Estimating the sampling error: Distribution of transition matrices and functions of transition matrices for given trajectory data

Philipp Metzner*
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA

Frank Noé† and Christof Schütte‡
Department of Mathematics and Computer Science, Free University Berlin, Arnimallee 6, D-14195 Berlin, Germany

(Received 22 November 2008; revised manuscript received 13 May 2009; published 13 August 2009)

The problem of estimating a Markov transition matrix to statistically describe the dynamics underlying an observed process is frequently found in the physical and economical sciences. However, little attention has been paid to the fact that such an estimation is associated with statistical uncertainty, which depends on the number of observed transitions between metastable states. In turn, this induces uncertainties in any property computed from the transition matrix, such as stationary probabilities, committor probabilities, or eigenvalues. Assessing these uncertainties is essential for testing the reliability of a given observation and also, if possible, to plan further simulations or measurements in such a way that the most serious uncertainties will be reduced with minimal effort. Here, a rigorous statistical method is proposed to approximate the complete statistical distribution of functions of the transition matrix provided that one can identify discrete states such that the transition process between them may be modeled with a memoryless jump process, i.e., Markov dynamics. The method is based on sampling the statistical distribution of Markov transition matrices that is induced by the observed transition events. It allows the constraint of reversibility to be included, which is physically meaningful in many applications. The method is illustrated on molecular dynamics simulations of a hexapeptide that are modeled by a Markov transition process between the metastable states. For this model the distributions and uncertainties of the stationary probabilities of metastable states, the transition matrix elements, the committor probabilities, and the transition matrix eigenvalues are estimated. It is found that the detailed balance constraint can significantly alter the distribution of some observables.

DOI: 10.1103/PhysRevE.80.021106 PACS number(s): 02.50.−r

I. INTRODUCTION

State transitions are essential to complex dynamical systems. In many such systems, the dynamics has multiple scales in time, and, on a given time scale of interest, the dynamics is metastable; i.e., there are regions in state space (metastable sets) within which transitions are rapid compared to the time scale of interest, while transitions between them are rare events. A prominent example for such systems is biomolecule [1], whose dynamics involve various processes such as binding of macromolecules and their ligands [2], complex conformational rearrangements switching between native protein substates [3,4] to the folding of proteins and RNA [5,6]. Further examples include Ising models [7], meteorological systems [8], and economic systems [9].

The slow transitions between metastable states are often well described by a memoryless jump process on a finite discrete state space, say, $S=\{1,\ldots,m\}$, i.e., with a memoryless master equation,

$$\frac{dp(t)}{dt} = p(t)L, \tag{1}$$

with $p(t)$ being an m-dimensional row vector containing the probability to find the system in each of its m states at time t. L is a rate matrix with $(L_{ij})_{i,j\in S}$ being the transition rate from state i to state j and the diagonal elements are $L_{ii}=-\Sigma_{j\neq i}L_{ij}$ to ensure probability mass conservation. Alternatively, the system dynamics can be described by a discrete-time Markov process using the transition matrix, $T(\tau)$, whose entries, T_{ij}, $i,j\in S$, provide the probability of the system to be found in state j at time $t+\tau$ given that it was in state i at time t. The time-discrete analog to Eq. (1) is the Chapman-Kolmogorov equation,

$$p(k\tau) = p(0)T^k(\tau). \tag{2}$$

Equations (1) and (2) provide equivalent results at discrete times $t=k\tau$, $k\in\mathbb{N}_0$ and are related by $T(\tau)=\exp(\tau L)$ [10]. Here, we will concentrate on the transition matrix $T(\tau)$ and Eq. (2). This transition matrix approach to molecular dynamics (MD) has been developed and successfully applied in a number of publications [11–16]. The memoryless ansatz implies that the dynamics $X(t)\in S$ between states is Markovian at time lag τ. In other words, the state of the system in the next time step, $t+\tau$, is assumed to only depend on the system's state at the current time t and not on its previous history,

$$p[X(t+\tau)|X(t)] = p[X(t+\tau)|X(t),X(t-\tau), X(t-2\tau),\ldots,X(0)].$$

In many cases it is not trivial to ensure Markovianity. The definition of states and the time lag τ need to be defined appropriately. However, this issue is beyond the present

*pmetzner@cims.nyu.edu
†noe@math.fu-berlin.de
‡schuette@math.fu-berlin.de

study and is addressed elsewhere [11,12,14,15].

Usually, $T(\tau)$ is not readily given but needs to be estimated from a set of observations or simulations such as molecular dynamics simulations. Since these simulations are of finite length, the estimated $\hat{T}(\tau)$ is associated with uncertainty. *For a given set of observed transitions from trajectory data, what is the uncertainty in $\hat{T}(\tau)$ and how does this affect the uncertainty of some function of $\hat{T}(\tau)$, say, $f[\hat{T}(\tau)]$? This question is addressed in the present paper.*

II. BAYESIAN FORMULATION

Consider one trajectory $Y=\{y_0=X(t=0),\ldots,y_N=X(t=N\tau)\}$ given. (The generalization to multiple trajectories is straightforward.) Let the frequency matrix $C(Y)=(C_{ij})_{i,j\in S}$ associated to a trajectory Y count the number of consecutively observed transitions between states, i.e., C_{ij} is the number of observed transitions from state i at time t to state j at time $t+\tau$, summed over all times t. In the limit of an infinitely long trajectory, the elements of the underlying transition matrix $T(\tau)$ are given by

$$T_{ij}(\tau)=\lim_{N\to\infty}\frac{C_{ij}}{\sum_{k\in S}C_{ik}},$$

where we dropped the dependency of the frequency matrix on the given trajectory Y for notational simplicity. For a trajectory of limited length, the underlying transition matrix $T(\tau)$ cannot be uniquely determined. The probability that a particular $T(\tau)$ would generate the observed trajectory is given by

$$p(Y|T)=\prod_{k=0}^{N-1}T_{y_k,y_{k+1}}=\prod_{i,j\in S}T_{ij}^{C_{ij}}. \tag{3}$$

In this paper we are interested in the opposite question: what is the probability $p(T|Y)$ that a particular transition matrix $T[=T(\tau)]$ has generated the observed data. By virtue of the Bayesian theorem it follows that the law of $p(T|Y)$ is proportional to $p(T)p(Y|T)$, where $p(T)$ is the prior probability of transition matrices. The particular choice of the prior reflects knowledge or reasonable assumptions on the set of all transition matrices *before* observing any data. Here we make the restriction that the prior probability can be written in the form $\Pi_{i,j\in S}T_{ij}^{B_{ij}}$ with some prior count matrix $B\in\mathbb{R}^{m\times m}$. Together with the likelihood $p(Y|T)$ in Eq. (3), the law of the posterior takes the form

$$P(T|Y)\propto p(T)p(Y|T)=\prod_{i,j\in S}T_{ij}^{B_{ij}+C_{ij}}=\prod_{i,j\in S}T_{ij}^{Z_{ij}},$$

where we have defined the effective count matrix $Z=B+C$. Notice that the posterior is fully characterized by the effective count matrix Z which we emphasize in the following by denoting the non-normalized probability density function (PDF) of $P(T|Y)$ by $p_Z(T)$, i.e.,

$$P(T|Y)\propto p_Z(T)=\prod_{i,j\in S}T_{ij}^{Z_{ij}}. \tag{4}$$

The specific form of the prior probabilities allows a number of common prior distributions to be used by simply adding a corresponding B matrix to the observed transition count matrix.

(1) *Uniform prior*. The uniform prior is simply given by using no prior counts, $B_{ij}=0$,

$$p_{Z,\text{uniform}}(T)=p_C(T)=\prod_{i,j\in S}T_{ij}^{C_{ij}}. \tag{5}$$

This prior distribution is used in *all* numerical experiments shown here. Notice again that for the uniform prior, $p_C(T)$ is fully characterized by the frequency matrix C associated with the observation Y.

(2) *Jeffrey's prior*. Jeffrey's prior is given by using $B_{ij}=-0.5$ for all i, j.

(3) *1/m prior*. The $1/m$ prior was suggested in [17] and ensures that there is a constant amount of information in the prior, independent of the size of the Markov model. It is given by using $B_{ij}=-1+m^{-1}$ for all i, j.

Defining $Z_i=\Sigma_{k\in S}Z_{ik}$ as the total number of effective transitions leaving the state i, it turns out that $\hat{T}(\tau)$, given by

$$\hat{T}_{ij}(\tau)=\frac{Z_{ij}}{Z_i}, \tag{6}$$

is the unique maximizer of $p_Z(T)$. In the case of a uniform prior ($B\equiv 0$), $\hat{T}(\tau)$ is also the unique maximizer of $P(Y|T)$ and, hence, is called *maximum likelihood estimator*. In the limit of infinite sampling, $p_Z(T)$ converges toward a delta distribution at $\hat{T}(\tau)$. When sampling is finite, the uncertainties of the entries of $\hat{T}(\tau)$ may be estimated by the element-wise standard deviations of $p_Z(T)$.

Example II.1. In a first example we illustrate the PDF $p_C(T)$ in Eq. (5) on a two-state Markov chain. Again, $p_C(T)$ is fully characterized by the frequency matrix C, e.g.,

$$C=\begin{pmatrix}5 & 2\\ 3 & 10\end{pmatrix}, \tag{7}$$

associated with a given finite observation Y. Let $T\in\mathbb{R}^{2\times 2}$ be a stochastic matrix, i.e.,

$$T=\begin{pmatrix}T_{11} & T_{12}\\ T_{21} & T_{22}\end{pmatrix}$$

with $T_{ij}\geq 0$, $1\leq i,j\leq 2$, and $T_{i1}+T_{i2}=1$, $i=1,2$. The non-normalized PDF $p_C(T)$ associated with the observation in Eq. (7) takes the form

$$p_C(T)=p_C(T_{11},T_{12},T_{21},T_{22})=T_{11}^5T_{12}^2T_{21}^3T_{22}^{10}.$$

Exploiting the stochasticity of T, $p_C(T)$ can be written as

$$p_C(T)=p_C(T_{12},T_{21})=(1-T_{12})^5T_{12}^2T_{21}^3(1-T_{21})^{10},$$

$$T_{12},T_{21}\in[0,1].$$

See Fig. 1 for an illustration of the transition matrix density function $p_C(T_{12},T_{21})$.

In general, one is interested to compute a particular property from the transition matrix, $f(T(\tau))$. f may represent any

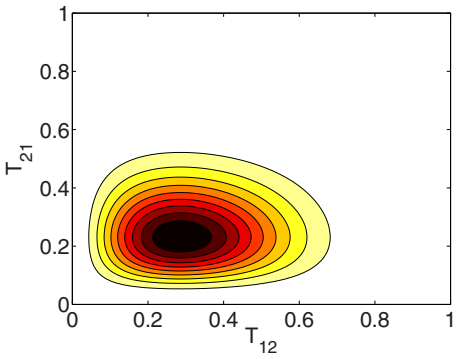

FIG. 1. (Color online) Probability distribution of 2×2 transition probability matrices for the observation given in Eq. (7). The resulting PDF, $p_C(T_{12},T_{21})=(1-T_{12})^5T_{12}^2T_{21}^3(1-T_{21})^{10}$, is shown in terms of the off-diagonal matrix elements. The color intensity encodes the probability density. The darker the color the higher the value of the probability density.

mathematical function, decomposition, or algorithm. In particular, we will consider the following properties here:

(1) The *stationary distribution*, π, which is the probability to be in each state in equilibrium and is given by the left one-normalized eigenvector of T to the eigenvalue 1.

(2) The *positive real eigenvalues*, $(\lambda_1,\ldots,\lambda_n)$, $n\leq m$ of $T(\tau)$, which indicate the time scales of the transition processes involved. The time scale implied by the ith positive real eigenvalue is given by

$$t_i^*=-\frac{\tau}{\ln(\lambda_i)}. \tag{8}$$

(3) The *committor*, q^{AB}, which is the probability, for each state i, that the system being in state i will go to state set B next rather than to state set A. In protein folding, if A and B correspond to the unfolded/folded state, q^{AB} denotes the probability of folding. The committor is computed *via*

$$\sum_j (T_{ij}-\delta_{ij})q_i^{AB}=0 \quad \forall\, i\in S\setminus(A\cup B),$$

$$q_i^{AB}=0 \quad \forall\, i\in A,$$

$$q_i^{AB}=1 \quad \forall\, i\in B, \tag{9}$$

with the Kronecker delta $\delta_{ij}=1$ for $i=j$ and 0 otherwise.

One is then interested how the uncertainty of the transition matrix, induced by the distribution $p(T|Y)$, carries over to uncertainties in the target function. In other words, for a given observation Y, what is the distribution of target functions, $p(f(T)|Y)$, and their variance?

An approach suggested in [17,18] is based on first-order perturbation theory: the posterior probability [Eq. (4)] is locally approximated by a multivariate Gaussian centered at the maximum, $\hat{T}(\tau)$, and the target function, $f(T)$, is approximated by a Taylor series truncated after the first term. The linear approximation of f preserves the Gaussian shape of the distribution, allowing the variance of $f(T)$ to be calculated analytically. This approach is very efficient in order to estimate the second moment of the sought distribution and thus in estimating the sampling error. Nevertheless, the approach makes some two approximations which may cause problems in practice:

(1) The method does not preserve stochasticity: the distribution of each transition matrix element, T_{ij}, is approximated by a Gaussian and thus allows for a finite probability for values <0 and >1, which are unphysical.

(2) It is unclear how well the first-order Taylor expansion will perform for various nonlinear target functions $f(T)$.

An alternative to employing linear error analysis is to generate an ensemble of transition matrices, according to the posterior probability [Eq. (4)] to compute the target functions $f(T)$ for each sampled T, thus sampling the distribution of $f(T)$. One approach, also suggested in [17,18], is to rewrite the posterior probability [Eq. (4)] as

$$p(T|Y)\propto\prod_{i\in S}\prod_{j\in S}T_{ij}^{Z_{ij}}=\prod_{i\in S}\prod_{j\in S}T_{ij}^{\alpha_{ij}-1},$$

where each factor $\Pi_{j\in S}T_{ij}^{\alpha_{ij}-1}$ has the form of a Dirichlet distribution with the parameters $\alpha_{ij}=Z_{ij}+1$. Efficient samplers for the Dirichlet distribution exist (see p. 594 in [19,20]). This approach ensures that all sampled transition matrices are stochastic matrices (i.e., $0\leq T_{ij}\leq1$ and $\Sigma_jT_{ij}=1$) drawn from the correct posterior probability. Unfortunately, this approach does not allow properties of T to be ensured which involve multiple rows. In particular, it is desirable to sample only transition matrices that fulfill detailed balance with respect to their stationary distribution, π,

$$\pi_iT_{ij}=\pi_jT_{ji}. \tag{10}$$

Such method was proposed in [21]. Here, an alternative and general method to sample transition matrices according to the posterior probability [Eq. (5)] based on Markov chain Monte Carlo (MCMC) is proposed. While it is computationally more expensive than the linear error analysis and the Dirichlet sampling, it allows the sampling to be restricted to transition matrices fulfilling additional constraints, such as Eq. (10).

III. MONTE CARLO SAMPLER FOR TRANSITION MATRICES

A Metropolis MCMC sampler is proposed. For notational convenience we denote the set of all transition matrices by

$$\mathcal{T}=\left\{T=(T_{ij})_{i,j\in\{1,\ldots,m\}}:T_{ij}\in[0,1],\ \sum_{k=1}^{m}T_{ik}=1 \quad \forall\, i,j\in\{1,\ldots,m\}\right\}.$$

The MCMC sampler will generate an ensemble of transition matrices drawn from $\mathcal{T}$ and distributed according to the posterior probability in Eq. (5).

Generally, a Metropolis MCMC scheme works as follows. Suppose that $p_Z(T)$ is the PDF to be sampled from and T_C is the current state. In the proposal step a new state T_N is generated with probability $p(T_C\rightarrow T_N)$. In the acceptance step the proposed state T_N is accepted with the probability

$$p_{acc} = \min\left\{1, \frac{p_Z(T_\mathcal{N})p(T_\mathcal{N} \to T_\mathcal{C})}{p_Z(T_\mathcal{C})p(T_\mathcal{C} \to T_\mathcal{N})}\right\}. \tag{11}$$

If the new state is accepted, then $T_\mathcal{N}$ is added to the ensemble and the scheme restarts with $T_\mathcal{N}$ as the current state. Otherwise, the current state $T_\mathcal{C}$ is added to the ensemble and is considered again in the next iteration of the scheme. This approach has a number of useful properties, including the following:

(1) The target density function $p_Z(T)$ does not need to be normalized as the normalization factor cancels in the ratio $p_Z(T_\mathcal{N})/p_Z(T_\mathcal{C})$ involved in the acceptance probability in Eq. (11). Thus, the proportionality factor in the posterior probability [Eq. (5)] does not need to be determined.

(2) In principle, any strategy for the generation of a new state in the proposal step may be used as long as the probabilities $p(T_\mathcal{C} \to T_\mathcal{N})$ and $p(T_\mathcal{N} \to T_\mathcal{C})$ can be evaluated and any two permitted states can be connected via a finite number of proposal steps. The choice of the proposal step strategy, however, is important for the efficiency and the convergence of the sampling procedure [22].

A. Monte Carlo in transition count matrix space

We will ensure the constraint $\Sigma_{j=1}^m T_{ij}=1$ while maintaining efficiency through a change in variables. For this, consider the matrices containing non-negative reals, $K \in \mathbb{R}_+^{m^2}$, and the transformation

$$T_{ij} = \frac{K_{ij}}{\sum_j K_{ij}} = \frac{K_{ij}}{K_i}$$

with $K_i=\Sigma_{j=1}^m K_{ij}$. This transformation maps K matrices to transition matrices. Considering Eq. (6), K may be interpreted as a matrix of fictitious transition counts and T as the corresponding maximum likelihood transition matrix. This mapping is formally written as the function

$$u(K) \overset{def}{=} \left(\frac{K_{11}}{K_1}, \ldots, \frac{K_{mm}}{K_m}\right) \in \mathcal{T} \tag{12}$$

such that $T=u(K)$.

The crucial idea is now to generate an ensemble of count matrices $\mathcal{K}=\{K \in \mathbb{R}_+^{m^2}\}$ via an MCMC procedure which is distributed according to the PDF $p_Z(T)$. For this approach to be valid, the mapping $u(K)$ must be such that the ensemble $\mathcal{T}=\{T=u(K), K \in \mathcal{K}\}$ is distributed according to $p_Z(T)$. This is indeed the case and it is established by Theorem V.1, stated in the Appendix. Particularly, we show that the ensemble of count matrices $\mathcal{K}$ has to be restricted on the subset

$$\mathfrak{K} = \left\{K \in R_+^{m^2} : k_i^- \le \sum_{j=1}^m K_{ij} \le k_i^+, \quad i = 1, \ldots, m\right\}, \tag{13}$$

with $0<k_i^-<k_i^+$, $i=1,\ldots,m$. The restriction on the set $\mathfrak{K}$ is independent of the proposal step and accounts for the noninvertibility of the transformation $u(K)$. Furthermore, it ensures the right statistical weights of the transition matrices in the ensemble $\mathcal{T}=\{T=u(K), K \in \mathcal{K}\}$.

Input: Count matrix $Z = (Z_{ij})_{i,j \in S}$,
set of boundary constants $\{(k_i^-, k_i^+)\}$, $1 \le i \le m$.

Output: Ensemble $\mathcal{T}$ of transition matrices.

(1) Initialize $K_\mathcal{C}$ with a nonnegative matrix, e.g., Z.

(2) Loop until convergence:

(2.1) Draw uniformly pair of indices $(i,j) : 1 \le i,j \le m$.

(2.2) Draw uniformly
$$\epsilon \in \left[\max\left\{-(K_\mathcal{C})_{ij}, k_i^- - (K_\mathcal{C})_i\right\}, k_i^+ - (K_\mathcal{C})_i\right].$$

(2.3) Generate proposal $K_\mathcal{N}$:
$$(K_\mathcal{N})_{kl} = \begin{cases} (K_\mathcal{C})_{ij} + \epsilon & \text{if } (k,l) = (i,j), \\ (K_\mathcal{C})_{kl} & \text{otherwise.} \end{cases}$$

(2.4) Accept $K_\mathcal{N}$ with acceptance probability:
$$p_{acc} = \min\left\{1, \frac{p_Z(u(K_\mathcal{N}))}{p_Z(u(K_\mathcal{C}))}\right\}.$$

(2.5) If $K_\mathcal{N}$ is accepted then set $K_\mathcal{C} \leftarrow K_\mathcal{N}$.

(2.6) Add $u(K_\mathcal{C})$ to the transition matrix ensemble $\mathcal{T}$.

FIG. 2. Metropolis algorithm: general case.

Let (i,j), $1 \le i,j \le m$, be a uniformly drawn pair of indices. We suggest the following proposal step scheme for $K_\mathcal{N} = [(K_\mathcal{N})_{kl}]$, $1 \le k, l \le m$:

$$(K_\mathcal{N})_{kl} = \begin{cases} (K_\mathcal{C})_{ij} + \epsilon & \text{if } (k,l) = (i,j) \\ (K_\mathcal{C})_{kl} & \text{otherwise,} \end{cases} \tag{14}$$

where the random variable ϵ is drawn such that the constraints,

$$0 \le (K_\mathcal{N})_{ij} \quad \text{and} \quad k_i^- \le \sum_{k=1}^m (K_\mathcal{N})_{ik} \le k_i^+, \tag{15}$$

are satisfied. This is achieved by drawing ϵ uniformly from the interval

$$[a,b] = [\max\{-(K_\mathcal{C})_{ij}, k_i^- - (K_\mathcal{C})_i\}, k_i^+ - (K_\mathcal{C})_i], \tag{16}$$

where $(K_\mathcal{C})_i=\Sigma_{k=1}^m (K_\mathcal{C})_{ik}$. Consequently, the proposal probabilities simply reduce to

$$p(K_\mathcal{C} \to K_\mathcal{N}) = p(K_\mathcal{N} \to K_\mathcal{C}) = \frac{1}{b-a}.$$

The algorithm in Fig. 2 summarizes our approach to generating an ensemble of transition matrices distributed according to $p_Z(T)$.

We end this section with a discussion of the computational cost of our proposed scheme. The computational cost of a single iteration step is dominated by the evaluation of the ratio $p_Z[u(K_\mathcal{N})]/p_Z[u(K_\mathcal{C})]$ for the acceptance probability. Since the update of $K_\mathcal{C}$ in the proposal step affects only one entry the evaluation of p_{acc} can efficiently be performed in $\mathcal{O}(m)$. The overall memory requirement scales with $\mathcal{O}(|\mathcal{T}|m^2)$, where $|\mathcal{T}|$ is the size of the transition matrix ensemble. However, the overall performance and memory re-

quirement depend crucially on the matrix function $f(T)$ under consideration. If both scale reasonably, e.g., with $\mathcal{O}(m)$, then the proposed MCMC scheme can even be applied on state spaces with $m \approx 1000$ states.

B. Sampling reversible transition matrices

In this section we present a MCMC sampling scheme which allows us to sample *reversible* transition matrices distributed according to posterior (5). The scheme is based on the following simple observation: if $K \in \mathbb{R}^{m \times m}$ is a symmetric non-negative matrix then the transition matrix $T=u(K)$ [see Eq. (12)] is reversible with respect to the probability distribution,

$$\pi = \left(\frac{\sum_{j=1}^{m} K_{1j}}{\sum_{i,j=1}^{m} K_{ij}}, \ldots, \frac{\sum_{j=1}^{m} K_{mj}}{\sum_{i,j=1}^{m} K_{ij}} \right).$$

Hence, the symmetry of the proposed count matrices needs to be ensured in the proposal step. Furthermore, to ensure the correct statistical weights, the MCMC sampling has to be restricted on the set

$$\mathfrak{K}_{sym} = \left\{ K \in R_+^{m^2} : K_{ij} = K_{ji} \quad \forall\, i,j \in \{1, \ldots, m\}, k^- \leq \sum_{i,j=1}^{m} K_{ij} \leq k^+ \right\}, \tag{17}$$

with $0 < k^- < k^+ < \infty$. For the particular choice of the set $\mathfrak{K}_{sym}$ see Theorem V.3 in the Appendix.

Let (i,j), $1 \leq i,j \leq m$, be a uniformly drawn pair of indices. We propose the following proposal step scheme for a symmetric proposal $K_{\mathcal{N}} = [(K_{\mathcal{N}})_{kl}]$, $1 \leq k, l \leq m$:

$$(K_{\mathcal{N}})_{kl} = \begin{cases} (K_{\mathcal{C}})_{ij} + \epsilon & \text{if } (k,l) \in \{(i,j),(j,i)\} \\ (K_{\mathcal{C}})_{kl} & \text{otherwise,} \end{cases} \tag{18}$$

where ϵ is uniformly drawn from the interval

$$[a,b] = \begin{cases} \left[\max\left\{ -(K_{\mathcal{C}})_{ij}, \frac{1}{2}(k^- - S_{\mathcal{C}}) \right\}, \frac{1}{2}(k^+ - S_{\mathcal{C}}) \right] & \text{if } i \neq j \\ \left[\max\{ -(K_{\mathcal{C}})_{ij}, k^- - S_{\mathcal{C}} \}, k^+ - S_{\mathcal{C}} \right] & \text{if } i = j, \end{cases} \tag{19}$$

with $S_{\mathcal{C}} = \Sigma_{k,l=1}^{m} (K_{\mathcal{C}})_{kl}$.

The proposal scheme in Eq. (18) together with Eq. (19) guarantees that $K_{\mathcal{N}} \in \mathfrak{K}_{sym}$ if $K_{\mathcal{C}} \in \mathfrak{K}_{sym}$. Analogously to Eq. (14), the proposal probability $p(K_{\mathcal{C}} \to K_{\mathcal{N}})$ is symmetric. Finally, the algorithm in Figure 3 summarizes our approach to generating an ensemble of reversible transition matrices distributed according to $p_Z(T)$.

Input: Count matrix $Z = (Z_{ij})_{i,j \in S}$,
boundary constants k^-, k^+.

Output: Ensemble $\mathcal{T}$ of reversible transition matrices.

(1) Initialize $K_{\mathcal{C}}$ with a symmetric and nonnegative matrix, e.g., $\frac{1}{2}(Z + Z^T)$.

(2) Loop until convergence:

(2.1) Draw uniformly pair of indices $(i,j) : 1 \leq i,j \leq m$.

(2.2) Draw uniformly $\epsilon \in [a,b]$ $(S_{\mathcal{C}} = \sum_{k,l=1}^{m} (K_{\mathcal{C}})_{kl})$

$$[a,b] = \begin{cases} \left[\max\left\{ -(K_{\mathcal{C}})_{ij}, \frac{1}{2}(k^- - S_{\mathcal{C}}) \right\}, \frac{1}{2}(k^+ - S_{\mathcal{C}}) \right] & \text{if } i \neq j, \\ \left[\max\{ -(K_{\mathcal{C}})_{ij}, k^- - S_{\mathcal{C}} \}, k^+ - S_{\mathcal{C}} \right] & \text{if } i = j. \end{cases}$$

(2.3) Generate proposal $K_{\mathcal{N}}$:

$$(K_{\mathcal{N}})_{kl} = \begin{cases} (K_{\mathcal{C}})_{ij} + \epsilon & \text{if } (k,l) \in \{(i,j),(j,i)\}, \\ (K_{\mathcal{C}})_{kl} & \text{otherwise.} \end{cases}$$

(2.4) Accept $K_{\mathcal{N}}$ with acceptance probability:

$$p_{acc} = \min\left\{ 1, \frac{p_Z(u(K_{\mathcal{N}}))}{p_Z(u(K_{\mathcal{C}}))} \right\}.$$

(2.5) If $K_{\mathcal{N}}$ is accepted then set $K_{\mathcal{C}} \leftarrow K_{\mathcal{N}}$.

(2.6) Add $u(K_{\mathcal{C}})$ to the transition matrix ensemble $\mathcal{T}$.

FIG. 3. Metropolis algorithm: reversible case.

IV. EXAMPLES

A. Distributions of nonreversible and reversible T matrices for a three-state system

To illustrate the sampling algorithms, first a three-state system is considered. Let C, given by

$$C = \begin{pmatrix} 1 & 10 & 2 \\ 2 & 26 & 3 \\ 15 & 20 & 20 \end{pmatrix}, \tag{20}$$

be a count matrix associated with a fictitious observation of a three-state Markov chain. We compare the exact PDF $p_C(T)$ in Eq. (5) with the sampled distribution of 3×3 transition matrices (compare Example II.1). In the first column of Fig. 4 we illustrate three different two-dimensional projections of $p_C(T)$. For example, panel A illustrates the marginal PDF,

$$\tilde{p}_Y(T_{12}, T_{13}) = [1 - (T_{12} + T_{13})]^1 T_{12}^{10} T_{13}^2,$$

with $0 \leq T_{12}, T_{13} \leq 1$, and $0 \leq 1 - (T_{12} + T_{13}) \leq 1$. The panels in the second column show the corresponding distributions resulting from sampling of $p_C(T)$. To be more precise, we generated an ensemble of 10^6 transition matrices by means of the algorithm in Fig. 2 and computed two-dimensional histograms, respectively, depicted as contour plots. As an initial transition count matrix we chose $K_{\mathcal{C}} \equiv 1/3$ and we set $k^- \equiv 0.9$ and $k^+ \equiv 1.1$.

Finally, we sampled $p_C(T)$ restricted on the set of reversible 3×3-transition matrices by means of the algorithm in Fig. 3. We chose the same initial transition count matrix as in the nonreversible case and generated an ensemble of 10^6 transition matrices $[k^- = \Sigma_{i,j=1}^{m} (K_{\mathcal{C}})_{i,j}$ and $k^+ = 1.05 \Sigma_{i,j=1}^{m} (K_{\mathcal{C}})_{i,j}]$. The resulting two-dimensional marginal distributions are depicted in the panels C, F, and I. One can clearly see that the marginal distributions illustrated in panels C and F significantly differ from the corresponding the nonreversible ones, respectively.

B. Nonreversible versus reversible sampling

To illustrate the effect of enforcing reversibility, let us consider the reversible five-state transition matrix shown in

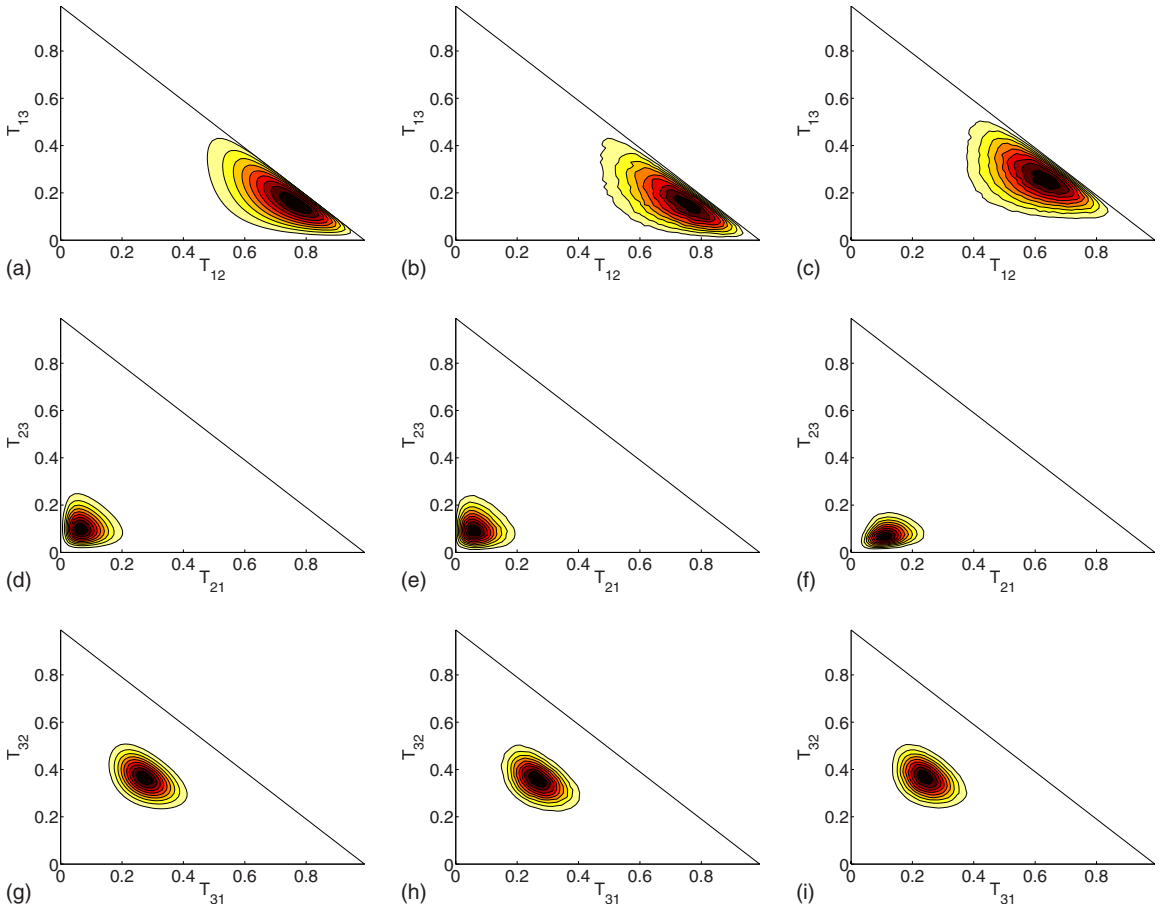

FIG. 4. (Color online) Visualization of the probability density of transition matrices to the observation in Eq. (20). Different two-dimensional marginal distributions (see axes) are shown in the columns. The exact and sampled distributions for stochastic matrices are shown in columns 1 and 2, respectively. Column 3 shows the sampled distribution for stochastic matrices fulfilling detailed balance.

Fig. 5(a). Each state in the series 1-2-3-4-5 has a stationary probability ten times greater than the previous one, such that the stationary probabilities of states 1 and 5 relate as 1:10 000. The 1000 transition counts are generated for each state and are distributed according to the transition probabilities [see Fig. 5(b)]. Based on these transition counts, the distribution of transition matrices is sampled with and without enforcing reversibility. These two distributions are almost identical in all transition matrix elements except for T_{51}. This transition probability is very low ($T_{51}=10^{-5}$) and no transition has been observed, such that the only pieces of information available to bound the value of T_{51} are the number of failed attempts to observe that transition (1000 times) and constraints on the ensemble of transition matrices (stochasticity and reversibility). Figure 5(c) shows that without reversibility, the distribution of T_{51} is very wide and its peak is much larger than the true value $T_{51}=10^{-5}$. Enforcing reversibility strongly sharpens the distribution and its peak is rather close to the true value. This is explained by the fact that enforcing reversibility allows us to estimate a transition probability $i \to j$ based on the estimate of the backward transition probability $j \to i$ if a good estimate for the relative stationary probabilities of i and j is available. In the present example,

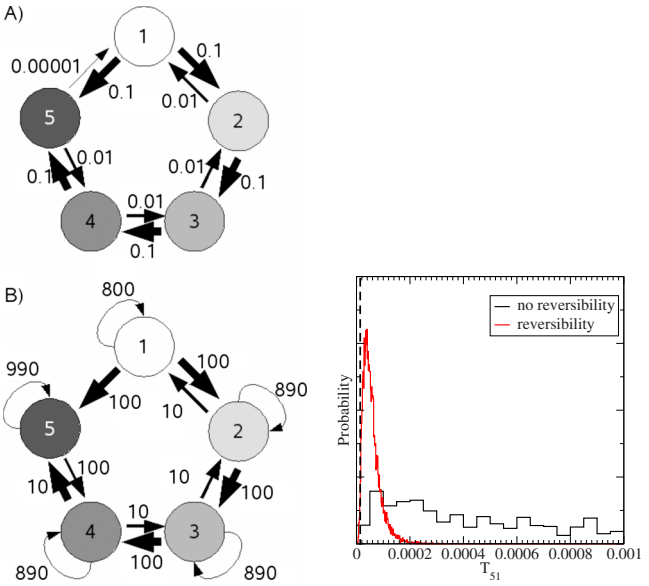

FIG. 5. (Color online) (A) a reversible five-state transition matrix. The states with higher stationary probability are shown in darker colors. (B) The transition counts used for sampling the distribution of transition matrices. (C) The distribution of the transition matrix element T_{51} with and without enforcing reversibility.

$$T_{15} = T_{51}\frac{\pi_5}{\pi_1}. \tag{21}$$

The relative probabilities π_5/π_1 are well estimated because the sufficient number of counts is available along the chain 1-2-3-4-5, and T_{51} can also be well estimated. Of course, enforcing reversibility is only allowed if it is known that the underlying system has reversible dynamics.

C. Probability distribution of spectrum and stationary distribution and their dependence on the observation length

In computer simulations of stochastic systems, such as MD simulations, the matrix of observed transition counts, C, depends on the length of the observations. Upon lengthening the simulation, more transitions will be observed and the implied distribution of transition matrices, $p_C(T)$, will become narrower. In a similar fashion, properties computed from T will become generally more accurate as the length of the simulation, and thus the C matrix, is increased. To study this effect, consider a four-state system defined by the true transition matrix,

$$T = \begin{pmatrix} 0.7 & 0.3 & & \\ 0.5 & 0.47 & 0.03 & \\ & 0.01 & 0.96 & 0.03 \\ & & 0.03 & 0.97 \end{pmatrix}. \tag{22}$$

Starting in state 4, we generated a realization according to T, with the total length of 2000, and considered the first 100, 500, 1000, 5000, and 20 000 steps. For the C matrices corresponding to each of these chains, the nonreversible transition matrix distribution is sampled to convergence using the algorithm in Fig. 2 where the parameters were chosen analogously as in the previous example (see Sec. IV A).

A particular interesting property of T is its spectrum, i.e., its eigenvalues λ_i with $i \in \{2,3,4\}$, which indicate the time scales of the transition processes in the system, t_i^*, via Eq. (8). The first eigenvalue, which is always $\lambda_1 = 1$, is irrelevant in this respect as it only represents the fact that the system as a whole is never left ($t_1^* = \infty$). The next time scales $t_2^*, t_3^*, \ldots$ correspond to the time scales of the slowest and next-slowest transition processes. Since there is a distribution of T, there is not a unique eigenvalue spectrum for a given observed transition count C but rather a spectral distribution. With an increased number of observed transition counts, the uncertainties of individual λ_i will decrease, thus allowing for some of these λ_i to be distinguished from the rest of the spectral distribution. Figure 6 shows the distribution of the two eigenvalues $1 > |\lambda_2| \geq |\lambda_3|$ with second- and third-highest orders of magnitude for the four simulations of different lengths. For the 100 step simulation, the spectral distributions do not exhibit any distinctive features except a broad peak. For 500 and 1000 steps, it is apparent that the distribution starts to separate into two distinct eigenvalues close to 1 and for 5000 steps these two modes of the distribution are clearly separated and closely located around the true eigenvalues of T which are $\lambda_2 = 0.9857$ and $\lambda_3 = 0.9336$ (indicated by a disk and a triangle on the x axis, respectively).

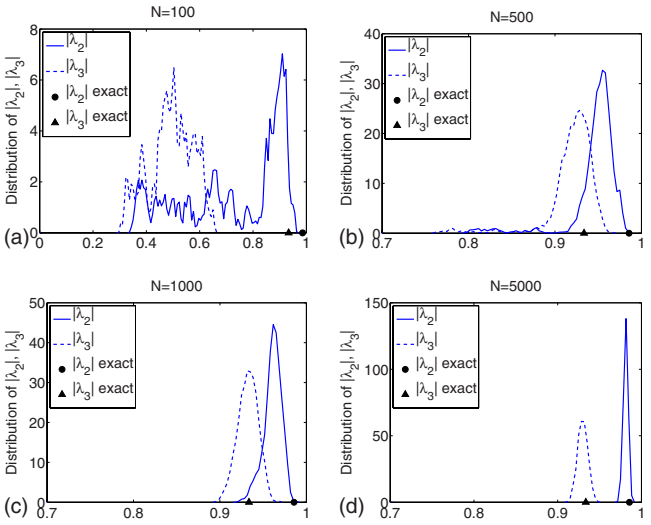

FIG. 6. (Color online) Distribution of eigenvalues close to 1 for different realizations of the Markov chain induced by the transition matrix T given in Eq. (22). (A) Realization of length $N=100$. (B) Length $N=500$. (C) $N=1000$. (D) $N=5000$. Observe that the spectrum becomes sharper distributed the larger N.

Next, the stationary distribution of T is estimated from the four differently long simulations (see Fig. 7). One can clearly see that the distributions shift toward the correct values and attain a Gaussian-like shape as the length of the realization increases from $N=500$ to $N=5000$. After 20 000 steps, the distributions are close to convergence and their peaks are located around the true values (indicated by triangles on the x axis).

D. Example from molecular dynamics: A 33-state system

In order to illustrate the transition matrix sampling on a realistic example, a 1 μs molecular dynamics (MD) simula-

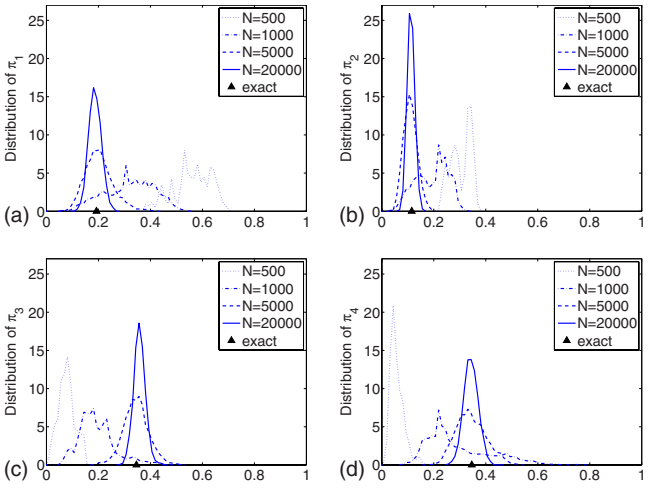

FIG. 7. (Color online) Distribution of stationary probabilities of states $i=1,\ldots,4$ for different realizations of length $N=500$, 1000, 5000, and 20 000 of the Markov chain induced by the transition matrix T given in Eq. (22). As the length of the realization increases the stationary distribution gradually shifts toward its limit values (indicated by triangles on the x axis).

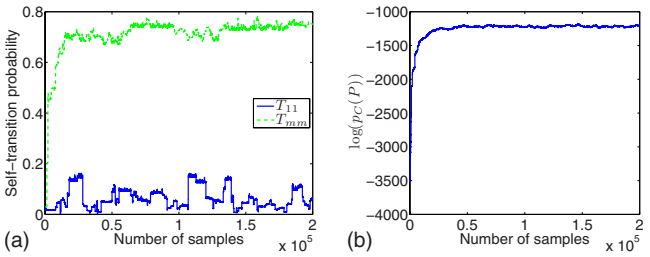

FIG. 8. (Color online) The time trace of the self-transition probabilities T_{11} and T_{mm} (panel A) and the log-likelihood function $\log[p_C(T)]$ (panel B). Both time traces suggest that the Markov chain is well mixed after 200 000 steps.

tion of the synthetic hexapeptide MR121-GSGSW peptide [23] in explicit water is used. The simulation setup is described in the Appendix. In order to concentrate on the metastable dynamics, 33 metastable states were identified and the interconversion between them was described with a Markov model using a lag time of $\tau=1$ ns. See Appendix for the detailed description of the Markov model construction. By counting the transitions between metastable conformations at time intervals of 1 ns along the trajectory, the transition count matrix, $C \in \mathbb{N}_0^{33\times 33}$, is obtained which serves as a test case for the sampling procedure.

Our proposed method estimates the distribution of transition matrices via Monte Carlo sampling and, hence, allows us to estimate the uncertainty of observables. Particularly, the following observables are chosen:

(i) the self-transition probabilities T_{11} and T_{mm} corresponding to the least and most-populated states, respectively,

(ii) the first two nontrivial dominant eigenvalues $1>|\lambda_2| \geq |\lambda_3|$ of T,

(iii) the stationary probabilities $\pi(1)$ and $\pi(m)$ corresponding to the least and most-populated states, and

(iv) the committor probability $q(i_{1/2})$ corresponding to a 1/2-committor state where $i_{1/2}$ is defined by

$$i_{1/2} = \arg\min_{i\in S}\{|\hat{q}(i) - 0.5|\}.$$

The committor function $\hat{q}(\cdot)$ satisfies Eq. (9) with respect to the maximum likelihood estimator $\hat{T}$ given in Eq. (6). The set A and B consists of the respective state which corresponds to the unfolded and folded states.

Now a couple of questions arise: (i) what is the burn-in time of the MCMC sampling scheme, i.e., how many samples are necessary to consider the underlying Markov chain to be stationary, and (ii) how many samples are necessary to consider the estimated distributions as "correct" (converged).

To answer the first question, we considered the time trace of the self-transition probabilities T_{11}, T_{mm} and the log-likelihood function $\log[p_C(T)]$ as a result of the algorithm in Fig. 2. As initial count matrix we chose $K \equiv 1/33$ and set $k^- \equiv 1$ and $k^+ \equiv 2$. The time traces depicted in the panels of Fig. 8 suggest that the Markov chain is well mixed after 200 000 steps. Moreover, to guarantee decorrelated samples, we took every 1000th sample of the continued sampling and stopped after a total of 100 000 samples.

Next, we address the question of convergence. Among different approaches to assess convergence of a MCMC simulation, we employed the method of Gelman and Rubin [24] where we followed the presentation in [25]. Suppose n different chains have been simulated each with different start points with respect to the distribution to be sampled. Roughly spoken, the idea of the Gelman-Rubin method is to infer convergence from comparing the between-chain statistics and the within-chain statistics in terms of the variance. In practice, the so-called *potential scale factor* $\hat{R}$ is computed which indicates convergence if $\hat{R}$ is close to 1. Instead of launching different simulations, here we divided one long single-run simulation into 100 equally sized pieces and calculated $\hat{R}$ for each observable. The resulting $\hat{R}$ values for some observables are given in the first row in Table I and indeed indicate convergence of the sampling.

The algorithm in Fig. 3 allows us to sample $p_C(T)$ subject to the detailed balance constraint in Eq. (10). Hence, we asked if restricting on reversible transition matrices results in significant differences of the observables' distributions. To make things comparable, we chose the same initial count matrix, burn-in time, thinning, and sampling length as in the nonreversible simulation where we used the boundary parameters $k^-=1$ and $k^+=1.01$.

E. Uncertainty in molecular dynamics simulations

Next, the effect of simulation length on the uncertainties of T itself is studied. For this, segments of the complete 1 μs trajectory were considered, starting at time 0 and having lengths between 10 and 1000 ns. For each segment, the transitions between states were counted using always the same definition of states. For each C matrix obtained in this way, the T matrices were sampled without and with the detailed balance constraint. Figure 9 shows the mean uncertainties of the diagonal elements in panel (a) and the off-diagonal elements in panel (b). All uncertainties become smaller with

TABLE I. Convergence diagnostic due to Gelman and Rubin [24] in the nonreversible case with respect to the observation C. We divided a single-run simulation with sampling size of 100 000 into 100 equally sized pieces and calculated the potential scale factor $\hat{R}$ for some observables, respectively. Since all $\hat{R}$ values are close to 1, we consider the sampling to be converged.

	$\hat{R}(T_{11})$	$\hat{R}(T_{mm})$	$\hat{R}(\pi_1)$	$\hat{R}(\pi_m)$	$\hat{R}[q(i_{1/2})]$
Nonreversible	1.0049	1.0183	1.0018	1.0087	1.0032
Reversible	1.0047	1.0177	1.0029	1.0578	1.0152

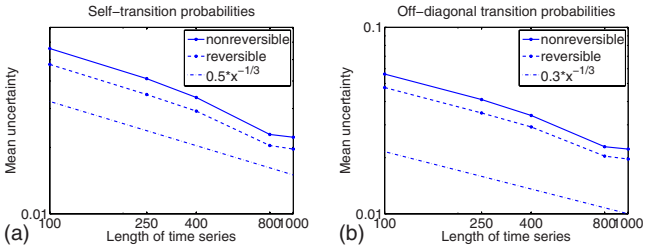

FIG. 9. (Color online) Mean uncertainties of (panel A) the diagonal and (panel B) the off-diagonal elements of the transition matrix for different simulation lengths. The uncertainties are shown for the ensembles of transition matrices (nonreversible) and transition matrices with detailed balance (reversible).

increasing trajectory length. The decay of the uncertainty follows roughly a power law with increasing simulation length and is slower than the $t^{-0.5}$ which would be expected from uncorrelated samples. Introducing detailed balance increases the uncertainty in the diagonal elements but decreases the uncertainty in the off-diagonal element. This is to be expected since detailed balance constraints only the ratio of symmetric off-diagonal elements.

Next, the effect of simulation length on the uncertainties of properties derived from the transition matrix is analyzed. First, consider the stationary probabilities of the metastable states, as provided by the first left eigenvector of $T(\tau)$. The stationary probabilities were computed for each sample $T(\tau)$. In order to avoid the average to be dominated by the few most-populated states, the mean relative uncertainty was computed via

$$\bar{\sigma}_r(\pi_i) = \frac{1}{m}\sum_{i=1}^{m} \frac{\sigma(\pi_i)}{\mu(\pi_i)}, \tag{23}$$

where $\mu(\pi_i)$ and $\sigma(\pi_i)$ are the means and standard deviations of the stationary probability of state i, respectively. These mean relative uncertainties are shown in Fig. 10. It is apparent that introducing detailed balance has almost no effect on the uncertainties in the stationary probabilities.

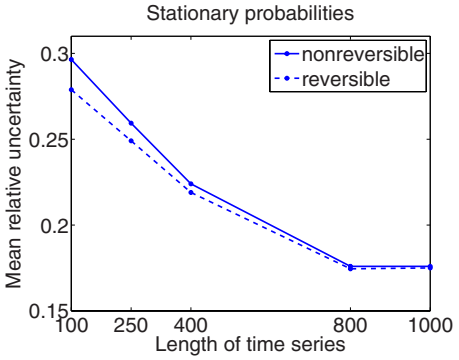

FIG. 10. (Color online) State-averaged standard deviations of the stationary probabilities for different simulation lengths. The uncertainties are shown for the ensembles of transition matrices (nonreversible) and transition matrices with detailed balance (reversible).

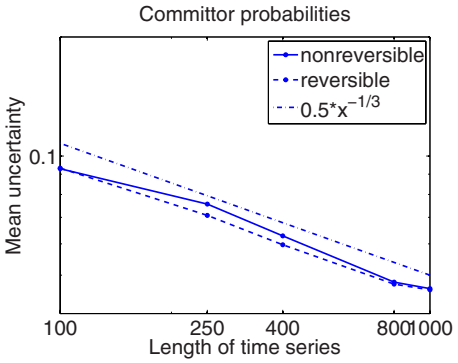

FIG. 11. (Color online) State-averaged relative standard deviations of the committor probabilities for different simulation lengths. The uncertainties are shown for the ensembles of transition matrices (nonreversible) and transition matrices with detailed balance (reversible).

Another property of interest is the committor probability of each state, with respect to two end states A and B. Here, the two metastable states with the most distant values in the second eigenvector were chosen as A and B, thus representing the states between which the slowest transition in the system occurs. The committor was computed for each sample of $T(\tau)$ using Eq. (9) and its mean relative uncertainties are shown in Fig. 11. This shows that the uncertainties of a property derived from $T(\tau)$ decrease with increasing simulation time, even if new states being found as the simulation proceeds. It is apparent that incorporating detailed balance somewhat reduces the uncertainties of the committor. This is expected since the committor is a dynamical property and thus benefits from the reduced uncertainty in the off-diagonal elements of the transition matrix observed in Fig. 9(b). Similarly as the off-diagonal transition matrix elements, the committor uncertainties decay with increasing simulation time approximately by a power law that is slower than $t^{-0.5}$.

Finally, another interesting property of T is its spectrum, i.e., its eigenvalues λ_i with $i \in \{1, \ldots, m\}$, which indicate the time scales of the transition processes in the system, t_i^*, via Eq. (8). Due to the distribution of T, there is not a unique eigenvalue spectrum for a given observed transition count C but rather a spectral distribution. With an increased number of observed transition counts, the uncertainties of individual λ_i will decrease, thus allowing for some of these λ_i to be distinguished from the rest of the spectral distribution. Figure 12 shows the spectral distribution for several simulation lengths. For simulation times up to 100 ns, the spectral distribution has no distinctive features. With increasing simulation time, some peaks at the large-eigenvalue region start to form. From 400 ns on, the slowest transition process at $\lambda_2 \approx 0.75$ can be clearly distinguished and continues to narrow with yet increasing simulation time. At 1000 ns, the spectrum exhibits a lot of structure in the range $\lambda \geq 0.5$, but apart from λ_2 no peaks are clearly separated. This indicates that even for a small peptide, 1 μs simulation time is rather short when good convergence of the kinetics is expected. Introducing detailed balance somewhat shifts large-eigenvalue (slow time) range of the spectrum to the right. In order to better see how the uncertainty of individual eigenvalues changes with

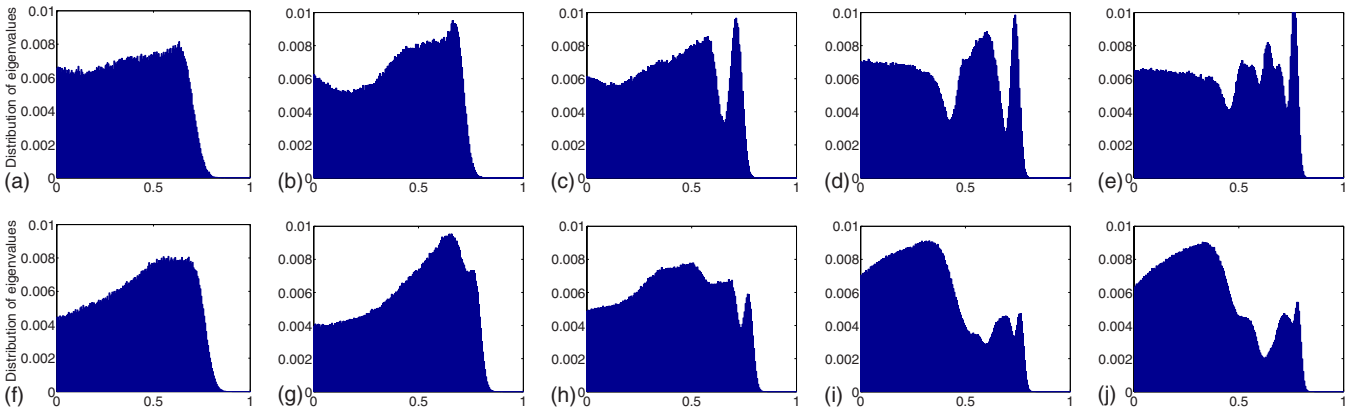

FIG. 12. (Color online) Distributions of the eigenvalue spectrum of T for different simulation lengths. The distributions are shown for the ensembles of transition matrices (nonreversible, first row) and transition matrices with detailed balance (reversible, second row).

sampling length, the relative uncertainties of second- and third-largest eigenvalues are plotted in Fig. 13. The uncertainties decay roughly with $t^{0.5}$ with increasing simulation time t, which is reasonable for a Markovian system as the estimate of the eigenvalues depends on the number of observed transitions along the corresponding eigenvector which itself is proportional to the sampling time. Again, introducing detailed balance significantly reduces the uncertainties since the eigenvalues are dynamic properties.

V. CONCLUSION

Methods were introduced for approximating the probability density of Markov transition matrices induced by observed transition counts. Algorithms are given for sampling stochastic matrices and stochastic matrices that fulfill detailed balance. The algorithms are based on Metropolis Monte Carlo, are easy to implement, and exhibit good convergence properties.

Molecular dynamics in equilibrium always fulfills detailed balance. It has been shown that including detailed balance can significantly alter the distribution of transition matrices. In particular, it may reduce the uncertainties of some transition matrix properties, which may be essential when computing kinetic properties, such as transition pathways or rates.

ACKNOWLEDGMENTS

We would like to thank the two anonymous referees for their constructive criticism and helpful comments. P.M. was supported by ONR Grant No. 25-74200-F6422. F.N. and C.S. acknowledge support by the Research Center MATHEON "Mathematics for Key Technologies," DFG (Grant No. FZT86) in Berlin.

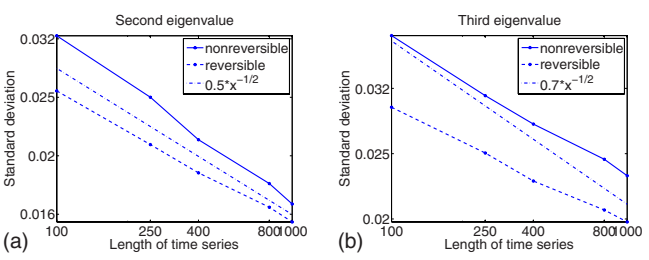

FIG. 13. (Color online) Relative uncertainties of the second- (panel A) and third-largest (panel B) eigenvalues of the transition matrix.

APPENDIX

1. Proof of correctness

The proposed MCMC scheme generates an ensemble of transition count matrices $\mathcal{K}$ distributed according to $p_Z[u(K)]$. It remains to prove that the ensemble of transition matrices $\mathcal{T}=\{u(\mathcal{K})\}$ resulting from the *transformation* $K \mapsto T=u(K)$ is indeed distributed according to $p_Z(T)$.

In order to motivate the need of a restriction of the MCMC scheme on a subset $\mathfrak{K} \subset \mathbb{R}^{m^2}$ notice that the transformation $u(K)$ is a projection. In particular, $u(K)$ is not injective because, e.g.,

$$u(K) = u(\alpha K) \quad \forall\, \alpha \in \mathbb{R}, \quad \alpha \neq 0,$$

which further shows that without any constraints on the ensemble $\mathcal{K}$ the statistical weight of $u(K)$ would not be well defined. Fortunately, the lack of invertibility can be compensated by a restriction of the MCMC scheme on an appropriate state space. In the nonreversible case it turns out that the restriction of the MCMC scheme on the set

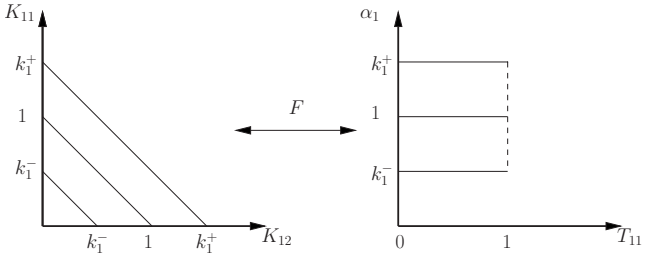

FIG. 14. Schematic illustration of the transformation F used in the proof of Theorem V.1 and the particular choice of the set $\mathfrak{K}$ in Eq. (24). A row-vector (K_{11}, K_{12}) of a 2×2 transition count matrix K is mapped via the transformation F into the (α_1, T_{11}) space and vice versa.

$$\mathfrak{K} = \left\{ K \in R_{0+}^{m^2} : k_i^- \le \sum_{j=1}^{m} K_{ij} \le k_i^+, \quad i = 1, \ldots, m \right\} \tag{A1}$$

with $0 < k_i^- < k_i^+$, $i = 1, \ldots, m$, leads to the right statistical weights. For a schematic illustration of Eq. (24) see Fig. 14.

Theorem V.1. Let $\mathcal{K} = \{K \in \mathfrak{K}\}$ be an ensemble of count matrices distributed according to $p_Z[u(K)]$. Then the ensemble $\mathcal{T} = \{u(K) : K \in \mathcal{K}\}$ is distributed according to $p_Z(T)$, i.e.,

$$\mathbb{P}[u(K) = T] = c p_Z(T) \quad \forall\, T \in \mathcal{T},$$

where $c > 0$ is a positive constant independent of the matrix T.

Proof. The probability $\mathbb{P}[u(K) = T]$ can formally be written as

$$\mathbb{P}[u(K) = T] = \int_{u^{-1}(T)} p_Z[u(K)] dK. \tag{A2}$$

Since the map $u(K)$ is a projection, the inverse $u^{-1}(T)$ is not unique. However, for an $T \in \mathcal{T}$ the inverse can be parameterized by

$$u^{-1}(T) = \{\mathrm{diag}(\alpha_1, \ldots, \alpha_m) T\}$$

with $\alpha_i \in [k_i^-, k_i^+]$, $i = 1, \ldots, m$. In order to evaluate the integral in Eq. (25) we change variables via

$$F : K \mapsto (\alpha_1, T_{11}, \ldots, T_{1,m-1}, \alpha_2, T_{21}, \ldots, T_{2,m-1}, \ldots, \alpha_m, T_{m1}, \ldots, T_{m,m-1}), \tag{A3}$$

where $\alpha_i = \sum_{j=1}^{m} K_{ij}$ and $T_{ij} = \frac{K_{ij}}{\sum_{n=1}^{m} K_{in}}$, $1 \le i \le m$. See Fig. 14 for a schematic illustration of $F(K)$.

Lemma V.2. Here

$$\int_{u^{-1}(T)} p_Z[u(K)] dK = \int_{k_1^-}^{k_1^+} \cdots \int_{k_m^-}^{k_m^+} p_Z(T) \alpha_1^{m-1} \cdots \alpha_m^{m-1} d\alpha_1 \ldots d\alpha_m.$$

Proof. The transformed integrand in the new variables is given in [26],

$$p_Z[u(F^{-1})] |\det[J(F^{-1})]|, \tag{A4}$$

where $J(F^{-1})$ is the Jacobian of the transformation F^{-1} which takes the form

$$F^{-1} : (\alpha_1, T_{11}, \ldots, T_{1,m-1}, \ldots, \alpha_m, T_{m1}, \ldots, T_{m,m-1}) \mapsto (K_{11}, \ldots, K_{mm}),$$

$$K_{ij} = \begin{cases} \alpha_i T_{ij} & \text{if } 1 \le i \le m, \ 1 \le j \le m-1 \\ \alpha_i \left(1 - \sum_{j=1}^{m-1} T_{ij}\right) & \text{if } j = m. \end{cases} \tag{A5}$$

Let $(\alpha_1, T_{11}, \ldots, T_{1,m-1}, \ldots, \alpha_m, T_{m1}, \ldots, T_{m,m-1}) \in F[u^{-1}(T)]$ then the first factor in Eq. (27) reduces to $p_Z(T)$. The Jacobian in Eq. (27) has a diagonal block structure since $(K_{i1}, \ldots, K_{im})$ for a fixed i depends only on $(\alpha_i, T_{i1}, \ldots, T_{i,m-1})$. Thus, $\det(J) = \det(J_1) \ldots \det(J_m)$ with

$$\det J_i = \begin{vmatrix} T_{i1} & \alpha_i & 0 & 0 & \ldots \\ T_{i2} & 0 & \alpha_i & 0 & \ldots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ T_{i,m-1} & 0 & \ldots & \ldots & \alpha_i \\ 1 - \sum_{j=1}^{m-1} T_{1j} & -\alpha_i & \ldots & \ldots & -\alpha_i \end{vmatrix}$$

$$= \begin{vmatrix} T_{i1} & \alpha_i & 0 & 0 & \ldots \\ T_{i2} & 0 & \alpha_i & 0 & \ldots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ T_{i,m-1} & 0 & \ldots & \ldots & \alpha_i \\ 1 & 0 & \ldots & \ldots & 0 \end{vmatrix} = (-1)^{(d-1)}$$

$$\times \begin{vmatrix} 1 & 0 & \ldots & \ldots & 0 \\ T_{i1} & \alpha_i & 0 & 0 & \ldots \\ T_{i2} & 0 & \alpha_i & 0 & \ldots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ T_{i,m-1} & 0 & \ldots & 0 & \alpha_i \end{vmatrix} = (-1)^{(m-1)} \alpha_i^{(m-1)}.$$

Thus, $|\det(J)| = \alpha_1^{(m-1)} \ldots \alpha_m^{(m-1)}$. ■

It follows that

$$\mathbb{P}[u(K)=T]=\int_{u^{-1}(T)} p_Z[u(K)]dK$$
$$=\int_{k_1^-}^{k_1^+}\cdots\int_{k_m^-}^{k_m^+} p_Z(T)\alpha_1^{m-1}\cdots\alpha_m^{m-1}d\alpha_1\cdots d\alpha_m = cp_Z(T),$$

where the constant c is independent of the matrix T. ■

Finally, we prove the reversible case. Recall the definition of the set

$$\mathfrak{K}_{sym}=\left\{K\in R_+^{m^2}:K \text{ symm.}, k^-\le\sum_{i,j=1}^{m}K_{ij}\le k^+\right\}$$

with $0<k^-<k^+<\infty$.

Theorem V.3. Let $\mathcal{K}=\{K\in\mathfrak{K}_{sym}\}$ be an ensemble of symmetric count matrices distributed according to $p_Z[u(K)]$. Then the ensemble $\mathcal{T}=\{u(K):K\in\mathcal{K}\}$ of reversible transition matrices is distributed according to $p_Z(T)$.

Proof. The proof follows the reasoning of the nonreversible case. The key observation is that the statistical weight of a reversible transition matrix $T\in\{u(K):K\in\mathcal{K}\}$ is given by

$$\int_{\{\alpha S\}} p_Z[u(K)]dK,$$

where

$$S=\mathrm{diag}(\pi_1,\ldots,\pi_m)T$$

is a symmetric matrix with $\alpha\in\{k^-,k^+\}$ and $\pi=(\pi_i)$, $i=1,\ldots,m$, is the unique stationary distribution of T.

To motivate the following transformation, notice that for symmetric K the stationary distribution of $T=u(K)$ is simply given by

$$\pi_i=\frac{\sum_{j=1}^{m}K_{ij}}{\sum_{k,j=1}^{m}K_{kj}}$$

and we conclude

$$\pi_i T_{ij}=\frac{K_{ij}}{\sum_{k,l=1}^{m}K_{kl}}.$$

A short calculation shows that by virtue of the transformation

$$K\mapsto\left(\sum_{k,l=1}^{m}K_{kl},\frac{K_{11}}{\sum_{k,l=1}^{m}K_{kl}},\ldots,\frac{K_{m,m-1}}{\sum_{k,l=1}^{m}K_{kl}}\right)$$

and the definition of the set $\mathfrak{K}_{sym}$ in Eq. (17) the integral evaluates to

$$\int_{\{\alpha S\}} p_Z[u(K)]dK=[(k^+)^{m^2-1}-(k^-)^{m^2-1}]p_Z(T).$$

■

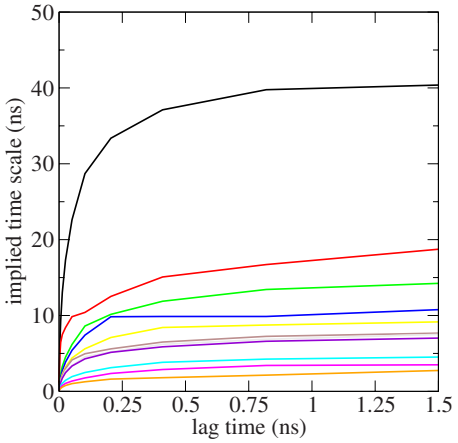

FIG. 15. (Color online) Implied time scales of the Markov model for MR121-GSGS-W depending on the lag time τ. The different curves represent the time scales implied by the eigenvalues of the transition matrix evaluated at lag time τ. This indicates that the dynamics becomes approximately Markovian at about $\tau\approx 1$ ns.

2. Molecular dynamics setup and Markov model for the model peptide

A molecular dynamics simulations of the MR121-GSGS-W peptide in water was performed with a simulation length of 1 μs. The simulation was performed in explicit water at 293 K using the GROMOS96 force field [27] and the GROMACS program [28]. Partial atomic charges for the dye MR121 were taken from [29]. One peptide molecule in an extended conformation was solvated with water and placed in a periodic rhombic dodecahedron box large enough to contain the peptide molecule and $\approx$1.0 nm of solvent on all sides at a liquid density of 55.32 mol/l ($\approx$1 g/cm^3), producing 1155 water molecules. Water was modeled by the simple point charge model [30]. Simulations were performed in the *NVT* ensemble using a Berendsen thermostat.

All bond lengths were fixed using the Lincs algorithm [31] and a time step of 2 fs for numerical integration was used. Periodic boundary conditions were applied to the simulation box and the long-range electrostatic interactions were treated with the particle mesh Ewald method [32] using a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation to compute the potential and forces in between grid points. The real space cutoff distance was set to 0.9 nm. The C-terminal end of the peptide was modeled as COO^- to reproduce a *p*H of about 7 as in the experimental conditions [23]. No counter ions were added since the simulation box was already neutral (one positive charge on MR121 and one negative charge on the terminal COO^-). The coordinates were saved every $\Delta t=0.2$ ps.

Next, a transition matrix model was built. To distinguish all relevant conformations of the system, the peptide coordinates were fitted to the extended configuration and then the state space was partitioned into small regions using a k-means clustering with $k=1000$. A fine-grained transition matrix, $T_{fine}(\tau)$, was estimated from the data at $\tau=1$ ns. In

order to concentrate on the metastable states in the system, the 1000 fine states were grouped together using the PCCA +method [33] as described in [4] into 33 metastable sets, which were used for the transition matrix sampling in the present paper. In order to determine the lag time τ, the implied time scale method proposed in [15] was employed, indicating that for $\tau=1$ ns the transitions are approximately Markovian (see Fig. 15).

[1] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).

[2] A. Ostermann, R. Waschipky, F. G. Parak, and G. U. Nienhaus, Nature (London) **404**, 205 (2000).

[3] S. Fischer, B. Windshuegel, D. Horak, K. C. Holmes, and J. C. Smith, Proc. Natl. Acad. Sci. U.S.A. **102**, 6873 (2005).

[4] F. Noé, D. Krachtus, J. C. Smith, and S. Fischer, J. Chem. Theory Comput. **2**, 840 (2006).

[5] M. Jäger, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. Noel, M. Gruebele, and J. Kelly, Proc. Natl. Acad. Sci. U.S.A. **103**, 10648 (2006).

[6] A. Y. Kobitski, A. Nierth, M. Helm, A. Jäschke, and G. U. Nienhaus, Nucleic Acids Res. **35**, 2047 (2007).

[7] P. Lenz, B. Zagrovic, J. Shapiro, and V. S. Pande, J. Chem. Phys. **120**, 6769 (2004).

[8] I. Horenko, S. Dolaptchiev, A. Eliseev, I. Mokhov, and R. Klein, J. Atmos. Sci. **65**, 3479 (2008).

[9] I. Horenko and C. Schütte, Adv. Data Anal. Class. (to be published).

[10] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 4th ed. (Elsevier, Amsterdam, 2006).

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

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

[13] F. Noé and S. Fischer, Curr. Opin. Struct. Biol. **18**, 154 (2008).

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

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

[16] Ch. Schütte, W. Huisinga, and P. Deuflhard, in *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, edited by B. Fiedler (Springer Verlag, Berlin, 2001), pp. 191–223.

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

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

[19] L. Devroye, *Non-Uniform Random Variate Generation* (Springer, New York, 1986).

[20] J. Ahrens and U. Dieter, Commun. ACM **25**, 47 (1982).

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

[22] P. Brémaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues* (Springer Verlag, Berlin, 1999).

[23] H. Neuweiler, M. Löllmann, S. Doose, and M. Sauer, J. Mol. Biol. **365**, 856 (2007).

[24] A. Gelman and D. B. Rubin, Stat. Sci. **7**, 457 (1992).

[25] S. Brooks and A. Gelman, J. Comput. Graph. Stat. **7**, 434 (1998).

[26] A. Levine, *Theory of Probability* (Addison-Wesley, Reading, MA, 1971).

[27] W. F. van Gunsteren and H. J. C. Berendsen, Angew. Chem., Int. Ed. Engl. **29**, 992 (1990).

[28] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, J. Comput. Chem. **26**, 1701 (2005).

[29] A. C. Vaiana, H. Neuweiler, A. Schulz, J. Wolfrum, M. Sauer, and J. C. Smith, J. Am. Chem. Soc. **125**, 14564 (2003).

[30] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).

[31] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).

[32] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).

[33] P. Deuflhard and M. Weber, ZIB Report, 2003 (unpublished), Vol. 03-09.