Pearce MJ, Mintseris J, Ferreyra J, Gygi SP, Darwin KH. Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. Science. 2008;322:1104–7.
Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA. Evolution of protein complexes by duplication of homomeric interactions. Genome Biol. 2007;8:R51.

Heteronuclear Multidimensional Nuclear Magnetic Resonance Spectroscopy

► Multidimensional NMR Spectroscopy
► Structure Determination by NMR: Overview

Hidden Markov Modeling in Single-Molecule Biophysics

Mario R. Blanco[1], Alexander E. Johnson-Buck[2] and Nils G. Walter[2]
[1]Department of Cellular and Molecular Biology, The University of Michigan, Ann Arbor, MI, USA
[2]Department of Chemistry, The University of Michigan, Ann Arbor, MI, USA

Synonyms

HMM; Markov model

Definition

A hidden Markov model (HMM) is a probabilistic model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states.

Introduction

The development of ► single-molecule spectroscopy has allowed for the investigation of a variety of biological questions previously inaccessible by ensemble techniques. The strength of single-molecule tools comes from the high-resolution data extracted from such experiments. Proper interpretation of these data requires efficient, unbiased analysis routines that are able to distinguish relevant signals from the intrinsically noisy measurements. The hidden Markov model, a statistical algorithm initially developed for speech recognition, has been adapted for the analysis of a variety of single-molecule signals. In this article, we will give a general introduction to the theoretical basis of hidden Markov modeling and the various single-molecule techniques in which they have been co-opted for signal analysis.

Basic Characteristics

Signals collected from single-molecule experiments can be described as a series of discrete states governed by an underlying physical property of the molecule(s) being interrogated. The discrete states are often obscured (hidden) by noise that is inherent to the experimental technique, making the identification and characterization of these states difficult. The hidden states are no longer efficiently detected by visual inspection or simple algorithms and doing so can introduce bias and an incomplete characterization of the underlying behavior(s). Probabilistic maximum-likelihood algorithms, like a hidden Markov model (HMM), have become the preferred method of analysis; they provide a mathematically derived routine that limits the possibility of user bias as well as providing a theoretical framework with which to interpret the quantitative results extracted from single-molecule experiments. A HMM describes a stochastic progression through a series of discrete states, where the likelihood of the next event in a series of observations can be predicted upon knowledge of the immediately preceding event and does not depend on knowledge of any of the prior events; that is, the process is Markovian (Fraser 2008). The main model assumptions of a HMM are as follows:

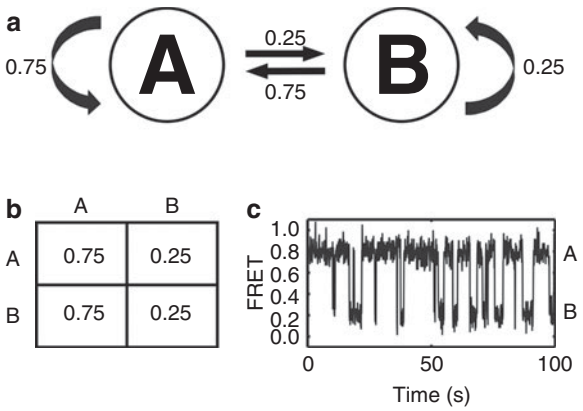
1. Given the current state, the probability of the current observation is independent of states and observations at all earlier times.
2. Given the current state, the probability of the next observed state occurring is independent of earlier states. More simply put, the future does not depend on the past.

HMMs are well suited for single-molecule analysis because of their ability to find discrete states, usually stable, biologically relevant conformations, within noisy time series data, and to reliably find the most probable path through these states. As a molecule

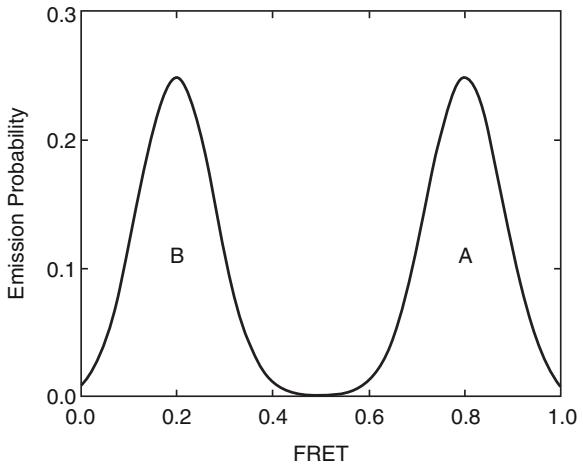
transitions from one stable conformation to another it is often the case that the process is Markovian and therefore governed by single exponential kinetics. Through the iterative optimization of the HMM parameters – the probability matrices of transition, emission, and initiation – a model is derived that best approximates the data. The transition probability matrix describes the probability of any one state changing to any other state or staying in the same state in the subsequent time step (Fig. 1). The emission probability distribution contains the probabilities of a specific signal value being emitted by each discrete state. Calculating an emission probability often requires an assumption of the noise in the system, usually shot noise that is efficiently approximated by simple Gaussian distributions (Fig. 2). The initiation probability matrix gives the probabilities of starting at each of the possible discrete states.

During the evaluation of single-molecule data one does not usually know what the idealized values of the states *A* and *B* or their probability matrices are. To begin the analysis, a Markov model is estimated with a given number of states and a probability transition matrix. The model is then optimized by determining the parameters that yield the maximum likelihood for the given trace. If the number of states has been determined, then the appropriate number of states can be entered and the rest of the parameters optimized. It is often the case that the number of states is also unknown, but can be determined by optimization of various Markov models with differing numbers of states and the results compared through a criterion such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) (Blanco and Walter 2010). Due to the iterative process required to find the most probable model parameters and sequence of states, HMMs are computationally expensive. The availability of a family of algorithms whose complexity scales only linearly with the length of the trajectory makes it possible to apply HMMs to time series that reach biologically relevant long time scales. The algorithms most commonly used include:

1. The forward algorithm, which calculates the conditional probability of being in a state *s* at time *t* given all of the observations up to that time. It also calculates the conditional probability of each observation given previous observations. Using these terms it calculates the probability of the entire data sequence given the model. This is also the first phase of the Baum–Welch algorithm (below).
2. The Viterbi algorithm, used when one needs to estimate a sequence of states from a sequence



Hidden Markov Modeling in Single-Molecule Biophysics, Fig. 1 *Simple hidden Markov model* (**a**) An example of a hidden Markov model with two states with the transition probabilities presented above individual arrows to represent the likelihood of transiting from one state to another or staying in the same state within the next time step. (**b**) The transition probability matrix of the Markov model from (**a**). (**c**) A simulated single-molecule FRET trajectory using the two-state model with states *A* (FRET = 0.8) and *B* (FRET = 0.2)



Hidden Markov Modeling in Single-Molecule Biophysics, Fig. 2 *Emission probabilities for the Markov process*. The emission probability distributions of the two states *A* and *B* from the simulated single-molecule FRET trajectory in Fig. 1 are plotted. Here the emission probabilities are calculated by assuming Gaussian noise distributions around the discrete mean FRET values ($A = 0.8$, $B = 0.2$), modeling the shot noise of the signal collection instrumentation

of observations. It finds the most probable state sequence.

3. The Baum–Welch algorithm (forward-backward algorithm) calculates, given a given sequence of observations and an initial set of model parameters, in a single pass based on the forward algorithm a new set of parameters that has higher likelihood of being correct. Running many iterations of the Baum–Welch algorithm yields a sequence that approaches a local maximum of the likelihood.

Although HMMs provide an excellent tool for analyzing single-molecule data it must be noted that not all data fulfill the assumptions of a Markov process. For example, the changes of biomolecules can exhibit time-dependent transition probabilities due to molecular memory effects. In these cases HMMs can still approximate the data but care must be taken in the interpretation of the results. For an in-depth discussion of the mathematical foundations underlying HMM, we refer the reader to Rabiner (1989) and Fraser (2008).

Applications of Hidden Markov Models in Single-Molecule Biophysics

The general framework of HMMs has been adapted to a variety of single-molecule techniques. Due to the differing character of the signals acquired from these techniques there is not a single HMM that can be utilized for the analysis of all types of data collection. However, HMMs have been modified and improved in the various fields to better model the data and noise of each particular field.

Ion Channel Recordings

Among some of the earliest single-molecule experiments came from the electrophysiology field, where the action potential across a single ion channel can be recorded over time with the use of ▶ patch clamp techniques. These techniques allow for the direct measurement of ionic currents through a single channel protein molecule. The amplitude of the signal collected describes the permeability of ions through the channel and the change in this permeability can be recorded in real time. HMMs can approximate the open and closed states of these pores effectively. This HMM implementation assumes the underlying signal is a Markov process whose noise is assumed to be Gaussian. QuB (available at http://www.qub.buffalo.edu/wiki/index.php/Main_Page) is a readily available software package routinely used for the analysis of ion channel recordings (Qin et al. 2000).

Fluorescence Microscopy

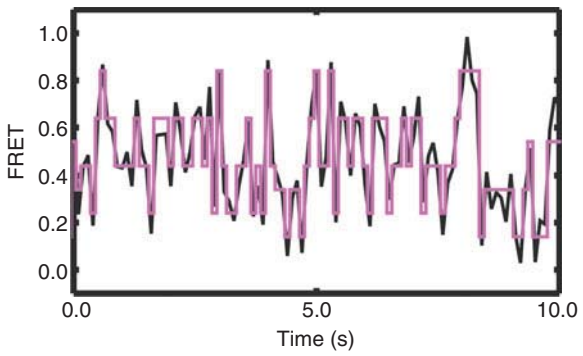
Single-molecule fluorescence microscopy has become one of the most popular single-molecule techniques due to the wide range of biomolecules that can be studied with this technique. HMMs have been adapted for various types of fluorescence microscopy to better model the different forms of intensity time traces collected.

Single-Molecule Fluorescence Resonance Energy Transfer (smFRET)

▶ smFRET can provide a real-time view of dynamics of biomolecules ranging from small catalytic RNAs (ribozymes) to large RNA-protein complexes such as the ribosome. The distance-dependent interaction between two fluorophores can report on intra- and intermolecular conformational changes. The application of HMM to the analysis of smFRET data became more accessible after the release of HaMMy, a user-friendly analysis software specifically designed for the analysis of smFRET data (McKinney et al. 2006). Previously, less sophisticated algorithms such as thresholding techniques were used that could often not handle complicated trajectories. The adaptation of HMM to smFRET has been advantageous for the field and the tool continues to be developed for the analysis of systems with more complicated behaviors. An example of the power of HMM analysis of smFRET trajectories is the trajectory of a single pre-mRNA molecule imaged under splicing conditions (Fig. 3) (Blanco and Walter 2010). The large number of states and rapid kinetics of transitions makes this type of trace difficult to analyze without the use of HMMs. In addition to HaMMy, programs such as QuB and vbFRET (Bronson et al. 2010) (available at http://vbfret.sourceforge.net) are available for smFRET analysis with advantages for more complex trajectories (Blanco and Walter 2010).

Switchable FRET

Switchable FRET is a combination of two techniques: smFRET and photoswitching, the reversible activation and deactivation of fluorophores commonly used in super-resolution imaging techniques such as



Hidden Markov Modeling in Single-Molecule Biophysics, Fig. 3 *A complex smFRET trajectory analyzed with HMM.* HMM can be utilized to characterize the dynamics of single pre-mRNA molecules during splicing that exhibit rapid kinetics and a large number of states without the need for smoothing which would eliminate these small rapid conformational changes. In black is the raw FRET trajectory, and in magenta the five-state HMM idealized fit. In this particular case, the QuB ion channel analysis software was utilized for the HMM analysis

► stochastic optical reconstruction microscopy (STORM). This technique utilizes multiple donor–acceptor fluorophore pairs to sequentially probe and obtain multiple distances within a single molecule. Traditional HMMs for smFRET cannot incorporate the stochastic photoswitching of the acceptor dye and therefore a linked hidden Markov model was developed where FRET and donor–acceptor stoichiometry are tracked (Uphoff et al. 2010). The linked HMM allows for the proper identification of states and determination of their transitions.

Multi-fluorophore Bleaching

Some single-molecule experiments utilize multiple fluorophores to help determine the number of subunits assembled in a particular biological complex. The sudden drops in intensity when single fluorophores each undergo an irreversible ► photobleaching event can be used as a measure of the number of particles present. HMMs have been developed for the unbiased determination of the discrete steps in the fluorescence intensity traces. Up to 30 fluorophores can be reliably detected through the use of HMMs (Messina et al. 2006). The ability to reliably visualize and count the number of single molecules present has applications ranging from self-assembly of biomolecules to tracking the assembly of *trans* factors in multi-component systems.

Molecular Motor Step Size

HMMs have been developed for the case of fluorescent measurements of molecular motors. An HMM variant, the variable-stepsize HMM (vsHMM) where the position of the motor is modeled as a large number of states, has been developed to more accurately track the movement of these motors as a result of their reaction cycle. This model differs in that it allows for an arbitrary distribution of step sizes that allows it to run as a robust algorithm with little user input. The algorithm has also been extended to the variable-stepsize integrating-detector HMM (VSI-HMM) which serves to better model the variation in signals during data acquisition such as random baseline fluctuations. Together, these HMMs have been utilized to characterize the movement of a myosin motor both in vitro and in vivo (Syed et al. 2010).

Single-Particle Tracking

► Single-particle tracking can be used to extract modes of diffusion for a single molecule. In the case of biomolecules diffusing through a cell this approach can provide insight into the regions of localized activity, concentration gradients, or sites of modification. A single-particle track, with certain assumptions, can be modeled with a two-state HMM. The two-state HMM is optimized through the diffusion coefficients of the states and the rates of transition between them. It has been shown that this HMM is sufficient to extract multiple states of diffusion within a single trajectory in a practical manner (Das et al. 2009).

Tethered Particle Microscopy (TPM)

► TPM experiments use light microscopy to measure the position of a bead tethered to a microscope slide via a polymer to infer the behavior of the polymer. An example is the use of DNA as the tether to measure DNA folding/unfolding dynamics and the effects of DNA-binding proteins on those dynamics. HMM analysis algorithms have been used to track the subtle changes in bead position and determine the relevant changes from those induced by Brownian motion by incorporating factors for the diffusive motion of the bead (Beausang et al. 2007). This approach has allowed for data analysis without the need for filtering. Another improvement to the HMM algorithms used for these experiments introduced factors to account for the nonlinear extension of DNA, allowing for a more accurate, quantitative assessment of the kinetics.

Limitations of Hidden Markov Models

HMMs can be a powerful tool for the unbiased analysis of single-molecule data, but the resulting models need to be carefully inspected. As noted previously, not all biological processes studied under single-molecule conditions fulfill the Markov property, thus violating one of the assumptions of HMMs. This can affect the ability of the model to fully recapitulate the underlying behaviors. Additionally, as can be seen by the various adaptations of HMMs, it is necessary to define the right set of parameters to properly simulate the noise present in the system as well as the discrete number of states of interest. Finally, as with any fitting technique, when using HMM a rigorous test for model selection is required. This testing is often complicated by the lack of a clear and decisive way of selecting the proper number of states for the model. Although several methods have been presented for an unbiased approach at state selection, there is little consensus regarding which is best.

Cross-References

- ▶ Patch-Clamp Recording of Single Channel Activity: Acquisition and Analysis
- ▶ Photoactivated Localization Microscopy (PALM)
- ▶ Single Fluorophore Photobleaching
- ▶ Single-Molecule Fluorescence Resonance Energy Transfer
- ▶ Single-Molecule Spectroscopy
- ▶ Single-Particle Tracking
- ▶ Stochastic Optical Reconstruction Microscopy
- ▶ Tethered Particle Microscopy

References

Beausang JF, et al. DNA looping kinetics analyzed using diffusive hidden Markov model. Biophys J. 2007;92(8): L64–6.

Blanco M, Walter NG. Analysis of complex single molecule 419FRET time trajectories. Method Enzymol. 2010;472: 153–78.

Bronson JE, et al. Graphical models for inferring single molecule dynamics. BMC Bioinform. 2010;11(Suppl 8):S2.

Das R, Cairo CW, Coombs D. A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. PLoS Comput Biol. 2009;5(11): e1000556.

Fraser AM. Hidden Markov models and dynamical systems. Philadelphia: SIAM; 2008.

McKinney SA, Joo C, Ha T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. Biophys J. 2006;91(5):1941–51.

Messina TC, et al. Hidden Markov model analysis of multichromophore photobleaching. J Phys Chem B. 2006; 110(33):16366–76.

Qin F, Auerbach A, Sachs F. A direct optimization approach to hidden Markov modeling for single channel kinetics. Biophys J. 2000;79(4):1915–27.

Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989;77: 257–86.

Syed S, et al. Improved hidden Markov models for molecular motors, part 2: extensions and application to experimental data. Biophys J. 2010;99(11):3696–703.

Uphoff S, et al. Monitoring multiple distances within a single molecule using switchable FRET. Nat Meth. 2010;7(10): 831–6.

Hierarchically Structured Lipid Systems

Chandrashekhar V. Kulkarni
Biological and Soft Systems, Cavendish Laboratory, University of Cambridge, Cambridge, UK

Synonyms

Lipid phase equilibria; Lipid self-organization; Lipid superstructures; Multiscale structural ordering of lipids

Definition

Amphiphilic lipids self-assemble into various thermodynamically stable nanostructures in the presence of water, which can be kinetically stabilized into hierarchically ordered lipid systems exhibiting multiple structural length scales.

Introduction

Amphiphilic molecules such as lipids have the inherent tendency of self-assembling in an aqueous environment. The hydrophobic effect acts upon minimizing interactions between the water and hydrophobic parts of lipid molecules, thus having a prime contribution