Supporting Material

Variational Bayes analysis of a photon-based hidden Markov model for single-molecule FRET trajectories

Kenji Okamoto and Yasushi Sako

S1. THEORETICAL DERIVATIONS

S1.1. Solution of hidden Markov model by variational Bayes

The time series of experimental data provided is given as $\mathbf{X} = \{x_1, \ldots, x_N\}$, where N is the total number of data points, with each individual data point in the time series assigned to one of the states. The latent variables $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ and $\mathbf{z}_n = \{z_{n1}, \ldots, z_{nK}\}$ are defined so that z_{nk} is equal to 1 if the molecule belongs to the k-th state at time n, and 0 otherwise, where K is the number of states, i.e. z_{nk} satisfies $z_{nk} \in \{0,1\}$ and $\sum_{k=1}^{K} z_{nk} = 1$. In the hidden Markov model (HMM), the simple Markov chain model is often assumed, that is, the probability distribution of the state at the n-th time step depends only on that at the (n-1)-th data point and is given by constants. The transition probability matrix $\mathbf{A}$ is introduced, the elements of which are described as $A_{ij} \equiv p(z_{nj} = 1 \mid z_{n-1,i} = 1)$ to represent the probability that the transition from the i-th state to the j-th state occurs ($i \neq j$) or the molecule stays in the i-th state ($i = j$). A_{ij} satisfies $0 \le A_{ij} \le 1$ and $\sum_{j}^{K} A_{ij} = 1$. The conditional distribution of transition probability can be written as

$$p(\mathbf{z}_n \mid \mathbf{z}_{n-1}, \mathbf{A}) = \prod_{i=1}^{K}\prod_{j=1}^{K} A_{ij}^{z_{n-1,i} z_{nj}} . \qquad \text{(S1)}$$

The initial state $\mathbf{z}_1$ is separately given by the conditional distribution as

$$p(\mathbf{z}_1 \mid \boldsymbol{\pi}) = \prod_{i=1}^{K} \pi_i^{z_{1i}} , \qquad \text{(S2)}$$

where π is a vector probability, the elements of which are defined as $\pi_k \equiv p(z_{1i} = 1)$ and satisfy $0 \le \pi_i \le 1$ and $\sum_{i}^{K} \pi_i = 1$. The experimental observable x_m at time m is dependent on the latent variable $\mathbf{z}_m$ and is connected via the emission probability distribution $p(x_m|\mathbf{z}_m, \phi)$, where ϕ is a set of parameters. The joint probability distribution can be written as:

$$p(\mathbf{X}, \mathbf{Z} \mid \Theta, M) = p(\boldsymbol{\pi}) p(\mathbf{A}) p(\phi) \times p(\mathbf{z}_1 \mid \boldsymbol{\pi}) \times \prod_{n=2}^{N} p(\mathbf{z}_n \mid \mathbf{z}_{n-1}, \mathbf{A}) \times \prod_{m=1}^{N} p(x_m \mid \mathbf{z}_m, \phi) , \qquad \text{(S3)}$$

where $\Theta = \{\boldsymbol{\pi}, \mathbf{A}, \phi\}$ denotes all the parameters related to the model and M represents the assumed model.

Parameters Θ and $\mathbf{Z}$ distribution are commonly optimized by maximum likelihood estimation (MLE) using the expectation-maximization (EM)-algorithm (1). In the E-step, $\mathbf{Z}$ distribution is optimized by a forward-backward algorithm, such as the Baum-Welch algorithm. Parameters are then optimized in the M-step. The E- and M-steps are iterated until the Θ and $\mathbf{Z}$ distributions converge. Finally, a max-sum algorithm, such as the Viterbi algorithm, is used to optimize the state transition trajectory.

In this paper, the HMM is solved by the variational Bayes (VB), which treats the marginal probability, so-called evidence, $p(\mathbf{X}|M)$ as shown in Eq. 5. The log evidence can be written as

$$\ln p(\mathbf{X} \mid M) = \mathcal{L}_q + \mathrm{KL}(q \,\|\, p) , \qquad \text{(S4)}$$

where

$$\mathcal{L}_q = \sum_{\mathbf{Z}} \int d\Theta q(\mathbf{Z},\Theta) \ln\left\{\frac{p(\mathbf{X},\mathbf{Z},\Theta \mid M)}{q(\mathbf{Z},\Theta)}\right\}, \tag{S5}$$

$$\mathrm{KL}(q \parallel p) = -\sum_{\mathbf{Z}} \int d\Theta q(\mathbf{Z},\Theta) \ln\left\{\frac{p(\mathbf{Z},\Theta \mid \mathbf{X},M)}{q(\mathbf{Z},\Theta)}\right\}. \tag{S6}$$

Terms $\mathcal{L}_q$ and KL(q||p) are functionals with respect to distribution functions p and q, where p is given with the model and q is to be fitted. The term KL(q||p) is the Kullback-Leibler divergence, which represents the similarity of two distribution functions, satisfies KL(q||p) ≥ 0 and vanishes when p and q are identical. Because the value of the term ln p(**X**|M) is fixed once the model M is assumed and the observable **X** is given, the other term $\mathcal{L}_q$ converges to ln p(**X**|M) as KL(q||p) decreases. In that sense, the term $\mathcal{L}_q$ represents the lower bound of the evidence. Eq. S5 can be maximized through an approximation procedure, similar to the mean field theory in physics. If we assume that **Z** and Θ are disjoint, the distribution function can be factorized, i.e.

$$q(\mathbf{Z},\Theta) = q(\mathbf{Z}) \times q(\Theta), \tag{S7}$$

and the distribution functions q(**Z**) and q(Θ) can be optimized by

$$\ln q^*(\mathbf{Z}) = \mathbb{E}_{\Theta}\left[\ln p(\mathbf{X},\mathbf{Z},\Theta \mid M)\right] + \text{const}, \tag{S8}$$

$$\ln q^*(\Theta) = \mathbb{E}_{\mathbf{Z}}\left[\ln p(\mathbf{X},\mathbf{Z},\Theta \mid M)\right] + \text{const}, \tag{S9}$$

where $\mathbb{E}_x[\cdots]$ denotes the expectation with respect to x (1).

Calculation of Eq. S8 can be simplified by a procedure similar to the E-step in the EM-algorithm. By taking the exponential of Eq. S8, we obtain

$$q^*(\mathbf{Z}) \propto \tilde{p}(\mathbf{z}_1 \mid \boldsymbol{\pi},M) \times \prod_{n=2}^{N} \tilde{p}(\mathbf{z}_n \mid \mathbf{z}_{n-1},\mathbf{A},M) \times \prod_{m=1}^{N} \tilde{p}(x_m \mid \mathbf{z}_m,\varphi,M) \tag{S10}$$

with

$$\tilde{p}(\mathbf{z}_1 \mid \boldsymbol{\pi},M) = \prod_{i}^{K} \exp\left(\overline{\ln \pi_i}\right)^{z_{1i}}, \tag{S11}$$

$$\tilde{p}(\mathbf{z}_n \mid \mathbf{z}_{n-1},\mathbf{A},M) = \prod_{i}^{K}\prod_{j}^{K} \exp\left(\overline{\ln A_{ij}}\right)^{z_{n-1,i}z_{nj}}, \tag{S12}$$

$$\tilde{p}(x_m \mid \mathbf{z}_m,\phi,M) = \exp\left(\overline{\ln p(x_m \mid \mathbf{z}_m,\phi,M)}\right), \tag{S13}$$

where overlines represent the mean. The optimized q^*(**Z**) can be calculated by the forward-backward algorithm using Eqs. S11–13 instead of Eqs. S2, S1 and $p(x_m|\mathbf{z}_m, \phi, M)$, respectively.

Calculation of the optimized q^*(Θ), which is given by Eq. S9, corresponds to the M-step of the EM-algorithm. q^*(Θ) can be factorized with respect to parameter groups disjoint to each other. Once q(Θ) is obtained, the expectation $\overline{\theta}$ for a parameter θ, together with $\overline{\ln\theta}$, which is often necessary for the E-step calculation, can be evaluated.

The variational lower bound defined by Eq. S5 can be written more simply as

$$\begin{aligned}\mathcal{L}_q = {} & \mathbb{E}\left[\ln p(\boldsymbol{\pi})\right] + \mathbb{E}\left[\ln p(\mathbf{A})\right] + \mathbb{E}\left[\ln p(\phi)\right] \\ & - \mathbb{E}\left[\ln q(\boldsymbol{\pi})\right] - \mathbb{E}\left[\ln q(\mathbf{A})\right] - \mathbb{E}\left[\ln q(\phi)\right] + \sum_{n=1}^{N} \ln \tilde{c}_n,\end{aligned} \tag{S14}$$

where $\tilde{c}_n$ is the scaling factor introduced for the forward-backward algorithm in the E-step calculation (1,2). The E- and M-steps are iterated until the lower bound $\mathcal{L}_q$ converges. Finally, $\mathcal{L}_q$ is compared among models to choose the likeliest model.

S1.2. VB-HMM-TS

The TS data are given by a single scalar value $x_n = \Delta t_n$, which represents the time lapsed after the previous photon. For ideal TS signals, Δt obeys the exponential distribution. The emission probability is then given as

$$p(x_m \mid \mathbf{z}_m, \mathbf{I}) = \prod_i^K \left\{ I_i \exp(-I_i x_m) \right\}^{z_{mi}} . \tag{S15}$$

In the HMM, the transition probability distribution $p(\mathbf{z}_n|\mathbf{z}_{n-1})$ is commonly given with the transition probability **A**-matrix under the assumption that time intervals between data points are uniform. In such cases, the transition probabilities represented by elements in the **A**-matrix are directly related to the transition rates, which we are essentially interested in physically, chemically and biologically. However, because the TS signals have variable time intervals between data points, their transition probabilities are not constant but depend on the data themselves. Therefore, we introduce the transition rates k_i and κ_{ij}, which are defined only for $i \neq j$, to replace the **A**-matrix. Under the assumption that the transition rates from the i-th to the j-th state ($i \neq j$) are constant during the experiment, k_i is defined as the rate that the molecule leaves from the i-th state. κ_{ij} ($i \neq j$) is the probability that the molecular state changes into the j-th state once the transition from the i-th state occurs with $0 \leq \kappa_{ij} \leq 1$ and $\sum_{j \neq i}^K \kappa_{ij} = 1$. The probability distribution is then written as

$$p(\mathbf{z}_n \mid \mathbf{z}_{n-1}, x_{n-1}, \mathbf{k}, \boldsymbol{\kappa}, \mathbf{I}) = \prod_i^K \left\{ \exp\left(-k_i x_{n-1}\right)^{z_{n-1,i} z_{ni}} \times \prod_{j \neq i}^K \left[\kappa_{ij} \{1 - \exp(-k_i x_{n-1})\} \right]^{z_{n-1,i} z_{nj}} \right\}. \tag{S16}$$

This is marginalized with respect to x_{n-1} to obtain

$$p(\mathbf{z}_n \mid \mathbf{z}_{n-1}, \mathbf{k}, \boldsymbol{\kappa}, \mathbf{I}) = \int dx_{n-1} p(x_{n-1} \mid \mathbf{z}_{n-1}, \mathbf{k}, \boldsymbol{\kappa}, \mathbf{I}) p(\mathbf{z}_n \mid \mathbf{z}_{n-1}, x_{n-1}, \mathbf{k}, \boldsymbol{\kappa}, \mathbf{I}) \tag{S17}$$

$$= \prod_i^K \left[\left(\frac{I_i}{k_i + I_i} \right)^{z_{n-1,i} z_{ni}} \times \prod_{j \neq i}^K \left(\frac{k_i \kappa_{ij}}{k_i + I_i} \right)^{z_{n-1,i} z_{nj}} \right]. \tag{S18}$$

Now, we obtain the joint probability distribution to represent the TS intensity signal:

$$p(\mathbf{X} \mid \mathbf{Z}, \Theta, M) = p(\boldsymbol{\pi}) p(k) p(\kappa) p(\mathbf{I}) \times \prod_i^K \pi^{z_{1i}} \times \prod_{n=2}^N \prod_i^K \left[\left(\frac{I_i}{k_i + I_i} \right)^{z_{n-1,i} z_{ni}} \times \prod_{j \neq i}^K \left(\frac{k_i \kappa_{ij}}{k_i + I_i} \right)^{z_{n-1,i} z_{nj}} \right]$$

$$\times \prod_m^N \prod_i^K \left\{ I_i \exp(-I_i x_m) \right\}^{z_{mi}} . \tag{S19}$$

The E-step can be calculated by using Eq. S11 and similarly derived

$$\tilde{p}(\mathbf{z}_n \mid \mathbf{z}_{n-1}, \mathbf{k}, \boldsymbol{\kappa}, \mathbf{I}) = \prod_i^K \left[\exp\left(\overline{\ln I_i} - \overline{\ln(k_i + I_i)} \right)^{z_{n-1,i} z_{ni}} \times \prod_{j \neq i}^K \exp\left(\overline{\ln k_i} + \overline{\ln \kappa_{ij}} - \overline{\ln(k_i + I_i)} \right)^{z_{n-1,i} z_{nj}} \right], \tag{S20}$$

$$\tilde{p}(x_m \mid \mathbf{z}_m, \mathbf{I}) = \prod_i^K \exp\left(\overline{\ln I_i} - \overline{I}_i x_m \right)^{z_{mi}} . \tag{S21}$$

For the M-step calculation to optimize parameters, the distribution functions for parameters must be obtained as follows.

The distribution function for the initial state distribution $\boldsymbol{\pi}$ is commonly given by the Dirichlet distribution as

$$q^*(\boldsymbol{\pi}) = \frac{\Gamma\left(u_0^\pi + 1\right)}{\prod_i^K \Gamma\left(u_i^\pi + \overline{z_{1i}}\right)} \prod_i^K \pi_i^{u_i^\pi + z_{1i} - 1} \tag{S22}$$

where $\Gamma(x)$ is the Gamma function. The Dirichlet distribution is also given as the prior probability distribution. u_i^π are the hyperparameters for the prior distribution and $u_0^\pi = \sum_i^K u_i^\pi$. From Eq. S22, expectations to estimate parameters and to be used in the E-step calculation are obtained as

$$\overline{\pi_i} = \frac{u_i^\pi + \overline{z_{1i}}}{u_0^\pi + 1} , \tag{S23}$$

$$\overline{\ln \pi_i} = \Psi\left(u_i^{\pi} + \overline{z_{1i}}\right) - \Psi\left(u_0^{\pi} + 1\right), \tag{S24}$$

respectively, where $\Psi(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the digamma function. We gave $u_i^{\pi} = 1$ for all states to assume a flat probability distribution.

The probability distributions for the transition rate k and the intensity $\mathbf{I}$ cannot be factorized, so that the distribution function must be written together :

$$q^*(k_i, I_i) = \frac{\Gamma(N_{ii} + N_{ij})\left(b_{ii}^I + T_i\right)^{a_{ii}^I N_i}}{\Gamma(a_i^I N_i)\Gamma(N_{ii})\Gamma(N_{ij})} \times \frac{I_i^{a_i^I + N_i + N_{ii} - 1} k_i^{N_{ij} - 1}}{(k_i + I_i)^{N_{ii} + N_{ij}}} \times \exp\left(-(b_i^I + T_i) I_i\right). \tag{S25}$$

where $N_i = \sum_{n=1}^{N} \overline{z_{ni}}$, $T_i = \sum_{n=1}^{N} \overline{z_{ni}} x_n$, $N_{ii} = \sum_{n=2}^{N} \overline{z_{n-1,i} z_{ni}}$, $M_{ij} = \sum_{n=2}^{N} \overline{z_{n-1,i} z_{nj}}$ and $N_{ij} = \sum_{j \neq i}^{K} M_{ij} = \sum_{n=2}^{N} \sum_{j \neq i}^{K} \overline{z_{n-1,i} z_{nj}}$. The scale-invariant prior distribution $p(k_i) = C_i^k \, k_i^{-1}$ and the Gamma distribution $p(I) = I_i^{a_i^I - 1} \exp(-b_i^I I_i)$ are given for k_i and I_i, respectively, with hyperparameters C_i^k, a_i^I and b_i^I. We gave $C_i^k = 1$ as described below, $a_i^I = 1$ and $b_i^I = 1/\overline{I}$ for all states, where $\overline{I}$ $= N/T$ is the average intensity over the whole data with the total time length T. Necessary expectations are obtained as

$$\overline{k_i} = \frac{(a_{ii}^I + N_i) N_{ij}}{(N_{ii} - 1)(b_{ii}^I + T_i)}, \tag{S26}$$

$$\overline{\ln k_i} = \Psi(a_i^I + N_i) + \Psi(N_{ij}) - \Psi(N_{ii}) - \ln(b_i^I + T_i), \tag{S27}$$

$$\overline{I_i} = \frac{a_{ii}^I + N_i}{b_{ii}^I + T_i}, \tag{S28}$$

$$\overline{\ln I_i} = \Psi(a_i^I + N_i) - \ln(b_i^I + T_i), \tag{S29}$$

$$\overline{k_i + I_i} = \overline{k_i} + \overline{I_i}, \tag{S30}$$

$$\overline{\ln(k_i + I_i)} = \Psi(a_i^I + N_i) + \Psi(N_{ii} + N_{ij}) - \Psi(N_{ii}) - \ln(b_i^I + T_i). \tag{S31}$$

A possible problem with using the transition rate k may be the choice of the prior probability distribution for k_i. We employed the scale-invariant prior distribution $p(k_i) \propto k_i^{-1}$, which requires no prior knowledge about the transition rates, not even their order. This can be advantageous, because it is usually difficult and not desirable to predict the transition rates in advance of experiments. However, it may be mathematically problematic that integration of the scale-invariant prior distribution does not converge. We introduced the normalizing coefficient C_i^k as a hyperparameter. C_i^k is cancelled in the derivation of $q(k_i, I_i)$ as shown in Eq. S25, while it remains in the calculation of the variational lower bounds given by Eq. S14. We expediently gave $C_i^k = 1$ and conclude that it is valid from the results of numerical experiments.

κ_{ij} are disjoint from other parameters and are given as the Dirichlet distribution. Similarly to π, the distribution function can be obtained as

$$q^*(\kappa_{ij}) = \frac{\Gamma(u_{i0}^{\kappa} + N_{ij})}{\prod_{j \neq i}^{\kappa} \Gamma(u_{ij}^{\kappa} + M_{ij})} \prod_{j \neq i}^{\kappa} \kappa_{ij}^{u_{ij}^{\kappa} + M_{ij} - 1}, \tag{S32}$$

where $u_{ij}{}^{\kappa}$ are hyperparameters and $u_{i0}^{\kappa} = \sum_{j \neq i}^{K} u_{ij}^{\kappa}$. Expectations are then calculated as

$$\overline{\kappa_{ij}} = \frac{u_{ij}^{\kappa} + M_{ij}}{u_{i0}^{\kappa} + N_{ij}}, \tag{S33}$$

$$\overline{\ln \kappa_{ij}} = \Psi(u_{ij}^{\kappa} + M_{ij}) - \Psi(u_{i0}^{\kappa} + N_{ij}). \tag{S34}$$

We set $u_{ij}{}^{\kappa} = 1/(K\text{-}1)$ for all i and j ($i \neq j$).

S1.3. VB-HMM-TS-FRET

The TS-FRET data analysis is based on the TS intensity analysis. We therefore need to extend the observable from a scalar x_n to a vector $\mathbf{x}_n$, adding a new quantity

$$\rho_n = \begin{cases} 0 & \text{(donor)} \\ 1 & \text{(acceptor)} \end{cases} \tag{S35}$$

so that $\mathbf{x}_n = \{\Delta t_n, \rho_n\}$. We also add a new parameter E, the FRET efficiency. Now each state is characterized by I and E, which can be written in principle as,

$$I = I_{\mathrm{A}} + I_{\mathrm{D}}\,, \tag{S36}$$

$$E = \frac{I_A}{I_A + I_D}\,, \tag{S37}$$

where I_{A} and I_{D} are the acceptor and the donor intensities, respectively. The relationships of Eqs. S36 and S37 are, however, not consistent in most of the experimental situations because of various factors, such as background signals, differences in the quantum yields and the detection efficiencies, crosstalk over the optical filter and direct excitation of the acceptor. It is usually necessary to compensate for these errors to obtain the true I and E values (3,4). However, we do not consider these compensations in this paper. As long as those errors are constant during the experiments, we need only to carry out the analysis with the apparent I and E and compensate for them after they are optimized.

The emission probability is factorized with respect to Δt_n, I and ρ_n, E and written as

$$p(\mathbf{x}_m \mid \mathbf{z}_m, \mathbf{I}, \mathbf{E}) = p(\rho_m \mid \mathbf{z}_m, \mathbf{E}) \times p(\Delta t_m \mid \mathbf{z}_m, \mathbf{I}) \tag{S38}$$

$$= \prod_i^K \left[\left\{ E_i^{\rho_m} \times (1 - E_i)^{1-\rho_m} \right\} \times I_i \exp(-I_i \Delta t_m) \right]^{z_{mi}}, \tag{S39}$$

which gives

$$\tilde{p}(x_m \mid \mathbf{z}_m, \mathbf{I}, \mathbf{E}) = \prod_i^K \left\{ \exp\left(\rho_n \overline{\ln E_i} + (1 - \rho_n) \overline{\ln(1 - E_i)} + \overline{\ln I_i} - \overline{I}_i \Delta t_n \right) \right\}^{z_{mi}}. \tag{S40}$$

As E is independent of the other parameters, its distribution function can be calculated separately. $q^*(E)$ is given by the binomial distribution in the range [0, 1] as

$$q^*(E_i) = \frac{\Gamma(u_i^E + v_i^E + N_i)}{\Gamma(u_i^E + \varepsilon_i)\Gamma(v_i^E + N_i - \varepsilon_i)} \times E_i^{u_i^E + \varepsilon_i - 1} \times (1 - E_i)^{v_i^E + N_i - \varepsilon_i - 1}, \tag{S41}$$

where the binomial prior distribution is given with hyperparameters u_i^E and v_i^E. $N_i = \sum_{n=1}^N \overline{z_{ni}}$ and $\varepsilon_i = \sum_{n=1}^N \overline{z_{ni}} \rho_n$ are defined. We gave $u_i^E = 1$ and $v_i^E = 1$ for all states to assume a flat probability distribution. Finally, necessary expectations are calculated as

$$\overline{E_i} = \frac{u_i^E + \varepsilon_i}{u_i^E + v_i^E + N_i}\,, \tag{S42}$$

$$\overline{1 - E_i} = \frac{v_i^E + N_i - \varepsilon_i}{u_i^E + v_i^E + N_i}\,, \tag{S43}$$

$$\overline{\ln E_i} = \Psi(u_i^E + \varepsilon_i) - \Psi(u_i^E + v_i^E + N_i)\,, \tag{S44}$$

$$\overline{\ln(1 - E_i)} = \Psi(v_i^E + N_i - \varepsilon_i) - \Psi(u_i^E + v_i^E + N_i)\,. \tag{S45}$$

S1.4. VB-HMM-PC

In time-binned SPC measurements, the number of photons is counted for each time bin with fixed width. Data x_n is the photon counts and the mean photon counts μ is given in unit of counts/bin. It is well known that this type of signal obeys the Poisson distribution, in which case the emission probability can be written as

$$p(x_m \mid \mathbf{z}_m, \mu) = \prod_{i=1}^{K} \left\{ \frac{\mu_i^{x_m}}{x_m!} \exp(-\mu_i) \right\}^{z_{mi}} . \tag{S46}$$

For the SPC signals, the transition probability matrix **A** can be employed to describe the transition probability distribution $p(\mathbf{z}_n|\mathbf{z}_{n-1})$. Elements of the **A**-matrix are independent of other parameters and are given as the Dirichlet distribution. Similarly to $\boldsymbol{\pi}$, the distribution $q^*(\mathbf{A})$ can be obtained as

$$q^*(\mathbf{A}) = \prod_{i}^{K} \frac{\Gamma(u_{i0}^A + M_i)}{\prod_{j}^{K} \Gamma(u_{ij}^A + N_{ij})} \prod_{j}^{K} A_{ij}^{u_{ij}^A + N_{ij} - 1} , \tag{S47}$$

where the Dirichlet prior distribution is given with hyperparameters u_{ij}^A and $u_{i0}^A = \sum_j^K u_{ij}^A$. $N_{ij} = \sum_{n=2}^{N} \overline{z_{n-1,i} z_{nj}}$ and $M_i = \sum_j^K N_{ij} = \sum_{n=2}^{N} \sum_j^K \overline{z_{n-1,i} z_{nj}}$ are defined. We set $u_{ij}^A = 100$ ($i = j$) or 1 ($i \neq j$) for all i and j to assume that transitions do not occur too frequently and every state has an equal chance as the destination after transition. In contrast to the transition rates **k**, the **A**-matrix is given the Dirichlet distribution as the prior distribution. While it is mathematically consistent, it may be a drawback that some preliminary knowledge about the transition rates is required to determine hyperparameters, especially, for the diagonal elements of the matrix. Expectations are calculated as

$$\overline{A_{ij}} = \frac{u_{ij}^A + N_{ij}}{u_{i0}^A + M_i} , \tag{S48}$$

$$\overline{\ln A_{ij}} = \Psi(u_{ij}^A + N_{ij}) - \Psi(u_{i0}^A + M_i) . \tag{S49}$$

The parameter μ is also independent of other parameters and its distribution $q^*(\mu_i)$ can be obtained as

$$q^*(\mu_i) = \frac{(b_i^\mu + N_i)^{a_i^\mu + C_i}}{\Gamma(a_i^\mu + C_i)} \times \mu_i^{a_i^\mu + C_i - 1} \times \exp\left(-(b_i^\mu + N_i)\mu_i\right) , \tag{S50}$$

where $C_i = \sum_n^N \overline{z_{ni}} x_n$ and $N_i = \sum_n^N \overline{z_{ni}}$. The Gamma distribution $p(\mu_i) = \mu_i^{a_i^\mu - 1} \exp(-b_i^\mu \mu_i)$ is given as the prior distribution, with hyperparameters $a_i^\mu = 1$ and $b_i^\mu = 1/\overline{I}$ for all states. Necessary expectations are calculated as

$$\overline{\mu_i} = \frac{a_i^\mu + C_i}{b_i^\mu + N_i} , \tag{S51}$$

$$\overline{\ln \mu_i} = \Psi(a_i^\mu + C_i) - \ln(b_i^\mu + N_i) , \tag{S52}$$

S1.5. VB-HMM-PC-FRET

The SPC signal is extended to dual-channel detection, and data are now given as vector $\mathbf{x}_m = \{d_m, a_m\}$, where d_m and a_m are the photon counts in the m-th bin on the donor and the acceptor detection channels, respectively. Both d_m and a_m obey the Poisson distribution and then the emission probability can be written with μ and E as

$$p(\mathbf{x}_m \mid \mathbf{z}_m, \mu, \mathbf{E}) = \prod_{i=1}^{K} \left\{ \frac{\{\mu_i(1-E_i)\}^{d_m}}{d_m!} \exp\left(-\mu_i(1-E_i)\right) \times \frac{(\mu_i E_i)^{a_m}}{a_m!} \exp(-\mu_i E_i) \right\}^{z_{mi}}$$

$$= \prod_{i=1}^{K} \left\{ \frac{(1-E_i)^{d_m} E_i^{a_m}}{d_m! a_m!} \times \mu_i^{d_m+a_m} \exp(-\mu_i) \right\}^{z_{mi}} . \quad \text{(S53)}$$

The terms on μ and E can be factorized and then the distribution $q^*(\mu_i)$ and expectations $\overline{\mu_i}$ and $\overline{\ln \mu_i}$ can be obtained by Eqs. S50–52, respectively. The distribution $q^*(E_i)$ is given as

$$q^*(E_i) = \frac{\Gamma(u_i^E + v_i^E + C_i)}{\Gamma(u_i^E + A_i)\Gamma(v_i^E + D_i)} \times E_i^{u_i^E + A_i - 1} \times (1-E_i)^{v_i^E + D_i - 1} , \quad \text{(S54)}$$

where the binomial prior distribution is given with hyperparameters u_i^E and v_i^E, which are both assumed to be 1 for all states. $D_i = \sum_n^N \overline{z_{ni}} d_n$, $A_i = \sum_n^N \overline{z_{ni}} a_n$ and $C_i = D_i + A_i$ are defined. Necessary expectations are then calculated as

$$\overline{E_i} = \frac{u_i^E + A_i}{u_i^E + v_i^E + C_i} , \quad \text{(S55)}$$

$$\overline{1-E_i} = \frac{v_i^E + D_i}{u_i^E + v_i^E + C_i} , \quad \text{(S56)}$$

$$\overline{\ln E_i} = \Psi(u_i^E + A_i) - \Psi(u_i^E + v_i^E + C_i) , \quad \text{(S57)}$$

$$\overline{\ln(1-E_i)} = \Psi(v_i^E + D_i) - \Psi(u_i^E + v_i^E + C_i) . \quad \text{(S58)}$$

S1.6. Step-by-step procedure

The actual procedure of the VB-HMM analysis is as follows:

1. Assume the number of states.
2. Apply initial values to all parameters of every state.
   Because the results may depend on initial values, a group of analyses should be repeated with different initial values, which may be applied with a degree of randomness.
3. Optimize the Z-distribution (Eq. S8).
   Eq. S10 should be maximized with, for the case of VB-HMM-TS-FRET, Eqs. S11, S20 and S40. Various numerical methods, such as the forward-backward algorithm (1), are available for this purpose.
4. Update parameters (Eq. S9).
   For the case of VB-HMM-TS-FRET, Eqs. S26–S31 and S42–45 should be calculated for use in step 3 of the subsequent cycle.
5. Calculate the lower bound $\mathcal{L}_q$ (Eq. S14).
   The expectations of prior probability distributions p can be calculated with hyperparameters and do not change during iteration. The probability distributions q can be calculated using S22, S25, S32 and S41 with S24, S27–S31, S34 and S44, respectively, for the case of VB-HMM-TS-FRET. The variables $\tilde{c}_n$ in the last term are calculated during the forward-backward algorithm in step 3.
6. Judge convergence of the lower bound $\mathcal{L}_q$.
   This can be judged by, for example, checking whether the difference from the previous cycle is small enough. If $\mathcal{L}_q$ has not yet converged, go on to the next cycle, repeating from step 3.
7. Optimize the state transition trajectory.
   Using a max-sum algorithm, such as Viterbi algorithm.

To find the optimum number of states, one must repeat the above procedure with different numbers of states to find the optimum number, which gives the highest lower bound.

S2. MATERIALS AND METHODS

S2.1. Numerical experiments

S2.1.1. Procedure

Each simulation proceeds as follows. The molecule is simulated to change its state at the given transition rates and emit photons according to the given intensity related to each state. It is assumed that the molecule has three states and that each of the two other states has an equal chance to be chosen as the next state after the transition occurs, i.e. $\kappa_{ij} = 0.5$ or $A_{ij} = (1 - A_{ii})/2$ for all i and j ($i \neq j$). During dual-channel FRET simulation, each photon is assigned to the donor or the acceptor detector channel, on which the photon is detected, according to the FRET efficiency related to the state. Once the TS signal is generated, the SPC signal is derived from it with three bin sizes, 1, 5 and 10 ms, for SPC analyses. The same data are then analyzed by several analytical methods with assumptions of the NoS from 2 to 6 and the likeliest NoS is determined. Cycles of data generation and analysis are repeated 1000 times and the statistics of the results are summarized.

The first set of simulation treats the single-channel TS intensity signals. Total signal length, T, is fixed to 10 s, while two parameters are varied. The first parameter varied is the intensity for states I_i (i = 1–3). We assumed that $I_1 < I_2 < I_3$ and the ratio of intensities between adjacent states was constant, i.e. $I_{\text{ratio}} \equiv I_2/I_3 = I_1/I_2$. The parameters I_3 = 10,000/s and $I_{\text{ratio}} = \{0.25, 0.5\}$ were then given. Another parameter varied is the transition rate. We defined transition rates with an average lifetime of states, λ_i. We assumed $\lambda_3/\lambda_2 = \lambda_2/\lambda_1 = I_{\text{ratio}}$ so that the average number of photons per state was constant, because the resolvability of states is thought to depend on the number of photons for TS signals (5). We fixed $\{\lambda_1, \lambda_2, \lambda_3\} = \{\{2, 1, 0.5\}, \{4, 2, 1\}, \{8, 4, 2\}, \{20, 10, 5\}, \{48, 24, 12\}, \{120, 60, 30\}\}$ ms. The probability of choosing one of two states as the destination for individual transitions is always 50%, as described above. The simulated data were analyzed by VB-HMM-TS and then converted to SPC data with bin sizes of 1, 5 and 10 ms for analysis by VB-HMM-PC. The TS signal was also analyzed by the CPD-MLE method (5), which was conducted with the probability of a type-I error $\alpha = 0.01$ and the probability for the confidence interval $\beta = 0.01$. We did not employ CPD-Bayes method (6) due to lack of clarity about how to pick the threshold for the Bayes factor.

The second set of simulations is based on the dual-channel TS-FRET signals. Analytical methods are all variants of VB-HMM for this set. The parameter of the intensity I is defined as the total photon rate, including the donor and the acceptor photons and is fixed to 10,000/s for all states. As a state-dependent parameter, the FRET efficiency E_i is given as $E_1 = 0.5 - \Delta E$, $E_2 = 0.5$ and $E_3 = 0.5 + \Delta E$ with $\Delta E = \{0.2, 0.1\}$. Transition rates are defined by λ, which is given the same value for all states this time and varied as $\lambda = \{6.25, 25, 100, 400\}$ ms. T = 10 s except for cases where $\Delta E = 0.2$, in which simulation is performed with T = 3 s and $\lambda = \{2, 4, 8, 16\}$ ms. This set of simulated data are first analyzed by VB-HMM-TS-FRET and then converted to SPC-FRET data with bin sizes of 1, 5, and 10 ms, to be analyzed by VB-HMM-PC-FRET. Finally, VB-HMM-TS and VB-HMM-PC methods are applied to only the donor channel data from TS-FRET and SPC-FRET signals, respectively.

S2.1.2. Dependence on the total signal length

Among the various simulation parameters, the total signal length T may be a sigificant parameter influencing the analysis result. The numbers of observed states and transitions between them are determined by T. The more events that are observed, the more robust and reliable the inference will be.

To investigate dependence on T, we conducted another series of simulations, in which T was a variable parameter, given as $T = \{0.625, 1.25, 2.5, 5.0, 10.0, 20.0\}$ s.

First, analyses of the intensity signals with VB-HMM-TS, CPD-MLE and VB-HMM-PC were given parameters I_3 = 10,000/s, $I_{\text{ratio}} = 0.5$ and $\{\lambda_1, \lambda_2, \lambda_3\} = \{20, 10, 5\}$ ms. Because the sum

of λ over three states is 35 ms, each state will be observed a few tens of times on average within the shortest T.

Second, FRET analysis using VB-HMM-TS-FRET, VB-HMM-PC-FRET, VB-HMM-TS and VB-HMM-PC were performed with parameters I = 10,000/s, ΔE = 0.1 and λ = 100 ms. The summed lifetime is 300 ms in this case. Because the transitions are made stochastically, some of simulated traces at short T may not include all three states.

The results are shown in Figs. S8–9.

S2.1.3. Influence of noise

SPC signals intrinsically contain "shot noise". However, while it is called "noise", it actually means ununiformity of photon distribution due to stochasticity of photon emission and it is therefore unavoidable. It is considered in our analyses through emission probabilities, such as in Eqs. 1 and 2.

Except for this, typical external noise sources in SPC signals are the dark count of the detector and contributions from the background light. Counts due to both noise sources, as well as the signal photon counts, obey Poisson statistics and their rates are constant if the experimental apparatus is carefully set up. When some Poissonian signals are mixed, the total distribution of counts is also Poissonian, whose count rate (mean count per bin) is the sum of each. Therefore, photon-based analyses, such as our VB-HMM methods or CPD, can be applied to the apparent intensity signal, including noise. Therefore, the user should be careful when interpreting results, i.e. the parameters I or E estimated by analysis include a noise component of $I' = I + B$ or $E' = (I_A + B_A)/(I_A + B_A + I_D + B_D)$, where I' and E' are the apparent intensity and the estimated FRET efficiency, and B, B_A and B_D are the noise components of intensity, on the acceptor and the donor channel, respectively.

As a matter of course, there is a disadvantage in treating signals with a low S/N ratio. Because the noise component becomes a bias to the signal, it lowers the signal contrast between states. For example, when two states with intensities I and $0.5I$ exist, addition of $0.25I$ noise degrades the contrast from 0.5 to $0.75I/1.25I = 0.6$. If a two-state signal with E_1=0.4 and E_2=0.6 is contaminated with $0.25I$ noise, the apparent FRET difference becomes 0.16, decreased from 0.2. Because a higher signal contrast improves the performance of analysis, as shown in Figs. 4–5 and S5–S6, noise should also be reduced as much as possible in photon-based signals.

To investigate the tolerance of the analyses to noise, we conducted another series of simulations. First, the intensity signals were simulated with parameters I_3 = 10,000/s, I_{ratio} = 0.5, $\{\lambda_1, \lambda_2, \lambda_3\} = \{20, 10, 5\}$ ms and T = 10 s and given extra counts with rates 0, $0.05I_3$, $0.1I_3$, $0.2I_3$, $0.5I_3$ and I_3 as a noise component, with which we defined the noise-to-signal (N/S) ratio as 0, 0.05, 0.1, 0.2, 0.5 and 1, respectively.

Second, FRET simulation used parameters I = 10,000/s, ΔE = 0.1, λ = 100 ms and T = 10 s. The additional noise count rates were 0, $0.05I_3$, $0.1I_3$, $0.2I_3$, $0.5I_3$ and I_3, again, but divided into two detector channels, half and half.

The results are shown in Figs. S10–11.

S2.2. Sample preparation

S2.2.1. Holliday junction DNA

A set of oligonucleotides (Japan Bio Services Co., Ltd., Saitama, Japan) was synthesized for the assembly of HJ. Four oligonucleotides were annealed to form a four-way junction structure, as described elsewhere (7). The sequences of oligonucleotides we used are the following:

1: biotin-5'-TCT TTT GAT AAG CTT GCA AGC ATA CAT ATC TCG TAA TTT CCG GTT AGG T-3'
2: 5'-ACC TAA CCG GAA ATT ACG AGA TAT CGA TGC ATG CAA GCT TCA CA-3'
3: 5'-TGT GAA GCT TGC AT$_D$G CAT CGA TTT AAT ACG TGA GGC CTA GGA TC-3'
4: 5'-GAT CCT AGG CCT CAC GTA TTA AAT GTA TGC T$_A$TG CAA GCT TAT CA-3'

T_D and T_A are thymines labeled with fluorescence dyes AlexaFluor488 and AlexaFluor594 as the FRET donor and the acceptor, respectively. Underlined bases represents the exchangeable parts in branch migration. This HJ is expected to migrate between 3 conformational states.

S2.2.2. Immobilization

A sample cell is made of cleaned coverslips with double-sided tapes as spacers. The cell was treated with 1 mg/ml biotinylated BSA (Sigma, St. Louis, MO) in TES buffer (pH 7.5, 10 mM Tirs-HCl, 150 mM NaCl and 1 mM EDTA) for 10 min, washed with TES buffer, treated with 0.2 mg/ml streptavidin (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA) in TES buffer for 10 min and washed with TES buffer, again. ~50 pM HJ solution in TES buffer was introduced and incubated for 10 min. Finally the solvent was exchanged to TN40M buffer (pH 7.5, 10m M Tris-HCl, 10 mM NaCl and 40 mM $MgCl_2$) for measurement. 1 mM Trolox and 10 mM 2-mercaptoethylamine (cysteamine) were added to prevent photobleaching of dyes and to scavenge the oxygen radicals (8).

S2.3. Apparatus

Single-molecule FRET measurements were conducted on a confocal microscope system based on a Ti-E system (Nikon, Tokyo, Japan) equipped with an oil-immersion objective lens Apo TIRF 60× NA 1.49 (Nikon, Tokyo, Japan) and a piezo-driven sample scanning stage P-542.2CL (Physik Instrumente GmbH, Karlsruhe, Germany). The excitation laser was Sapphire 488-100 CDRH (Coherent, Santa Clara, CA) with wavelength of 488 nm. Laser light was attenuated down to ~44 nW at the sample by ND filters and introduced to the sample by a custom-made dichroic cube beam splitter (c/o 540nm (S-polarization; excitation)/470nm (P-polarization); Sigma-Koki, Tokyo, Japan), which prevents astigmatism. Fluorescence was collected through a pinhole with 100 μm diameter and a double notch filter NF03-488E-25 (Semrock, Lake Forest, IL) to remove residual laser light. After a dichroic filter Q570LP (Chroma Technology Corp., Bellows Falls, VT) to separate fluorescence into a donor and an acceptor channels, each light was focused onto an avalanche photodiode (APD) single-photon-counting detectors (SPCM-AQR, Perkin Elmer, Optoelectronics, Freemont, CA). (Dark counts are typically < 150/s for both APDs.) The time stamp signal was acquired by a PC equipped with a counter board LPC-632104 (Interface Corp., Hiroshima, Japan), which detects photon pulses based on a 10 MHz clock. The measurement system was controlled by a LabVIEW (National Instruments Corp., Austin, TX) program written in-house.

S2.4. FRET calculation

S2.4.1. Calculation of true FRET

The FRET efficiency was calculated as

$$E_{\mathrm{FRET}} = \frac{I_{\mathrm{A}} - \beta I_{\mathrm{D}}}{I_{\mathrm{A}} + (\alpha - \beta) I_{\mathrm{D}}}, \tag{S59}$$

where I_A and I_D are the detected fluorescence intensities after subtraction of the background signals on the acceptor and the donor detectors, respectively. α is a factor correcting differences in the dyes' quantum yields and the detection efficiencies between dyes and detectors, while β is the coefficient for leakage of the donor fluorescence onto the acceptor channel. We determined the values of α and β experimentally from the acceptor's photobleaching events (3,4) for molecules immobilized on the coverslip surface. The typical values for α and β in our system were ~3.6 and ~0.35, respectively.

S2.4.2. Calculation of compensated intensity

In single-molecule FRET measurements, fluorescence signals are modulated not only by FRET changes, but also by unexpected photochemical effects, such as blinking or quenching. One may

expect that the FRET changes do not affect the total fluorescence intensity ($I_A + I_D$) and therefore the real FRET dynamics can be distinguished from unexpected fluctuations. However, it is not correct because I_A and I_D are affected by the quantum yield, the detection efficiency and the fluorescence leakage as mentioned above. Here, we introduce the compensated fluorescence intensity I_C:

$$I_C = \frac{I_A}{\alpha} + \left(1 - \frac{\beta}{\alpha}\right) I_D , \tag{S60}$$

where α and β are identical to those used in Eq. (S59). I_C is proportional to the total excitation energy given to the donor and is kept constant while fluorescence intensities are modulated only by the FRET changes.

Once we obtain fluorescence intensities, we can calculate I_C. If I_C trace fluctuates beyond the noise level, the fluctuation must be caused by some unexpected effects other than FRET changes. If we want to focus only on the FRET dynamics, we can discard such data selectively. We can choose data showing constant I_C except for shot-noise fluctuations as shown in Fig. 6 *C* and make further detailed analysis of FRET dynamics.

S3. SUPPLEMENTARY DATA

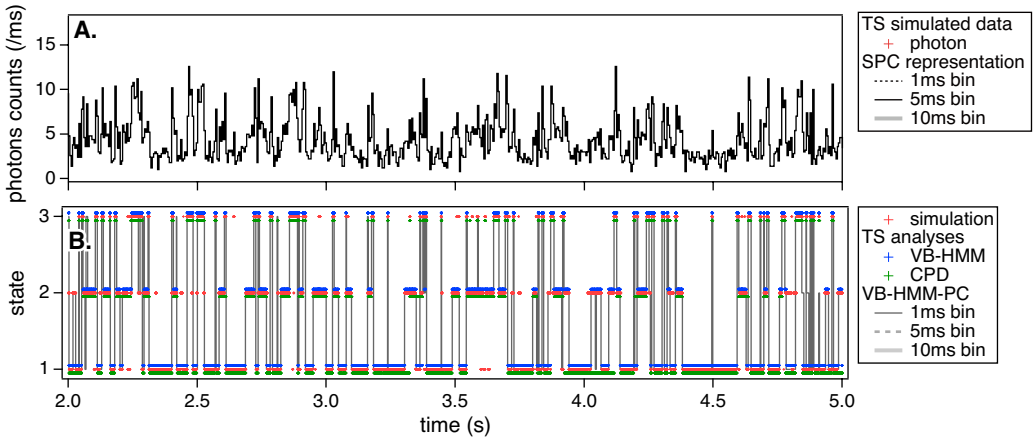

Figure S1: Example results of analyses of an intensity signal. Full length data of Fig. 2. (*A*) The SPC-signal with 5 ms time bin (black) is converted from a simulated TS signal. (*B*) The trajectories of the state transition. The original simulated data (red), the results of VB-HMM-TS (blue) and CPD (green) assign a state to each photon. The VB-HMM-PC result (gray; only result with 1 ms bin is plotted) is time bin-based. $I_{\mathrm{ratio}} = 0.5$ and $\lambda_i = \{20, 10, 5\}$ ms were applied for this simulation.

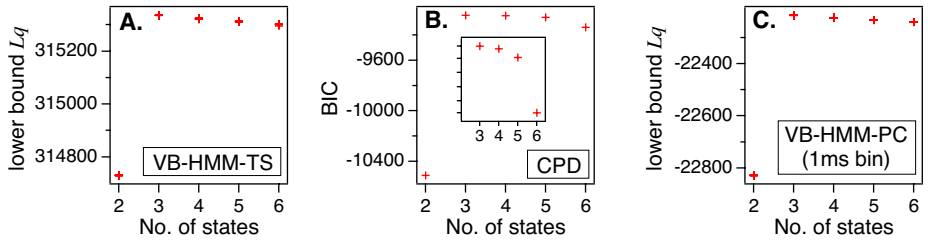

Figure S2: Example results of inference scores (the variational lower bound for VB-HMM methods, the Bayesian information criterion for CPD) obtained for the data in Fig. S1. Dependence on the assumed NoS is shown for (*A*) VB-HMM-TS, (*B*) CPD and (*C*) VB-HMM-PC with a 1 ms bin, respectively. The inset of (*B*) emphasizes the difference at NoS ≥ 3. All analyses gave a maximum value at NoS of 3.

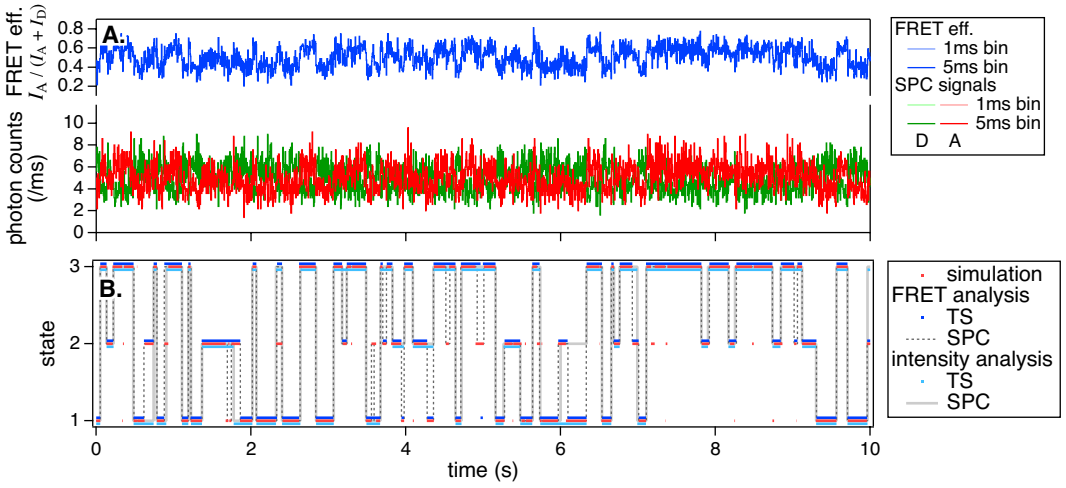

Figure S3: Example results of analyses of a dual-channel FRET signal. Full length data of Fig. 3. (*A*) The SPC-signals with 5 ms time bin (donor: green, acceptor: red) are converted from the simulated TS signal. The FRET trajectory (blue) is calculated from the SPC trajectories. (*B*) The trajectories of the state transition. The original simulated data (red), the results of VB-HMM-TS-FRET (blue) and VB-HMM-TS (light blue) assign a state to each photon. The results of VB-HMM-PC-FRET (dashed gray) and VB-HMM-PC (light gray) only with 1 ms bin are shown. VB-HMM-TS/-PC analyses treat only the donor photons. $\Delta E = 0.1$, $I = 10{,}000$ and $\lambda = 100$ ms for all states are given for this simulation.

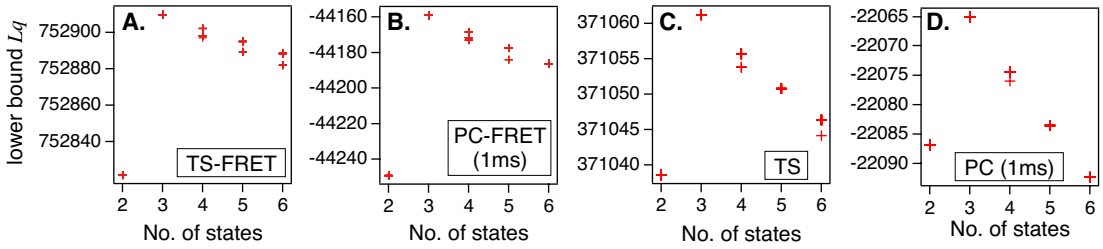

Figure S4: Example results of the variational lower bounds obtained for the data in Fig. S3. Dependence on the assumed NoS is shown for (*A*) VB-HMM-TS-FRET, (*B*) VB-HMM-PC-FRET (1 ms bin), (*C*) VB-HMM-TS and (*D*) VB-HMM-PC (1 ms bin), respectively. All analyses gave a maximum value at NoS of 3.

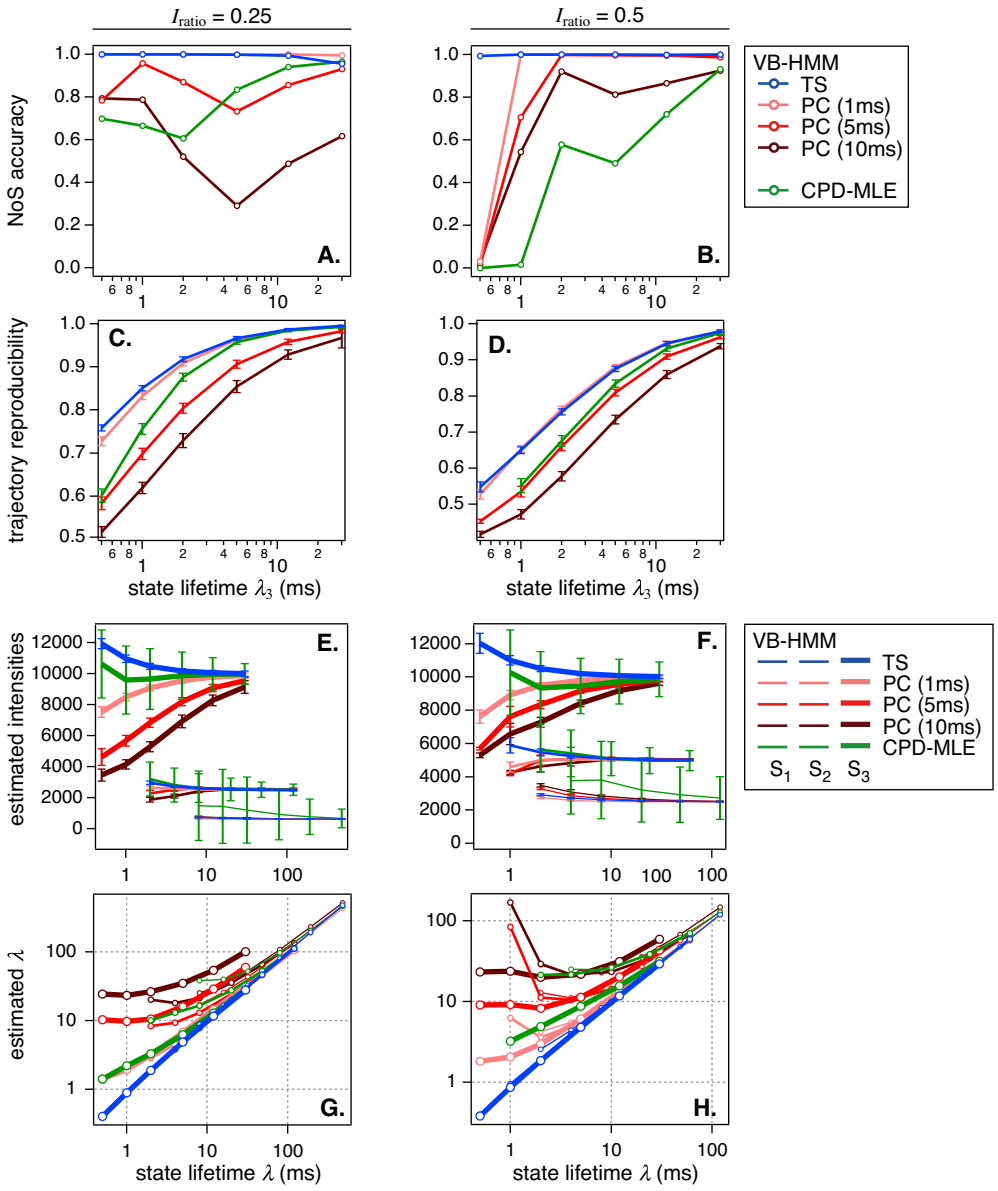

Figure S5: Statistics of an intensity analysis on 1000 Monte-Carlo-simulated time series data with I_{ratio} = 0.25 and 0.5. (*A*)(*B*) The accuracy of the NoS estimation. VB-HMM-TS achieves almost 100% accuracy over this parameter range, while VB-HMM-PC with 1 ms bin is close. (*C*)(*D*) The accuracy in reproducing the state transition trajectory. VB-HMM-TS and VB-HMM-PC (1 ms) are again superior to others. CPD, another TS-based analysis, is equivalent at large λ. (*E*)(*F*) The accuracy of parameter estimation of the intensities. (*G*)(*H*) The accuracy of parameter estimation of the transition rates given by the state lifetim λ. Estimated λ should be same as the assumed λ given for simulation. In that sense, VB-HMM-TS shows almost perfect results. Parameter estimation seems sufficiently accurate in the parameter range giving high NoS accuracy. Statistics of (*C*–*H*) are evaluated from results that estimated NoS correctly. Error bars designate the standard deviation.

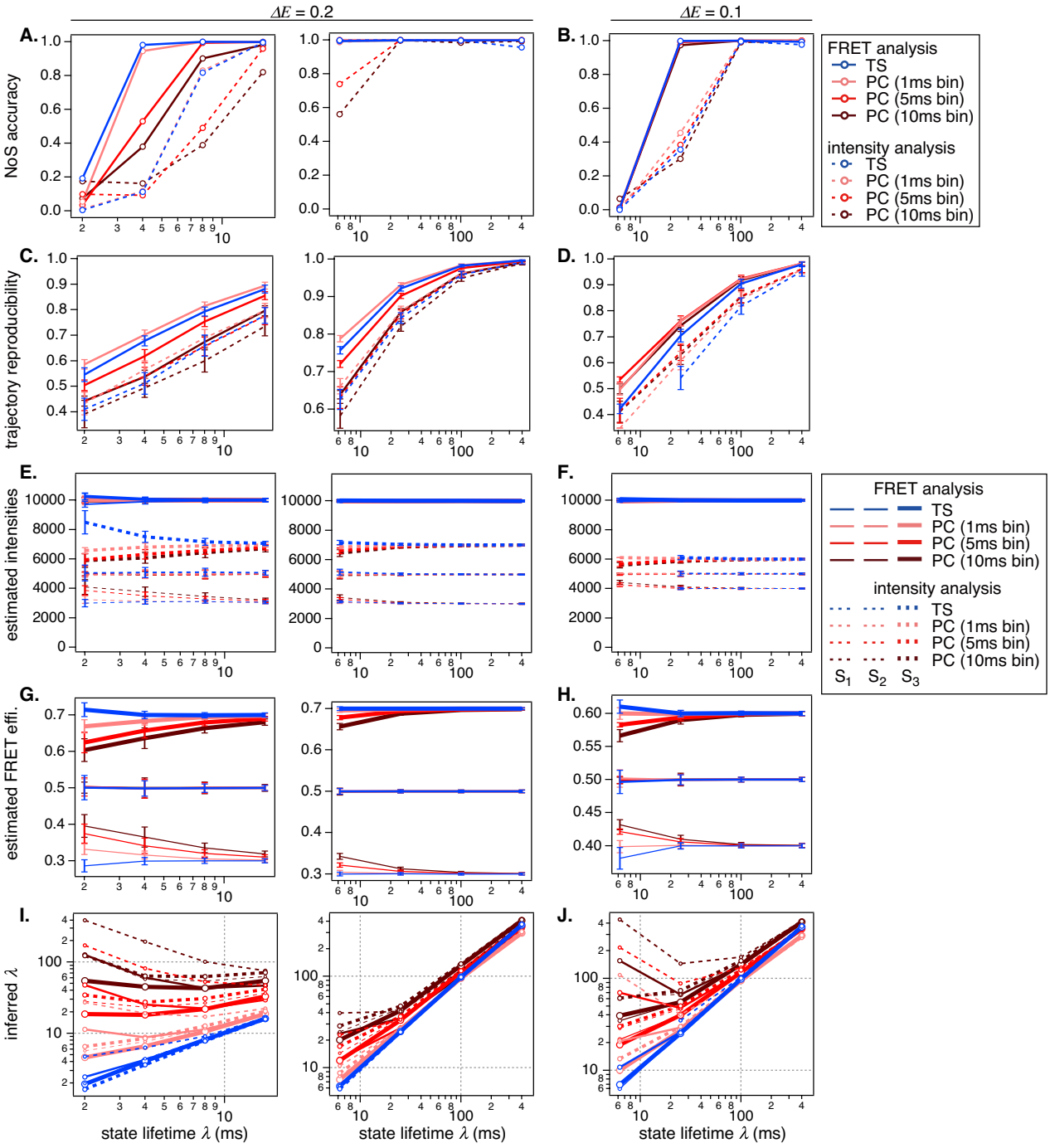

Figure S6: Evaluation of the FRET signal analysis on 1000 Monte-Carlo-simulated time series data with $\Delta E = 0.2$ and 0.1. Solid lines show the results of the FRET analyses, while dashed lines are of the intensity analyses. (*A*)(*B*) The accuracy of the NoS estimation. VB-HMM-TS-FRET as well as VB-HMM-PC-FRET (1 ms) show the best performance. (*C*)(*D*) The reproducibility of the state transition trajectory. VB-HMM-PC-FRET (1 ms) appears to be the best, while VB-HMM-TS-FRET is close to it. (*E*)(*F*) Estimation of the intensities. It is natural that FRET analyses estimate the intensities exactly, because I was constant during simulation. (*G*)(*H*) Estimation of the FRET efficiencies. Estimation is almost exact when the time bin is not too large and λ is not too small. (*I*)(*J*) Estimation of the transition rates, given by the state lifetime λ. VB-HMM-TS-FRET again shows almost perfect results. Statistics of (*C*–*J*) are evaluated from results that estimated NoS correctly. Overall, the FRET analyses are superior to the intensity-based analyses. Error bars designate the standard deviation.

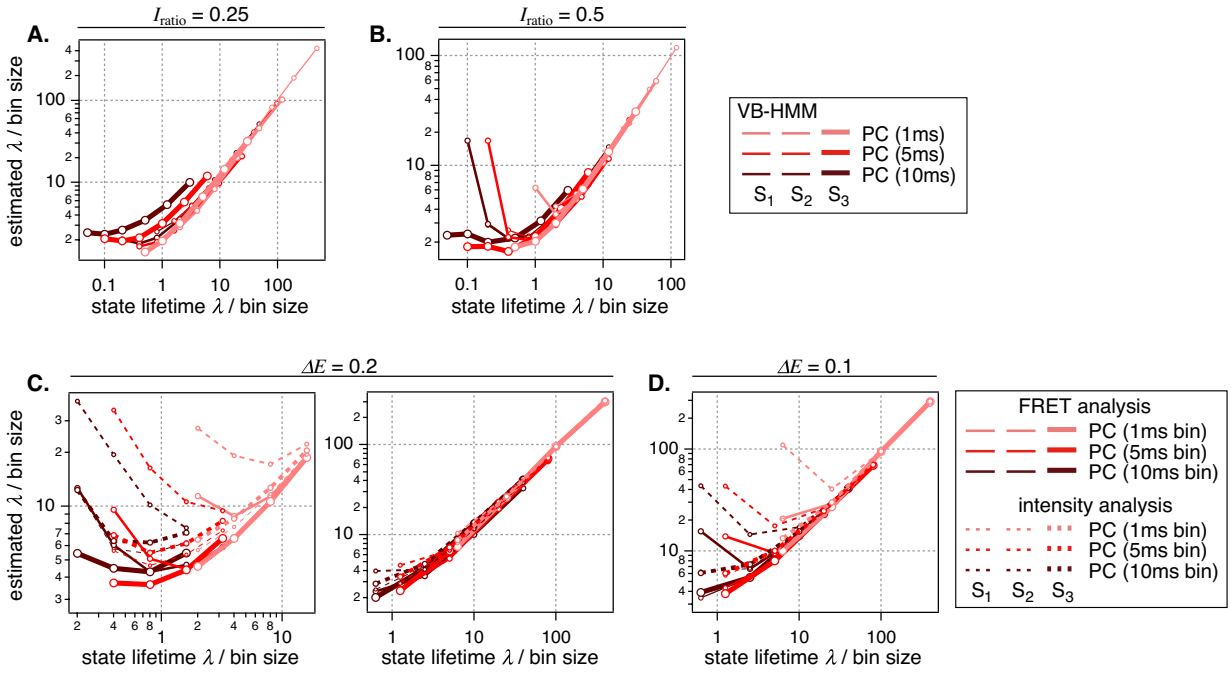

Figure S7: Results of the state lifetime λ estimation are replotted from (*A*) Fig. S5 *G*, (*B*) Fig. 4 *C* (S5 *H*), (*C*) Fig. S6 *I* and (*D*) Fig. 5 *C* (S6 *J*), respectively, so that λ is rescaled as the ratio to the bin size on both the horizontal and the vertical axes. In each graph, the curves almost overlap, which indicates that correctness of analysis seems to depend on the relative bin size to the state lifetime rather than the absolute bin size. When the ratio is large enough (> ~10), the curves are on the diagonal lines, meaning correct estimations. With the ratio getting smaller than 10, the curves begin to deviate from the diagonal lines and the errors are significant where the ratio < 1. There are some curves showing the substantial deviation from the diagonal line even with the ratio > 1. Most of them are VB-HMM-PC results for state 2, which is the intermediate state among three states. Those errors are presumably caused by frequently missed short-lived stays of the state 2.

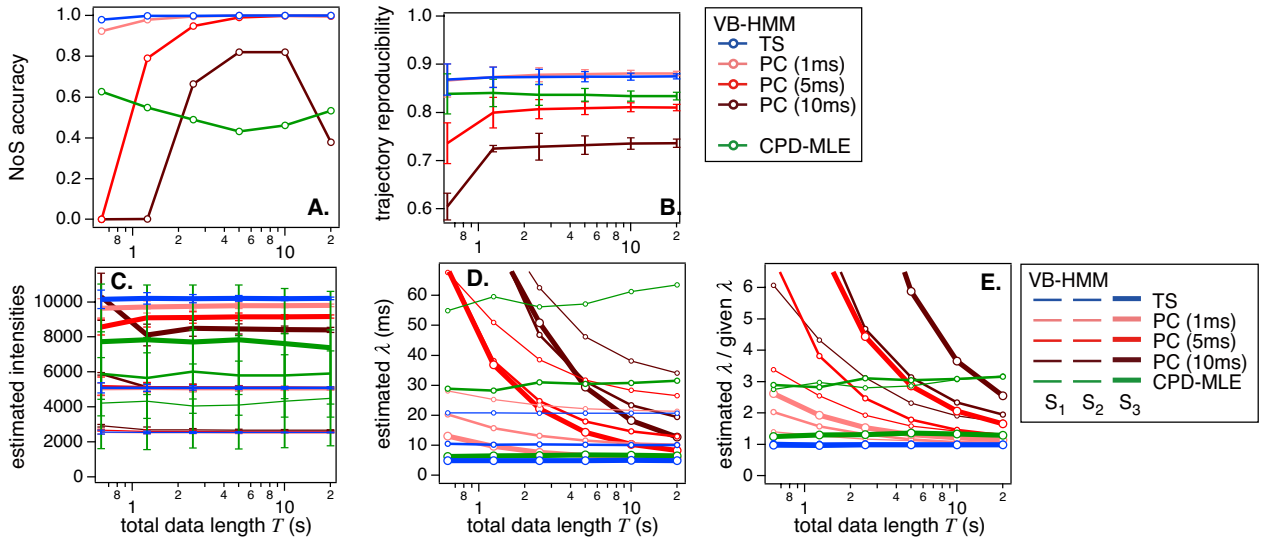

Figure S8: Dependence of intensity analyses on the total signal length T. 1000 Monte-Carlo-simulations were conducted with conditions described in S.2.1.2. (*A*) The accuracy of the NoS estimation. (*B*) The reproducibility of the state transition trajectory. (*C*) The accuracy of parameter estimation of the intensities. (*D*) The accuracy of parameter estimation of the transition rates given by the state lifetime, λ, which should be $\{\lambda_1, \lambda_2, \lambda_3\} = \{20, 10, 5\}$ ms. (*E*) The estimated λ is normalized by the correct λ given to simulation.

The shortest $T = 0.625$s, at which each state appears ~20 times in average, is long enough to give the correct result for VB-HMM-TS. There seems to be little dependence on T except for VB-HMM-PC with large bins. It may be noticeable that estimation of λ by VB-HMM-TS again appears to be perfect.

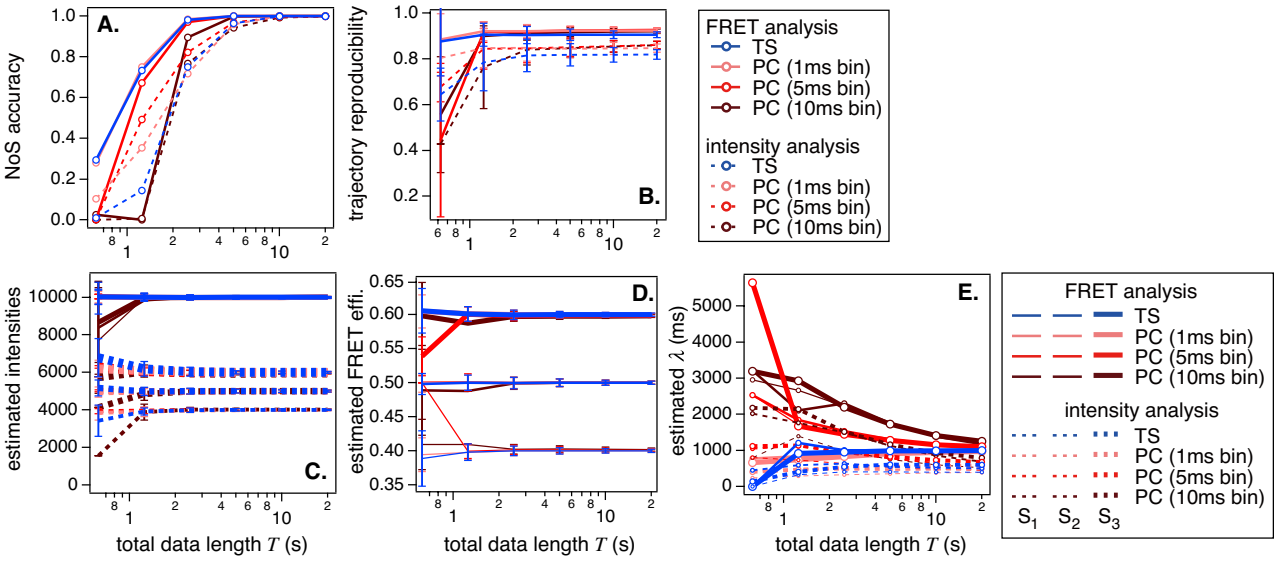

Figure S9: Dependence of FRET analyses on the total signal length T. 1000 Monte-Carlo simulations were conducted with conditions described in S.2.1.2. (*A*) The accuracy of the NoS estimation. (*B*) The reproducibility of the state transition trajectory. (*C*) Estimation of the intensities. (*D*) Estimation of the FRET efficiencies. (*E*) Estimation of the transition rates, given by the state lifetime, λ, which should be 1000 ms.

The NoS estimation become severely erroneous when $T < 2$ s, within which each state appears < 7 times on average. Because the state transition is stochastically generated during simulation, some of these signal traces may not include all three states. The resultant incompleteness of the signal presumably causes part of the error. VB-HMM-TS-FRET and VB-HMM-PC-FRET with a 1 ms bin appears to be satisfactory with $T > 2$ s, i.e. a signal longer than a few seconds is sufficient.

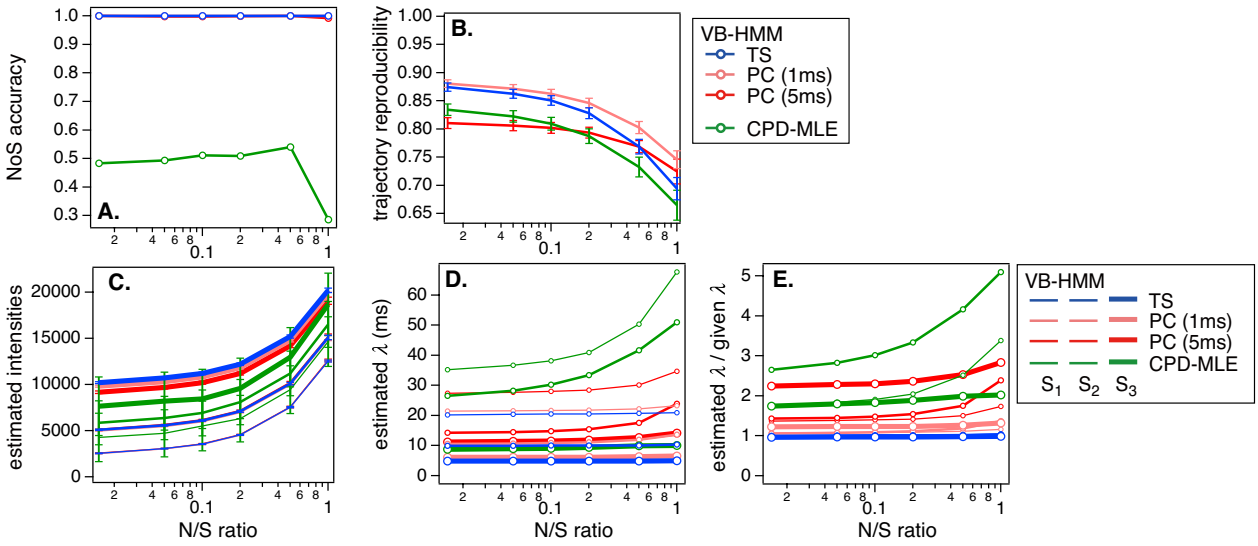

Figure S10: Dependence of intensity analyses on the noise-to-signal (N/S) ratio. 1000 Monte-Carlo-simulations were conducted with conditions described in S.2.1.3. The leftmost points in the graphs represent results with a N/S ratio = 0. (*A*) The accuracy of the NoS estimation. (*B*) The reproducibility of the state transition trajectory. (*C*) Estimation of the intensities. Estimated values increase with N/S ratio because the analyses estimate the apparent intensities including noise photons. (*D*) Estimation of the transition rates, given by the state lifetime, λ, which should be $\{\lambda_1, \lambda_2, \lambda_3\} = \{20, 10, 5\}$ ms. (*E*) The estimated λ is normalized by the correct λ given to simulation.

There appears to be little influence from noise, except for a gradual degradation in the trajectory reproducibility.

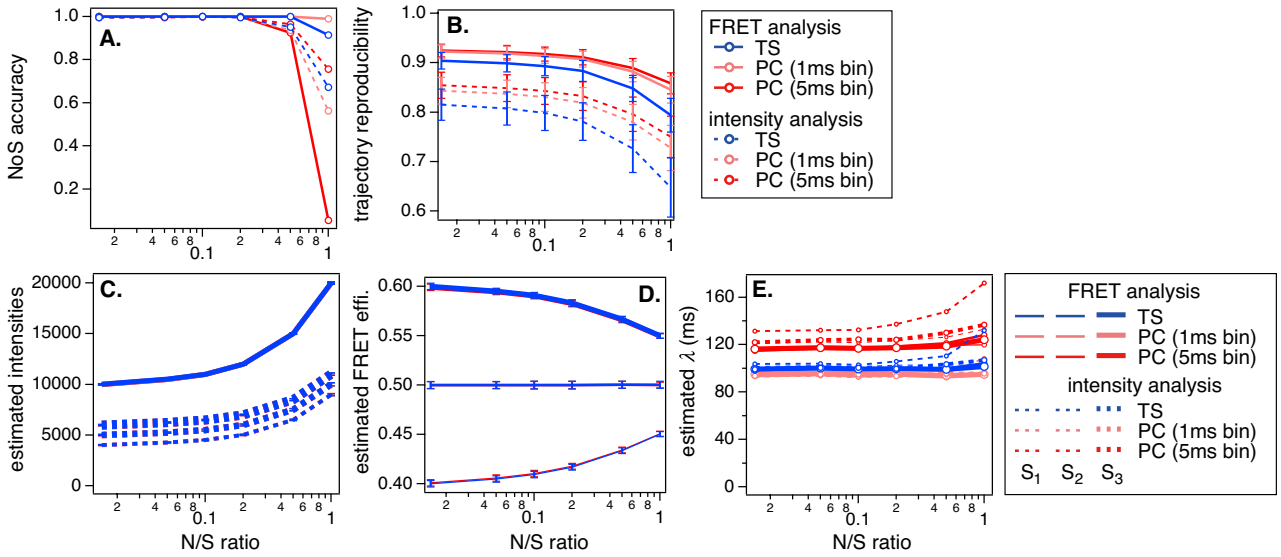

Figure S11: Dependence of FRET analyses on the N/S ratio. 1000 Monte-Carlo-simulations were conducted with conditions described in S.2.1.3. The leftmost points in the graphs represent results with a N/S ratio = 0. (*A*) The accuracy of the NoS estimation. (*B*) The reproducibility of the state transition trajectory. (*C*) Estimation of the intensities. (*D*) Estimation of the FRET efficiencies. Estimated intensities and FRET efficiencies change with N/S ratio because the analyses estimate the apparent values of those parameters, including noise photons. (*E*) Estimation of the transition rates, given by the state lifetime, λ, which should be 1000 ms.

There appears to be little influence from noise, except for a gradual degradation in the trajectory reproducibility, especially for VB-HMM-TS-FRET and VB-HMM-PC-FRET with a 1 ms bin.

Because, under our experimental conditions, the N/S ratio was typically less than 5%, noise may not be worth considering. However, care should be taken that the estimated parameters (intensity and FRET efficiency) include the counts originating from noise and should be compensated for later, if necessary.

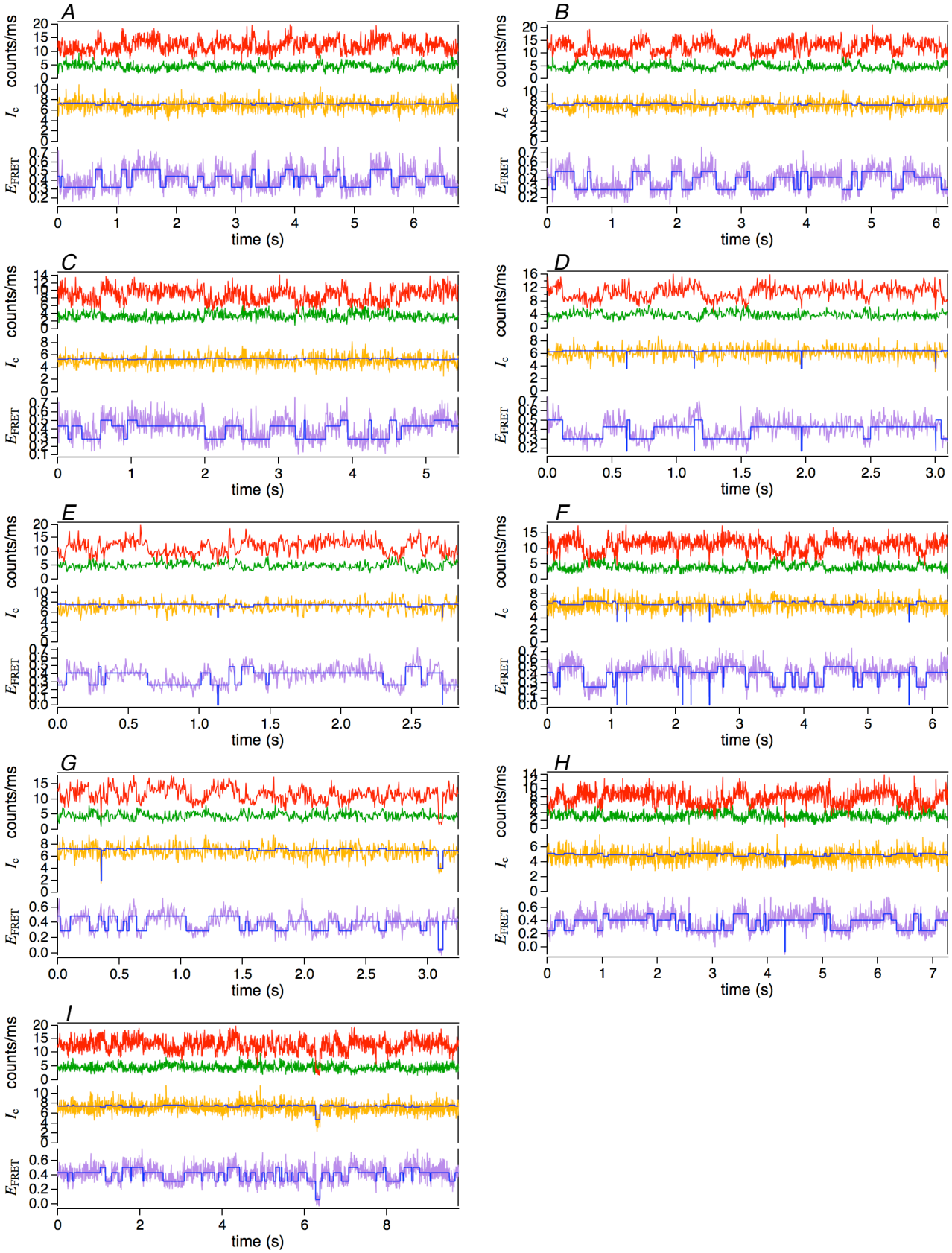

Figure S12: Examples of experimental FRET data from single HJs in the PC-representation with a bin size of 5 ms. Donor (green) and acceptor (red) fluorescence signals, compensated intensities (yellow) and FRET efficiencies (purple). VB-HMM-TS-FRET results (blue) are also shown. (*A*) is identical to Fig. 6 *B–D*.

index	intensity	FRET eff.	transition rate
3	6.47±0.96	0.502±0.011	15.6±5.8
2	6.64±0.96	0.429±0.013	8.9±4.1
1	6.85±0.97	0.286±0.027	11.6±5.5

Table S1: VB-HMM-TS-FRET analysis result: Estimated parameters for states reproduced from the smFRET observations of HJ branch migration shown in Figs. S12.

SUPPORTING REFERENCES

1. Bishop, C. M. 2006. Pattern recognition and machine learning. Springer Science+Business Media, LLC, New York, USA.
2. Bronson, J. E., J. Fei, J. M. Hofman, R. L. Gonzalez Jr., C. H. Wiggins. 2009. Learning rates and states from biophysical time series: A Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* 97:3196–3205.
3. Ha, T., A. Y. Ting, J. Liang, W. B. Caldwell, A. A. Deniz, D. S. Chemla, P. G. Schultz, and S. Weiss. 1999. Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc. Natl. Acad. Sci.* 96:893–898.
4. Sabanayagam, C. R., J. S. Eid, A. Meller. 2004. High-throughput scanning confocal microscope for single molecule analysis. *Appl. Phys. Lett.* 84:1216–1218.
5. Watkins, L. P., and H. Yang. 2005. Detection of intensity change points in time-resolved single-molecule measurements. *J. Phys. Chem. B* 109:617–628.
6. Ensign, D. L., and V. S. Pande. 2010. Bayesian detection of intensity changes in single molecule and molecular dynamics trajectories. *J. Phys. Chem. B* 114:280–292.
7. Karymov, M., D. Daniel, O. F. Sankey, and Y. L. Lyubchenko. 2005. Holliday junction dynamics and branch migration: Single-molecule analysis. *Proc. Natl. Acad. Sci. U.S.A.* 102:8186–8191.
8. Campos, L. A., J. Liu, X. Wang, R. Ramanathan, D. S. English, and V. Munñoz. 2011. A photoprotection strategy for microsecond- resolution single-molecule fluorescence spectroscopy. *Nat. Methods* 8:143–147.