FRET Efficiency Distributions of Multistate Single Molecules

Irina V. Gopich* and Attila Szabo

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States

Received: June 10, 2010; Revised Manuscript Received: September 10, 2010

A simple analytic theory is developed to describe FRET efficiency histograms constructed from a photon trajectory generated by a molecule with multiple conformational states. The histograms are approximated by a sum of Gaussians with the parameters explicitly determined by the FRET efficiencies of the states and the rates of the transitions between the states. The theory, which has been tested against exact histograms for two conformational states and simulated data for three and four conformational states, accurately describes how the peaks in the histograms collapse as the bin time or the transition rates increase.

I. Introduction

In the past decade, single-molecule fluorescence experiments emerged as a powerful tool to study conformational dynamics of biological molecules.[1−15] Förster resonance energy transfer (FRET) between donor and acceptor fluorophores can be exploited to probe fluctuations in the interdye distances between ∼1 and 10 nm. The output of single-molecule FRET experiments is a sequence of photons of different colors emitted by the donor and acceptor fluorophores. Changes in intramolecular distances modulate the observed pattern of photon colors.

The principal tool for analyzing such photon trajectories is a FRET efficiency histogram. This is constructed by binning the photons and calculating a FRET efficiency in each bin as the fraction of acceptor photon counts. The shape of a histogram potentially provides information about both the structure and the dynamics of the molecule. However, with few notable exceptions,[15−19] FRET efficiency histograms have been primarily used as a tool to resolve separate species. More extensive application of these histograms has been impeded by the lack of a simple analytic theory. An exact expression for the FRET efficiency histograms does exist for a two-state molecule.[20−22] However, even in this case, it is extremely complicated (involving integrals over Bessel functions), and it is unlikely that it can be generalized to the molecules with more than two states.

In this paper, we derive an accurate approximate analytic expression for the FRET efficiency histogram of a molecule that has multiple conformational states. It is assumed that the distribution of the donor and acceptor photons emitted by each state is Poissonian, and the mean total number of photons is the same for all states. The approximation is a sum of Gaussians, each with a weight, mean, and variance that are not adjustable but rather are explicit functions of the model parameters (the rate constants and the apparent FRET efficiencies) and the bin time. The basic idea of our approximation is to describe the contribution of the bins that contain no transitions and those that involve transitions between just two states by Gaussians with the appropriate mean and variance. The contribution of all other bins is described by a single Gaussian with parameters chosen to ensure that the entire FRET efficiency histogram is normalized and has the exact mean and variance for all bin times.

* Corresponding author. E-mail: irinag@niddk.nih.gov.

The above strategy is implemented in section II to derive our multi-Gaussian approximation to FRET efficiency histograms. In section III, this approximation is tested for two-, three-, and four-state systems. In the last section, we discuss how the Gaussian approximation can be used to analyze free diffusion experiments in which the distribution of the total number of photons is non-Poissonian.

II. Theory

Consider a sequence of photons divided into bins of duration T. Each bin contains a different number of N_A acceptor and N_D donor photons. FRET efficiency in a bin is defined as $E = N_A/(N_A + N_D)$. This is a random quantity that fluctuates due to shot noise and conformational dynamics on a time scale slower than or comparable to the interphoton time. We are interested in how the shape of the FRET efficiency histograms depends on the rate of conformational changes and on the bin time. When the bin time is short so that the interdye distance does not change during the bin time, the histogram is a superposition of several peaks, each broadened by shot noise. At long bin times, the interdye distance is averaged out, and the peaks collapse into a single peak. This collapse is a signature of conformational dynamics.

In the absence of conformational dynamics, the interdye distance is fixed, and the donor and acceptor photons are uncorrelated and obey Poisson statistics. Poisson statistics is characterized by the acceptor and donor count rates, n_A and n_D, which are the mean numbers of acceptor and donor photons detected per unit time. The ratio of the acceptor count rate to the total count rate is the apparent mean FRET efficiency, denoted by $\varepsilon = n_A/(n_A + n_D)$. The count rates and, therefore, the apparent FRET efficiency are related to the interdye distance. The precise relationship depends on the dye's photophysics, detection efficiencies, dynamics on the submicrosecond time scale, etc.

When the interdye distance fluctuates, so do the photon count rates. We describe these fluctuations by a kinetic scheme. The molecule emitting photons has M conformational states. Each state i is characterized by the apparent FRET efficiency $\varepsilon_i = n_{Ai}/(n_{Ai} + n_{Di})$. We assume that the total count rate $n = n_{Ai} + n_{Di}$ does not depend on conformation. The transitions between the states are described by the rate matrix $\mathbf{K}$. Its element K_{ij} is the rate of the $j \rightarrow i$ transition. The diagonal elements are negative, $K_{ii} = -\sum_{m\neq i} K_{mi}$. The equilibrium populations of states

10.1021/jp105359z © 2010 American Chemical Society
Published on Web 10/28/2010

are found by solving $\mathbf{Kp} = 0$, where $\mathbf{p}$ is the vector of equilibrium populations with elements p_i. These populations are normalized, $\Sigma_{i=1}^{M} p_i = 1$. The detailed balance condition, $K_{ij}p_j = K_{ji}p_i$, imposes a constraint on the transition rates.

The FRET efficiency distribution is simply related to the joint probability, $P(N_A, N_D)$, of detecting N_A acceptor and N_D donor photons in bin time T. When the interdye distance does not change, this distribution is a product of Poisson distributions of acceptor and donor photons. When the interdye distance fluctuates, the photon count rates are random functions of time. In this case, the Poisson distribution is generalized to

$$P(N_A, N_D) = \left\langle \frac{(\bar{n}_A T)^{N_A}}{N_A!} \frac{(\bar{n}_D T)^{N_D}}{N_D!} e^{-(\bar{n}_A+\bar{n}_D)T} \right\rangle_c \quad (1)$$

where $\bar{n}_{A(D)} \equiv \int_0^T n_{A(D)}(t)\, dt/T$, which are still random since they depend on the conformational states visited during time T. $\langle \cdots \rangle_c$ denotes an average over all conformational trajectories starting from the equilibrium distribution.

Our approximation requires the first two moments of the FRET efficiency, and we start by finding them. The mean of the FRET efficiency is found by multiplying the above joint distribution (with a proper normalization factor) by $N_A/(N_A + N_D)$ and evaluating the sums over N_A and N_D subject to $N_A + N_D \geq N_T$, where N_T is the threshold imposed on the total number of detected acceptor and donor photons.[27] When the total count rate does not fluctuate, $n_A(t) + n_D(t) = n$, one can show that the mean FRET efficiency is

$$\langle E \rangle = \langle \bar{\varepsilon} \rangle_c = \langle \varepsilon \rangle_{eq} \quad (2)$$

where $\bar{\varepsilon} \equiv \int_0^T \varepsilon(t)\, dt/T$ with $\varepsilon(t) \equiv n_A(t)/n$ and $\langle \varepsilon \rangle_{eq} = \Sigma_{i=1}^{M} \varepsilon_i p_i$ is the equilibrium average of the FRET efficiency. The first equality follows directly from the definition after evaluating the sums over the numbers of photons. The second equality is obtained by evaluating the average over the conformational trajectories.

Analogously, the second moment can be expressed as

$$\langle E^2 \rangle = \langle \bar{\varepsilon} \rangle_c \langle N^{-1} \rangle + \langle \bar{\varepsilon}^2 \rangle_c (1 - \langle N^{-1} \rangle) \quad (3)$$

where $\langle N^{-1} \rangle$ is the reciprocal of the total number of photons in a bin averaged over the distribution of the total number of photons, $\langle N^{-1} \rangle = [\Sigma_{N=N_T}^{\infty} (nT)^N/NN!]/\Sigma_{N=N_T}^{\infty}(nT)^N/N!$. Note that $\langle N^{-1} \rangle$ depends on both laser intensity (through the total count rate, n) and the bin time, T.

The average $\langle \bar{\varepsilon}^2 \rangle_c$ can be related to the autocorrelation function of the apparent efficiency by[20]

$$\begin{aligned} \langle \bar{\varepsilon}^2 \rangle_c &= \frac{2}{T^2} \int_0^T (T - t) \langle \varepsilon(t)\, \varepsilon(0) \rangle_{eq}\, dt \\ &= \frac{2}{T^2} \int_0^T (T - t) \sum_{i,j} \varepsilon_i [e^{\mathbf{K}t}]_{ij} \varepsilon_j p_j\, dt \end{aligned} \quad (4)$$

where $[\exp(\mathbf{K}t)]_{ij}$ is the ij element of the matrix exponential of $\mathbf{K}$. Simple analytic expressions for $\langle \bar{\varepsilon}^2 \rangle_c$ can be found for two- and three-state systems (see below). In general, it can be directly obtained numerically using a program such as Mathematica.[23]

With the two moments in hand, we can now derive the approximation for FRET efficiency histograms. Let us divide the bins that contribute to the FRET efficiency histogram into those that (1) have no transitions during the bin time, (2) have transitions only between two states, and (3) have transitions between three or more states. The contributions to the FRET efficiency histogram made by bins of the first two types are approximated by Gaussians labeled by index ij, $i, j = 1, ..., M$, $i \leq j$. The diagonal terms ii correspond to the bins with no transitions during the bin time. The off-diagonal terms ij ($i < j$) correspond to the bins with transitions only between states i and j. Finally, the Gaussian labeled by the index 0 corresponds to the remaining bins. Thus, the FRET efficiency histogram (FEH) is approximated as

$$\begin{aligned} \mathrm{FEH}(E) = {} & c_0 (2\pi\sigma_0^2)^{-1/2} \exp\left(-\frac{(E - \varepsilon_0)^2}{2\sigma_0^2} \right) + \\ & \sum_{i \leq j}^{M} c_{ij} (2\pi\sigma_{ij}^2)^{-1/2} \exp\left(-\frac{(E - \varepsilon_{ij})^2}{2\sigma_{ij}^2} \right) \end{aligned} \quad (5)$$

Let us first consider the parameters of the diagonal terms in the sum. The iith Gaussian describes the contribution of those bins in which the molecule is in state i during the entire bin time. Its weight, c_{ii}, is the fraction of such bins, which is the probability to be in state i during the entire bin time, $p_i \exp(K_{ii}T)$, where the matrix element $K_{ii} = -\Sigma_{m \neq i} K_{mi}$ is the negative of the sum of all rates out of state i. The mean FRET efficiency calculated from the subset of such bins is equal to the apparent efficiency of the state, ε_i, and the second moment is $\varepsilon_i \langle N^{-1} \rangle + \varepsilon_i^2 (1 - \langle N^{-1} \rangle)$ (see eq 3). Equating these with the moments of the iith Gaussian (ε_{ii} and $\varepsilon_{ii}^2 + \sigma_{ii}^2$), we have

$$c_{ii} = p_i e^{K_{ii}T} \quad (6a)$$

$$\varepsilon_{ii} = \varepsilon_i \quad (6b)$$

$$\sigma_{ii}^2 = \varepsilon_i (1 - \varepsilon_i) \langle N^{-1} \rangle \quad (6c)$$

for $i = 1, ..., M$. Note that σ_{ii}^2 is the variance due solely to shot noise.[22]

Now consider the off-diagonal terms. The ijth Gaussian corresponds to the bins that have transitions only between states i and j. Its weight, c_{ij}, is the fraction of such bins, which is equal to the probability to be in states i and j during the bin time (at least one transition between i and j must occur, but none to any other state). To find these weights, let us first obtain the probability, a_{ij}, to be in state i at time T starting from equilibrium when only states i or j were visited during the bin time. The subscript ij specifies the two states visited during the bin time, with the first index being the state at time T. These probabilities satisfy the rate equations

$$\frac{d}{dt} \begin{pmatrix} a_{ij} \\ a_{ji} \end{pmatrix} = \begin{pmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{pmatrix} \begin{pmatrix} a_{ij} \\ a_{ji} \end{pmatrix} \quad (7)$$

subject to equilibrium initial conditions $a_{ij}(0) = p_i$, $a_{ji}(0) = p_j$. The diagonal elements K_{ii} (K_{jj}) are the negative of the sum of all rates out of state i (j). They describe both the reversible relaxation of states i and j due to mutual interconversion as

well as the irreversible decay due to transitions to other states. Solving eq 10 and using the detailed balance condition $K_{ij}p_j = K_{ji}p_i$, we find that

$$a_{ij} = p_i e^{(K_{ii}+K_{jj})T/2}\left(\cosh(\lambda T) + \lambda^{-1}[(K_{ii} - K_{jj})/2 + K_{ji}]\sinh(\lambda T)\right) \quad (8)$$

where $\lambda = ((K_{ii} - K_{jj})^2/4 + K_{ij}K_{ji})^{1/2}$. The sum of a_{ij} and a_{ji} equals the fraction of bins in which the molecule visits only states i or j, and thus, $c_{ii} + c_{jj} + c_{ij} = a_{ij} + a_{ji}$. This condition determines the weights c_{ij}.

The other two parameters of the off-diagonal Gaussians, ε_{ij} and σ_{ij}^2, are found from the requirement that the first two moments of the FRET efficiency of the sum of the iith, jjth, and ijth Gaussians are equal to those obtained from the bins during which the molecule visits only states i or j. Using the same procedure that we used previously when all bins were considered (e.g., the way we obtained eqs 2 and 3), we now find

$$c_{ii}\varepsilon_{ii} + c_{jj}\varepsilon_{jj} + c_{ij}\varepsilon_{ij} = (a_{ij} + a_{ji})\langle\bar{\varepsilon}\rangle_{i\vee j} \quad (9a)$$

$$c_{ii}(\varepsilon_{ii}^2 + \sigma_{ii}^2) + c_{jj}(\varepsilon_{jj}^2 + \sigma_{jj}^2) + c_{ij}(\varepsilon_{ij}^2 + \sigma_{ij}^2) = (a_{ij} + a_{ji})(\langle\bar{\varepsilon}\rangle_{i\vee j}\langle N^{-1}\rangle + \langle\bar{\varepsilon}^2\rangle_{i\vee j}(1 - \langle N^{-1}\rangle)) \quad (9b)$$

where $\langle\bar{\varepsilon}^m\rangle_{i\vee j}$, $m = 1, 2$, is the conditional average of the mth power of $\bar{\varepsilon} = \int_0^T \varepsilon(t)\,dt/T$ over the subset of bins during which the molecule visits only states i or j. Although these averages can be found exactly, they can be well approximated by simpler expressions. The first moment, $\langle\bar{\varepsilon}\rangle_{i\vee j}$, is obtained from the mean $\langle\bar{\varepsilon}\rangle_c$ for an isolated two-state system consisting of states i and j, $\varepsilon_i p_i + \varepsilon_j p_j$, by replacing p_i and p_j with q_{ij} and q_{ji}, where $q_{ij} = a_{ij}/(a_{ij} + a_{ji})$ and $q_{ji} = 1 - q_{ij}$. Thus,

$$\langle\bar{\varepsilon}\rangle_{i\vee j} \approx \varepsilon_i q_{ij} + \varepsilon_j q_{ji} \quad (10)$$

It can be shown from eq 4 that the second moment, $\langle\bar{\varepsilon}^2\rangle_c$, for the isolated two-state system is $(\varepsilon_i p_i + \varepsilon_j p_j)^2 + p_i p_j(\varepsilon_i - \varepsilon_j)^2(1 - \phi((K_{ij} + K_{ji})T))$, where

$$\phi(\tau) = 1 - 2(\tau + \exp(-\tau) - 1)/\tau^2 \quad (11)$$

Replacing p_i and p_j with q_{ij} and q_{ji}, we have

$$\langle\bar{\varepsilon}^2\rangle_{i\vee j} \approx \langle\bar{\varepsilon}\rangle_{i\vee j}^2 + q_{ij}q_{ji}(\varepsilon_i - \varepsilon_j)^2(1 - \phi((K_{ij} + K_{ji})T)) \quad (12)$$

These approximations are especially good when the interconversion rates between i and j are much larger than other rates involving these states, which is when the contribution of the ijth Gaussian is the most significant.

Substituting the above expressions for the conditional averages into eqs 9 and solving these sequentially, we find that the parameters of the off-diagonal Gaussians ($i < j$) are given by

$$c_{ij} = a_{ij} + a_{ji} - c_{ii} - c_{jj} \quad (13a)$$

$$\varepsilon_{ij} = \varepsilon_i f_{ij} + \varepsilon_j f_{ji} \quad (13b)$$

$$\sigma_{ij}^2 = \varepsilon_{ij}(1 - \varepsilon_{ij})\langle N^{-1}\rangle + \sigma_{cij}^2(1 - \langle N^{-1}\rangle) \quad (13c)$$

$$\sigma_{cij}^2 = (\varepsilon_i - \varepsilon_j)^2\left(f_{ij}f_{ji} - \frac{a_{ij}a_{ji}\phi((K_{ij} + K_{ji})T)}{c_{ij}(a_{ij} + a_{ji})}\right) \quad (13d)$$

where $\phi(\tau)$ is given in eq 11 and $f_{ij} = (a_{ij} - c_{ii})/c_{ij} = 1 - f_{ji}$. In deriving eqs 13, we used the identity $(a_{ij} + a_{ji})(\varepsilon_i q_{ij} + \varepsilon_j q_{ji})^2 - c_{ii}\varepsilon_i^2 - c_{jj}\varepsilon_j^2 - c_{ij}\varepsilon_{ij}^2 = (\varepsilon_i - \varepsilon_j)^2(c_{ij}f_{ij}f_{ji} - (a_{ij} + a_{ji})q_{ij}q_{ji})$.

Finally, the parameters for the zeroth Gaussian, c_0, ε_0, and σ_0^2, are found from the requirements that the entire distribution is normalized and the first two moments are exact (i.e., equal to those in eqs 2 and 3). In this way, we find that

$$c_0 + \sum_{i\le j} c_{ij} = 1 \quad (14a)$$

$$c_0\varepsilon_0 + \sum_{i\le j} c_{ij}\varepsilon_{ij} = \langle\varepsilon\rangle_{eq} \quad (14b)$$

$$c_0(\varepsilon_0^2 + \sigma_0^2) + \sum_{i\le j} c_{ij}(\varepsilon_{ij}^2 + \sigma_{ij}^2) = \langle\varepsilon\rangle_{eq}\langle N^{-1}\rangle + \langle\bar{\varepsilon}^2\rangle_c(1 - \langle N^{-1}\rangle) \quad (14c)$$

where $\langle\varepsilon\rangle_{eq} = \sum_{i=1}^M \varepsilon_i p_i$ and $\langle\bar{\varepsilon}^2\rangle_c$ is given in eq 4. These equations are solved sequentially to find c_0, ε_0, and σ_0^2.

The sum of Gaussians in eq 5 with the parameters in eqs 6, 13, and 14 describes the FRET efficiency histogram of a molecule with M conformational states for arbitrary bin times. At bin times short compared with the time between transitions, only the diagonal terms contribute. At long bin times, all Gaussians vanish except the zeroth Gaussian. The mean and variance of FRET efficiency are exact at all times.

For two conformational states, there are only three Gaussians: 11, 22, and 12 ($c_0 = 0$ for all T in this case). The diagonal terms ($i = 1, 2$) are (see eq 6)

$$c_{11} = p_1 e^{-K_{21}T} \qquad c_{22} = p_2 e^{-K_{12}T} \quad (15a)$$

$$\varepsilon_{ii} = \varepsilon_i \quad (15b)$$

$$\sigma_{ii}^2 = \varepsilon_i(1 - \varepsilon_i)\langle N^{-1}\rangle \quad (15c)$$

where K_{21} (K_{12}) is the rate of $1 \rightarrow 2$ ($2 \rightarrow 1$) transition, $p_1 = K_{12}/(K_{12} + K_{12}) = 1 - p_2$. The coefficients that determine the 12 Gaussian are (see eq 13)

$$c_{12} = 1 - c_{11} - c_{22} \tag{16a}$$

$$\varepsilon_{12} = \varepsilon_1 f_{12} + \varepsilon_2 f_{21} \tag{16b}$$

$$\sigma_{12}{}^2 = \varepsilon_{12}(1 - \varepsilon_{12})\langle N^{-1}\rangle + \sigma_{c12}{}^2(1 - \langle N^{-1}\rangle) \tag{16c}$$

$$\sigma_{c12}{}^2 = (\varepsilon_1 - \varepsilon_2)^2\left(f_{12}f_{21} - \frac{p_1 p_2}{c_{12}}\phi((K_{12} + K_{21})T)\right) \tag{16d}$$

where $\phi(\tau)$ is given in eq 16 and $f_{12} = (p_1 - c_{11})/c_{12} = 1 - f_{21}$. This special case[24] has been recently applied to extract transition rates from experimental data on a two-state protein.[19]

FRET efficiency histograms for three conformational states with transitions among all states are described by seven Gaussians. The parameters of the Gaussians can be found from eqs 6, 13, and 14 using the formulas given in the Appendix. Our theory has the important property that the Gaussian approximation of the FRET efficiency histogram for a three-state system reduces analytically to the appropriate two-state approximation in all limiting cases when the three-state kinetic scheme reduces to a two-state one (e.g., transitions between any two states become sufficiently fast).

III. Illustrative Calculations

To test the approximation, we compare the distributions calculated using the Gaussian approximation with those obtained from the exact histograms for a two-state model[21,22] and from simulated photon trajectories for three- and four-state models. Details of the simulations are described in ref 25.

Figure 1 shows results for two interconverting conformational states, $1 \rightleftharpoons 2$. The approximation is the sum of three Gaussians with weights, means, and variances defined in eqs 15 and 16. The left and right columns show the performance at equal (the rates are $K_{12} = K_{21} = k$) and unequal ($2K_{12} = K_{21} = k$) populations of the states. For the purposes of illustration, we vary the rate of the transitions between the states, k, keeping the total count rate and the bin size constant.

The shape of the histograms is sensitive to the product of the bin time and conformational relaxation rate (which is $2k$ and $3k/2$ for the left and right columns, respectively). When the bin time is short compared with the conformational relaxation time ($kT \ll 1$), the histogram is a sum of two peaks centered on the efficiencies of the states (the first row). The peaks are broadened only by shot noise. The shot noise width depends on the total number of photons detected during the bin time through $\langle N^{-1}\rangle$. As the relaxation rate increases, these peaks do not move, but their amplitude drops, and a new peak (located at approximately the mean FRET efficiency) appears. At large kT (the bottom row), only this peak survives, which is broadened by shot noise and conformational dynamics. The approximation is very accurate at small and large kT. The deviation from the exact histograms is most noticeable when the amplitudes of the three Gaussians are comparable.

If k is fixed at $k = 1$ ms^{-1} and the bin size T is varied, the resulting histograms look similar (of course, identical for $T = 1$ ms) to those shown in Figure 1, with k being replaced by T. However, the peaks are slightly wider for $T < 1$ ms and narrower for $T > 1$ ms because of the different average number of photons in a bin.

Figure 2 compares the Gaussian approximation and simulated histograms for a linear three-state model, $1 \rightleftharpoons 2 \rightleftharpoons 3$, with equal equilibrium populations. The approximation consists of six Gaussians (there are no transitions between states 1 and 3 so that $c_{13} = 0$). The parameters of the Gaussians are explicit functions of the efficiencies of the three states (ε_1, ε_2, and ε_3) and the rate k (see eqs 6, 13, and 14, and Appendix). In the first column, the transitions rates are the same. It can be seen that the collapse of the three peaks into one as the transition rate increases is well described. In the second column, the transitions between states occur on different time scales, that is, the transitions between states 1 and 2 are much faster than those between 2 and 3. The histogram has three well resolved peaks when the bin time is much shorter than the relaxation time of the fast transitions between states 1 and 2 (the first row). As the transition rate increases, the first and second peaks collapse (the diagonal 11th and 22th Gaussians disappear and the 12th Gaussian appears). The remaining two peaks eventually collapse to a single peak when the bin time becomes longer than the relaxation time of states 2 and 3.

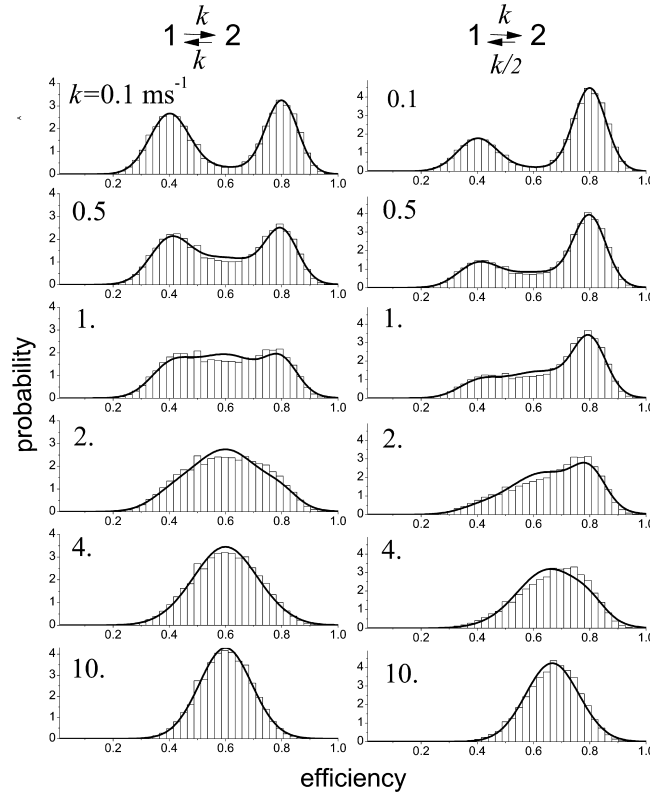

Figure 1. The Gaussian approximation (full curves) tested against exact FRET efficiency histograms (bars) for two conformational states with FRET efficiencies $\varepsilon_1 = 0.4$ and $\varepsilon_2 = 0.8$ and different values of the transition rate given in the upper left corner of each histogram. Left column: the equilibrium probabilities of the two states are the same. Right column: the probabilities differ by a factor of 2. The total count rate is $n = 50$ ms^{-1}, the bin time is $T = 1$ ms, threshold value is $N_T = 30$, histogram step is $h = 1/41$. The Gaussian approximation, eq 5, consists of three Gaussians with parameters in eqs 15 and 16.

Figure 3 shows the results for a linear four-state model, $1 \rightleftharpoons 2 \rightleftharpoons 3 \rightleftharpoons 4$, when the transitions occur on three different time scales. The approximation, which consists of eight Gaussians, describes the shape of the histograms remarkably well, except in the right column when $k = 25$ ms^{-1}. In this case, which was specifically chosen because it is the most demanding for our approximation, most bins involve transitions among states 1, 2, and 3, or no transitions at all (i.e., the system stays in state 4 during the entire bin time). There are very few bins in which all the four states are visited. Our theory does not properly describe contributions from bins when more that two, but not all, states are visited during the bin time. Note that this problem does not arise for the kinetic

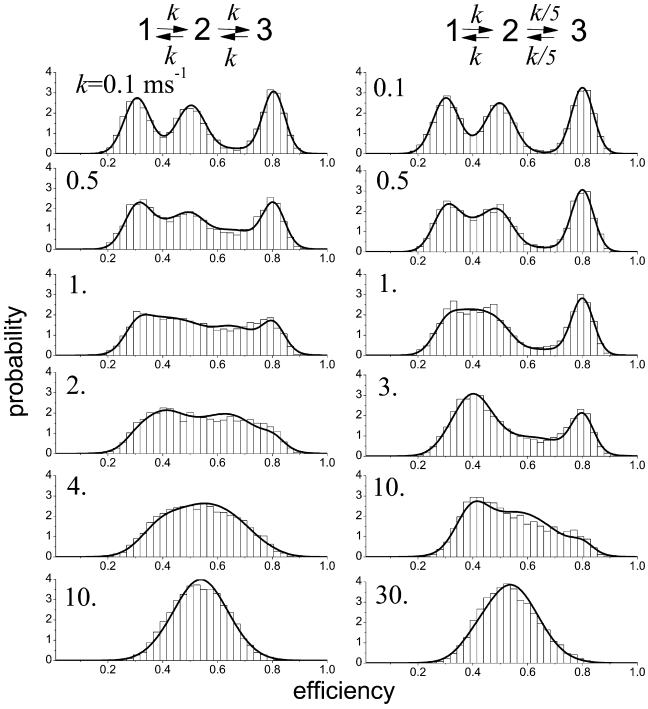

Figure 2. The Gaussian approximation (full curves) tested against simulated FRET efficiency histograms (bars) for a three-state model with FRET efficiencies $\varepsilon_1 = 0.3$, $\varepsilon_2 = 0.5$, $\varepsilon_3 = 0.8$ and different values of the transition rate given in the upper left corner of each histogram. Left column: equal transition rates. Right column: the transitions between states 1 and 2 are faster than those between 2 and 3. The simulated histograms were obtained from 5000 bins of duration $T = 1$ ms, the total count rate is $n = 100$ ms^{-1}, and the other parameters are the same as in Figure 1.

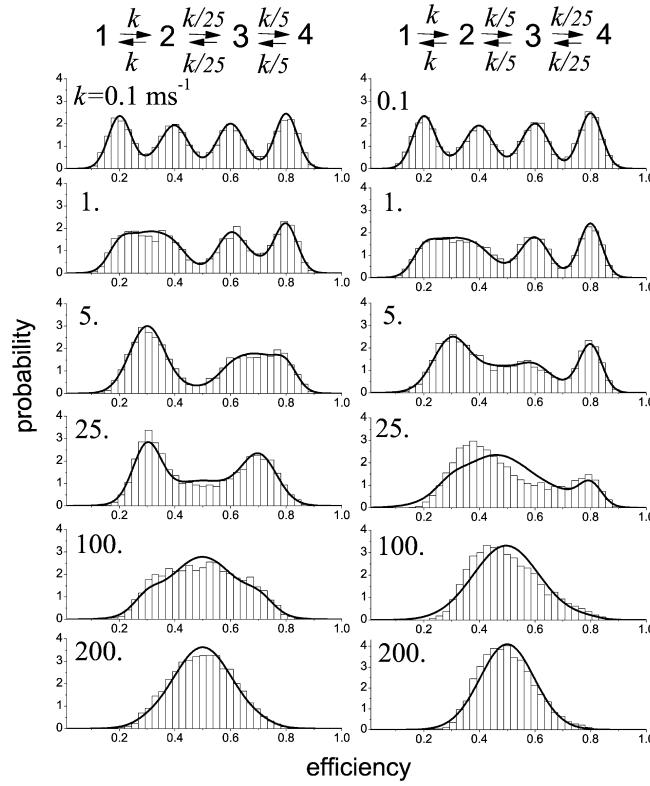

Figure 3. Same as Figure 2, for four conformational states with FRET efficiencies $\varepsilon_1 = 0.2$, $\varepsilon_2 = 0.4$, $\varepsilon_3 = 0.6$, $\varepsilon_4 = 0.8$ and transitions between the states on three different time scales. Left column: the fastest transitions are those between states 1 and 2, and the slowest are between states 2 and 3. Right column: the slowest transitions are between states 3 and 4.

scheme in the left column. Here, when $k = 25$ ms^{-1}, most bins have transitions between a pair of states (i.e., 1 and 2 or 3 and 4) and this is correctly handled by the theory. Thus, for a large number of states, even though it is always correct at short and long times, our approximation can deteriorate when transitions occur on many different time scales.

IV. Concluding Remarks

We have derived a simple expression for FRET efficiency histograms obtained in single-molecule measurements. For M interconverting conformational states, the FRET efficiency histogram is approximated by a sum of Gaussians (three for a two-state, seven for a general three-state system, etc.). The weights, means and variances of the Gaussians are not adjustable parameters, but are given by eqs 6, 13, and 14, in terms of the FRET efficiencies of the states and the transition rates. The variances of the Gaussians also involve the average of the reciprocal of the sum of donor and acceptor photons, $\langle N^{-1} \rangle$. This depends on the laser intensity (through the total count rate), bin size, and threshold value and can be obtained directly from the experimental data.

The Gaussian approximation can be readily used to analyze experimental histograms and extract the rates of conformational changes and the FRET efficiencies of the conformers. The efficiencies can be then related to the interdye distances after correcting for crosstalk, background noise, and fast polymer dynamics.[24,26] Such complications do not influence the rates. The shape of FRET efficiency histograms is sensitive to the product of the bin time T and the conformational relaxation rates (i.e., the eigenvalues of the rate matrix), especially when one of these products is close to unity. To show the effect of conformational dynamics on the shape of the histogram, one can change the bin size (by simply rebinning the photon trajectory). Alternatively, one can change the conformational relaxation rate (e.g., by adding denaturant when studying protein folding). Chung et al.[19] have recently applied the three-Gaussian approximation with parameters given by eqs 15 and 16 to two-state protein folding and compared the extracted rates and apparent FRET efficiencies with those obtained by a likelihood method that does not require binning.[25]

Our derivation assumes that the total count rate is the same for all conformational states and that the distribution of the sum of acceptor and donor photons is Poissonian. When only the energy transfer rate depends on the conformation, this is equivalent to the requirement that the γ-factor (which is the product of the ratios of the acceptor and donor quantum yields and detection efficiencies) is equal to one.[28] For arbitrary total count rates, the extension of the Gaussian approximation is a challenging problem that we are trying to solve.

In free diffusion experiments, even when the total count rate does not depend on conformation, it fluctuates because laser intensity changes as the molecule diffuses through the confocal volume. Nevertheless, our results can be applied to diffusing molecules when the bins or bursts are preselected to optimize the probability that the molecule is quasi-immobilized; that is, during the observation time, it explored only a limited region of the observation volume where the intensity did not change significantly.[22] The parameters of the Gaussians in the case of free diffusion are still given by eqs 6, 13, and 14, but now $\langle N^{-1} \rangle$ is the reciprocal of the sum of acceptor and donor photons averaged over the experimentally observed non-Poissonian distribution. This average can be obtained directly from the bins or bursts selected to construct the FRET efficiency histograms. When a photon

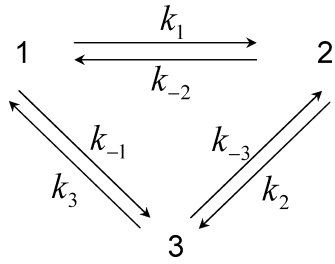

Figure 4. Kinetic scheme for three interconverting conformational states.

trajectory is divided into time bins of the same duration, the approximation should be used only in conjunction with high threshold values which eliminate bins during which the molecule leaves and then reenters the laser spot. In other words, the bins selected for analysis should not contain significant gaps where only a few photons are detected. When bursts of photons of variable duration are selected using a burst search algorithm, the bursts should be trimmed into segments of the same duration T in such a way that intensity is fairly uniform. The best strategy for selecting bins or bursts of photons during which the molecule is quasi-immobilized still needs to be established. In this regard, the rates obtained by fitting the histograms constructed from trajectory fragments using different burst/bin selection criteria could be compared with those obtained using an alternative procedure that we have developed[25] that does not require the quasi-immobilization assumption.

Acknowledgment. This work was supported by the Intramural Research Program of the National Institutes of Health, NIDDK.

Appendix: Three-State System.

In this Appendix, we present the formulas for the quantities that are required to calculate the 21 parameters of the 7 Gaussians ($ij = 11,\ 22,\ 33,\ 12,\ 23, 13,\ 0$) for a three-state system using eqs 6, 13, and 14. The rate matrix **K** for the three-state kinetic scheme shown in Figure 4 is

$$\begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{pmatrix} = \begin{pmatrix} -(k_1 + k_{-1}) & k_{-2} & k_3 \\ k_1 & -(k_2 + k_{-2}) & k_{-3} \\ k_{-1} & k_2 & -(k_3 + k_{-3}) \end{pmatrix} \tag{A1}$$

Detailed balance imposes the constraint that $k_1k_2k_3 = k_{-1}k_{-2}k_{-3}$.

The absolute values of the nonzero eigenvalues of this matrix are

$$\kappa_\pm = B/2 \pm \sqrt{B^2/4 - C} \tag{A2}$$

where

$$\begin{aligned} B &= k_1 + k_2 + k_3 + k_{-1} + k_{-2} + k_{-3} \\ C &= k_1k_2 + k_2k_3 + k_3k_1 + k_{-1}(k_2 + k_{-2}) + \\ &\quad k_{-2}(k_3 + k_{-3}) + k_{-3}(k_1 + k_{-1}) \end{aligned} \tag{A3}$$

The normalized equilibrium populations are

$$\begin{aligned} p_1 &= [k_2k_3 + k_{-2}(k_3 + k_{-3})]/C \\ p_2 &= [k_3k_1 + k_{-3}(k_1 + k_{-1})]/C \\ p_3 &= 1 - p_1 - p_2 \end{aligned} \tag{A4}$$

By evaluating the matrix exponential in eq 4, it can be shown that $\langle\bar{\varepsilon}^2\rangle_c$ is given by

$$\langle\bar{\varepsilon}^2\rangle_c = \langle\varepsilon\rangle_{eq}{}^2X + \langle\varepsilon^2\rangle_{eq}(1 - X) + Y[(\varepsilon_1 - \varepsilon_2)^2k_1p_1 + (\varepsilon_2 - \varepsilon_3)^2k_2p_2 + (\varepsilon_3 - \varepsilon_1)^2k_3p_3] \tag{A5}$$

where $\langle\varepsilon^m\rangle_{eq} = \sum_{i=1}^{3} \varepsilon_i^m p_i$, $X = (\kappa_+\phi_- - \kappa_-\phi_+)/(\kappa_+ - \kappa_-)$, $Y = (\phi_- - \phi_+)/(\kappa_+ - \kappa_-)$, and $\phi_\pm = \phi(\kappa_\pm T)$, where $\phi(\tau)$ is given in eq 11. This is required to calculate σ_0^2 using eq 14.

References and Notes

(1) Jia, Y.; Talaga, D. S.; Lau, W. L.; Lu, H. S. M.; DeGrado, W. F.; Hochstrasser, R. M. *Chem. Phys.* **1999**, *247* (1), 69–83.

(2) Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (10), 5179–5184.

(3) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419* (6908), 743–747.

(4) Rothwell, P. J.; Berger, S.; Kensch, O.; Felekyan, S.; Antonik, M.; Wöhrl, B. M.; Restle, T.; Goody, R. S.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (4), 1655–1660.

(5) Rhoades, E.; Cohen, M.; Schuler, B.; Haran, G. *J. Am. Chem. Soc.* **2004**, *126* (45), 14686–14687.

(6) Coban, O.; Lamb, D.; Zaychikov, E.; Heumann, H.; Nienhaus, G. *Biophys. J.* **2006**, *90* (12), 4605–4617.

(7) Moerner, W. E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (31), 12596–12602.

(8) Mukhopadhyay, S.; Krishnan, R.; Lemke, E. A.; Lindquist, S.; Deniz, A. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (8), 2649–2654.

(9) Kinoshita, M.; Kamagata, K.; Maeda, A.; Goto, Y.; Komatsuzaki, T.; Takahashi, S. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (25), 10453.

(10) Wozniak, A.; Schröder, G.; Grubmüller, H.; Seidel, C.; Oesterhelt, F. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (47), 18337.

(11) Roy, R.; Hohng, S.; Ha, T. *Nat. Methods* **2008**, *5* (6), 507–516.

(12) Chung, H. S.; Louis, J. M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (29), 11837–11844.

(13) Santoso, Y.; Joyce, C.; Potapova, O.; Le Reste, L.; Hohlbein, J.; Torella, J.; Grindley, N.; Kapanidis, A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (2), 715.

(14) Hofmann, H.; Hillger, F.; Pfeil, S. H.; Hoffmann, A.; Streich, D.; Haenni, D.; Nettels, D.; Lipman, E. A.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (26), 11793–11798.

(15) Hanson, J.; Duderstadt, K.; Watkins, L.; Bhattacharyya, S.; Brokaw, J.; Chu, J.; Yang, H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (46), 18055.

(16) Pljevaljcic, G.; Millar, D. P.; Deniz, A. A. *Biophys. J.* **2004**, *87* (1), 457–467.

(17) Kalinin, S.; Valeri, A.; Antonik, M.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. B* **2010**, *114* (23), 7983–7995.

(18) Santoso, Y.; Torella, J. P.; Kapanidis, A. N. *ChemPhysChem* **2010**, *11*, 2209–2219.

(19) Chung, H. S.; Gopich, I. V.; McHale, K.; Cellmer, T.; Louis, J. M.; Eaton, W. A. *J. Phys. Chem. A* **2010**, DOI: 10.1021/jp1009669.

(20) Gopich, I. V.; Szabo, A. *J. Phys. Chem. B* **2003**, *107* (21), 5058–5063.

(21) Gopich, I.; Szabo, A. *J. Chem. Phys.* **2005**, *122* (1), 14707−1–18.

(22) Gopich, I. V.; Szabo, A. *J. Phys. Chem. B* **2007**, *111* (44), 12925–12932.

(23) Wolfram Research, Inc. *Mathematica ed.: Version 7.0*; Wolfram Research, Inc., Champaign, IL, 2008.

(24) Gopich, I. V. ; Szabo, A. *Adv. Chem. Phys.*, in press.

(25) Gopich, I. V.; Szabo, A. *J. Phys. Chem. B* **2009**, *113* (31), 10965–10973.

(26) Makarov, D. E.; Plaxco, K. W. *J. Chem. Phys.* **2009**, *131*, 085105-1-6.

(27) A threshold is usually applied in free diffusion experiments to minimize the number of bins where the molecule is not inside the spot during the entire bin time. However, even in the case of immobilized molecules, a threshold $N_T = 1$ is obviously required to avoid division by zero in the FRET efficiency.

(28) It is possible that the γ-factor is not equal to 1, but the total count rate is the same for all states when some parameter other than the energy transfer rate (e.g., the quantum yield[19]) also depends on conformation.

JP105359Z