Bayesian hidden Markov model analysis of single-molecule biophysical experiments

John D. Chodera,[1,*] Phillip Elms,[2,†] Frank Noé,[3,‡] Bettina Keller,[4,‡] Aaron Ewall-Wice,[4,§] and Nina Singhal Hinrichs[5,¶]

[1]*California Institute of Quantitative Biosciences (QB3), University of California, Berkeley, 260J Stanley Hall, Berkeley, CA 94720, USA*
[2]*Biophysics Graduate Group, Berkeley, CA 94720, USA*
[3]*DFG Research Center Matheon, FU Berlin, Arnimallee 6, 14195 Berlin, Germany*
[4]*University of Chicago, IL 60637, USA*
[5]*Departments of Statistics and Computer Science, University of Chicago, IL 60637, USA*

(Dated: August 19, 2010)

Single-molecule experiments are now able to probe the dynamics of single biological macromolecules or macromolecular assemblies using a variety of techniques, including fluorescence measurements, fluorescent energy transfer (FRET), and optical or atomic force microscopy. While able to report on the behavior of single molecules, the observed signal is usually only an *indirect* probe of molecular conformation, without the guarantee of a unique correspondence between the observed signal and molecular conformation. This can lead to conformational states with strongly overlapping spectroscopic signatures, making resolution of the instantaneous molecular configuration difficult. Hidden Markov models (HMMs), now a standard approach in machine learning, have been employed to solve this problem by using kinetic information to aid resolution of the observed temporal signal into a sequence of distinct conformational states. These methods suffer from an important drawback: maximum-likelihood fitting procedures do not give a clear picture of how well the model parameters are determined by the data due to instrument noise and finite-sample statistics. Here, we propose a solution to this problem through a simple *Bayesian* extension of hidden Markov model analysis that allows both the uncertainties in the transition rates and hidden state assignments to be characterized. The method is based on Gibbs sampling, allowing it to be easily extended to other models of observables or to multiple observables by simply "plugging in" new components of the model. As there is significant cost in, for example, doubling the size of the dataset by the collection of additional data, the method also provides a straightforward way to assess how further data acquisition experiments will reduce model parameter uncertainties, so that the experimenter may judge whether additional experiments are worth the cost. We illustrate the method for the analysis of biomolecules by examining two- and three-state behavior of an RNA hairpin measured by optical force microscopy.

Keywords: hidden Markov model (HMM); Bayesian analysis; single-molecule experiments; statistical error; statistical uncertainty

I. INTRODUCTION

Recent advances in biophysical measurement have led to an unprecendented ability to monitor the dynamics of single biomolecules, such as proteins and nucleic acids [1]. These experiments aim to probe the statistical, heterogeneous dynamics relevant to folding and function. Recent studies have examined the conformational dynamics of large RNA molecules under equilibrium and nonequilibrium conditions by monitoring the energy transfer between two covalently attached fluorophores [2, 3]; the turnover of individual molecules of substrate by enzymes [4]; permeation and gating events of single ion channels [5]; and the fluctuation of nucleic acids under external forces in optical traps [6, 7] or atomic force microscopes [8].

Unlike corresponding *ensemble* experiments, where spectroscopic observables appear to evolve deterministically after an external perturbation (such as a laser-induced temperature jump) and spectroscopic fluctuations cannot generally be observed at equilibrium, the behavior of spectroscopic probes of *single molecules* both in and out of equilibrium contains a great deal of stochastic fluctuation. While some of this fluctuation is undoubtedly due to measurement noise, some large component of this fluctuation is due to conformational dynamics of the molecule under study. Often, the dynamics appears to be dominated by stochastic interconversions between two or more strongly *metastable states* [?], regions of conformation space in which the system persists for long times before making a transition (often accurately described by first-order kinetics) to another state (a situation also observed in NMR relaxation-dispersion experiments [9] and molecular dynamics simulations [10]).

While visual inspection of the dynamics may suggest the clear presence of multiple metastable states, characterization of these states is often difficult. First, the spectroscopic observable is unlikely to correspond to a true reaction coordinate able to easily separate all metastable states, and second, measurement noise may further broaden the spectral signatures of individual states. As a result, there is often a large degree of spectral overlap in their signatures [? ?]. Attempting to separate these states with simple separation points can often lead to a high degree of state misassignment that corrupts both the

*Corresponding author; jchodera@berkeley.edu
†elms@berkeley.edu
‡frank.noe@fu-berlin.de
§olivercromwell@uchicago.edu
¶nshinrichs@uchicago.edu