A Comparative Study of Multivariate and Univariate Hidden Markov Modelings in Time-Binned Single-Molecule FRET Data Analysis

Yang Liu,[†] Jeehae Park,[†,‡] Karin A. Dahmen,[†] Yann R. Chemla,[†,‡] and Taekjip Ha*[,†,‡]

Center for the Physics of Living Cell, Department of Physics, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801 and Center for Biophysics and Computational Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801

Received: June 18, 2009; Revised Manuscript Received: November 23, 2009

We compare two different types of hidden Markov modeling (HMM) algorithms, e.g., multivariate HMM (MHMM) and univariate HMM (UHMM), for the analysis of time-binned single-molecule fluorescence energy transfer (smFRET) data. In MHMM, the original two channel signals, i.e., the donor fluorescence intensity (I_D) and acceptor fluorescence intensity (I_A), are simultaneously analyzed. However, in UHMM, only the calculated FRET trajectory is analyzed. On the basis of the analysis of both synthetic and experimental data, we find that, if the noise in the signal is described with a proper probability distribution, MHMM generally outperforms UHMM. We also show that, in the case of multiple trajectories, analyzing them simultaneously gives better results than averaging over individual analysis results.

Introduction

Single-molecule fluorescence energy transfer (smFRET) has been regarded as one of the most powerful and adaptable single-molecule techniques.[1] In smFRET experiments, usually two fluorescent dye molecules (termed donor and acceptor) are attached to a single molecule. By measuring the extent of nonradiative energy transfer between the donor and acceptor, we get information on their intervening distance. In other words, FRET acts as a spectroscopic ruler. By tracking FRET changes over time, conformational dynamics of single molecules can then be observed in real time. The original signals obtained, i.e., the time traces, have two channels: donor fluorescence intensity (I_D) and acceptor fluorescence intensity (I_A). Usually, those two traces are converted to one FRET trajectory by calculating the ratio of acceptor intensity to the total emission intensity, i.e.,

$$\mathrm{FRET} = I_A/(I_A + I_D) \tag{1}$$

Ideally, well-defined FRET values can be observed corresponding to stable conformational states of the system (in this paper, by "conformation", we just mean the distance of the donor−acceptor pair rather than general conformations of the molecule).

Typically, the data obtained in smFRET experiments are quite noisy due to instrumental noise, e.g., shot noise. To analyze noisy FRET data, many schemes have been suggested.[2−8] Among them, the hidden Markov modeling (HMM) turns out to be the most accurate and reliable one, because it is conducted in a fully probabilistic way.[5,8] Historically, HMM was first developed in speech recognition in the mid-1970s.[9] Since then, it has been widely used as a workhorse for temporal pattern recognition in many fields. In biophysics, HMM has been extensively used in the analysis of biological sequences (in particular DNA),[10] single ion channel recordings,[11,12] molecule motors,[13] DNA looping,[14] nucleosome unwrapping,[15] and smFRET trajectories.[5,6,8,16]

Even though HMM has been shown to be very successful in analyzing smFRET data, we are aware that most of those HMM analyses consider the FRET trajectory alone.[6,8] In other words, a univariate HMM (UHMM) is used, where the analyzed univariate time series is just the FRET trajectory calculated from the original two channel signals (I_D and I_A) according to eq 1. Presumably, if we could directly analyze the original two channel signals simultaneously, we would be able to extract more information simply because we fully utilize the data. A joint statistical analysis of multichannel time series based on a joint maximum likelihood estimation method has been performed in FRET data from systems consisting of a single quantum dot-$(Cy5)_n$ hybrid.[17] However, to our knowledge, a multivariate HMM (MHMM) has not been developed for the fully probabilistic multichannel time-binned smFRET data analysis. Moreover, conformational changes during biomolecular interactions are seldom one-dimensional, so there is an increasing interest in extending the reach of FRET to higher dimensions,[1] for instance, using the so-called three-color FRET scheme.[18,19] From the HMM point of view, more colors means more channels. Therefore, an MHMM would be necessary in this case.

In this work, we develop an MHMM for the time-binned data analysis of the original two channel signals in two-color smFRET experiments. We compare the performance of MHMM with that of UHMM in analyzing both synthetic and experimental data and find that in general MHMM outperforms UHMM if the noise is characterized with a correct probability distribution.

Hidden Markov Modeling

Basic Assumptions. As a statistical model, HMM assumes that the state sequence in the system being modeled is a Markov process with unknown (hidden) parameters. The state sequence is not directly visible to the observer or just masked by the instrumental noise. For example, in the HMM analysis of smFRET, we first assume that the conformational state-to-state transitions are governed by single exponential decay kinetics; i.e., the conformational state sequence is a first-order Markov

* To whom correspondence should be addressed. E-mail: tjha@illinois.edu.
[†] Center for the Physics of Living Cell.
[‡] Center for Biophysics and Computational Biology.

10.1021/jp9057669
Published on Web 04/02/2010

chain. This state sequence is always buried in the quite noisy FRET trajectory. In other words, it is hidden. To analyze the noisy data in a probabilistic way, we further assume that at each conformational state the system will emit a random signal with a well-defined observation probability distribution. This obtained signal series is called the observation sequence. The challenge is to extract the hidden model parameters which describe the state-transition probabilities and the observation probability distribution, from the noisy observation sequence. The extracted model parameters can then be used to perform further analysis to find the optimal state sequence which best explains the observations, i.e., the experimental data.

For a standard HMM, we denote the state sequence of length T as $q = (q_1, q_2, \ldots, q_T)$, which is a first-order Markov chain:

$$P(q_t|q_{t-1}, O_{t-1}; \ldots; q_1, O_1) = P(q_t|q_{t-1}) \tag{2}$$

And the observation sequence associated with the state sequence is denoted as $O = (O_1, O_2, \ldots, O_t)$. Note that the observation at time t depends on the state at t only:

$$P(O_t|q_t; q_{t-1}, O_{t-1}; \ldots; q_1, O_1) = P(O_t|q_t) \tag{3}$$

Such a standard HMM is characterized by the following elements. (1) N: **the number of hidden states**. We denote the set of states as $\mathcal{S} = \{1, 2, \ldots, N\}$. (2) π: **the initial state distribution**. $\pi_i = P(q_1 = i)$ with $1 \le i \le N$. (3) A: **the state-transition probability matrix**. $a_{i,j} = P(q_t = j|q_{t-1} = i)$ with $1 \le i, j \le N$. Here, we assume $a_{i,j}$ is time-independent; i.e., the hidden Markov chain is time-homogeneous. (4) B: **the observation probability distribution** (OPD). $b_i(O) = P(O_t = O|q_t = i)$ with $1 \le i \le N$. Conventionally, we use the compact notation $\lambda = (\pi, A, B)$ to denote the whole parameter set of the HMM.

In smFRET data analysis, if the instrument noise is mainly due to shot noise, and the noises of the two channels (acceptor and donor) are uncorrelated, one can suggest a simple two-dimensional Poisson OPD:

$$b_i(\mathbf{O}_t) = \prod_{k=1}^{d} \frac{e^{-\mu_{i,k}} \mu_{i,k}^{O_{t,k}}}{O_{t,k}!} \tag{4}$$

with $d = 2$. This suggests that, at time t, the system at hidden state $q_t = i$ emits a random two-dimensional Poisson vector $\mathbf{O}_t = (O_{t,1}, O_{t,2}) = (I_A(t), I_D(t))$ with mean vector given by $\boldsymbol{\mu}_i = (\mu_{i,1}, \mu_{i,2}) = (\langle I_A\rangle_i, \langle I_D\rangle_i)$. Such a HMM with two-dimensional Poisson OPDs will be called a multivariate Poisson HMM (MPHMM).

Note that the multivariate HMM (MHMM) actually has been well developed in speech recognition.[9] In general, the MHMM regards an ordered sequence of vectors as noisy multivariate observations of a Markov chain. The most general representation of the OPD is a finite mixture of multivariate Gaussian distributions:

$$b_i(\mathbf{O}_t) = \sum_{m=1}^{M} C_{im} \mathcal{N}(\mathbf{O}_t, \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im}) \tag{5}$$

with $\mathcal{N}(\mathbf{O}, \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im}) = [1/(2\pi)^{d/2}|\boldsymbol{\Sigma}_{im}|^{1/2}] \exp\{-1/2(\mathbf{O} - \boldsymbol{\mu}_{im})' \cdot \boldsymbol{\Sigma}_{im}^{-1} \cdot (\mathbf{O} - \boldsymbol{\mu}_{im})\}$. Here, $\mathbf{O}$ is the vector being modeled, M is the number of mixture components, and C_{im} is the mixture coefficient for the mth mixture in state i. The prime denotes vector transpose. $\boldsymbol{\mu}_{im}$ and $\boldsymbol{\Sigma}_{im}$ are the mean vector and covariance matrix for the mth mixture component in state i. Note that the mixture coefficients C_{im} satisfy the stochastic constraint $\sum_{m=1}^{M} C_{im} = 1$ and $C_{im} > 0$ for all components so that the OPD is properly normalized.

In smFRET data analysis with two-channel signal ($d = 2$), if the instrument noise has complicated components rather than simple shot noise, the above finite mixture of multivariate Gaussians with $d = 2$ will be a good candidate of the OPD. The number of mixture components M can be a tunable parameter. Such a HMM will be called a multivariate Gaussian-mixture HMM (MGmHMM). A special case of MGmHMM is that $M = 1$; i.e., there is just one Gaussian component. It will be called a multivariate Gaussian HMM (MGHMM).

In analyzing the smFRET trajectory alone with a univariate HMM, one still has different candidates of OPD. If the instrument noise is mainly shot noise, then it has been argued that the FRET distribution can be described by a Beta OPD. The reason follows. If the mean values of two independent Poisson distributions are large enough, e.g., $\langle I_A\rangle_i, \langle I_D\rangle_i \ge 5$, then the random variable $\mathrm{FRET}(t) \equiv I_A(t)/[I_A(t) + I_D(t)]$ can be well described by a Beta distribution:[20,21]

$$b_i(O_t; \alpha_i, \beta_i) = \frac{O_t^{\alpha_i - 1}(1 - O_t)^{\beta_i - 1}}{B(\alpha_i, \beta_i)} \tag{6}$$

Here, the mean value $\alpha_i/(\alpha_i + \beta_i) = \langle I_A\rangle_i/(\langle I_A\rangle_i + \langle I_D\rangle_i) = \mathrm{FRET}_i$ is the idealized FRET value for the conformational state i. The normalization factor $B(\alpha_i, \beta_i) = \Gamma(\alpha_i)\Gamma(\beta_i)/\Gamma(\alpha_i + \beta_i)$, with $\Gamma(x)$ being the Gamma function, is called the Beta function. Such a HMM with one-dimensional Beta OPDs will be called a univariate Beta HMM (UBHMM).

In a previous work, it was assumed that the FRET distributions can be approximated well by Gaussian distributions with mean μ_i and standard deviation σ_i:[8]

$$b_i(O_t) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(O_t - \mu_i)^2}{2\sigma_i^2}\right\} \tag{7}$$

Here, $\mu_i = \mathrm{FRET}_i$ is the idealized FRET value for the conformational state i and $O_t = \mathrm{FRET}_t$ is the observed FRET value. We call such a HMM with one-dimensional Gaussian OPDs a univariate Gaussian HMM (UGHMM).

Up to now, we have defined five different HMMs. On the basis of whether their OPDs are multivariate or univariate, they can be classified as two types: (1) multivariate—MPHMM, MGHMM, and MGmHMM—or (2) univariate—UBHMM and UGHMM. A natural question arises here: Which HMM can best explain the noisy smFRET data? In principle, one expects that an MHMM should work better than a UHMM simply because MHMM fully utilizes the data. However, this is just an intuitive conjecture, which has to be systematically tested.

Basic Problems. There are three basic problems for HMMs.[9] First, "evaluation": Given the observation sequence O and a model λ, calculate the probability of the observation sequence given the model $P(O|\lambda)$. This problem is efficiently solved by the **forward−backward algorithm** with time complexity $\mathcal{O}(N^2T)$. Second, "decoding": Given the observation sequence O and a model λ, choose the optimal state sequence q which best explains (fits) the observation sequence O. This is solved by the **Viterbi algorithm**, which finds the state sequence q

maximizing $P(q|O, \lambda)$: the probability of the state sequence given the model and the observation sequence. The optimal state sequence q is called the **Viterbi path**. Third, "learning": Given the observation sequence O, adjust the model parameters λ to maximize the likelihood function $P(O|\lambda)$. This is solved by the **Baum−Welch algorithm**, which iteratively reestimates the hidden parameters by their expected values $\tilde{\lambda}$ until some limiting point is reached. In practice, this is implemented by setting a stop criterion, i.e., the log-likelihood difference $\Delta \log P \equiv \log P(O|\tilde{\lambda}) - \log P(O|\lambda)$ is smaller than a constant, e.g., 10^{-4}. Though this procedure only leads to local maxima of the likelihood function, it has been found that as long as reasonable initial guesses are made with respect to the parameters the algorithm converges to the true values.[9]

Those algorithms mentioned above have been well described in the literature. Here, we want to emphasize the reestimation formulas used in the Baum−Welch algorithm. Those reestimation formulas can be derived directly by maximizing Baum's auxiliary function:

$$Q(\lambda, \tilde{\lambda}) \equiv \sum_{q \in \mathcal{Q}} P(O, q|\lambda) \log P(O, q|\tilde{\lambda}) \tag{8}$$

(In case there are some constraints, the standard constrained optimization technique, i.e., the Lagrange multiplier method, can be used.) Here, $\lambda = (\pi, A, B)$ are our initial (or previous) estimates of the parameters. $\tilde{\lambda} = (\tilde{\pi}, \tilde{A}, \tilde{B})$ are being optimized. $\mathcal{Q}$ is the space of all state sequences of length T. An elegant proof using Jensen's inequality shows that, if $Q(\lambda, \tilde{\lambda}) > Q(\lambda, \lambda)$, then $P(O|\tilde{\lambda}) > P(O|\lambda)$; i.e., the maximization of $Q(\lambda, \tilde{\lambda})$ leads to increased likelihood.[22]

For a general HMM, the reestimation formulas of the model parameters $\tilde{\pi}$ and $\tilde{A}$ can be easily derived:[9,22,23]

$$\tilde{\pi}_i = \gamma_1(i) \tag{9}$$

$$\tilde{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{10}$$

Here,

$$\begin{aligned} \xi_t(i,j) &= P(q_t = i, q_{t+1} = j|O, \lambda) \\ &= \frac{P(q_t = i, q_{t+1} = j, O|\lambda)}{P(O|\lambda)} \end{aligned} \tag{11}$$

is the probability of being in state i at time t and state j at time $t + 1$, which can be easily calculated using the forward−backward algorithm.[9] From $\xi_t(i, j)$, we can easily calculate

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) = P(q_t = i|O, \lambda) = \frac{P(q_t = i, O|\lambda)}{P(O|\lambda)} \tag{12}$$

which is the probability of being in state i at time t, given the observation sequence and the model.

Obviously, the reestimation formulas of the model parameter $\tilde{B}$ depend on the OPD of the HMM. For MPHMM with OPD given by eq 4, we have

$$\tilde{\mu}_{i,k} = \frac{\sum_{t=1}^{T} o_{t,k} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)} \tag{13}$$

See Appendix section 1.2 for the derivation.

For MGmHMM with OPD given by eq 5, we have

$$\tilde{C}_{ik} = \frac{\sum_{t=1}^{T} \gamma_t(i,m)}{\sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_t(i,m)} \tag{14}$$

$$\tilde{\boldsymbol{\mu}}_{im} = \frac{\sum_{t=1}^{T} \gamma_t(i,m) \mathbf{O}_t}{\sum_{t=1}^{T} \gamma_t(i,m)} \tag{15}$$

$$\tilde{\boldsymbol{\Sigma}}_{im} = \frac{\sum_{t=1}^{T} \gamma_t(i,m)(\mathbf{O}_t - \tilde{\mu}_{im})(\mathbf{O}_t - \tilde{\mu}_{im})'}{\sum_{t=1}^{T} \gamma_t(i,m)} \tag{16}$$

where

$$\gamma_t(i,m) = \gamma_t(i) \frac{C_{im} \mathcal{N}(\mathbf{O}_t, \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im})}{\sum_{m=1}^{M} C_{im} \mathcal{N}(\mathbf{O}_t, \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im})} \tag{17}$$

is the probability of being in state i at time t with the mth mixture component accounting for $\mathbf{O}_t$.

For UBHMM with OPD given by eq 6, it can be shown that the reestimation formulas for the model parameter B are implicitly given by

$$\tilde{f}(\tilde{\alpha}_i, \tilde{\beta}_i) = \frac{\sum_{t=1}^{T} \log O_t \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)} \tag{18}$$

$$\tilde{g}(\tilde{\alpha}_i, \tilde{\beta}_i) = \frac{\sum_{t=1}^{T} \log(1 - O_t) \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)} \tag{19}$$

See Appendix section 1.1 for the derivation and the definitions of $f(x, y)$ and $g(x, y)$.

For UGHMM with OPD given by eq 7, one simply has

$$\tilde{\mu}_i = \frac{\sum_{t=1}^{T} \gamma_t(i) O_t}{\sum_{t=1}^{T} \gamma_t(i)} \tag{20}$$

$$\tilde{\sigma}_i^2 = \frac{\sum_{t=1}^{T} \gamma_t(i)(O_t - \tilde{\mu}_i)^2}{\sum_{t=1}^{T} \gamma_t(i)} \tag{21}$$

It is often the case that we have multiple observation sequences (traces). For example, smFRET experiments routinely generate about 20 traces. In the presence of multiple independent traces from the same system, the modification of the reestimation procedure is straightforward. The formal derivation is shown in Appendix section 2.

Application to Synthetic Data. ***Determining the Number of Hidden States.*** In analyzing smFRET data, the number of underlying FRET states is the first quantity we want to extract. In a previous work, this was done by plotting a two-dimensional pseudohistogram graph from compiling hundreds of fit traces.[8] This graph is often called the transition density plot (TDP), which is obtained by summing up Gaussian functions for every transition found, with centers corresponding to the initial and final FRET value for the transition. The number of underlying FRET states is determined by counting the peaks in the TDP. For a general N-state system which is ergodic or fully connected, i.e., every state of the system could be reached from every other state in one transition, $N(N-1)$ peaks should appear in the TDP.

Here, we consider a much simpler procedure to determine the number of underlying FRET states, which is suitable even in the presence of only one trace. This procedure is based on information criteria which have been used in the HMM analysis of smFRET data[5,24] and molecular motor data.[13] The information criteria are tools for model selection. The goal is to best explain the data with a minimum of free parameters, i.e., select the most parsimonious model. We know that adding parameters to any model will always improve the fit but not necessarily the statistical significance. All of the information criteria are essentially penalized log-likelihood scores, which take into account the trade-off between bias and variance in model selection.[25] Here, we use the three most popular ones: Akaike information criterion (AIC), Bayesian information criterion (BIC), and Hannan-Quinn information criterion (HIC) defined as

$$\mathrm{AIC} = -2 \log L + 2k \tag{22}$$

$$\mathrm{BIC} = -2 \log L + k \log n \tag{23}$$

$$\mathrm{HIC} = -2 \log L + 2k \log(\log n) \tag{24}$$

with L being the maximized value of the likelihood function for the estimated model, k the number of free parameters, and n the number of data points. In our studies, $\log L = \max[\log P(O|\lambda)]$ which can be approximately replaced by $\log P(O|\lambda_{\mathrm{BW}})$, i.e., the (local) maximum log-likelihood found by the Baum−Welch algorithm. It is easy to count that $k = N^2 + 2N - 1$ for MPHMM, UGHMM, and UBHMM and $k = N^2 + (6M-1)N - 1$ for MGmHMM with M mixture components. In both cases, $n = T$, i.e., the length of the trace. Given a data set, we can vary N within a reasonable range and perform the Baum−Welch reestimation procedure for each N. Then, all of the competing models with different N can be ranked according to their information criteria, with the one having the lowest AIC (or BIC, HIC) being the best.

For example, let us consider a system with five states. Its model parameters (π, A, B) are given by

(1) Initial probability $\pi = (0.2, 0.2, 0.2, 0.2, 0.2)$

(2) Transition matrix

$$A = \begin{pmatrix} 0.9 & 0.025 & 0.025 & 0.025 & 0.025 \\ 0.025 & 0.9 & 0.025 & 0.025 & 0.025 \\ 0.025 & 0.025 & 0.9 & 0.025 & 0.025 \\ 0.025 & 0.025 & 0.025 & 0.9 & 0.025 \\ 0.025 & 0.025 & 0.025 & 0.025 & 0.9 \end{pmatrix}$$

(3) Observation probability distribution $b_i(\mathbf{O})$ (two-dimensional Poisson distribution) with mean vectors $\boldsymbol{\mu}_1 = (10, 90)$, $\boldsymbol{\mu}_2 = (30, 70)$, $\boldsymbol{\mu}_3 = (50, 50)$, $\boldsymbol{\mu}_4 = (70, 30)$, and $\boldsymbol{\mu}_5 = (90, 10)$. Correspondingly, $\mathrm{FRET}_1 = 0.1$, $\mathrm{FRET}_2 = 0.3$, $\mathrm{FRET}_3 = 0.5$, $\mathrm{FRET}_4 = 0.7$, and $\mathrm{FRET}_5 = 0.9$.

Figure 1 shows the state and observation sequence of length $T = 500$ generated from the above model with five states. Numerically, the observations (I_A, I_D) are generated from two-dimensional Poisson distributions randomly. And we assume the absence of all other fluctuations due to variable background photon levels, laser intensity fluctuations, and changes in collection efficiency, etc. For example, if the system stays in the fourth state with $\boldsymbol{\mu}_4 = (70, 30)$ for $T = 500$ time steps, the observation sequence and corresponding histograms are shown in Figure 2. The two-channel signal (I_A, I_D) is of course described by the two-dimensional Poisson distribution where it is generated from. More interestingly, it is found that the FRET distribution can be approximated by both Beta and Gaussian distributions very well. Here, the parameters of the Beta distribution $b(O; \alpha, \beta)$ are set as $\alpha = \langle I_A \rangle$ and $\beta = \langle I_D \rangle$, with $\langle I_A \rangle$ (or $\langle I_D \rangle$) being the mean value of the I_A (or I_D) sequence. For the Gaussian distribution, mean is just the mean value of the FRET trajectory ($\langle \mathrm{FRET} \rangle$) and standard deviation is just the standard deviation of the FRET trajectory ($\sigma(\mathrm{FRET})$).

For this given observation sequence, with reasonable initial estimation and varying N, UGHMM, UBHMM, and MPHMM analyses are performed. The corresponding Viterbi paths at $N = 5$ are plotted on top of the hidden state sequence and noisy FRET data. One sees that all three Viterbi paths match the hidden state sequence very well.

Figure 3 shows the information criteria as functions of N in MPHMM, UGHMM, and UBHMM analysis for this system with different trace lengths $T = 100$ and 800. The maximum log-likelihood found by the Baum−Welch algorithm ($\log P(O|\lambda_{\mathrm{BW}})$) is also plotted. We find that, even for a short trace with $T = 100$, the correct number of hidden states ($N = 5$) can be obtained by observing where the information criteria reach their minima. In particular, we find that the BIC often shows a relatively more significant minimum, comparing to AIC and HIC. At first sight, the maximum log-likelihood $\log P(O|\lambda_{\mathrm{BW}})$ itself does not serve as a good criterion in determining N. However, we find that its slope, i.e., $\partial \log P(O|\lambda_{\mathrm{BW}})/$

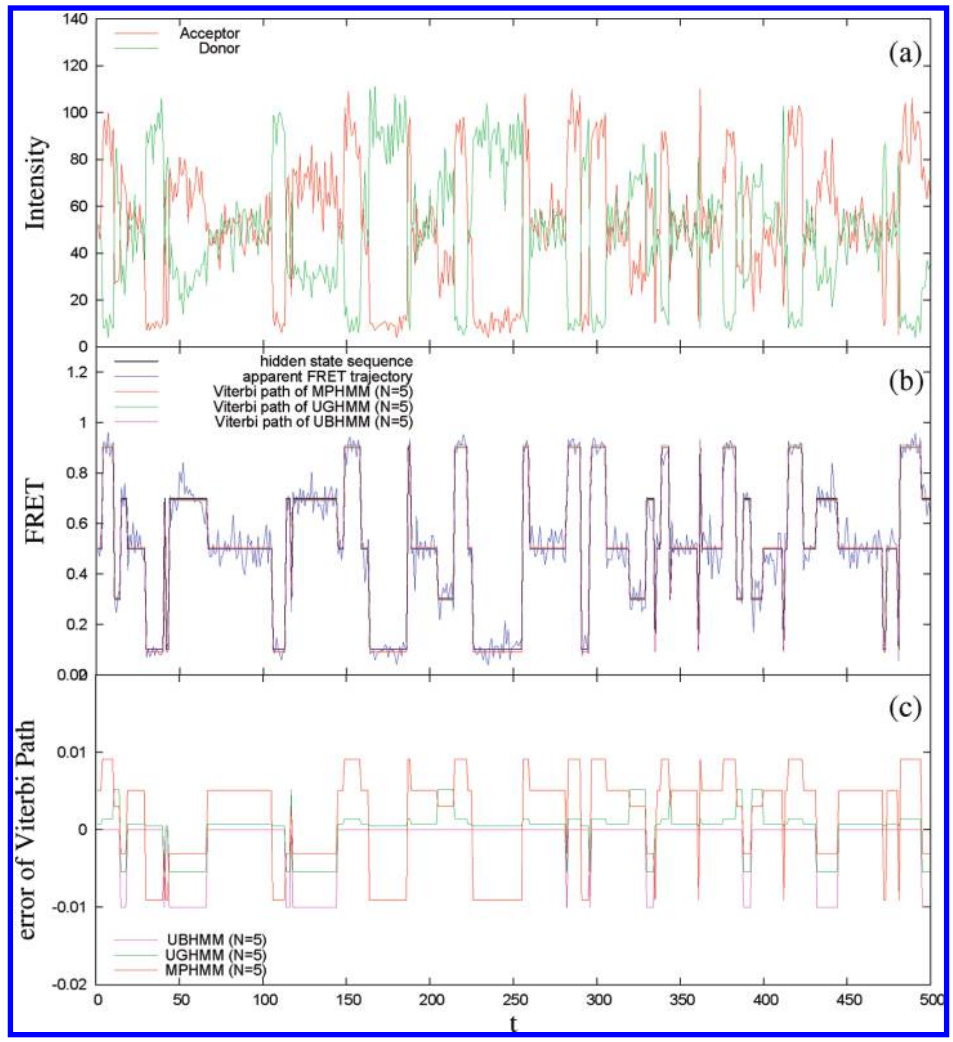

Figure 1. Synthetic observation and state sequences of the tested five-state system. (a) The two-channel signal (acceptor and donor), i.e., the shot noise, is generated from two-dimensional Poisson distributions of the five-state system. (b) The FRET trajectory is calculated from the acceptor and donor signals using eq 1. The "hidden" state sequence is buried under the noisy FRET trajectory. Using MPHMM, UGHMM, and UBHMM, we can extract the model parameter and calculate the Viterbi paths which fit the original state sequence very well. (c) The deviations of the calculated Viterbi paths from the "hidden" state sequence.

∂N changes dramatically at $N = 5$, which suggests that it can also be used to determine N. Therefore, we have two different types of measures to determine N consistently. Moreover, we find that, in UGHMM, those measures sometimes display large fluctuations as varying N. (For example, see Figure 3b, $T = 100$, around $N = 6$.) We have tested for longer traces with T from 200 up to 10 000 and found that those fluctuations appear often in UGHMM, while in UBHMM and MPHMM, all measures show a well-defined trend for all of the trace lengths we have tested. Detailed analysis finds that this is due to the fact that UGHMM gets trapped in local maxima of the likelihood function more often than UBHMM and MPHMM, an observation which will be further studied in the next section. Considering this, we conclude that, for the synthetic data generated from two-dimensional Poisson distributions, UBHMM and MPHMM are better than UGHMM in determining the number of hidden states.

Reliability. To compare the performance of MPHMM, UGHMM, and UBHMM further, we test their reliability for a simple two-state system. Simulated acceptor (and donor) time traces are generated by adding Poisson noise to a series of idealized values. FRET trajectories are then calculated according to eq 1. The standard parameters (π, A, B) are as follows:

(1) Initial probability

$$\pi = (0.5, 0.5) \tag{25}$$

(2) Transition matrix

$$A = \begin{pmatrix} 0.95 & 0.05 \\ 0.02 & 0.98 \end{pmatrix} \tag{26}$$

(3) Observation probability distribution $b_i(\mathbf{O})$ (two-dimensional Poisson distribution) with mean vectors

$$\boldsymbol{\mu}_1 = (300, 700), \qquad \boldsymbol{\mu}_2 = (700, 300) \tag{27}$$

Correspondingly, $\mathrm{FRET}_1 = 0.3$ and $\mathrm{FRET}_2 = 0.7$.

We vary different parameters (or their combinations) but keep the others constant to test different impacts on the algorithm performance. For each given parameter set, 100 traces are generated and analyzed using UGHMM (eq 7), UBHMM (eq 6), and MPHMM (eq 4). The standard trace length is $T = 10\,000$.

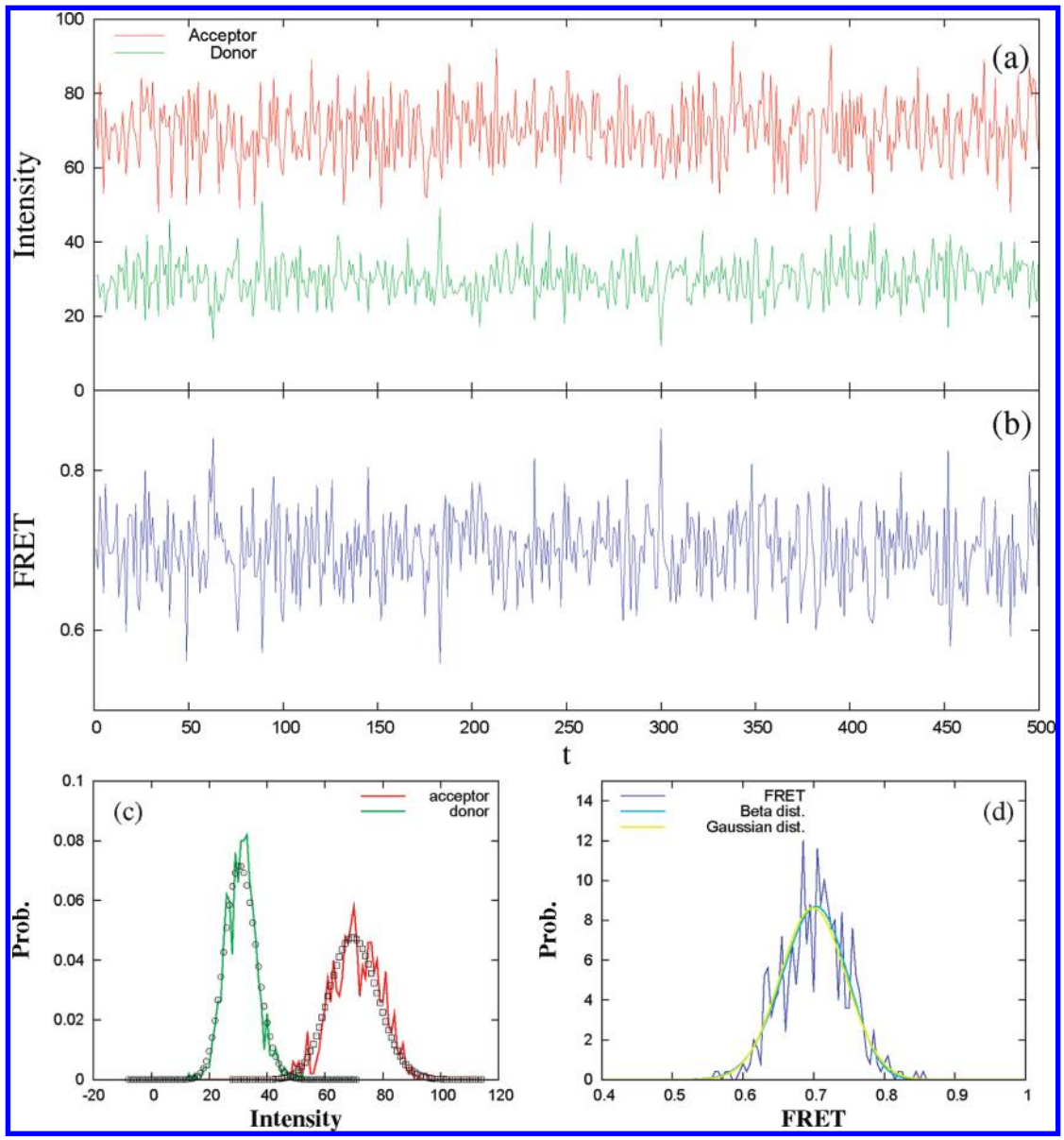

Figure 2. Synthetic observation sequences of the fourth state of the tested five-state system. (a) The two-channel signal (acceptor and donor), i.e., the shot noise, is generated from a two-dimensional Poisson distribution according to the state parameter $\boldsymbol{\mu}_4 = (70, 30)$. (b) The FRET trajectory is calculated from the acceptor and donor signals using eq 1. (c) The (normalized) histograms of I_A and I_D. Symbols represent components of the two-dimensional Poisson distribution. (d) The (normalized) histogram of FRET. Both Beta and Gaussian distributions fit the FRET distribution very well.

To be fair, the UGHMM, UBHMM, and MPHMM reestimation procedures (based on the Baum−Welch algorithm) start from the **same** initial random guesses of the transition matrix A and initial state distribution π. And the initial estimates of the observation probability distribution B for UGHMM, UBHMM, and MPHMM are related to each other according to eq 1. Also, the UGHMM, UBHMM, and MPHMM reestimation procedures are stopped according to the same criterion, i.e., $\Delta \log P = 10^{-4}$.

As pointed out in ref 9, either random (subject to stochastic constraints) or uniform initial estimates of π and A parameters would be adequate for giving meaningful reestimates of those parameters in almost all cases. However, for B parameters, good initial estimates are essential.[15] For simplicity, in our tests, the initial estimates for the B parameters are obtained by analyzing the histograms of signals. For example, in analyzing the FRET trajectories of the two-state systems using UGHMM, the initial guesses of the mean FRET values for the two states are assigned to be $\langle \text{FRET} \rangle \pm \sigma_{\text{FRET}}$. Here, $\langle \text{FRET} \rangle$ and σ_{FRET} are the mean and standard deviation of the FRET distribution, respectively. Similar initial estimates are made for I_A and I_D in MPHMM.

Similarly to what was done in a previous work,[8] the successes of MPHMM, UGHMM, and UBHMM are quantified in two different reliability measures. (1) P_f: the fraction of the 100 traces that returned the true FRET values: $\text{FRET}_1 \pm 0.05$ and $\text{FRET}_2 \pm 0.05$. (2) $|\log(k/k^*)|$: the systematic error in the transition rate, with $k = a_{12}$ being the deduced transition rate from state 1 to state 2 and $k^* = a_{12}^*$ the true input value. Obviously, when the input model parameters are perfectly recovered, $P_f = 1$ and $|\log(k/k^*)| = 0$. These two measures are shown in Figure 4 with solid and open symbols, respectively.

Figure 4a shows the effect of changing $\Delta\text{FRET} \equiv \text{FRET}_1 - \text{FRET}_2$, i.e., the spacing between the two FRET states. It is clearly seen that both the MPHMM and UBHMM respond very well with $P_f \sim 1$ and $|\log(k/k^*)| \sim 0$ for the whole ΔFRET range [0.02, 0.5] we have tested. However, for UGHMM, its

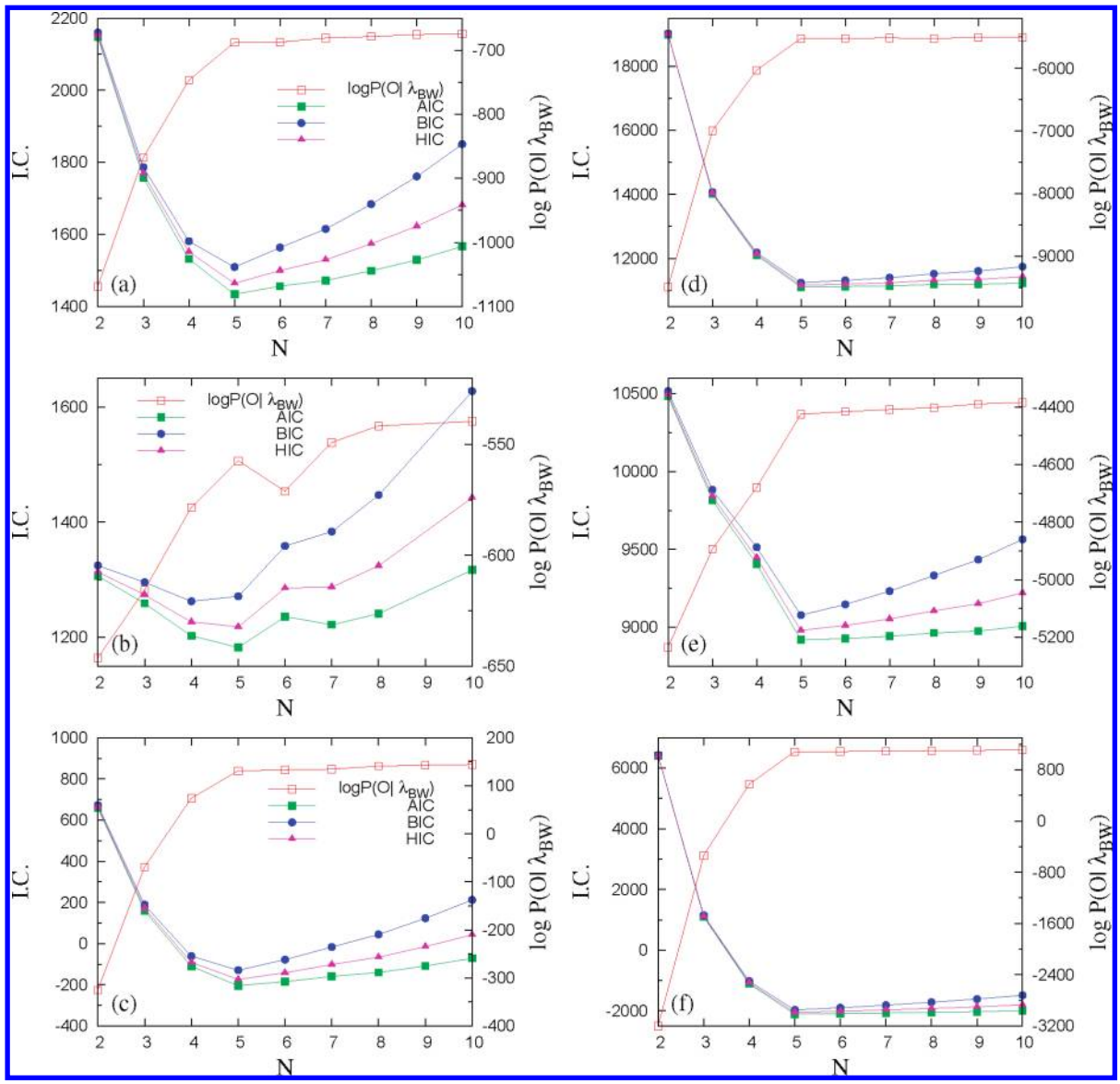

Figure 3. Information criteria (solid symbols) and maximized log-likelihood (open squares) as functions of N (number of hidden states) in MPHMM (a,d), UGHMM (b,e), and UBHMM (c,f) analysis for a five-state system. The analysis is performed on a single trace with length $T = 100$ (a−c) and 800 (d−f).

reliability generally decreases with decreasing ΔFRET. Note that, for ΔFRET ≤ 0.05, P_f for UGHMM goes to 1, which naively suggests that it works well in this regime. However, this is just due to the fact that in defining P_f the size of the error bar is set to be 0.05. In fact, the reliability of UGHMM becomes worse for ΔFRET ≤ 0.05, which can be clearly seen from the finite $|\log(k/k^*)|$ values in this regime.

Figure 4b shows the effect of FRET noise level, parametrized by the average FRET peak width δ. For the simple two-state system, we have $\delta \equiv \sigma_1 + \sigma_2$ with σ_i ($i = 1, 2$) being the standard deviation of the Gaussian FRET distribution for state i (see eq 7). Since the simulated FRET trajectories are calculated from the acceptor and donor time traces, σ_i and consequently δ cannot be tuned explicitly in our simulation. However, they can be read from the UGHMM output. In our tests, we find that $\sigma_1 \approx \sigma_2$ in all cases and the FRET peak width δ is highly correlated with the total intensity (I^{tot}) of acceptor and donor signals in such a way: $\delta \approx 0.92/(I^{tot})^{1/2}$. Therefore, we can vary I^{tot} but keep other parameters fixed to study the impact of FRET noise level. In our tests, we tune I^{tot} from 1000 down to 10. We find that MPHMM responds very well to increased noise level (or decreased I^{tot}), with slightly decreased reliability only when $\delta > 0.2$ (corresponding to $I^{tot} < 20$). (In a previous work,[8] it was found that the univariate HMM breaks down when $\delta > 0.4$. We notice that δ was explicitly tunable there, while, in our tests, it can be tuned only by varying I^{tot}. Nevertheless, our result is consistent with theirs.) The UBHMM also responds fairly well to increased noise level but with decreased reliability for $\delta > 0.15$ (corresponding to $I^{tot} < 40$). On the other hand, the UGHMM does not respond well for all of the noise levels we have tested.

Figure 4c shows the effect of state lifetime τ_1, i.e., the mean dwell time of state 1. It is easy to derive that $\tau_1 = (1 - a_{11})^{-1}$. Therefore, we can tune τ_1 by varying a_{11}. However, to study the impact of τ_1 only, we should consider traces with (almost) the same number of transitions (N_{tr}). In our tests, we find that $N_{tr} \approx 0.6Ta_{12}$. This suggests that, to keep $N_{tr} =$ const, we have to vary the trace length T accordingly as we vary a_{12}. In our tests, we set $N_{tr} = 90$, which is large enough to see all of the state transitions. We find that both UBHMM and MPHMM are reliable for the whole range of τ_1 we have tested. For UGHMM,

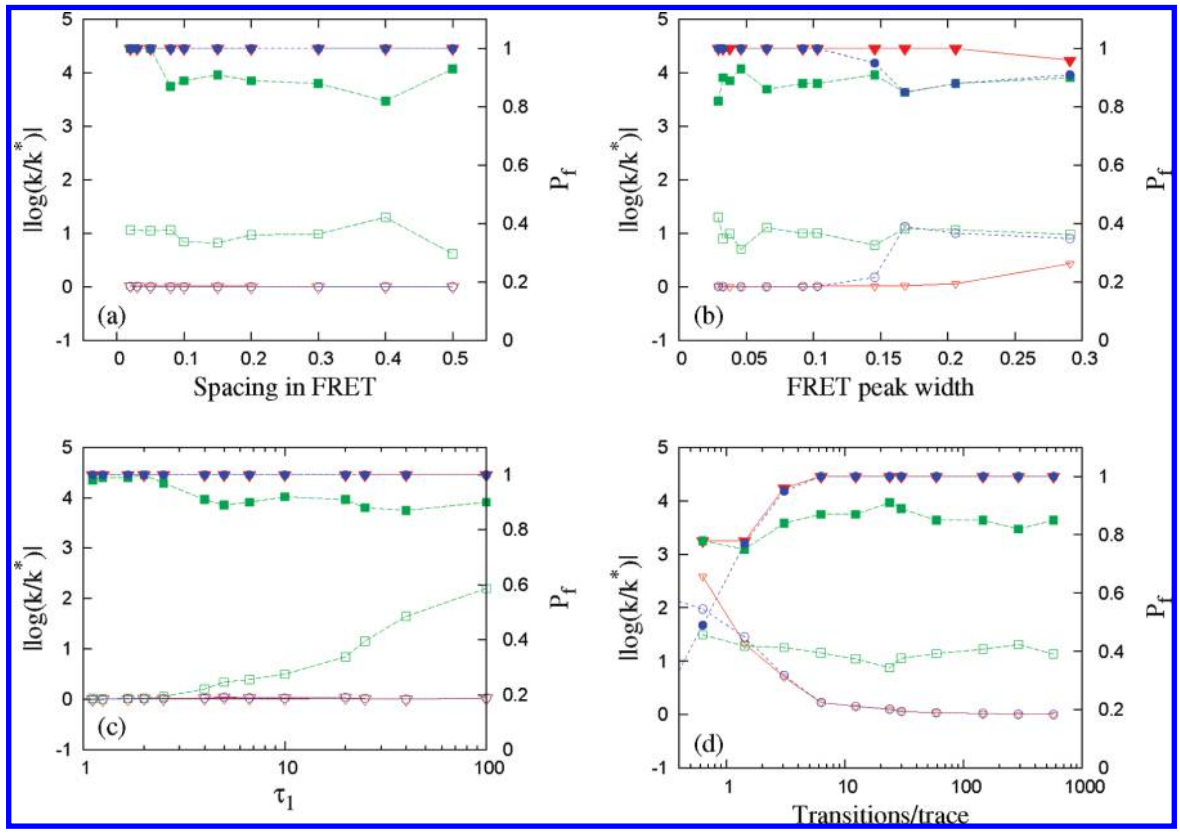

Figure 4. Algorithm responses to changes in model and trace parameters for MPHMM (red triangles), UGHMM (green squares), and UBHMM (blue cycles) with 100 traces. Open symbols: systematic error $|\log(k/k^*)| = |\log(a_{12}/a_{12}^*)|$. Solid symbols: the probability (the fraction of the 100 traces) that obtained FRET values match the true values of both states. (a) Varying spacing between the two FRET states. (b) Varying FRET peak width (δ) by simply changing the total intensity ($\langle I_A \rangle + \langle I_D \rangle$). (c) Varying the dwell time of FRET state 1 by tuning a_{12}^* (also keeping the number of state transitions constant by changing the length of trance accordingly). (d) Varying the number of state transitions by simply changing the length of trace.

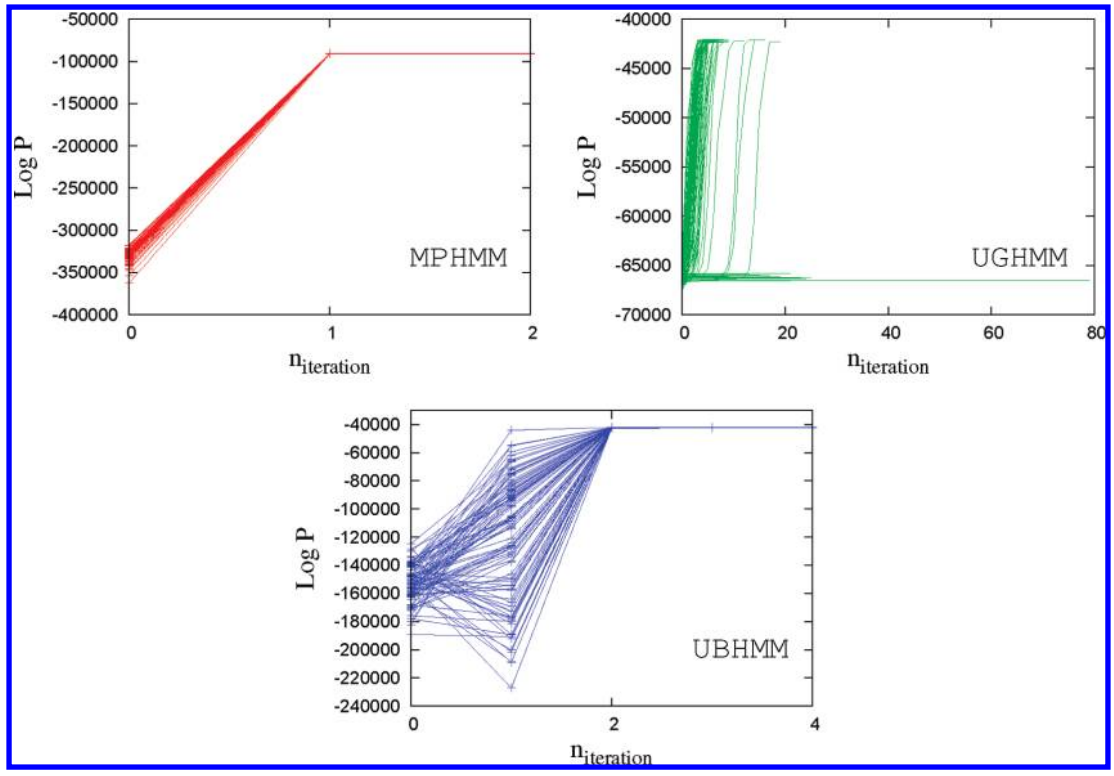

Figure 5. The increasing log-likelihood during the Baum−Welch iterations. Those analyses are conducted on the traces generated with the standard input model parameters (eqs 25 and 27). (left) With MPHMM, all 100 trace analyses return the true model parameters after only two iterations ($n_{\text{iteration}} = 2$). (middle) With UGHMM, a significant fraction of the 100 trace analyses get trapped in local maxima. Generally, $n_{\text{iteration}}$ used in UGHMM is much larger than that used in MPHMM. (right) With UBHMM, all 100 trace analyses return the true model parameters after up to four iterations ($n_{\text{iteration}} \leq 4$). Note that there are some traces showing decreasing log(P), which is due to the fact that there is no closed form of reestimation formula for UBHMM. For details, see Appendix section 1.1.

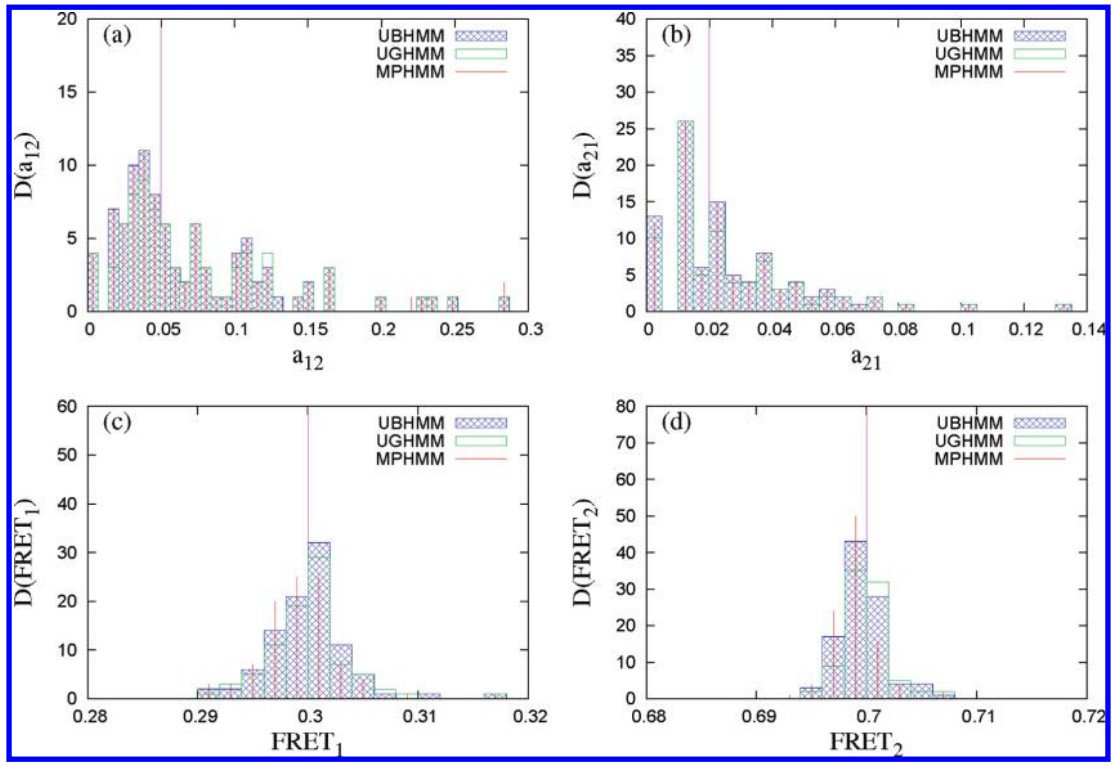

Figure 6. Histograms of a_{12}, a_{21}, $FRET_1$, and $FRET_2$ obtained from 100 traces with length $T = 100$, by using MPHMM (red pulses), UGHMM (green boxes), and UBHMM (blue boxes). Number of states: $N = 2$. The true (input) values of those model parameters are plotted in purple lines.

TABLE 1: The Difference between the Results of Two Methods in Analyzing Multiple Traces[a]

	result of method I			result of method II		
input	MPHMM	UGHMM	UBHMM	MPHMM	UGHMM	UBHMM
$a_{12} = 0.05$	0.06841 (36.8%)	0.07096 (41.9%)	0.06661 (33.2%)	0.05262 (5.2%)	0.05262 (5.2%)	0.05262 (5.2%)
$a_{21} = 0.02$	0.02502 (25.1%)	0.03070 (53.5%)	0.02503 (25.1%)	0.02099 (4.95%)	0.02099 (4.95%)	0.02099 (4.95%)
$FRET_1 = 0.3$	0.3147 (4.9%)	0.3216 (7.2%)	0.3158 (5.3%)	0.3003 (0.1%)	0.3001 (0.03%)	0.3003 (0.1%)
$FRET_2 = 0.7$	0.6993 (0.1%)	0.6936 (0.9%)	0.6996 (0.06%)	0.6997 (0.04%)	0.6999 (0.01%)	0.6997 (0.04%)

[a] Method I: average over individual trace analysis. Method II: analysis of multiple traces simultaneously. For each method, we also compare the performance of using MPHMM, UGHMM, and UBHMM. The tested system is a simple two-state system with model parameters given by eqs 25−27. Trace length $T = 100$. Number of traces $M = 100$. Data shown in parentheses present the percentage errors.

it responds very well for $\tau_1 \leq 3$ only. Its reliability is significantly reduced for $\tau_1 > 10$.

Figure 4d shows the effect of the number of transitions N_{tr}, which is tuned by simply varying T while keeping all other parameters fixed. We see that, as long as $N_{tr} > 5$, both UBHMM and MPHMM respond very well. For $N_{tr} < 5$, the reliabilities of UBHMM and MPHMM decrease dramatically. This is reasonable because, without enough state transitions, any HMM would not work at all. We also find that for the N_{tr} range UBHMM and MPHMM work very well and UGHMM always shows reduced reliability.

All of the tests shown in Figure 4 suggest that, for the synthetic data generated from two-dimensional Poisson distributions, given the same initial guess and the same stop criterion, MPHMM and UBHMM outperform UGHMM in response to various model and trace parameters. Why is UGHMM not as reliable as UBHMM and MPHMM? Detailed analysis finds that, in contrast to UBHMM and MPHMM, UGHMM gets trapped more easily in local maxima of the likelihood function $P(O|\lambda)$ during the reestimation procedure. Generally, it is also found that the number of iterations ($n_{iteration}$) in UGHMM is much larger than that in UBHMM and MPHMM (see Figure 5).

Fortunately, in our tests of the simple two-state system, trapping in local maxima can be easily detected. We find that UGHMM sometimes obtains two almost identical FRET values for the two states with $\Delta FRET < 0.001$. This is apparently wrong, since in all our tests we have $\Delta FRET \geq 0.01$. In this case, trapping can easily be avoided by resetting the transition matrix A and restarting the reestimation procedure until the algorithm jumps away from the local maxima. Of course, this takes extra computing time. However, with this effort, it can be shown that the modified UGHMM is as reliable as UBHMM and MPHMM in response to different model and trace parameters (data not shown here). Generally, for complex systems with more states, trapping would be much more difficult to detect than in the simple two-state system. Therefore, it is hard to improve UGHMM to avoid local maxima. MPHMM or UBHMM would be the better choice.

Analyzing Multiple Traces. In previous work, multiple traces obtained in smFRET experiments were analyzed one by one and then averaged in the following way.[8] Transition rates were found to be distributed asymmetrically and argued to obey the log-normal distribution. To obtain the representative average values, one can average over their logarithms. The mean values are then converted back by exponentiation. For FRET values, a simple average is good enough. We call this method I.

We mentioned that, when multiple traces are independent from each other, we can analyze them simultaneously using either UGHMM, UBHMM, or MPHMM. In this way, no extra average is needed at all. We call this method II.

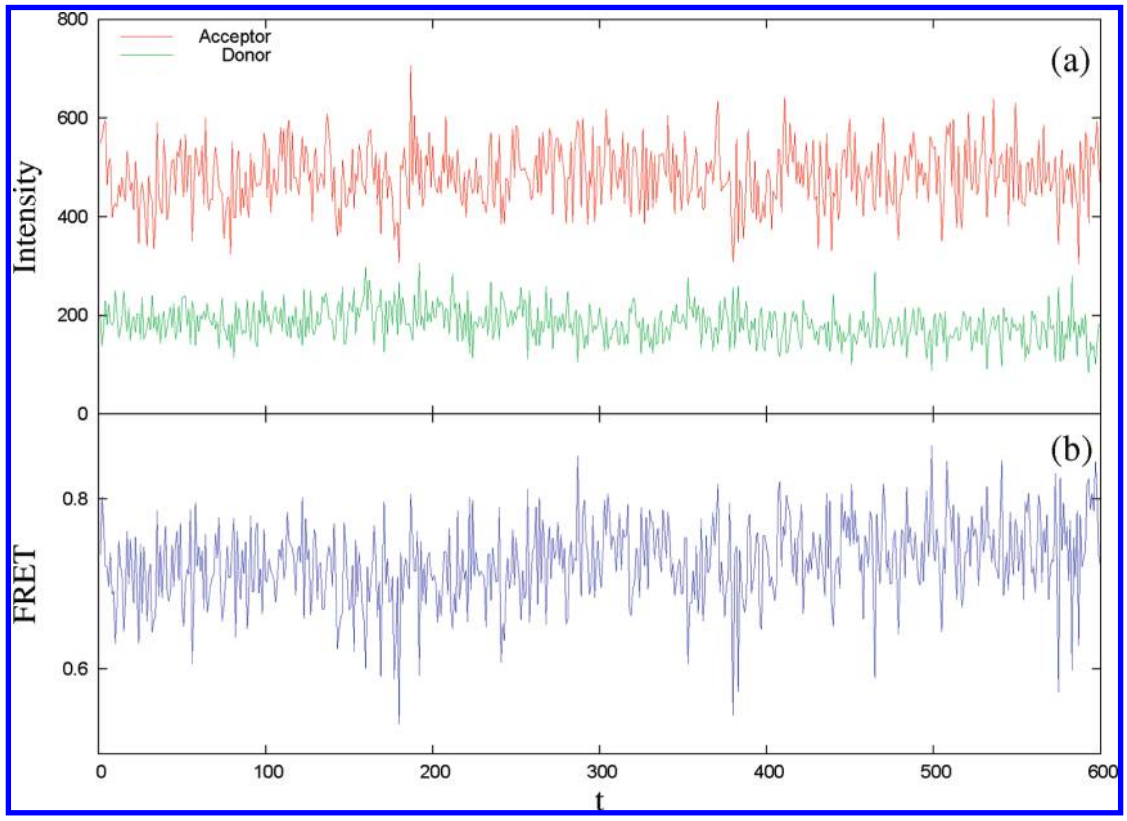

Figure 7. Experimental observation sequences of the bare DNA. (a) The two-channel signal (acceptor and donor) suffers from instrument noise, e.g., shot noise, spurious noise (also called clock induced charge), and amplification noise. (b) The FRET trajectory is calculated from the acceptor and donor signals using eq 1.

We compare the performance of those two methods for the simple two-state system used in the previous section. We generate 100 traces of length $T = 100$ randomly according to the standard model parameters. Then, we analyze those traces with the two methods mentioned above. The results are shown in Figure 6 and Table 1.

For method I, we get distributions of parameter values (see Figure 6). After averaging, the mean values deviate from the true values significantly, especially for the transition rates (see Table 1). Note that the transition rates are averaged in the special way as mentioned above. If we average them directly, the deviations from the true values are even larger. Moreover, we notice that, for method I, MPHMM gives the best result, while UGHMM gives the worst.

Comparing with method I, the parameter values obtained in method II are much closer to the true input values. Also, we notice that, in this method, MPHMM, UBHMM, and UGHMM give almost the same high-quality result. This can be well explained by noticing that, in analyzing $M = 100$ traces simultaneously, a much larger data set (actually 100 times larger than the single trace case) is taken into account. Simply because we have more data points, it is harder for UGHMM, UBHMM, and MPHMM to get trapped in local maxima. The difference in their performances will be negligible.

Application to Experimental Data. In the previous section, we assume that the shot noise is the only source of noise, so the synthetic two-channel signals are generated from the two-dimensional Poisson distribution (eq 4). (For example, the histograms of I_A, I_D, and FRET for a particular state in the tested five-state system are shown in Figure 2.) Considering this, the better performances of MPHMM and UBHMM than UGHMM are easily understood. However, in analyzing real experimental data, will MPHMM and UBHMM still be better than UGHMM? Shall we use more general MHMM such as MGHMM or MGmHMM?

Here, we apply all the different HMMs to the analysis of experimental data on RecA binding and dissociation. We have previously reported that binding and dissociation RecA proteins on a ssDNA can be observed with single monomer resolution using smFRET.[26] ssDNA is a highly flexible polymer, and when RecA proteins bind and form a filament, ssDNA is stretched; therefore, its end to end distance is increased. We labeled two positions on a ssDNA with donor and acceptor fluorophore so that we can detect three different RecA bound states at the 5′ filament end: no RecA bound state with highest FRET value ~0.8, one RecA bound state with FRET ~0.6, and two RecA bound state with FRET ~0.45.

Figure 7 shows the experimental observation sequences of the bare ssDNA measured by the EMCCD camera. The histograms of I_A and I_D are shown in Figure 8a. It is clearly seen Gaussians fit the histograms while Poissonians do not. The reason why Poissonians do not fit the histograms of intensities is that there is an additional scaling factor during the conversion from the actual photon count to the measured intensity processed by the EMCCD camera. This scaling process can actually introduce additional noise, such as spurious noise (also called clock induced charge) and amplification noise. For the same reason, the Beta distribution does not fit the histogram of the apparent FRET calculated from the donor and acceptor intensities using eq 1. However, we find that Gaussian fits the FRET distribution very well (see Figure 8b). Those results suggest that naive application of MPHMM or UBHMM on the experimental data measured by EMCCD will definitely fail due to the bad choice of OPDs. Instead, we expect MGHMM (or more generally MGmHMM) and UGHMM will work.

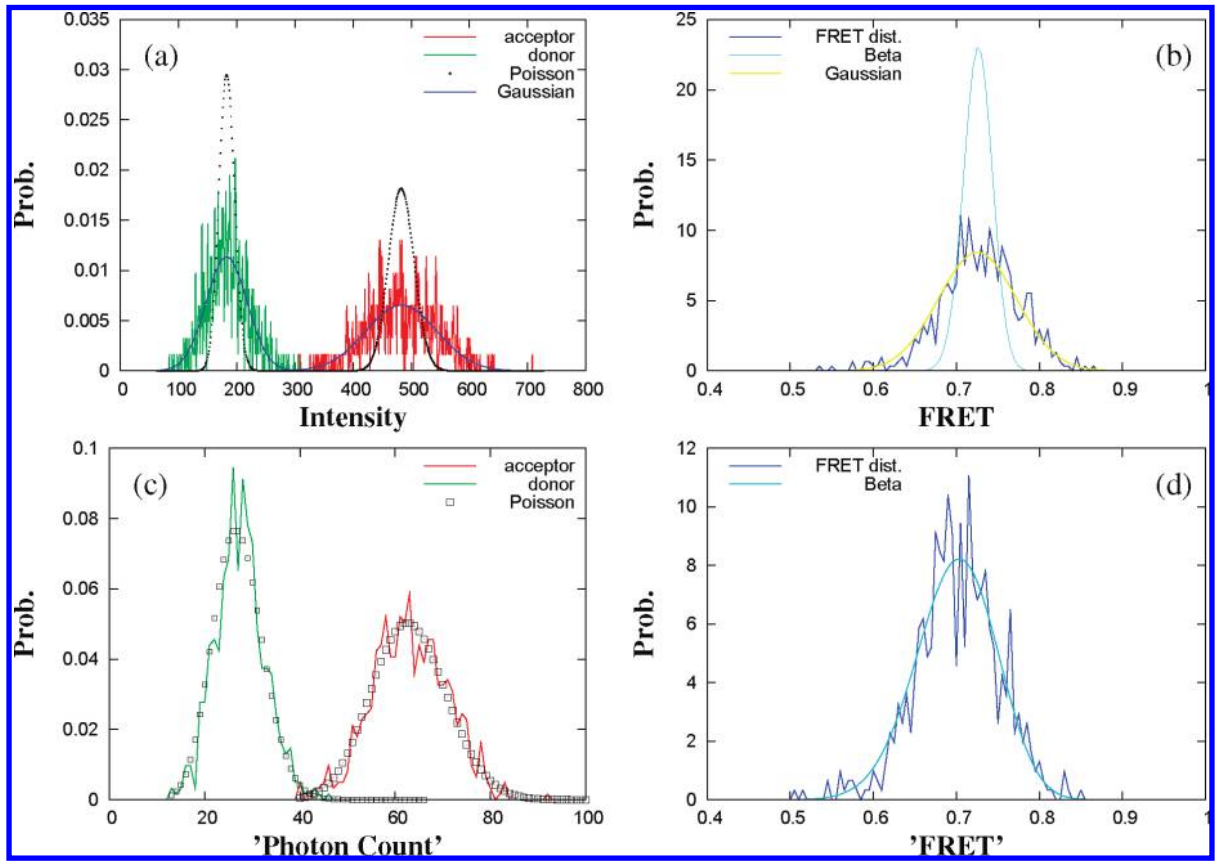

Figure 8. (a) The (normalized) histograms of I_A and I_D of the experimental observation sequence of the bare DNA. Symbols represent Poissonians, while blue lines represent Gaussians. Note that those histograms cannot be fit by Poissonians due to the scaling factor during the conversion from the actual photon count to the measured intensity processed by EMCCD. Instead, Gaussians fit the histograms of intensities very well. (b) The (normalized) histogram of FRET. Again, due to the scaling factor, Beta distribution does not fit the FRET distribution, while Gaussian does. (c) With the linear scaling assumption, one can get back the "photon count" data from the intensity data. Then, the histograms of the approximate "photon count" can be fit by Poissonians. (d) The (normalized) histograms of the "FRET" values calculated from the approximate "photon count" data can be fit by Beta distribution.

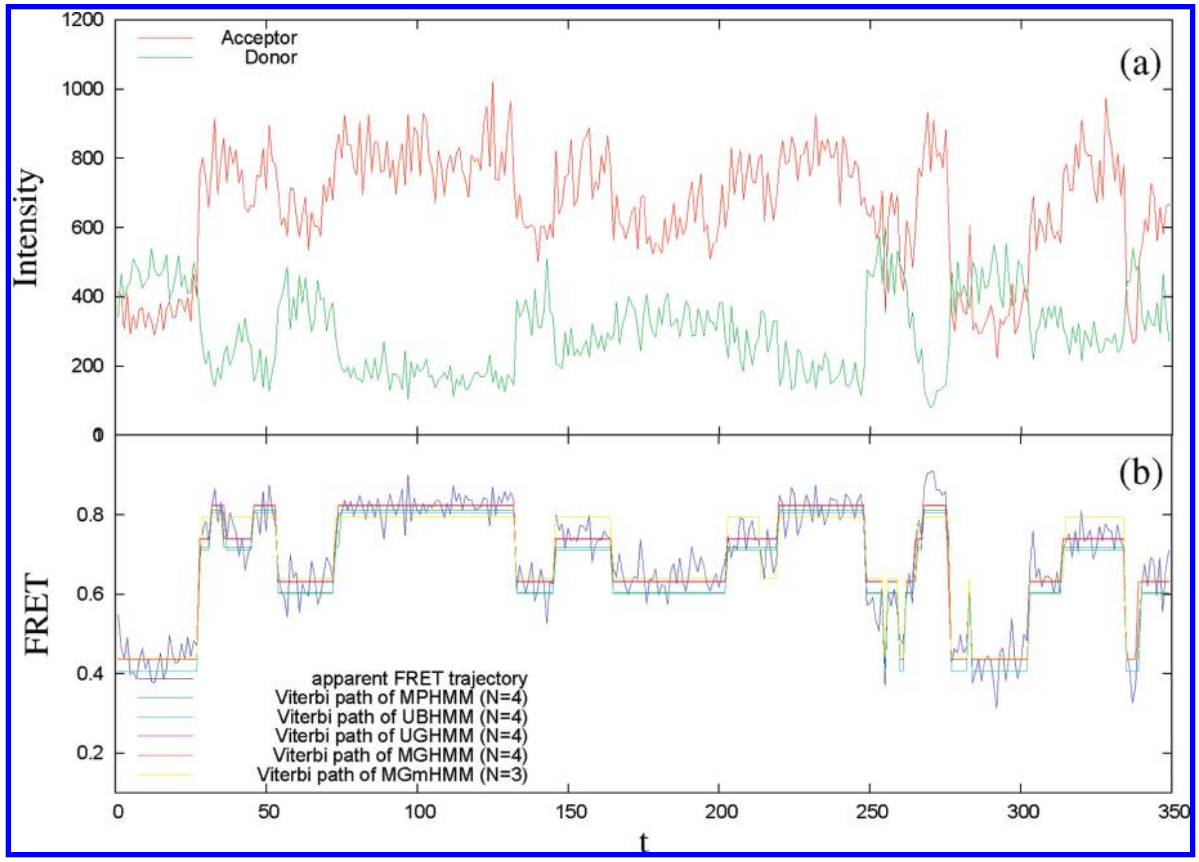

Figure 9. Experimental observation sequence of RecA binding. (a) Acceptor and donor channel signals. (b) The FRET trajectory is calculated from the acceptor and donor signals using eq 1.

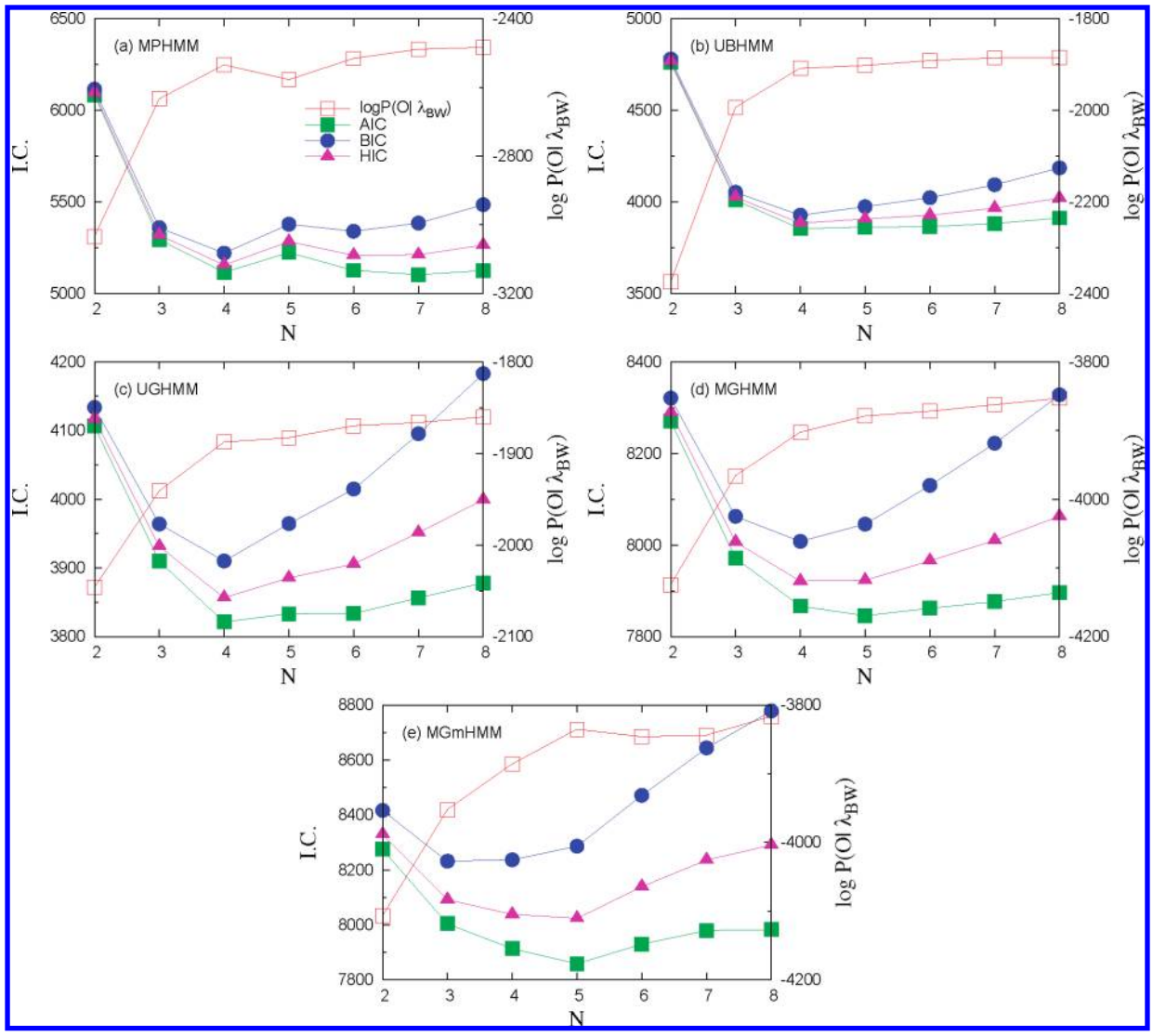

Figure 10. Information criteria (solid symbols) and maximized log-likelihood (open squares) as functions of N (number of hidden states) in MPHMM (a), UBHMM (b), UGHMM (c), MGHMM (d), and MGmHMM (e) analysis of the experimental observation sequence of RecA binding.

If one can get back the actual photon count from the measured intensity, then we expect MPHMM and UBHMM will work. However, the conversion from the intensity to photon count is highly nontrivial. Here, we make a simple linear approximation, i.e., assuming the scaling between the intensity I and the actual photon count C is linear: $f = I/C$. Then, the scaling factor f can be estimated easily. We assume the noise in photon count is just shot noise, so we have

$$\sigma_C = \sqrt{v_C} = \sqrt{\langle C \rangle} \tag{28}$$

Here, v represents variance and σ represents standard deviation. Due to the linear assumption, we have $\sigma_C = \sigma_I/f$ and $\langle C \rangle = \langle I \rangle/f$. Plugging those two equations into eq 28, one has $\sigma_I/f = (\langle I \rangle/f)^{1/2}$, which yields

$$f = \frac{\sigma_I^2}{\langle I \rangle} = \frac{v_I}{\langle I \rangle} \tag{29}$$

Note that the scaling factor could be different for different channels. For example, for sequences shown in Figure 7, we get a scaling factor of $f_A = 7.73$ for the acceptor channel and $f_D = 6.78$ for the donor channel. With f_A and f_D, we can then get back the "photon count" data from the intensity data. Here, quotes mean that the "photon count" we get is just an approximation of the actual photon count. To check the quality of the linear scaling approximation, we fit the histograms of the approximate photon count with Poissonians. As shown in Figure 8c, the fits are very good. Moreover, the histogram of "FRET" values calculated from "photon count" data can be fit by Beta distribution very well. Those self-consistent checks indicate that the linear scaling is a good approximation, at least from the point of view of the noise distribution.

On the basis of the analysis of noise distributions, we have the following choices in performing HMM on the EMCCD data: (1) Analyze the measured two-channel intensity data (I_A and I_D) either using MGHMM or more generally MGmHMM. (2) Analyze the calculated FRET trajectory alone using UGHMM. (3) Analyze the calculated two-channel "photon count" data (C_A and C_D) using MPHMM. (4) Analyze the calculated "FRET" trajectory alone using UBHMM.

Figure 9a shows a typical experimental observation sequence of RecA binding to ssDNA measured by EMCCD. The Viterbi paths found by the five different HMMs are shown in Figure

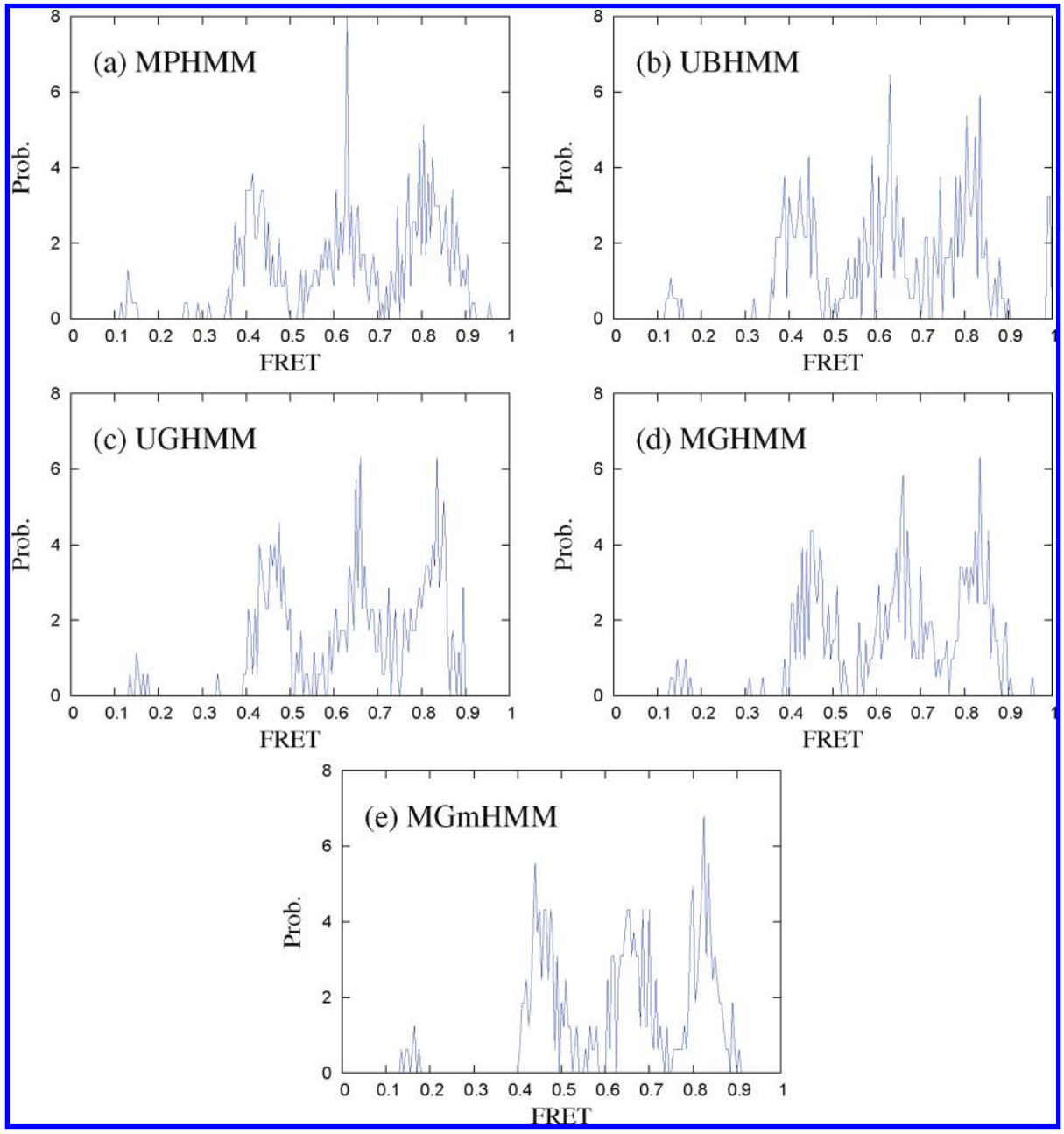

Figure 11. Normalized histograms of FRET values found from a total of 96 traces of RecA binding using different HMMs: MPHMM (a), UBHMM (b), UGHMM (c), MGHMM (d), and MGmHMM (e).

9b. Physically, there are three possible states in the system ($N = 3$), as mentioned above. To determine N from the observation sequence, we calculate the information criteria. The results are shown in Figure 10. We find that, for MGmHMM (with $M = 3$ components), the BIC gives the correct number of states, while AIC and HIC overestimate N. For all other HMMs, all three information criteria overestimate N. It is interesting to mention that BIC has been shown to provide a rigorous upper limit on the true number of states for an infinitely long sequence.[27] Study on the performance differences of different information criteria in model selection is interesting itself but is beyond the scope of the current work. Actually, additional information criteria such as peak localization error and chi-square probability-based goodness-of-fit have been studied in HMM analysis of short single-molecule intensity trajectories.[24]

To compare the performances of the five HMMs further, we plot the normalized histograms of FRET values found from the optimal Viterbi paths of the 96 traces of RecA binding using the five different HMMs. The results are shown in Figure 11. We find that FRET histograms of all five HMMs vividly show three main peaks corresponding to the three states in the system.

Transition density plots (TDPs) calculated from the total 96 traces of RecA binding using different HMMs are shown in Figure 12. Here, the x-axis (or y-axis) represents the FRET value before (or after) transition, respectively. We find that all five HMMs except UBHMM give clear state transitions at expected positions, represented by the peaks around (0.45, 0.6), (0.6, 0.45), (0.6, 0.8), and (0.8, 0.6) on the TDPs. We also notice that the spanning of the peaks in MGmHMM's TDP are slightly narrower than all other HMMs. Nevertheless, it is hard to draw a conclusion that MGmHMM is the best choice in analyzing the RecA binding data measured by EMCCD. What we learn is that, for smFRET data measured by EMCCD, the performances of MHMM and UHMM are very comparable and MGmHMM seems to slightly outperform other HMMs.

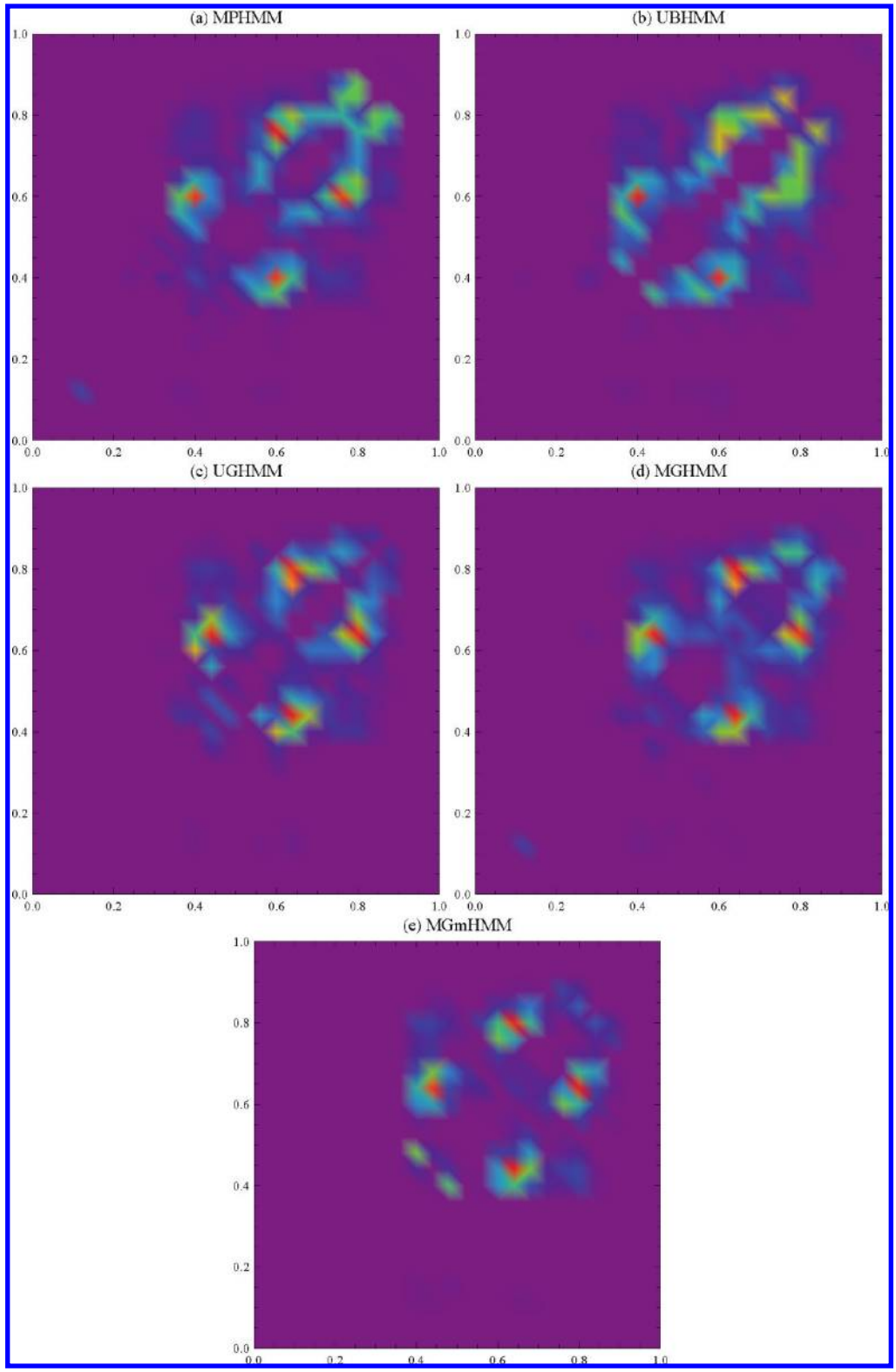

Figure 12. Transition density plots (TDPs) calculated from a total of 96 traces of RecA binding using different HMMs: MPHMM (a), UBHMM (b), UGHMM (c), MGHMM (d), and MGmHMM (e). Here, the *x*-axis (or *y*-axis) represents the FRET value before (or after) transition, respectively.

Summary

In sum, we compared two different types of HMM analysis algorithms for the time-binned smFRET data analysis: multivariate HMM and univariate HMM. For a multivariate HMM, at each conformational state, the two-channel signal (I_A, I_D) can be described by a two-dimensional distribution, e.g., Poisson, Gaussian, or a finite mixture of Gaussian distributions. The corresponding HMMs are denoted as MPHMM, MGHMM, or MGmHMM, respectively. For a univariate HMM, the calculated FRET trajectory is analyzed alone. At each conformational state, the signal (FRET) is described with a one-dimensional distribution, e.g., a Beta or Gaussian distribution. We denote the corresponding HMMs as UBHMM or UGHMM, respectively. We find that generally MHMM outperforms UHMM. For

synthetic data, with a two-channel signal generated from two-dimensional Poisson distributions, numerical tests in (1) determining number of hidden states and (2) reliability in response to varying model parameters show that MPHMM and UBHMM are much better than UGHMM. We also show that, in the case of multiple trajectories, analyzing them simultaneously gives much better results than averaging over individual analysis results. For experimental data, in particular the data measured by EMCCD, due to the complicated noise source, we find that generally UHMM and MHMM work comparably well. However, MGmHMM seems to be the best one, according to the transition density plot. Those studies clearly show that choosing correct observation distribution functions are very important in conducting HMM analysis for smFRET data.

Acknowledgment. The authors acknowledge the support of NSF Grant No. A3502 NSF 08-22613 PFC: Center for Physics of Living Cells. Y.L. thanks Sean A. McKinney, Yuji Ishitsuka, Ibrahim Cisse, Jiajie Diao, Reza Vafabakhsh, and Victor Caldas for valuable discussions. K.A.D. thanks Jonathan T. Uhl for helpful discussions.

Appendix

1. Reestimation Formulas for Single Observation Sequence

1.1. UBHMM. Here, we consider a UHMM with Beta observation probability distribution:

$$b_i(o;\alpha_i,\beta_i)=\frac{o^{\alpha_i-1}(1-o)^{\beta_i-1}}{B(\alpha_i,\beta_i)}$$

Given a particular state sequence q and model parameter $\tilde{\lambda}$, $P(O,q|\tilde{\lambda})$ can be easily written as

$$P(O,q|\tilde{\lambda})=\tilde{\pi}_{q_1}\tilde{b}_{q_1}(o_1)\prod_{t=1}^{T-1}\tilde{a}_{q_tq_{t+1}}\tilde{b}_{q_{t+1}}(o_{t+1}) \tag{30}$$

Then $Q(\lambda,\tilde{\lambda})$ becomes

$$Q(\lambda,\tilde{\lambda})=\sum_{q\in\mathcal{Q}}\log\tilde{\pi}_{q_1}P(O,q|\lambda)+\sum_{q\in\mathcal{Q}}\sum_{t=1}^{T-1}\log\tilde{a}_{q_tq_{t+1}}P(O,q|\lambda)+\sum_{q\in\mathcal{Q}}\sum_{t=1}^{T}\log\tilde{b}_{q_t}(O_t)P(O,q|\lambda) \tag{31}$$

We can optimize each term individually, leading to the reestimation formulas for π, A, and B, respectively. The optimizations of the first two terms are very simple and general for any HMM.[28]

Here, we show the optimization of the third term. Consider a Beta distribution for the UHMM (eq 6). The third term of eq 42 can be written as

$$\begin{aligned}\sum_{q\in\mathcal{Q}}\sum_{t=1}^{T}\log\tilde{b}_{q_t}(O_t)\cdot P(O,q|\lambda)&=\sum_{i=1}^{N}\sum_{t=1}^{T}\log\tilde{b}_i(O_t)\cdot P(O,q_t=i|\lambda)\\&=\sum_{i=1}^{N}\sum_{t=1}^{T}\log\left[\frac{O_t^{\tilde{\alpha}_i-1}(1-O_t)^{\tilde{\beta}_i-1}}{B(\tilde{\alpha}_i,\tilde{\beta}_i)}\right]\cdot P(O,q_t=i|\lambda)\\&=\sum_{i=1}^{N}\sum_{t=1}^{T}[-\log B(\tilde{\alpha}_i,\tilde{\beta}_i)+(\tilde{\alpha}_i-1)\log O_t+(\tilde{\beta}_i-1)\log(1-O_t)]\cdot P(O,q_t=i|\lambda)\end{aligned} \tag{32}$$

Taking the derivative with respect to $\tilde{\alpha}_i$, setting it to be zero, and doing the same thing for $\tilde{\beta}_i$, we have

$$\begin{cases}\sum_{t=1}^{T}[\psi(\tilde{\alpha}_i+\tilde{\beta}_i)-\psi(\tilde{\alpha}_i)+\log O_t] & P(O,q_t=i|\lambda)=0\\ \sum_{t=1}^{T}[\psi(\tilde{\alpha}_i+\tilde{\beta}_i)-\psi(\tilde{\beta}_i)+\log(1-O_t)] & P(O,q_t=i|\lambda)=0\end{cases} \tag{33}$$

where $\psi(x)=\Gamma'(x)/\Gamma(x)$ is the digamma function. We define

$$\begin{cases}\tilde{f}(\tilde{\alpha}_i,\tilde{\beta}_i)\equiv\psi(\tilde{\alpha}_i)-\psi(\tilde{\alpha}_i+\tilde{\beta}_i)\\ \tilde{g}(\tilde{\alpha}_i,\tilde{\beta}_i)\equiv\psi(\tilde{\beta}_i)-\psi(\tilde{\alpha}_i+\tilde{\beta}_i)\end{cases} \tag{34}$$

Then, rearranging the terms in eq 33 yields

$$\begin{cases}\tilde{f}(\tilde{\alpha}_i,\tilde{\beta}_i)=\dfrac{\sum_{t=1}^{T}\log O_tP(O,q_t=i|\lambda)}{\sum_{t=1}^{T}P(O,q_t=i|\lambda)}=\dfrac{\sum_{t=1}^{T}\log O_t\gamma_t(i)}{\sum_{t=1}^{T}\gamma_t(i)}\\ \tilde{g}(\tilde{\alpha}_i,\tilde{\beta}_i)=\dfrac{\sum_{t=1}^{T}\log(1-O_t)P(O,q_t=i|\lambda)}{\sum_{t=1}^{T}P(O,q_t=i|\lambda)}=\dfrac{\sum_{t=1}^{T}\log(1-O_t)\gamma_t(i)}{\sum_{t=1}^{T}\gamma_t(i)}\end{cases} \tag{35}$$

These reestimation formulas implicitly contain $\tilde{\alpha}_i$ and $\tilde{\beta}_i$. Though we cannot get closed forms for $\tilde{\alpha}_i$ and $\tilde{\beta}_i$, we can numerically calculate them by solving eq 34 for given $\tilde{f}(\tilde{\alpha}_i,\tilde{\beta}_i)$ and $\tilde{g}(\tilde{\alpha}_i,\tilde{\beta}_i)$ values. In practice, a look-up table and bisection search can be used to speed up the calculation. However, we should mention that the numerically inverse is not an exact method to get $\tilde{\alpha}_i$ and $\tilde{\beta}_i$. It sometimes can cause decreased likelihood during the Baum−Welch reestimation, as shown in Figure 5. This is just a numerical effect.

Note that sometimes, to capture the anticorrelated feature of the donor and acceptor mean intensities, we use the constraint

$$I_i^{\mathrm{tot}}=I^{\mathrm{tot}}=\mathrm{const} \tag{36}$$

Here, $I_i^{\mathrm{tot}}=\alpha_i+\beta_i=\langle I_{\mathrm{A}}\rangle_i+\langle I_{\mathrm{D}}\rangle_i$ is the total emission intensity at state i. ΔI_{D} and ΔI_{A} between different states are not

identical unless the quantum yield (ϕ) and detection frequency (η) are comparable. For the various FRET pairs, the correction factor γ is defined as ($\eta_A\phi_A/\eta_D\phi_D$) and used to correct for absolute FRET as follows:[29]

$$E = I_A/(I_D + \gamma I_A) \tag{37}$$

For the most popular FRET pair, Cy3 and Cy5, the γ factor is 1 (ΔI_D and ΔI_A); therefore, we can assume that I_i^{tot} is a constant.

With the constraint (eq 36), the reestimation formula can be easily derived. Consider the third term of eq 42. Taking the derivative with respect to $\tilde{\alpha}_i$ and setting it to be zero, we have

$$\sum_{t=1}^{T}[\psi(I_i^{tot} - \tilde{\alpha}_i) - \psi(\tilde{\alpha}_i) + \log O_t - \log(1 - O_t)]P(O, q_t = i|\lambda) = 0 \tag{38}$$

We define

$$\tilde{h}(\tilde{\alpha}_i) \equiv \psi(I_i^{tot} - \tilde{\alpha}_i) - \psi(\tilde{\alpha}_i) \tag{39}$$

Then, rearranging the terms in eq 38 yields

$$\tilde{h}(\tilde{\alpha}_i) = \frac{\sum_{t=1}^{T}[\log(1 - O_t) - \log O_t]P(O, q_t = i|\lambda)}{\sum_{t=1}^{T}P(O, q_t = i|\lambda)} = \frac{\sum_{t=1}^{T}[\log(1 - O_t) - \log O_t]\gamma_t(i)}{\sum_{t=1}^{T}\gamma_t(i)} \tag{40}$$

Then, $\tilde{\alpha}_i$ can be calculated numerically by inversing $\tilde{h}(\tilde{\alpha}_i)$:

$$\tilde{\alpha}_i = \tilde{h}^{-1}\left[\frac{\sum_{t=1}^{T}[\log(1 - O_t) - \log O_t]\gamma_t(i)}{\sum_{t=1}^{T}\gamma_t(i)}\right] \tag{41}$$

We want to emphasize that, from the HMM point of view, the constraint (eq 36) is not necessary at all. It is not a basic element of HMM but just appropriate for the usual two-color smFRET data analysis. For a more complicated FRET scheme, e.g., the three-color FRET scheme, different constraints should be considered.

1.2. MPHMM. Without loss of generality, we consider an MHMM with multivariate observations. As shown in the previous section, Baum's auxiliary function can be written as

$$Q(\lambda, \tilde{\lambda}) = \sum_{q\in\mathcal{Q}}\log\tilde{\pi}_{q_1}P(O, q|\lambda) + \sum_{q\in\mathcal{Q}}\sum_{t=1}^{T-1}\log\tilde{a}_{q_tq_{t+1}}P(O, q|\lambda) + \sum_{q\in\mathcal{Q}}\sum_{t=1}^{T}\log\tilde{b}_{q_t}(\mathbf{O}_t)P(O, q|\lambda) \tag{42}$$

We optimize each term individually. Here, we consider the third term. Consider a d-dimensional Poisson distribution for the MHMM (eq 4); the third term of eq 42 can be written as

$$\begin{aligned}\sum_{q\in\mathcal{Q}}\sum_{t=1}^{T}\log\tilde{b}_{q_t}(\mathbf{O}_t)\cdot P(O, q|\lambda) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\log\tilde{b}_i(\mathbf{O}_t)\cdot P(O, q_t = i|\lambda)\\ &= \sum_{i=1}^{N}\sum_{t=1}^{T}\log\left(\prod_{k=1}^{d}\frac{e^{-\tilde{\mu}_{i,k}}\tilde{\mu}_{i,k}^{o_{t,k}}}{o_{t,k}!}\right)\cdot P(O, q_t = i|\lambda)\\ &= \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{d}\log\left(\frac{e^{-\tilde{\mu}_{i,k}}\tilde{\mu}_{i,k}^{o_{t,k}}}{o_{t,k}!}\right)\cdot P(O, q_t = i|\lambda)\\ &= \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{d}[-\tilde{\mu}_{i,k} + o_{t,k}\log\tilde{\mu}_{i,k} - \log(o_{t,k}!)]\cdot P(O, q_t = i|\lambda)\end{aligned} \tag{43}$$

Without any constraints, this term can be optimized by setting the derivative with respect to $\tilde{\mu}_{i,k}$ to be zero, which yields

$$\sum_{t=1}^{T}\left(-1 + \frac{o_{t,k}}{\tilde{\mu}_{i,k}}\right)P(O, q_t = i|\lambda) = 0 \tag{44}$$

Rearranging the terms yields the reestimation formula

$$\tilde{\mu}_{i,k} = \frac{\sum_{t=1}^{T}o_{t,k}P(O, q_t = i|\lambda)}{\sum_{t=1}^{T}P(O, q_t = i|\lambda)} = \frac{\sum_{t=1}^{T}o_{t,k}\gamma_t(i)}{\sum_{t=1}^{T}\gamma_t(i)} \tag{45}$$

If we consider the constraint $\sum_{k=1}^{d}\tilde{\mu}_{i,k} = I_i^{tot} = \text{const}$, we can add a Lagrange multiplier θ, and again setting the derivative with respect to $\tilde{\mu}_{i,k}$ to be zero:

$$\frac{\partial}{\partial\tilde{\mu}_{i,k}}\{\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{d}[-\tilde{\mu}_{i,k} + o_{t,k}\log\tilde{\mu}_{i,k} - \log(o_{t,k}!)]\cdot P(O, q_t = i|\lambda) + \theta(\sum_{k=1}^{d}\tilde{\mu}_{i,k} - I_i^{tot})\} = 0 \tag{46}$$

Taking the derivative yields

$$\sum_{t=1}^{T}\left(-1 + \frac{o_{t,k}}{\tilde{\mu}_{i,k}}\right)P(O, q_t = i|\lambda) + \theta = 0 \tag{47}$$

Rearranging the terms and summing over k yields

$$\theta = \sum_{t=1}^{T}(1 - I_{tot\,i}^{-1}\sum_{k=1}^{d}o_{t,k})P(O, q_t = i|\lambda) \tag{48}$$

One then solves for $\tilde{\mu}_{i,k}$:

$$\tilde{\mu}_{i,k} = \frac{I_i^{\text{tot}} \sum_{t=1}^{T} o_{t,k} P(O, q_t = i|\lambda)}{\sum_{t=1}^{T} (\sum_{k=1}^{d} o_{t,k}) P(O, q_t = i|\lambda)} = \frac{I_i^{\text{tot}} \sum_{t=1}^{T} o_{t,k} \gamma_t(i)}{\sum_{t=1}^{T} (\sum_{k=1}^{d} o_{t,k}) \gamma_t(i)} \tag{49}$$

2. Reestimation Formulas for Multiple Observation Sequences

Now, we consider a set of observation sequences

$$\mathbf{O} = \{O^{(1)}, O^{(1)}, \dots, O^{(M)}\} \tag{50}$$

where

$$O^{(m)} = o_1^{(m)} o_2^{(m)}, \dots, o_{T_m}^{(m)} \tag{51}$$

with $1 \le m \le M$. Generally, the multiple observation probability given the model can be expressed as

$$\begin{aligned} P(\mathbf{O}|\lambda) &= P(O^{(1)}|\lambda) P(O^{(2)}|O^{(1)}, \lambda) \dots \\ &\quad P(O^{(M)}|O^{(M-1)}, \dots, O^{(1)}, \lambda) \\ P(\mathbf{O}|\lambda) &= P(O^{(2)}|\lambda) P(O^{(3)}|O^{(2)}, \lambda) \dots \\ &\quad P(O^{(1)}|O^{(M)}, \dots, O^{(2)}, \lambda) \\ &\vdots \\ P(\mathbf{O}|\lambda) &= P(O^{(M)}|\lambda) P(O^{(1)}|O^{(M)}, \lambda) \dots \\ &\quad P(O^{(M-1)}|O^{(M-2)}, \dots, O^{(1)}, O^{(M)}, \lambda) \end{aligned} \tag{52}$$

If we assume

$$P(\mathbf{O}|\lambda) = \sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \tag{53}$$

then the weights are given by

$$\begin{aligned} \omega_1 &= \frac{1}{M} P(O^{(2)}|O^{(1)}, \lambda) \dots P(O^{(M)}|O^{(M-1)}, \dots, O^{(1)}, \lambda) \\ \omega_2 &= \frac{1}{M} P(O^{(3)}|O^{(2)}, \lambda) \dots P(O^{(1)}|O^{(M)}, \dots, O^{(2)}, \lambda) \\ &\vdots \\ \omega_K &= \frac{1}{M} P(O^{(1)}|O^{(M)}, \lambda) \dots P(O^{(M-1)}|O^{(M-2)}, \dots, O^{(1)}, O^{(M)}, \lambda) \end{aligned} \tag{54}$$

In practice, those weights are difficult to calculate. Therefore, the reestimation procedure is hard to implement. Nevertheless, it can be formally derived.[30] The basic idea follows.

The Baum auxiliary function can be constructed as

$$Q(\lambda, \tilde{\lambda}) = \sum_{m=1}^{M} \omega_m Q_m(\lambda, \tilde{\lambda}) \tag{55}$$

with

$$Q_m(\lambda, \tilde{\lambda}) = \sum_{q^{(m)} \in \mathcal{Q}} P(O^{(m)}, q^{(m)}|\lambda) \log P(O^{(m)}, q^{(m)}|\tilde{\lambda}) \tag{56}$$

Note that since ω_m are not functions of $\tilde{\lambda}$, the Lagrange multiplier method can be used to maximize the Baum auxiliary function. Since now

$$P(O^{(m)}, q^{(m)}|\tilde{\lambda}) = \tilde{\pi}_{q_1^{(m)}} \tilde{b}_{q_1^{(m)}}(\mathbf{O}_1^{(m)}) \prod_{t=1}^{T_m - 1} \tilde{a}_{q_t^{(m)} q_{t+1}^{(m)}} \tilde{b}_{q_{t+1}^{(m)}}(\mathbf{O}_{t+1}^{(m)}) \tag{57}$$

it follows that

$$\begin{aligned} Q(\lambda, \tilde{\lambda}) = \sum_{m=1}^{M} \omega_m \sum_{q^{(m)} \in \mathcal{Q}} (\log \tilde{\pi}_{q_1^{(m)}} + \sum_{t=1}^{T_m - 1} \log \tilde{a}_{q_t^{(m)} q_{t+1}^{(m)}} + \\ \sum_{t=1}^{T_m} \log \tilde{b}_{q_t^{(m)}}(\mathbf{O}_t^{(m)})) P(O^{(m)}, q^{(m)}|\lambda) \end{aligned} \tag{58}$$

We can optimize each term individually just as we did for the single observation sequence case. The reestimated formulas are as follows:

$$\tilde{\pi}_i = \frac{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \gamma_1^{(m)}(i)}{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda)} \tag{59}$$

$$\tilde{a}_{ij} = \frac{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m - 1} \xi_t^{(m)}(i,j)}{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m - 1} \gamma_t^{(m)}(i)} \tag{60}$$

For UGHMM,

$$\tilde{\mu}_i = \frac{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m} \gamma_t^{(m)}(i) o_t^{(m)}}{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m} \gamma_t^{(m)}(i)} \tag{61}$$

$$\tilde{\sigma}_i^2 = \frac{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m} \gamma_t^{(m)}(i) (o_t^{(m)} - \tilde{\mu}_i)^2}{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m} \gamma_t^{(m)}(i)} \tag{62}$$

For UBHMM (eq 6) with constraint (eq 36),

$$f(\tilde{\alpha}_i) = \frac{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m} [\log(1 - o_t^{(m)}) - \log o_t^{(m)}]\gamma_t^{(m)}(i)}{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m} \gamma_t^{(m)}(i)} \quad (63)$$

For two-dimensional Poisson HMM (eq 4) with constraint (eq 36),

$$\tilde{\mu}_{i,k} = \frac{I_i^{\text{tot}} \sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m} o_{t,k}^{(m)} \gamma_t^{(m)}(i)}{\sum_{m=1}^{M} \omega_m P(O^{(m)}|\lambda) \sum_{t=1}^{T_m} (\sum_{k=1}^{d} o_{t,k}^{(m)}) \gamma_t^{(m)}(i)} \quad (64)$$

Assuming that these observation sequences are **independent** of each other, i.e.,

$$P(\mathbf{O}|\lambda) = \prod_{m=1}^{M} P(O^{(m)}|\lambda) \quad (65)$$

which is a reasonable assumption in many cases, then the weights become

$$\omega_m = \frac{1}{M} \frac{P(O|\lambda)}{P(O^{(m)}|\lambda)} \quad (66)$$

and eqs 59, 60, 61, 62, 63, and 64 become

$$\tilde{\pi}_i = \frac{1}{M} \sum_{m=1}^{M} \gamma_1^{(m)}(i) \quad (67)$$

$$\tilde{a}_{ij} = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T_m - 1} \xi_t^{(m)}(i,j)}{\sum_{m=1}^{M} \sum_{t=1}^{T_m - 1} \gamma_t^{(m)}(i)} \quad (68)$$

$$\tilde{\mu}_i = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T_m} \gamma_t^{(m)}(i) o_t^{(m)}}{\sum_{m=1}^{M} \sum_{t=1}^{T_m} \gamma_t^{(m)}(i)} \quad (69)$$

$$\tilde{\sigma}_i^2 = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T_m} \gamma_t^{(m)}(i)(o_t^{(m)} - \tilde{\mu}_i)^2}{\sum_{m=1}^{M} \sum_{t=1}^{T_m} \gamma_t^{(m)}(i)} \quad (70)$$

$$f(\tilde{\alpha}_i) = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T_m} [\log(1 - o_t^{(m)}) - \log o_t^{(m)}]\gamma_t^{(m)}(i)}{\sum_{m=1}^{M} \sum_{t=1}^{T_m} \gamma_t^{(m)}(i)} \quad (71)$$

$$\tilde{\mu}_{i,k} = \frac{I_i^{\text{tot}} \sum_{m=1}^{M} \sum_{t=1}^{T_m} o_{t,k}^{(m)} \gamma_t^{(m)}(i)}{\sum_{m=1}^{M} \sum_{t=1}^{T_m} (\sum_{k=1}^{d} o_{t,k}^{(m)}) \gamma_t^{(m)}(i)} \quad (72)$$

References and Notes

(1) Roy, R.; Hohng, S.; Ha, T. *Nat. Methods* **2008**, *5*, 507–516.

(2) McKinney, S. A.; Declais, A.-C.; Lilley, D. M.; Ha, T. *Nat. Struct. Biol.* **2002**, *10*, 93–97.

(3) Zhuang, X.; Kim, H.; Pereira, M. J. B.; Babcock, H. P.; Walter, N. G.; Chu, S. *Science* **2002**, *296*, 1473–1476.

(4) Tan, E.; Wilson, T. J.; Nahas, M. K.; Clegg, R. M.; Lilley, D. M. J.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9308–9313.

(5) Andrec, M.; Levy, R. M.; Talaga, D. S. *J. Phys. Chem. A* **2003**, *107*, 7454–7464.

(6) Schröder, G. F.; Grubmüller, H. *J. Chem. Phys.* **2003**, *119*, 9920–9924.

(7) McKinney, S. A.; Freeman, A. D.; Lilley, D. M.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5715–5720.

(8) McKinney, S. A.; Joo, C.; Ha, T. *Biophys. J.* **2006**, *91*, 1941–1951.

(9) Rabiner, L. R. *Proc. IEEE* **1989**, *77*, 257–286.

(10) Durbin, R.; Eddy, S. R.; Krogh, A.; Mitchison, G. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1998.

(11) Chung, S. H.; Gage, P. W. *Methods Enzymol.* **1998**, *293*, 420–437.

(12) Qin, F.; Auerbach, A.; Sachs, F. *Biophys. J.* **2000**, *79*, 1915–1927.

(13) Milescu, L. S.; Yildiz, A.; Selvin, P. R.; Sachs, F. *Biophys. J.* **2006**, *91*, 3135–3150.

(14) Beausang, J. F.; Zurla, C.; Manzo, C.; Dunlap, D.; Finzi, L.; Nelson, P. C. *Biophys. J.* **2007**, *92*, L64–L66.

(15) Kruithof, M.; van Noort, J. *Biophys. J.* **2009**, *96*, 3708–3715.

(16) Lee, T.-H. *J. Phys. Chem. B* **2009**, *113*, 11535–11542.

(17) Xu, C. S.; Kim, H.; Hayden, C. C.; Yang, H. *J. Phys. Chem. B* **2008**, *112*, 5917–5923.

(18) Hohng, S.; Joo, C.; Ha, T. *Biophys. J.* **2004**, *87*, 1328–1337.

(19) Roy, R.; Kozlov, A. G.; Lohman, T. M.; Ha, T. *Nature* **2009**, *461*, 1092–1097.

(20) Dahan, M.; Deniz, A. A.; Ha, T.; Chemla, D. S.; Schultz, P. G.; Weiss, S. *Chem. Phys.* **1999**, *247*, 85–106.

(21) Talaga, D. S. *J. Phys. Chem. A* **2006**, *110*, 9743–9757.

(22) Baum, L. E.; Petrie, T.; Soules, G.; Weiss, N. *Ann. Math. Stat.* **1970**, *41*, 164–171.

(23) Liporace, L. A. *IEEE Trans. Inf. Theory* **1982**, *28*, 729–734.

(24) Jung, S.; Dickson, R. M. *J. Phys. Chem. B* **2009**, *113*, 13886–13890.

(25) Konishi, S.; Kitagawa, G. *Information Criteria and Statistical Modeling*; Springer Series in Statistics; Springer: New York, 2008.

(26) Joo, C.; McKinney, S. A.; Nakamura, M.; Rasnik, I.; Myong, S.; Ha, T. *Cell* **2006**, *126*, 515–527.

(27) Lanterman, A. D. *International Statistical Review* **2001**, *69*, 185–212.

(28) Bilmes, J. A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models; Technical Report, 1998.

(29) Ha, T.; Ting, A. Y.; Liang, J.; Caldwell, W. B.; Deniz, A. A.; Chemla, D. S.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 893–898.

(30) Li, X. L.; Parizeau, M.; Plamondon, R. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 371.

JP9057669