\Theta}_{xx} - \hat{\Theta}_{xY} - \hat{\Theta}_{Xx} + \hat{\Theta}_{XY}\right] \tag{26}$$

with the covariance matrix estimate $\mathbf{\Theta}$ computed as

$$\hat{\mathbf{\Theta}} \equiv \mathbf{W}^T[\mathbf{I}_N - \mathbf{W}\mathbf{N}\mathbf{W}^{\mathrm{T}}]^{+}\mathbf{W}, \tag{27}$$

where $\mathbf{I}_N$ is the identity matrix of rank N and $\mathbf{N} = \mathrm{diag}(N_1, \ldots, N_K, 0, 0, 0)$. The $[]^+$ here denotes the generalized inverse. Using Eq. (4) the sensitivity of $\hat{T}_{ij}$ to the correlation $\hat{C}_{ab}$ is given by

$$\frac{\partial\hat{T}_{ij}}{\partial\hat{C}_{ab}} = \frac{\delta_{aj}\delta_{bi} + \delta_{ai}\delta_{bj} - \delta_{ab}}{\hat{C}_i} - \frac{\hat{C}_{ij}\left(-M\delta_{ab} + \delta_{ai} + \delta_{bi}\right)}{\hat{C}_i^2}, \tag{28}$$

with $\hat{C}_i = \sum_i \hat{C}_{ij}$ being the row sum of the correlation matrix $\mathbf{C}$. The final expression for the covariance matrix of transition probabilities $\delta^2 T_{ij}$ is complex but still calculable. A detailed description of this procedure can be found in [31, 32].

F. Modified parallel tempering protocol

We employ a modified parallel tempering protocol in which a set of Hamiltonian trajectory segments $\boldsymbol{z}_{kn}(t)$ of uniform length $T \geq \tau$ is generated, with the initial phase space points $\boldsymbol{z}_{kn}(0)$ sampled from the canonical (NVT) ensemble at corresponding inverse temperatures $\beta_1, \ldots, \beta_K$ [32]. We start by assuming that some process was used to generate the initial phase space points $\boldsymbol{z}_{k0}(0)$ from equilibrium within the canonical ensemble at each corresponding inverse temperature β_k

$$\mathbb{P}\left(\boldsymbol{z}_{k0}(0)\right) = [Z(\beta_k)]^{-1}\, e^{-\beta_k \mathcal{H}(\boldsymbol{z}_{k0}(0))} \tag{29}$$

This initial phase space points may be obtained, for example, by a standard parallel tempering protocol, or by running the modified protocol for a number of iterations starting from one or more arbitrary initial configurations.

Consider iteration n of the algorithm. For each temperature index $k = 1, \ldots, K$, Hamilton's equations of motion are propagated using a symplectic integrator with sufficiently small timesteps to generate trajectories of $\boldsymbol{z}_{kn}(t)$ of length T. Finally, we propose exchanges between the final configurations $\boldsymbol{z}_{in}(T)$ and $\boldsymbol{z}_{jn}(T)$ of neighboring temperatures

β_i and $\beta_{i\pm1}$, starting from the highest temperature down to the lowest one in odd iterations and in reverse order in even ones [17].[1] The Metropolis-like probability [13] of accepting or rejecting the exchange depends on the final potential energies of the configurations U_i and U_j with

$$P_{\text{exch}}(U_i, \beta_i; U_j, \beta_j) = \min\{1, \exp[-(\beta_i - \beta_j)(U_j - U_i)]\}$$

Regardless of whether the exchange is accepted or rejected, we reassign the velocities according to the Maxwell-Boltzmann distribution [38] at the new (or old, if rejected) temperatures, and denote the new phase space points from which the next iteration can be carried out as $\boldsymbol{z}_{k(n+1)}(0)$ (see proof in Appendix B). This satisfies the conditions defined by Okamoto [13] in order for the kinetic energies to not appear in P_{exch} and is equivalent to rescaling the velocities for accepted exchanges and then applying a massive collision for the Andersen thermostat [38]. The reason for reassignment of velocities instead of rescaling is that when using Hamiltonian trajectories minimal thermostatting would otherwise take place.

G. Bayesian estimation of transition probabilities from a single temperature

We also consider two Bayesian methods for estimation of the transition matrices and rate matrices using data collected from a single temperature. Both methods sample transition probabilities or rates according to the same likelihood function, but employ different model parameterizations and, more importantly, different prior probability distributions.

1. *Reversible Transition Matrices*

We use the approach described in [39]. Starting with an observation represented by the fractional count matrix $\mathbf{B}$, the posterior probability of a transition matrix $\mathbf{T}(\tau)$ given

[1] Note that other exchange proposal schemes can be used, provided the resulting algorithm satisfies the condition of "balance" (not detailed balance) [37]

this observation is

$$\mathbb{P}(\mathbf{T}|\mathbf{B}) \propto \mathbb{P}(\mathbf{B}|\mathbf{T})\,\mathbb{P}(\mathbf{T}) = \prod_{i,j\in S} T_{ij}^{B_{ij}}\,\mathbb{P}(\mathbf{T}). \tag{30}$$

As the prior, $\mathbb{P}(\mathbf{T})$, we choose a Dirichlet distribution for each row which adds no additional observations to the likelihood probability

$$\mathbb{P}(T) \equiv \prod_{i,j\in S} T_{ij}^{-1}. \tag{31}$$

Furthermore, we restrict ourselves to transition matrices that fulfill detailed balance, i.e., are reversible with respect to the stationary distribution π:

$$\pi_i T_{ij} = \pi_j T_{ji} \tag{32}$$

Here, the distribution in Eq. (30) was sampled using a Markov chain Monte Carlo procedure described in Ref. [39].

2. *Reversible Rate Matrices*

To sample rate matrices $\mathbf{K}$ with elements $K_{ij} > 0$ for $i \neq j$ and $K_{ii} = -\sum_{j\neq i} K_{ij}$ we use the approach proposed in Ref. [25], which does not estimate the transition probabilities directly, but uses a parametric form of a reversible rate matrix $\mathbf{K}$, that uses the logarithms of the elements in the upper-right triangular matrix K_{ij} for $j > i$ (without diagonal entries) and the equilibrium distribution π_i, $i = 1 \ldots M$ thus assuring a rate matrix with non-positive eigenvalues and also positive off-diagonal rates. The posterior in Eq. 30, written in terms of the rate matrix $\mathbf{K}$, is given by

$$\mathbb{P}(\mathbf{K}|\mathbf{B}) \propto \mathbb{P}(\mathbf{B}|\mathbf{K})\mathbb{P}(\mathbf{K}) = \prod_{i,j\in S} \exp\left(\tau\mathbf{K}\right)_{ij}^{B_{ij}}\,\mathbb{P}(\mathbf{K}) \tag{33}$$

where the prior is uniform in $\ln K_{ij}, j > i$ and $\ln \pi_i, i = 1 \ldots (N-1)$ and the detailed balance constraint

$$\pi_i K_{ij} = \pi_j K_{ji} \tag{34}$$

holds. After sampling rate matrices with a Metropolis Monte Carlo scheme [25] the related set of transition matrices with the lag time τ is computed by

$$\mathbf{T}(\tau) = \exp\left(\tau\mathbf{K}\right) \tag{35}$$

All methods with their abbreviations and colors used consistently in the text and figures in this article are listed in Table I.

Method	Abbreviation	Color
Transition matrix estimation [39]	[TE]	RED
Rate matrix estimation [25]	[RE]	GREEN
Dynamical reweighting [32]	[DR]	BLUE
Shooting trajectories at 302 K [17]	[ST]	BLACK

TABLE I. Table of methods used for transition probability or rate estimation with their corresponding abbreviations and colors used consistently throughout this paper

III. APPLICATION TO TERMINALLY-BLOCKED ALANINE DIPEPTIDE

A. System Setup

To illustrate the construction of temperature-dependent Markov models using dynamical reweighting, we estimated the transition probabilities between conformational states for the terminally-blocked alanine peptide (Ace-Ala-Nme) (s. Figure 2) in explicit solvent from a parallel tempering molecular dynamics (PT) simulation. An ensemble of 501 Hamiltonian trajectories 20 ps in length at each of 40 temperatures were generated, spanning the range from 273 to 600 K, with peptide configurations stored every 0.1 ps. The temperatures were exponentially spaced, thus providing a good overlap in the potential and total energy distributions between neighboring temperatures and resulting in an average exchange acceptance probability of $\sim 50\%$ (see Figure 1). All details of the simulation were given previously [17].

A leapfrog Verlet integrator [40–42] (with bonds involving hydrogen atoms constrained) was used to produce the dynamical trajectories. The fluctuation in total energy averaged over all 20 ps trajectories at each temperature was minimal and the drift negligible (see Table II). The production run followed a 1 ns equilibration phase during which exchanges were attempted at 1 ps intervals, ensuring that all initial configurations were drawn from equilibrium at their respective temperatures. Previous work has demonstrated that a Markov model based on a six-state decomposition, as depicted in Figure 2, can accurately describe the dynamics of this peptide for lagtimes longer than $\tau = 6$ ps [17]. We employ the same state decomposition for all temperatures with the

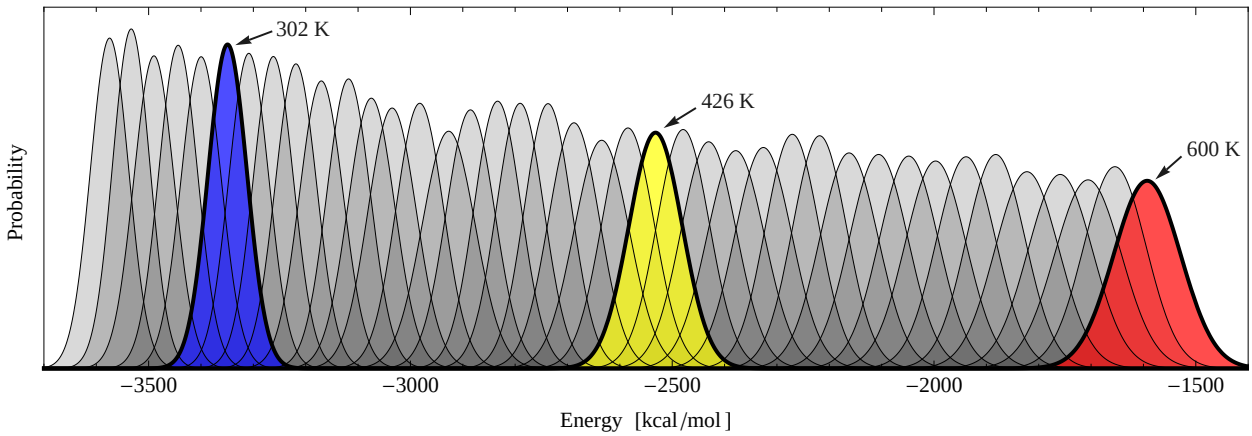

FIG. 1. Distribution of total trajectory energies in kcal/mol for all 40 temperatures from parallel tempering simulation. Highlighted are temperatures 302 K, 425 K, and 600 K the single-temperature Bayesian analysis convergence properties of which are shown in Figure 8.

Temp [K]	ASD [kcal/mol]	Drift [kcal/(ps mol)]
302	0.214 ± 0.014	0.0056 ± 0.0006
426	0.280 ± 0.019	0.0073 ± 0.0006
600	0.376 ± 0.026	0.0097 ± 0.0011

TABLE II. Average of the standard deviation in the energies (ASD) and average drift energies of leapfrog integrator computed over all 20 ps trajectories of given temperature

suggested minimal lagtime of $\tau = 6$ ps.

To evaluate the accuracy of the methods for estimating transition probabilities, we compare the separate estimates obtained using dynamical reweighting (DR), transition matrix estimation (TE), and rate matrix estimation (RE) with a simulation of $6 \times 10\,000$ short (10 ps) trajectories (ST) initiated from the equilibrium ensemble within each state at 302 K. The PT simulation, in comparison, furnishes a total of 501 independent trajectories at that temperature.

The system is small enough that reasonable statistics can be obtained with moderate CPU requirements, while complex enough that some transitions (and even some states) are sampled only at high temperatures. In what follows, the results from the Markov model obtained from the dynamical reweighting method are compared to the model

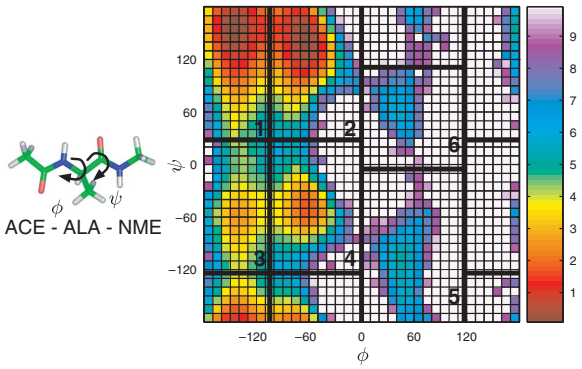

FIG. 2. Terminally-blocked alanine peptide potential of mean force and Markov state definitions. Left: The terminally blocked alanine peptide with (ϕ, ψ) torsions labeled. Right: The potential of mean force as a function of (ϕ, ψ) torsions at 302 K in units of $k_B T$, estimated from the parallel tempering simulation using WHAM [28, 43], truncated at 10 $k_B T$ (white regions). The six manually identified states are labelled in black [17].

computed by Bayesian analyses using data from a single temperature only, as in Buchete and Hummer [25].

B. Estimated transition probabilities as a function of temperature

A comparison of the transition probabilities between all 6×6 pairs of states as a function of temperature is given in Figure 3. The blue solid lines give the estimates from dynamical reweighting (DR) [17] using all available data at all temperatures, as described in Section II. To obtain the dimensionless free energy estimates $\hat{f}_i$, we solved the set of self-consistent equations in Eq. (22) with a relative convergence tolerance in the residual of 10^{-7} (see Appendix A). Transition probabilities were also estimated at one intermediate temperature between each pair of simulated temperatures.

The red dashed lines in Figure 3 show transition probabilities for the reversible single temperature estimation of transition matrices (TE) [39]. For each of the 40 temperatures the sampler was run to collect a total of 10 000 samples. For the sampling of reversible rate matrices (RE), depicted by dotted green lines, the sampling as proposed in Ref. [25] was used. After an equilibration phase of 1 ns with exchange attempts every 1 ps, a total of 10 000 samples was stored for each temperature separately. Diagnostics of convergence

for both methods appear as supplementary Figure 8. The black cross-hair in Figure 3 refers to the reference values at 302 K.

Qualitatively, all methods agree, and especially for transitions among highly populated states (1 to 4). However, the reweighting estimate, which uses the combined data from all temperatures, has smaller uncertainties than the estimators that use only individual temperatures. The general agreement with the reference simulation is best for dynamical reweighting. The two Bayesian methods have almost identical predictions for transitions among states with many observed transitions, which is as expected since they use the same likelihood functions and the influence of the choice of the prior probability distribution is minimal here. However, the differences in the transition probabilites among states with few observations (states 5 and 6) arise from the influence of different prior probability distributions. See Table III for comparison.

C. Detailed comparison of transition probability estimates at 302 K

For a detailed comparison with precisely known transition probabilites, the Bayesian analysis method with reversibility constraint for transition matrices[39] (TE) was also applied to a large set of shooting trajectories at 302 K, in which many trajectories are initiated from an equilibrium distribution within each state. The results of the comparison at 302 K and a lag time of $\tau = 6$ ps between the different estimation methods are shown in Figure 4. All colors are the same as in Figure 3 and Table I.

For transitions that are not sampled at certain temperature ranges, the maximum-likelihood estimates obtained with the present reweighting method are close to zero (see Figure 3). Generally, for transition probabilities close to zero or unity, the normal distribution is a poor approximation to this highly asymmetric distribution and therefore leads to too large uncertainties in these cases (see Figure 4). Thus we suspect that reweighting will significantly overestimate the uncertainty in these cases since this is derived from the locally estimated Hessian of the probability distribution.

Overall the reweighting method performs very well compared to the single-temperature estimates. Even transition probabilities that are sampled very poorly at 302 K (such as for transitions involving states 5 and 6) have a good agreement with the reference values at 302 K. Table III shows the standard deviation in the absolute difference of the estimation

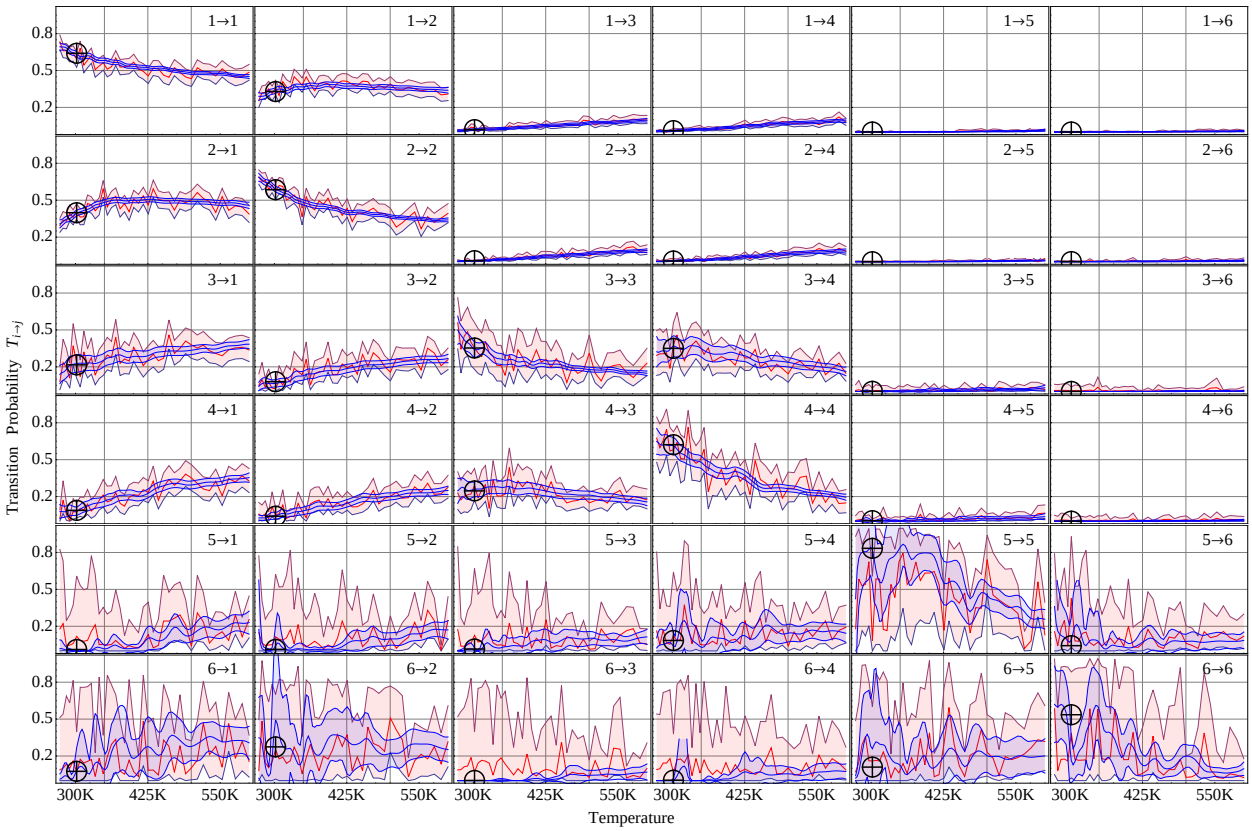

FIG. 3. Comparison of all 6x6 inter-state transition probabilities at a lagtime of $\tau = 6$ ps as a function of temperature with error bars showing 95% confidence intervals. Blue lines show the transition probabilities estimated using dynamical reweighting (DR). Red lines show the estimates from transition matrix estimation (TE) computed from only single temperature data. The black cross-hair indicates the reference using the shooting trajectory data (ST) at 302 K only.

methods compared to the reference simulation (ST) using a lag time of $\tau = 6$ ps. The dynamical reweighting has a smaller deviation than both Bayesian methods for both high and low free-energy states.

D. Comparison of temperature dependence of eigenvalues

Dynamical reweighting can also be applied to estimate properties derived from the transition probabilities. For example, the eigenvalues λ_i of a transition matrix are related to the timescales of processes t_i^* by

$$t_i^* = -\tau / \ln(\lambda_i), \tag{36}$$

	RMS in Estimation		
	(DR)	(TE)	(RE)
Low-energy states (1 to 4)	0.007	0.023	0.055
High-energy states (5 and 6)	0.140	0.267	0.288
All Transitions	0.079	0.151	0.167

TABLE III. RMS in the absolute difference of transition probabilities for $\tau = 6$ ps compared to the reference simulation (ST) at 302 K for the three methods of Markov model estimation and high- and low-energy subsets of transitions. Dynamical reweighting (DR) shows the smallest RMS error compared to the reference simulation (ST).

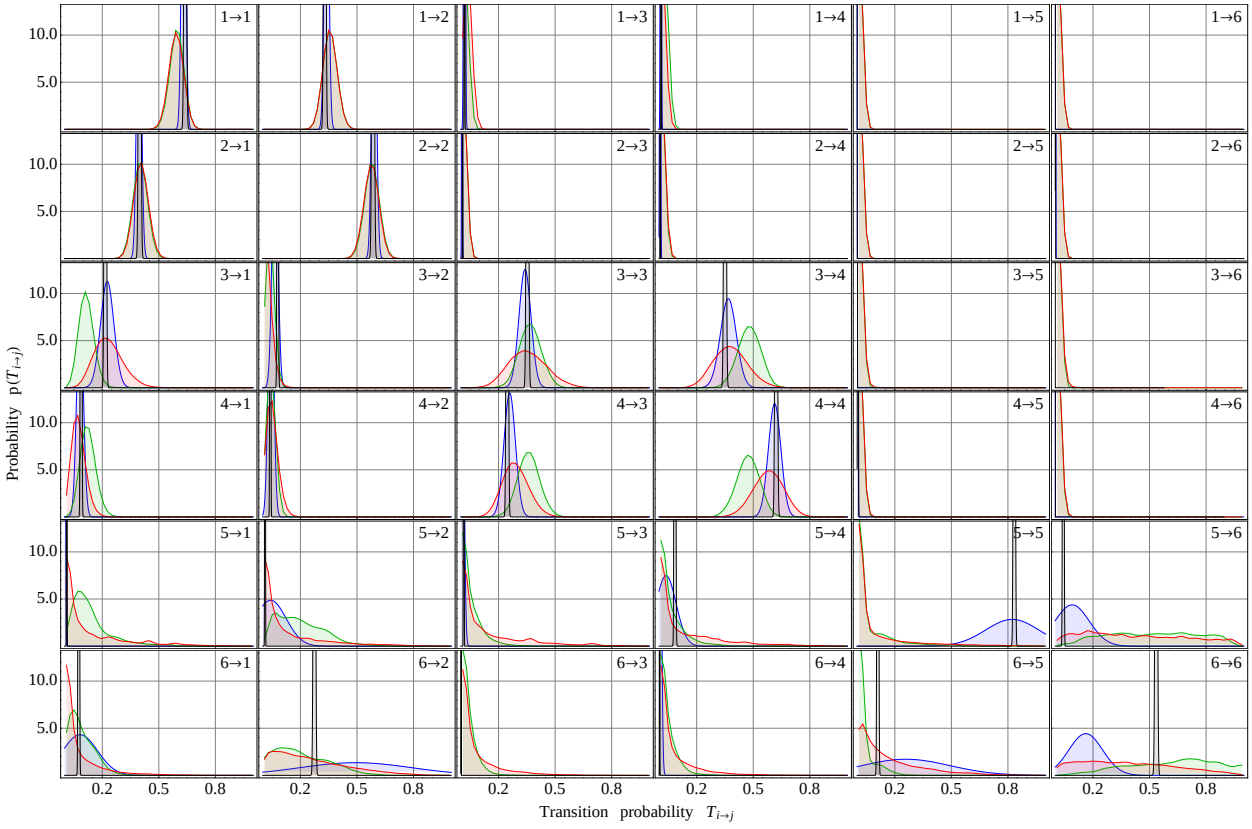

FIG. 4. Detailed comparison of transition probabilities for $\tau = 6$ ps and uncertainty estimations at 302 K. Red: Single temperature estimation of transition matrix (TE); Green: Single temperature estimation of rate matrix (RE); Blue: dynamical reweighting estimation (DR), Black: reference using shooting trajectories (ST).

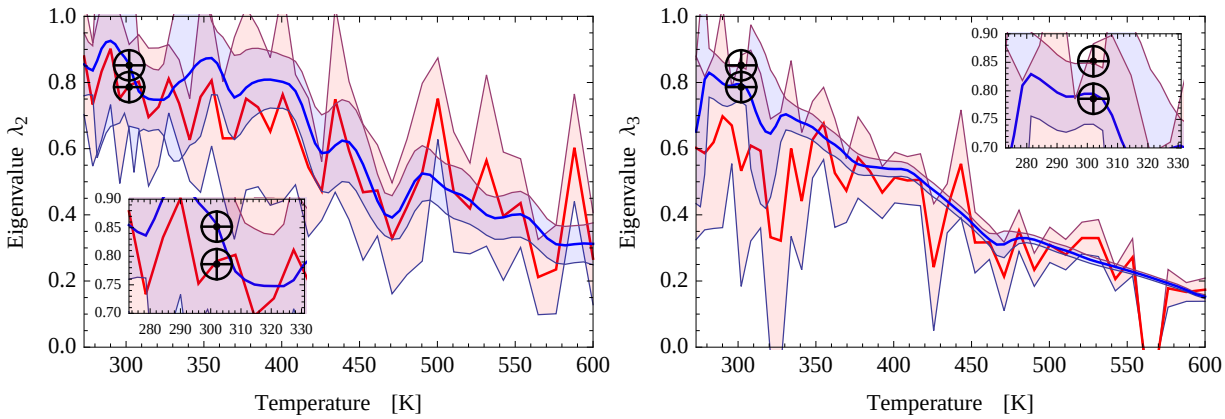

FIG. 5. Temperature dependence of estimated eigenvalues. Red: Single temperature estimation of transition matrix (TE), Blue: dynamical reweighting estimation (DR), Black: reference 2nd and 3rd eigenvalue at 302 K using shooting trajectories (ST). Left: Comparison of the second largest eigenvalue vs temperature, Right: Comparison for third eigenvalue. The third reference eigenvalue is well predicted by both estimation methods at all temperatures although it matches only the second eigenvalue in the transition matrix estimation (TE). The second reference eigenvalue at low temperatures (below 350 K) is only detected by dynamical reweighting (DR).

where we assume that the eigenvalues λ_i are sorted in order of descending modulus ($\lambda_1 = 1 > |\lambda_2| > \ldots > |\lambda_M| > 0$) [23]. Hence, eigenvalues close to unity imply slow processes i.e., those we are mostly interested in.

We investigated the dependence of the eigenvalues on the temperature in the present system. Figure 5 compares estimates for the second and third eigenvalues (λ_2, λ_3) of the transition matrix at each temperature with the different methods. The variance in the (TE) case was estimated from the set of eigenvalues of each sampled transition matrix. To estimate the errors of (DR), we used linear error propagation of the uncertainties in the transition matrix to the errors in the eigenvalues [20]. At low temperatures (below 350 K), the second eigenvalue is estimated correctly by dynamical reweighting (DR), but not by single-temperature estimations. This is due to the fact that the transition process corresponding to this slowest timescale is not sampled at these low temperatures. Thus, estimates using only data collected at that temperature are erroneous. The agreement of dynamical reweighting timescales with the reference simulation is very good, although the error bars of the reweighted estimate are still very large compared to the good agree-

ment of the estimated values with the reference values from the shooting trajectories. We assume that the inappropriate approximation of the asymmetric distributions with normal distributions used for the linear error propagation lead here as well to an overestimation of the errors in the transition probabilities.

The third largest eigenvalue is predicted by both methods equally well (Fig. 5), although it occurs as the second-largest eigenvalue in the single-temperature estimates, which missed sampling the slowest process (described by λ_2) completely. A direct comparison of the predicted eigenvectors (Fig. 6) reveals that the slowest process (given by the second eigenvector of the reference transition matrix (ST)) is not detected by any of the single temperature methods. However, dynamical reweighting successfully finds all the processes, although the matching eigenvalues, and thus timescales, are permuted for faster processes.

The comparison of Markov models is a nontrivial task [44], for which we use a symmetrized form of the transition matrix $\mathbf{T}^{sym}$ and expand it into a sum of rank one matrices $\mathbf{Q}_i$ spanned by an outer product of the eigenvectors of $\mathbf{T}^{sym}$ by

$$\mathbf{T}^{sym} = \mathrm{diag}(\boldsymbol{\pi}^{1/2})\ \mathbf{T}\ \mathrm{diag}(\boldsymbol{\pi}^{-1/2}) \tag{37}$$

$$= \mathbf{R}\ \mathrm{diag}(\boldsymbol{\pi}^{1/2})\ \mathbf{R}^{\mathbf{T}} \tag{38}$$

$$= \sum_{i=1}^{M} \lambda_i \boldsymbol{r}_i\ \boldsymbol{r}_i^T \equiv \sum_{i=1}^{M} \mathbf{Q}_i. \tag{39}$$

Here, $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_M\}$ is the equilibrium distribution, $\mathbf{R} = \{\boldsymbol{r_1}, \ldots, \boldsymbol{r_M}\}$ the matrix of normalized eigenvectors of $\mathbf{T}^{sym}$ and λ_i the corresponding eigenvalues, which are equal to the eigenvalues of $\mathbf{T}$. Each of the submatrices $\mathbf{Q}_i$ can be considered as a part of the full transition matrix working on a timescale given by the respective eigenvalue. These subprocesses are similar independent of their respective eigenvalues if their eigenvectors match, as can be determined by the scalar product between the spanning eigenvectors r_i:

$$\mathrm{Sim}(\mathbf{Q}_i, \mathbf{Q}'_j) = \boldsymbol{r}_i^T \boldsymbol{r'}_j. \tag{40}$$

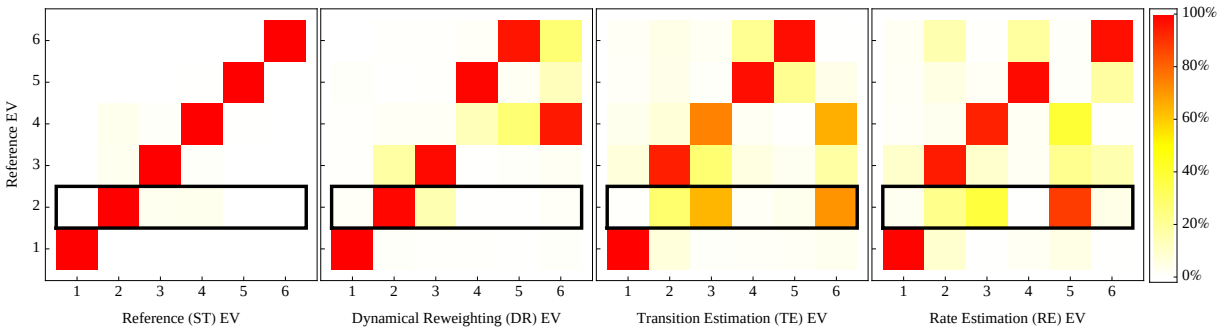

FIG. 6. Similarity matrices (scalar product) of eigenvectors (EV) from symmetrized transition matrices estimated with different methods (ST, DR, TE, RE) at 302 K. The eigenvectors indicate the states involved in the process, and thus high similarity (red) indicates a good approximation to the reference process (ST). Eigenvectors are sorted as descending eigenvalues. The 2nd eigenvector is found correctly only by dynamical reweighting, meaning that the single temperature estimations are unable to correctly predict the slowest process at 302 K.

E. Contributions from different temperatures to the estimates of expectation values

The contribution of each trajectory segment to the estimation of any expectation value at any given temperature is illustrated in Fig. 7. The left plot shows the average normalized weights from trajectories sampled from the distribution at β_k reweighted to β_l:

$$\bar{w}(\beta_l|\beta_k) = \frac{1}{\sum_{n\in Q(k)} 1} \sum_{n\in Q(k)} w_n(\beta_l) \tag{41}$$

with $Q(k) = \{n \in 1 \dots N \mid e_n = k\}$ being the set of trajectories sampled from the distribution at β_k and the unnormalized trajectory weights $w_n(\beta)$ given in (20), where, on average, seven temperatures contribute more than 1% to the expectation. The right hand plot illustrates the contribution from the sampled data at β_k to the transition counts $B_{ij}(\beta_l)$ for the transition $6 \to 5$ given by

$$B_{ij}(\beta_l \mid \beta_k) = \frac{1}{\sum_{n\in Q(k)} 1} \sum_{n\in Q(k)} \tilde{w}_n(\beta_l)\hat{C}_{ij}^{(n)}, \tag{42}$$

with k indicating the temperature contributed from, l the temperature estimated at and i, j representing the transition $i \to j$. As expected, most information contributing to a specific temperature estimate is always contained in the simulations performed at the nearest temperatures.

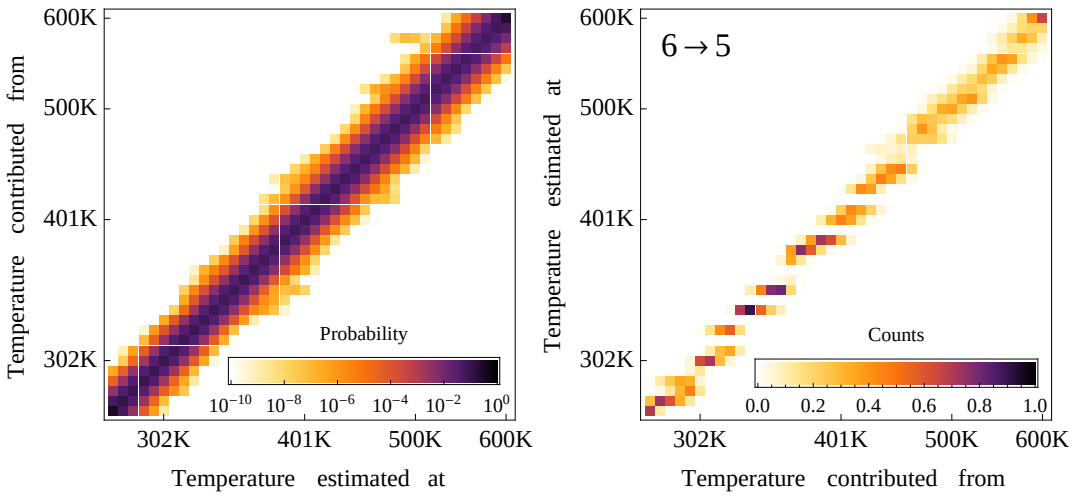

FIG. 7. Upper: Relative contribution $\bar{w}_{kl}$ in Eq. 41 to the estimates at inverse temperature β_k from simulations at inverse temperature β_l averaged over all trajectories at the same temperature. On average, seven temperatures contribute more than 1% each to the estimation. Lower: Contributions to the estimation of transition counts for the transition $6 \to 5$. The sum of one row equals to the total counts estimated by the method at the desired temperature.

IV. DISCUSSION

The present method provides a means of generating an estimate of transition probabilities from parallel tempering MD simulations of biomolecules as a continuous function of temperature. Even at intermediate temperatures not included in the simulation the estimate is much more precise than that obtained with either single-temperature methods. At low temperatures, at which some transitions are not observed at all, low transition probabilities can still be estimated.

Additionally, the estimates of transition probabilities can be differentiated with respect to the inverse temperature β, because the trajectory weights $w_n(\beta)$ are differentiable functions of temperature (20). This allows, in principle, thermodynamic properties to be computed (e.g., heat capacities), provided caution is taken in dealing with numerical issues since the trajectory weights $w_n(\beta)$ can easily span hundreds of orders of magnitude.

Although the method can be applied to several dynamical models [32], we chose a set of parallel tempering simulations with a modified protocol to produce a series of NVE trajectories with initial configurations drawn from the NVT ensemble. For very large sys-

tems, the PT simulation might not be long enough to globally converge. In this case we cannot use reversible counting as in Eq. (18) to enhance the statistics, but the method can still be applied without the detailed balance constraint as long as we draw from equilibrium inside each set Γ_i. The PT requirement of good exchange rates also ensures good overlap in the contribution to the dynamical reweighting for neighboring temperatures. The approach itself is not limited to Hamiltonian trajectories but can be extended to other dynamics (e.g. Brownian and Langevin dynamics) as long as an analytical connection between the weighting factors and the temperature exists [32].

The degree to which the use of parallel tempering can enhance thermodynamic sampling efficiency has been a matter of much discussion. While activated processes will be sampled more often at higher temperatures, entropic barriers become less and less probable to pass, effectively limiting the possible improvement in sampling. The same problem limits the range of contributing temperatures, too. Lower temperatures increase the sampling of entropic barriers, while at the same time decreasing probabilities for enthalpic barriers. However, information about activated processes is transferred from higher to lower temperatures and, for entropic barriers, from lower to higher temperatures. Transitions in the alanine peptide test system are dominated by activated processes (i.e., enthalpic barriers).

Both single-temperature methods give similar results for transitions with good statistics, differing mostly for transitions that have only rarely been sampled due presumably to the influence of different prior probability distributions. Surprisingly, the Bayesian estimates provide a reasonable bound on transition probabilities to and from a state even when the state is not sampled at all. This is most likely due to the reversibility constraint, which seems to provide information even in cases where there are few transitions to or from a state. There is, however, a dependence on the Bayesian prior, which leads to different predictions in cases where the state is not sampled or only rarely sampled. The rate matrix estimation (RE) assumes, in addition to the detailed balance constraint, positivity of all eigenvalues and non-negative off-diagonal entries. The uniform distribution of parameters in logarithmic space leads most likely to favoring of low transition probabilities in states with poor transition statistics.

The way in which the transition probabilities are estimated in terms of equilibrium correlation functions requires that the trajectory segments sampled during the parallel

tempering simulation be drawn from the equilibrium distribution and that the trajectories to be reweighted be uncorrelated so as to permit an estimation of the statistical error.

The predictions of mean values are very good, while the quality of the error estimation is limited to a Gaussian approximation, which is problematic in cases where transition probabilities are close to the extremes (0 or 1) and the probability distribution is thus very asymmetric. Some combination of Bayesian and reweighting methods (such as T-WHAM[29]) may provide the best of both types of estimators by yielding more accurate uncertainties at the expense of introducing some bias from the introduction of energy histograms or some other parametric distributions for describing the energy density of states. Finally, the enhanced estimates of mean values and their respective statistical uncertainties may be used to guide subsequent (potentially adaptive) sampling strategies, as described in Ref. [21].

V. ACKNOWLEDGMENTS

The authors would like to thank Jed W. Pitera (IBM Almaden), Nicolae-Viorel Buchete (UCD Dublin), and Gerhard Hummer (NIH) for stimulating conversations during the execution of this work. JHP gratefully acknowledges funding from the German Research Foundation (DFG) through the award of a doctoral scholarship in the International Graduiertenkolleg IGK 710: "Complex processes: Modeling, Simulation and Optimization". JDC gratefully acknowledges support from HHMI and IBM predoctoral fellowship programs, NIH grant GM34993 through Ken A. Dill (UCSF), and NSF grant for Cyberinfrastructure (NSF CHE-0535616), and a California Institute for Quantitative Biosciences (QB3) Distinguished Postdoctoral Fellowship at various points throughout this work. VSP acknowledges support from NIH RO1 GM062868. JHP and FN acknowledge support from DFG Research Center Matheon and DFG grant No. 725/2. JCS acknowledges funding from the U.S. Department of Energy "Multiscale Mathematics" SciDAC and Genomes-to life program under grant No. ERKJE84/ERKPE84.

---

[1] E. Z. Eisenmesser, D. A. Bosco, M. Akke, and D. Kern, Science **295**, 1520 (2002).

[2] E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay, and D. Kern, Nature **438**, 117 (2005).

[3] H. Feng, Z. Zhou, and Y. Bai, Proc. Nat. Acad. Sci. USA **102**, 5026 (2005).

[4] W. Min, G. Luo, B. Cherayil, S. Kou, and X. Xie, Phys. Rev. Lett **94**, 1 (2005).

[5] G. Smith, K. Lee, X. Qu, Z. Xie, J. Pesic, T. Sosnick, T. Pan, and N. Scherer, J. Mol. Biol. **378**, 941 (2008).

[6] X. Zhuang and M. Rief, Curr. Opin. Struct. Biol. **13**, 88 (2003).

[7] A. Matagne, S. Radford, and C. Dobson, J. Mol. Biol. **267**, 1068 (1997).

[8] H. Neuweiler, S. Doose, and M. Sauer, Proc. Nat. Acad. Sci. USA **102**, 16650 (2005).

[9] R. Goldbeck, Y. Thomas, E. Chen, R. Esquerra, and D. Kliger, Proc. Nat. Acad. Sci. USA **96**, 2782 (1999).

[10] C. J. Geyer and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).

[11] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[12] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).

[13] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).

[14] M. Jäger, H. Nguyen, J. C. Crane, J. W. Kelly, and M. Gruebele, J. Mol. Biol. **311**, 373 (2001).

[15] N.-V. Buchete and G. Hummer, J. Phys. Chem. B **112**, 6057 (2008).

[16] F. Noé, I. Horenko, C. Schütte, and J. C. Smith, J. Chem. Phys **126**, 155102 (2007).

[17] J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, Multiscale Model. Sim. **5**, 1214 (2006).

[18] S. Muff and A. Caflisch, J. Phys. Chem. B **113**, 3218 (2009).

[19] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, J. Chem. Phys **126**, 155101 (2007).

[20] N. S. Hinrichs and V. S. Pande, J. Chem. Phys **126**, 244101 (2007).

[21] N. Singhal and V. S. Pande, J. Chem. Phys **123**, 204909 (2005).

[22] W. C. Swope, J. W. Pitera, and F. Suits, J. Phys. Chem. B **108**, 6571 (2004).

[23] W. C. Swope, J. W. Pitera, F. Suits, M. Pitman, M. Eleftheriou, B. G. Fitch, R. S. Germain, A. Rayshubskiy, T. C. Ward, and Y. Zhestkov, J. Phys. Chem. B **108**, 6582 (2004).

[24] M. Sarich, F. Noé, and C. Schütte, Multiscale Model. Sim. (2009).

[25] N.-V. Buchete and G. Hummer, Phys. Rev. E **77**, 4 (2008).

[26] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[27] C. Bartels and M. Karplus, J. Comput. Chem. **18**, 1450 (1997).

[28] S. Kumar, J. M. Rosenberg, D. Bouzida, and R. H. Swendsen, J. Comp. Chem. **13**, 1011 (1992).

[29] E. Gallicchio, M. Andrec, A. K. Felts, and R. M. Levy, J. Phys. Chem. B **109**, 6722 (2005).

[30] C. Bartels, Chem. Phys. Lett. **331**, 446 (2000).

[31] M. R. Shirts and J. D. Chodera, J. Chem. Phys **129**, 124105 (2008).

[32] J. D. Chodera, M. R. Shirts, J.-H. Prinz, W. C. Swope, F. Noé, and V. S. Pande, in preparation (2009).

[33] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).

[34] A. F. Voter and J. D. Doll, J. Chem. Phys **82**, 80 (1985).

[35] J. Adams and J. D. Doll, Surf. Sci. **111**, 492 (1981).

[36] J. D. Chodera, W. Swope, J. Pitera, and K. Dill, Multiscale Model. Sim. **5**, 1214 (2006).

[37] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).

[38] H. C. Andersen, J. Chem. Phys **72**, 2384 (1980).

[39] F. Noé, J. Chem. Phys **128**, 244103 (2008).

[40] L. Verlet, Phys. Rev. **159**, 98 (1967).

[41] L. Verlet, Phys. Rev. **165**, 201 (1968).

[42] J. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[43] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, J. Chem. Theo. Comp. **3**, 26 (2007).

[44] G. R. Bowman, D. L. Ensign, and V. S. Pande, J. Chem. Theo. Comp. **6**, 787 (2010).

[45] M. R. Shirts and J. D. Chodera, J. Chem. Phys. **129**, 124105 (2008).

[46] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).

[47] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).

[48] A. I. Kinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, USA, 1949).

Appendix A: Efficient solution of the self-consistent equations for canonical distribution of Hamiltonian trajectories

For the case of a canonical distribution of Hamiltonian trajectories, the normalilazion constants Z_k or alternatively the dimensionless free energies $\hat{f}_i \equiv -\ln \hat{Z}_i$ are defined through a set of K coupled nonlinear equations involving the trajectory energies $E_{kn} \equiv$

$H(z_{kn}(0))$:

$$\hat{f}_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_k} \left[\sum_{k=1}^{K} N_k \, \exp[\hat{f}_k - (\beta_k - \beta_i) E_{kn}] \right]^{-1} \tag{A1}$$

Any numerically stable method for solving a set of coupled nonlinear equations can, in principle, be used to obtain the $\hat{f}_i$. A scheme for solving a more general form of these equations by self-consistent iteration or Newton-Raphson is described in Appendix C of [45].

Because of the structure of this specific case, we can rapidly obtain a close initial guess for the $\hat{f}_i$ by using a form inspired by the weighted histogram analysis method (WHAM) [46]. By instead constructing M bins in the total energy E spanning a range $(E_{\min}, E_{\max})$, we can approximate Eq. A1) with a sum over histograms (as in Eqs. 19–20 of [46]):

$$\hat{f}_i^{(n+1)} = -\ln \sum_{m=1}^{M} H_m \left[\sum_{k=1}^{K} N_k \, \exp[\hat{f}_k^{(n)} - (\beta_k - \beta_i) E_m \right]^{-1} \tag{A2}$$

where H_m denotes the number of samples E_{kn} falling in histogram bin m, and E_m represents the energy at the midpoint of that bin. Typically, a value of $M \approx 100$ can be used. One value, say f_1, is generally fixed to 0 by subtracting off the computed value of $f_1^{(n+1)}$ after each iteration in order to avoid numerical drift.

After an initial guess has been reached, self-consistent iteration can rapidly refine the free energies to the desired tolerance while eliminating the bias arising from the use of histograms:

$$\hat{f}_i^{(n+1)} = -\ln \sum_{k=1}^{K} \sum_{n=1}^{N} \left[\sum_{k=1}^{K} N_k \, \exp[\hat{f}_k^{(n)} - (\beta_k - \beta_i) E_{kn}] \right]^{-1} \tag{A3}$$

Again, we fix $\hat{f}_1 = 0$ and terminate iterations when a relative tolerance $\max_{i=2,\ldots,K} |f_i^{(n+1)} - f_i^{(n)}|/|f_i^{(n+1)} + f_i^{(n)}|$ is less than some given tolerance that ensures the computed expectations of properties of interest are no longer changing. We find that 10^{-7} is often a safe choice.

Cautions observed in Appendix C of [45] regarding sums of logarithms and numerical over/underflow in the evaluation of exponentials should be observed in implementation of this, or any, algorithm for obtaining the $\hat{f}_i$.

Appendix B: Proof that modified parallel tempering protocol generates canonical distribution

Here, we prove that the modified parallel tempering protocol described in Section II F samples from the canonical stationary distribution at all temperatures.

Define stationary distributions for momenta $\boldsymbol{p}$ and coordinates $\boldsymbol{q}$ in Cartesian space $\Re^{3N}$ at inverse temperature β:

$$\begin{aligned}\pi_p(\boldsymbol{p}|\beta) &= [P(\beta)]^{-1}\, e^{-\beta T(\boldsymbol{p})}; P(\beta) = \int d\boldsymbol{p}\, e^{-\beta T(\boldsymbol{p})} \\ \pi_q(\boldsymbol{q}|\beta) &= [Q(\beta)]^{-1}\, e^{-\beta U(\boldsymbol{q})}; Q(\beta) = \int d\boldsymbol{q}\, e^{-\beta U(\boldsymbol{q})}\end{aligned} \tag{B1}$$

where $T(\boldsymbol{p})$ denotes the kinetic energy and $U(\boldsymbol{q})$ the potential energy function. Suppose we have two replicas whose current phase space points are denoted by $\boldsymbol{z}_1 = (\boldsymbol{q}_1, \boldsymbol{p}_1)$ and $\boldsymbol{z}_2 = (\boldsymbol{q}_2, \boldsymbol{p}_2)$, initially at equilibrium at their respective inverse temperatures β_1 and β_2, such that

$$\begin{aligned}\boldsymbol{p}_1 &\sim \pi_p(\boldsymbol{p}_1|\beta_1); \boldsymbol{q}_1 \sim \pi_q(\boldsymbol{q}_1|\beta_1) \\ \boldsymbol{p}_2 &\sim \pi_p(\boldsymbol{p}_2|\beta_2); \boldsymbol{q}_2 \sim \pi_q(\boldsymbol{q}_2|\beta_2).\end{aligned} \tag{B2}$$

We now consider what happens to the distributions of $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ after an exchange attempt. Define "post-exchange attempt" coordinates and momenta for inverse temperature β_1:

$$\begin{aligned}\boldsymbol{q}_1' &\leftarrow \begin{cases} \boldsymbol{q}_1 & \text{with prob.} \quad 1 - \theta(\boldsymbol{q}_1, \boldsymbol{q}_2|\beta_1, \beta_2) \quad \text{(rejected)} \\ \boldsymbol{q}_2 & \text{with prob.} \quad \theta(\boldsymbol{q}_1, \boldsymbol{q}_2|\beta_1, \beta_2) \qquad \text{(accepted)} \end{cases} \\ \boldsymbol{p}_1' &\sim \pi_p(\boldsymbol{p}_1'|\beta_1) \quad \text{(velocity randomization)}\end{aligned}$$

where the exchange acceptance probability $\theta(\boldsymbol{q}_1, \boldsymbol{q}_2|\beta_1, \beta_2)$ is given by

$$\theta(\boldsymbol{q}_1, \boldsymbol{q}_2|\beta_1, \beta_2) = \min\{1, \exp[-\beta_1 U(\boldsymbol{q}_2) - \beta_2 U(\boldsymbol{q}_1) + \beta_1 U(\boldsymbol{q}_1) + \beta_2 U(\boldsymbol{q}_2)\} \tag{B3}$$

We now compute the distribution of $\boldsymbol{q}_1'$, the configuration supposedly at temperature β_1

after the exchange attempt:

$$
\begin{aligned}
\rho_1(\boldsymbol{q}_1') &= \int d\boldsymbol{q}_2 \, [1 - \theta(\boldsymbol{q}_1', \boldsymbol{q}_2|\beta_1, \beta_2)] \, \pi_q(\boldsymbol{q}_1'|\beta_1) \, \pi_q(\boldsymbol{q}_2|\beta_2) + \int d\boldsymbol{q}_2 \, \theta(\boldsymbol{q}_2, \boldsymbol{q}_1'|\beta_1, \beta_2) \, \pi_q(\boldsymbol{q}_2|\beta_1) \, \pi_q(\boldsymbol{q}_1'|\beta_2) \\
&= \int d\boldsymbol{q}_2 \, [1 - \min\{1, e^{-\beta_1 U(\boldsymbol{q}_2)} \, e^{-\beta_2 U(\boldsymbol{q}_1')} \, e^{+\beta_1 U(\boldsymbol{q}_1')} \, e^{+\beta_2 U(\boldsymbol{q}_2)}\}] \, \frac{e^{-\beta_1 U(\boldsymbol{q}_1')}}{Q(\beta_1)} \, \frac{e^{-\beta_2 U(\boldsymbol{q}_2)}}{Q(\beta_2)} \\
&+ \int d\boldsymbol{q}_2 \, \min\{1, e^{-\beta_1 U(\boldsymbol{q}_1')} \, e^{-\beta_2 U(\boldsymbol{q}_2)} \, e^{+\beta_1 U(\boldsymbol{q}_2)} \, e^{+\beta_2 U(\boldsymbol{q}_1')}\} \, \frac{e^{-\beta_1 U(\boldsymbol{q}_2)}}{Q(\beta_1)} \, \frac{e^{-\beta_2 U(\boldsymbol{q}_1')}}{Q(\beta_2)} \\
&= \frac{e^{-\beta_1 U(\boldsymbol{q}_1')}}{Q(\beta_1)} - \int d\boldsymbol{q}_2 \, \min\left\{ \frac{e^{-\beta_1 U(\boldsymbol{q}_1')}}{Q(\beta_1)} \, \frac{e^{-\beta_2 U(\boldsymbol{q}_2)}}{Q(\beta_2)}, \frac{e^{-\beta_1 U(\boldsymbol{q}_2)}}{Q(\beta_1)} \, \frac{e^{-\beta_2 U(\boldsymbol{q}_1')}}{Q(\beta_2)} \right\} \\
&+ \int d\boldsymbol{q}_2 \, \min\left\{ \frac{e^{-\beta_1 U(\boldsymbol{q}_2)}}{Q(\beta_1)} \, \frac{e^{-\beta_2 U(\boldsymbol{q}_1')}}{Q(\beta_2)}, \frac{e^{-\beta_1 U(\boldsymbol{q}_1')}}{Q(\beta_1)} \, \frac{e^{-\beta_2 U(\boldsymbol{q}_2)}}{Q(\beta_2)} \right\} \\
&= \pi_q(\boldsymbol{q}_1'|\beta_1)
\end{aligned}
\tag{B4}
$$

Therefore, after the exchange attempt, the new configuration $\boldsymbol{q}_1'$ is still at equilibrium at inverse temperature β_1. (A similar series of steps can be applied for the temperature β_2.)

Redrawing the momentum from the Maxwell-Boltzmann distribution at inverse temperature β_1 will, of course, not change the equilibrium distribution, and can be shown to only support the canonical distribution at inverse temperature β_1, and no other stationary distribution [47]. Evolution by Hamiltonian dynamics for any length of time does not alter the stationary canonical distribution [48]. Therefore, the proposed protocol samples from the canonical distribution at the desired temperatures, provided sufficient time is allowed for equilibration. ∎

Appendix C: Convergence of transition probabilities in Bayesian Methods

The convergence of transition probabilities from the Bayesian sampling methods is presented in Supplementary Figure 8 for various temperatures. Therefore, the proposed protocol generates the canonical distribution at the desired temperature.

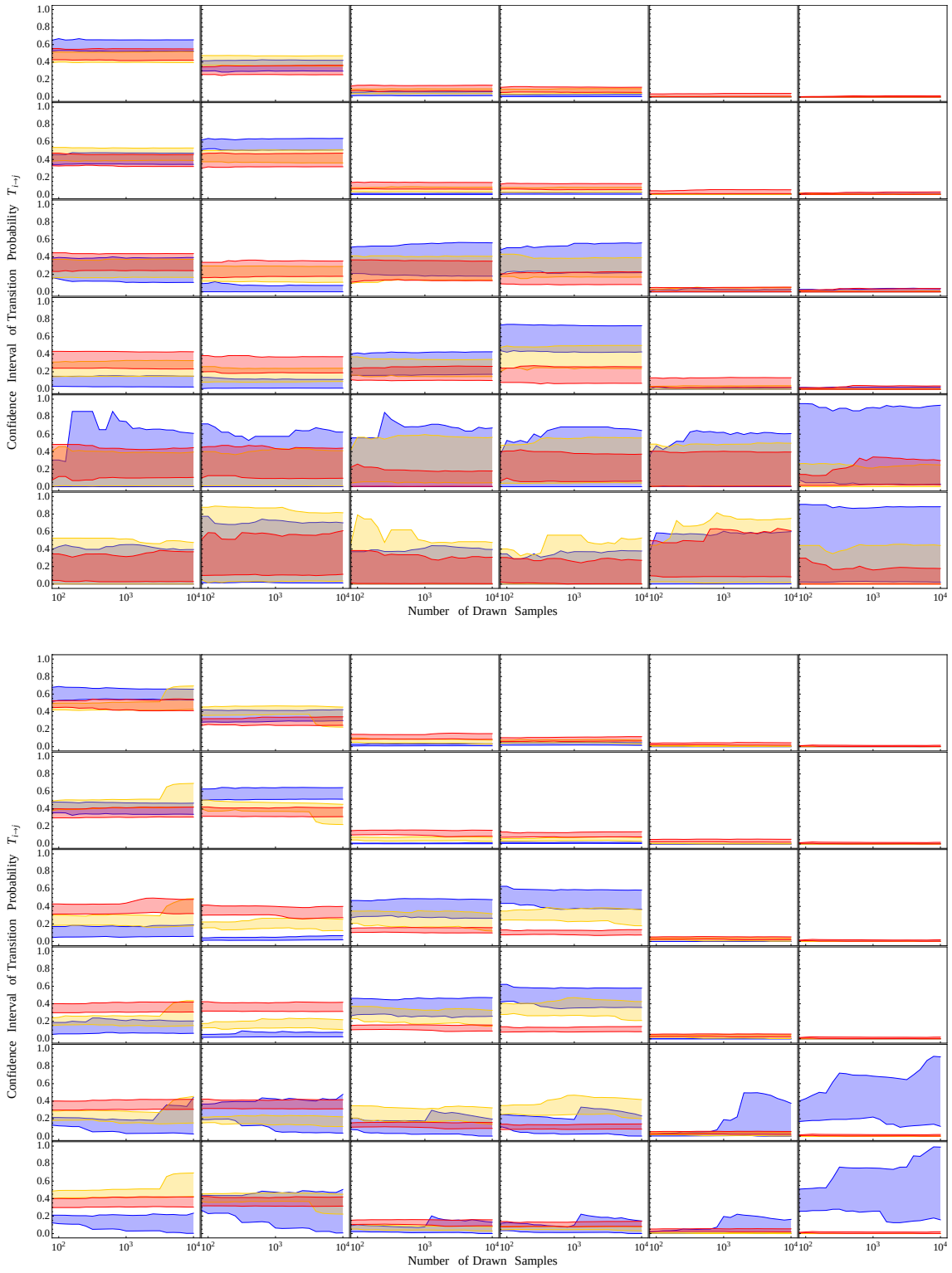

FIG. 8. 95% confidence intervals of Transition Probabilities sampled by the transition matrix estimation (upper Plot) (TE) and rate matrix estimation (lower plot) (RE) versus number of drawn samples. Color indicates performance by temperature. Blue: 273 K, Yellow: 426 K, Red: 600 K. After about 5 000 samples the confidence intervals stabilize suggesting reasonably well sampled transition probabilities.