Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules

Frank Noé, Hao Wu, Jan-Hendrik Prinz, and Nuria Plattner

Citation: The Journal of Chemical Physics **139**, 184114 (2013); doi: 10.1063/1.4828816
View online: http://dx.doi.org/10.1063/1.4828816
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/139/18?ver=pdfcov
Published by the AIP Publishing

Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules

Frank Noé,[a] Hao Wu,[b] Jan-Hendrik Prinz,[b] and Nuria Plattner
Department of Mathematics and Computer Science, FU Berlin, Arnimallee 6, 14159 Berlin, Germany

(Received 3 September 2013; accepted 22 October 2013; published online 13 November 2013)

Markov state models (MSMs) have been successful in computing metastable states, slow relaxation timescales and associated structural changes, and stationary or kinetic experimental observables of complex molecules from large amounts of molecular dynamics simulation data. However, MSMs approximate the true dynamics by assuming a Markov chain on a clusters discretization of the state space. This approximation is difficult to make for high-dimensional biomolecular systems, and the quality and reproducibility of MSMs has, therefore, been limited. Here, we discard the assumption that dynamics are Markovian on the discrete clusters. Instead, we only assume that the full phase-space molecular dynamics is Markovian, and a projection of this full dynamics is observed on the discrete states, leading to the concept of Projected Markov Models (PMMs). Robust estimation methods for PMMs are not yet available, but we derive a practically feasible approximation via Hidden Markov Models (HMMs). It is shown how various molecular observables of interest that are often computed from MSMs can be computed from HMMs/PMMs. The new framework is applicable to both, simulation and single-molecule experimental data. We demonstrate its versatility by applications to educative model systems, a 1 ms Anton MD simulation of the bovine pancreatic trypsin inhibitor protein, and an optical tweezer force probe trajectory of an RNA hairpin.
 [http://dx.doi.org/10.1063/1.4828816]

I. INTRODUCTION

Conformational transitions are essential to the function of proteins and nucleic acids. With the ever increasing time resolution of ensemble kinetics experiments and the more recent maturation of sensitive single-molecule techniques in biophysics, experimental evidence supporting the near-universality of the existence of multiple metastable conformational substates and complex kinetics in biomolecules has continued to accumulate.[11,13,14,32,44] Markov (state) models (MSMs) are a very successful approach to deal with such a multitude of metastable states that has emerged from the simulation community.[5,6,24,26,34,38,39] A MSM consists of a discretization of the molecular state space into n clusters, and a $n \times n$ transition probability matrix containing the conditional probabilities that the system will, given that it is in one of its n discrete substates, be found in any of these n discrete substates a fixed lag time τ later. Because only *conditional* transition probabilities are needed, an MSM can be estimated from ensembles of short trajectories, computed distributedly on clusters or volunteer networks.[4,26,40] This circumvents the need for ultralong trajectories that can only be computed by special-purpose supercomputers.[20,37] Additionally, MSMs have been so successful because they permit many important thermodynamic, kinetic, and mechanistic molecular quantities to be computed much more directly and unambiguously than with conventional MD analyses.

However, a key approximation of MSMs is that they assume a Markov chain on the discrete clusters—although these discrete dynamics are not Markovian. It has been rigorously shown that the MSM approximation can be very precise if the molecular coordinates relevant for the slow transitions are finely discretized.[29,33] In practice the discretization quality will depend on the subset molecular coordinates and metric used as input, and the method used to cluster this coordinate space. The sheer high-dimensionality of solvated biomolecular systems, and the necessity to neglect many coordinates (velocities, solvent positions), limits the practical ability to produce a very fine discretization. Therefore, MSM results may significantly differ depending on the choice of input coordinates and clustering methods.[5,6,27,29,35] Moreover, the assumption that the dynamics of the clustered molecular observables is Markovian prohibits the use of MSMs for the analysis of experimental single-molecule trajectories, where the molecular coordinate traced is determined by what is experimentally observable and cannot be arbitrarily chosen.

Here, we introduce a new framework that altogether discards the assumption that dynamics are Markovian on the observed clusters. Instead we only make very basic physical assumptions: The full phase-space dynamics are Markovian, and in thermodynamic equilibrium. This full-space dynamics becomes projected onto the discrete clusters whose discrete dynamics is observed. This leads to the concept of Projected Markov Models (PMMs). We show that if the dynamics are metastable, having a number m slow relaxation processes, and if there is a separation of timescales to the next-faster relaxation processes, then PMMs can be

[a] Author to whom correspondence should be addressed. Electronic mail: frank.noe@fu-berlin.de
[b] H. Wu and J.-H. Prinz contributed equally to this work.

approximated by Hidden Markov Models (HMMs) with m hidden states. We describe an MSM→HMM transformation that provides a good starting point to estimate the HMM via Baum-Welch Expectation Maximization (EM) algorithm. It is shown how various molecular observables of interest that are often computed from MSMs can be computed from HMMs. The new method is applicable to both, simulation and single-molecule experimental data. Moreover, all important thermodynamic, kinetic, and mechanistic molecular quantities computable from MSMs can also be computed from HMMs. We demonstrate the versatility of our approach by applications to various systems—including model systems that demonstrate the superiority of PMM/HMM models over MSMs, a 1 ms Anton MD simulation of the bovine pancreatic trypsin inhibitor (BPTI) protein where a three-state rate matrix with metastable sets of structures is readily obtained, and an optical tweezer force probe trajectory of an RNA hairpin, where a hidden and yet unreported transition state is found.

II. PROJECTED MARKOV MODELS

We assume that there is a Markov process $\{\mathbf{z}_t\}$ in state space Ω (a state may consist of both positions and velocities—depending on the model of the dynamics). This Markov process is assumed to be ergodic and reversible with respect to a unique stationary distribution $\mu(\mathbf{z})$. Often, a canonical ensemble is employed, and then the stationary distribution is the Boltzmann distribution $\mu(\mathbf{z}) = Z^{-1}\mathrm{e}^{-\beta H(\mathbf{z})}$ with $H(\mathbf{z})$ the total energy and $\beta = 1/k_B T$ the inverse temperature. For such a process, we can write the ensemble dynamics as follows: We consider a probability distribution of states p_0. At a later time τ, the distribution has evolved according to the Markov propagator $\mathcal{P}$,

$$p_\tau(\mathbf{z}_\tau) = \mathcal{P}(\tau)\, p_0(\mathbf{z}_0) = \int_{\mathbf{z}_0} d\mathbf{z}_0\; p_\tau(\mathbf{z}_0, \mathbf{z}_\tau)\; p_0(\mathbf{z}_0), \qquad (1)$$

where the conditional transition probability $p_\tau(\mathbf{z}_0, \mathbf{z}_\tau)$ characterizes the dynamics of the system. With the ergodicity, we can expand the propagation density into basis functions

$$p_\tau(\mathbf{z}_0, \mathbf{z}_\tau) = \mu(\mathbf{z}_\tau) + \sum_{i=2}^{\infty} \mathrm{e}^{-\kappa_i \tau} \frac{\phi_i(\mathbf{z}_0)}{\mu(\mathbf{z}_0)} \phi_i(\mathbf{z}_\tau), \qquad (2)$$

where κ_i is the relaxation rate of the ith-slowest process and $t_i = \kappa_i^{-1}$ is the corresponding relaxation timescale. We can also consider the corresponding correlation density, i.e., the joint probability density to observe the system at position $\mathbf{z}_0$ at time 0 and at position $\mathbf{z}_\tau$ at time τ,

$$c_\tau(\mathbf{z}_0, \mathbf{z}_\tau) = \mu(\mathbf{z}_0)\mu(\mathbf{z}_\tau) + \sum_{i=2}^{\infty} \mathrm{e}^{-\kappa_i \tau} \phi_i(\mathbf{z}_0)\phi_i(\mathbf{z}_\tau). \qquad (3)$$

Note that for $\tau \to \infty$, the joint probability density is simply given by the stationary probabilities: $c_\infty(\mathbf{z}_0, \mathbf{z}_\infty) = \mu(\mathbf{z}_0)\mu(\mathbf{z}_\infty)$. For the rest of the paper we assume that our system of interest has m slow processes and a timescale separation to the faster processes. Thus, at lag times significantly larger than $t_m = \kappa_m^{-1}$, the correlation density is approximately given by

$$c_\tau(\mathbf{z}_0, \mathbf{z}_\tau) \approx \mu(\mathbf{z}_0)\mu(\mathbf{z}_\tau) + \sum_{i=2}^{m} \mathrm{e}^{-\kappa_i \tau} \phi_i(\mathbf{z}_0)\phi_i(\mathbf{z}_\tau). \qquad (4)$$

Now we assume that the molecular state space (typically only configurations, not velocities) is completely partitioned into a set of n clusters $\{S_i\}$, which might be rather coarse. What happens to the dynamics when we observe it on the space of clusters? From Eq. (4) we can compute the correlation matrix between clusters,

$$\begin{aligned} C_{ij} &= \int_{\mathbf{z}_0 \in S_i} d\mathbf{z}_0 \int_{\mathbf{z}_\tau \in S_j} d\mathbf{z}_\tau\; c_\tau(\mathbf{z}_0, \mathbf{z}_\tau) \\ &= \pi_i \pi_j + \sum_{k=2}^{m} \mathrm{e}^{-\kappa_k \tau} q_{ki} q_{kj}, \end{aligned}$$

where $\pi_i = \int_{\mathbf{z} \in S_i} d\mathbf{z}\, \mu(\mathbf{z})$ are the stationary probabilities and $\mathbf{q}_k$ is the kth discretized eigenfunction,

$$q_{ki} = \int_{\mathbf{z}_0 \in S_i} d\mathbf{z}_0\, \phi_k(\mathbf{z}_0).$$

We can also express the correlation matrix as

$$\mathbf{C} = \mathbf{Q}^\top \widetilde{\mathbf{\Lambda}} \mathbf{Q}, \qquad (5)$$

where $\mathbf{Q} \in \mathbb{R}^{m \times n}$ contains the discretized projected eigenfunctions $\mathbf{q}_k$, and $\widetilde{\mathbf{\Lambda}} \in \mathbb{R}^{m \times m}$ contains the m dominant eigenvalues. We will use the tilde in order to annotate "small" matrices or vectors related to the m dominant eigenvalues $\lambda_1 = 1, \lambda_2 = \mathrm{e}^{-\kappa_2 \tau} \ldots, \lambda_m = \mathrm{e}^{-\kappa_m \tau}$, and thus to the m most metastable processes. If we write the stationary probability vector $\boldsymbol{\pi}$ on the diagonal of the matrix $\mathbf{\Pi}$, we can write the transition matrix between clusters as

$$\mathbf{T}(\tau) = \mathbf{\Pi}^{-1} \mathbf{Q}^\top \widetilde{\mathbf{\Lambda}}(\tau) \mathbf{Q}. \qquad (6)$$

This is the transition matrix that is estimated when building a Markov model at lag time τ. Now we can easily illustrate the problem of MSMs: The dynamics between clusters are not Markovian, i.e., the transition matrix estimated at τ cannot be used to predict long-timescale behavior,

$$\begin{aligned} \mathbf{T}(2\tau) &= \mathbf{\Pi}^{-1} \mathbf{Q} \widetilde{\mathbf{\Lambda}}(2\tau) \mathbf{Q}^T = \mathbf{\Pi}^{-1} \mathbf{Q} [\widetilde{\mathbf{\Lambda}}(\tau)]^2 \mathbf{Q}^T \\ &\neq \mathbf{\Pi}^{-1} \mathbf{Q} \widetilde{\mathbf{\Lambda}}(\tau) \mathbf{Q}^T \mathbf{\Pi}^{-1} \mathbf{Q} \widetilde{\mathbf{\Lambda}}(\tau) \mathbf{Q}^T = \mathbf{T}^2(\tau). \qquad (7) \end{aligned}$$

The first row is not equal to the second row because projected eigenvectors $\mathbf{Q}$ of the full-space dynamics are *not* eigenvectors of $\mathbf{T}$, and are, therefore, *not* orthonormal with respect to the observed stationary distribution: $\mathbf{Q}^T \mathbf{\Pi}^{-1} \mathbf{Q} \neq \mathbf{Id}$. Therefore, in order to estimate the cluster dynamics in a way that is unbiased, and that allows the long-time dynamics to be predicted, one needs to estimate the PMM quantities

$$\{\mathbf{Q}, \widetilde{\mathbf{\Lambda}}\} \qquad (8)$$

separately.

III. APPROXIMATING PMMS VIA HIDDEN MARKOV MODELS

In general, estimating the matrices $\{\mathbf{Q}, \widetilde{\mathbf{\Lambda}}\}$ is difficult, especially for large m and n. Therefore, we consider a slightly different model that we can efficiently estimate: A HMM. A hidden Markov model consists of a transition matrix between m hidden (here metastable) states, $\widetilde{\mathbf{T}}(\tau)$, and an associated stationary distribution $\widetilde{\boldsymbol{\pi}}$. This hidden dynamics has a joint probability (correlation) matrix $\widetilde{\mathbf{C}} = \widetilde{\mathbf{\Pi}}\widetilde{\mathbf{T}}$. Each hidden state i will output to one of the observable states j with a probability χ_{ij}, such that the vector $\boldsymbol{\chi}_i$ is the output probability distribution of hidden state i. We can write the correlation matrix on the observed states as

$$\begin{aligned}\mathbf{C} &= \boldsymbol{\chi}^\top \widetilde{\mathbf{C}} \boldsymbol{\chi} = \boldsymbol{\chi}^\top \widetilde{\mathbf{\Pi}} \widetilde{\mathbf{T}} \boldsymbol{\chi} \\ &= \boldsymbol{\chi}^\top \widetilde{\mathbf{L}}^\top \widetilde{\mathbf{\Lambda}} \widetilde{\mathbf{L}} \boldsymbol{\chi} \\ &= \mathbf{Q}^\top \widetilde{\mathbf{\Lambda}} \mathbf{Q}, \end{aligned} \tag{9}$$

where $\widetilde{\mathbf{L}} = [\tilde{\mathbf{l}}_1, \ldots, \tilde{\mathbf{l}}_m]^\top$ contains the dominant m eigenvalues of the hidden transition matrix $\widetilde{\mathbf{T}}(\tau)$, i.e., $\tilde{\mathbf{l}}_i^\top \widetilde{\mathbf{T}} = \lambda_i \tilde{\mathbf{l}}_i^\top$. By comparing the last row to (5), it is apparent that a HMM has a similar structure like a PMM. Here, the vectors in $\mathbf{Q}$ are given by the HMM eigenvectors $\widetilde{\mathbf{L}}$ projected onto the observable states via $\boldsymbol{\chi}$. However, we want to use HMM estimation algorithms to estimate the slow molecular kinetics of a Markov process observed on a cluster space, which is a PMM (5)—and therefore we must show that a PMM can also be represented as a HMM. This is not obvious: A given PMM, defined by the slow process eigenfunctions and the chosen discretization has a certain $\mathbf{Q}$. It is not *a priori* clear whether this $\mathbf{Q}$-matrix can be represented by decomposition into the two matrices $\boldsymbol{\chi}\widetilde{\mathbf{L}}$, because these matrices have to fulfill the constraints that the columns of $\boldsymbol{\chi}$ are probability distributions and the rows of $\widetilde{\mathbf{L}}$ form a set of eigenvectors which are orthonormal with respect to $\widetilde{\boldsymbol{\pi}}^{-1}$. Appendix A contains a proof that modeling a PMM with an HMM is valid in a special, but interesting case. We summarize it as follows:

Given a Markov process $\{\mathbf{x}_t\}$ *that is ergodic and reversible with respect to the unique stationary distribution* $\mu(\mathbf{x})$. *Given that this process has m metastable states, such that there is a gap in the relaxation timescales,* $t_{m+1} \ll t_m$, *and the stationary distribution* $\mu(\mathbf{x})$ *almost decomposes into m modes, such that almost all stationary probability is in the metastable states and the intervening transition states have vanishing populations. We further consider an arbitrary discretization of the state space* $\mathbf{x}$ *into n clusters. Then, the dynamics on the n discrete states is described by a discrete hidden Markov model with m hidden and n observed states.*

This is an important result: in many applications, especially in biomolecular dynamics, we have a few metastable states with rarely populated transition regions. The theorem above says, that even using a *poor* discretization of the state space of such a system, we can still describe its metastable dynamics *exactly* with an HMM. Of course, our practical ability to find the true HMM will depend on the amount of statistics at hand, and may very well depend on the quality of the discretization. However, we will show in Sec. V that HMMs perform very well in this setting, and almost exclusively better than MSMs.

A. Initializing a hidden Markov model from a Markov model

Estimating hidden Markov models is more difficult than estimating directly observed Markov models, because in contrast to the MSM likelihood, the HMM likelihood does not necessarily have a unique optimum. Therefore, it is important to start the HMM estimation "close" to the optimal result. How do we get a good initial guess for the hidden transition matrix $\widetilde{\mathbf{T}}(\tau)$ and the output probability matrix $\boldsymbol{\chi}$?

Hence, we propose an initial HMM based on a direct Markov model. Given the simulation trajectories, discretized in n states, we estimate a Markov model transition matrix at some lag time τ, $\mathbf{T}(\tau) \in \mathbb{R}^{n\times n}$. In order to ensure that this Matrix fulfills detailed balance, we use the reversible transition matrix estimator described in Ref. 29 and implemented in the EMMA software.[36]

Next, we fix a number of hidden states, m, and obtain an initial estimate of the output probability matrix $\boldsymbol{\chi}$. For this, we first employ the PCCA+ method[8] implemented in the EMMA software.[36] PCCA+ is a clustering method, which provides, for each observed state i a degree of membership to a metastable state j, m_{ij}. This is done first fixing m states as representatives of the metastable states, each obtaining membership 1 to the respective metastable states and 0 to the others. The choice of these states depends on the implementation of PCCA+ (Ref. 42 optimizes metastability, while Ref. 31 maximizes the crispness of the memberships). Some current codes do not implement the full PCCA+ algorithm and select m representatives from the available set of states, often providing spurious negative membership functions. However, these differences are unimportant for the present context, as they will be taken care of by the subsequent HMM optimization. The full membership matrix $\mathbf{M} \in \mathbb{R}^{n\times m}$ is obtained by solving a linear system of equations, as described in Ref. 8.

Now, the membership m_{ij} can be interpreted as a probability of being in a metastable state j, given that the system is observed in discrete state i. We can use the membership matrix to coarse-grain the stationary probabilities to the hidden metastable states,

$$\widetilde{\boldsymbol{\pi}} = \mathbf{M}^\top \boldsymbol{\pi} \tag{10}$$

and we can use Bayesian statistics in order to transform $\mathbf{M}$ to the desired output probabilities,

$$\begin{aligned}\mathbb{P}[\text{cluster } j \,|\, \text{hidden } i] &= \frac{\mathbb{P}[\text{cluster } j]}{\mathbb{P}[\text{hidden } i]} \mathbb{P}[\text{hidden } i \,|\, \text{cluster } j] \\ \chi_{ij} &= \frac{\pi_j}{\widetilde{\pi}_i} m_{ji}. \end{aligned} \tag{11}$$

In matrix form,

$$\boldsymbol{\chi} = \widetilde{\mathbf{\Pi}}^{-1} \mathbf{M}^\top \mathbf{\Pi}. \tag{12}$$

Finally, we need a hidden transition matrix $\widetilde{\mathbf{T}}$ which fulfills Eq. (9), i.e., which produces, together with $\boldsymbol{\chi}$, the observable correlation matrix $\mathbf{C}$. A method for computing such a matrix

is given in Ref. 17. Using their Eq. (12) and performing some algebraic transformations given in Appendix B 1, we obtain the result,

$$\widetilde{\mathbf{T}} = \mathbf{M}^\top \mathbf{T} \mathbf{M} (\mathbf{M}^\top \mathbf{M})^{-1}, \tag{13}$$

which has a nice interpretation: $\mathbf{M}^\top \mathbf{T} \mathbf{M}$ performs a coarse-graining of $\mathbf{T}$, and $\mathbf{M}^\top \mathbf{M}$ is an overlap matrix needed for normalization. $\widetilde{\mathbf{T}}$ has the nice property that it preserves the dominant kinetics of $\mathbf{T}$: the eigenvalues of $\widetilde{\mathbf{T}}$ are identical to the dominant m eigenvalues of $\mathbf{T}$. In some cases $\widetilde{\mathbf{T}}$ may have (usually only slightly) negative elements. Moreover, $\widetilde{\mathbf{T}}$, when computed from Eq. (13), may no longer exactly fulfill detailed balance. As we want to obtain a starting guess for the transition matrix in a reversible HMM estimation framework, we enforce reversibility of $\widetilde{\mathbf{T}}$. For this, we compute the correlation matrix, $\widetilde{\mathbf{C}} = \widetilde{\mathbf{\Pi}} \widetilde{\mathbf{T}}$, symmetrizing it $\widetilde{\mathbf{C}} \leftarrow 1/2(\widetilde{\mathbf{C}} + \widetilde{\mathbf{C}}^\top)$, setting negative elements to 0, and then renormalizing the rows of $\widetilde{\mathbf{C}}$ to obtain $\widetilde{\mathbf{T}}$. Note that this intervention is important to make sure that the HMM optimization is seeded with a $\widetilde{\mathbf{T}}$ matrix that has a meaningful structure, but should not strongly affect the results as $\widetilde{\mathbf{T}}$ will be subsequently optimized.

B. Hidden Markov model estimation

Consider the observed trajectory $\{s_t\}$ and hidden trajectory $\{h_t\}$. The HMM likelihood is given by

$$\mathbb{P}(\{s_t\}|\widetilde{\mathbf{T}}, \boldsymbol{\chi}) = \sum_{\substack{\text{hidden paths} \\ h_0, \ldots, h_{t_{max}}}} \widetilde{\pi}_{h_0} \chi_{s_0 h_0} \prod_{t=1}^{t_{max}} \widetilde{T}_{h_{t-1} h_t} \chi_{s_t h_t}. \tag{14}$$

Obviously, the product over all possible hidden paths cannot be directly computed due to a combinatorial explosion of possibilities. The likelihood (14) can be maximized by a Expectation-Maximization algorithm, more precisely by the Baum-Welch method.[18,43] See Ref. 30 for a thorough and educative description of HMMs and the Baum-Welch method.

Since the EM method is established, we give a brief summary of our implementation in Appendix B 3. EM iterates two steps, called expectation and maximization step. While the expectation step is general, the maximization step must be designed for the specific HMM implementation. Here, we use the Baum-Welch algorithm to estimate a count matrix $\widetilde{\mathbf{Z}}(\tau)$ containing the estimated numbers of transitions between the m hidden states, and then estimate the maximum likelihood transition matrix $\widetilde{\mathbf{T}}(\tau)$ that fulfills detailed balance using the algorithm described in Ref. 29 and implemented in the EMMA software.[36] The HMM is assumed to be in equilibrium, i.e., it uses the stationary probability distribution of $\widetilde{\mathbf{T}}(\tau)$ as an initial distribution. The output probabilities $\boldsymbol{\chi}$ are estimated through straightforward histograms of the expected counts on the clusters.

C. Implied timescale plot

A commonly used approach to assess the quality of a MSM introduced in Ref. 39 is the implied timescale plot. Here, one asks how much the dynamics on the discretized state space deviates from a Markov chain. For a Markov chain, the Chapman-Kolmogorow equality $[\mathbf{T}(\tau_0)]^n = \mathbf{T}(n\tau_0)$ holds, and therefore for every eigenvalue $[\lambda_i(\tau_0)]^n = [\lambda_i(n\tau_0)]$. This condition is equivalent to the condition that the relaxation timescales (or implied timescales)

$$t_i(\tau) = -\frac{\tau}{\ln|\lambda_i(\tau)|} \tag{15}$$

are constant in $\tau = n\tau_0$. Because the dynamics on the discretized state space are not Markovian, the timescales (15) are *not* constant in τ. In the limit of good statistics they are guaranteed to be smaller than the true relaxation timescales,[10,25] and the error between the estimated relaxation timescale $t_i(\tau)$ and the true relaxation timescale decays slowly, as τ^{-1}.[28]

Here, we also conduct implied timescale plots in order to get a first assessment of the quality and robustness of the PMM estimation. However, instead of computing $\lambda_i(\tau)$ from a diagonalization of the transition matrix on the discretized state space, we use the eigenvalues of the hidden transition matrix, i.e., the timescales

$$\tilde{t}_i(\tau) = -\frac{\tau}{\ln|\tilde{\lambda}_i(\tau)|}. \tag{16}$$

If we are in a setting valid for PMM's, i.e., $\tau \gg t_{m+1}$ (all timescales that are not resolved by the PMM have decayed, where "$\gg$" is already given by a factor of 2–3), and we are in the limit of good statistics, then the PMM/HMM estimate of $\tilde{t}_i(\tau)$ should indeed be constant in τ.

D. Hidden Markov model validation

Finally, we estimate the HMM at a lag time τ_0 that has been selected such that the relaxation timescales $\tilde{t}_i(\tau)$ are constant at lag times $\tau > \tau_0$ or larger. Validation of the model consists of using it to compute kinetic quantities at a series of lag times τ and comparing them with the directly computed quantities at these different lag times. An example is the set-based Chapman-Kolmogorow test suggested in Ref. 29. Here, we suggest a very simple and direct test based on the relaxation timescales. Given the HMM estimated at τ_0, we compute the predicted transition matrices for the discretized state space at lag times $\tau = n\tau_0$,

$$\mathbf{T}^{\text{pred}}(\tau) = \mathbf{\Pi}^{-1} \boldsymbol{\chi}^\top \widetilde{\mathbf{\Pi}} [\widetilde{\mathbf{T}}(\tau_0)]^n \boldsymbol{\chi} \tag{17}$$

and compare their relaxation timescales with the relaxation timescales computed directly from MSM transition matrices estimated at τ,

$$t_i^{\text{pred}}(\tau) = t_i(\tau). \tag{18}$$

This test must succeed in order for the estimated HMM to be a valid description of the metastable kinetics.

Note that a more general comparison is possible by comparing appropriate norms of $\mathbf{T}^{\text{pred}}(\tau)$ and $\mathbf{T}(\tau)$, or, alternatively, of the correlation matrices $\mathbf{C}^{\text{pred}}(\tau) = \boldsymbol{\chi} \widetilde{\mathbf{\Pi}} [\widetilde{\mathbf{T}}(\tau_0)]^n \boldsymbol{\chi}^\top$ and $\mathbf{C}(\tau) = \mathbf{\Pi} \mathbf{T}(\tau)$. A practically feasible comparison could

be constructed in a similar way as the Chapman-Kolmogorow test in Ref. 29.

IV. QUANTITATIVE ANALYSIS

One of the reasons for the great success of direct (non-hidden) Markov models is that various quantities related to the molecule's thermodynamics, kinetics, and mechanisms can be calculated easily from an MSM transition matrix. Therefore, although it will be shown in Sec. V that HMMs can be much superior to MSMs in modeling the kinetics, their use needs to be motivated by showing that they are equally versatile as MSM. This section goes through a number of commonly used molecular observables and discusses how they can be computed from the HMM quantities $\widetilde{\mathbf{T}}$ and $\boldsymbol{\chi}$. Interestingly, for some observables, the computation from HMMs is even more straightforward than from MSMs.

Some important quantities can be directly accessed through an eigenvalue decomposition of the hidden transition matrix $\tilde{\mathbf{T}}(\tau)$. Such a decomposition provides the eigenvalues $\tilde{\lambda}_i(\tau)$, the right eigenvector matrix $\widetilde{\mathbf{R}}$, which contains the right eigenvectors $\widetilde{\mathbf{r}}_i$ as columns, and the left eigenvector matrix $\widetilde{\mathbf{L}} = \widetilde{\mathbf{R}}^{-1}$ which contains the left eigenvectors $\widetilde{\mathbf{l}}_i^T$ as rows. The first left eigenvector can be normalized to a sum of 1, yielding the stationary distribution of hidden states, $\widetilde{\boldsymbol{\pi}}$.

The stationary distribution $\widetilde{\boldsymbol{\pi}}$ provides the probability of observing one of the metastable states. The free energy of state i with respect to an arbitrary reference state 0 is given by

$$\Delta F_i = -k_B T \ln \frac{\tilde{\pi}_i}{\tilde{\pi}_0}. \tag{19}$$

Note that these free energy differences are associated to the weights of the metastable states, even when the state space discretization is poor. This is not the case when computing the free energy of metastable states from an MSM, where a poor discretization can lead to significant confusion which microstate should be associated to what degree to a metastable state. However, when the stationary distribution is sought on the microstates, it can be easily computed by transforming the stationary distributions of metastable states through the output probabilities,

$$\boldsymbol{\pi} = \boldsymbol{\chi}^{\top} \widetilde{\boldsymbol{\pi}}. \tag{20}$$

A quantity of particular interest is the definition of the metastable states themselves. In particular, which set of molecular structures is metastable? This question has been an important driving force in the development of MSMs. The original contribution in this field was made by Schütte, Deuflhard, and co-workers by noticing that for m most metastable states, the signs of the dominant m MSM eigenvectors are indicative.[9,34] Hence, they have proposed the clustering method PCCA, which finds m metastable sets according to the signs of the elements in the first m MSM transition matrix eigenvectors. A few years later, Weber and Deuflhard have invented an improved algorithm PCCA+,[8] which is numerically and conceptually superior and assigns to each microstate i a fuzzy membership m_{ij} to each metastable state j based on the proximity of microstate i to a representative state that is representative for metastable state j in the space of the dominant m eigenvectors of the MSM transition matrix. While PCCA and PCCA+ have nice theoretical properties, they are both unsatisfactory from a statistical point of view. As the PCCA(+) metastable states are defined based on the transition matrix eigenvectors, any information of the statistical significance is lost. Reference 2 has taken a Bayesian standpoint and proposed the agglomerative clustering method BACE, which defines the quality of a clustering based on information in the MSM transition count matrix.

HMMs directly provide information of the metastable states. The output matrix $\boldsymbol{\chi} = [\chi_{ij}]$ directly provides the probability that a given metastable state i is observed in a given microstate j. Its row vectors $\boldsymbol{\chi}_i$, therefore, are probability distributions of metastable states on the space of clusters. With the weight $\widetilde{\pi}_i$ these probability distributions can be weighted, such that these vectors sum up to the overall probability distribution of microstates: $\pi_j = \sum_i \widetilde{\pi}_i \chi_{ij}$. Using Bayesian inversion, the $\boldsymbol{\chi}$ matrix can be transformed into a membership matrix $\mathbf{M} = [m_{ij}]$ which contains the information "how much" microstate i belongs to metastable state j,

$$\mathbf{M} = \boldsymbol{\Pi}^{-1} \boldsymbol{\chi}^T \widetilde{\boldsymbol{\Pi}}, \tag{21}$$

where $\boldsymbol{\Pi} = \mathrm{diag}(\boldsymbol{\pi})$ and $\widetilde{\boldsymbol{\Pi}} = \mathrm{diag}(\widetilde{\boldsymbol{\pi}})$. This approach of defining metastable states unifies the advantages of PCCA+ and of statistically driven methods such as BACE: (1) As in PCCA+, the membership vectors of $\mathbf{M}$ lie in the subspace of the slow dynamics, and are therefore a mathematically meaningful approach for characterizing metastability. (2) In contrast to PCCA+, one does not need to define an objective function that will determine the location of the representative states, and therefore the memberships. The membership matrix $\mathbf{M}$ is a natural result of the HMM estimation itself. (3) Since HMMs include $\boldsymbol{\chi}$ as a direct parameter, the quantity $\boldsymbol{\chi}$ is directly amenable to statistical treatment like in BACE. When estimated via EM, $\boldsymbol{\chi}$ is the result of a likelihood (or posterior) maximization. When using Monte-Carlo sampling of the HMM likelihood,[7] the statistical uncertainty of elements in $\boldsymbol{\chi}$ can be directly assessed.

Let us turn to kinetic quantities. The m slowest relaxation rates, or phenomenological rates, of the molecule are given by

$$\tilde{\kappa}_i = -\frac{\ln |\tilde{\lambda}_i(\tau)|}{\tau}. \tag{22}$$

These rates, and their inverses, the relaxation timescales $\tilde{t}_i = \tilde{\kappa}_i$ are of central interest in kinetics, because they can be probed by various experimental techniques. A main advantage of PMMs is that—in stark contrast to MSMs—the rates $\tilde{\kappa}_i$ can be estimated without systematic error. This is also true for HMMs, when they are employed for metastable systems (see discussion in Sec. III). From these relaxation rates, and the eigenvectors of the hidden transition matrix, we can compute the rate matrix between metastable states,

$$\widetilde{\mathbf{K}} = \widetilde{\boldsymbol{\Pi}}^{-1} \widetilde{\mathbf{L}}^{\top} \begin{bmatrix} 0 & & & \\ & -\tilde{\kappa}_2 & & \\ & & \ddots & \\ & & & -\tilde{\kappa}_m \end{bmatrix} \widetilde{\mathbf{L}}. \tag{23}$$

In contrast to MSMs, the transformation to a rate matrix is possible, because the first m processes are metastable, and therefore $\lambda_1 \ldots \lambda_m$ are positive such that the rates (22) exit.

Besides the decomposition into metastable states, and the rate or transition matrix switching between them, the eigenvectors themselves provide a quite direct understanding of the metastable dynamics: The sign changes in $\widetilde{\mathbf{r}}_i$ and $\widetilde{\mathbf{l}}_i$ indicate structural changes that occur at the associated rates $\tilde{\kappa}_i$ or timescales $\tilde{t}_i$. On the discretized state space, these eigenvectors occur as projections from the hidden states through the output probability matrix,

$$\begin{aligned} \mathbf{s}_i &= \mathbf{M}\widetilde{\mathbf{r}}_i, \\ \mathbf{q}_i &= \boldsymbol{\chi}^\top \widetilde{\mathbf{l}}_i. \end{aligned} \tag{24}$$

Note that these projected eigenvectors may significantly differ from the right and left eigenvectors that are directly computed from an MSM transition matrix on the cluster space. The projected eigenvectors and the relaxation rates are the key components for calculating kinetic experimental observables. In Refs. 16 and 23, we have derived general expressions for computing correlation and relaxation experiments that can be straightforwardly extended to HMMs. In Ref. 19, we have extended this theory to scattering experiments.

An important source of kinetic information is time-correlation experiment. This may be realized by taking trajectories from time-resolved single molecule experiments, such as single molecule fluorescence or pulling experiments, and computing time correlations from these trajectories. Moreover, several ensemble kinetic experiments effectively measure time-correlation functions, for example, dynamical neutron scattering. A general expression for modeling these experiments is that of the time cross-correlation, of two experimentally observable quantities. Given a partition into states S_i, let us denote by a_i and b_i the averages of the two experimentally observable quantities over the discrete state S_i. $\mathbf{a}$, $\mathbf{b}$ are the vectors with these averages as elements. The cross-correlation for time τ can be expressed as

$$\begin{aligned} \mathbb{E}[a(t)b(t+\tau)] &= \sum_{i=1}^{m} \mathrm{e}^{-\tau\kappa_i} \langle \mathbf{a}, \mathbf{q}_i \rangle \langle \mathbf{b}, \mathbf{q}_i \rangle \\ &= \langle \mathbf{a}, \boldsymbol{\pi} \rangle \langle \mathbf{b}, \boldsymbol{\pi} \rangle + \sum_{i=2}^{m} \mathrm{e}^{-\tau\kappa_i} \langle \mathbf{a}, \mathbf{q}_i \rangle \langle \mathbf{b}, \mathbf{q}_i \rangle. \end{aligned} \tag{25}$$

Autocorrelation experiments can be modeled by simply setting $a = b$.

Alternatively, relaxation experiments can be used to probe the molecules' kinetics. In these experiments, the system is allowed to relax from a nonequilibrium starting state with probability distribution. Examples are temperature-jump, pressure-jump, or pH-jump experiments, rapid mixing experiments, or experiments where measurement at $t = 0$ starts from a synchronized starting state, such as in processes that are started by an external trigger like a photoflash. We consider initial distributions that are modeled on the metastable states, $\widetilde{\mathbf{p}}(0)$. For example, in an ideal two-state folder, the relaxation experiment shifts probabilities between the two metastable states, and a meaningful value of $\widetilde{\mathbf{p}}(0)$ could be computed from available experimental titration curves. The time evolution of such an initial distribution can be computed by propagating it with the transition or rate matrix that describe the dynamics for the conditions after the trigger,

$$\begin{aligned} \mathbf{p}_\tau^\top &= \tilde{\mathbf{p}}_0^\top [\widetilde{\mathbf{T}}(\tau_0)]^n \boldsymbol{\chi} \\ &= \tilde{\mathbf{p}}_0^\top \exp[\tau \widetilde{\mathbf{K}}] \boldsymbol{\chi} \end{aligned} \tag{26}$$

with $\tau = n\tau_0$. The ensemble average $\mathbb{E}_{\boldsymbol{p}(0)}[a(\tau)]$ of an experimentally measurable quantity, a, is recorded while the system relaxes from the initial distribution $\widetilde{\mathbf{p}}(0)$ to the new equilibrium distribution $\widetilde{\boldsymbol{\pi}}$. The expectation value of the signal at time τ is then given by

$$\mathbb{E}_{\tilde{\mathbf{p}}_0}[a(\tau)] = \sum_{i=1}^{m} \mathrm{e}^{-\tau\kappa_i} \langle \mathbf{a}, \mathbf{q}_i^l \rangle \langle \tilde{\mathbf{l}}_i, \widetilde{\mathbf{p}}_0^* \rangle, \tag{27}$$

where $\widetilde{\mathbf{p}}_0^*$ is the *excess probability distribution* $\widetilde{\mathbf{p}}_0^* = \Pi^{-1}\widetilde{\mathbf{p}}_0$. $\mathbb{E}_{\tilde{\mathbf{p}}_0}[a(\tau)]$ is again a multiexponential decay function with amplitudes $\langle \mathbf{a}, \mathbf{q}_i^l \rangle \langle \tilde{\mathbf{l}}_i, \widetilde{\mathbf{p}}_0^* \rangle$. Each of the amplitudes is associated with an eigenvector of the transition matrix, and therefore readily interpretable in terms of structural changes.

The combination of Markov models and the spectral theory given is useful to compare simulations and experiments via the *dynamical fingerprint* representation of the system kinetics.[23] Furthermore, this approach permits to design experiments that are optimal to probe individual relaxations.[23]

Finally, detailed molecular mechanisms of a particular process that transitions between two states A and B can be calculated with transition path theory (TPT).[41] Here, A and B may be associated to the unfolded and folded ensembles in protein folding, or to the dissociated and associated states in protein-ligand binding. TPT can be straightforwardly applied on the level of metastable states. This is done either by directly applying to the transition matrix $\widetilde{\mathbf{T}}$ (see Ref. 26), or by computing rate matrix $\widetilde{\mathbf{K}}$ (see above) and conducting TPT as described in Ref. 21.

V. APPLICATIONS

Figure 1 compares the performances of MSMs and HMMs on a diffusion in a metastable two-well potential (model details see supplementary material of Ref. 29). The state space is discretized into two clusters, comparing results for a good discretization separating the two metastable states in the transition region (Fig. 1(a)) and a bad discretization that splits one of the metastable states (Fig. 1(c)). For the good discretization, the MSM estimate converges to the true timescale when the lag time τ is increased, although slowly with an error that vanishes as τ^{-1}—see Ref. 28 for derivation. For the poor discretization, the convergence of the MSM is so slow that it does not come close to the true timescale before hitting the "forbidden" region $\tau > t$ at which no numerically robust MSM estimate is possible.[1] In contrast, the PMM/HMM converges quickly to the true timescale, and the timescale estimate then stays constant in τ. The speed of this

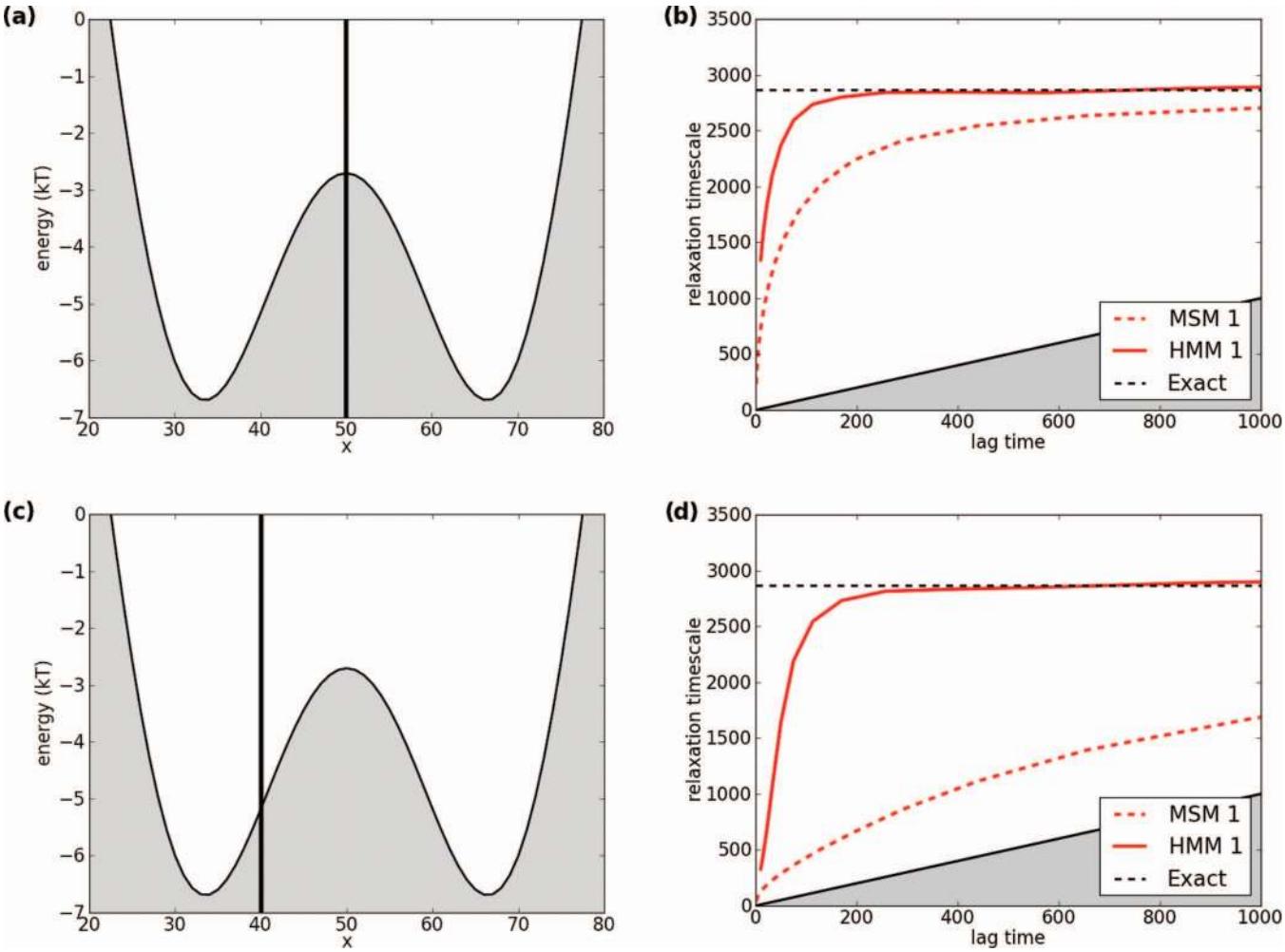

FIG. 1. Comparison of MSM and PMM/HMM for modeling the diffusion in a bistable potential, using (a) good and (c) poor discretization into two states. (b) and (d) τ-dependence of relaxation timescales computed with MSMs and HMMs. The grey region is the $\tau > t$ region where no numerically robust estimation of the relaxation timescale t is possible.

convergence goes exponential with the greatest neglected timescale, as $\exp(-\tau/t_3)$. Thus, the HMM behaves as a multi-τ estimator analyzed in Ref. 28. Obtaining a good model for the slow kinetics for a short lag time τ is very important for ensemble simulations, because it allows to keep the length of the individual simulations short as well. Shorter trajectory lengths also permit a more rapid turnover in adaptive sampling simulations,[3,15,38] thus allowing to get statistically converged estimates of the slow kinetics with lesser total sampling effort.

Figure 2 compares the performances of MSMs and HMMs when constructing the model on a subspace of conformation space that neglects important degrees of freedom. For the diffusive dynamics in the two-dimensional three-well potential shown in Fig. 2(a), both dimensions are needed in order to separate the three metastable states from one another. The projections of the probability density onto either the x or y coordinate (grey distributions) only exhibit two poorly separated modes. The slowest relaxation timescale is associated to the transition between the two deep wells, and therefore mostly with the x-axis, while the second-slowest relaxation timescale is associated to the transition between the lower left minimum and the shallow upper minimum. An MSM is able to estimate the slowest relaxation timescale from the x-projections (Figs. 2(b) and 2(c)), and the second-slowest relaxation timescale from the y-projection (Fig. 2(d)). However, an MSM is not able to estimate both slow processes simultaneously. The HMM performs similarly on the x-projection: When using a two-state HMM, the slowest timescale is estimated very accurately, and with a shorter lagtime than the MSM (Fig. 2(b)). When using a three-state HMM, the result for the slowest timescale actually gets worse (Fig. 2(c)), while still being unable to estimate the second-slowest timescale. This shows a limitation of the method in the worst-case scenario that a process is completely hidden: For the x-projection, the stable lower left state and the less stable upper state are projected to exactly the same observable values. Since the upper state only exchanges with the lower state and has much shorter lifetimes, its presence does not even affect the kinetics significantly. Therefore, the projection onto x really behaves like a system with two-state kinetics, and using in a three-state HMM will compromise the estimation.

How can the estimate become worse when *too many* hidden states are used? The answer lies in the structure the hidden

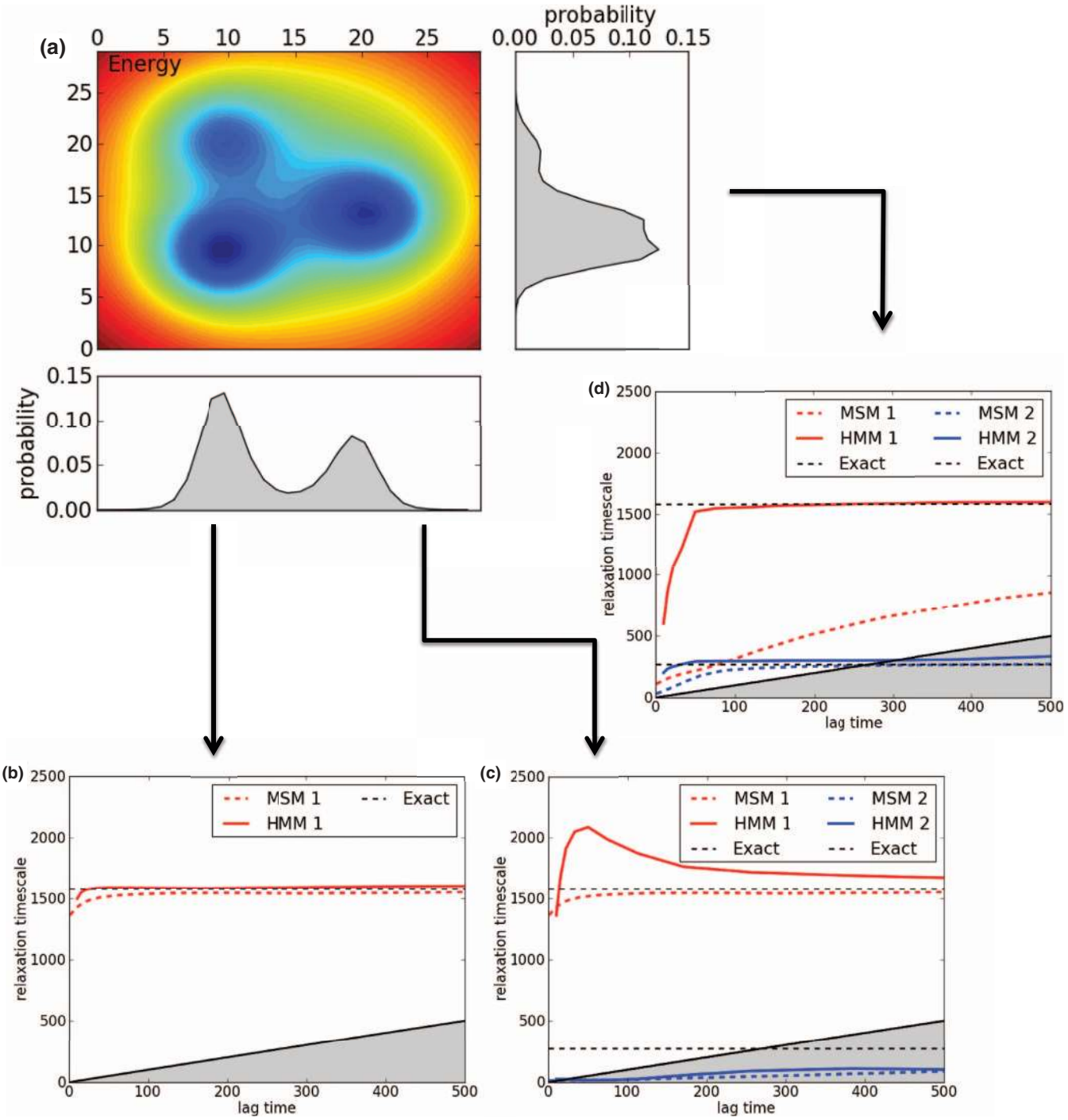

FIG. 2. Comparison of MSM and PMM/HMM for modeling the diffusion in a bistable potential (model details see supplementary material of Ref. 29) from projections onto the *x*- and *y*-coordinates, respectively. A fine (30-state) discretization in the respectively observed coordinate is used in order to build the MSM or PMM/HMM. (a) Energy landscape and observed probability densities in *x* and *y*. (b) and (c) 1- and 2-timescale estimates for the projection onto *x*. (d) Timescale estimation for the projection onto *y*.

HMM transition matrix which has eigenvectors $\tilde{\mathbf{l}}_i$ associated with the slowest processes. When analyzing a two-state system with two hidden states, the HMM transition matrix will have two two-element vectors $\tilde{\mathbf{l}}_1 = \tilde{\boldsymbol{\pi}}$ and $\tilde{\mathbf{l}}_2$, associated with the stationary distribution and the slowest relaxation process, respectively, and these eigenvectors will fulfill the orthogonality condition $\langle \tilde{\mathbf{l}}_2, \tilde{\boldsymbol{\pi}} \rangle = 0$. When analyzing a two-state system with three eigenvectors, the transition matrix will have a third eigenvector $\tilde{\mathbf{l}}_3$, but there is no relaxation process in the data associated to that. Therefore, the HMM estimate will produce a random vector for $\tilde{\mathbf{l}}_3$. Unfortunately, this also affects the quality of the other eigenvectors $\tilde{\mathbf{l}}_1 = \tilde{\boldsymbol{\pi}}$ and $\tilde{\mathbf{l}}_2$, because these eigenvectors are linked by pairwise orthogonality constraints. The MSM is less affected by this problem, because it has many more (n) eigenvectors, so errors in estimating the fast process eigenvectors do not necessarily compromise the

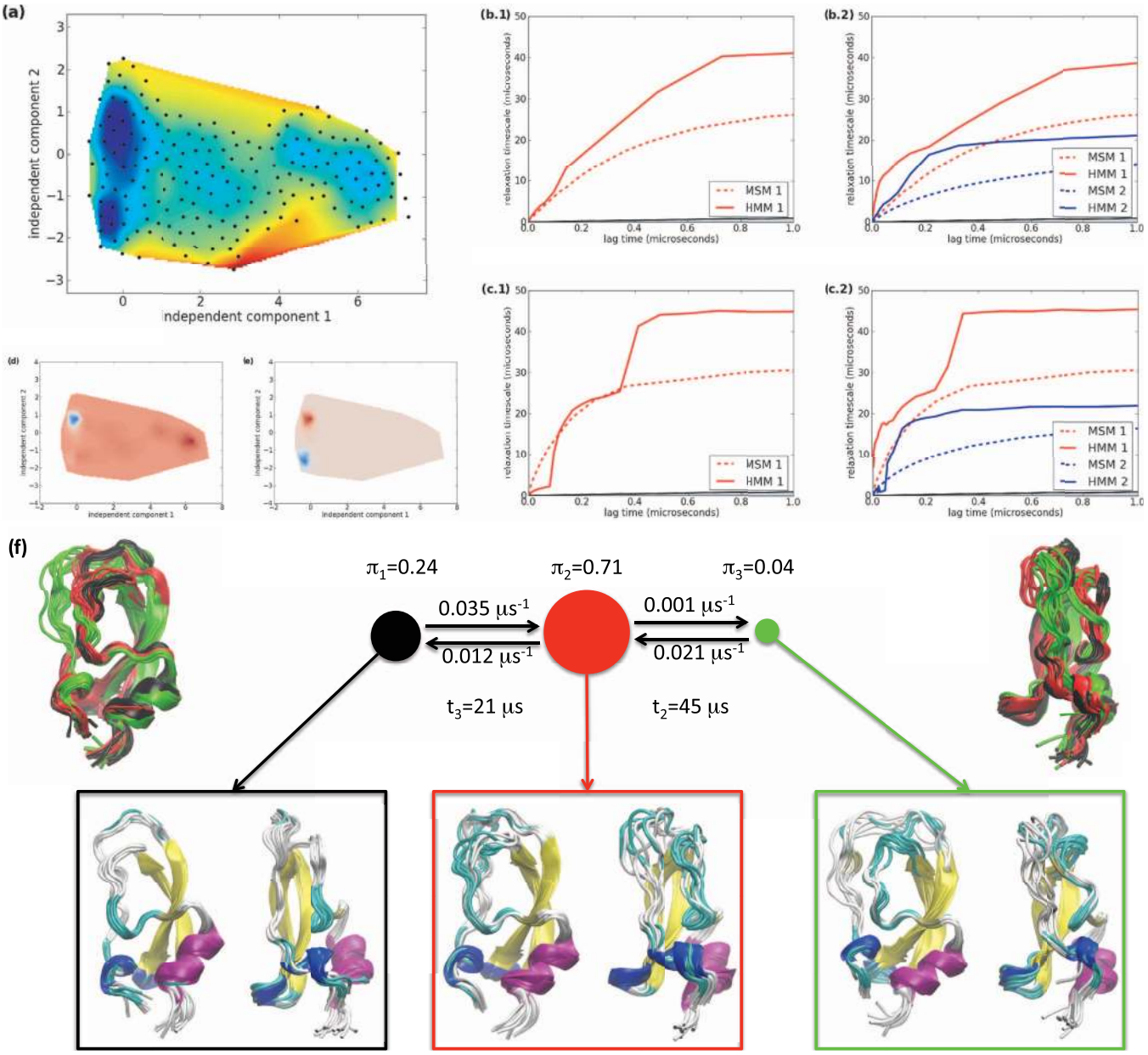

FIG. 3. Comparison of MSM and PMM/HMM for modeling the conformation dynamics of BPTI using a 1 ms simulation trajectory generated by the Anton supercomputer[37] and kindly provided by D.E. Shaw research. The discretization consisted of either 13 or 191 clusters, approximately uniformly distributed on the data projected onto the space of the two dominant independent components of a TICA analysis (see Ref. 27 for details), using the EMMA software implementation.[36] (a) 191 clusters and a visualization of their reduced free energy, $-\ln \pi_i$. (b) and (c) Comparison of the slowest MSM timescales with the timescales of the 2 and 3-states HMM, respectively, using 13 clusters ((b.1) and (c.1)) or 191 clusters ((b.2) and (c.2)). (d) and (e) Visualization of the second and third eigenvectors, $\mathbf{l}_2$ and $\mathbf{l}_3$ for the 191 cluster discretization. (f) Three-state rate matrix corresponding to the 3-state HMM. The structures are overlays of 10 frames drawn from the state distributions $\boldsymbol{\chi}_1$, $\boldsymbol{\chi}_2$, and $\boldsymbol{\chi}_3$.

slow process eigenvectors. This emphasizes that it is important to use HMMs in the right setting: estimating an HMM with m states requires m relaxation processes to be present in the data, *and* having a timescale separation to the $(m+1)$th process.

Figure 2(d) shows that the three-state HMM is able to accurately estimate both relaxation timescales from the y-projection, and is therefore superior to the MSM in this case.

Figure 3 shows the analysis of a 1 ms MD simulation of the BPTI produced on the Anton supercomputer[37] and kindly provided by D.E. Shaw research. We again consider two and three hidden states, because an MSM analysis suggested gaps after the slowest and second-slowest timescale. To obtain a cluster discretization, we first computed the slowest independent components with time-lagged independent component analysis (TICA)[22] as described in Ref. 27 using the EMMA 1.4 software implementation. The data were then projected onto the two slowest components and we considered two cluster discretizations into 13 and 191 clusters. Figure 3 shows a scatter plot of the 191 cluster centers in the two dominant independent components. The color code is a map of the

logarithmized probability map of the clusters, indicating a free energy surface. Note that such free-energy surfaces generally suffer from overlap of states in the directions not resolved in this plot, and only serves to provide a qualitative impression where the regions with most statistical weight are. Figures 3(b) and 3(c) show that the MSMs slowly converge towards slowest timescale estimates of around 30 and 15 μs, while the HMMs converge to robust and nearly τ-constant estimates of timescales around 40 and 20 μs—at lagtimes of 0.7 μs for the 13-cluster partition and at a lagtime of 0.3 μs for the 191-cluster partition. The HMMs, therefore, estimate somewhat larger relaxation timescales and do that robustly for shorter lag times. Figure 3(c.1) nicely shows what happens when employing a two-state HMM in a three-state kinetics system: for short lagtimes, the HMM first finds the faster timescale, and after a lagtime of about $\tau = 0.3$ μs then jumps to the slower timescale.

Figures 3(d) and 3(e) illustrate the two slow processes by plotting the projected eigenvectors $\mathbf{q}_1$ and $\mathbf{q}_2$ (Eq. (24)) on the two dominant independent components. The slowest process is associated with probability exchange along the first independent component (Fig. 3(d)) and the second-slowest process along the second independent component (Fig. 3(e)). Figure 3(f) illustrates the structures associated with the three corresponding metastable states by plotting overlays of 10 structures each selected from the metastable state output distributions $\{\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \boldsymbol{\chi}_3\}$ that are directly estimated by the HMM. Here, the black state is associated with the lower left minimum in Fig. 3(a), and is the most ordered structure. The red state is associated with the top left minimum in Fig. 3(a), and is a slightly less ordered structure, while the green state is associated with the rightmost minimum in Fig. 3(a) and exhibits a re-folded loop on the *N*-terminal side of the backbone.

Figure 3(f) also shows the 3×3 rate matrix between metastable states computed from Eq. (23). This shows that the three metastable states are linearly connected, with the black and red states exchanging on the faster 20 μs timescale, while the red state and the green state exchange on the slower 40 μs timescale. Note that the green state is rather unstable, and actually only one transition into and back out of the green state occurs in the 1 ms trajectory, while the red and black states interchange more frequently. Therefore, the 40 μs timescale is dominated by the relatively short exit time from the green state, and this process is statistically unreliable—thus, the 40 μs timescale is a rough estimate. It is possible to extend the present HMM estimations towards a fully Bayesian approach (analogously to Ref. 7). Thus, in the future, it will be possible to compute error bars on the HMM estimates.

As shown in the second example (Fig. 2), the HMM estimation can also deal with projections of higher-dimensional dynamical systems onto low-dimensional observables, provided these projections do not hide some slow relaxation processes completely. Therefore, we also illustrate the performance of our method on experimental single-molecule data. Note that a key assumption of our approach is that dynamics are reversible, thus only experiments probing equilibrium fluctuations, such as passive-mode force probe experiments or equilibrium single-molecule Förster resonance energy transfer (FRET) experiments, are suitable for this. We have chosen optical tweezer measurements of the extension fluctuations of two biomolecules examined in a recent optical force spectroscopy study: the p5ab RNA hairpin.[12] The p5ab hairpin forms stem-loop structure with a bulge under native conditions (Fig. 4(a)) and zips/unzips repeatedly under the conditions used to collect data (Fig. 4(b)). Experimental force trajectory data were generously provided by the authors of Ref. 12; experimental details are given therein. The instrument used to collect both datasets was a dual-beam counter-propagating optical trap. The molecule of interest was tethered to polystyrene beads by means of dsDNA handles, with one bead suctioned onto a pipette and the other held in the optical trap. A piezoactuator controlled the position of the trap and allowed position resolution to within 0.5 nm, with the instrument operated in passive (equilibrium) mode such that the trap was stationary relative to the pipette during data collection. The force on the bead held in the optical trap was recorded at 50 kHz, with each recorded force trajectory 60 s in duration. The trajectory shown in Fig. 4(b) that was chosen for analysis has a relative similar population in the open and closed states. This experimental construct suffers from a slow drift in the recorded force trajectory. Although the drift is very small for the selected trajectory, it may interfere with an analysis of the slow kinetics; and it will be seen below how the HMM analysis deals with this.

For the analysis, we discretized the observed force coordinate into 30 regularly spaced bins. Figures 4(c) and 4(d) compare the performances of 30-state MSMs with two- or three-state HMMs, respectively. While the MSMs converge only very slowly towards the $\sim$17 ms timescale associated with the exchange of open and closed states, both HMMs estimate this timescale robustly after a lag time of 0.7 ms. Interestingly, for the three-state HMM, there is a switch of eigenvectors at lag times of 2–2.5 ms. While the open/close transition is now estimated as the second-slowest timescale, the slowest timescale vastly increases to a timescale on the order of the entire trajectory length. Inspection of the corresponding eigenvector has confirmed that the process found by this second timescale indeed corresponds to a slight shift of the output distributions that captures the small drift that is present in the trajectory and is associated to a slight shift of the open and closed force distributions between the beginning and the end of the trajectory.

Figure 4(d) shows that the three-state HMM also finds a faster process of less than 1 ms for short lag times. Clearly, this fast process disappears at long lag times, and therefore the blue curve in Fig. 4(d) leaves this initial plateau after $\tau > 0.7$ ms. However, at $\tau = 0.7$ ms both processes are present in the data, and a three-state HMM can be successfully constructed. Figures 4(e) and 4(f) show the corresponding HMM output distributions $\{\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \boldsymbol{\chi}_3\}$, weighted by the stationary probabilities $\{\widetilde{\pi}_1, \widetilde{\pi}_2, \widetilde{\pi}_3\}$, thus illustrating where the two or three metastable states are located in the force coordinate. As expected, the most stable black state (small forces) and the less stable green state (higher forces) correspond to the open and closed states of the hairpin. Interestingly, the three-state HMM identifies a third (red) state that lies "in between" open and closed. This state has so far not been reported. The

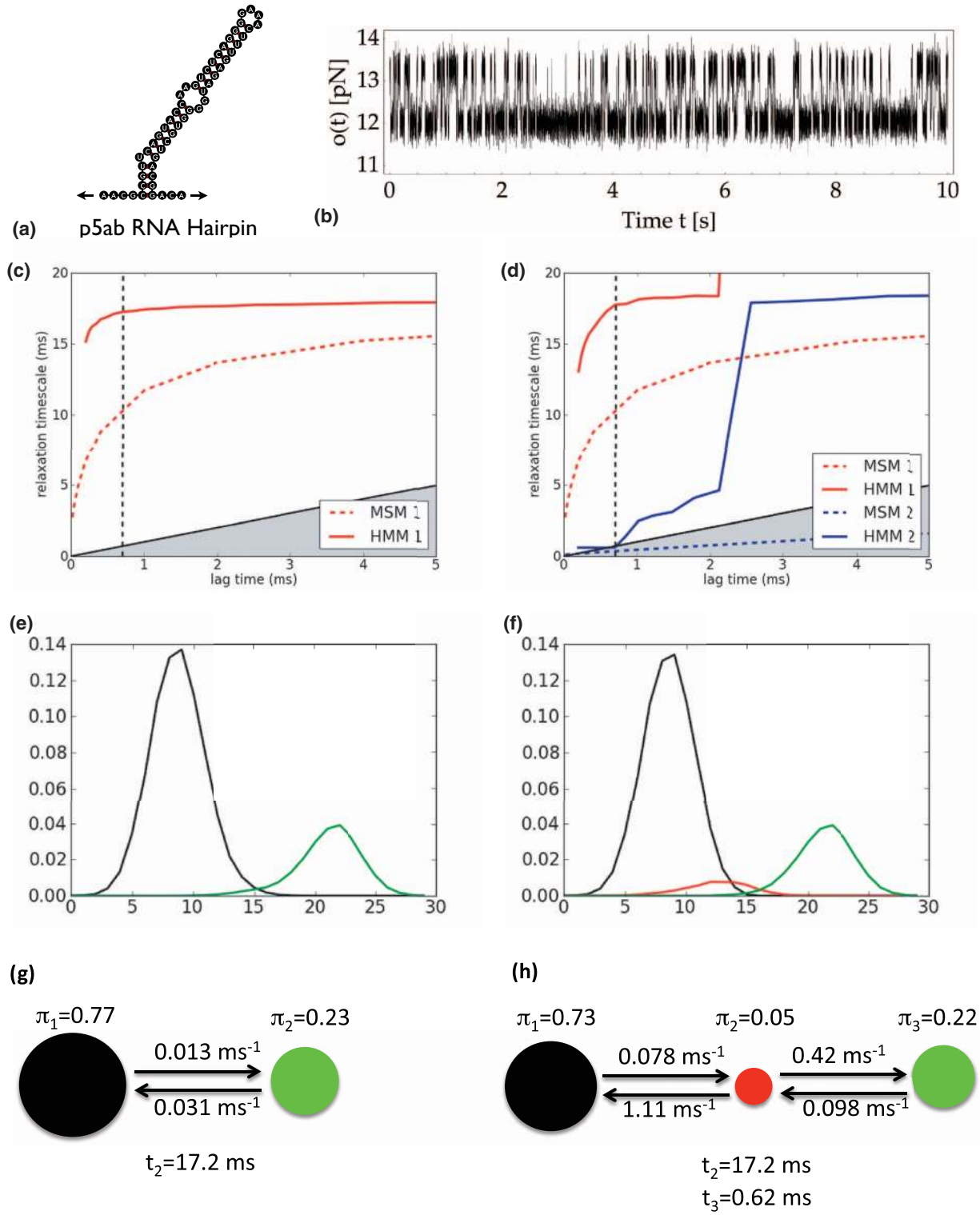

FIG. 4. Comparison of MSM and HMM for analyzing single-molecule force-probe data of the RNA hairpin p5ab.[12] (a) Sketch of the folded secondary structure, (b) the optical tweezer trace. (c) and (d) Relaxation timescales computed by MSMs compared with HMMs with 2 and 3 states, respectively. (e) and (f) HMM output distributions for the 2- and 3-states HMM, respectively. (g) and (h) rate matrices of the 2- and 3-states HMM, respectively.

rate matrix and stationary probabilities shown in Fig. 4(h) reveal that the three states are linearly connected, and the low-populated red state is a transition state. This rate matrix also indicates that the intermediate state has a lifetime of about 0.65 ms.

VI. CONCLUSIONS

We have introduced the concept of PMMs and established a connection between conformation dynamics, PMMs, HMMs, and the widely used MSMs. When observing the

continuous and full-phase space dynamics on some (possibly coarse) set of discrete clusters, the true kinetics is described by PMMs, rather than with MSMs, although MSMs are very widely used on this discrete dynamics. MSMs are, therefore, just an approximation of the discrete dynamics, which are not actually Markovian.

Currently, no efficient approach for directly estimating PMMs is available. Here, we have shown that in the important setting where the dynamics are metastable, with rarely populated transition regions, and there is a timescale separation after the first $m-1$ slow relaxation timescales (m could be any number, but is usually small), PMMs can be approximated with HMMs. This is an important result, because HMMs can be efficiently estimated by maximum-likelihood or Bayesian techniques, and will in this setting give the correct estimate of the slow molecular kinetics—without the systematic bias induced by the Markovianity assumption of MSMs.

HMMs are then estimated with an $m \times m$ transition matrix describing the dynamics between the m hidden states, and each hidden state associated with an n-element probability vector containing the probability that the hidden state will appear in one of the n discrete clusters. In order to successfully and reliably conduct the HMM estimation for large values of n, it is important to have a good starting guess of the transition matrix and the output probability matrix. Here, we have also made a new connection between MSMs and HMMs and shown that the initial HMM transition matrix and output probability matrix can be computed from a set of established algebraic transformations of an MSM transition matrix.

We have shown that a vast number of relevant thermodynamics, kinetic, and mechanistic quantities that are commonly computed from MSMs can also be computed from HMMs. Notably, this includes kinetic experimental observables such as time-correlation functions and time-dependent expectation values of triggered dynamics. These experimentally observable quantities occur in a functional form that can be readily interpreted by assigning experimentally measurable relaxation timescales to the experimentally not directly measurable structural changes.

Thereby, PMMs, and their HMM-approximations, are invoked as a new modeling framework for slow molecular kinetics, and a real alternative to MSMs. Future studies will extend this framework, e.g., by addressing the computation of the statistical error of the present HMMs via a full Bayesian analysis.

ACKNOWLEDGMENTS

We are grateful to the authors of Ref. 12 for providing us with the force-probe optical tweezer data and D.E. Shaw research (especially the authors of Ref. 37) for publishing the 1 ms BPTI trajectory. We would also like to thank Gerhard Hummer (MPI Frankfurt, Germany) for inspiring discussions. We are grateful for the continuous support of our work by Christof Schütte (FU Berlin). This work was funded by Deutsche Forschungsgemeinschaft (Grant Nos. WU 744/1-1 and NO 825/2, research center Matheon, SFB 958), the Einstein foundation Berlin (international postdoctoral fellow program), and the European commission (ERC starting grant "pcCell").

APPENDIX A: PROOF THAT AN OBSERVED METASTABLE MARKOV PROCESS WITH *m* SLOW RELAXATION PROCESSES IS EQUIVALENT TO A *m*-STATE HMM

We consider the dynamics of the Markov process $\mathbf{z}_t \in \Omega$, in the full-dimensional phase space. In this section, we do *not yet* consider any projection to an observation space. The purpose of this section is to investigate if the following two processes are equivalent:

*Definition: **m-timescale Markov process:** A reversible and ergodic Markov process with m dominant slow processes. We assume that we work at a lag time τ, at which the all other processes have decayed. Thus the spectrum is assumed to be* $1, \lambda_2 \ldots, \lambda_m, 0, \ldots, 0)$.

*Definition: **m-state hybrid process:** A $m \times m$ Markov chain where each state has a fixed output distribution $\rho_k(\mathbf{z})$, $\mathbf{z} \in \Omega$. The process consists of propagating the Markov chain in time. At every time instant, we draw an independent random number from $\rho_k(\mathbf{z})$ where k is the current discrete state.*

These two are equivalent if their transition kernels are identical,

$$p_\tau^{(1)}(\mathbf{z}_0, \mathbf{z}_\tau) = p_\tau^{(2)}(\mathbf{z}_0, \mathbf{z}_\tau) \tag{A1}$$

or, equivalently, if their correlation densities are identical,

$$\begin{aligned} \mu^{(1)}(\mathbf{z}_0) p_\tau^{(1)}(\mathbf{z}_0, \mathbf{z}_\tau) &= \mu^{(2)}(\mathbf{z}_0) p_\tau^{(2)}(\mathbf{z}_0, \mathbf{z}_\tau), \\ c_\tau^{(1)}(\mathbf{z}_0, \mathbf{z}_\tau) &= c_\tau^{(2)}(\mathbf{z}_0, \mathbf{z}_\tau), \end{aligned} \tag{A2}$$

where μ is the stationary distribution of the respective process. We write down the corresponding correlation densities:

1. ***m*-timescale Markov process:** see Ref. 25,

$$c_\tau(\mathbf{z}_0, \mathbf{z}_\tau) = \mu(\mathbf{z}_0)\mu(\mathbf{z}_\tau) + \sum_{k=2}^{m} \mathrm{e}^{-\kappa_k \tau} \phi_k(\mathbf{z}_0)\phi_k(\mathbf{z}_\tau). \tag{A3}$$

2. ***m*-state hybrid process:** We use the $m \times m$ transition matrix $\widetilde{\mathbf{T}}(\tau)$ that is reversible with respect to its stationary distribution $\widetilde{\boldsymbol{\pi}}$, and the corresponding correlation matrix $\widetilde{\mathbf{C}}(\tau) = \widetilde{\mathbf{\Pi}}\widetilde{\mathbf{T}}(\tau)$. At every time step, the process generates output by drawing independent random variables from the continuous output functions ρ_i associated to the current state i,

$$\begin{aligned} c_\tau(\mathbf{z}_0, \mathbf{z}_\tau) &= \sum_{i,j} \rho_i(\mathbf{z}_0) \widetilde{c}_{ij}(\tau) \rho_j(\mathbf{z}_\tau) \\ &= \sum_k \sum_{i,j} \rho_i(\mathbf{z}_0) \tilde{l}_{ki} \lambda_k(\tau) \tilde{l}_{kj} \rho_j(\mathbf{z}_\tau) \\ &= \sum_{i,j} \pi_i \rho_i(\mathbf{z}_0) \pi_j \rho_j(\mathbf{z}_\tau) \\ &\quad + \sum_k \mathrm{e}^{-\kappa_k \tau} \sum_{i,j} \tilde{l}_{ki} \rho_i(\mathbf{z}_0) \tilde{l}_{kj} \rho_j(\mathbf{z}_\tau). \end{aligned} \tag{A4}$$

In order to show (1) ≡ (2), we must show that the expansion of eigenfunctions into a basis of state output functions,

$$\phi_k(\mathbf{z}) = \sum_i \tilde{l}_{ki}\rho_i(\mathbf{z}) \tag{A5}$$

is feasible. We immediately see that this implies a **necessary condition**: in the expansion above, the normalization conditions of eigenfunctions imply,

$$\begin{aligned}\langle\phi_k|\phi_o\rangle_{\mu^{-1}} &= \left\langle \sum_i \tilde{l}_{ki}\rho_i(\mathbf{z})| \sum_j \tilde{l}_{oj}\rho_j(\mathbf{z})\right\rangle_{\mu^{-1}} \\ &= \sum_{i,j}\tilde{l}_{ki}\tilde{l}_{oj}\langle\rho_i|\rho_j\rangle_{\mu^{-1}} \\ &= \sum_{i,j}\tilde{l}_{ki}\tilde{l}_{oj}s_{ij} \\ &= \tilde{\mathbf{l}}_k^T\mathbf{S}\tilde{\mathbf{l}}_o,\end{aligned} \tag{A6}$$

where

$$s_{ij} := \langle\rho_i|\rho_j\rangle_{\mu^{-1}} \tag{A7}$$

is the overlap matrix of basis functions. We have to fulfill

$$\begin{aligned}\mathbf{LSL}^T &= \mathbf{Id} \\ \mathbf{S} = (\mathbf{L}^T\mathbf{L})^{-1} &= (\mathbf{L}^T\mathbf{R}^T\mathbf{\Pi})^{-1} = \mathbf{\Pi}^{-1},\end{aligned} \tag{A8}$$

but that means that **S** has to be a diagonal matrix. Since the output distributions ρ_i are non-negative, **S** can only be diagonal if the sets on which the ρ_i are non-zero do not overlap in the full state space. This condition is necessary for both directions of the proof.

This observation suggests that the two processes ***m*-timescale Markov process** and ***m*-state hybrid process** are generally not equivalent, but equivalence is possible when the nonoverlap condition $\langle\rho_i|\rho_j\rangle = 0$ for $i \neq j$ is used as a condition. Additionally, it has been observed that the weighted eigenfunction $\psi_i = \mu^{-1}\phi_i$ is approximately constant on the metastable sets, a property that will be required later. Therefore, we define a variation of the m-process Markov

*Definition: **m-metastable Markov process:** is a m-timescale Markov process with the following additional properties: Let $\{\rho_i\}_{i=1}^m$ be a set of non-overlapping probability density functions, and let $\{A_1, \ldots, A_m\}$ be a partition of Ω defined as*

$$\mathbf{z} \in A_i \Leftrightarrow \rho_i(\mathbf{z}) > 0 \tag{A9}$$

and

$$\mathbf{z} \in A_i \Leftrightarrow \frac{\phi_i(\mathbf{z})}{\mu(\mathbf{z})} = const. \tag{A10}$$

In this definition, the sets $A_1, \ldots, A_m$ are metastable sets and the boundaries between them are the transition states. The definition represents an idealized metastable Markov process: The decomposability of μ into m distinct modes implies that transition states between the metastable sets have no probability density: $\mu(\mathbf{z}) = 0$. Furthermore, the assumption that the weighted eigenfunctions $\phi_i\mu^{-1}$ are constant on the sets $A_1, \ldots, A_m$ is an idealization of the fact that these eigenfunctions have been observed to be almost constant on metastable sets.[34]

Therefore, no classical physical system can be an m-metastable Markov process—whenever transitions between the sets $A_1, \ldots, A_m$ are possible, the dynamical process must travel through the transition regions, and therefore, $\mu(\mathbf{z})$ will not be exactly zero in these regions. However, a real metastable system may have transition states that are rarely populated, and thereby approximate the idealized m-metastable Markov process.

Below, we will show the following:

1. *m-hybrid process* ⇒ *m-metastable Markov process*
2. *m-metastable Markov process* ⇒ *m-hybrid process*
3. m-metastable PMM ⇔ m-state HMM

From 1 and 2 it is obvious that *m-metastable Markov process* ⇔ *m-hybrid process.* The third step follows from a projection of these full phase-space processes on the observed discrete clusters.

1. A *m*-hybrid process is a *m*-metastable Markov process

Suppose that the stochastic process $\{\mathbf{z}_t\}$ in the state space Ω is generated by a m-hybrid dynamics with transition matrix $\widetilde{\mathbf{T}}(\tau)$ and output distribution functions $\{\rho_i\}_{i=1}^m$, where $\widetilde{\mathbf{T}}(\tau)$ is reversible with respect to its invariant distribution $\widetilde{\boldsymbol{\pi}}$ which can be decomposed as

$$\widetilde{\mathbf{T}}(\tau) = \widetilde{\mathbf{\Pi}}^{-1}\sum_{i=1}^m e^{-\kappa_i\tau}\tilde{\mathbf{l}}_i\tilde{\mathbf{l}}_i^{\mathrm{T}}, \tag{A11}$$

where $\widetilde{\mathbf{\Pi}} = \mathrm{diag}(\boldsymbol{\pi})$, $\tilde{\mathbf{l}}_i$ denotes the ith left eigenvector of $\widetilde{\mathbf{T}}(\tau)$. The eigenvectors are normalized such that they satisfy $\tilde{\mathbf{l}}_1 = \widetilde{\boldsymbol{\pi}}$ and $\tilde{\mathbf{l}}_i^{\mathrm{T}}\widetilde{\mathbf{\Pi}}^{-1}\tilde{\mathbf{l}}_j = \delta_{ij}$. We now prove that $\{\mathbf{z}(t)\}$ is also a m-metastable Markov process. Note that for any $t_1 > t_2 \geq 0$ and $B \subset \Omega$,

$$\begin{aligned}\mathbb{P}\left(\mathbf{z}_{t_1} \in B|\{\mathbf{z}_t\}_{t=0}^{t_2}\right) &= \sum_i \mathbb{P}\left(\mathbf{z}_{t_1} \in B, s_{t_2} = i|\{\mathbf{z}_t\}_{t=0}^{t_2}\right) \\ &= \sum_i \mathbb{P}\left(\mathbf{z}_{t_1} \in B|s_{t_2} = i, \{\mathbf{z}_t\}_{t=0}^{t_2}\right) \\ &\quad\times \mathbb{P}\left(s(t_2) = i|\{\mathbf{z}_t\}_{t=0}^{t_2}\right) \qquad \text{(A12)} \\ &= \sum_i \mathbb{P}\left(\mathbf{z}_{t_1} \in B|s_{t_2} = i\right) \\ &\quad\times \mathbb{P}(s_{t_2} = i|\mathbf{z}_{t_2}) \qquad \text{(A13)} \\ &= \mathbb{P}(\mathbf{z}_{t_1} \in B|\mathbf{z}_{t_2}), \qquad \text{(A14)}\end{aligned}$$

where s_t denotes the state of the HMM at time t. Therefore, $\{\mathbf{z}_t\}$ is a Markov process.

Furthermore, we have the correlation density given in (A4) with the eigenfunction representation (A5). Using the non-overlap condition (A6), these eigenfunctions have the correct normalization:

$$\langle\phi_i|\phi_j\rangle_{\mu^{-1}} = \mathbf{l}_i^T\mathbf{S}\mathbf{l}_j = \delta_{ij}. \tag{A15}$$

Therefore, $\{\mathbf{z}_t\}$ is a m-metastable Markov process.

2. A *m*-metastable Markov process is a *m*-hybrid process

Suppose that the stochastic process $\{\mathbf{z}_t\}$ is a m-metastable Markov process in state space Ω. Then its correlation density is given by (A3) and the propagator eigenfunctions $\{\phi_i\}_{i=1}^m$, where $\phi_1 = \mu$ is the stationary distribution of $\{\mathbf{z}(t)\}$, satisfy the orthogonality conditions $\langle\phi_i|\phi_j\rangle_{\mu^{-1}} = \delta_{ij}$. From (A9), we can directly follow

$$\langle\rho_i|\rho_j\rangle_{\mu^{-1}} = 0 \quad \forall i \neq j \tag{A16}$$

and thus, every density can be described as a linear combination of basis functions,

$$\mu(\mathbf{z}) \in \mathrm{span}\,(\rho_1, \ldots, \rho_m)\,. \tag{A17}$$

Combining this result with (A10), it follows that the entire set of propagator eigenfunctions must be expressable in terms of such linear combinations:

$$\phi_1, \ldots, \phi_m \in \mathrm{span}\,(\rho_1, \ldots, \rho_m)\,. \tag{A18}$$

We call the coefficients required to represent the eigenfunctions $\{\phi_i\}_{i=1}^m$ in the basis $\{\rho_j\}_{j=1}^m$, $\tilde{l}_{ij}$,

$$\phi_i = \sum_j \tilde{l}_{ij}\rho_j \tag{A19}$$

and define,

$$\widetilde{\boldsymbol{\pi}} = [\tilde{\pi}_i] := \tilde{\mathbf{l}}_1, \tag{A20}$$

$$\widetilde{\mathbf{T}}(\tau) = [\widetilde{T}_{ij}(\tau)] := \widetilde{\boldsymbol{\Pi}}^{-1} \cdot \sum_{k=1}^m e^{-\kappa_k\tau}\tilde{\mathbf{l}}_k\tilde{\mathbf{l}}_k^{\mathrm{T}}, \tag{A21}$$

with $\widetilde{\boldsymbol{\Pi}} = \mathrm{diag}\,(\widetilde{\boldsymbol{\pi}})$. From these definitions, it follows that

$$\begin{aligned}\tilde{\pi}_i &= \int_{A_i}\sum_j \widetilde{\pi}_j\rho_j(\mathbf{z})\,\mathrm{d}\mathbf{z}\\ &= \mathbb{P}(\mathbf{z}_t \in A_i)\end{aligned} \tag{A22}$$

and

$$\begin{aligned}\tilde{\pi}_i\tilde{T}_{ij}(\tau) &= \sum_{k=1}^m \exp(-\kappa_k\tau)\,\tilde{l}_{ki}\tilde{l}_{kj}\\ &= \sum_{k=1}^m \exp(-\kappa_k\tau)\int_{A_i}\tilde{l}_{ki}\rho_i(\mathbf{z}_0)\,\mathrm{d}\mathbf{z}_0\\ &\quad\times\int_{A_j}\tilde{l}_{kj}\rho_j(\mathbf{z}_\tau)\,\mathrm{d}\mathbf{z}_\tau\end{aligned} \tag{A23}$$

$$\begin{aligned}&= \sum_{k=1}^m \exp(-\kappa_k\tau)\int_{A_i}\left(\sum_a \tilde{l}_{ka}\rho_a(\mathbf{z}_0)\right)\mathrm{d}\mathbf{z}_0\\ &\quad\times\int_{A_j}\left(\sum_b \tilde{l}_{kb}\rho_b(\mathbf{z}_\tau)\right)\mathrm{d}\mathbf{z}_\tau\end{aligned} \tag{A24}$$

$$\begin{aligned}&= \sum_{k=1}^m \exp(-\kappa_k\tau)\int_{A_i}\int_{A_j}\phi_k(\mathbf{z}_0)\,\phi_k(\mathbf{z}_\tau)\,\mathrm{d}\mathbf{z}_\tau\,\mathrm{d}\mathbf{z}_0\\ &= \int_{A_i}\int_{A_j} c(\mathbf{z}_0, \mathbf{z}_\tau)\,\mathrm{d}\mathbf{z}_\tau\,\mathrm{d}\mathbf{z}_0\\ &= \mathbb{P}(\mathbf{z}_0 \in A_i, \mathbf{z}_\tau \in A_j).\end{aligned} \tag{A25}$$

Therefore, $\widetilde{\boldsymbol{\pi}}$ is a discrete distribution and $\widetilde{\mathbf{T}}$ is a reversible transition matrix with respect to $\widetilde{\boldsymbol{\pi}}$, and we can construct a m-state hybrid Markov process with transition matrix $\widetilde{\mathbf{T}}$ and output distributions $\{\rho_i\}_{i=1}^m$. Noting that

$$\begin{aligned}\langle\tilde{\mathbf{l}}_i|\tilde{\mathbf{l}}_j\rangle_{\tilde{\boldsymbol{\pi}}^{-1}} &= \int\frac{\left(\sum_a \tilde{l}_{ia}\rho_a(\mathbf{z})\right)\left(\sum_b \tilde{l}_{jb}\rho_b(\mathbf{z})\right)}{\sum_c \widetilde{\pi}_c\chi_c(\mathbf{z})}\mathrm{d}\mathbf{z}\\ &= \langle\phi_i|\phi_j\rangle_{\mu^{-1}}\\ &= \delta_{ij}\end{aligned} \tag{A26}$$

and according to the conclusion in Appendix A 1, we can conclude that the dynamics of $\{\mathbf{z}_t\}$ can be exactly described by a m-state hybrid Markov process.

3. *m*-metastable PMM ≡ *m*-state HMM

We now consider that the dynamics are observed on a set of n discrete states $\{S_1, \ldots, S_n\}$.

It is straightforward to classify the processes after projecting them onto an observable y:

1. When projecting a m-timescale Markov process onto the discrete partition $\{S_1, \ldots, S_n\}$, we obtain a **PMM** with m relaxation timescales (5). Therefore, when ***m*-metastable Markov process** onto the partition $\{S_1, \ldots, S_n\}$, we also obtain a PMM. We call this specific PMM a m-metastable PMM.
2. When projecting a ***m*-hybrid process** onto sets $\{S_i\}$, we obtain a m-state **HMM** with m hidden states, the $m \times m$ transition matrix of the m-hybrid process as a hidden transition matrix, and the output probability matrix

$$\chi_{ki} = \int_{\mathbf{z}\in S_i}\rho_k(\mathbf{z})\,d\mathbf{z}. \tag{A27}$$

In Appendices A 1 and A 2 we have shown that for the metastable case, we have the equality

***m*-metastable Markov dynamics ≡ *m*-hybrid dynamics**

and thus we have shown

***m*-metastable PMM ≡ *m*-state HMM**

APPENDIX B: ALGORITHMS AND DERIVATIONS

1. Estimation algorithm

We summarize by sketching the PMM/HMM estimation algorithm

ALGORITHM I. PMM/HMM estimation.

Input:
- N trajectories, discretized into n clusters: $S = \{\{s_t^{(1)}\}, \ldots, \{s_t^{(N)}\}\}$
- lag time: τ
- number of slow relaxation processes considered: m

Algorithm:
1. Estimate reversible Markov transition matrix $\mathbf{T}(\tau) \in \mathbb{R}^{n\times n}$ from the discrete trajectory S
2. Decompose $\mathbf{T}(\tau)$ into an initial guess for the HMM matrices: $\boldsymbol{\chi} \in \mathbb{R}^{n\times m}$ and $\widetilde{\mathbf{T}}(\tau) \in \mathbb{R}^{m\times m}$ using PCCA and Eqs. (12) and (13).
3. Optimize $\boldsymbol{\chi}$ and $\widetilde{\mathbf{T}}(\tau)$ using the EM algorithm.
4. Validate model by comparing correlation matrices $\mathbf{C}^{\mathrm{pred}}(\tau) = \boldsymbol{\chi}\widetilde{\boldsymbol{\Pi}}[\widetilde{\mathbf{T}}(\tau_0)]^n\boldsymbol{\chi}^\top$ and $\mathbf{C}(\tau) = \boldsymbol{\Pi}\mathbf{T}(\tau)$, or the apparent relaxation timescales computed from $\mathbf{T}^{\mathrm{pred}}(\tau) = \boldsymbol{\Pi}^{-1}\boldsymbol{\chi}\widetilde{\boldsymbol{\Pi}}[\widetilde{\mathbf{T}}(\tau_0)]^n\boldsymbol{\chi}^\top$ and the direct MSM $\mathbf{T}(\tau)$.

2. Computing the HMM transition matrix from PCCA memberships

We use the definition of the coarse-grained transition matrix derived in Ref. 17,

$$\widetilde{\mathbf{T}} = (\mathbf{RI})^{-\top}\mathbf{I}^\top\mathbf{PR}^\top \tag{B1}$$

with restriction and interpolation operators,

$$\begin{aligned}\mathbf{R} &= \mathbf{M}^\top\\ \mathbf{I} &= \boldsymbol{\Pi}\mathbf{M}\widetilde{\boldsymbol{\Pi}}^{-1}.\end{aligned} \tag{B2}$$

By a series of algebraic transformations we obtain:

$$\begin{aligned}\widetilde{\mathbf{T}} &= (\mathbf{RI})^{-\top}\mathbf{I}^\top\mathbf{PR}^\top\\ &= (\mathbf{M}^T\boldsymbol{\Pi}\mathbf{M}\widetilde{\boldsymbol{\Pi}}^{-1})^{-T}(\boldsymbol{\Pi}\mathbf{M}\widetilde{\boldsymbol{\Pi}}^{-1})^T\mathbf{PM}\\ &= \mathbf{M}^T\mathbf{P}^T\boldsymbol{\Pi}\mathbf{M}\widetilde{\boldsymbol{\Pi}}^{-1}(\mathbf{M}^T\boldsymbol{\Pi}\mathbf{M}\widetilde{\boldsymbol{\Pi}}^{-1})^{-1}\\ &= \mathbf{M}^T\mathbf{C}\mathbf{M}\widetilde{\boldsymbol{\Pi}}^{-1}(\mathbf{M}^T\boldsymbol{\Pi}\mathbf{M}\widetilde{\boldsymbol{\Pi}}^{-1})^{-1}\\ &= \mathbf{M}^T\mathbf{C}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Pi}\mathbf{M}\widetilde{\boldsymbol{\Pi}}^{-1}\widetilde{\boldsymbol{\Pi}})^{-1}\\ &= \mathbf{M}^T\mathbf{C}\mathbf{M}(\mathbf{M}^T\boldsymbol{\Pi}\mathbf{M})^{-1}\\ \mathbf{M}\widetilde{\mathbf{T}}\mathbf{M}^T\boldsymbol{\Pi}\mathbf{M}\mathbf{M}^T &= \mathbf{M}\mathbf{M}^T\mathbf{C}\mathbf{M}\mathbf{M}^T\\ \widetilde{\mathbf{T}}\mathbf{M}^T &= (\mathbf{TM})^T\\ \widetilde{\mathbf{T}} &= \mathbf{M}^T\mathbf{TM}(\mathbf{M}^T\mathbf{M})^{-1}.\end{aligned} \tag{B3}$$

3. EM implementation

In order to estimate a discrete HMM using the Baum-Welch EM method, we iterate the following two steps

1. Expectation step: Estimate the hidden path probabilities $\{\boldsymbol{\alpha}_t\}^{(k)}$ and $\{\boldsymbol{\beta}_t\}^{(k)}$,

$$\begin{aligned}&\{\{\boldsymbol{\alpha}_t\}^{(k)}, \{\boldsymbol{\beta}_t\}^{(k)}\}\\ &\quad= \arg\max_{\{\boldsymbol{\alpha}_t\},\{\boldsymbol{\beta}_t\}} \mathbb{P}(\{s_t\}|\{\boldsymbol{\alpha}_t\}, \{\boldsymbol{\beta}_t\}, \widetilde{\mathbf{T}}^{(k)}, \boldsymbol{\chi}^{(k)})\end{aligned} \tag{B4}$$

and compute the log-likelihood $L = \log\mathbb{P}(\{s_t\}|\widetilde{\mathbf{T}}, \boldsymbol{\chi})$.

2. Maximization step: Estimate $\widetilde{\mathbf{T}}^{(k+1)}$ and $\boldsymbol{\chi}^{(k+1)}$:

$$\{\widetilde{\mathbf{T}}^{(k+1)}, \boldsymbol{\chi}^{(k+1)}\} = \arg\max_{\widetilde{\mathbf{T}},\boldsymbol{\chi}} \mathbb{P}(\widetilde{\mathbf{T}}, \boldsymbol{\chi}|\{s_t\}, \{\boldsymbol{\alpha}_t\}^{(k)}, \{\boldsymbol{\beta}_t\}^{(k)}) \tag{B5}$$

until the increase of the likelihood (14) falls below a user-defined threshold. While the expectation step is general, the maximization step must be designed for the specific HMM implementation. Here, we estimate the quantities $\widetilde{\mathbf{T}}$, $\boldsymbol{\chi}$ as follows:

1. From the expectation step, we compute the Baum-Welch count matrix between hidden states,[18,43]

$$\widetilde{\mathbf{Z}}_t = g^{-1}\boldsymbol{\alpha}_t\widetilde{\mathbf{T}}^{(k)}\boldsymbol{\beta}_{t+1}\boldsymbol{\chi}_{s_t}, \tag{B6}$$

$$g = \mathbf{1}^T\boldsymbol{\alpha}_t\widetilde{\mathbf{T}}^{(k)}\boldsymbol{\beta}_{t+1}\boldsymbol{\chi}_{s_t}\mathbf{1}, \tag{B7}$$

where S is a normalization factor ensuring that we only count 1 transition per time step. The total count matrix is given by the sum over all single-step count matrices,

$$\widetilde{\mathbf{Z}} = \sum_t \widetilde{\mathbf{Z}}_t, \tag{B8}$$

which may run over multiple trajectories. Given the count matrix $\widetilde{\mathbf{Z}}$, we estimate the maximum likelihood transition matrix that fulfills detailed balance using the algorithm described in Ref. 29 and implemented in the EMMA software,[36]

$$\begin{aligned}\widetilde{\mathbf{T}} &= \arg\max_{\widetilde{\mathbf{T}}} \mathbb{P}(\widetilde{\mathbf{Z}}|\widetilde{\mathbf{T}})\\ &\quad\text{such that } \widetilde{\boldsymbol{\Pi}}\widetilde{\mathbf{T}} = \widetilde{\mathbf{T}}^\top\widetilde{\boldsymbol{\Pi}}.\end{aligned} \tag{B9}$$

2. The stationary probability is computed from the hidden transition matrix

$$\widetilde{\boldsymbol{\pi}}^\top = \widetilde{\boldsymbol{\pi}}^\top\widetilde{\mathbf{T}}. \tag{B10}$$

3. The output probability distributions are computed by first estimating histograms,

$$y_{ij} = y_{ij}^0 + \sum_t \frac{\alpha_{t,i}\beta_{t,i}1(s_t = j)}{\sum_k \alpha_{t,k}\beta_{t,k}}, \tag{B11}$$

where y_{ij} is the estimated number of times that hidden state i has produced cluster j as an output. $1(s_t = j)$ is an indicator function that is 1 if the cluster trajectory is at state j at time t, and 0 otherwise. y_{ij}^0 is a prior count, here uniformly set to n^{-1}. Then, the histograms are normalized to

$$\chi_{ij} = \frac{y_{ij}}{\sum_k y_{kj}}. \tag{B12}$$

4. Derivation of experimental observables

Correlation function between observables a and b for lag time $\tau = n\tau_0$,

$$\begin{aligned}\mathbb{E}[a(t)b(t+\tau)] &= \mathbf{a}^\top \boldsymbol{\chi}^\top \widetilde{\boldsymbol{\Pi}}[\widetilde{\mathbf{T}}(\tau_0)]^n \boldsymbol{\chi}\mathbf{b} \\ &= \mathbf{a}^\top \boldsymbol{\chi}^\top \sum_{i=1}^{m} \lambda_i(\tau)\tilde{\mathbf{l}}_i\tilde{\mathbf{l}}_i^\top \boldsymbol{\chi}\mathbf{b} \\ &= \sum_{i=1}^{m} \mathrm{e}^{-\tau\kappa_i}\langle \mathbf{a}, \mathbf{q}_i^l\rangle\langle \mathbf{b}, \mathbf{q}_i^l\rangle. \end{aligned} \tag{B13}$$

Relaxation function of observable a, starting from the hidden-state probability distribution $\widetilde{\mathbf{p}}_0$,

$$\begin{aligned}\mathbb{E}_{\widetilde{\mathbf{p}}_0}[a(\tau)] &= \mathbf{a}^\top(\widetilde{\mathbf{p}}_0^\top[\widetilde{\mathbf{T}}(\tau_0)]^n\boldsymbol{\chi})^\top \\ &= \mathbf{a}^\top \boldsymbol{\chi}^\top(\widetilde{\mathbf{p}}_0^\top[\widetilde{\mathbf{T}}(\tau_0)]^n)^\top \\ &= \mathbf{a}^\top \boldsymbol{\chi}^\top(\widetilde{\mathbf{p}}_0^\top \widetilde{\boldsymbol{\Pi}}^{-1}\sum_{i=1}^{m} \lambda_i(\tau)\tilde{\mathbf{l}}_i\tilde{\mathbf{l}}_i^\top)^\top \\ &= \sum_{i=1}^{m} \lambda_i(\tau)\mathbf{a}^\top \boldsymbol{\chi}^\top\tilde{\mathbf{l}}_i\tilde{\mathbf{l}}_i^\top \widetilde{\boldsymbol{\Pi}}^{-1}\widetilde{\mathbf{p}}_0 \\ &= \sum_{i=1}^{m} \mathrm{e}^{-\tau\kappa_i}\langle \mathbf{a}, \mathbf{q}_i^l\rangle\langle \tilde{\mathbf{l}}_i^\top, \widetilde{\mathbf{p}}_0^*\rangle, \end{aligned} \tag{B14}$$

where $\widetilde{\mathbf{p}}_0^*$ is the *excess probability distribution* $\widetilde{\mathbf{p}}_0^* = \Pi^{-1}\widetilde{\mathbf{p}}_0$.

[1] K. A. Beauchamp, R. McGibbon, Y.-S. Lin, and V. S. Pande, "Simple few-state models reveal hidden complexity in protein folding," Proc. Natl. Acad. Sci. U.S.A. **109**, 17807–17813 (2012).

[2] G. R. Bowman, "Improved coarse-graining of Markov state models via explicit consideration of statistical uncertainty," J. Chem. Phys. **137**, 134111 (2012).

[3] G. R. Bowman, D. L. Ensign, and V. S. Pande, "Enhanced modeling via network theory: Adaptive sampling of Markov state models," J. Chem. Theory Comput. **6**(3), 787–794 (2010).

[4] G. R. Bowman and P. L. Geissler, "Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites," Proc. Natl. Acad. Sci. U.S.A. **109**, 11681–11686 (2012).

[5] N. V. Buchete and G. Hummer, "Coarse master equations for peptide folding dynamics," J. Phys. Chem. B **112**, 6057–6069 (2008).

[6] J. D. Chodera, K. A. Dill, N. Singhal, V. S. Pande, W. C. Swope, and J. W. Pitera, "Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics," J. Chem. Phys. **126**, 155101 (2007).

[7] J. D. Chodera, P. Elms, F. Noé, B. Keller, C. M. Kaiser, A. Ewall-Wice, S. Marqusee, C. Bustamante, and N. Singhal Hinrichs, "Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty," e-print arXiv:1108.1430.

[8] P. Deuflhard and M. Weber, "Robust Perron cluster analysis in conformation dynamics," in *Linear Algebra Applications*, edited by M. Dellnitz, S. Kirkland, M. Neumann, and C. Schütte (Elsevier, New York, 2005), Vol. 398C, pp. 161–184.

[9] P. Deulfhard, W. Huisinga, A. Fischer, and C. Schütte, "Identification of almost invariant aggregates in reversibly nearly uncoupled Markov chains," Linear Algebr. Appl. **315**, 39–59 (2000).

[10] N. Djurdjevac, M. Sarich, and C. Schütte, "Estimating the eigenvalue error of Markov State Models," Multiscale Model. Simul. **10**, 61–81 (2012).

[11] 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, "Intrinsic dynamics of an enzyme underlies catalysis," Nature (London) **438**(7064), 117–121 (2005).

[12] P. J. Elms, J. D. Chodera, C. Bustamante, and S. Marqusee, "The limitations of constant-force-feedback experiments," Biophys. J. **103**, 1490 (2012).

[13] A. Gansen, A. Valeri, F. Hauger, S. Felekyan, S. Kalinin, K. Tóth, J. Langowski, and C. A. M. Seidel, "Nucleosome disassembly intermediates characterized by single-molecule FRET," Proc. Natl. Acad. Sci. U.S.A. **106**(36), 15308–15313 (2009).

[14] J. C. Gebhardt, T. Bornschlögl, and M. Rief, "Full distance-resolved folding energy landscape of one single protein molecule," Proc. Natl. Acad. Sci. U.S.A. **107**(5), 2013–2018 (2010).

[15] N. S. Hinrichs and V. S. Pande, "Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics," J. Chem. Phys. **126**, 244101 (2007).

[16] B. Keller, J.-H. Prinz, and F. Noé, "Markov models and dynamical fingerprints: Unraveling the complexity of molecular kinetics," Chem. Phys. **396**, 92–107 (2012).

[17] S. Kube and M. Weber, "A coarse graining method for the identification of transition rates between molecular conformations," J. Chem. Phys. **126**(2), 024103 (2007).

[18] G. Soules, L. E. Baum, T. Petrie, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," Ann. Math. Stat. **41**, 164–171 (1970).

[19] B. Lindner, Z. Yi, J.-H. Prinz, J. C. Smith, and F. Noé, "Dynamic neutron scattering from conformational dynamics I: Theory and Markov models," J. Chem. Phys. **139**, 175101 (2013).

[20] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," Science **334**, 517–520 (2011).

[21] P. Metzner, C. Schütte, and E. Vanden-Eijnden, "Illustration of transition path theory on a collection of simple examples," J. Chem. Phys. **125**(8), 084110 (2006).

[22] L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," Phys. Rev. Lett. **72**, 3634–3637 (1994).

[23] F. Noé, S. Doose, I. Daidone, M. Löllmann, J. D. Chodera, M. Sauer, and J. C. Smith, "Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments," Proc. Natl. Acad. Sci. U.S.A. **108**, 4822–4827 (2011).

[24] F. Noé, I. Horenko, C. Schütte, and J. C. Smith, "Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states," J. Chem. Phys. **126**, 155102 (2007).

[25] F. Noé and F. Nüske, "A variational approach to modeling slow processes in stochastic dynamical systems," SIAM Multiscale Model. Simul. **11**, 635–655 (2013).

[26] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, "Constructing the full ensemble of folding pathways from short off-equilibrium simulations," Proc. Natl. Acad. Sci. U.S.A. **106**, 19011–19016 (2009).

[27] G. Perez-Hernandez, F. Paul, T. Giorgino, G. de Fabritiis, and Frank Noé, "Identification of slow molecular order parameters for Markov model construction," J. Chem. Phys. **139**, 015102 (2013).

[28] J.-H. Prinz, J. D. Chodera, and F. Noé, "Spectral rate theory for projected two-state kinetics," Phys. Rev. X (in press); e-print arXiv:1207.0225.

[29] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, "Markov models of molecular kinetics: Generation and validation," J. Chem. Phys. **134**, 174105 (2011).

[30] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE **77**, 257–286 (1989).

[31] S. Röblitz, "Statistical error estimation and grid-free hierarchical refinement in conformation dynamics," Ph.D. thesis (FU Berlin, 2009).

[32] Y. Santoso, C. M. Joyce, O. Potapova, L. Le Reste, J. Hohlbein, J. P. Torella, N. D. F. Grindley, and A. N. Kapanidis, "Conformational transitions in DNA polymerase I revealed by single-molecule FRET," Proc. Natl. Acad. Sci. U.S.A. **107**(2), 715–720 (2010).

[33] M. Sarich, F. Noé, and C. Schütte, "On the approximation error of Markov state models," SIAM Multiscale Model. Simul. **8**, 1154–1177 (2010).

[34] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, "A direct approach to conformational dynamics based on hybrid Monte Carlo," J. Comput. Phys. **151**, 146–168 (1999).

[35] C. R. Schwantes and V. S. Pande, "Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9," J. Chem. Theory Comput. **9**, 2000–2009 (2013).

[36]M. Senne, B. Trendelkamp-Schroer, A. S. J. S. Mey, C. Schütte, and F. Noé, "EMMA - A software package for Markov model building and analysis," J. Chem. Theory Comput. **8**, 2223–2238 (2012).
[37]D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, "Atomic-level characterization of the structural dynamics of proteins," Science **330**(6002), 341–346 (2010).
[38]N. Singhal and V. S. Pande, "Error analysis and efficient sampling in Markovian state models for molecular dynamics," J. Chem. Phys. **123**, 204909 (2005).
[39]W. C. Swope, J. W. Pitera, F. Suits, M. Pitman, and M. Eleftheriou, "Describing protein folding kinetics by molecular dynamics simulations: 2. Example applications to alanine dipeptide and beta-hairpin peptide," J. Phys. Chem. B **108**, 6582–6594 (2004).
[40]V. A. Voelz, G. R. Bowman, K. A. Beauchamp, and V. S. Pande, "Molecular simulation of *ab initio* protein folding for a millisecond folder NTL9," J. Am. Chem. Soc. **132**(5), 1526–1528 (2010).
[41]W. E. and E. Vanden-Eijnden, "Towards a theory of transition paths," J. Stat. Phys. **123**(3), 503–523 (2006).
[42]M. Weber, "Meshless methods in conformation dynamics," Ph.D. thesis (FU Berlin, 2006).
[43]L. R. Welch, "Hidden Markov models and the Baum-Welch algorithm," IEEE Inf. Theory Soc. Newsl. **53**, 1–13 (2003).
[44]B. G. Wensley, S. Batey, F. A. C. Bone, Z. M. Chan, N. R. Tumelty, A. Steward, L. G. Kwa, A. Borgia, A. Garrido, and J. Clarke, "Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family," Nature (London) **463**(7281), 685–688 (2010).