Expectation-Maximization of the Potential of Mean Force and Diffusion Coefficient in Langevin Dynamics from Single-Molecule FRET Data Photon by Photon

Kevin Haas,[†] Haw Yang,[*,‡] and Jhih-Wei Chu[*,†,¶,§]

University of California-Berkeley, Department of Chemical and Biomolecular Engineering, Berkeley, CA 94720, USA, Princeton University, Department of Chemistry, Princeton, NJ 08544, USA, Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan, and Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan

E-mail: hawyang@princeton.edu; jwchu@nctu.edu.tw

Abstract

The dynamics of a protein along a well-defined coordinate can be formally projected onto the form of an overdamped Lagevin equation. Here, we present a comprehensive statistical-learning framework for simultaneously quantifying the deterministic force (the potential of mean force, PMF) and the stochastic force (characterized by the diffusion coefficient, D) from single-molecule Förster-type resonance energy transfer (smFRET) experiments. The likelihood functional of the Langevin parameters, PMF and D, is expressed by a path integral of the latent smFRET distance that follows Langevin dynamics and realized by the donor and the acceptor photon emissions. The solution is made possible by an eigen decomposition of the time-symmetrized form of the corresponding Fokker-Planck equation coupled with photon statistics. To extract the Langevin parameters from photon arrival time data, we advance the expectation-maximization algorithm in statistical learning, originally developed for and mostly used in discrete-state systems, to a general form in the continuous space that allows for a variational calculus on the continuous PMF function. We also introduce the regularization of the solution space in this Bayesian inference based on a maximum trajectory-entropy principle. We use a highly nontrivial example with realistically simulated smFRET data to illustrate the application of this new method.

Introduction

A fundamental property of biomolecules such as proteins is their conformational flexibility which allows for a diverse set of physical and chemical processes. The physical origin of the dynamics generally consists of two components—the deterministic mean forces as a function of configurational variation and the stochastic forces due to the thermal energy and environmental noises. Resolving the manner by which these two components contribute to governing dynamical behaviors is thus at the core of elucidating the structure-dynamics-function relationship of protein conformational changes.

The direct observation of individual proteins is possibly the most straightforward way of dissecting the forces that drive their conformational dynamics. From the physical chemistry view point, therefore, the objective of a single-molecule analysis is to clarify and to quantify from the measured experimental data the dynamics parameters of the probed degree of freedom. Ideally, one would like to capture both the underlying free-energy landscape of the protein conformation and the stochastic diffusion coming from thermal fluctuations and interactions with the unobserved degrees of freedom. Indeed, the direct observation of the distribution and the dynamics that a molecular system exhibits—which are scrambled in ensemble-averaged experiments—are the two unique pieces of information that only single-molecule experiments can provide.[1,2] Yet, despite the vigorous development of single-molecule spectroscopy to date,[3–5] the quantitative determination of single-molecule dynamics (not *kinetics*) has not been achieved.

For an explicit illustration of the challenges in analyzing single-molecule experiments, let us consider the time-dependent single-molecule Förster-type resonance energy transfer (smFRET) experiment of a protein.[6–8] A typical setup uses a pair of fluorescent dyes, a donor and an acceptor, to attach to the ends of a surface-immobilized protein, Fig. 1. Following laser excitation, an electronically excited donor dye can relax to its ground state by emitting a "green" photon or by transferring the energy to the nearby acceptor dye that may then emit a "red" photon to go back to its ground state. Under favorable circumstances,[9] the energy-transfer efficiency between dyes depends on the donor-acceptor distance r as $\zeta(x) = 1/(1+x^6)$ with $x = r/R_0$ and R_0 being the Förster radius for the acceptor-donor pair at which the energy-transfer efficiency is 50%. In this case, the donor-acceptor distance r, or equivalently, x, is a measure of the protein conformation and is the experimentally accessible degree of freedom onto which the dynamics of the protein is projected. The dynamics of x is naturally stochastic owing to the omnipresent thermal agitations from the experimentally unaccessible protein degrees of freedom and the environment—the first layer of stochasticity in a single-molecule experiment comes from thermal fluctuations.

The signals from an smFRET experiment are the photons emitted from the tagged protein, which can be captured by confocal microscopy and recorded by a pair of avalanche photodiodes.[10] The statistics of photon arrival times follows that of a Poisson process with the emission intensity depending on x parametrically:

$$I_a(x) = I_a^0 \zeta(x) + B_a \quad \text{(acceptor)} \tag{1}$$

$$I_d(x) = I_d^0 (1 - \zeta(x)) + B_d \quad \text{(donor)}. \tag{2}$$

*To whom correspondence should be addressed

†University of California-Berkeley, Department of Chemical and Biomolecular Engineering, Berkeley, CA 94720, USA

‡Princeton University, Department of Chemistry, Princeton, NJ 08544, USA

¶Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan

§Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan

Here, $I^0_{a,d}$ are the maximum intensities of the donor (subscript d) and acceptor (subscript a), and $B_{a,d}$ are the background signals including the donor-acceptor cross talk.[11] The background, cross talk, and the Poisson photon-counting statistics represent the three main sources of the apparent "noise" in fluorescence single-molecule signals. Therefore, the "noise" is explicitly taken into account in the theoretical framework as well in the numerical simulations presented in this work.

Since the arrival time of each emitted photon is recorded, the waiting-time distributions of the acceptor (Δt_a) and donor (Δt_d) photons follow the exponential probability density function with intensities $I_{a,d}$ describing the coupling between the latent variable, x, and the observed signals, Δt_a and Δt_d, through photon statistics:

$$p(\Delta t_{a,d}|I_{a,d}) = I_{a,d}e^{-I_{a,d}\Delta t_{a,d}}. \quad (3)$$

More specifically, within an infinitesimal time slice dt, one of the three observations would occur with the probability densities depending on the latent variable of the system state at the moment, x_t, which involve the parameter of total intensity defined as $I(x_t) = I_d(x_t) + I_a(x_t)$:

1. an acceptor photon arrives, and the probability density of this event is:

$$p(\Delta t_a = dt|x_t)p(\Delta t_d > dt|x_t) = I_a(x_t)e^{-I(x_t)dt}; \quad (4)$$

2. a donor photon arrives, and the probability density of this event is:

$$p(\Delta t_d = dt|x_t)p(\Delta t_a > dt|x_t) = I_d(x_t)e^{-I(x_t)dt}; \quad (5)$$

3. no photon arrives, in which the particular dt instance is considered "dark," and the probability of this event is:

$$P(\Delta t_a > dt, \Delta t_d > dt|x_t) = e^{-I(x_t)dt}. \quad (6)$$

The probability of observing both acceptor and donor photons in $dt \to 0$ is extremely small and this event is hence ignored. The information of protein dynamics along x is encoded in the sequence of the colors and arrival times of photons that depend on the system state probabilistically according to Eqs. 4-6. The photon-detection statistics adds another layer of stochasticity to the quantification of single-molecule dynamics from smFRET. With the two layers of stochasticity explained above—the thermal fluctuations and the random photon-detection events—the core difficulty of learning the protein dynamics along x through such indirect measurements as smFRET is now apparent: There is no explicit probabilistic structure to relate the measured data with the dynamics parameters that characterize the time propagation of the latent variable.

In principle, the dynamics of the latent variable x can be recovered using a Bayesian-inference model. However, such developments have so far been limited to a coarse-grained description in which the system dynamics are treated as "jumps" between discrete states.[12–18] In fact, it is necessary to assume an *ad hoc* number of states in order to construct a Bayesian-inference model. Though the number of stable states along the x coordinate is in general unknown *a priori*, and is actually one of the primary goals of a single-molecule analysis. Furthermore, the discrete-state and the "jump" assumptions imply a timescale separation in that the each "jump" is considered instantaneous and that the dynamics within each discrete state is ignored. Treatments like this in essence mix the contributions from the deterministic and the stochastic forces of protein dynamics into a rate constant matrix connecting different states. As a result, the dynamics are completely omitted and the ensemble averaged *kinetics* is obtained instead. Analysis methods that are objective and driven by data rather than by the more subjective modeling would be more satisfactory for learning continuous stochastic dynamics from indirect measurements for the model-independent methods could afford unexpected discoveries from the otherwise noisy single-molecule data.

The Maximum Information Method (MIM) is one such approach that explicitly takes into account photon-counting statistics but does not require any presumed model about the underlying dynamics or modes of states.[11] For a given time-stamped photon trajectory, the method operates under the assumption that the unknown x is stagnant until a sufficient number of photons is collected such that the latent variable can be evaluated with a satisfactory precision. In other words, the photon trajectory is binned adaptively (information binding) to produce a distance-time trajectory (thus the x dynamics is followed) in which all the distance measures have the same uncertainty related to photon-counting statistics. This approach in turn permits the quantitative removal of the photon-counting uncertainty in the distance histogram to unambiguously determine the number of states as well the quantitative evaluation of the entire distribution.[19] Although this approach is general, free from limiting x to a set of discrete states, and readily applicable to processing experimental data,[10,20] resolution loss is inevitable because of the coarse-graining to a single distance value within each time bin, thereby limiting the capacity to rigorously quantify the dynamics. In fact, any binning of the trajectory (time averaging) will inevitably lose information for the dynamics. Furthermore, it is generally difficult to quantify dynamics from fluorescent single-molecule data based on the correlation-based approach because of the poor statistics due to limited trajectory lengths.[21,22] Nevertheless, the maximum-information method represents the current state-of-the-art in the quantitative evaluation of distance fluctuations in smFRET experiments, and has allowed the direct comparison with molecular mechanics modeling[23] as well the development of empirical force fields for coarse-grained modeling.[24] Therefore, it is used for comparing with the results of the new path-integral statistical learning method presented here.

From the discussions above, it is clear that the dynamics and the quantitative evaluation thereof from smFRET data are the key missing pieces toward realizing the full potential of time-dependent single-molecule spectroscopy. The primary goal of this work is to show the feasibility of solving this problem. Particularly, we aim to go beyond the simple correlation analysis, and provide a framework for the quantitative evaluation of the deterministic and stochastic forces in a molecular system from the indirect measurement of dynamics.

To begin, one recognizes that the dynamics of x is the projection of the movements of all degrees of freedom of the system, including those of the single molecule in question and its surrounding solvent. Following Zwangzig's projection-operator formalism,[25] the dynamics of x can be described by the Langevin-type equation of motion,

$$dx_t = DF(x_t)dt + \sqrt{2D}dW_t, \quad (7)$$

in which the over-damped form implies a separation of time scale between the slow smFRET accessible x and the other fast unobserved degrees of freedom, and is consistent with

the dynamics in low Reynolds number media. In this model, The mean force $F(x) = -\nabla_x V(x)$ constitutes the deterministic component in the equation of motion. The PMF, $V(x)$, is related to the equilibrium probability density of x, $p_{\rm eq}(x)$, as $V(x) = -\ln(p_{\rm eq}(x))$. The stochastic force component is parameterized by the diffusion coefficient D. The Weiner process dW_t has a mean $\langle dW_t \rangle = 0$ and variance $\langle dW_t \cdot dW_{t'} \rangle = \delta(t-t')dt$.[1] The Langevin equation in Eq. 7 captures the spatial and temporal continuity of molecular mechanics and dynamics. The study of single-molecule dynamics thus becomes the quantitative eduction of the $F(x)$ profile and the diffusion coefficient D in Eq. 7 from the experimentally recorded photon-arrival time trajectory. In practice, however, the difficulties of two-layers of randomness in such experiments, the infinite dimensionality, the non-differentiability in time, and the path integral implied in Eq. 7 must be overcome for it to be useful for analyzing realistic experimental data.

This work presents our analytical, numerical, and statistical developments that make possible this goal by overcoming the aforementioned difficulties. Although the method was devised for the specific case of using smFRET to study protein conformational changes, the foundation for statistical leaning of continuous stochastic dynamics established here may also be applicable to other single-molecule methods such as pulling using atomic force microscope or optical tweezer through molecular tags to transmit forces. Since the free-energy landscape and diffusion coefficient of conformational dynamics can also be constructed from the bottom up via computer molecular dynamics simulations with path-based methods of sampling and optimization,[26,27] the availability of the same type of data from experiments could greatly facilitate experiment-theory cross-validation, tracing the atomistic origin of protein dynamics and ultimately the control thereof.

The rest of this paper is organized as the following. We first present the Bayesian inference framework we employed for the statistical learning of Langevin dynamics from smFRET data. Theoretical developments for calculating the likelihood functional of PMF and D through a trajectory path integral are presented next. This procedure can also be used to infer the probability densities of the latent trajectory, $X(t)$, from a recorded photon trajectory, $Y(t)$. $X(t)$ is a continuous function of time that gives the value of x at a specific time t, i.e., x_t. On the other hand, $Y(t)$ is a function of time that gives discrete outcomes. At a specific time, the readout of the photon trajectory, y_t, is either a donor photon, an acceptor photon, or darkness. We then derive the functional derivatives of the likelihood function with respect to the Langevin dynamics parameters given the observed photon trajectory. With these elements established, an expectation-maximization (EM) optimization of the Langevin model can be devised to deduce the optimal PMF and diffusion coefficient that best describe the data of photon sequence. This work thus advances the applicability of the EM statistical learning algorithm from discrete-state systems to extracting a continuous profile from data. Application of this method to a highly non-trivial test case is presented at the end.

[1] Throughout the text, the physical variables presented are nondimensionalized by the thermal energy $k_{\rm B}T$ at a fixed temperature T as the characteristic energy and the Föroster radius R_0 as the characteristic length. That is, $\bar{V}(r/R_0)/k_{\rm B}T \to V(x)$, $\bar{F}(r/R_0)R_0/k_{\rm B}T \to F(x)$, $\bar{D}/R_0^2 \to D$, and $\overline{I_{a,d}}(r/R_0) \to I_{a,d}(x)$. Variables with an overbar are the actual quantities before nondimensionalization. D and $I_{a,d}(x)$ have the unit of $\rm s^{-1}$.

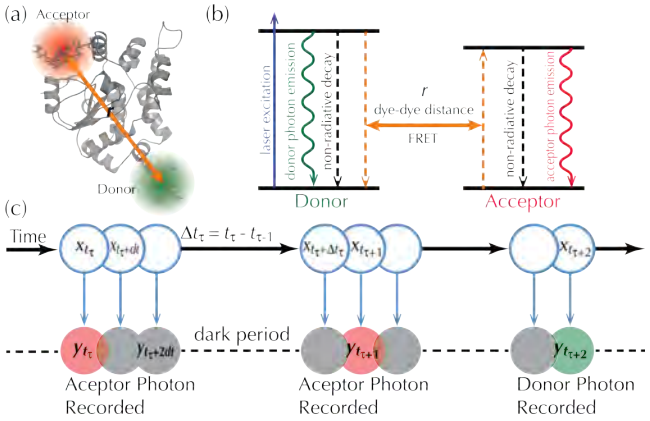

Figure 1: A schematic representation of photon-by-photon smFRET experiments. (a) A typical smFRET experimental scheme for a protein using the dye-attached structure of *Mycobacterium tuberculosis* protein tyrosine phosphatase, PtpB,[20] as a generic example for graphical illustration purposes. (b) The Jablonski diagram of the energy states in the FRET flourophores and the energy transfer event. The efficiency of energy transfer depends on the inter-dye distance r and the Förster radius R_0. The dimensionless distance x is r/R_0. (c) The graphical model of the continuous Bayesian inference for Langevin dynamics from smFRET measurements. Clear circles represent the latent dynamic variables of the system trajectory, $X(t)$ that gives the value of x at a specific time t, i.e., x_t. The filled circles represent $Y(t)$, the experimentally realized and recorded photon trajectory. At a specific time, the readout of the photon trajectory, y_t, is either a donor photon, an acceptor photon, or darkness. Horizontal arrows represent the conditional probability densities of the time evolution of x, $p(x_{t+dt}|x_t)$, and vertical arrows represent the probabilities of photon emission, $p(y_t|x_t)$.

The Bayesian Inference Framework for Continuous Stochastic Dynamics with smFRET

The trajectory of the tagged protein degree of freedom, $X(t)$, is not observed directly. The statistics regarding the PMF profile and diffusion coefficient are thus not explicit in the photon trajectory. The structure of this convolution is best represented via a Bayesian Graphical Model (BGM) as shown in Fig. 1c. The vertical arrows in the BGM link the conditional probability density of $p(y_t|x_t)$ with the experimental observable at time t, y_t=donor, acceptor, or darkness, and the latent protein conformation variable at the same time, x_t. Following Eqs. 4–6, there are two classes of observations. The instantaneous event of observing a photon is represented by taking the limit of $dt \to 0$, and the position-dependent probability density functions of $p(y_t|x_t)$ are:

$$p(y_t|x_t) = \begin{cases} I_a(x_t) & y_t = \text{acceptor photon} \\ I_d(x_t) & y_t = \text{donor photon.} \end{cases} \tag{8}$$

Alternatively, if the state of darkness was observed over the infinitesimally small, but nonzero interval dt, performing time integration in the BGM framework paints a dark duration of the specified size along the trajectory. This observation also depends on x with the probability of:

$$P(y_t|x_t) = \left\{ e^{-I(x_t)dt} \quad y_t = \text{darkness.} \right. \tag{9}$$

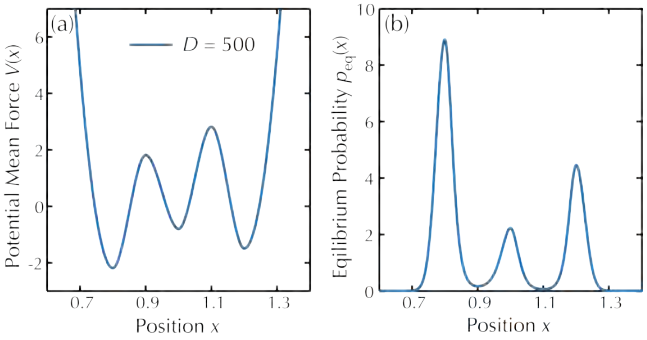

Figure 2: The test case potential. (a) The potential of mean force and (b) the corresponding equilibrium probability density of x used for the simulated smFRET trajectories.

We note that it is important to explicitly incorporate the intermediate times of "dark" periods because the "dark" periods also carry information about the latent dynamics. The horizontal arrows in the BGM of Fig. 1 indicate the conditional probability densities $p(x_{t+dt}|x_t)$ for the time propagation of the indirectly observed degree of freedom.

With this construction, the inversion of smFRET data into PMF and D comes down to solving the following two problems consecutively in an iteration loop:

Inference What is the probability density of the dynamic trajectory of the protein degree of freedom of interest, i.e., $X(t)$, given a sequence of photon arrival times and colors recorded via smFRET, $Y(t)$? In other words, with a trial mean force profile $F(x)$ and diffusion coefficient D of the Langevin equation, one aims to calculate:

$$\mathcal{P}(X(t)|Y(t);F(x),D). \tag{10}$$

Optimization What is the optimal profile of the mean force $F(x)$ and diffusion coefficient D for describing the observed photon trajectory? The answer is finding the supremum of (maximizing) the likelihood functional:

$$\sup_{F(x),D} \mathcal{P}(Y(t);F(x),D). \tag{11}$$

Solving the inference and optimization problems stated above requires a path integral over the coordinate space of the probability density of a system trajectory $X(t)$ given the smFRET observation of $Y(t)$:

$$\mathcal{P}(Y(t);F(x),D) = \int \mathcal{D}X(t)\mathcal{P}(X(t),Y(t);F(x),D). \tag{12}$$

The differential volume of the trajectory space is $\mathcal{D}X(t)$. The integrand is the joint probability density of $X(t)$ and $Y(t)$ given by the BGM of Fig. 1, which is also the complete likelihood functional of $F(x)$ and D. The theoretical development presented later illustrates how to perform such calculations given the time stamps and colors of the recorded photons.

Based on the BGM of Fig. 1 and the conditional independencies of the Markov probabilities prescribed therein, the complete likelihood functional can be factorized as:

$$\mathcal{P}(X(t),Y(t);F(x),D) = \mathcal{P}(Y(t)|X(t))\,\mathcal{P}(X(t);F(x),D). \tag{13}$$

The capability of performing smFRET inference with the entire continuous profile of $F(x)$ as the basis eliminates the requirement of subjectively assuming the number of metastable conformational states; this information would simply emerge as a result of the optimization.

Table 1: The default simulation parameters of smFRET measurements employed in this work. These values were motivated by the typically encountered conditions in experiments. N_P is the number of photons observed before the first photo bleaching event occurred and $\langle t_{exp}\rangle$ is the average duration of the trajectories with these intensities and the number of photons.

Intensity		
I_d^0	30000	s^{-1}
I_a^0	16000	s^{-1}
B_d	10	s^{-1}
B_a	20	s^{-1}
N_P	80000	
$\langle t_{exp}\rangle$	3.3	s

Without loss of generality, we perform analysis and illustration of the statistical learning algorithm with the model potential shown in Fig. 2. The PMF contains two barriers of around 5 k_BT, a magnitude that is biologically relevant for protein conformational changes. The two barriers connecting two metastable states correspond to a short and long inter-dye distance with an intermediate region locating at the value of Föster radius, $x = r/R_0 \approx 1$. The diffusion coefficient is set to $D = 500\ s^{-1}$. Photon trajectories of smFRET experiments are simulated by propagating the Langevin equation with the aforementioned PMF and D coupled with a Kinetic Monte-Carlo scheme for simulating the processes of photon emission; the Supplementary Information contains more details of this numerical procedure.

Table 1 lists the default parameters for simulating typical smFRET experiments. In this work, the resolution of a photon trajectory is specified by referencing to the default values of dye intensities listed in Table 1. If the intensities are 2x (two times) of the values in Table 1, twice numbers of photon per time will be received on average, and the resolution is thus doubled. Furthermore, since the dimensionality of the likelihood functional of Eq. 12, or, the information content of the latent dynamics, is dictated by the total number of photons, the comparison of data with different resolutions was conducted with the same of number of photons. That is, the values of dye intensities timing the duration of trajectory recording are constant. For the case of 2x resolution, $\langle t_{exp}\rangle$ would become 1.65 s as compared to that in Table 1.

Calculation of the Likelihood Functional of PMF and D and Inference of the Latent Trajectory

Eq. 13 indicates that the joint probability density of $X(t)$ and $Y(t)$ can in theory be calculated based on the Langevin equation of motion that sets $\mathcal{P}(X(t);F(x),D)$ and the waiting time distributions of experimental photon arrival events of Eq. 8 and Eq. 9. At the time instances at which a photon is detected, $\{t_\tau \mid \forall\tau \in [0,N_P]\}$ where N_P is the total number of recorded photons, the conditional probability density of Eq. 8 is used to represent the likelihood of such an event. To express the complete likelihood function, or $\mathcal{P}(X(t),Y(t))$, for a trajectory of duration t_{exp}, we employ the following

expansion based on the BGM of Fig. 1:

$$\mathcal{P}(X(t),Y(t)) = p(x_{t_0})\prod_{\tau=1}^{N_P} p(y^{a,d}_{t_\tau}|x_{t_\tau})p(x_{t_\tau},y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}|x_{t_{\tau-1}}). \quad (14)$$

The notation of $y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}$ in this equation indicates that during the time window between the arrivals of photon τ and photon $\tau-1$, $\Delta t_\tau = t_\tau - t_{\tau-1}$, the recorded observation in the smFRET experiment is darkness. On the other hand, $y^{a,d}_{t_\tau}$ denotes the photon color (acceptor or donor) observed at the instance of time t_τ. A similar construction was offered in [28] for a discrete-state Markov representation.

An important message from the above analysis is that for the statistical learning of smFRET measurements, observing the dark period of Δt_τ before receiving the τ^{th} photon also contains certain dynamics information on the latent variable. Essential to incorporating the complete information in the photon sequence for extracting continuous stochastic dynamics is thus the efficient and accurate calculation of $p(x_{t_\tau},y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}|x_{t_{\tau-1}})$ that inevitably involves a path integral.

The calculation involves dividing the dark period into $\Delta t_\tau/dt$ slices, and the coordinate at each slice is set to x_{t_i} with $i = (1,...,(\Delta t_\tau/dt-1))$ and $t_i = t_{\tau-1} + i(dt)$. One thus needs to perform integration for all time slices between the photon arrival events:

$$p(x_{t_\tau},y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}|x_{t_{\tau-1}}) = \int\cdots\int\prod_{i=1}^{\Delta t_\tau/dt-1} dx_{t_i} \; p(y^{\text{dark}}|x_{t_i})p(x_{t_i}|x_{t_{i-1}})\delta(X(\Delta t_\tau)-x_{t_\tau}). \quad (15)$$

Here, $p(x_{t_i}|x_{t_{i-1}})$ is the latent dynamics propagator over a time step dt. Taking the limit of $dt\to 0$, the dimensionality of the path integral in Eq. 15 becomes infinity. How to overcome this seemly intractable task is a key challenge in inferring the continuous latent trajectory from a recorded photon sequence. Below, we overcome this challenge by devising a scheme—the first of its kind—to evaluate $p(x_{t_\tau},y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}|x_{t_{\tau-1}})$ via Eq. 15.

Since both the Langevin equation and the dark snapshot probability (Eq. 9) do not have explicit time dependence, we seek to propagate them forward in time together for calculating the path integral of Eq. 15. Considering that $p(y^{\text{dark}}|x_{t_i}) = \exp(-I(x_{t_i})dt)$ is the exponential of a Reimann integral over time, Eq. 15 is transformed into a path expectation form:

$$p(x_{t_\tau},y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}|x_{t_{\tau-1}}) = \mathbb{E}_{X(t)}\left[e^{-\int_0^{\Delta t_\tau} dt' I(X(t'))}\delta(X(\Delta t_\tau)-x_{t_\tau}) \mid X(0)=x_{t_{\tau-1}}\right]. \quad (16)$$

Following the Feynman-Kac theorem in a similar context,[29,30] the probability density defined in Eq. 16 can be obtained by solving the following partial differential equation (PDE):

$$\frac{\partial p(x,t)}{\partial t} = \Big(D\nabla^2 - \nabla DF(x) - I(x)\Big)\,p(x,t). \quad (17)$$

Here, $p(x,t)$ is a shorthand notation of $p(x_{t_\tau},y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}|x_{t_{\tau-1}})$. The variable x at time t corresponds to x_{t_τ}, i.e., $t_\tau \equiv t$, and is the object of the gradient operators in Eq. 17. The initial distribution of probability density $p(x,0)$ represents the condition at $t_{\tau-1}$ in Eq. 16, and $t_{\tau-1}\equiv 0$.

Eq. 17 is essentially the Fokker-Plank equation of the Langevin equation of motion of Eq. 7 with the additional $I(x)$ term acting as an indicator the observation of darkness. The incorporation of the dark operator with the Langevin time propagation allows the path integral in Eq. 15 to be accomplished by solving the partial differential equation of Eq. 17. It is thus not necessary to explicitly go through the infinite dimensionality for including the information of each dark period. This development for $p(x_{t_\tau},y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}|x_{t_{\tau-1}})$ calculation is one of the critical aspects devised this work for making possible the statistical learning of continuous stochastic dynamics.

Another essential component is the recognition that a symmetric version of the PDE in Eq. 17 can drastically simplify the calculation of the likelihood functional of Eq. 12 through the path integral over $X(t)$. In particular, a new dependent variable is defined as $\rho(x,t) = p(x,t)/\sqrt{p_{\text{eq}}(x)}$ to transform the PDE to a symmetric form with the Hermitian operator of time propagation given below:

$$\frac{\partial}{\partial t}\rho(x,t) = -\boldsymbol{H}\rho(x,t) \quad (18)$$

$$\boldsymbol{H} = -D\nabla^2 + D\frac{\nabla F(x)}{2} + D\frac{F^2(x)}{4} + I(x). \quad (19)$$

The formal solution of this Hermitian PDE can be written as:

$$\rho(x,t) = e^{-\boldsymbol{H}t}\rho(x,0). \quad (20)$$

Along the same token, the photon arrival probability densities of Eq. 8 can be expressed in an operator form. The bright operator (photon detection even), $\boldsymbol{y}_\tau$, would appear N_P times at the time instances of $\{t_\tau, \tau\in[1,N_P]\}$:

$$\boldsymbol{y}_\tau = \begin{cases} \boldsymbol{y}^a \equiv I_a(x) & y_{t_\tau} = \text{acceptor photon;} \\ \boldsymbol{y}^d \equiv I_d(x) & y_{t_\tau} = \text{donor photon.} \end{cases} \quad (21)$$

Performing the path integral of Eq. 12 via the factorization scheme in Eq. 14 in the evaluation of the likelihood functional of the Langevin parameters, $(F(x),D)\equiv\theta$, can now be represented via the Dirac notation [31] as a series of time propagations in the dark followed by the events of recording a photon:

$$\mathcal{P}(Y(t);F(x),D) = \mathcal{P}(Y(t);\theta) = \mathcal{L}[\theta] = \langle\alpha_{t_0}|e^{-\boldsymbol{H}\Delta t_1}\boldsymbol{y}_1 e^{-\boldsymbol{H}\Delta t_2}\boldsymbol{y}_2\ldots e^{-\boldsymbol{H}\Delta t_{N_P}}\boldsymbol{y}_{N_P}|\beta_{t_{\text{exp}}}\rangle. \quad (22)$$

In this representation, the "bra" state $\langle\alpha_{t_\tau}|$ carries the probability amplitude of the system state at t_τ given all the photon data in past between $[0,\tau]$ and the "ket" state $|\beta_{t_\tau}\rangle$ contains the probability density amplitude of the latent variable at the same time given all of the photons arriving in the future of the smFRET recording, $[\tau,t_{\text{exp}}]$. Path integral across the entire duration from smFRET initiation to the collection of the last photon is just the inner product of these "bra-ket" pairs. Therefore, inferring the trajectory of the latent variable x via all of the recorded photon data, i.e., solving the inference problem defined in Eq. 10, one can follow the Copenhagen interpretation in quantum mechanics [32] to obtain:

$$p(x_t|Y(t);\theta) = \frac{\alpha(x_t,y_{[0,t]})\beta(y_{(t,t_{\text{exp}}]}|x_t)}{\mathcal{P}(Y(t))} = \frac{1}{\mathcal{L}[\theta]}\langle\alpha_t|x\rangle\langle x|\beta_t\rangle. \quad (23)$$

The dependence of the terms in Eq. 23 on the parameter set θ has been omitted to avoid over-complication of the notation.

Since there is no external forces, the initial and final state, $\langle\alpha_{t_0}|$ and $|\beta_{t_{\text{exp}}}\rangle$, respectively, are assumed to follow the equi-

librium distribution $\rho_{\text{eq}}(x) = \sqrt{p_{\text{eq}}(x)}$; that is,

$$\langle \alpha_{t_0} | x \rangle = \rho_{\text{eq}}(x) \quad \text{and} \quad \langle x | \beta_{t_{\text{exp}}} \rangle = \rho_{\text{eq}}(x). \tag{24}$$

The initial and final states can also be constructed by using the Hamiltonian propagator of the Langevin equation without the dark operator, denoted as $\boldsymbol{H}^0$, and extending the temporal domain to infinite times since

$$\left\langle 1 \middle| e^{-\boldsymbol{H}^0 t} \middle| x \right\rangle \Big|_{t\to\infty} \to \sqrt{p_{\text{eq}}(x)}. \tag{25}$$

As such, the likelihood function can be written as:

$$\mathcal{L}[\theta] = \text{tr}\left[e^{-\boldsymbol{H}^0 \infty} e^{-\boldsymbol{H}\Delta t_0} \left(\prod_{\tau}^{N_{\text{P}}} \boldsymbol{y}_\tau e^{-\boldsymbol{H}\Delta t_\tau} \right) e^{-\boldsymbol{H}^0 \infty} \right]. \tag{26}$$

Much of this formulation resembles the structure of quantum dynamics in the form of the density matrix.[33]

Progress in evaluating the path integral of Eq. 22 or the trace operation of Eq. 26 can be made by seeking an eigendecomposition of the Hermitian operator of Eq. 18 to obtain the eigenbasis $\psi_i(x)$ that is consistent with the completeness relationship, $\mathbf{1} = \sum_i |\psi_i(x)\rangle\langle\psi_i(x)|$ with $\mathbf{1}$ being the identity operator. Inserting this identity in between each photon-arrival operator in Eq. 22 transforms the path integral or trace operation into matrix multiplications as:

$$\mathcal{L}[\theta] = \sum_{\{i_\tau\}} \sum_{\{j_\tau\}} \prod_{\tau}^{N_{\text{P}}} \langle \psi_{i_{\tau-1}} | \boldsymbol{y}_\tau | \psi_{j_\tau} \rangle \langle \psi_{j_\tau} | e^{-\boldsymbol{H}\Delta t_\tau} | \psi_{i_{\tau+1}} \rangle. \tag{27}$$

In the summation at the arrival time of the τ^{th} photon, i_τ and j_τ both vary from one to the total number of eigenvectors.[2] The i_τ and j_τ sets in Eq. 27 thus include the indices of all of the eigenvectors associated with the system state at the arrival of each of the N_{P} photons. The initial condition of Eq. 24 is also implied in the summation. In Eq. 27, the joint probability density $\mathcal{P}(Y, X; \theta)$ appears as the product of braket groupings inside the double summation with the specific eigenstate at each time dictated by the elements in the i_τ and j_τ sets. Therefore, $\prod_\tau \langle | \rangle$ is used as a shorthand notation for the summand in Eq. 27, and

$$\mathcal{P}(X, Y; \theta) = \prod_\tau \langle | \rangle. \tag{28}$$

This exploitation of the Hermitian nature of the time propagator plays a critical role in making possible the statistical learning of continuous stochastic dynamics. Although the operation in Eq. 22 or Eq. 26 can be performed forward or backward in time, we generally start from time zero with the vector given by Eq. 24 and perform the matrix operation to the right with Eq. 27. Next, we present the procedure we devised for eigen-decomposing the Hermitian time propagator, evaluating the likelihood functional, and inferring the latent trajectory given a recorded photon sequence.

Determination of Eigenbasis from $\boldsymbol{H}^0$ and $\boldsymbol{H}$

The procedure presented below for eigen-decomposition is not unique but nonetheless allows for computational feasibility of the path integral. Diagonalization of the operator was performed by using a spectral finite element method.[34] First, we solve the symmetric Fokker-Planck equation of the Langevin dynamics,

$$\frac{\partial \rho(x,t)}{\partial t} = -\boldsymbol{H}^0 \rho(x,t), \tag{29}$$

$$\boldsymbol{H}^0 = -D\nabla^2 + V_{\text{eff}}(x), \tag{30}$$

$$V_{\text{eff}} = \frac{DF'(x)}{2} + \frac{DF^2(x)}{4}. \tag{31}$$

We then use the resulting eigenbasis to solve the PDE of Eq. 18 with the dark operator.

The Hermitian $\boldsymbol{H}^0$ gives a set of orthonormal basis $\langle \psi_i^0 |$ that satisfies the completeness relationship, $\mathbf{1} = \sum_i |\psi_i^0\rangle\langle\psi_i^0|$. The time dependence of $\rho(x,t)$ in Eq. 29 can be accounted for by:

$$\rho(x,t) = \sum_i c_i \langle x | \psi_i^0 \rangle e^{-\lambda_i^0 t}. \tag{32}$$

The coefficients c_i's are time invariant and can be determined, for example, based on the initial distribution of $\rho(x,0)$. The eigenvalues satisfy:

$$\boldsymbol{H}^0 |\psi_i^0\rangle = \lambda_i^0 |\psi_i^0\rangle. \tag{33}$$

The finite-time propagation of $\boldsymbol{H}^0$ can be represented by constructing the matrix Ψ^0 that contains the eigenvectors as the columns and the diagonal matrix Λ^0 of the eigenvalues:

$$e^{-\boldsymbol{H}^0 \Delta t} = \Psi^0 e^{-\Lambda^0 \Delta t} \Psi^{0\dagger}. \tag{34}$$

Given a set of PMF and diffusion coefficient of the Langevin equation, we determine the eigenvalues and eigenvectors via a highly accurate spectral finite element method with localized polynomials as the interpolation function in the elements. Details of this numerical solution are provided in the Supplementary Information. We found robust convergence with $N_E = 256$ elements with $N_L = 7$ order polynomials for all of the systems we have analyzed. In particular, the spectral elements, $u_n(x)$'s, are used to expand the scaled eigenvectors by the square root of the equilibrium probability density, $\rho_{\text{eq}}(x)$:

$$\psi_i^0(x) = \rho_{\text{eq}}(x) \phi_i^0(x) \tag{35}$$

$$\phi_i^0(x) = \sum_n c_n^0 u_n(x). \tag{36}$$

In this case, a generalized eigenvalue problem is solved:

$$\sum_n \langle u_m \rho_{\text{eq}} | \boldsymbol{H}^0 | \rho_{\text{eq}} u_n \rangle c_n^0 = \lambda^0 \sum_n \langle u_m | p_{\text{eq}} | u_n \rangle c_n^0. \tag{37}$$

The Hamiltonian matrix $K_{nm}^0 = \langle u_m \rho_{\text{eq}} | \boldsymbol{H}^0 | \rho_{\text{eq}} u_n \rangle$ and the overlap matrix $S_{nm}^0 = \langle u_m | p_{\text{eq}} | u_n \rangle$ are then calculated with analytical differentiation of the interpolation functions, and numerically integrate to solve the algebraic equation of $K^0 c^0 = \lambda^0 S^0 c^0$. Some representative eigenbases are graphically displayed in Fig. 3.

The eigenbasis of $\boldsymbol{H}^0$ is then used to solve the eigenvalue problem that involves the dark operator required for the smFRET likelihood calculation:

$$\left(\boldsymbol{H}^0 + \boldsymbol{H}^I \right) |\psi_i\rangle = \lambda_i |\psi_i\rangle. \tag{38}$$

Here, $\boldsymbol{H}^I = I(x)$, and the new eigenvector is then constructed as a linear combination of $|\psi_i^0\rangle$, $|\psi_i\rangle = \sum_j c_{ij} |\psi_j^0(x)\rangle$.[3] The al-

[2] In this work, 64 eignevectors where used in all calculations. However, 16 eigenvectors would also be adequate for numerical solutions because the precision of the eigen decomposition converges with spectral accuracy.

[3] The summplemenary information details how the Jeffery's prior, or square root of fisher information for the dye intensities of a smFRET experiment, can be added to $I(x)$ to account for the disparity in the information content from the photons emitted at different positions in the domain when the acceptor and

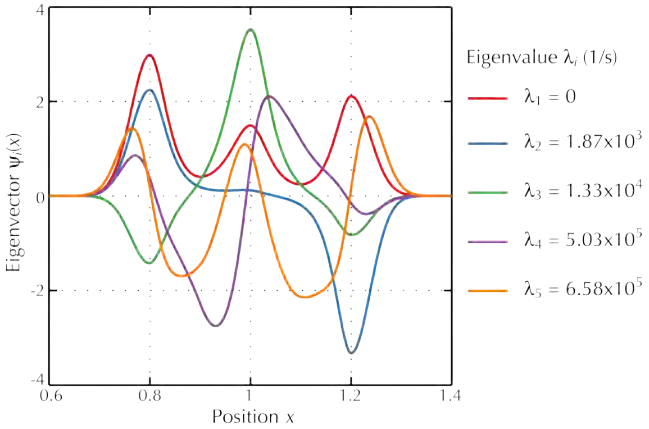

Figure 3: The eigenvectors and eigenvalues of the Langevin Hamiltonian with the reference PMF ($V(x)$) shown in Fig. 2 and $D = 500\ \text{s}^{-1}$. The first eigenvector is the equilibrium density $\rho_{\text{eq}}(x)$ with eigenvalue $\lambda_1 = 0$. The second eigenvector is the slowest reaction of the system that is the transition between the state at $x = 0.8$ and $x = 1.2$, and the eigenvalue of this process sets the apparent time-scale of system relaxation, $1/\lambda_2 = \tau \approx 0.5\,\text{ms}$. The third eigenvector is entrance and escape from the intermediate state at $x = 1.0$.

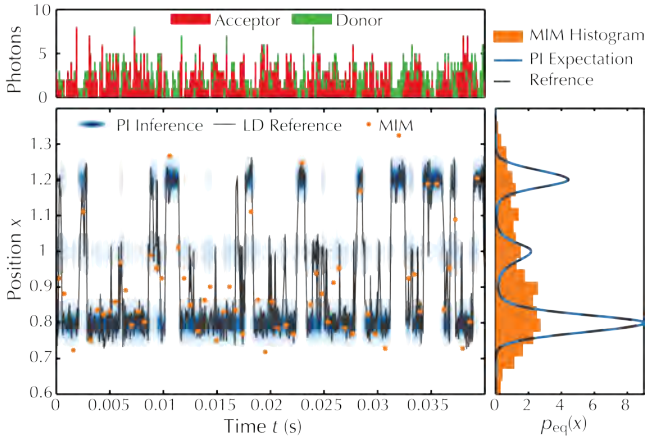

Figure 4: Numerical validation of using Eq. 23 for inferring the dynamics of x from the photon sequence recorded in a smFRET experiment. In this case, PMF and D are assumed to be known a *a priori*, and the EM algorithm was thus not applied. Note the quantitative match of the inferred distance-time trajectory (dark-blue shades) with respect to the "true" trajectory (solid black line). The resolution used in simulating the photon emission follows the intensity values specified in Table 1. The reference model parameters shown in Fig. 2 were used in simulating the photon-arrival trajectory. Only the initial section of 0.04 s of the 3.3 s total trajectory is shown. (top) The trace of photon arrivals per millisecond recorded in the donor and acceptor channels. (bottom-right) Time-averaged probability density of x in the inferred probability density of trajectories, $p_{\text{eq}}(x) = 1/t \int_0^t \delta(x - x')$. The dashed line is the reference distribution in Fig. 2a. (bottom-left) The contours of the product of the inferred $\langle \alpha(t)|x\rangle$ and $\langle x|\beta(t)\rangle$ vectors, i.e., $\mathcal{P}(X(t)|Y(t);F(x),D)$, with the color intensity using the log-scale. Orange crosses are the MIM estimates via adaptive time binning with a relative standard deviation of $\sigma = 0.1$.[11,19]

gebraic equation of this problem, $Kc = \lambda c$, is then solved. The matrix elements of K are:

$$K_{ij} = \langle \psi_i^0 | \boldsymbol{H}^0 + \boldsymbol{H}^I | \psi_j^0(x) \rangle = \lambda_i \delta_{ij} + \langle \psi_i^0 | \boldsymbol{H}^I | \psi_j^0 \rangle. \tag{39}$$

After obtaining the eigenbasis of $\boldsymbol{H}$, the photon arrival operators would have the matrix elements as:

$$\langle \psi_i | \boldsymbol{y}_{a,d} | \psi_j \rangle = \int dx\ \psi_i(x) I_{a,d}(x) \psi_j(x). \tag{40}$$

With the theoretical and numerical tools developed thus far, the likelihood function of Eq. 22 can now be evaluated via a series of N_{P} matrix operations starting at either the α or β end via Eq. 27. The calculation of the likelihood functional is thus proportional to the number of the collected photons. After each matrix-vector multiplication, the state vector α_t or β_t is normalized to prevent numerical underflow and these normalizations are collected according to Eq. 41 to record the log-likelihood as well as the inferred trajectory of the latent variable:[35]

$$\ell[\theta] = \ln \mathcal{L}[\theta] = \sum_{\tau=1} \ln \frac{\|\alpha_{t_\tau}\|}{\|\alpha_{t_{\tau-1}}\|} + \ln \frac{\langle \alpha_{t_{N_{\text{P}}}} | \beta_{t_{N_{\text{P}}}} \rangle}{\|\alpha_{t_{N_{\text{P}}-1}}\|}, \tag{41}$$

where $\|\cdot\|$ indicates vector norm.

Inferring the Probability Density of the Latent Dynamic Variable as a Function of Time

Given the sequence of photon colors and arrival times in a specific smFRET measurement, the probability density of the latent variable at different times can be evaluated via Eq. 23. A simulated sequence of photon emission of a smFRET process using the PMF and diffusion coefficient outlined in Fig. 2 was used as the data set to perform the inference calculation. The resolution used in simulating the photon emission follows the intensity values specified in Table 1 with a total length of 3.3 s. With the knowledge of the underlying PMF and D and a sufficiently high resolution in the smFRET experiment, Fig. 4 shows that the latent trajectory can indeed be accurately inferred. The inferred probability density shown as contours closely overlaps with the trace of the actual Langevin trajectory. The figure also shows that the resulting statistics of the equilibrium distribution quantitatively reproduces that given by the underlying PMF. This result thus numerically validates the inference scheme of Eq. 23 developed in the previous section. In contrast, using the MIM method discussed earlier that involves time binning gives a blurred histogram due to the resulting information loss of the MIM method. (cf. the orange bars in Fig. 4) that deviates from the right answer. Other details of the smFRET simulation and MIM analysis are reported in Supplementary Information.

Obviously, without a prior knowledge of the true $F(x)$ and/or D, the inference would not be accomplished as accurately. Using a default trial profile for the equilibrium probability density, $p_0(x) = \cos^2(x/L)$, where L is the size of spatial domain of x, in Eq. 23, Fig. 5 shows that the inferred probability densities of the latent trajectory has significant differences in comparison to the actual trajectory of x, although the instances of the transitions between metastable states can be captured rather accurately. The simulated photon trajectories and x are the same as those in generating the results of Fig. 4. The time-averaged distribution of x from the inferred probability densities, $\overline{p}(x) \sim p_{\text{eq}}(x)$, also differs significantly

donor intensities are not equivalent.

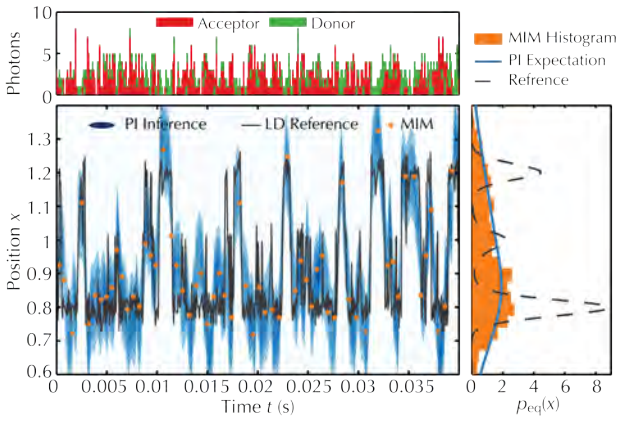

Figure 5: Inferring the dynamics of x from the photon sequence recorded in a smFRET experiment with the maximum likelihood $D = 82\text{s}^{-1}$ is the case of having no information on PMF. The EM algorithm was not yet applied. The trial PMF used for inference is $p_0(x) = \cos^2(x/L)$ that gives least informative dynamics with a fixed domain.[37] The trial PMF is thus different from the actual profile used to generate the photon trajectory. Only the initial section of 0.04 s of the 3.3 s total trajectory is shown. (top) The trace of photon arrivals per millisecond recorded in the donor and acceptor channels. (bottom-right) Time-averaged probability density of x in the inferred probability density of trajectories, $p_{\text{eq}}(x) = 1/t \int_0^t \delta(x - x')$. (bottom-left) The contours of the product of the inferred $\langle \alpha(t)|x\rangle$ and $\langle x|\beta(t)\rangle$ vectors, i.e., $\mathcal{P}(X(t)|Y(t);F(x),D)$, with the color intensity using the log-scale. The solid line is the "true" x trajectory based on the free energy surface shown in Fig. 2b. Points (x) are the MIM estimates via adaptive time binning with a relative standard deviation of $\sigma = 0.1$.[11,19]

from the three-well PMF of the actual dynamics of the latent variable. The default trial profile of $p_0(x) = \cos^2(x/L)$ is the least informative dynamics model for a system of a fixed domain without any prior knowledge of $F(x)$.[36] In this case, Fig. 5 shows that the MIM histogram with the information loss due to adaptive time-binning gives a similar profile of x histogram as that of the inference with $p_0(x) = \cos^2(x/L)$. In this illustration, the maximum likelihood diffusion coefficient of $D = 82\text{s}^{-1}$ given the p_0 distribution was used.

The agreement between the inferred and the actual latent trajectory without the prior knowledge of the underlying PMF and D can be systematically improved by performing the maximization step of statistical learning in Eq. 11. The remaining sections of this paper detail the developments in this direction.

Extracting the Langevin Parameters embedded in smFRET data via Expectation-Maximization

Maximization for the optimal parameter set requires taking derivatives of the log-likelihood functional in Eq. 22 and solves the resulting Euler-Lagrange equation:

$$\frac{\delta \ell[\theta]}{\delta \theta} = \frac{\delta}{\delta \theta} \ln \int \mathcal{D}X(t) \mathcal{P}(X(t), Y(t); \theta) = 0. \tag{42}$$

This task appears to be nearly impossible because the functional dependence on the Langevin parameters is implicitly buried within the path integral. To make the calculation tractable, we generalize the expectation-maximization (EM) algorithm[38,39] developed for discrete-state systems[40] to handle the continuous space of Langevin dynamics.

Firstly, taking the natural log outside the integral in Eq. 42 is moved inside to optimize a lower bound of the likelihood function due to Jensen's inequality. As shown later, this expectation step of EM ensures that optimizing this lower bound also improves the likelihood function itself. In particular, the expectation is performed via the auxiliary function $\mathcal{P}(X(t)|Y(t);\theta^k)$, the inferred probability density of the latent trajectory with the parameter set at step k of the EM iteration, θ^k.[39] The expectation is thus defined as: $\mathbb{E}^k_{X|Y}[\cdot] = \int \mathcal{D}X(t)(\cdot)\mathcal{P}(X(t)|Y(t);\theta^k)$. In this way, Eq. 42 is recast into:

$$\frac{\delta}{\delta \theta} \mathbb{E}^k_{X|Y} [\ln \mathcal{P}(X(t), Y(t); \theta)] = 0. \tag{43}$$

The key here is treating the parameter set θ^k in the weighting function of Eq. 43 as constants in functional derivatives. Only the complete likelihood being expected in Eq. 43 contains the θ set as independent variables for increasing the likelihood function via maximization. Solving Eq. 43 determines the parameter set of the next EM iteration, θ^{k+1}.

The connection between Eq. 43 and Eq. 42 can be expressed alternatively as the following. First, the log-likelihood function we aim to optimize can be split up into:

$$\ell[\theta] = \ln \mathcal{P}(Y;\theta) = \ln \mathcal{P}(Y,X;\theta) - \ln \frac{\mathcal{P}(Y,X;\theta)}{\mathcal{P}(Y;\theta)} \tag{44}$$

$$= \ln \mathcal{P}(Y,X;\theta) - \ln \mathcal{P}(X|Y;\theta). \tag{45}$$

The notation has been collapsed via setting $Y = Y(t)$ and $X = X(t)$. The expectation over the latent trajectories with the auxiliary function $\mathcal{P}(X|Y;\theta^k)$ described earlier is now taken to both sides of Eq. 45. The log-likelihood in the lefthand side remains unchanged with no X dependence. The expectation transforms the first term on the righthand side into the expected log of the complete likelihood in Eq. 43. The second term on the righthand side of Eq. 45 after the expectation is recognized as an "entropy" function for X. Therefore, the following decomposition can be achieved:

$$\ell[\theta] = Q^k(\theta) + S^k(\theta) \tag{46}$$

$$Q^k[\theta] = \int \mathcal{D}X \, \mathcal{P}(X|Y;\theta^k) \ln \mathcal{P}(Y,X;\theta) \tag{47}$$

$$S^k[\theta] = -\int \mathcal{D}X \, \mathcal{P}(X|Y;\theta^k) \ln \mathcal{P}(X|Y;\theta). \tag{48}$$

By setting $\theta = \theta^k$ in Eq. 46 and taking the difference of $\Delta\ell^k = \ell - \ell^k$, one obtains:

$$\Delta\ell^k[\theta] = \Delta Q^k[\theta] + \Delta S^k[\theta]. \tag{49}$$

Here, $\Delta Q^k[\theta] = Q[\theta] - Q^k[\theta^k]$, and $\Delta S^k[\theta]$ is defined similarly. Since the Gibbs inequality ensures that $\Delta S^k[\theta] \geq 0 \; \forall \; \theta$, the following inequality holds:

$$\Delta\ell^k[\theta] \geq \Delta Q^k[\theta]. \tag{50}$$

Therefore, the update of EM optimization for systematically improving $\ell[\theta]$ can be achieved via:

$$\theta^{k+1} = \arg\max_{\theta} Q^k[\theta]. \tag{51}$$

Next, the EM theory for the continuous stochastic dynamics via Eq. 51 is translated into a practical algorithm via the

eigenbasis decomposition presented earlier.

Following the transformation of Eq. 22 into Eq. 27, Eq. 45 is written with the eigenbasis decomposition as:

$$\ell[\theta] = \ln \prod\nolimits_{\tau} \langle | \rangle - \ln \frac{\prod_{\tau} \langle | \rangle}{\mathcal{L}[\theta]}. \tag{52}$$

The expectation over the latent trajectories via the parameter set at the k^{th} iteration, θ^k, now involves taking the following sum for the righthand side of Eq. 52:

$$\mathbb{E}^k_{X|Y}[\cdot] \equiv \sum_{\{i_\tau\}} \sum_{\{j_\tau\}} \frac{\prod^k_{\tau} \langle | \rangle}{\mathcal{L}^k} \cdot . \tag{53}$$

Here, the likelihood and bra-opperator-ket terms are indexed by k to indicate that it is the expectation step of the EM algorithm. As such, Eq. 51 is translated into the following expression based on Eq. 47:

$$\theta^{k+1} = \arg\max_{\theta} \sum_{\{i_\tau\}} \sum_{\{j_\tau\}} \frac{\prod^k_{\tau} \langle | \rangle}{\mathcal{L}^k} \ln \prod\nolimits_{\tau'} \langle | \rangle. \tag{54}$$

In this equation, the dependence on θ is implied in the $\prod_{\tau'} \langle | \rangle$ term as in Eq. 28.

The representation of $Q^k[\theta]$ via the eigenbasis in Eq. 54 can be more concretely expressed as:

$$Q^k[\theta] = \frac{1}{\mathcal{L}^k} \sum_{\tau} \langle \alpha^k_{t_\tau} | e^{-\boldsymbol{H}\Delta t_\tau} | \beta^k_{t_{\tau+1}} \rangle. \tag{55}$$

The θ dependence is now implied in $\boldsymbol{H}$. The expected states in Eq. 55 are:

$$\langle \alpha^k_{t_\tau} | = \sum_j a^k_j(\tau) \langle \psi_j(x) | \tag{56}$$

$$| \beta^k_{t_\tau} \rangle = \sum_i b^k_i(\tau) | \psi_i(x) \rangle. \tag{57}$$

The coefficients in these states for the system at the arrival of the τ^{th} photon are inferred from the eigen-representation of the path integral:

$$a^k_j(\tau) = \sum_{\{i_{\tau'}, \tau' < \tau\}} \prod\nolimits^k_{\tau' < \tau} \langle | \rangle \langle \psi_j | \boldsymbol{y}_{\tau'} | \psi_{i_{\tau'}} \rangle \tag{58}$$

$$b^k_i(\tau+1) = \sum_{\{j_{\tau'}, \tau' > \tau\}} \prod\nolimits^k_{\tau' > \tau} \langle | \rangle \langle \psi_{j_{\tau'}} | \boldsymbol{y}_{\tau'} | \psi_i \rangle. \tag{59}$$

A similar construction has been developed for Markov state models and the forward-backward Baum-Welch algorithm.[41]

Finally, we derived in the Supplementary Information that the Gibbs inequality of $\Delta S^k(\theta) \geq 0 \; \forall \; \theta$ still holds in the eigendecomposition form of the path integral. It is different from the typical analysis with probability distributions that only positive values are involved because the coefficients of eigenvectors at a particular time can be negative.

Evaluation of Functional Derivatives

With the parameter dependence implied in $\boldsymbol{H}$ in Eq. 55, the maximization step of the EM algorithm comes down to solving the following equation for the parameter set θ^{k+1} of the next iteration:

$$0 = \frac{1}{\mathcal{L}^k} \sum_{\tau} \frac{\delta}{\delta \theta} \langle \alpha^k_{t_\tau} | e^{-\boldsymbol{H}\Delta t_\tau} | \beta^k_{t_{\tau+1}} \rangle \Big|_{\theta^{k+1}}. \tag{60}$$

Unfortunately, the functional dependence on the θ set that involves $F(x)$ and D is buried within the exponential of the time propagation operator $\boldsymbol{H}$ and a direct extraction of the functional derivatives is prohibitive. To overcome this difficulty, we derived an line-integral approach to evaluate the derivative kernel[42] for which the details are provided in Supplementary Information. In summary, the functional derivatives are calculated as:

$$\frac{\delta}{\delta \theta} \langle \alpha^k_{t_\tau} | e^{-\boldsymbol{H}\Delta t_\tau} | \beta^k_{t_{\tau+1}} \rangle = - \int_0^{\Delta t_\tau} dt' \langle \alpha^k_{t_\tau} | e^{-\boldsymbol{H}t'} \frac{\delta \boldsymbol{H}}{\delta \theta} e^{-\boldsymbol{H}(\Delta t_\tau - t')} | \beta^k_{t_{\tau+1}} \rangle. \tag{61}$$

The operator derivative is thus defined through a moving window from τ to $\tau + 1$. By inserting $\mathbf{1} = \sum_i |\psi_i\rangle\langle\psi_i|$ and using $\langle \psi_i | e^{-\boldsymbol{H}t} | \psi_j \rangle = \delta_{ij} e^{-\lambda_i t}$, we obtain the derivatives as:

$$- \sum_{i,j} \int_0^{\Delta t_\tau} dt' \langle \alpha^k_{t_\tau} | \psi_i \rangle e^{-\lambda_i t'} \frac{\delta \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{\delta \theta} e^{-\lambda_j (\Delta t_\tau - t')} \langle \psi_j | \beta^k_{t_{\tau+1}} \rangle. \tag{62}$$

Since only the two exponentials in the above equation have dependence on t', we define the transfer functions Γ^τ_{ij} after performing the time integration as:

$$\Gamma^\tau_{ij} = \begin{cases} \Delta t_\tau e^{-\lambda_i \Delta t_\tau} & i = j \\ \dfrac{e^{-\lambda_i \Delta t_\tau} - e^{-\lambda_j \Delta t_\tau}}{\lambda_j - \lambda_i} & i \neq j. \end{cases} \tag{63}$$

Putting the result of Eq. 62 in Eq. 60 and applying Eqs. 56 and 57, one can recognize that the sum over τ in Eq. 62 leads to the following term:

$$\mathbb{E}^k_{X|Y}[a_i b_j] = \sum_{\tau} \Gamma^\tau_{ij} a^k_i(\tau) b^k_j(\tau+1). \tag{64}$$

Finally, we obtain the equation of derivatives required for performing the maximization step in EM:

$$\frac{\delta Q^k[\theta]}{\delta \theta} = - \frac{1}{\mathcal{L}^k} \sum_{i,j} \frac{\delta \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{\delta \theta} \mathbb{E}^k_{X|Y}[a_i b_j]. \tag{65}$$

With the eigenbasis, the functional derivatives in Eq. 65 can be performed with the Euler-Lagrange equation:

$$\frac{\delta \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{\delta F(x)} = \frac{\partial (\psi_i(x) \boldsymbol{H} \psi_j(x))}{\partial F(x)} - \frac{d}{dx} \frac{\partial (\psi_i(x) \boldsymbol{H} \psi_j(x))}{\partial F'(x)}. \tag{66}$$

Imposing the form of the Hamiltonian of Eq. 19 gives the functional derivative with respect to the mean-force profile:

$$\frac{\delta \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{\delta F(x)} = \frac{D}{2} \left(F(x) \psi_i(x) \psi_j(x) - \frac{d}{dx} \Big(\psi_i(x) \psi_j(x) \Big) \right). \tag{67}$$

Since the Hamiltonian and its eigenvalues scale linearly with the diffusion coefficient, the derivative is simply

$$\frac{d \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{dD} = \frac{\lambda_i}{D} \delta_{ij}. \tag{68}$$

The EM Algorithm for Learning Langevin Dynamics from smFRET Measurements

With the functional derivatives attained, the maximization step is accomplished by setting the derivatives of Eq. 65 to zero. First, the functional derivatives with respect to the mean force $F(x)$ in Eq. 67 is applied to Eq. 65 to reach the final form of the maximization step:

$$\mathbb{E}^k_{X|Y}\left[F^{k+1}(x)\frac{D}{2}\psi_i(x)\psi_j(x) - \frac{D}{2}\frac{d}{dx}\left(\psi_i(x)\psi_j(x)\right)\right] = 0. \tag{69}$$

Therefore, the update equation for the mean-force profile in an EM iteration is:

$$F^{k+1}(x) = \frac{\frac{d}{dx}(\mathbb{E}^k_{X|Y}[\delta(x-X)])}{\mathbb{E}^k_{X|Y}[\delta(x-X)]}. \tag{70}$$

In this case, it is convenient to update the equilibrium probability density in EM:

$$p^{k+1}_{\mathrm{eq}}(x) = \mathbb{E}^k_{X|Y}[\delta(x-X)] \tag{71}$$

$$= p^k_{\mathrm{eq}}(x)\sum_{i,j}\mathbb{E}^k_{X|Y}[a_i b_j]\phi_i(x)\phi_j(x). \tag{72}$$

The optimization for the scalar D is then performed by a line search[43] for the maximal likelihood with the new $p^{k+1}_{\mathrm{eq}}(x)$ and Eq. 68. This EM scheme is summarized in Algorithm 1.

Algorithm 1 Expectation-Maximization Statistical Learning for $F(x)$ and D from smFRET

procedure EM(Photon arrival times $\{t_\tau | \forall\tau \in [0, N_{\mathrm{P}}]\}$)
  Initialize $p^0_{\mathrm{eq}} \propto \cos^2(x)$
  repeat
    Eigen-decompose the Hermitian $\boldsymbol{H}$ for Ψ and Λ
    Construct photon arrival operator $Y_{a,d} = \Psi \boldsymbol{y}_{a,d}\Psi^\dagger$
    Evaluate likelihood via $Ye^{-\Lambda t}Y\ldots e^{-\Lambda t}Y$
    Determine $\ell[\theta] = \sum_\tau |\alpha_\tau|$ via normalization
    Infer the latent states $\langle\alpha(t)|, |\beta(t)\rangle$
    Collect statistics of $\mathbb{E}^k_{X|Y}[a_i b_j] = \sum_\tau \Gamma^\tau_{ij}$
    Update $p^{k+1}_{\mathrm{eq}} = p^k_{\mathrm{eq}}(x)\sum_{ij}\mathbb{E}^k_{X|Y}[a_i b_j]\phi_i(x)\phi_j(x)$
    Line search $D^{k+1} = \arg\max_D \ell[p^{k+1}_{\mathrm{eq}}, D]$
  until $\ell[\theta^k] - \ell[\theta^{k-1}] < 1\times 10^{-5}$
end procedure

The proposed statistical learning problem of Langevin dynamics from smFRET data is naturally underdetermined because we attempt to extract a continuous profile from a finite number of photons. A Bayesian prior is thus required to break the degeneracy in the parameter set (such a device was used in a different context for the number of discrete states.[40,44] With the prior, the posterior function for parameter optimization becomes:

$$\mathcal{L}(\theta) = \mathcal{P}(Y(t);\theta)\frac{\mathcal{P}(\theta)}{\mathcal{P}(Y(t))}. \tag{73}$$

A criterion for choosing the prior is to guide the optimization towards $F(x)$ profiles that imply the least amount information of dynamics. The goal is to prevent the statistical learning from over-fitting and over-interpreting the measured data. As such, we select the prior based on maximum trajectory entropy. For Langevin dynamics at equilibrium, the trajectory entropy has been derived[36] and the prior is hence:

$$\mathcal{P}(F(x), D) = e^{-\eta_F S[F(x),D)]} = e^{-\eta_F D\langle F^2(x)\rangle_{\mathrm{eq}}}. \tag{74}$$

$S[\cdot]$ is the trajectory entropy functional, and η_F is the effective temperature specifying the regularization weighting in the optimization. The temperature is set heuristically to the lowest possible value sufficient for maintaining a numerically stable EM iteration. For the test cases examined in this work, $\eta_F = 2\times 10^{-7}$ was found to be sufficient to ensure numerical stability and an order of magnitude higher or lower provides nearly equivalent results.

To incorporate the prior into the EM framework, the ensemble average of force squared in Eq. 74 is approximated by the path expectation at each EM iteration:

$$-\eta_F D\langle F^2(x)\rangle_{\mathrm{eq}} \simeq -\eta_F \mathbb{E}^k_{X|Y}\left[DF^2(x)\right]. \tag{75}$$

The modified update function for the log posterior is then found by setting the expected functional derivative with respect to force to zero

$$0 = \frac{\delta}{\delta F}\mathbb{E}^k_{X|Y}\left[\ell[F(x)] - \eta_F DF^2(x)\right] \tag{76}$$

We then solve for the modified EM update equation for equilibrium probability

$$p^{k+1}_{\mathrm{eq}}(x) = \left(\mathbb{E}^k_{X|Y}[\delta(x-X)]\right)^{1/(1+\eta_F D)}. \tag{77}$$

In practice, only the equilibrium probability is required to construct the eigenvector basis at each iteration and Eq. 77 shows that the net effect of the prior is to smooth the maximum likelihood EM equation by taking the profile to an exponent just slightly less than 1. For the finite domain used in the spectral finite element method, the probability is initialized to a $\propto\cos^2(x\pi/2L)$ distribution of maximum trajectory entropy with zero probability at the boundary of the domain.

The calculation of $\mathbb{E}_{X|Y}[\delta(x-X)]$ as well as its derivative if needed can be performed simply by matrix multiplication if the matrix Ψ is constructed with the eigenvectors oriented in the columns and the inferred state matrix $\mathbb{E}^k_{X|Y}[a_i b_j]$:

$$p^{k+1}_{\mathrm{eq}}(x) = \mathrm{tr}(\Psi\mathbb{E}^k_{X|Y}[a_i b_j]\Psi^\dagger). \tag{78}$$

For trajectories composed of 80,000 photons, convergence in likelihood typically requires an exhaustive number of 50,000 iterations due to the sub-linearity of the EM algorithm.[45]

Fig. 6 shows a typical sequence of photon data and the corresponding trajectory of the latent variable in the simulation. The comparison of the optimized profile of equilibrium distribution with that corresponds to the reference PMF indicates the ability of the EM scheme we devised to learn about the continuous profile of PMF from the photon sequence. Fig. 6 also shows that the latent trajectory can be inferred accurately with the optimized parameter set. With an explicit consideration of each arrived photon, including the dark period waited before, much of the information of the underlying dynamics can indeed be extracted from the indirect measurement of smFRET. On the other hand, time-binning of any kind inevitably leads to information loss; even the MIM method cannot retrieve this level of mechanistic details as seen in Fig. 6.

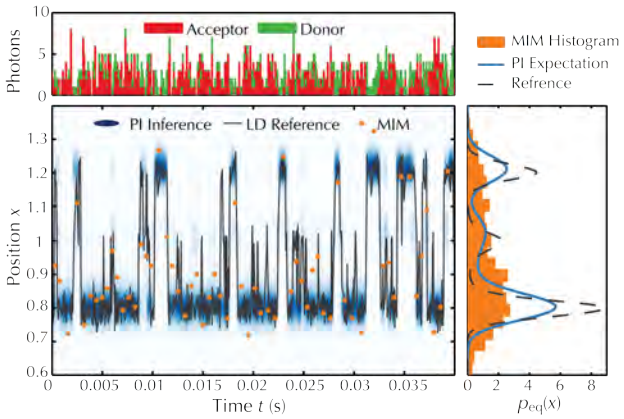

Figure 6: EM statistical learning of PMF and D from the photon-arrival time trajectory of a simulated smFRET experiment. Only the initial section of 0.04 s of the 3.3 s total trajectory is shown. (top) The trace of photon arrivals per millisecond recorded in the donor and acceptor channels. (bottom-right) Time-averaged probability density of x in the inferred probability density of trajectories, $p_{\rm eq}(x) = 1/t \int_0^t \delta(x - x')$. (bottom-left) The contours of the product of the inferred $\langle \alpha(t)|x\rangle$ and $\langle x|\beta(t)\rangle$ vectors, i.e., $\mathcal{P}(X(t)|Y(t);F(x),D)$, with the color intensity using the log-scale. The solid line is the "true" distance trajectory. The orange crosses are the MIM estimates via adaptive time binning with a relative standard deviation of $\sigma = 0.1$.[11,19]

Results and Discussions

The EM algorithm we developed for continuous stochastic dynamics has many similar features as in the self-consistent mean field theory in polymer physics that both aim to search for the extrema of a functional. The fundamental property of EM that the likelihood is a strictly increasing quantity ensures that the optimization is stable, robust and reliable, despite the fact that the convergence rate is sub-linear.[45]

To illustrate the robustness of EM against the stochasticity in smFRET data, twelve independent photon trajectories were simulated at four different resolutions to compare their results of statistical learning. The total number of EM optimization for generating the results of this section is thus forty eight. The behaviors of learning the Langevin parameters and kinetic behaviors are summarized in Figs. 7–10.

The optimized probability densities of the equilibrium distribution and PMFs for each of the twelve trajectories are plotted in Fig. 7 and Fig. 8, respectively for the cases of 1x and 5x intensity, cf. Table 1. The averaged and reference profiles are also shown in the figures for comparison. It can be seen that at both resolutions, there exhibits considerable variation in the results of statistical learning due to the noise in stochastic trajectories and photon statistics. Nonetheless, the number of meta-stable states and their locations are consistently reproduced. Since resolving the dynamics associated with the middle state requires a higher temporal resolution, its inference shows more significant deviation from the "true" values as compared to the short-distance ($x = 0.8$) and the long-distance ($x = 1.2$) states as seen in Fig. 7 and Fig. 8. It is also clear from Fig. 7(a) and Fig. 8(a) that at the typical resolution of smFRET experiments (1x intensity), the middle state can barely be resolved. Although each EM optimization for an individual trajectory does predict higher barriers and have the middle-state more resolved, variances amongst the 12 trajectories, and hence their inferred locations of meta-stable states and barrier heights, muddle these information in the averaged result. If the resolution of the photon data was reduced to 0.5x intensities, the ability of resolving the middle state in Fig. 2 disappears. The corresponding EM results for the other resolutions examined in this work are shown in Supplementary Information. With 5x intensities, Fig. 7(b) and Fig. 8(b) show that the same number of photons in a 5 times shorter duration carries more information for resulting the finer details of the local shapes and curvatures of the profiles of the PMF and equilibrium probability density. The averaged profiles resulting from the 5x EM capture the answer quite closely. Next, we present the results of learning dynamical properties of D, mean first passage times, and kinetic rates.

The statistically learned diffusion coefficients for the trajectory sets simulated at different resolutions are summarized in the boxplot of Fig. 9. The boxplot represents the variation in a data set by showing the average as the red horizontal line, the 25 % quartile of the data above and below the average by the unfilled blue bar, the upper and lower bounds of the continuous spread of data as black caps, and the outliers as red crosses. It can be seen that the maximum entropy prior for regularizing the EM optimization causes systematic bias of D towards lower values. Without sufficient information in the data, the trajectory entropy penalty of $\langle F^2(x)\rangle$ tends to lower PMF barriers and the diffusion coefficient is hence underestimated. On the other hand, Fig. 9 also shows that if the data does provide information for resolving the PMF barriers, such as in the case of 5x intensities, the resulting D would have a higher value. The PMF and D work in balance to control the transition time-scales of the system and the estimation of kinetic rates is actually much less biased (as presented later). Although the Bayesian prior approach overcomes the degeneracy issue associated with learning a continuous function,

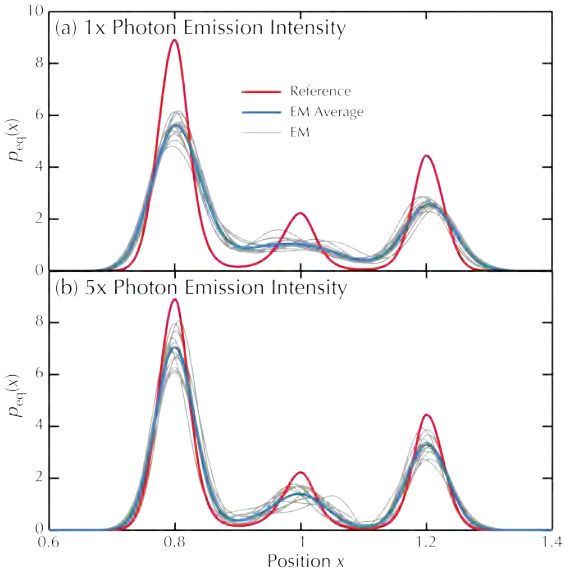

Figure 7: The $p_{\rm eq}(x)$ profiles of the converged results of EM optimization for 12 independent trajectories of 80,000 photons. The trajectories were simulated with the PMF and D shown in Fig. 2 at different resolutions of the smFRET experiment. (a) The results using the default smFRET parameters listed in Table 1, i.e., the 1x intensity. (b) The results using the 5x intensity. $\eta_F = 2 \times 10^{-7}$ was employed for all runs of EM optimization

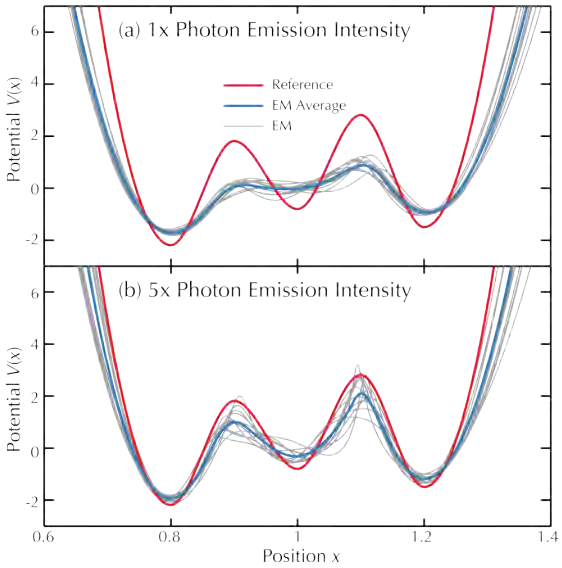

Figure 8: The PMF profiles of the converged results of EM optimization for 12 independent trajectories of 80,000 photons. The trajectories were simulated with the PMF and D shown in Fig. 2 at different resolutions of the smFRET experiment. (a) The results using the default smFRET parameters listed in Table 1, i.e., the 1x intensity. (b) The results using the 5x intensity. $\eta_F = 2 \times 10^{-7}$ was employed for all runs of EM optimization.

it does introduce bias into the final solution under data deficiency.[46] Further investigation indicates that the converged diffusion coefficient is relatively insensitive to the η_F parameter over a wide range of values from 10^{-5} to 10^{-7} after the establishment of numerical stability. Fig. 9 shows that with a 5x intensity, the bias in D can start to be overcome by the richer dynamics information carried in the photon data.

The kinetic rates of transition between the meta-stable states can be determined by calculating the mean first passage times (MFPTs) as a post-processing step after the PMF and D being optimized with EM. It is important to emphasize the number and locations of meta-stable states are read off from the EM converged profiles without assuming prior knowledge. We calculate the MFPT from state A at position $x_A \approx 0.8$ to state B at position $x_B \approx 1.2$ via:[47]

$$\mathrm{MFPT}(x_\mathrm{A} \to x_\mathrm{B}) = k^{-1}_{\mathrm{A}\to\mathrm{B}} = \frac{1}{D}\int_{x_\mathrm{A}}^{x_\mathrm{B}} dx\, e^{V(x)} \int_{x_\mathrm{L}}^{x} dx'\, e^{-V(x')}, \tag{79}$$

where the subscripts L and R denote the left and right ends of the domain of the system dynamics, respectively. For the reverse transition, the formula reads

$$\mathrm{MFPT}(x_\mathrm{B} \to x_\mathrm{A}) = k^{-1}_{\mathrm{B}\to\mathrm{A}} = \frac{1}{D}\int_{x_\mathrm{B}}^{x_\mathrm{A}} dx\, e^{V(x)} \int_{x_\mathrm{R}}^{x} dx'\, e^{-V(x')}. \tag{80}$$

Or, recognizing $\int_x^{x_\mathrm{R}} = \int_{x_\mathrm{L}}^{x_\mathrm{R}} - \int_{x_\mathrm{L}}^{x}$, $k^{-1}_{\mathrm{B}\to\mathrm{A}}$ can be calculated alternatively as:

$$k^{-1}_{\mathrm{B}\to\mathrm{A}} = \frac{1}{D}\int_{x_\mathrm{A}}^{x_\mathrm{B}} dx\, e^{V(x)} \left(\int_{x_\mathrm{L}}^{x_\mathrm{R}} dx'\, e^{-V(x')} - \int_{x_\mathrm{L}}^{x} dx'\, e^{-V(x')} \right). \tag{81}$$

By using the reference PMF and diffusion coefficient of

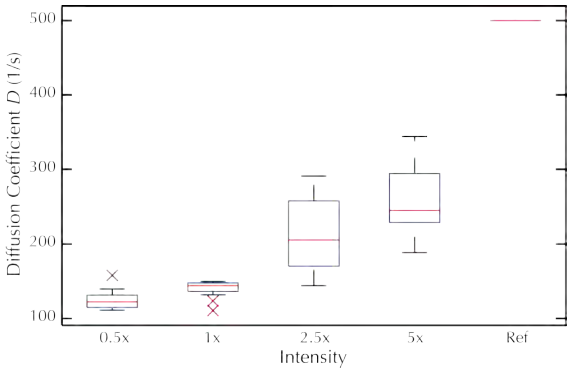

Figure 9: Boxplots of the converged diffusion coefficients from EM optimization for the 12 independent trajectories of 80,000 photons simulated at each photon resolution. At each level of intensity, the averaged value of D over the statistical learning of 12 trajectories is represented as the red horizontal line, the 25 % quartile of the values above and below the average is shown by the unfilled blue bar, the upper and lower bounds of the continuous spread of D are labeled as black caps, and the outliers are denoted as red crosses. $\eta_F = 2 \times 10^{-7}$ was employed for all runs of EM optimization.

Table 2: Mean-first-passage-times (MFPTs) and reaction rates for the reference potential shown in Fig. 2 with $D = 500\ \mathrm{s}^{-1}$.

$\mathrm{MFPT}(x_\mathrm{A} \to x_\mathrm{B})$	1.959×10^{-3}	s
$\mathrm{MFPT}(x_\mathrm{B} \to x_\mathrm{A})$	0.939×10^{-3}	s
$k_{\mathrm{A}\to\mathrm{B}}$	0.510×10^{3}	s^{-1}
$k_{\mathrm{B}\to\mathrm{A}}$	1.064×10^{3}	s^{-1}
$k_{\mathrm{A}\to\mathrm{B}}/k_{\mathrm{B}\to\mathrm{A}}$	0.480	

Fig. 2, we calculate the MFPTs and kinetic rates for the transition between states A and B. These values of the actual latent dynamics are summarized in Table 2. The same post-processing evaluation of MFPT is also performed on the converged PMF and D from the EM optimization of the aforementioned set of 12 trajectories at each data resolution. The results are summarized in the boxplots of Fig. 10. Despite the variance in the diffusion coefficients deduced from these trajectories as seen in Fig. 9, the kinetic rates are quantitatively reproduced with high accuracy even at the lowest resolution. The little bias, if any, in the inferred kinetic rate can be understood as a nice consequence of the balance between PMF and D in the EM statistical learning discussed earlier.

Conclusion

In this work, we have developed a Bayesian inference framework to learn about the continuous stochastic dynamics of Langevin equation from time-dependent single-molecule FRET experiments. Our theory explicitly and rigorously incorporates the two layers of stochasticity separating the dynamics information of interest from the fluorescence single-molecule data—the statistical photon detections and the stochastic thermal fluctuations. The resulting EM algorithm hence allows the entire PMF profile and diffusion coefficient of protein conformational changes to be extracted from the photon colors and arrival times recorded in a smFRET experiment, without any presumed profile shape nor kinetic models; this method thus

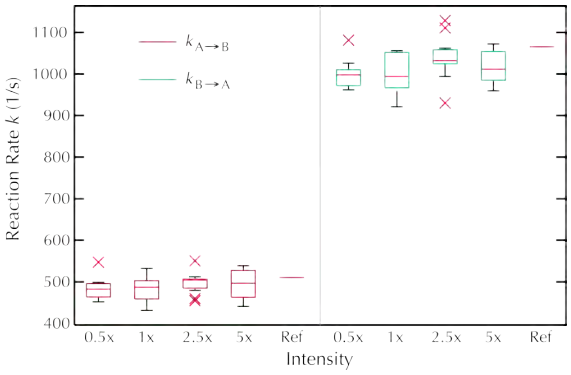

Figure 10: Boxplots of the kinetic rate $k_{A\to B}$ and $k_{B\to A}$ from mean-first-passage-time calculations using the converged PMF and D of EM optimization. The data include 12 independently simulated trajectories with 80,000 photons at each data resolution. For each level of intensity, the average values of kinetic rates over 12 trajectories are represented as the red horizontal lines, the 25% quartile of the rate data above and below the average is shown by the unfilled blue bar, the upper and lower bounds of the continuous spread of the rates are labeled as black caps, and the outliers are denoted as red crosses. $\eta_F = 2 \times 10^{-7}$ was employed for all runs of EM optimization.

enables unanticipated discoveries. The capability to extract the conformational diffusion coefficient means that the forces that govern the stochastic dynamics can now be quantified directly from single-molecule data and is a significant milestone. Together with the deterministic force given by the PMF profile, the conformational *dynamics* can now be quantitatively determined at the single-molecule level. This work thus represents an important step forward advancing single-molecule spectroscopy.

A series of analytical and numerical advances has been accomplished to achieve the capability of Bayesian inference for continuous dynamics. Firstly, the numerical path integration of the likelihood functional, which would have involved bookkeeping infinite terms for continuous dynamics, is made possible by integrating forward the time-independent terms and incorporating the operator of observing darkness into the Fokker-Planck equation of Langevin dynamics, Eq. 17. Secondly, the Fokker-Planck equation with the dark operator is transformed into a time-symmetrized form of Eqs. 18 and 19. The Hermitian property of this representation allows eigenvectors with orthonormality and completeness to be acquired for convenient decomposition of the Langevin time propagation coupled with photon statistics in Eq. 22. The eigen decomposition reduces the otherwise infinite operations in continuous space to the finite number of basis sets. Combining these two aspects transforms the path-integral calculations of the likelihood functional into matrix operations. Thirdly, we generalized the EM scheme originally developed for discrete-state statistical learning to the space of continuous profiles by deriving the analytical derivatives of the likelihood functional with respect to the PMF profile, Eqs. 67 and 68. Our EM algorithm for continuous stochastic dynamics, Algorithm 1, is also analogous to generalizing the Kalman filter for the statistical inference on linear dynamical systems (linearity in the sense that the time propagator is independent of the system position[48,49]) to handle arbitrary potentials of x and arbitrary probabilistic information from the experimental observations. Lastly and equally importantly, a trajectory-entropy motivated prior is imposed to ensure the numerical stability of EM for Langevin dynamics by breaking the solution degeneracy within the space of continuous PMF profiles.

As a result, extracting the governing PMF and diffusion coefficient of protein dynamics from smFRET experiments can now be accomplished. Conversely, an experimentalist can also use this framework to establish the data quality required for resolving such mechanistic features as the number and location of meta-stable states as well the kinetic rates connecting them. The ability to acquire the mechanistic details of protein dynamics may facilitate the engineering of the functionalities of enzymes and protein machines.[50]

The derivation presented in this work exemplifies the manner by which the specific features of smFRET experiments are utilized to construct the operators associated with photon statistics. Although the theoretical development and numerical illustration presented in this work focus on the Langevin dynamics with a constant diffusion coefficient, the generalization to x-dependent diffusion is expected to be relatively uncomplicated. The framework can also be extended straightforwardly into multiple dimensions if several separate signals of the system could be measured simultaneously.[51,52] An essential requirement for such applications is that time propagation of the system can be made symmetric so that an eigenbasis set can be constructed for transforming a continuous path integral into matrix multiplications. Extension of our developments to other classes of data types obtained experimentally or computationally can also be achieved readily provided that the information (observation) operator $\boldsymbol{y}$ is defined. Examples of other data types include the force and position trajectories measured in single-molecule pulling experiments,[53] the many short bursts of trajectories in specific types of molecular simulations,[54,55] and potentially quantum dynamics measurements due to the similarities in the path-integral framework.[33]

Acknowledgement We would like to thank Attila Szabo for helpful discussions and Berend Smit for advices in preparing the manuscript. We thank Jeffrey A. Hanson and Thomas E. Morrell for the assistance in understanding the experimental setup of smFRET. This work is supported by the Princeton University and also by the University of California at Berkeley.

Supporting Information Available: Details on the numerical methods used in the simulation of the Langevin Dynamics and smFRET trajectories, formula for the Maximum Information method, and framework of the spectral finite element method are provided. Also included is a symbol list. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Moerner, W. E.; Orrit, M. Illuminating Single Molecules in Condensed Matter. *Science* **1999**, *283*, 1670–1676.

(2) Weiss, S. Fluorescence Spectroscopy of Single Biomolecules. *Science* **1999**, *283*, 1676–1683.

(3) Barkai, E.; Brown, F. L.; Orrit, M.; Yang, H. *Theory and Evaluation of Single-Molecule Signals*; Scientific Publishing: Singapore, 2008.

(4) Selvin, P. R.; Ha, T. *Single-Molecule Techniques: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: New York, 2008.

(5) Komatsuzaki, T.; Kawakami, M.; Takahashi, S.; Yang, H.; Silbey, R. J. *Single-Molecule Biophysics Experiment and Theory Advances in Chemical Physics Volume 146*; John Wiley & Sons, Inc.: New York, 2012.

(6) Michalet, X.; Weiss, S.; Jäger, M. Single-Molecule Fluorescence Studies of Protein Folding and Conformational Dynamics. *Chem. Rev.* **2006**, *106*, 1785–1813.

(7) Sisamakis, E.; Valeri, A.; Kalinin, S.; Rothwell, P. J.; Seidel, C. A. M. Accurate Single-Molecule FRET Studies Using Multiparameter Fluorescence Detection. *Methods in Enzymology, Vol 475: Single Molecule Tools, Pt B* **2010**, *475*, 455–514.

(8) Tan, Y. W.; Yang, H. Seeing the Forest for the Trees: Fluorescence Studies of Single Enzymes in the Context of Ensemble Experiments. *Phys. Chem. Chem. Phys.* **2011**, *13*, 1709–1721.

(9) Yang, H. The Orientation Factor in Single-Molecule Förster-Type Resonance Energy Transfer, with Examples for Conformational Transitions in Proteins. *Israel J. Chem.* **2009**, *49*, 313–321.

(10) Hanson, J. A.; Duclerstacit, K.; Watkins, L. P.; Bhattacharyya, S.; Brokaw, J.; Chu, J. W.; Yang, H. Illuminating the Mechanistic Roles of Enzyme Conformational Dynamics. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18055–18060.

(11) Watkins, L. P.; Yang, H. Information Bounds and Optimal Analysis of Dynamic Single Molecule Measurements. *Biophys. J.* **2004**, *86*, 4015–4029.

(12) Andrec, M.; Levy, R. M.; Talaga, D. S. Direct Determination of Kinetic Rates from Single-Molecule Photon Arrival Trajectories Using Hidden Markov Models. *J. Phys. Chem. A* **2003**, *107*, 7454–7464.

(13) Kou, S. C.; Xie, X. S.; Liu, J. S. Bayesian Analysis of Single-Molecule Experimental Data. *J. Roy. Statist. Soc. Ser. C* **2005**, *54*, 469–506.

(14) Zucchini, W.; MacDonald, I. L. *Hidden Markov Models for Time Series*; An Introduction Using R; Chapman and Hall/CRC, 2009.

(15) Bronson, J. E.; Fei, J.; Hofman, J. M.; Gonzalez, R. L., Jr; Wiggins, C. H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophys. J.* **2009**, *97*, 3196–3205.

(16) Gopich, I. V.; Szabo, A. Decoding the Pattern of Photon Colors in Single-Molecule FRET. *J. Phys. Chem. B* **2009**, *113*, 10965–10973.

(17) Bronson, J. E.; Hofman, J. M.; Fei, J.; Gonzalez, R. L.; Wiggins, C. H. Graphical Models for Inferring Single Molecule Dynamics. *BMC Bioinformatics* **2010**, *11*, S2.

(18) Liu, Y.; Park, J.; Dahmen, K. A.; Chemla, Y. R.; Ha, T. A Comparative Study of Multivariate and Univariate Hidden Markov Modelings in Time-Binned Single-Molecule FRET Data Analysis. *J. Phys. Chem. B* **2010**, *114*, 5386–5403.

(19) Watkins, L. P.; Chang, H.; Yang, H. Quantitative Single-Molecule Conformational Distributions: A Case Study with Poly-(L-proline). *J. Phys. Chem. A* **2006**, *110*, 5191–5203.

(20) Flynn, E. M.; Hanson, J. A.; Alber, T.; Yang, H. Dynamic Active-Site Protection by the *M. tuberculosisProtein* Tyrosine Phosphatase PtpB Lid Domain. *J. Am. Chem. Soc.* **2010**, *132*, 4772–4780.

(21) Hanson, J. A.; Yang, H. A General Statistical Test for Correlations in a Finite-Length Time Series. *J. Chem. Phys.* **2008**, *128*, 214101.

(22) Hanson, J. A.; Yang, H. Quantitative Evaluation of Cross-Correlation between Two Finite-Length Time Series with Applications to Single-Molecule FRET. *J. Phys. Chem. B* **2008**, *112*, 13962–13970.

(23) Hanson, J. A.; Brokaw, J.; Hayden, C. C.; Chu, J.-W.; Yang, H. Structural Distributions from Single-Molecule Measurements as a Tool for Molecular Mechanics. *Chem. Phys.* **2012**, *396*, 61–71.

(24) Wang, Y.; Gan, L. F.; Wang, E. K.; Wang, J. Exploring the Dynamic Functional Landscape of Adenylate Kinase Modulated by Substrates. *J. Chem. Theory Comput.* **2013**, *9*, 84–95.

(25) Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press: New York, NY, 2001.

(26) Brokaw, J. B.; Haas, K. R.; Chu, J.-W. Reaction Path Optimization with Holonomic Constraints and Kinetic Energy Potentials. *J. Chem. Theory Comput.* **2009**, *5*, 2050–2061.

(27) Haas, K.; Chu, J.-W. Decomposition of Energy and Free Energy Changes by Following the Flow of Work along Reaction Path. *J. Chem. Phys.* **2009**, *131*, 144105–144105–11.

(28) Cao, J.; al, e. Generic Schemes for Single-Molecule Kinetics. 1: Self-Consistent Pathway Solutions for Renewal Processes. *J. Phys. Chem. B* **2008**, *112*, 12867–12880.

(29) Gopich, I.; Szabo, A. Theory of Photon Statistics in Single-Molecule Förster Resonance Energy Transfer. *J. Chem. Phys.* **2005**, *122*, 014707.

(30) Gopich, I. V.; Szabo, A. Single-Macromolecule Fluorescence Resonance Energy Transfer and Free-Energy Profiles. *J. Phys. Chem. B* **2003**, *107*, 5058–5063.

(31) Song, L.; Huang, J.; Smola, A.; Fukumizu, K. Hilbert Space Embeddings of Conditional Distributions with Applications to Dynamical Systems. *ICML '09 Proceedings of the 26th Annual International Conference on Machine Learning* **2009**, 961–968.

(32) Sakurai, J. J. *Modern Quantum Mechanics*; Addison Wesley Publishing Company, 1985.

(33) Feynman, R. P.; Hibbs, A. R.; Styer, D. F. *Quantum Mechanics and Path Integrals*; Emended Edition; Courier Dover Publications, 2012.

(34) Pozrikidis, C. *Introduction To Finite And Spectral Element Methods Using Matlab*; CRC Press, 2005.

(35) Turner, R. Direct Maximization of the Likelihood of a Hidden Markov Model. *Comput. Stat. Data Anal.* **2008**, *52*, 4147–4160.

(36) Haas, K. R.; Yang, H.; Chu, J.-W. Elements of the Trajectory Entropy in Continuous Stochastic Processes at Equilibrium. *Phys. Rev. Lett.* **2013**, under review.

(37) Haas, K. R.; Yang, H.; Chu, J.-W. Fisher Information Metric for the Langevin Equation and Least Informative Models of Continuous Stochastic Dynamics. *J. Chem. Phys.* **2013**, under review.

(38) Dempster, A. P.; Laird, N. M.; Rubin, D. B. Maximum Likelihood from Incomplete Data Via Em Algorithm. *J. Roy. Stat. Soc. Ser. B* **1977**, *39*, 1–38.

(39) Little, R. J. A.; Rubin, D. B. *Statistical Analysis with Missing Data*; Wiley-Interscience, 2002.

(40) Watkins, L. P.; Yang, H. Detection of Intensity Change Points in Time-Resolved Single-Molecule Measurements. *J. Phys. Chem. B* **2005**, *109*, 617–628.

(41) Chodera, J. D.; Elms, P.; Noé, F.; Keller, B.; Kaiser, C. M.; Ewall-Wice, A.; Marqusee, S.; Bustamante, C.; Hinrichs, N. S. Bayesian Hidden Markov Model Analysis of Single-Molecule Force Spectroscopy: Characterizing Kinetics under Measurement Uncertainty. *arXiv* **2011**,

(42) Wilcox, R. M. Exponential Operators and Parameter Differentiation in Quantum Physics. *J. Math. Phys.* **1967**, *8*, 962.

(43) Nocedal, J.; Wright, S. J. *Numerical Optimization*; Springer, 2006.

(44) Bayarri, M.; García Donato, G. Generalization of Jeffreys Divergence-Based Priors for Bayesian Hypothesis Testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2008**, *70*, 981–1003.

(45) Xu, L.; Jordan, M. I. On Convergence Properties of the EM Algorithm for Gaussian Mixtures. *Neural Comput.* **1996**, *8*, 129–151.

(46) Firth, D. Bias Reduction of Maximum Likelihood Estimates. *Biometrika* **1993**, *80*, 27–38.

(47) Gardiner, C. W. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*; Springer Verlag, 2004.

(48) Wainwright, M. J.; Jordan, M. I. Graphical Models, Exponential Families, and Variational Inference. *FNT in Machine Learning* **2007**, *1*, 1–305.

(49) Krishnamurthy, V.; Moore, J. B. On-line Estimation of Hidden Markov Model Parameters Based on the Kullback-Leibler Information Measure. *IEEE Trans. Signal Proccess.* **1993**, *41*, 2557–2573.

(50) Wang, J.; Xu, L.; Wang, E. Potential Landscape and Flux Framework of Nonequilibrium Networks: Robustness, Dissipation, and Coherence of Biochemical Oscillations. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 12271.

(51) Uphoff, S.; Holden, S. J.; Le Reste, L.; Periz, J.; van de Linde, S.; Heilemann, M.; Kapanidis, A. N. Monitoring Multiple Distances within a Single Molecule Using Switchable FRET. *Nat. Meth.* **2010**, *7*, 831–836.

(52) Chodera, J. D.; Pande, V. S. Splitting Probabilities as a Test of Reaction Coordinate Choice in Single-Molecule Experiments. *Phys. Rev. Lett.* **2011**, *107*, 098102.

(53) Mossa, A.; de Lorenzo, S.; Huguet, J. M.; Ritort, F. Measurement of Work in Single-Molecule Pulling Experiments. *J. Chem. Phys.* **2009**, *130*, 234116.

(54) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(55) Wells, B. H.; Smith, E. B.; Zare, R. N. The Stability of the RKR Inversion Procedure to Errors in the Spectroscopic Data: Origin of the Inner-wall Ripple. *Chem. Phys. Lett.* **1983**, *99*, 244–249.