THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Classic Maximum Entropy Recovery of the Average Joint Distribution of Apparent FRET Efficiency and Fluorescence Photons for Single-Molecule Burst Measurements

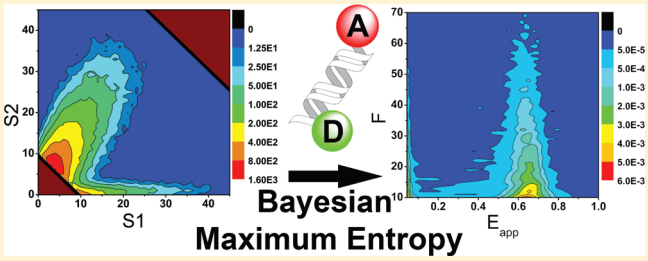
Matthew S. DeVore,[†] Stephen F. Gull,[‡] and Carey K. Johnson*[,†]

[†]Department of Chemistry, University of Kansas, Lawrence, Kansas 66045, United States

[‡]Astrophysics Group, Department of Physics, Cambridge University, Cambridge CB3 0HE, United Kingdom

Supporting Information

ABSTRACT: We describe a method for analysis of single-molecule Förster resonance energy transfer (FRET) burst measurements using classic maximum entropy. Classic maximum entropy determines the Bayesian inference for the joint probability describing the total fluorescence photons and the apparent FRET efficiency. The method was tested with simulated data and then with DNA labeled with fluorescent dyes. The most probable joint distribution can be marginalized to obtain both the overall distribution of fluorescence photons and the apparent FRET efficiency distribution. This method proves to be ideal for determining the distance distribution of FRET-labeled biomolecules, and it successfully predicts the shape of the recovered distributions.



INTRODUCTION

Single-molecule Förster resonance energy transfer (FRET) is being applied to characterize the conformational states and dynamics of biomolecules in increasingly powerful ways.[1−6] FRET trajectories of immobilized molecules can track the internal dynamics of single molecules over periods of seconds,[7] but the molecules must be trapped or tethered at a surface. The detection of fluorescence bursts from biomolecules freely diffusing in solution avoids the need for their immobilization. The underlying molecular parameters may be hard to recover, however, due to the stochastic nature of molecular paths through the probe volume and the statistics of photon detection. As a result, the analysis of single-molecule FRET data has become increasingly sophisticated.[8−12]

In single-molecule analysis, as in other fields, wildly different models can fit the experimental data adequately, based on goodness-of-fit measures such as least-squares. The experimenter must then decide which model to prefer. To make this decision systematically, researchers in the single-molecule field have begun to use Bayesian data-fitting approaches.[13−20]

In Bayesian analysis, the probability $P(\boldsymbol{f}|\boldsymbol{D},I)$ of a set of parameters $\boldsymbol{f}$, given the experimental data set $\boldsymbol{D}$ and all other assumed information I, is calculated via Bayes' theorem. This probability is called the inference. Bayes' theorem relates the inference $P(\boldsymbol{f}|\boldsymbol{D},I)$ to the prior probability $P(\boldsymbol{f}|I)$, the likelihood for the data $P(\boldsymbol{D}|\boldsymbol{f},I)$, and the evidence $P(\boldsymbol{D}|I)$, as follows:[21]

$$P(\boldsymbol{f}|\boldsymbol{D},I) = \frac{P(\boldsymbol{f}|I)P(\boldsymbol{D}|\boldsymbol{f},I)}{P(\boldsymbol{D}|I)} \quad (1)$$

The evidence carries information concerning model selection, since it is the probability of the data given all assumed information including whatever model is chosen. A model-free, or nonparametric, approach dispenses with a predetermined shape (Gaussian or any other) of an entire distribution. This is in contrast to model-fitting approaches in which the experimenter specifies the form of a distribution in terms of a few parameters that are to be estimated from the data. Even a model-free approach does not mean that all prior assumptions have been eliminated, since it inevitably depends on the mathematical form of the prior.

Our objective is to analyze single-molecule FRET data on freely diffusing molecules ("burst" measurements) by Bayesian maximum entropy theory. The power of maximum entropy to recover FRET distributions from single-molecule trajectories of immobilized molecules has already been demonstrated by Yang and co-workers,[22,23] who used the method to uncover the underlying FRET states for polyproline constructs of varying length[22] and to detect open and closed states of adenylate kinase.[23] Maximum entropy methods have been applied to burst analysis by Seidel and co-workers to recover the distribution of FRET efficiencies $P(E)$ for freely diffusing molecules[8] given a model for the distribution of total fluorescence counts $P(F)$ per time bin.[24]

The approach we present extends the application of maximum entropy to single-molecule burst data in four ways.

Received: October 13, 2011
Revised: February 8, 2012
Published: February 16, 2012

First, it implements the so-called "classic" maximum entropy method,[25] in which Bayesian analysis is used to evaluate the weight of the information entropy relative to the data likelihood. Second, the maximum entropy analysis is applied directly to the two-dimensional histogram of intensities in two data channels detecting green and red emission rather than to the experimental FRET distribution. This approach makes no assumption as to the character of the fluorescence decay, $P(F)$. Third, the Bayesian analysis permits evaluation of the Bayesian evidence, allowing for model selection. Fourth, intrinsic correlation functions are used as a model for introducing smoothness into the recovered distribution with the degree of blurring optimized by maximizing the Bayesian evidence.[25,26]

BACKGROUND AND THEORY

Bayesian Maximum Entropy. One of the most widely used model-free data fitting techniques is maximum entropy. The information entropy of a probability distribution was first described by Shannon; it is a measure of the amount of information contained in a probability distribution.[27] Jaynes[28] and Shore and Johnson[29] have shown that the only consistent way to recover a probability distribution, given a set of constraints upon it, is to choose the distribution with the maximal information entropy. Maximum entropy was initially employed as a regularization technique[25,30−32] in which the entropy is maximized subject to a constraint on the data, typically the χ^2 constraint. This is done by maximizing the function Θ defined as[25]

$$\Theta = \alpha S(\boldsymbol{f}, \boldsymbol{m}) - L(\boldsymbol{D}, \boldsymbol{f}) \tag{2}$$

$$S(\boldsymbol{f}, \boldsymbol{m}) = \sum_i f_i - m_i - f_i \log \frac{f_i}{m_i} \tag{3}$$

where $S(\boldsymbol{f},\boldsymbol{m})$ is the entropy of a positive, additive, function as defined by Skilling,[25,26] $L(\boldsymbol{D},\boldsymbol{f})$ is a goodness-of-fit parameter (such as $(1/2)\chi^2$ for Gaussian-distributed errors), $\boldsymbol{D}$ is a vector comprising the measured data, $\boldsymbol{f}$ is the distribution to be recovered, and $\boldsymbol{m}$ is an initial model that encodes prior knowledge of the distribution. Although the entropy formalism requires an initial model in the prior, the shape of the resulting distribution $\boldsymbol{f}$ has not been parametrized as a Gaussian or any other form. The weight of the entropy relative to the goodness-of-fit constraint is controlled by the Lagrange multiplier α.

Choice of the Lagrange Multiplier α. All maximum entropy methods face the problem of how to set the value of the Lagrange multiplier α. In non-Bayesian maximum entropy, α is typically selected such that the value of χ^2 for the recovered distribution is equal to the number of data points.[25,26,32] With the value of α selected, the distribution $\boldsymbol{f}$ which maximizes expression 2 is determined. Selection of α by this means has been termed historic maximum entropy.[25,26,32] This choice for α is convenient, but χ^2 is equal to the number of data points in the fit only on average. For a given measurement the value of χ^2 is a random variable that follows its own distribution. The situation is similar with other goodness-of-fit parameters.

To optimize the choice of α Gull and Skilling developed classic maximum entropy, which uses Bayesian analysis for selection of the value of the Lagrange multiplier α. In classic maximum entropy the information entropy determines a prior probability of the form[25,33]

$$P(\boldsymbol{f}|\alpha, \boldsymbol{m}) = \frac{e^{\alpha S(\boldsymbol{f},\boldsymbol{m})}}{Z(\alpha)} \tag{4}$$

Here, $Z(\alpha)$ is a normalization factor. This prior is inserted into Bayes' theorem, together with the likelihood, to determine the probability $P(\boldsymbol{f}|\boldsymbol{D},\boldsymbol{m})$ of any positive additive distribution $\boldsymbol{f}$, given the experimental data $\boldsymbol{D}$ and the initial model $\boldsymbol{m}$. The Lagrange multiplier α is determined by maximizing $P(\boldsymbol{D}|\alpha,\boldsymbol{m})$, the probability of the data given α, which is referred to as the α-evidence.[26] The α-evidence is brought into the analysis through the full joint probability of $\boldsymbol{D}$, $\boldsymbol{f}$, and α. This treatment of α constitutes a rigorous way to select α and includes information concerning the entire distribution of the likelihood function by returning not one recovered $\boldsymbol{f}$ distribution, but rather the probability of any $\boldsymbol{f}$ distribution, $P(\boldsymbol{f}|\boldsymbol{D},\boldsymbol{m})$.[25,26,32] A full description of the classic maximum entropy approach can be found in the original papers,[25,32,33] the MemSys5 user's manual,[26] the book by Sivia,[21] and a paper by Brochon.[34]

The Bayesian choice of α usually results in a value of χ^2 that is less than the number of data points, as there will be a number of statistically significant features in the reconstructed $\boldsymbol{f}$. These features constitute "good" degrees of freedom that do not contribute to the misfit. It should be emphasized, however, that the actual value of χ^2 is not explicitly used in classic maximum entropy, though it can be used to make a self-consistent estimate of the noise level.

The Bayesian approach to maximum entropy also allows quantification of $\boldsymbol{f}$ over regions of its domain. This can be done either through the use of a mask function or by directly calculating the average and standard deviation of each peak or region from random samples of the inference distribution, $P(\boldsymbol{f}|\boldsymbol{D},\boldsymbol{m})$.[32] For example, if a mask function is chosen such that its value is unity over a range in the domain of $\boldsymbol{f}$ and zero elsewhere, the inference gives the average amplitude and error of $P(\boldsymbol{f}|\boldsymbol{D},\boldsymbol{m})$ contained within the masked region.[32]

Finally, intrinsic correlation functions (ICFs) can be used to allow correlations between elements of the reconstructed image, $\boldsymbol{f}$.[25,26] ICFs consist of a blurring function that acts upon a hidden distribution, $\boldsymbol{h}$. The blurring reduces the effective number of degrees of freedom in the model, and can therefore increase the value of the evidence, provided that the fit to the data is not compromised. On the other hand, if the blurring is too severe, the increase in the value of χ^2 will decrease the evidence. The hidden distribution can also be constructed to have multiple layers, where each layer increases the amount of blurring between elements. The visible distribution, $\boldsymbol{f}$, is then calculated from a weighted sum of the layers of $\boldsymbol{h}$. The entropy formalism is applied to the hidden distribution. The amount of blurring required by the data is controlled by the fit, although the experimenter must choose how many layers to place into the hidden distribution. This can again be determined using the Bayesian evidence. An increase in the evidence suggests that the blurring model is more appropriate for the data. The use of an ICF introduces smoothness into the recovered distribution, while allowing the overall shape of the distribution to be dictated by the data.

Photon Distribution Analysis. Single-pair FRET burst measurements (spFRET) are used to determine the efficiency of energy transfer between two fluorophores. In many cases the fluorophores are linked to specific locations on a biomolecule,

dx.doi.org/10.1021/jp209861u | J. Phys. Chem. B 2012, 116, 4006−4015

and the FRET efficiencies are then related to the conformational distribution of the biomolecule. This distribution is an important property of the biomolecule relating to its structure and function; it is what the experimenter wishes to determine.

A very dilute solution of the FRET species is first placed on a microscope. Individual molecules diffuse through the focus of a laser, and photons that are emitted by the FRET dye pair are detected. The number of detectors used can vary, but at least two are needed. Each detector monitors the emission of one of the dye probes.

To analyze spFRET data, a stochastic approach is used. The probability of measuring a particular combination of signals in two detectors is related to the FRET efficiency distributions of the molecules in the sample.[8,10] For measurements of multiple molecular species diffusing freely in solution, Seidel and co-workers showed that the probability of detecting S1,S2 counts in two detectors in a time bin is given by[8,24]

$$P(S1, S2) = \int \sum_{F1+B1=S1;F2+B2=S2} P(F1|F, E_{app}) P_{\langle B1\rangle}(B1) P_{\langle B2\rangle}(B2)[\sum_j P_s(j) P_j(F, E_{app})] \, dE_{app} \quad (5)$$

where $F1$ and $F2$ are the numbers of fluorescence photons detected in each detection channel, $F = F1 + F2$, $B1$, and $B2$ are the numbers of background counts detected in each channel, and $P(F1|F,E_{app})$ is the probability of detecting $F1$ photons in channel one given that F photons were detected in both channels and emitted by a molecule with apparent FRET efficiency E_{app}. Also, $P(F1|F,E_{app})$ is the binomial distribution with probability of success $1 - E_{app}$, and $P_{\langle B1\rangle}(B1)$ and $P_{\langle B2\rangle}(B2)$ are the Poisson probabilities of detecting $B1$ and $B2$ background counts in the time bin, with averages $\langle B1\rangle$ and $\langle B2\rangle$; $P(F,E_{app})$ is the joint distribution for the total number of fluorescence photons and apparent FRET efficiency, j is an index denoting the molecular species, and $P_s(j)$ is the probability of observing the jth molecular species. Equation 5 is obtained in the single-molecule limit in which only one molecule is present in the focal volume. Weiss and co-workers found a similar equation that describes the distribution of counts per single-molecule burst.[10] In our instrumental configuration, detection channel 1 measures fluorescence from the donor dye, and channel 2 detects photons from the acceptor dye. By explicitly accounting for measurement noise and Poisson statistics in burst measurements, eq 5 and the analogous equation for burst-wise probabilities[10] can account for the rugged appearance of experimental FRET histograms.[8,10]

In previous treatments, the joint distribution $P(F,E_{app})$ was factorized as $P(F)P(E_{app})$ according to the assumption that the distributions are independent.[8,9,24,35] The distribution $P(F)$ was then expanded into a sum over volume elements of the point spread function. Each element was defined by an average brightness q_i and the probability that a molecule resides within a particular volume element, $P(q_i)$, an approach analogous to the fluorescence intensity distribution analysis (FIDA).[35,36] The distribution $P(F)$ was determined as an intermediate result by recovering the distribution $P(q_i)$ for a preselected range of q_i. A historic maximum entropy approach was used with

$$P(S) = P(F) \otimes P(B) \quad (6)$$

where $\otimes$ denotes convolution, $P(S)$ is the experimentally measured distribution $P(S1 + S2)$, and $P(B)$ is a Poisson distribution for the background counts, with a mean determined by a background measurement. After $P(q_i)$ had been determined, it was used in a second historic maximum entropy fit to recover $P(E_{app})$.[24,35]

Recovery of $P_A(E_{app},F)$. We describe in this paper the direct model-free recovery of the average joint distribution $P_A(F,E_{app})$ by the classic maximum entropy technique. This approach avoids the use of an expansion over volume elements and recovers the entire joint distribution directly. Classic maximum entropy has not previously been used to analyze burst measurements. It has the advantages of returning the inference distribution, the evidence, and the value of α determined via the α-evidence. Furthermore, ICFs with multiple blurring layers can be used to incorporate smoothness into the model, and the evidence can be used to determine the number of layers appropriate for the data.

The average joint distribution can be obtained as a sum over all molecular species present in solution

$$P_A(F, E_{app}) = \sum_j P_s(j) P_j(F, E_{app}) \quad (7)$$

where $P_A(F,E_{app})$ is the average joint distribution. After recovering $P_A(F,E_{app})$, we determine the overall apparent FRET efficiency distribution $P_A(E_{app})$ and the overall average fluorescence distribution $P_A(F)$ by marginalization over F and E_{app}, respectively

$$P_A(E_{app}) = \sum_F P_A(F, E_{app}) \quad (8)$$

$$P_A(F) = \int P_A(F, E_{app}) dE_{app} \quad (9)$$

The apparent FRET efficiency is related to the true FRET efficiency E between the dyes by[8]

$$E_{app}(E) = 1 - \frac{1}{1 + \frac{g_2\varphi_2}{g_1\varphi_1}\left(\frac{E}{1-E}\right) + \frac{g_2}{g_1}c} \quad (10)$$

where g_2/g_1 is a ratio of detection efficiencies in the red and green channels, φ_1 and φ_2 respectively denote the quantum yields of the donor and acceptor dyes, and c is the fraction of donor dye emission detected in the acceptor detection channel. For a single diffusing molecular species, the distribution $P_A(E_{app})$ can be converted to a distribution of distances $P_A(R)$ by[8]

$$P_A(R) = P_A(E_{app}(R)) \left|\frac{dE_{app}(R)}{dR}\right|_R \quad (11)$$

When the joint distribution of a single species can be determined, or when the contributions of each species to the average joint distribution are intrinsically separate, a potential of the mean force can be obtained from the distance distribution. This is done by taking the negative logarithm of the distance distribution

$$-\ln P(R) = \beta V(R) + \ln\left(\int e^{-\beta V(R)} \, dR\right) \quad (12)$$

where β is the inverse temperature and $V(R)$ is the potential of the mean force. The integral amounts to a normalization

dx.doi.org/10.1021/jp209861u | J. Phys. Chem. B 2012, 116, 4006–4015

constant. This means the shape of the estimated potential of the mean force can be calculated from single-molecule burst measurements provided the molecules do not exhibit conformational dynamics during a time bin.[37]

MATERIALS AND METHODS

Simulations. Simulations were performed by generating random numbers from appropriate probability distributions, based on the theory of photon distribution analysis (PDA), eq 5. A detailed description of our simulations is as follows. Two arrays of integers were generated to represent the number of counts detected by two detectors in time bins of equal length. In all simulations, two molecular species with differing distributions $P(F,E_{app})$ were simulated such that $P(F,E_{app}) = P(F)P(E_{app})$. For each time bin, photon counts were generated for one of the species. The contribution of each species to the simulated $P_A(F,E_{app})$ was controlled by the number of array elements associated with each species. The total number of photons F detected by both detectors was simulated by selecting a random integer from a Poisson distribution with a given mean. After selection of the total number of photons, the FRET efficiency of the molecule for that time bin was chosen by selecting a random number from a Gaussian distribution having specified width and mean. Once the efficiency for a bin had been selected, a random number was chosen from the binomial distribution with the appropriate total number of photons and apparent FRET efficiency. This number of photons was input to the donor array, representing $F1$, and a number $(F - F1)$ was placed into the acceptor array. Finally, two random numbers representing $B1$ and $B2$ from independent Poisson distributions were added to the corresponding $F1$ and $F2$ arrays. The resulting signal arrays were then ordered into a two-dimensional histogram of counts $N(S1,S2)$, which differs from $P(S1,S2)$ only by a normalization constant. All simulations displayed in this paper contained 70,000 total counts in the data histogram. Random numbers were generated using programming code from Numerical Recipes.[38]

Sample Preparation. Unlabeled and dye-labeled single-stranded DNA was purchased from Eurofins MWG Operon (Huntsville, AL). The forward strand was labeled on the 5′ end with Alexa Flour 594 or Cy5 through an amino C6 linkage (sequence: 5′ −TGTAAAACGAGAGAGCCTAAAACGATC-3′), or was unlabeled. The reverse strand was labeled with Alexa Flour 488 via a deoxyguanosinethymidine (dT) base with a six carbon linker (sequence: 5′-GATCGTTTTAGGC-dTCTCTCGTTTTACA-3′), resulting in a 12 base pair separation between the dye labeling sites. All DNA species were dissolved in buffer EB (Qiagen, Valencia, CA), 10 mM Tris-HCl, pH 8.5. Single labeled strands were mixed with unlabeled complementary strands in the ratios 2:1 and 1:1. The FRET species was formed by mixing Cy5 labeled single-stranded DNA with its complementary strand containing AF488 in a 2:1 ratio (Cy5 strand:AF488 strand), or by mixing AF594 labeled single strands with the AF488 strand in a 1:1 ratio. All mixed samples were placed in a PCR Sprint thermal cycler for annealing. The temperature of each mixture was raised to 95 °C for 3 min and then lowered in steps to 58.1 °C, 53.1 °C, and 20 °C, in 3 min intervals. The final double-stranded DNA solutions were stored in a freezer at −20 °C prior to analysis. The resulting samples included double-stranded DNA labeled with AF488 and AF594 (DA594), double-stranded DNA labeled with AF488 and Cy5 (DACy5), double-stranded DNA labeled with AF488 (D488), and double-stranded DNA labeled with Cy5 (ACy5).

Single-Molecule Burst Measurements. All single-molecule burst measurements were performed with an inverted fluorescence microscope system. Light produced from an argon-ion laser at 488 nm (2201-20SL, JDS Uniphase, San Jose, CA) was polarized circularly and directed to the sample through a Nikon TE2000 microscope fitted with a 1.2 N.A. water immersion objective (UPLSAPO, Olympus, Center Valley, PA). Fluorescence from the sample was collected and separated from the excitation light using a z488/594rpc dichroic beam splitter (Chroma Technology Corp., Rockingham, VT). The emission of the dyes was then separated into two detection channels by a 565 DCLP dichroic beam splitter (Chroma). Emission from each channel was focused through a 75 μm pinhole before being refocused onto an avalanche photodiode (SPCM AQR-14, Perkin-Elmer). Emission filters were placed in front of the detectors, to further select for dye fluorescence: ET535/50m for donor emission (Chroma), and HQ 650/100m-2p or HQ 667LP (Chroma) for acceptor emission from DA594 or DACy5.

TTL pulses produced by the avalanche photodiode detectors were logged with a National Instruments PCI-6602 counter timer card, which was custom programmed in LabVIEW. Each pulse was time stamped with an 80 MHz clock (12.5 ns per tick), giving rise to two data arrays containing the arrival times of photons relative to the beginning of the experiment. These arrival times were binned into 300-μs time bins and ordered as a histogram. This is the experimental $N(S1,S2)$ histogram for subsequent classic maximum entropy analysis. Only histogram bins corresponding to a total signal greater than 10 counts and less than or equal to a selected maximum signal were included in the fit. The maximum signals were 70 and 80 counts for DA594 and DACy5 respectively.

A Petri dish fitted with a glass coverslip was used as a sample cell (MatTek Corp., Ashland, MA). The sample cell was prepared by placing 2 mL of 2 mg/mL Bovine Serum Albumin (BSA) in the dish for approximately 5 min. The BSA was used to prevent DNA from sticking to the glass coverslip. The cell was then emptied and dried with nitrogen gas. The FRET-labeled DNA sample, 2 mL at 20 pM, was added to the cell and placed on the microscope. The focus of the laser was adjusted to be positioned approximately 30 μm into the solution. The laser power was set at 40 μW for DA594, and 60 μW for DACy5 experiments. The laser power was measured before the beam entered the microscope. Three sets of burst data were collected for approximately 1 h each for DA594, and two sets of an hour each for DACy5. The average background count rate was determined for each measurement with a buffer-only sample. The average number of background counts in each channel in a 300-μs time bin was then used in eq 5. The average numbers of background counts measured for the green and red channels in a 300-μs time bin were 0.137 and 0.0783 with the DA594 filter set and 0.154 and 0.112 for the DACy5 filter set.

Instrumental Detection Factors and Quantum Yields. A Cary Eclipse fluorimeter was used to measure the fluorescence spectra for DACy5, ACy5, D488, and DA594. The fluorescence spectrum of AF594 with double-stranded DNA was obtained by exciting DA594 at 575 nm; this wavelength does not excite the donor dye. The emission spectra were corrected for the wavelength-dependent detection efficiencies of the fluorimeter by using the standard emission spectra of Coumarin 153, LDS 751, DCM, and fluorescein.[39]

The shapes of the emission spectra for DA594 (acceptor excitation), D488, and ACy5 were fitted to sums of Gaussians. The bulk FRET efficiencies with DACy5 and DA594 were determined from fits to fixed shapes of the spectra from the individual dyes while allowing the amplitude for each dye species to vary. The resulting contribution to the fit from each dye was integrated and was used to calculate the bulk FRET efficiency, weighted appropriately by the quantum yields.

D488 was used to determine the cross talk of the donor dye into the acceptor channel on the microscope. The sample, comprising 30.0 μL of 8 nM D488, was placed on a BSA-treated coverslip for data collection. The average intensity was measured in both detection channels. At low laser intensities and large sample concentrations the background intensity was negligible. The ratio of the average emission intensity in the acceptor channel to that in the donor channel gives the term $(g_2/g_1)c$ directly (eq 10). The cross talk was found to be 0.046 for the DA594 filter set, and 0.0077 for the DACy5 filter set.

To measure the bulk apparent FRET efficiency, 30 μL of 8 nm DA594 or of 10 nm DACy5 was placed on BSA-treated coverslips. Measurements were performed three times between burst data sets. The bulk E_{app} from each of the measurements was averaged; values were 0.619 ± 0.003 for DA594 and 0.303 ± 0.002 for DACy5.

The quantum yields of dyes bound covalently to DNA were measured using either the standard reference technique[39] or the Strickler–Berg method.[40,41] Relative to fluorescein (quantum yield 0.925 in 0.1 M NaOH[42]), the quantum yield of AF488 on DNA, D488, was determined to be 0.95. The quantum yield of DNA-AF594 and DNA-Cy5 were 0.71 and 0.45, respectively, by the Strickler–Berg method.[40,41] All quantum yield values determined have an uncertainty of approximately 5%. Finally, the bulk E_{app} measured on the microscope, the bulk FRET efficiency measured on the fluorimeter, the quantum yields and the cross talk were used in eq 8 to determine the ratio of detector efficiencies g_2/g_1, which was 1.05 for the DA594 filter set and 0.71 for the DACy5 filter set.

Recovery of the Joint Distribution. MemSys5 in C was obtained from Maximum Entropy Data Consultants (Cambridge, U.K.). The MemSys5 kernel contains a set of functions that solves the classic maximum entropy data problem in an iterative manner. We custom-programmed function calls to the kernel so as to interface its minimization calculations with the photon distribution analysis. A Poisson likelihood was chosen for all classic maximum entropy fits. The initial model m in the entropy functional (eq 3) was chosen to be flat, with the dimensions of the joint distribution $P_A(E_{app},F)$ multiplied by the number of blurring layers. Each layer increases the blurring by a factor of 2 in both dimensions. Fits were performed with up to 8 layers. The continuous distribution in E_{app}, $P(E_{app}|F)$, was approximated by 100 equally spaced points. The value of F ranged from 5 to 55 in the simulations, 10 to 70 for DA594 measurements, and 10 to 80 for DACy5 measurements.

RESULTS

Simulations. Simulations were used to test the reliability of classic maximum entropy analysis in recovering the distribution $P_A(E_{app},F)$. Simulated data were generated directly from the PDA theory with model distributions $P(E_{app})$ and $P(F)$.[8] The distribution $P_A(E_{app},F)$ recovered from classic maximum entropy analysis could then be compared with the original distribution. Accurate recovery of the initially assumed distribution indicates that the analysis can reliably determine the average distribution that is exhibited by molecules during an experiment.

Four simulations were performed as described in Materials and Methods, each with two molecular species. (For simulations with three molecular species, see the Supporting Information.) The centers of the input Gaussian $P(E_{app})$ for each species were changed in successive simulations, to test the ability to recover different multimodal distributions. Table 1 summarizes the parameters used in each simulation.

Table 1. Parameters Used in Four Simulations[a]

simulation		$P(E)$		$P(F)$
run	species	mean	σ	average
1	1	0.7	0.07	20
	2	0.3	0.05	35
2	1	0.6	0.07	20
	2	0.4	0.05	35
3	1	0.575	0.07	20
	2	0.425	0.05	35
4	1	0.55	0.07	20
	2	0.45	0.05	35

[a]Each simulation was performed with two species. $P(E)$ for each species was a Gaussian distribution, and $P(F)$ was a Poisson distribution.

Classic maximum entropy fits were performed with various numbers of blurring layers in the hidden distribution. The recovered distribution is compared with the simulated data in Figure 1. The distribution $P_A(E_{app},F)$ in panel A was used to create the simulated data histogram, $N(S1,S2)$, shown in panel B. The value of $N(S1,S2)$ was read into the maximum entropy routine for fitting. Although classic maximum entropy returns the probability, known as the inference, of any $P_A(E_{app},F)$ distribution, we plot the most probable $P_A(E_{app},F)$ in panel C.

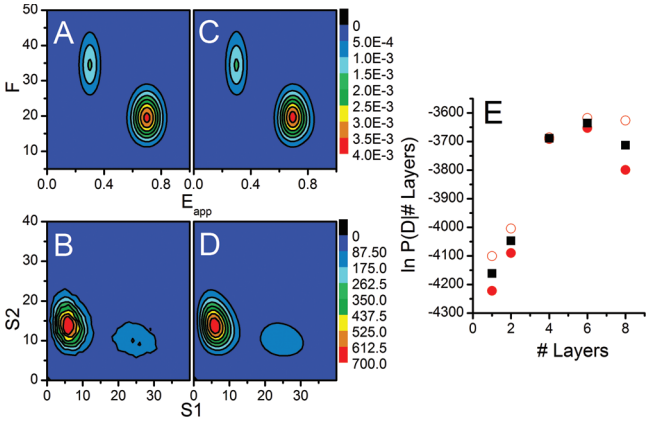


Figure 1. Plots corresponding to the first simulation are shown in panels A and B. Output from the maximum entropy recovery with six layers of blurring is shown in panels C and D. (A) Distribution $P_A(E_{app},F)$ from which random samples were taken to create the simulated data. (B) Data histogram, $N(S1,S2)$, from simulation run 1. (C) Most probable $P_A(E_{app},F)$ recovered from the maximum entropy fit. The most probable distribution was used to generate $M(S1,S2)$. (D) Mock data histogram generated from the maximum entropy fit, $M(S1,S2)$. (E) The Bayesian evidence conditional upon the number of blurring layers in the hidden distribution (black squares). Red open circles and solid circles depict the upper and lower error bounds of the evidence.

dx.doi.org/10.1021/jp209861u | J. Phys. Chem. B 2012, 116, 4006–4015

The MemSys5 algorithm approximates the inference distribution in a second-order Taylor expansion around the most probable $P_A(E_{app},F)$. In this approximation the most probable distribution must also be the average distribution. Mock data $M(S1,S2)$, generated from the most probable $P_A(E_{app},F)$, is displayed in panel D. There is close resemblance of both the mock data and the recovered $P_A(E_{app},F)$ to the simulated data and actual $P_A(E_{app},F)$. Six layers of blurring maximized the evidence for the first simulation, Figure 1E. ICFs account for the fact that a model incorporating "neighboring" two-dimensional histograms of FRET efficiency E and intensity F accounts better for the experimental histogram of counts in $S1$ and $S2$ than a model without blurring, as demonstrated by the increased value of the evidence. The recovery without ICFs is displayed in the Supporting Information along with the Bayesian evidence.

The contribution to $P_A(E_{app},F)$ from each species was determined with a mask function. The mask was chosen to determine the average probability contained within each peak, along with its error, and was applied to the nonblurred recovery for the sake of simplicity. The total probability and error in the apparent efficiency ranges of 0.101 to 0.467 and 0.475 to 0.909, summed over all F, were respectively 0.285 ± 0.003 and 0.711 ± 0.004. These values are almost exactly the amplitudes of the species from the simulation, which were 0.286 and 0.714.

A helpful way to visualize the fit is to calculate a one-dimensional histogram from a ratio such as the experimental FRET efficiency. For a given time bin, this is simply a histogram of the number of counts detected in the acceptor channel divided by the total number of counts for that time bin, $E_{exp} = S_2/(S_1 + S_2)$. This type of histogram is commonly plotted to present single-molecule burst data. Figure 2B shows the experimental FRET efficiency histogram for the simulated data (gray) and corresponding maximum entropy fit (red). Figure 2A shows the weighted residuals calculated from the histograms. Figure 2C shows the weighted residuals calculated from the two-dimensional histograms $N(S1,S2)$, $M(S1,S2)$ used in the fitting procedure. The randomness and amplitude in the residuals in Figure 2, panels B and C show that the mock data accurately reproduces the simulated data. The slightly negative background region in Figure 2C (light green) notably arises where there are no counts in a particular bin in the data histogram. This is due to the inclusion of histogram bins with few or no data points but with a slight positive amplitude in the mock data. The solid yellow area at the upper right of the plot is outside the fitting range for the data set and was not used in the calculations. The range for the recovered distribution of F was chosen to cover nearly all of the expected range of $P_A(F)$. The range in F selected has little effect on the recovered distribution as discussed in the Supporting Information. Variation in the background level also has little effect, as shown in the Supporting Information.

The distribution $P_A(E_{app},F)$ was marginalized over E_{app} and F according to eq 7 and the resulting marginals were compared against the distribution from which random numbers were generated for the simulation; see Figure 3. The marginal distributions recovered both with and without ICFs reproduce the actual marginal distributions very well. The recovery without blurring is not as smooth as the visible blurred distribution. This is because each point in the nonblurred distribution is independent.

The recovered $P_A(E_{app},F)$ distributions for the last three simulations were marginalized over E_{app} and F; the results are shown in Figure 4. Panels A, C, and E show the recovered $P_A(F)$, and Panels B, D, and F show the corresponding $P_A(E_{app})$. In all panels, solid lines indicate the actual marginalized distributions used in the simulations. Red squares

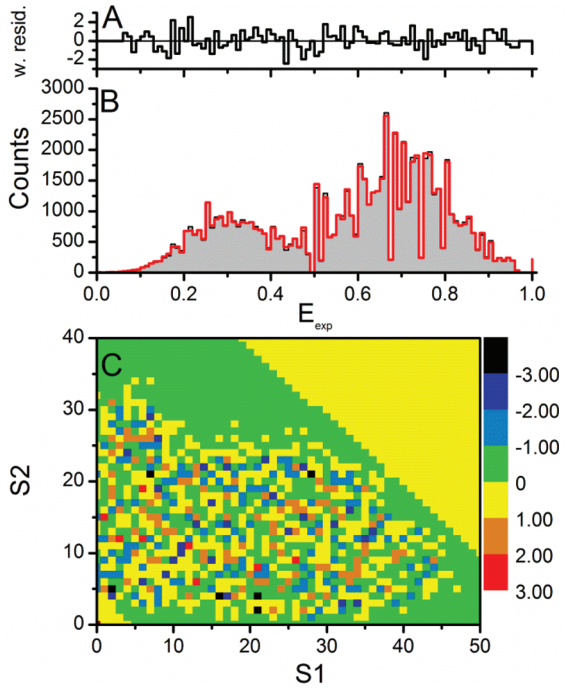


Figure 2. (A) Weighted residuals between the data and mock data for the experimental FRET efficiency histogram. The residuals were calculated as $(D_i - M_i)/(D_i)^{1/2}$, where D_i is a data point and M_i is the corresponding data point calculated from the recovered $P_A(E_{app},F)$. Histogram bins with less than 10 counts were excluded from the residuals. (B) A one-dimensional experimental FRET efficiency histogram corresponding to the data in Figure 1. Data are shown with a black line and solid gray fill. Mock data generated from the fit are shown in red. (C) Residuals calculated by the maximum entropy routine for the two-dimensional signal histogram. Residuals were calculated by the MemSys5 kernel as $(D_i - M_i)/(D_i + 1)^{1/2}$, and all bins were included.

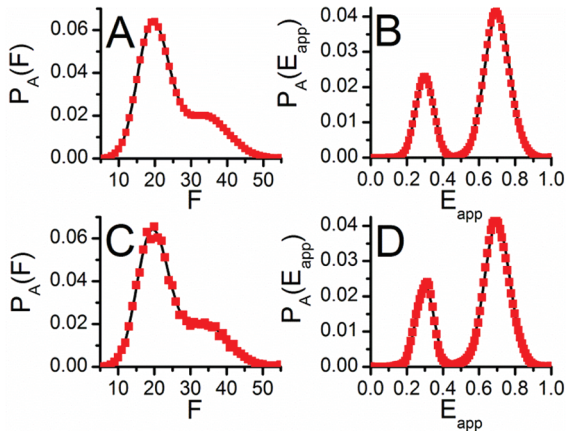


Figure 3. Comparison of fits to data from the first simulation with and without ICFs. In all panels, red squares indicate the recovered marginalized distribution, and the solid line is the marginalized distribution used in the simulation. (A) and (B) show the recovered $P_A(F)$ and $P_A(E_{app})$ with six layers of blurring. (C) and (D) show the recovered $P_A(F)$ and $P_A(E_{app})$ without any use of ICFs.

dx.doi.org/10.1021/jp209861u | J. Phys. Chem. B 2012, 116, 4006–4015

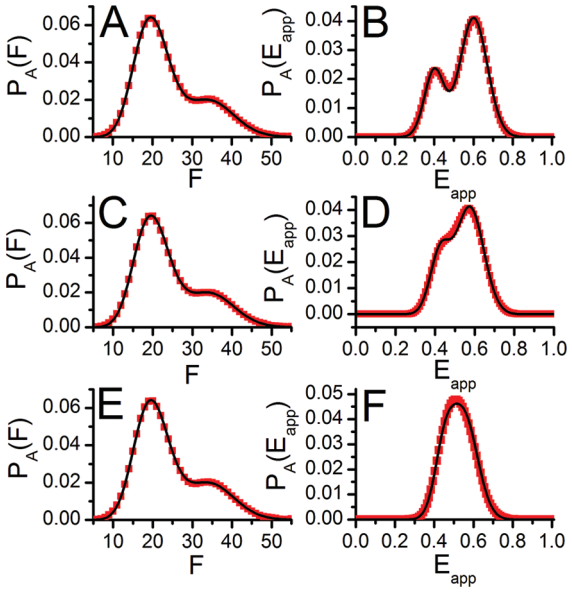


Figure 4. (A) and (B) Recovered $P_A(F)$ and $P_A(E_{app})$ for simulation 2. (C) and (D) Recovered $P_A(F)$ and $P_A(E_{app})$ for simulation 3. (E) and (F) Recovered $P_A(F)$ and $P_A(E_{app})$ for simulation 4. In all panels, red squares show the recovered marginalized distribution with six layers of blurring. The solid line corresponds to the underlying marginalized distribution used in the simulation. Table 1 gives the parameters used for each simulation.

indicate the fit. In all simulations, the recovered most probable distributions are in excellent agreement with the theoretical distributions. As overlap increases between states in the underlying distribution, the recovered distribution begins to blur the two states together. Furthermore, the recoveries of the proper shapes for $P_A(F)$ and $P_A(E_{app})$ are not strongly interdependent. As shown in Figures 3 and 4, $P_A(F)$, which is the same in all simulations, is recovered accurately even though the overall shape of $P_A(E_{app})$ is different in each simulation. Recovery of a distribution with three FRET states is demonstrated in the Supporting Information.

FRET Labeled DNA Measurements. Burst measurements were performed on DA594 and DACy5 as described in Materials and Methods. The recovered $P_A(F,E_{app})$ and $P_A(E_{app})$ for both samples show two major peaks. Figure 5 displays the recovered distribution for one DA594 measurement. The peak at an apparent FRET efficiency of 0.044 is due to the presence of a donor-only species. This peak may come from single stranded DNA labeled with AF488 that remains after annealing. The peak may also include contributions from DA594 with a photobleached acceptor. The peak at an apparent FRET efficiency of 0.643 is due to the FRET species. Figure 5B compares $P_A(F)$ and $P(S)$. It is apparent that $P_A(F)$ follows $P(S)$ closely. This is as expected from the low average background count rate. A low-signal threshold value of 10 counts per bin was used to ensure that a molecule was present in the focal volume during a time bin. For this threshold value, with typical average background count rates, it is unlikely that the background would contribute significantly to the detected signal. As a result, $P_A(F)$ is expected to follow $P(S)$ closely. Simulations to test the effect of background level on the recovered distributions show little effect with background levels up to 20 times that expected in experiments (see Supporting Information). An advantage of this approach is that $P_j(F)$ can be determined for a single species, j, if its distribution does not overlap with the distribution of another species in $P_A(E_{app},F)$. Specifically, $P_j(F)$ is determined by marginalizing the joint distribution over E_{app} in the range associated with the single species. Any $P_j(F)$ determined in this manner need not resemble $P(S)$ closely; only the species averaged $P_A(F)$ is expected to resemble $P(S)$. The experimental FRET efficiency histogram is shown in Figure 5D, together with its weighted residuals. We see that the recovery is an acceptable fit to the data.

Each of three data sets of DA594 was fitted using classic maximum entropy. Figure 6, panels A and B, shows overlays of $P_A(E_{app})$ and $P_A(F)$ from each measurement. Each recovery used four layers of blurring in the hidden distribution. The distributions were highly reproducible. The total probability amplitude contained within the donor-only and FRET-species peaks was estimated from the inference distribution for each

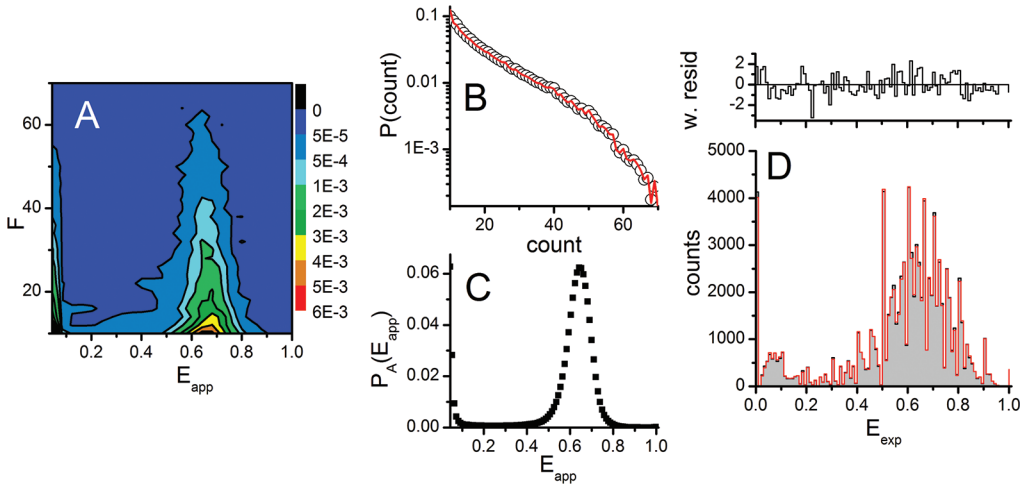


Figure 5. Most probable recovered distribution for double-stranded DNA labeled with AF488 and AF594. (A) Joint distribution $P_A(E_{app},F)$. (B) $P_A(F)$ obtained by marginalization of the joint distribution over E_{app} (red line) and $P(S)$ obtained from the experimental data (black circles). (C) $P_A(E_{app})$ obtained by marginalization of the joint distribution over F. (D) Experimental FRET efficiency histogram, with $N(S1,S2)$ in gray and $M(S1,S2)$ in red (lower panel). The weighted residuals are shown in the upper panel. Four layers of blurring were used in the hidden distribution.

dx.doi.org/10.1021/jp209861u | J. Phys. Chem. B 2012, 116, 4006–4015

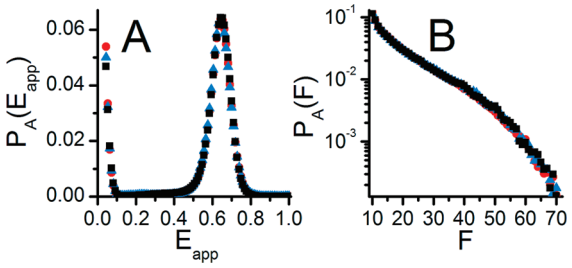


Figure 6. (A) Overlay of $P_A(E_{app})$ recovered from three consecutive data collection intervals for the DA594 sample. (B) Overlay of $P_A(F)$ for the same data sets as in panel A. In all panels, black squares indicate the same data set as displayed in Figure 5.

data set from the nonblurred recovery. The mask was selected to have value 1 in the range covering most of each peak; the results are summarized in Table 2. The amplitude and error

Table 2. Average Probability for Two Ranges of E_{app} Marginalized over All F

run	E_{app} range 0.411 to 0.817	E_{app} range: 0.044 to 0.208
1	0.865 ± 0.005	0.122 ± 0.003
2	0.860 ± 0.006	0.119 ± 0.003
3	0.858 ± 0.005	0.124 ± 0.003

associated with all three data sets show very good agreement. Plots corresponding to the recoveries of DACy5 can be found in the Supporting Information.

The averages of the most probable $P_A(E_{app})$ for DA594 and DACy5 were calculated from repeated data sets and transformed into distributions in distance, $P(R)$. This $P(R)$ was then used to determine an estimated potential of the mean force (PMF), shown in Figure 7. The Förster distance (R_o) calculated for DA594 was 58.8 Å, and for DACy5 it was 54.3 Å. Details of the calculation are set out in the Supporting Information. The maximum amplitude of the FRET species occurs at 51.4 ± 0.5 Å for DA594 and 53.7 ± 0.5 Å for DACy5.

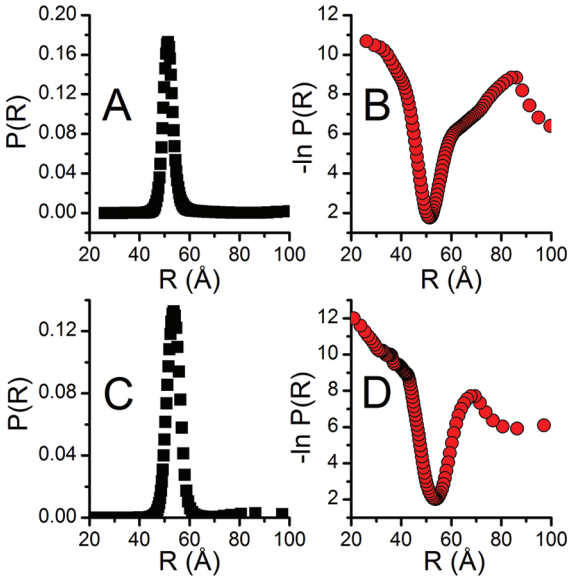


Figure 7. (A) Distance distribution for DA594. (B) Estimated potential of the mean force for DA594. (C) Distance distribution from DACy5. (D) Estimated potential of the mean force for DACy5.

We estimated the error in the peak position by propagating the estimated 5% error associated with our quantum yield measurements to R_o. The average $P_A(E_{app})$ distributions were then transformed to $P(R)$ type using the upper and lower error bounds for R_o. For both dye pairs, the PMF is actually a combination of two potentials, one for the donor-only species, which dominates at long distances, and one for the FRET species. This is an insightful test of the estimation of PMFs from model-free spFRET analysis, because it demonstrates the exaggeration of small details in the distance distribution when taking the logarithm to create the PMF. In particular, structure having low probability amplitude within the distribution is exaggerated, giving rise to visible structure in the PMF. The shape of this structure is not well determined. Caution is therefore necessary when interpreting the potential corresponding to structures with low probability amplitude. Furthermore, the sensitivity of the FRET efficiency to distance is greatly reduced at long and short distances. Sharp structures in $P_A(E_{app})$ located at very high or low values of E_{app} are transformed to features with broad widths in the distance distribution and in the PMF.

DISCUSSION

Distributions of single-molecule distances in DNA have been reported by several groups.[8,10,43,44] In particular, our results are consistent with published results from the Seidel group for the same sequence of double stranded DNA with the dye pair Alexa 488 and Cy5.[8] In their work, the DNA and dye pair data were fitted to a Gaussian distribution of distances, resulting in a width of 2.3 Å centered at 53.9 Å. To compare the width of our classic maximum entropy approach with the Gaussian distribution of distances reported previously, we fitted the average $P(R)$ distribution for DA594 and DACy5 by a Gaussian distribution. The width obtained from these fits was 2.6 Å for DACy5 and 2.0 Å for DA594, in good agreement with the previously published results.[8] Seidel and co-workers showed that the distribution of distances for short strands of DNA is probably due to effects of the local environment on the properties of the dyes, and is not a consequence of conformational heterogeneity of the DNA.[44] This may also be the case for DA594 and DACy5. The absolute distance measured between the transition dipole moments of the dyes does not strictly relate to the distance between the attachment points of the DNA base pairs to which the dyes are attached.[45] With similar linker lengths and sufficient orientational averaging, however, the absolute distance measured with different dye pairs should be similar. The absolute distances measured for DA594 and DACy5 are in good agreement, differing by only 2.3 Å. The distance measured for DACy5 is also in agreement with that published previously,[8] with a difference of only 0.2 Å.

Because of possible dye-biomolecule interactions, care is necessary in interpreting distributions for distance and the estimated PMF. Any broadening of distributions due to heterogeneous properties of the dyes will add to the width of the PMF estimated by the classic maximum entropy approach. Consequently, we can view the distance distribution and PMF calculated from experiment as a limiting width for the true distributions. If dye heterogeneities are present, the actual distributions must in fact be narrower than those recovered by our estimate.

Maximum entropy recovers the widest distribution consistent with the data. Far from being a drawback of the technique, this

dx.doi.org/10.1021/jp209861u | J. Phys. Chem. B 2012, 116, 4006–4015

is its strength: it never overcommits, since information theory guarantees that we cannot learn more than this. A narrow distribution would imply that we could predict future results confidently; every measurement would result in virtually the same outcome. The distribution recovered by maximum entropy analysis approaches the width of the true distribution as more data is collected. Happily, extremely large data sets to determine discrete states do not appear to be needed with the application of classic maximum entropy to fit $P_A(E_{app},F)$. This is due to the extreme sensitivity of FRET when looking for conformations near to R_o. A broad distribution in E_{app} near to R_o can collapse to a quite narrow distribution when transformed into distance.

CONCLUSIONS

We have demonstrated that the underlying average joint probability distribution $P_A(F,E_{app})$ can be recovered directly from spFRET data using classic maximum entropy analysis. The present technique is ideal for measurements in which a distribution of distances is expected for the molecule of interest. The distributions $P_A(F)$ and $P_A(E_{app})$ can be obtained by marginalizing the joint distribution over E_{app} and F respectively. For molecules with conformational distributions that are static on the time scale of photon binning, the potential of mean force can be estimated directly from the marginal distribution $P(E_{app})$. A great benefit of utilizing classic maximum entropy is the calculation of the Bayesian evidence. This procedure allows the experimenter to compare, quantitatively, the accuracy of various assumptions, such as the Bayesian priors chosen. Furthermore, the probability amplitude, and its average and error, can readily be calculated within an interval of E_{app}. These benefits are a consequence of the Bayesian framework, which determines the probability of any $P_A(E_{app},F)$ distribution given the data. Finally, ICFs can be used to induce smoothness in the recovered distribution.

We expect that the advantages of this approach will best be realized when working with flexible proteins where an arbitrary distribution of distances is likely. In that case there is no reason to suppose that the distance distribution conforms to any particular model, such as a sum of Gaussians, and it is better to tackle the problem of determining an arbitrary distribution using a model-free Bayesian approach. Regardless of whether the Bayesian approach is model-free, the evidence $P(\boldsymbol{D}|I)$ can be used to discriminate between analyses.

ASSOCIATED CONTENT

Supporting Information

(1) Plots of the Bayesian evidence for differing numbers of blurring layers in the hidden distributions of simulations; (2) recovery without ICFs for simulated data; (3) recovery of three FRET states; (4) demonstration of the dependence of the recovered distribution on the range of F; (5) effect of background level on the recovered distribution; (6) data and maximum entropy analysis for DNA labeled with AF488 and Cy5; (7) description of R_o calculation. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ckjohnson@ku.edu.

Notes

The authors declare the following competing financial interest(s): Professor Steve Gull is the Managing Director of Maximum Entropy Data Consultants, which supplies maximum entropy software (MEMSYS5) used in this project.

ACKNOWLEDGMENTS

We thank Prof. Matthew Antonik, Prof. Claus Seidel, and Stanislav Kalinin for valuable discussion concerning the photon distribution analysis and for providing sample programming code. We thank Prof. Emily Scott and Prof. Rich Givens for use of instrumentation in their laboratories. This work was supported by the National Science Foundation (CHE-0710515) and the American Heart Association (AHA-07557117Z). M.S.D. also thanks the NIH Dynamic Aspects of Chemical Biology Training Grant (GM08545) for support.

REFERENCES

(1) Ha, T.; Enderle, T.; Ogletree, D. F.; Chemla, D. S.; Selvin, P. R.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6264−6268.
(2) Rothenberg, E.; Ha, T. *Methods Mol. Biol.* **2010**, *587*, 29−43.
(3) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2754−2759.
(4) Deniz, A. A.; Laurence, T. A.; Dahan, M.; Chemla, D. S.; Schultz, P. G.; Weiss, S. *Annu. Rev. Phys. Chem.* **2001**, *52*, 233−253.
(5) Michaelis, J. Quantitative distance and position measurement using single-molecule FRET. In *Single Particle Tracking and Single Molecule Energy Transfer*; Brauchle, C., Lamb, D. C., Michaelis, J., Eds.; Wiley-VCH: Weinheim, Germany, 2010; pp 191−214.
(6) Sisamakis, E.; Valeri, A.; Kalinin, S.; Rothwell, P. J.; Seidel, C. A. M. *Methods Enzymol.* **2010**, *475*, 455−514.
(7) Roy, R.; Hohng, S.; Ha, T. *Nat. Methods* **2008**, *5*, 507−516.
(8) Antonik, M.; Felekyan, S.; Gaiduk, A.; Seidel, C. A. M. *J. Phys. Chem. B* **2006**, *110*, 6970−6978.
(9) Kalinin, S.; Valeri, A.; Antonik, M.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. B* **2010**, *114*, 7983−7995.
(10) Nir, E.; Michalet, X.; Hamadani, K. M.; Laurence, T. A.; Neuhauser, D.; Kovchegov, Y.; Weiss, S. *J. Phys. Chem. B* **2006**, *110*, 22103−22124.
(11) Gopich, I. V.; Szabo, A. *J. Phys. Chem. B* **2005**, *109*, 17683−17688.
(12) Gopich, I. V.; Szabo, A. *J. Phys. Chem. B* **2007**, *111*, 12925−12932.
(13) Bronson, J. E.; Fei, J. Y.; Hofman, J. M.; Gonzalez, R. L.; Wiggins, C. H. *Biophys. J.* **2009**, *97*, 3196−3205.
(14) Gopich, I. V.; Szabo, A. *J. Phys. Chem. B* **2009**, *113*, 10965−10973.
(15) Backovic, M.; Price, E. S.; Johnson, C. K.; Ralston, J. P. *J. Chem. Phys.* **2011**, *134*, 15.
(16) Kou, S. C.; Sunney Xie, X.; Liu, J. S. *J. R. Stat. Soc.: Ser. C* **2005**, *54*, 469−506.
(17) Witkoskie, J. B.; Cao, J. *J. Chem. Phys.* **2004**, *121*, 6373−6379.
(18) Witkoskie, J. B.; Cao, J. *J. Phys. Chem. B* **2008**, *112*, 5988−5996.
(19) Taylor, J. N.; Makarov, D. E.; Landes, C. F. *Biophys. J.* **2010**, *98*, 164−173.
(20) Ensign, D. L.; Pande, V. S. *J. Phys. Chem. B* **2010**, *114*, 280−292.
(21) Sivia, D. S. *Data analysis: a Bayesian tutorial*; Oxford University Press: New York, 1996.
(22) Watkins, L. P.; Chang, H. Y.; Yang, H. *J. Phys. Chem. A* **2006**, *110*, 5191−5203.
(23) Hanson, J. A.; Duderstadt, K.; Watkins, L. P.; Bhattacharyya, S.; Brokaw, J.; Chu, J. W.; Yang, H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18055−18060.
(24) Kalinin, S.; Felekyan, S.; Valeri, A.; Seidel, C. A. M. *J. Phys. Chem. B* **2008**, *112*, 8361−8374.

(25) Gull, S. F. Developments in Maximum Entropy Data Analysis. In *Maximum Entropy and Bayesian Methods*; Skilling, J., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989; pp 53−71.
(26) Gull, S. F.; Skilling, J. Quantified Maximum Entropy Memsys5 Users' Manual; Maximum Entropy Data Consultants Ltd., 1999.
(27) Shannon, C. E. *Bell Syst. Tech. J.* **1948**, *27*, 379−423.
(28) Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620−630.
(29) Shore, J. E.; Johnson, R. W. *IEEE Trans. Inf. Theory* **1980**, *26*, 26−37.
(30) Frieden, B. R. *J. Opt. Soc. Am.* **1972**, *62*, 511.
(31) Gull, S. F.; Daniell, G. J. *Nature* **1978**, *272*, 686−690.
(32) Skilling, J. Quantified Maximum Entropy. In *Maximum Entropy and Bayesian Methods*; Fougere, P. F., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1990; pp 341−350.
(33) Skilling, J. Classic maximum entropy. In *Maximum Entropy and Bayesian Methods*; Skilling, J., Ed.; Kluwer Academic Publishers: Dordrecht, Netherlands, 1989; pp 45−52.
(34) Brochon, J. C. *Methods Enzymol.* **1994**, *240*, 262−311.
(35) Kalinin, S.; Felekyan, S.; Antonik, M.; Seidel, C. A. M. *J. Phys. Chem. B* **2007**, *111*, 10253−10262.
(36) Kask, P.; Palo, K.; Ullmann, D.; Gall, K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13756−13761.
(37) Gopich, I. V.; Szabo, A. *J. Phys. Chem. B* **2003**, *107*, 5058−5063.
(38) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing*, 3rd ed.; Cambridge University Press: New York, 2007.
(39) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer Science+Business Media, LLC: New York, 2006.
(40) Strickler, S. J.; Berg, R. A. *J. Chem. Phys.* **1962**, *37*, 814−822.
(41) Birks, J. B.; Dyson, D. J. *Proc. R. Soc. London A-Math. Phys. Sci.* **1963**, *275*, 135−148.
(42) Magde, D.; Wong, R.; Seybold, P. G. *Photochem. Photobiol.* **2002**, *75*, 327−334.
(43) Torella, J. P.; Holden, S. J.; Santoso, Y.; Hohlbein, J.; Kapanidis, A. N. *Biophys. J.* **2011**, *100*, 1568−1577.
(44) Kalinin, S.; Sisamakis, E.; Magennis, S. W.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. B* **2010**, *114*, 6197−6206.
(45) Sidebert, S.; Kalinin, S.; Hien, N.; Kienzler, A.; Clima, L.; Bannwarth, W.; Appel, B.; Muller, S.; Seidel, C. A. M. *J. Am. Chem. Soc.* **2011**, *133*, 2463−2480.