Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty

John D. Chodera,[1,*] Phillip Elms,[1,2,3] Frank Noé,[4] Bettina Keller,[4] Christian M. Kaiser,[1,5] Aaron Ewall-Wice,[6] Susan Marqusee,[1,7,3] Carlos Bustamante,[7,3,5,8,9] and Nina Singhal Hinrichs[10]

[1]*California Institute of Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA*
[2]*Biophysics Graduate Group, University of California, Berkeley, CA 94720, USA*
[3]*Jason L. Choy Laboratory of Single Molecule Biophysics, Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720, USA*
[4]*DFG Research Center Matheon, FU Berlin, Arnimallee 6, 14195 Berlin, Germany*
[5]*Department of Physics, University of California, Berkeley, CA 94720, USA*
[6]*University of Chicago, IL 60637, USA*
[7]*Department of Molecular & Cell Biology, University of California, Berkeley, CA 94720, USA*
[8]*Department of Chemistry, University of California, Berkeley, CA 94720, USA*
[9]*Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA*
[10]*Departments of Statistics and Computer Science, University of Chicago, IL 60637, USA*

(Dated: July 30, 2011)

Single-molecule force spectroscopy has proven to be a powerful tool for studying the kinetic behavior of biomolecules. Through application of an external force, conformational states with small or transient populations can be stabilized, allowing them to be characterized and the statistics of individual trajectories studied to provide insight into biomolecular folding and function. Because the observed quantity (force or extension) is not necessarily an ideal reaction coordinate, individual observations cannot be uniquely associated with kinetically distinct conformations. While maximum-likelihood schemes such as hidden Markov models have solved this problem for other classes of single-molecule experiments by using temporal information to aid in the inference of a sequence of distinct conformational states, these methods do not give a clear picture of how precisely the model parameters are determined by the data due to instrument noise and finite-sample statistics, both significant problems in force spectroscopy. We solve this problem through a *Bayesian* extension that allows the experimental uncertainties to be directly quantified, and build in detailed balance to further reduce uncertainty through physical constraints. We illustrate the utility of this approach in characterizing the three-state kinetic behavior of an RNA hairpin in a stationary optical trap.

INTRODUCTION

Recent advances in biophysical measurement have led to an unprecedented ability to monitor the dynamics of single biological macromolecules, such as proteins and nucleic acids [3]. As a new approach to probing the behavior of biological macromolecules, these experiments promise to change the way we study folding, dynamics, catalysis, association, transcription, translation, and motility, providing otherwise-inaccessible information about microscopic kinetics, energetics, mechanism, and the stochastic heterogeneity inherent in these processes. Advances in instrumentation for optical force spectroscopy in particular have produced instruments of extraordinary stability, precision, and temporal resolution [4, 5] that have already demonstrated great utility in the study of biomolecules in the presence of externally perturbative forces [6–8]. Under external force, it becomes possible to stabilize and characterize short-lived conformational states, such as protein folding and unfolding intermediates [9–11].

In a typical single-molecule optical trapping experiment, a protein or nucleic acid is tethered to two polystyrene beads by dsDNA handles that prevent the molecule under study from interacting with the beads (see Figure 1). The handle-biomolecule-handle assembly—referred to as a *fiber*—is associated with the beads through tight noncovalent interactions, with one bead held in an optical trap and the other either suctioned to a micropipette (as in Figure 1) or held in a second optical trap. During an experiment, the position of the bead within the laser trap is monitored, and either the relative displacement from the trap center or the total force on the bead is recorded, providing a series of displacement or force measurements equally spaced in time, resulting in a timeseries such as the one depicted in Figure 2. The instrument can generally be operated in several modes: a *force ramp* mode, in which the trap is translated rapidly enough to potentially carry the system out of equilibrium; an equilibrium *passive* mode, in which the trap is held fixed; and a constant *force-feedback* mode, in which the trap is continually repositioned to maintain a set constant force on the fiber. Here, we concern ourselves with the latter two classes of experiment, though nonequilibrium experiments remain an exciting topic of current research [12].

Often, the dynamics observed in these experiments appears to be dominated by stochastic interconversions between two or more strongly metastable conformational states [13, 14]—regions of conformation space in which the system remains for long times before making a transition to another conformational state. These transitions are generally well-described by first-order kinetics [15]. While visual inspection of the dynamics may suggest the clear presence of multiple metastable states, quantitative characterization of these states is often difficult. First, the observed force or extension is unlikely to correspond to a true reaction coordinate easily able to separate all metastable states [16–19], and second, measurement noise may further broaden the force or extension signatures of individual states, increasing their overlap. Attempting to separate these states by simply dividing the observed force or extension range into regions, following current practice [20, 21], can of-

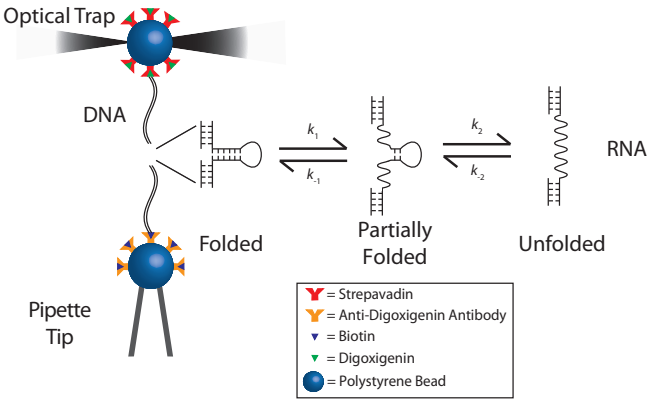

FIG. 1. **Single-molecule optical trapping configuration for p5ab RNA hairpin.** The biomolecule of interest—here, the p5ab RNA hairpin—is tethered to two polystyrene beads by dsDNA handles. The fluctuating force on one bead held in an optical trap is monitored, while the other bead is held suctioned to a micropipette tip. Conformational transitions of the hairpin—such as transitions among the three kinetically metastable states illustrated here—are observed indirectly through motion of the bead in the trap.

ten lead to a high degree of state mis-assignment that results in the estimated rate constants and state distributions containing a significant amount of error [22] (see *Supporting Information*).

Hidden Markov models (HMMs) [23], which use temporal information in addition to the instantaneous value of the observable (force or extension) to determine which conformational states the system has visited during the experiment, have provided an effective solution to the hidden state problem in many other classes of single-molecule experiments, such as ion channel currents [24–27], single-molecule FRET [28–32], and the stepping of motor proteins [33–35]. In applying hidden Markov modeling to the analysis of single-molecule force spectroscopy data, the observed force or extension trace is assumed to come from a realization of an underlying Markov chain, where the system makes history-independent transitions among a set of discrete conformational states with probabilities governed by a transition or rate matrix. Data, in the form of force or bead-to-bead extension measurements, is sampled at an interval that ensures that sequential observations satisfy the Markov property of history-independence, though the appropriate interval depends on the properties of the experimental configuration. Under a given set of external force conditions, each state has a distribution of forces or extensions associated with it. Given observed timeseries data for forces or extensions, the maximum likelihood estimate (MLE) of the model parameters (transition rates and state force or extension distributions) and sequence of hidden states corresponding to the observed data can be determined by standard methods [36, 37], as demonstrated in recent work [38].

Unfortunately, this approach has a number of significant drawbacks. Due to technical limitations, experiments often suffer from limited statistics—the events of interest (transitions between states or visits to rare states) may occur only a few times during the course of the measurement, and data for additional fibers is time-consuming to collect. As a result, while the MLE yields the most *likely* set of model parameters, there may be enormous uncertainty in some of these parameters, and the uncertainty in multiple parameters may be correlated in complex nonlinear ways. While methods exist for estimating the standard error or confidence intervals from MLHMMs [39], these schemes can be prohibitively costly for long traces, and may still significantly underestimate the statistical error for short traces due to the normally-distributed error approximation inherent in the approach. The high cost (both in terms of instrument and experimenter time) of collecting additional data also means that it is not a simple task to judge *how much* data need be collected to test a particular hypothesis in a statistically meaningful way. Worse yet, the standard algorithms employed to find the MLE may not even find the true maximum likelihood solution, instead converging to a local maximum in likelihood that is far from optimal [40].

Here, we resolve this issue through the use of a *Bayesian* extension of hidden Markov models [41–44] applicable to single molecule force experiments. By sampling over the posterior distribution of model parameters and hidden state assignments instead of simply finding the most likely values, the experimenter is able to accurately characterize the correlated uncertainties in both the model parameters (transition rates and state force or extension distributions) and hidden state sequences corresponding to observed data. Additionally, prior information (either from additional independent measurements or physical constraints) can be easily incorporated. We also include a reversibility constraint on the transition matrix—in which microscopic detailed balance is imposed on the kinetics—which has been shown to significantly reduce statistical uncertainties in data-poor conditions [45, 46]. The framework we present is based on Gibbs sampling [47, 48], allowing simple swap-in replacement of models for observable distributions, extension to multiple observables, and alternative models for state transitions. Additionally, the Bayesian method provides a straightforward way to model the statistical outcome and assess the utility of additional experiments given some preliminary data, allowing the experimenter a powerful tool for assessing whether the cost of collecting additional data is outweighed by their benefits. A Matlab implementation of this approach is available online [http://simtk.org/home/bhmm].

HIDDEN MARKOV MODELS FOR FORCE SPECTROSCOPY

Suppose the temporal behavior of some observable $O(x)$ that is a function of molecular configuration x—here, generally force or molecular extension—is observed at temporal intervals Δt to produce a timeseries o_t, where $t = 0, 1, \ldots, L$. An instantaneous observation o_t does not necessarily contain enough information to unambiguously identify the current conformational state the molecule occupies; to infer the hidden state, we must also make use of the temporal information in the observed trace. We restrict ourselves to consideration of scalar functions $O(x)$, but the generalization to multidimensional probes (or multiple probes, such as combined force and fluorescence measurements [49]) and multiple observed temporal traces is straightforward.

We presume the system under study has M kinetically distinct states, in the sense that the system generally remains in a given state for several observation intervals Δt, but these

states may not necessarily represent highly populated states of the system at equilibrium. We treat these conformational states as the *hidden states* of the model, because we cannot *directly* observe the identity of the metastable state in which the system resides. The hidden Markov model presumes the observed data $\mathbf{O} \equiv \{o_t\}$ was generated according to the following model dependent on parameters $\mathbf{\Theta} \equiv \{\mathbf{T}, \mathbf{E}\}$, where $\mathbf{T}$ is an $M \times M$ row-stochastic transition matrix and $\mathbf{E}$ a set of emission parameters governing the observable (force or extension) distributions for each of the M hidden states, and prior information about the initial state distribution $\boldsymbol{\rho}$,

$$\begin{aligned} \mathbb{P}(s_0) &= \rho_{s_0} \\ \mathbb{P}(s_t \mid s_{t-1}, \mathbf{T}) &= T_{s_{t-1} s_t}, \;\; t \geq 1 \\ \mathbb{P}(o_t \mid s_t, \mathbf{e}_{s_t}) &= \varphi(o_t \mid \mathbf{e}_{s_t}). \end{aligned} \tag{1}$$

In diagrammatic form, the observed state data $\{o_t\}$ and corresponding hidden state history $\{s_t\}$ can be represented

$$\begin{array}{ccccccccc} \xrightarrow{\boldsymbol{\rho}} & s_0 & \xrightarrow{\mathbf{T}} & s_1 & \xrightarrow{\mathbf{T}} & s_2 & \xrightarrow{\mathbf{T}} & \cdots & \xrightarrow{\mathbf{T}} & s_L \\ & \downarrow \varphi & & \downarrow \varphi & & \downarrow \varphi & & & & \downarrow \varphi \\ & o_0 & & o_1 & & o_2 & & & & o_L \end{array} \tag{2}$$

The initial state distribution $\boldsymbol{\rho}$ reflects our knowledge of the initial conditions of the experiment that collected data $\mathbf{o}$. In the case that the experiment was prepared in equilibrium, $\boldsymbol{\rho}$ corresponds to the equilibrium distribution $\boldsymbol{\pi}$ of the model transition matrix $\mathbf{T}$; if the experiment was prepared out of equilibrium, $\boldsymbol{\rho}$ may be chosen to reflect some other prior distribution (e.g. the uniform prior).

State transitions $(s_{t-1} \to s_t)$ are governed by the discrete transition probability $T_{s_{t-1} s_t}$. The *Markov property* of HMMs prescribes that the probability that a system originally in state i at time t is later found in state j at time $t+1$ is dependent only on knowledge of the state i, and given by the corresponding matrix element T_{ij} of the (row-stochastic) transition matrix $\mathbf{T}$. Alternatively, one could instead use the rate matrix $\mathbf{K}$, related to the transition matrix $\mathbf{T}$ through the equation $\mathbf{T} = e^{\mathbf{K}\Delta t}$. If the processes described by $\mathbf{T}$ or $\mathbf{K}$ are slow compared to the observation interval Δt, then we can easily estimate the rate matrix from the associated transition matrix in a way that avoids the matrix logarithm, through the expansion $\mathbf{K} \approx (\mathbf{T} - \mathbf{I})/\Delta t$, where $\mathbf{I}$ denotes the $M \times M$ identity matrix.

The probabilistic "emission" of observables from each state $(s_t \to o_t)$ is governed by the continuous emission probability $\varphi(o_t \mid \mathbf{e}_{s_t})$, parametrized by observable *emission* parameters $\mathbf{e}$. For example, in the force spectroscopy applications described here, $\varphi(o \mid \mathbf{e}_s)$ is taken to be a univariate normal (Gaussian) distribution parameterized by a mean μ and variance σ^2 that characterize each state, such that $\mathbf{e}_i \equiv \{\mu_i, \sigma_i^2\}$. Other choices of observable distribution can easily be substituted in a modular way without affecting the structure of the algorithms presented here.

Given the HMM process specified in Eq. 1, the probability of observing data $\mathbf{O}$ given the model parameters $\mathbf{\Theta}$ is then,

$$P(\mathbf{O} \mid \mathbf{\Theta}) = \sum_{\mathbf{S}} \rho_{s_0} \varphi(o_0 \mid \mathbf{e}_{s_0}) \prod_{t=1}^{L} T_{s_{t-1} s_t} \varphi(o_t \mid \mathbf{e}_{s_t}), \tag{3}$$

where the sum over hidden state histories $\mathbf{S}$ is shorthand for

$$\sum_{\mathbf{S}} \equiv \sum_{s_0=1}^{M} \sum_{s_1=1}^{M} \cdots \sum_{s_L=1}^{M}. \tag{4}$$

If multiple independent traces $\{o_t\}$ are available, the probability $P(\mathbf{O} \mid \mathbf{\Theta})$ is simply the product of Eq. 3 for the independent traces.

Maximum likelihood hidden Markov model (MLHMM)

The standard approach to construct an HMM from observed data is to compute the *maximum likelihood estimator* (MLE) for the model parameters $\mathbf{\Theta} \equiv \{\mathbf{T}, \mathbf{E}\}$, which maximize the probability of the observed data $\mathbf{O}$ given the model,

$$\hat{\mathbf{\Theta}} = \arg\max_{\mathbf{\Theta}} P(\mathbf{O} \mid \mathbf{\Theta}), \tag{5}$$

yielding MLE estimates of transition matrix $\hat{\mathbf{T}}$ and state emission parameters $\hat{\mathbf{E}}$. Typically, determination of the model parameters $\mathbf{\Theta}$ is carried out using the Baum-Welch algorithm [36].

Once the MLE parameters $\hat{\mathbf{\Theta}}$ are determined, the most likely hidden state history that produced the observations $\mathbf{O}$ can be determined using these parameters:

$$\hat{\mathbf{S}} = \arg\max_{\mathbf{S}} P(\mathbf{S} \mid \mathbf{O}, \hat{\mathbf{\Theta}}). \tag{6}$$

This is typically carried out using the Viterbi algorithm [37], a classic example of dynamic programming.

Both the Baum-Welch and Viterbi schemes are described in detail in *Algorithms*.

Bayesian hidden Markov model (BHMM)

Instead of simply determining the model that maximizes the likelihood of observing the data $\mathbf{O}$ given the model parameters $\mathbf{\Theta}$, we can make use of Bayes' theorem to compute the *posterior* distribution of model parameters given the observed data:

$$P(\mathbf{\Theta} \mid \mathbf{O}) \propto P(\mathbf{O} \mid \mathbf{\Theta}) P(\mathbf{\Theta}). \tag{7}$$

Here, $P(\mathbf{\Theta})$ denotes a *prior* distribution that encodes any *a priori* information we may have about the model parameters $\mathbf{\Theta}$. This prior information might include, for example, physical constraints (such as ensuring the transition matrix satisfies detailed balance) or prior rounds of inference from other independent experiments.

Making use of the likelihood (Eq. 3), the model posterior is then given by,

$$P(\mathbf{\Theta} \mid \mathbf{O}) \propto P(\mathbf{\Theta}) \sum_{\mathbf{S}} \rho_{s_0} \varphi(o_0 \mid \mathbf{e}_{s_0}) \prod_{t=1}^{L} T_{s_{t-1} s_t} \varphi(o_t \mid \mathbf{e}_{s_t}) \tag{8}$$

Drawing samples of $\mathbf{\Theta}$ from this distribution will, in principle, allow the *confidence* with which individual parameters and combinations thereof are known, given the data (and subject to the validity of the model of Eq. 1 in correctly representing the process by which the observed data is generated). While the posterior $P(\mathbf{\Theta}|\mathbf{O})$ is complex, we could in principle use a Markov chain Monte Carlo (MCMC) approach [48] to sample it. In its current form, however, this would be extremely expensive due to the sum over all hidden state histories $\mathbf{S}$ appearing in ratios involving Eq. 8. Instead, we introduce

the hidden state histories $\mathbf{S}$ as an auxiliary variable, sampling from the augmented posterior,

$$P(\mathbf{\Theta}, \mathbf{S} \mid \mathbf{O}) \propto \left[\rho_{s_0} \varphi(o_0 \mid \mathbf{e}_{s_0}) \prod_{t=1}^{L} T_{s_{t-1} s_t} \varphi(o_t \mid \mathbf{e}_{s_t}) \right] P(\mathbf{\Theta}). \tag{9}$$

which makes it much less costly to compute the ratios required for MCMC on the augmented $(\mathbf{\Theta}, \mathbf{S})$ parameter space.

If we presume the prior is separable, such that $P(\mathbf{\Theta}) \equiv P(\mathbf{T})P(\mathbf{E})$, we can sample from the augmented posterior (Eq. 9) using the framework of *Gibbs sampling* [48], in which the augmented model parameters are updated by sampling from the conditional distributions,

$$\begin{aligned} P(\mathbf{S} \mid \mathbf{T}, \mathbf{E}, \mathbf{O}) &\propto \rho_{s_0} \varphi(o_0 \mid \mathbf{e}_{s_0}) \prod_{t=1}^{L} T_{s_{t-1} s_t} \varphi(o_t \mid \mathbf{e}_{s_t}) \\ P(\mathbf{T} \mid \mathbf{E}, \mathbf{S}, \mathbf{O}) = P(\mathbf{T} \mid \mathbf{S}) &\propto P(\mathbf{T}) \prod_{t=1}^{L} T_{s_{t-1} s_t} \\ P(\mathbf{E} \mid \mathbf{S}, \mathbf{T}, \mathbf{O}) = P(\mathbf{E} \mid \mathbf{S}, \mathbf{O}) &\propto P(\mathbf{E}) \prod_{t=0}^{L} \varphi(o_t \mid \mathbf{e}_{s_t}). \end{aligned} \tag{10}$$

The equalities on the second and third lines reflect the conditional independence of the hidden Markov model defined by Eq. 1. When only the model parameters $\mathbf{\Theta} \equiv \{\mathbf{T}, \mathbf{E}\}$ or the hidden state histories $\mathbf{S}$ are of interest, we can simply marginalize out the uninteresting variables by sampling from the augmented joint posterior for $\{\mathbf{T}, \mathbf{E}, \mathbf{S}\}$ and examine only the variables of interest. In addition, the structure of the Gibbs sampling scheme above allows individual components (such as the observable distribution model $\varphi(o \mid \mathbf{e})$ or transition probability matrix $\mathbf{T}$) to be modified without affecting the structure of the remainder of the calculation.

In the illustrations presented here, we employ a Gaussian observable distribution model for $\varphi(o \mid \mathbf{e})$,

$$\varphi(o \mid \mathbf{e}) = \varphi(o \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(o-\mu)^2}{\sigma^2} \right], \tag{11}$$

where μ is the mean force or extension characterizing a particular state, and σ is the standard deviation or width of forces or extensions corresponding to that state. We note that marginal posterior distributions of each mean $P(\mu_i|\mathbf{O})$ reflect the statistical uncertainty in how well the mean force or position is determined, and need not correspond to the standard deviation σ_i, which may be much broader (or narrower, depending on the situation).

ALGORITHMS

Generating an initial model

To initialize either computation of the MLHMM or sampling from the posterior for the BHMM, an initial model that respects any constraints imposed in the model prior $P(\mathbf{\Theta})$ must be selected. Here, we employ a Gaussian observable distribution model for $\varphi(o \mid \mathbf{e})$,

$$\varphi(o \mid \mathbf{e}) = \varphi(o \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(o-\mu)^2}{\sigma^2} \right], \tag{12}$$

and enforce that the transition matrix $\mathbf{T}$ satisfy detailed balance[1].

Observable parameter estimation

We first initialize the observed distributions of each state by fitting a Gaussian mixture model with M states to the pooled observed data $\mathbf{O}$, ignoring temporal information:

$$P(\mathbf{O} \mid \boldsymbol{\pi}, \mathbf{E}) = \prod_{t=0}^{L} \sum_{m=1}^{M} \pi_m \varphi(o_t \mid \mu_m, \sigma_m^2), \tag{13}$$

where the state observable emission probability vector $\mathbf{E} \equiv \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ and $\mathbf{e}_m \equiv \{\mu_m, \sigma_m^2\}$ with μ_m denoting the observable mean and σ_m^2 the variance for state m for the Gaussian mixture model. The vector $\boldsymbol{\pi}$ is composed of equilibrium state populations $\{\pi_1, \ldots, \pi_M\}$ with $\pi_m \geq 0$ and $\sum_{m=1}^{M} \pi_m = 1$.

A first approximation to $\boldsymbol{\pi}$ and $\mathbf{E}$ is computed by pooling and sorting the observed o_t, and defining M indicator functions $h_m(o)$ that separate the data into M contiguous regions of the observed range of o of roughly equal population. Let $N_m \equiv \sum_{t=0}^{L} h_m(o_t)$ denote the total number of observations falling in region m, and $N_{\mathrm{tot}} = \sum_{m=1}^{M} N_m$. The initial parameters are then computed as,

$$\begin{aligned} \pi_m &= N_m / N_{\mathrm{tot}} \\ \mu_m &= N_m^{-1} \sum_{t=0}^{L} o_t \, h_m(o_t) \qquad (14) \\ \sigma_m^2 &= N_m^{-1} \sum_{t=0}^{L} (o_t - \mu_m)^2 \, h_m(o_t). \qquad (15) \end{aligned}$$

This approximation is then improved upon by utilizing the expectation-maximization procedure described by Bilmes [51],

$$\begin{aligned} \pi'_m &= N_{\mathrm{tot}}^{-1} \sum_{t=0}^{L} \chi_m(o_t, \mathbf{E}, \boldsymbol{\pi}) \\ \mu'_m &= (\pi'_m N_{\mathrm{tot}})^{-1} \sum_{t=0}^{L} o_t \, \chi_m(o_t, \mathbf{E}, \boldsymbol{\pi}) \\ \sigma'^2_m &= (\pi'_m N_{\mathrm{tot}})^{-1} \sum_{t=0}^{L} (o_t - \mu'_m)^2 \, \chi_m(o_t, \mathbf{E}, \boldsymbol{\pi}) \end{aligned} \tag{16}$$

where the function $\chi_m(o, \mathbf{E}, \boldsymbol{\pi})$ is given by the fuzzy membership function,

$$\chi_m(o, \mathbf{E}, \boldsymbol{\pi}) = \frac{\pi_m \, \varphi(o \mid \mathbf{e}_m)}{\sum_{l=1}^{M} \pi_l \, \varphi(o \mid \mathbf{e}_l)}. \tag{17}$$

This iterative procedure is terminated at iteration j when the change in the parameters $\{\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2\}$ falls below a certain relative threshold, such as $\|\boldsymbol{\pi}^{[j]} - \boldsymbol{\pi}^{[j-1]}\|_2 / \|\boldsymbol{\pi}^{[j]}\|_2 < 10^{-4}$.

[1] Physical systems that are not driven in time by an external force or fed by an energy reservoir should satisfy detailed balance [50], and its use has been shown to provide a large reduction in transition matrix uncertainty in data-poor conditions [45, 46]. Detailed balance specifies that $\pi_i T_{ij} = \pi_j T_{ji}$ for all i, j, where $\boldsymbol{\pi}$ is the equilibrium distribution of the row-stochastic transition matrix $\mathbf{T}$.

Transition matrix estimation

Once initial state observable emission parameters $\mathbf{E}$ are determined, an initial transition matrix is estimated using an iterative likelihood maximization approach that enforces detailed balance [52]. First, a matrix of fractional transition counts $\mathbf{C} \equiv (c_{ij})$ is estimated using the membership function:

$$c_{ij} = \sum_{t=1}^{L} \chi_i(o_{t-1}, \mathbf{E}, \boldsymbol{\pi})\, \chi_j(o_t, \mathbf{E}, \boldsymbol{\pi}) \tag{18}$$

A symmetric $M \times M$ matrix $\mathbf{X} \equiv (x_{ij})$ is initialized by

$$x_{ij} = x_{ji} = c_{ij} + c_{ji} \tag{19}$$

and a vector of row sums

$$x_{i*} = \sum_{j=1}^{M} x_{ij}. \tag{20}$$

Then, the iterative procedure described in Algorithm 1 of [52] is applied. For each update iteration, we first update the diagonal elements of $\mathbf{X}$:

$$x'_{ii} = \frac{c_{ii}(x_{i*} - x_{ii})}{c_{i*} - c_{ii}} \tag{21}$$

where

$$c_{i*} = \sum_{j=1}^{M} c_{ij} \tag{22}$$

followed by the off-diagonal elements:

$$x'_{ij} = x'_{ji} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{23}$$

where the quantities a, b, and c are computed from $\mathbf{X}$ and $\mathbf{C}$ as

$$\begin{aligned} a &\equiv c_{i*} - c_{ij} + c_{j*} - c_{ji} \\ b &\equiv c_{i*}(x_{j*} - x_{ji}) + c_{j*}(x_{i*} - x_{ij}) \\ &\quad - (c_{ij} + c_{ji})(x_{i*} - x_{ij} + x_{j*} - x_{ji}) \\ c &\equiv -(c_{ij} + c_{ji})(x_{i*} - x_{ij})(x_{j*} - x_{ji}) \end{aligned} \tag{24}$$

Once a sufficient number of iterations j have been completed to compute a stable estimate of $\mathbf{X}$ (such as the relative convergence criteria $\|\mathbf{X}^{[j]} - \mathbf{X}^{[j-1]}\|_2 / \|\mathbf{X}^{[j]}\|_2 < 10^{-4}$, the maximum likelihood transition matrix estimate $\mathbf{T}$ is computed as

$$T_{ij} = \frac{x_{ij}}{x_{i*}}. \tag{25}$$

Note that the equilibrium probability vector $\boldsymbol{\pi}$ computed during the Gaussian mixture model fitting is not respected during this step.

Fitting a maximum likelihood HMM

The HMM model parameters $\boldsymbol{\Theta} \equiv \{\mathbf{T}, \mathbf{E}\}$ are fit to the observed data $\mathbf{O}$ through use of the expectation-maximization (EM) algorithm [53]. This is an iterative procedure, where the model parameters are subsequently refined through successive iterations. The initial HMM is usually quick to compute, and can give the experimenter a rough idea of the model parameters, as well as providing a useful starting point for sampling models from the Bayesian posterior.

During each iteration, the Baum-Welch algorithm [36] is used to compute $\boldsymbol{\Xi} \equiv (\xi_{tij})$, which represents the probability that the system transitions from hidden state i at time $t-1$ to hidden state j at time t, and γ_{ti}, the probability that the system occupied state i at time t. This is accomplished by first executing the *forward algorithm*,

$$\alpha_{tj} = \begin{cases} \rho_j\, \varphi(o_0 \mid \mathbf{e}_j) & t = 0 \\ \varphi(o_t \mid \mathbf{e}_j) \sum_{i=1}^{M} \alpha_{(t-1)i} T_{ij} & t = 1, \ldots, L \end{cases} \tag{26}$$

followed by the *backward algorithm*,

$$\beta_{ti} = \begin{cases} 1 & t = L \\ \sum_{j=1}^{M} T_{ij} \varphi(o_{t+1} \mid \mathbf{e}_j) \beta_{(t+1)j} & t = (L-1), \ldots, 0 \end{cases} \tag{27}$$

The $L \times M \times M$ matrix $\boldsymbol{\Xi}$ is then computed for $t = 0, \ldots, (L-1)$ as,

$$\xi_{tij} = \alpha_{ti} \varphi(o_{t+1} \mid \mathbf{e}_i) T_{ij} \beta_{(t+1)j} / \sum_{i=1}^{M} \alpha_{Ti} \tag{28}$$

$$\gamma_{ti} = \sum_{j=1}^{M} \xi_{tij} \tag{29}$$

In practice, the logarithms of these quantities are computed instead to avoid numerical underflow.

The aggregate matrix of expected transition counts $\mathbf{C} \equiv (c_{ij})$ is then computed from $\boldsymbol{\Xi}$ as,

$$c_{ij} = \sum_{t=0}^{L-1} \xi_{tij}. \tag{30}$$

This count matrix is used to update the maximum-likelihood transition matrix $\mathbf{T}$ using the method of Prinz et al. [52] described in the previous section.

The state observable distribution parameters $\mathbf{E}$ are then updated from the γ_{ti}. For the univariate normal distribution applied to force spectroscopy data here, we update the mean μ_i and variance σ_i^2 for state i using the scheme,

$$\mu'_i = \frac{\sum_{t=0}^{L} o_t \gamma_{ti}}{\sum_{t=0}^{L} \gamma_{ti}} \;\; ; \;\; \sigma'^2_i = \frac{\sum_{t=0}^{L} (o_t - \mu'_i)^2 \gamma_{ti}}{\sum_{t=0}^{L} \gamma_{ti}}. \tag{31}$$

Once the model parameters have been fitted by iteration of the above update procedure to convergence (which may only converge to a local maximum of the likelihood), the most likely hidden state sequence can be determined given the observations $\mathbf{O}$ and the MLE model $\hat{\boldsymbol{\Theta}}$ using the Viterbi algorithm [37]. Like the forward-backward algorithm employed in the Baum-Welch procedure, the Viterbi algorithm also has a forward recursion component,

$$\epsilon_{jt} = \begin{cases} \rho_j \varphi(o_t \mid \mathbf{e}_j) & t = 0 \\ \varphi(o_t \mid \mathbf{e}_j) \max_i \epsilon_{i(t-1)} T_{ij} & t = 1, \ldots, L \end{cases} \tag{32}$$

$$\Phi_{jt} = \begin{cases} 1 & t = 0 \\ \arg\max_i \epsilon_{i(t-1)} T_{ij} & t = 1, \ldots, L \end{cases}$$

as well as a reverse reconstruction component to compute the most likely state sequence $\hat{\mathbf{S}}$,

$$\hat{s}_t = \begin{cases} \arg\max_i \epsilon_{it} & t = L \\ \Phi_{\hat{s}_{t+1}(t+1)} & t = (L-1), \ldots, 0 \end{cases} \tag{33}$$

Sampling from the posterior of the BHMM

Sampling from the posterior of the BHMM (Eq. 8) proceeds by rounds of Gibbs sampling, where each round consists of an update of the augmented model parameters $\{\mathbf{T}, \mathbf{E}, \mathbf{S}\}$ by sampling

$$\begin{aligned} \mathbf{S}' \mid \mathbf{T}, \mathbf{E}, \mathbf{O} &\sim P(\mathbf{S}' \mid \mathbf{T}, \mathbf{E}, \mathbf{O}) \\ \mathbf{T}' \mid \mathbf{S}' &\sim P(\mathbf{T}' \mid \mathbf{S}') \\ \mathbf{E}' \mid \mathbf{S}', \mathbf{O} &\sim P(\mathbf{E}' \mid \mathbf{S}', \mathbf{O}) \end{aligned}$$

where the conditional probabilities are given by Eq. 10.

Updating the hidden state sequences

We use a modified form of the Viterbi process to generate an independent sample of the hidden state history $\mathbf{S}$ given the transition probabilities $\mathbf{T}$, state observable distribution parameters $\mathbf{E}$, and observed data $\mathbf{O}$. Like the Viterbi scheme, a forward recursion is applied to each observation trace $\mathbf{o}$, but instead of computing the most *likely* state history on the reverse pass, a new hidden state history $\mathbf{S}$ is drawn from the distribution $P(\mathbf{S} \mid \mathbf{O}, \mathbf{T}, \mathbf{E})$. The forward recursion uses the same forward algorithm as used in Baum-Welch [36],

$$\alpha_{tj} = \begin{cases} \rho_j \, \varphi(o_0 \mid \mathbf{e}_j) & t = 0 \\ \varphi(o_t \mid \mathbf{e}_j) \sum_{i=1}^{M} \alpha_{(t-1)i} T_{ij} & t = 1, \ldots, L \end{cases} \tag{34}$$

In the reverse recursion, we now *sample* a state sequence by sampling each hidden state from the conditional distribution $s_t \sim P(s_t \mid s_{t+1}, \ldots, s_L)$ starting from $t = L$ and proceeding down to $t = 0$, where the conditional distribution is given by,

$$\begin{aligned} &P(s_t = i \mid s_{t+1}, \ldots, s_L) \\ &\propto \begin{cases} \alpha_{ti} / \sum_{j=1}^{M} \alpha_{tj} & t = L \\ \alpha_{ti} T_{i s_{t+1}} / \sum_{j=1}^{M} \alpha_{tj} T_{j s_{t+1}} & t = (L-1), \ldots, 0 \end{cases} \end{aligned} \tag{35}$$

It is straightforward to show the result of these sampling steps reconstitutes the probability distribution $P(\mathbf{S}|\mathbf{T}, \mathbf{E}, \mathbf{O})$ (see *Supplementary Material*).

Updating the transition probabilities

If no detailed balance constraint is used and the prior $P(\mathbf{T})$ is Dirichlet in each row of the transition matrix $\mathbf{T}$, it is possible to generate an independent sample of the transition matrix from the conditional distribution $P(\mathbf{T}' \mid \mathbf{S}')$ by sampling each row of the transition matrix from the conjugate Dirichlet posterior using the transition counts from the sampled state sequence $\mathbf{S}'$ [45]. However, because physical systems in the absence of energy input through an external driving force should satisfy detailed balance, we make use of this constraint in updating our transition probabilities, since this has been demonstrated to substantially reduce parameter uncertainty in the data-limited regime [45].

The transition matrix is updated using the reversible transition matrix sampling scheme of Noé [45, 54]. Here, an adjusted count matrix $\mathbf{C} \equiv (c_{ij})$ is computed using the updated hidden state sequence $\mathbf{S}'$,

$$c_{ij} = b_{ij} + \sum_{t=1}^{L} \delta_{i s_{t-1}} \delta_{j s_t}, \tag{36}$$

where the Kronecker $\delta_{ij} = 1$ if $i = j$ and zero otherwise, and $\mathbf{B} \equiv (b_{ij})$ is a matrix of prior pseudocounts, which we take to be zero following the work of Noé et al. [15]. Using the adjusted count matrix $\mathbf{C}$, a Metropolis-Hastings Monte Carlo procedure [55] is used to update the matrix and produce a new sample from $P(\mathbf{T}' \mid \mathbf{S}')$. Two move types are attempted, selected with equal probability, and 1000 moves are attempted to generate a new sample $\mathbf{T}'$ that is approximately uncorrelated from the previous $\mathbf{T}$. Prior to starting the Monte Carlo procedure, the vector of equilibrium probabilities for all states $\boldsymbol{\pi}$ is computed according to

$$\mathbf{T}^{\mathrm{T}} \boldsymbol{\pi} = \boldsymbol{\pi}. \tag{37}$$

The first move type is a *reversible element shift*. A pair of states (i, j), $i \neq j$, are selected with uniform probability, and a random number Δ is selected uniformly over the interval,

$$\Delta \in [\max(-T_{ii}, -\frac{\pi_j}{\pi_i} T_{jj}), T_{ij}].$$

The changed elements in the proposed transition matrix $\mathbf{T}'$ are then given by:

$$\begin{aligned} T'_{ij} &= T_{ij} - \Delta \;\; ; \;\; T'_{ji} = T_{ji} - \frac{\pi_i}{\pi_j} \Delta \\ T'_{ii} &= T_{ii} + \Delta \;\; ; \;\; T'_{jj} = T_{jj} + \frac{\pi_i}{\pi_j} \Delta. \end{aligned}$$

This move is accepted with probability

$$\begin{aligned} P_{\mathrm{accept}}(\mathbf{T} \to \mathbf{T}') = \min \Bigg\{ 1, \sqrt{\frac{(T'_{ij})^2 + (T'_{ji})^2}{(T_{ij})^2 + (T_{ji})^2}} \\ \times \left(\frac{T'_{ii}}{T_{ii}}\right)^{c_{ii}} \left(\frac{T'_{ij}}{T_{ij}}\right)^{c_{ij}} \left(\frac{T'_{jj}}{T_{jj}}\right)^{c_{jj}} \left(\frac{T'_{ji}}{T_{ji}}\right)^{c_{ji}} \Bigg\} \end{aligned} \tag{38}$$

This move will leave the vector of stationary probabilities $\boldsymbol{\pi}$ unchanged.

The second move type is a *row shift*. A row i of $\mathbf{T}$ is selected with uniform probability, and a random number η chosen uniformly over the interval

$$\eta \in \left[0, \frac{1}{1 - T_{ii}}\right]$$

and used to update row i of $\mathbf{T}$ according to

$$T'_{ij} = \begin{cases} \eta T_{ij} & j = 1, \ldots, M, \;\; j \neq i \\ \eta(T_{ii} - 1) + 1 & j = i \end{cases} \tag{39}$$

This move is accepted with probability

$$\begin{aligned} &P_{\mathrm{accept}}(\mathbf{T} \to \mathbf{T}') \\ &= \min \left\{ 1, \eta^{(M-2)} \eta^{(c_{i*} - c_{ii})} \left(\frac{1 - \eta(1 - T_{ii})}{T_{ii}}\right)^{c_{ii}} \right\} \end{aligned} \tag{40}$$

The row shift operation will change the stationary distribution of $\boldsymbol{\pi}'$, but it may be efficiently updated with

$$\pi_i' = \frac{\pi_i}{\pi_i + \eta(1-\pi_i)} \;;\; \pi_j' = \frac{\eta\,\pi_j}{\pi_i + \eta(1-\pi_i)}.$$

Since this update scheme is incremental, it will accumulate numerical errors over time that cause the updated $\boldsymbol{\pi}$ to drift away from the stationary distribution of the current transition matrix. To avoid this, $\boldsymbol{\pi}$ is recomputed from the current sample of the transition matrix in regular intervals (here, every 100 sampling steps).

Updating the observable distribution parameters

Following the update of the transition matrix $\mathbf{T}$, the observable distribution parameters $\mathbf{E}$ are updated by sampling $\mathbf{E}$ from the conditional probability $P(\mathbf{E}' \mid \mathbf{S}', \mathbf{O})$. The conditional probability for the observable distribution parameters for state m, denoted $\mathbf{e}_m$, is given in terms of the output model $\varphi(o \mid \mathbf{e})$ by Bayes' theorem,

$$P(\mathbf{E} \mid \mathbf{O}, \mathbf{S}) = \left[\prod_{t=0}^{L} \varphi(o_t \mid \mathbf{e}_{s_t})\right] P(\mathbf{E}). \tag{41}$$

An important choice must be made with regards to the prior, $P(\mathbf{E})$. If the prior is chosen to be composed of independent priors for each state, as in

$$P(\mathbf{E}) = \prod_{m=1}^{M} P(\mathbf{e}_m), \tag{42}$$

then the full BHMM posterior (Eq. 8) will be invariant under any permutation of the states. This behavior might be undesirable, as the states may switch labels during the posterior sampling procedure; this will require any analysis of the models sampled from the posterior to account for the possible permutation symmetry in the states. On the other hand, breaking this symmetry (e.g., by enforcing an ordering on the state mean observables) can artificially restrict the confidence intervals of the states, which might additionally complicate data analysis.

Here, we make the choice that the prior be separable (Eq. 42), which has the benefit of allowing the conditional probability for $\mathbf{E}$ (Eq. 41) to be decomposed into a separate posterior for each state. For each state m, collect all the observations o_t whose updated hidden state labels ${s_t}' = m$ into a single dataset $\mathbf{o} \equiv \{o_n\}_{n=1}^{N_m}$, where N_m is the total number of times state m is visited, for the purposes of this update procedure. Then, the observable parameters $\mathbf{e}$ for this state are given by

$$P(\mathbf{e} \mid \mathbf{o}) = P(\mathbf{o} \mid \mathbf{e})P(\mathbf{e}) = \left[\prod_{n=1}^{N_m} \varphi(o_n \mid \mathbf{e})\right] P(\mathbf{e}). \tag{43}$$

In the application presented here, we use a Gaussian output model (Eq. 12) for the state observable distributions $P(o \mid \mathbf{e})$, where $\mathbf{e} \equiv \{\mu, \sigma^2\}$, with μ the state mean observable and σ^2 the variance (which will include both the distribution of the observable characterizing the state and any broadening from measurement noise). Other models (including multidimensional or multimodal observation models) are possible, and require replacing only the observation model $\varphi(o \mid \mathbf{e})$ and corresponding prior $P(\mathbf{e})$.

We use the (improper) Jeffreys prior [56] which has the information-theoretic interpretation as the prior that maximizes the information content of the data [57], (suppressing the state index subscript m),

$$P(\mathbf{e}) \propto \sigma^{-1}, \tag{44}$$

which produces the posterior

$$P(\mathbf{e} \mid \mathbf{o}) \propto \sigma^{-(N+1)} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(o_n - \mu)^2\right], \tag{45}$$

where we remind the reader that here and in the remainder of this section, the symbols $\mathbf{e}$, $\mathbf{o}$, σ, μ, and N refer to $\mathbf{e}_m$, $\mathbf{o}_m$, σ_m, μ_m, and N_m, respectively.

Updating $\{\mu, \sigma^2\}$ also proceeds by a Gibbs sampling scheme, alternately updating μ and σ, as earlier described in Ref. [54],

$$\begin{aligned} \mu &\sim P(\mu \mid \sigma^2, \mathbf{o}) \\ \sigma^2 &\sim P(\sigma^2 \mid \mu, \mathbf{o}) \end{aligned} \tag{46}$$

The conditional distribution of the mean μ is then given by

$$\begin{aligned} P(\mu \mid \sigma^2, \mathbf{o}) &\equiv \frac{P(\mu, \sigma^2 \mid \mathbf{o})}{\int d\mu\, P(\mu, \sigma^2 \mid \mathbf{o})} \\ &\propto \exp\left[-\frac{1}{2(\sigma^2/N)}(\mu - \hat{\mu})^2\right] \end{aligned} \tag{47}$$

where $\hat{\mu}$ is the sample mean for $\mathbf{o}$, the samples in state m,

$$\hat{\mu} \equiv \frac{1}{N}\sum_{n=1}^{N} o_n \tag{48}$$

This allows us to update μ according to

$$\mu' \sim \mathcal{N}(\hat{\mu}, \sigma^2/N) \tag{49}$$

The conditional distribution of the variance σ^2 is given by

$$\begin{aligned} P(\sigma^2 \mid \mu, \mathbf{o}) &= \frac{p(\mu, \sigma^2 \mid \mathbf{o})}{\int d\sigma^2\, p(\mu, \sigma^2 \mid \mathbf{o})} \\ &\propto \sigma^{-(N+1)} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(o_n - \mu)^2\right] \\ &\propto \sigma^{-(N+1)} \exp\left[-\frac{N\hat{\sigma}^2}{2\sigma^2}\right] \end{aligned} \tag{50}$$

where the quantity $\hat{\sigma}^2$, which is *not* in general identical to the sample variance, is given by

$$\hat{\sigma}^2 \equiv \frac{1}{N}\sum_{n=1}^{N}(o_n - \mu)^2. \tag{51}$$

A convenient way to update $\sigma^2 \mid \mu, \mathbf{o}$ is to sample a random variate y from the chi-square distribution with $N-1$ degrees of freedom,

$$y \sim \chi^2(N-1) \tag{52}$$

and then update σ^2 as

$$\sigma'^2 = \frac{N\hat{\sigma}^2}{y}. \tag{53}$$

Note that μ and σ^2 can be updated in either order, but the *updated* values of μ or σ^2 must be used in sampling the not-yet-updated σ^2 or μ, and vice-versa.

Other output probabilities, such as *mixtures* of normal distributions or other distributions, can be substituted by simply changing $P(\mathbf{E} \mid \mathbf{O}, \mathbf{S})$ and the scheme by which $\mathbf{E}$ is updated.

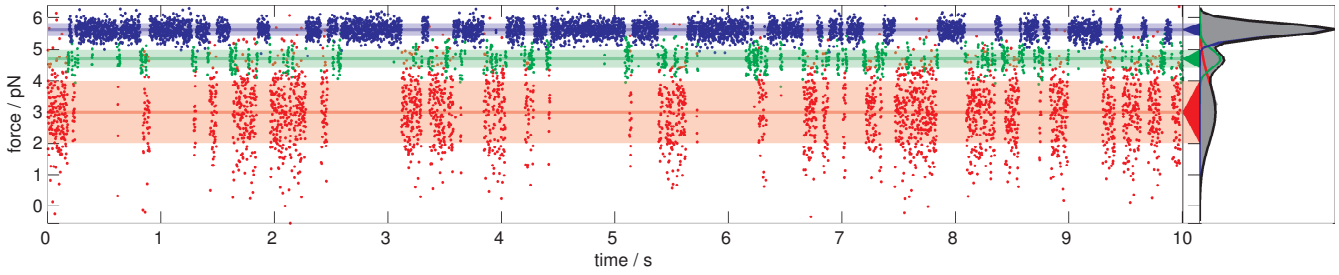

FIG. 2. **Synthetic force trajectory and inferred state assignments in MLHMM.** Observed samples are colored by their hidden state assignments. Dark horizontal lines terminating in triangles to the right denote state means, while lightly colored bands indicate one standard deviation on either side of the state mean. The gray histogram on the right side shows the total observed probability of samples, while the colored peaks show the weighted Gaussian output contribution from each state, and the black outline the weighted sum of the Gaussian output contributions from the HMM states.

TABLE I. **Estimated mean model parameters and confidence intervals for synthetic timeseries data**

			Estimated Model Parameters		
Property		**True Value**	**1 000 observations**	**10 000 observations**	**100 000 observations**
stationary probability	π_1	0.308	$0.228\,^{0.480}_{0.074}$	$0.318\,^{0.407}_{0.244}$	$0.324\,^{0.355}_{0.292}$
	π_2	0.113	$0.093\,^{0.172}_{0.042}$	$0.124\,^{0.155}_{0.098}$	$0.112\,^{0.121}_{0.104}$
	π_3	0.579	$0.679\,^{0.870}_{0.415}$	$0.558\,^{0.648}_{0.455}$	$0.564\,^{0.599}_{0.531}$
transition probability	T_{11}	0.980	$0.970\,^{0.987}_{0.945}$	$0.972\,^{0.978}_{0.966}$	$0.979\,^{0.981}_{0.978}$
	T_{12}	0.019	$0.023\,^{0.045}_{0.009}$	$0.026\,^{0.032}_{0.021}$	$0.020\,^{0.021}_{0.018}$
	T_{13}	0.001	$0.007\,^{0.018}_{0.001}$	$0.002\,^{0.003}_{0.001}$	$0.001\,^{0.001}_{0.001}$
	T_{21}	0.053	$0.054\,^{0.106}_{0.018}$	$0.067\,^{0.082}_{0.053}$	$0.057\,^{0.061}_{0.052}$
	T_{22}	0.900	$0.868\,^{0.931}_{0.790}$	$0.890\,^{0.907}_{0.870}$	$0.897\,^{0.903}_{0.892}$
	T_{23}	0.050	$0.078\,^{0.136}_{0.035}$	$0.043\,^{0.056}_{0.033}$	$0.046\,^{0.050}_{0.042}$
	T_{31}	0.001	$0.002\,^{0.006}_{0.000}$	$0.001\,^{0.002}_{0.000}$	$0.001\,^{0.001}_{0.000}$
	T_{32}	0.009	$0.010\,^{0.019}_{0.004}$	$0.010\,^{0.012}_{0.007}$	$0.009\,^{0.010}_{0.008}$
	T_{33}	0.990	$0.988\,^{0.995}_{0.978}$	$0.990\,^{0.992}_{0.987}$	$0.990\,^{0.991}_{0.989}$
state mean force (pN)	μ_1	3.000	$2.947\,^{3.082}_{2.812}$	$2.998\,^{3.033}_{2.963}$	$3.001\,^{3.013}_{2.990}$
	μ_2	4.700	$4.666\,^{4.721}_{4.612}$	$4.699\,^{4.716}_{4.683}$	$4.702\,^{4.707}_{4.696}$
	μ_3	5.600	$5.597\,^{5.614}_{5.583}$	$5.602\,^{5.607}_{5.596}$	$5.602\,^{5.603}_{5.600}$
state std dev force (pN)	σ_1	1.000	$1.037\,^{1.134}_{0.951}$	$0.992\,^{1.018}_{0.967}$	$0.999\,^{1.007}_{0.991}$
	σ_2	0.300	$0.254\,^{0.300}_{0.217}$	$0.287\,^{0.300}_{0.275}$	$0.301\,^{0.305}_{0.296}$
	σ_3	0.200	$0.200\,^{0.211}_{0.190}$	$0.203\,^{0.207}_{0.199}$	$0.201\,^{0.203}_{0.200}$

VALIDATION USING SYNTHETIC DATA

To verify that our BHMM posterior sampling scheme reflects the *true* uncertainty in the model parameters, we tested the scheme on synthetic data generated from a model with known parameters $\mathbf{\Theta}^*$. Given observed data $\mathbf{O}$ generated from $P(\mathbf{O} \mid \mathbf{\Theta}^*)$, sampling from the posterior $P(\mathbf{\Theta} \mid \mathbf{O})$ using the scheme described in *Sampling from the posterior of the BHMM* will provide us with confidence intervals $[\theta_{\mathrm{low}}, \theta_{\mathrm{high}}]$ for a specified confidence interval size $\alpha \in [0, 1]$. If these computed confidence intervals are accurate, we should find that the true model parameter θ^* lies in the computed confidence interval $[\theta^{(\alpha)}_{\mathrm{low}}, \theta^{(\alpha)}_{\mathrm{high}}]$ with probability α. This can be tested by generating synthetic observed data $\mathbf{O}$ from $P(\mathbf{O} \mid \mathbf{\Theta}^*)$ and verifying that we find $\theta^* \in [\theta^{(\alpha)}_{\mathrm{low}}, \theta^{(\alpha)}_{\mathrm{high}}]$ in a fraction α of these synthetic experiments.

As an example synthetic model, consider the three-state system intended to mimic a protein with (1) a highly-compliance, low-force unfolded state, (2) a moderately compliant low-population intermediate at intermediate force, and (3) a low-compliance, high-force folded state. Here, the term "compliance" refers to the width of the force or extension distribution characterizing the state. Parameters of the model are given in Table I, and the observation interval was taken to be $\tau = 1$ ms. An example realization of a model trajectory, along with the MLHMM state assignment, is shown in Figure 2. We generated a trajectory of 100 000 observations, and characterized the BHMM mean parameter estimate and 95% confidence intervals for a subset of this trajectory of varying lengths. The results, shown in Table I, show that the confidence intervals contract as trajectory length increases, as expected, and the BHMM-computed 95% confidence intervals contain the true model parameters with the expected statistics. In contrast, a model created from simply segmenting the observed forces into disjoint region and assigning state membership based on the force value alone estimates model parameters with significant bias even for 1 000 000 observations (see *Supporting Information*).

As a more rigorous test, we sampled 50 random models from the prior $P(\mathbf{\Theta})$ with two to six states, generated a 10 000 observation synthetic trajectory for each, and accumulated statistics on the observed fraction of time the true model pa-

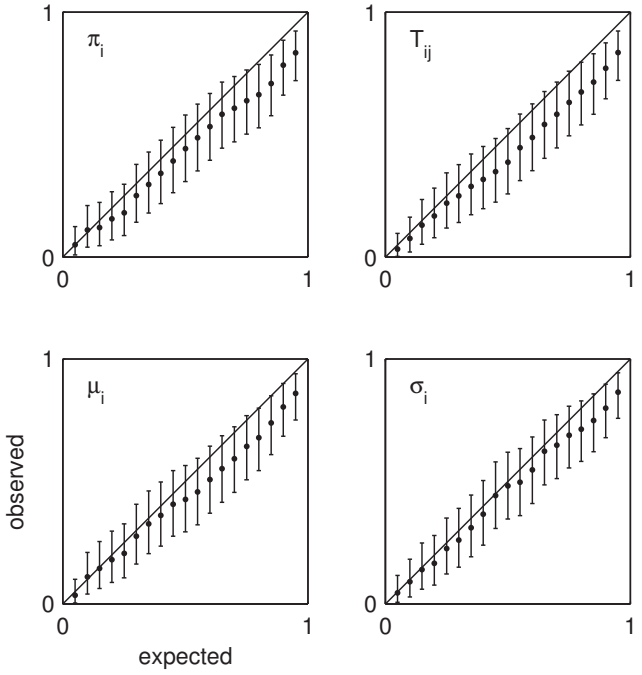

FIG. 3. **Validation of confidence intervals using randomly-generated synthetic test data.** Observed confidence intervals (points) are plotted as a function of the desired confidence intervals for equilibrium probabilities (π_i), transition probabilities (T_{ij}), state means (μ_i), and state standard deviations (σ_i). The black diagonal line indicates perfect agreement between expected and observed confidence intervals, while observed confidence intervals above the diagonal indicate overestimates of the uncertainty, and below the diagonal indicate underestimates. Because only 50 random models were evaluated, error bars denote a 95% confidence interval in the estimated observed confidence intervals.

rameters were within the BHMM confidence intervals for various values of the confidence interval width α. The results of this test are depicted in Figure 3. We expect that the plot traces the diagonal if the observed and expected confidence intervals are identical; an overestimate of the confidence interval will be above the diagonal, and an underestimate will fall below it. Because only a finite number of independent replicates of the experiment are conducted, there is some associated uncertainty with the observed confidence intervals. The results show that the observed confidence intervals line up with the expected confidence intervals to within statistical error, suggesting the BHMM confidence intervals neither underestimate nor overestimate the actual uncertainty in model parameters.

RNA HAIRPIN KINETICS IN A PASSIVE OPTICAL TRAP

We illustrate the BHMM approach applied to real force spectroscopy data by characterizing the average forces and transition rates among kinetically distinct states of the p5ab RNA hairpin in an optical trap under passive (equilibrium) conditions.

The p5ab RNA hairpin from *Tetrahymena thermophilia* was provided by Jin-Der Wen, and prepared as previously described [58]. Within the population of RNA hairpin molecules in the examined sample, there were two chemically distinct species present in the sample (i.e. as a result of post-transcriptional or other covalent modification during sample storage), exhibiting either apparent two-state (as reported previously [58]) or three-state behavior (studied here). For the purposes of testing this method, we examined a fiber that appeared to consistently exhibit three-state behavior upon visual inspection of the force timeseries data.

TABLE II. **BHMM model estimates for p5ab hairpin data.**

Property		**Value**
Equilibrium probability	π_1	$0.215\,^{0.236}_{0.193}$
	π_2	$0.046\,^{0.050}_{0.041}$
	π_3	$0.740\,^{0.762}_{0.717}$
Transition probability ($\Delta t = 1$ ms)	T_{11}	$0.954\,^{0.959}_{0.950}$
	T_{12}	$0.033\,^{0.037}_{0.029}$
	T_{13}	$0.013\,^{0.015}_{0.011}$
	T_{21}	$0.154\,^{0.169}_{0.139}$
	T_{22}	$0.650\,^{0.673}_{0.627}$
	T_{23}	$0.196\,^{0.216}_{0.180}$
	T_{31}	$0.004\,^{0.004}_{0.003}$
	T_{32}	$0.012\,^{0.013}_{0.011}$
	T_{33}	$0.984\,^{0.985}_{0.983}$
State force mean (pN)	μ_1	$12.549\,^{12.552}_{12.544}$
	μ_2	$13.016\,^{13.027}_{13.006}$
	μ_3	$13.849\,^{13.852}_{13.848}$
State force std dev (pN)	σ_1	$0.210\,^{0.213}_{0.207}$
	σ_2	$0.201\,^{0.208}_{0.193}$
	σ_3	$0.213\,^{0.214}_{0.211}$
Transition rate (s^{-1})	k_{12}	$41.4\,^{46.6}_{36.3}$
	k_{13}	$9.1\,^{11.3}_{7.2}$
	k_{21}	$194.7\,^{216.7}_{173.1}$
	k_{23}	$243.7\,^{271.5}_{219.0}$
	k_{31}	$2.6\,^{3.2}_{2.1}$
	k_{32}	$15.0\,^{16.6}_{13.4}$
State mean lifetime (ms)	τ_1	$21.9\,^{24.1}_{20.0}$
	τ_2	$2.9\,^{3.1}_{2.7}$
	τ_3	$63.1\,^{68.5}_{58.4}$

The instrument used in this experiment was a dual-beam counter-propagating optical trap with a spring constant of 0.1 pN/nm. A piezoactuator controlled the position of the trap and allowed position resolution to within 0.5 nm [59]. Drift in the instrument was less than 1 nm/minute resulting in a constant average force within 0.1 pN over the course of a typical 60 s experiment. For these constant trap position experiments, higher frequency data was recorded at 50 kHz recording the voltage corresponding to the force on the tether directly from the position-sensitive detectors. To ensure sequential samples obeyed Markovian statistics, these data were subsampled down to 1 kHz for analysis by the BHMM framework after examination of autocorrelation functions for trap positions where the hairpin appeared to remain in a single conformational state (see *Supplementary Information*).

A single observed force trajectory at a fixed trap position adequate to cause hopping among multiple states is shown in Figure 4. The most likely state trajectory from the MLHMM fit with three states is shown by coloring the observations most likely to be associated with each state, with bands of color indi-

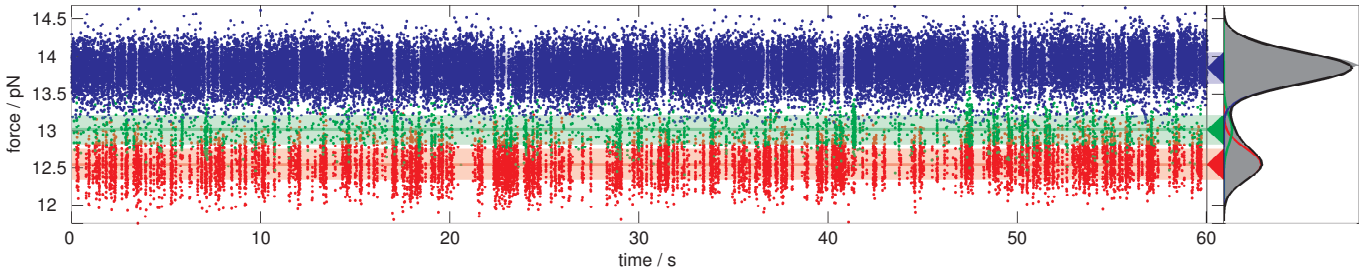

FIG. 4. **Experimental force trajectory of the p5ab hairpin and MLHMM state assignments.** Observed samples are colored by their hidden state assignments. Dark horizontal lines terminating in triangles to the right denote state means, while lightly colored bands indicate one standard deviation on either side of the state mean. The gray histogram on the right side shows the total observed probability of samples, while the colored peaks show the weighted Gaussian output contribution from each state, and the black outline the weighted sum of the Gaussian output contributions from the HMM states.

cating the mean and standard deviation about the mean force characterizing each state.

Table II lists the BHMM posterior means and confidence intervals characterizing the three-state model extracted from this single 60 s observed force trace. Several things are notable about the estimated model parameters. Surprisingly, while there is a clearly-resolved intermediate-force state (*state 2*) through which most of the flux from the high- and low-force states passes (as seen from large K_{12} and K_{23}), there are non-trivial rate constants connecting the high and low force states directly (K_{13}), indicating that while a sequential mechanism involving passing through the intermediate state is preferred, it may not be an obligatory step in hairpin formation under these conditions. While the state mean forces are clearly distinct, the state standard deviations—which reflect the width of the observed force distribution characterizing each state, rather than the uncertainty in state means—possess overlapping confidence intervals. These standard deviations reflect not only contributions from both the distribution of extensions sampled by the hairpin in each conformational state, but also from fluctuations in the handles and beads, and other sources of mechanical and electrical noise in the measurement. As we would expect the unfolded hairpin to be more compliant (i.e. possess a wider distribution of forces) than the folded hairpin, the inability to distinguish the standard deviations among states is suggestive that, for this experimental configuration and observation time, the predominant contribution to the observed force distributions for each state may be in the form of handle or bead fluctuations or other sources of measurement noise.

Finally, the lifetime of the intermediate-force state is significantly shorter than for the low- and high-force states by nearly an order of magnitude, and only a few times longer than the observation interval of 1 ms—despite this, the lifetime appears to be well-determined, as indicated by the narrow confidence intervals.

DISCUSSION

We have described an approach to determining the first-order kinetic parameters and observable (force or extension) distributions characterizing conformational states in single-molecule force spectroscopy. By use of a Bayesian extension of hidden Markov models, we are able to characterize the experimental uncertainty in these parameters due to instrument noise and finite-size datasets. The use of a detailed balance constraint additionally helps reduce the experimental uncertainty over standard hidden Markov models, as both transitions into and out of conformational states provide valuable information about state kinetics and populations in data-poor conditions. Additionally, the Gibbs sampling framework used to sample from the Bayesian posterior can be easily extended to incorporate additional nuisance parameters, such as stochastic models of instrument drift or laser power fluctuations.

We have opted to make use of a reversible transition matrix to describe the statistical kinetic behavior between the observation intervals Δt, but it is possible to use a reversible rate matrix instead by substituting a rate matrix sampling scheme [60] in the appropriate stage of the Gibbs sampling updates.

While the experimenter must currently choose the number of conformational states by hand, a number of extensions of Bayesian hidden Markov models can be used to automatically determine the number of states best supported by the data, including reversible-jump schemes [61, 62] and variational Bayes methods [63, 64].

We note that the experimenter in principle has access to the full posterior distribution of models given the observed data, so that instead of looking at the confidence of single parameters, confidence intervals in more complex functions of parameters—such as the rates or lifetimes in Table II—can be computed, or joint posterior distributions of multiple parameters examined. It is also possible to generate *synthetic* data from the current model, or family of models, to examine how the collection of additional data will further reduce uncertainties or allow discrimination among particular hypotheses. The field of *Bayesian experimental design* [65] holds numerous possibilities for selecting how future experiments can maximize information gain, and whether the information gain from the collection of additional data will be of sufficient utility to justify the expense.

ACKNOWLEDGMENTS

The authors thank Sergio Bacallado (Stanford University) for helpful feedback on this manuscript, and Steve Presse (UCSF) for engaging discussions on this topic. JDC acknowl-

edges support from a QB3-Berkeley Distinguished Postdoctoral Fellowship. FN acknowledges DFG Grant 825/2. This work was supported in part by a grant from the NSF (SM).

---

* Corresponding author

[1]

[2] 00, 0. 0.

[3] Ritort, F., 2006. Single-molecule experiments in biological physics: methods and applications. *J. Phys.: Condens. Matter* 18:R531–R583.

[4] Neuman, K. C., and S. M. Block, 2004. Optical trapping. *Rev. Sci. Instrum.* 75:2787.

[5] Moffitt, J. R., Y. R. Chemla, S. B. Smith, and C. Bustamante, 2008. Recent advances in optical tweezers. *Annu. Rev. Biochem.* 77:205–228.

[6] Bustamante, C., Z. Bryant, and S. B. Smith, 2003. Ten years of tension: single-molecule DNA mechanics. *Nature* 421:423–427.

[7] Woodside, M. T., C. García-Garcá, and S. M. Block, 2008. Folding and unfolding single RNA molecules under tension. *Curr. Opin. Chem. Biol.* 12:640–646.

[8] Bustamante, C., W. Cheng, and Y. X. Mejia, 2011. Revisiting the central dogma one molecule at a time. *Cell* 144:480–497.

[9] Fowler, S. B., R. B. Best, J. L. T. Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, and J. Clarke, 2002. Mechanical unfolding of a titin Ig domain: Structure of unfolding intermediates revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J. Mol. Biol.* 322:841–849.

[10] Cecconi, C., E. A. Shank, C. Bustamante, and S. Marqusee, 2005. Direct observation of the three-state folding of a single protein molecule. *Science* 309:2057–2060.

[11] Gebhardt, J. C. M., T. Bornschlögl, and M. Rief, 2010. Full distance-resolved folding energy landscape of one single protein molecule. *Proc. Natl. Acad. Sci. USA* 107:2013–2018.

[12] Ritort, F., 2008. Nonequilibrium fluctuations in small systems: From physics to biology. *Adv. Chem. Phys.* 137:31–123.

[13] Schütte, C., and W. Huisinga, 2002. Biomolecular conformations can be identified as metastable states of molecular dynamics. *In* P. G. Ciaret, and J.-L. Lions, editors, Handbook of Numerical Analysis - special volume on computational chemistry, Elsevier, volume X.

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

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

[16] Best, R. B., E. Paci, G. Hummer, and O. K. Dudko, 2008. Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J. Phys. Chem. B* 112:5968–5976.

[17] Yew, Z. T., M. Schlierf, M. Rief, and E. Paci, 2010. Direct evidence of the multidimensionality of the free-energy landscapes of proteins revealed by mechanical probes. *Phys. Rev. E* 81:031923.

[18] Morrison, G., C. Hyeon, M. Hinczewski, and D. Thirumalai, 2011. Compaction and tensile forces determine the accuracy of folding landscape parameters from single molecule pulling experiments. *Phys. Rev. Lett.* 106:138102.

[19] Chodera, J. D., and V. S. Pande, 2011. Splitting probabilities as a test of reaction coordinate choice in single-molecule experiments. *arχiv preprint* arXiv:1105.0710.

[20] Woodside, M. T., P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block, 2006. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314:1001–1004.

[21] Forns, N., S. de Lorenzo, M. Manosas, J. M. Huguet, and F. Ritort, 2011. Improving signal/noise resolution in single-molecule experiments using molecular constructs with short handles. *Biophys. J.* 100:1765–1774.

[22] Chodera, J. D., P. Elms, W. C. Swope, F. Noé, and V. S. Pande, 2011. A robust approach to estimating rates from time-correlation functions. *in preparation* .

[23] Rabiner, L. R., 1989. A tutorial on Hidden Markov models and selected applications in speech recognitiion. *Proceedings of the IEEE* 77:257–286.

[24] Becker, J. D., J. Honerkamp, J. Hirsch, U. Fröbe, E. Schlatter, and R. Greger, 1994. Compaction and tensile forces determine the accuracy of folding landscape parameters from single molecule pulling experiments. *Pflügers Arch.* 426:328–332.

[25] Qin, F., A. Auerbach, and F. Sachs, 2000. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys. J.* 79:1915–1927.

[26] De Gunst, M. C. M., H. R. Künsch, and J. G. Schouten, 2001. Statistical analysis of ion channel data using hidden Markov models with correlated state-dependent noise and filtering. *J. Am. Stat. Assoc.* 96:805–815.

[27] Qin, F., 2004. Restoration of single-channel currents using the segmental k-means method based on hidden Markov modeling. *Biophys. J.* 86:1488–1501.

[28] Andrec, M., R. M. Levy, and D. S. Talaga, 2003. Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden Markov models. *J. Phys. Chem. A* 107:7454–7464.

[29] McKinney, S., C. Joo, and T. Ha, 2006. Analysis of single-molecule FRET trajectories using hidden Markov models. *Biophys. J.* 91:1941–1951.

[30] Lee, T.-H., 2009. Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden Markov models. *J. Phys. Chem. B* 113:11535–11542.

[31] Li, Y., X. Qu, A. Ma, G. J. Smith, N. F. Scherer, and A. R. Dinner, 2009. Models of single-molecule experiments with periodic perturbations reveal hidden dynamics in RNA folding. *J. Phys. Chem. B* 113:7579–7590.

[32] Gopich, I. V., and A. Szabo, 2009. Decoding the pattern of photon colors in single-molecule FRET. *J. Phys. Chem. B* 113:10965–10973.

[33] Smith, D. A., W. Steffen, R. M. Simmons, and J. Sleep, 2001. Hidden-Markov methods for the analysis of single-molecule actomoysin displacement data: The variance-hidden-Markov method. *Biophys. J.* 81:2795–2816.

[34] Milescu, L. S., A. Yildiz, P. R. Selvin, and F. Sachs, 2006. Extracting dwell time sequences from processive molecular motor data. *Biophys. J.* 91:3135.

[35] Müllner, F. E., S. Syed, P. R. Sevin, and F. J. Sigworth, 2010. Improved hidden Markov models for molecular motors,

part 1: Basic theory. *Biophys. J.* 99:3684–3695.

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

[37] Viterbi, A. J., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Info. Theory* 13:260–269.

[38] Kruithof, M., and J. van Noort, 2009. Hidden Markov analysis of nucleosome unwrapping under force. *Biophys. J.* 96:3708–3715.

[39] Aittokallio, T., and E. Uusipaikka, 2008. Computation of standard errors for maximum-likelihood estimates in hidden Markov models. Technical report, University of Turku. Technical Report No. 379.

[40] Merialdo, B., 1993. On the locality of the forward-backward algorithm. *IEEE Trans. Speech and Audio Proc.* 1:255–257.

[41] Robert, C. P., G. Celeux, and J. Diebolt, 1993. Bayesian estimation of hidden Markov chains: A stochastic implementation. *Stat. Prob. Lett.* 16:77–83.

[42] Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. *J. Econometrics* 75:79–97.

[43] Scott, S. L., 2002. Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *J. Am. Stat. Assoc.* 97:337–351.

[44] Rydén, T., 2008. EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective. *Bayesian Analysis* 3:659–688.

[45] Noé, F., 2008. Probability distributions of molecular observables computed from Markov models. *J. Chem. Phys.* 128:244103.

[46] Metzner, P., F. Noé, and C. Schütte, 2009. Estimating the sampling error: Distribution of transition matrices and functions of transition matrices for given trajectory data. *Phys. Rev. E* 80:021106.

[47] Geman, S., and D. Geman, 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal.* 6:721–741.

[48] Liu, J. S., 2002. Monte Carlo strategies in scientific computing. Springer-Verlag, New York, 2nd ed. edition.

[49] Comstock, M. J., T. Ha, and Y. R. Chemla, 2011. Ultrahigh-resolution optical trap with single-fluorophore sensitivity. *Nature Methods* 8:335–340.

[50] van Kampen, N. G., 1997. Stochastic processes in physics and chemistry. Elsevier, second edition.

[51] Bilmes, J. A., 1998. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Technical report, University of California, Berkeley.

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

[53] Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977. Maximum-likelihood from incomplete data via the EM algortihm. *J. Royal Statist. Soc. B* 39:1–38.

[54] Chodera, J. D., and F. Noé, 2010. Probability distributions of molecular observables computed from Markov models: II. uncertainties in observables and their time-evolution. *J. Chem. Phys.* 133:105012.

[55] Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.

[56] Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. *Proc. Royal Soc. A* 186:453–461.

[57] Goyal, P., 2005. Prior probabilities: An information-theoretic approach. *In* K. H. Knuth, A. E. Abbas, R. D. Morriss, and J. P. Castle, editors, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, American Institute of Physics, 366–373.

[58] Wen, J.-D., M. Manosas, P. T. X. Li, S. B. Smith, C. Bustamante, F. Ritort, and J. Ignacio Tinoco, 2007. Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophys. J.* 92:2996–3009.

[59] Bustamante, C. J., and S. B. Smith, 2006. Light-force sensor and method for measuring axial optical-trap forces from changes in light momentum along an optic axis.

[60] Hummer, G., 2005. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics* 7:34.

[61] Robert, C. P., T. Rydén, and D. M. Titterington, 2000. Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *J. R. Statist. Soc. B* 62:57–75.

[62] De Gunst, M. C. M., and B. Schouten, 2003. Model selection for hidden Markov models of ion channel data by reversible Markov chain Monte Carlo. *Bernoulli* 9:373–393.

[63] Beal, M. J., 2003. Variational algorithms for approximate Bayesian inference. Master's thesis, University of Cambridge, UK.

[64] Bronson, J. E., J. Fei, J. M. Hofman, R. L. G. Jr., and C. H. Wiggins, 2009. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* 97:3196–3205.

[65] Chaloner, K., and I. Verdinelli, 1995. Bayesian experimental design: A review. *Statist. Sci.* 10:273–204.