



Published in final edited form as:

J Phys Chem B. 2006 November 9; 110(44): 22103–22124. doi:10.1021/jp063483n.

Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments†

Eyal Nir[‡], Xavier Michalet[‡], Kambiz M. Hamadani[‡], Ted A. Laurence[§], Daniel Neuhauser[‡], Yevgeniy Kovchegov^{||}, and Shimon Weiss^{*,‡,⊥,#}

Department of Chemistry and Biochemistry, Department of Physiology, and California NanoScience Institute, University of California at Los Angeles, Los Angeles, California 90095, Physical Biosciences Institute, Lawrence Livermore National Laboratory, Livermore, California 94550, and Department of Mathematics, Oregon State University, Corvallis, Oregon 97331

Abstract

We describe a simple approach and present a straightforward numerical algorithm to compute the best fit shot-noise limited proximity ratio histogram (PRH) in single-molecule fluorescence resonant energy transfer diffusion experiments. The key ingredient is the use of the experimental burst size distribution, as obtained after burst search through the photon data streams. We show how the use of an alternated laser excitation scheme and a correspondingly optimized burst search algorithm eliminates several potential artifacts affecting the calculation of the best fit shot-noise limited PRH. This algorithm is tested extensively on simulations and simple experimental systems. We find that dsDNA data exhibit a wider PRH than expected from shot noise only and hypothetically account for it by assuming a small Gaussian distribution of distances with an average standard deviation of 1.6 Å. Finally, we briefly mention the results of a future publication and illustrate them with a simple two-state model system (DNA hairpin), for which the kinetic transition rates between the open and closed conformations are extracted.

1. Introduction

Forty years after Stryer and Haugland proposed to use fluorescence resonance energy transfer (FRET) as a molecular ruler,¹ single-molecule fluorescence resonance energy transfer (smFRET) has become a widely used tool for monitoring relative changes in distances but also absolute distances between a donor and acceptor dye attached to either freely diffusing, vesicle-encapsulated or surface-bound biomolecules.^{2,3} These approaches have been used to study (i) the conformation and conformational dynamics of nucleic acids, ribozymes,⁴ and proteins,⁵ (ii) the interaction of proteins with nucleic acids or other proteins,^{6,7} (iii) the dimerization of membrane receptors in live cells,⁸ (iv) the folding and unfolding of proteins,^{5,9} and many other questions described in recent reviews.^{2,3,5,10–13}

†Part of the special issue “Charles M. Knobler Festschrift”.

©2006 American Chemical Society

[‡]Department of Chemistry and Biochemistry, UCLA.

[§]Lawrence Livermore National Laboratory.

^{||}Oregon State University.

[⊥]Department of Physiology, UCLA.

[#]California NanoScience Institute, UCLA.

Supporting Information Available: Additional experimental data and figures. This material is available free of charge via the internet at <http://pubs.acs.org>.

Although FRET in principle provides a way to obtain distances between dyes, and therefore to obtain structural information on molecular species, in practice extracting accurate distances from smFRET measurements can be delicate due to the difficulty of measuring critical experimental parameters such as the dyes' rotational freedom of motion (and hence the orientational factor κ^2), detection efficiencies, dye quantum yields, donor fluorescence leakage in the acceptor channel, or direct excitation of the acceptor, to cite only a few.^{14,15} Luckily, even without a complete knowledge of all these parameters, smFRET measurements have demonstrated their utility to detect structural changes and quantify conformational subpopulations, as well as, in certain conditions, giving access to the time scales of various fluctuations taking place in biomolecules. A large amount of recent theoretical and experimental work has been put into trying to obtain more quantitative information from these single-molecule experiments, and this article will describe new tools for this purpose.

Before delineating the domain of application of our work, it is worth remembering that there are two very different types of smFRET experiments (or in general, single-molecule spectroscopy experiments): (i) those which provide information on a few individual molecules over long periods of time (they involve immobilized molecules, slowly diffusing molecules, or tracked molecules); and (ii) those which provide only instantaneous snapshots of single-molecule properties, but do so over a very large number of molecules (these experiments usually take advantage of the Brownian diffusion of molecules in a solution or in a membrane, in and out of a small excitation volume). In both cases, it is always possible to extract typical time scales of the molecular fluctuations using (fluorescence) correlation methods, but this aspect will not be treated here.

For molecules attached to a surface or enclosed in slowly moving vesicles, donor and acceptor fluorescence fluctuations occurring in an individual molecule can in principle be monitored continuously. Once the optimal, shot-noise limited time binning of the photon stream has been determined,¹⁶ standard single-molecule time-trace analysis can be performed. For instance, the distribution of values reached by an observable can be characterized by histogram analysis, possible molecular states identified (corresponding to specific values of one or more observables), and the conversion rates between these states quantified.⁹ Time traces, therefore, provide unique information on the dynamics of the system, which could hardly be obtained from ensemble of molecules.

A different situation is encountered in diffusion-based experiments. Here, single molecules diffuse through a femtoliter (fL) excitation volume during a very short time (on the order of a millisecond in solution), producing small bursts of donor (D) and acceptor (A) emitted photons of a few dozen to a few thousand photons each. Such a small number of photons is enough to spectrally identify the dye molecules present in the burst, to extract their stoichiometry,^{6,17,18} and to compute other simple quantities on the basis of the total number of photons collected in each detection channel (such as FRET efficiency¹⁹ or polarization anisotropy²⁰). Other ensemble quantities, such as diffusion coefficient, concentration, and time scales of fluctuations, can also be extracted from such experiments using FCS types of analyses.²¹ If pulsed-laser excitation is used rather than continuous-wave laser excitation, time-correlated single-photon counting (TCSPC) measurements can also provide lifetime information, which, in addition to helping identify the molecular species in a single burst, can also be used to measure FRET efficiency.^{22,23} In general, however, the number of photons in a single burst is usually insufficient to perform any time-trace analysis, except in a few rare cases of long bursts.²⁴ Instead, in smFRET diffusion experiments, quantities are computed at the burst level using all photons in each burst and then histogrammed.

Diffusion-based measurements have several advantages over immobilized-molecule measurements: there is no need to (i) develop immobilization chemistry (neither surface treatment nor molecule functionalization), (ii) worry about surface-interaction effects, and (iii) worry about photophysical artifacts due to the proximity of the surface (higher background, nonradiative transfer, etc.); (iv) finally, diffusion-based measurements allow the accumulation of data coming from several thousands of different molecules in a matter of a few minutes, a feat that cannot be attained with surface immobilization, where at most a few hundreds of molecules can be monitored at a time.

In summary, in both approaches (immobilized or diffusing molecules), the common products of smFRET data acquisition are histograms of observed values computed from a generally small number of photons and can be therefore severely affected by shot noise.

To analyze these histograms, early smFRET experiments used simple fitting functions, such as Gaussian or β distributions, providing one (or more) mean value(s) of the measured observable.^{19,25–27} The question remains regarding what information to extract from the second parameter of these fitting functions, namely, their standard deviation. It was clear that the width of these distributions was strongly dependent on the average signal collected in each burst (small bursts yielding noisier measurements than large ones due to shot noise). On the other hand, the broad and diverse distributions of burst sizes characteristic of actual experiments made it problematic to define a simple relation between average burst size and histogram width. Early attempts to understand this relation were therefore limited to estimating a lower bound of the histogram width due to shot noise.^{26,28,29} Analyses of the shapes of the histograms were essentially limited to counting the number of peaks (indicative of the number of distinct populations of molecules) and the areas under these peaks (representing the fraction of the molecules belonging to these identified populations). These populations could, for instance, correspond to different biomolecules¹⁹ or to different conformations of the same single molecule²⁵ depending on the experimental situation.

Recently, attempts were made to convert FRET efficiency distributions into distance distributions using the well-known Förster dependence of the energy transfer on distance (step 1) and sometimes into free energy landscapes assuming a Gibbs–Boltzman distribution of distances (step 2) (for example, see refs 25 and 30). Even though these steps were usually made with cautionary wording, criticisms have appeared in the literature rightly questioning their legitimacy.³¹ To understand whether it is possible to extract more information than the mere average FRET value (and in particular whether there is any hope of extracting distance distribution or even energy landscapes from such smFRET experiments), it appears necessary to first account for the exact contribution of the shot noise to the shape of the experimental histogram, taking into account all available experimental information.

In this article, we show that it is possible not only to compute the exact histogram width due to shot noise (and not only an upper bound) but also to account for the global shape of experimental histograms, assuming a few simple hypotheses on the experimental system. The only required ingredient of our analysis is the observed burst size distribution, information that is readily available from the experiment although usually not exploited. We focus on a simple FRET observable, the proximity ratio PR, which is related to the FRET efficiency by a simple relation. In particular, we show that PR histograms obtained from simulated smFRET data can be fully accounted for by our approach. Deviations from this result are characterized for different nonideal situations. We also study the PR histograms of typical single-molecule FRET experiments and discuss possible sources of the discrepancies between observed and predicted histograms.

While this manuscript was being completed, a similar and independent analysis was published by Antonik et al.³² Although both works use the same basic premises, i.e., that the observed signal is principally due to the combination of shot noise and burst size distribution, our work provides a complementary result, numerical and experimental verifications and practical algorithms to compute the theoretical histograms. Although shot-noise issues that will be discussed in this article also occur for time-trace analyses, we will not deal with these issues in this article. A recent publication by the Yang group presents an interesting alternative analysis of the influence of shot-noise applied specifically to smFRET time-trace data of immobilized molecules.³³

This article is organized as follows. In section 2, we define the proximity ratio histogram (PRH) and introduce basic concepts and notations used in the remainder of the article. Section 3 discusses the possible sources of broadening of the PRH and presents a model of the shot noise contribution to the PRH and an algorithm to compute it. Section 4 describes our material and methods. In section 5, we test our model using numerical simulations of smFRET experiments, studying various possible artifacts that may result in deviation from the model. Having checked the validity of our approach on simulated data, we use it on several experimental model systems in section 6. We discuss our results and the possible extensions of this work to the study of dynamical systems in section 7 and conclude in section 8.

2. The smFRET Proximity Ratio Histogram (PRH)

smFRET Photon Streams

Figure 1 of refs 14 and 18 shows a schematic of a typical diffusion-based smFRET experimental setup using continuous-wave laser excitation. The raw data extracted from such a setup consists of one or more streams of photons collected by separate detectors (usually single-photon counting avalanche photodiodes, or SPADs). When a single laser excitation is used, one detector collects mainly photons emitted by the donor molecule of the Förster pair, while the other collects mainly photons emitted by the acceptor molecule. When alternated laser excitation (ALEX) is used, each detector collects two temporally intertwined photon streams, corresponding to donor excitation periods and acceptor excitation periods, respectively.^{6,14,18} In both cases, the signals used to compute the proximity ratio are the signals collected during donor excitation. Adopting the notations of ref 14, we will denote by F_X^E the signal (number of photons) collected by channel E (donor D or acceptor A) during excitation with the laser line specific of dye X (donor D or acceptor A). As discussed in the reference cited above, this signal contains contributions from the fluorescent molecule of interest, but also background signal, and in some cases, cross-talk from the other dye molecule. Limiting ourselves to donor excitation

$$F_D^D = {}^D F_D^D + {}^A F_D^D + {}^{D \rightarrow A} F_D^D + B_D^D$$

$$= {}^D F_D^D + 0 + 0 + B_D^D \quad (1)$$

$$F_D^A = {}^D F_D^A + {}^A F_D^A + {}^{D \rightarrow A} F_D^A + B_D^A$$

$$= Lk + Dir + F^{FRET} + B_D^A \quad (2)$$

Where the left upper index indicates the origin of the signal: D denotes the donor dye, A the acceptor dye, and D→A denotes acceptor signal originating from FRET. B_D^D is the background signal in the donor channel. There is usually no cross-talk of the acceptor dye in

the donor detection channel; therefore, the second and third terms of the right-hand side (RHS) of eq 1 are equal to zero.

The first term of the RHS of eq 2 is the signal coming from the donor dye detected in the acceptor channel, or leakage (Lk). Its importance can be estimated from the theoretical emission spectrum of the donor and the wavelength-dependent detection efficiency of the acceptor channel or, experimentally, from the observation of donor-only labeled molecules. The second term corresponds to the direct excitation of the acceptor dye by the donor-excitation laser. Its importance can also be estimated theoretically or measured experimentally on an acceptor-only labeled sample, as explained in ref 14. The third term is due to FRET, and the final term corresponds to background signal in the acceptor channel.

Definition of Bursts

In the absence of any molecules, only background signal is detected in each channel. This occurs a significant part of the time in diffusion-based single-molecule experiments. Every now and then, however, when a molecule transits through the diffraction-limited excitation volume, a finite number of excitation/emission processes take place, resulting in additional photons being detected during a brief period of time (typically on the order of one millisecond or less). These bursts of photons are naturally those we are interested in, and the question arises regarding how to define their beginning and end. Several algorithms have been proposed, based either on time-trace binning and thresholding²⁶ or interphoton time averaging and thresholding,²³ and, more recently, maximum likelihood statistical analysis.³⁴ Here, we propose a novel burst detection scheme taking advantage of ALEX, that allows us to separate donor-only and acceptor-only bursts from doubly labeled bursts, as well as to interrupt a burst when one of the dyes bleaches.

Let us first consider a simple burst definition inspired from that proposed by Seidel and co-workers.²³ In this definition, photons belong to a burst if at least L successive photons have at least M neighboring photons within a time window of length T centered on their own arrival time. Typical parameters used in this work are $L = 50$, $M = 30$, and $T = 500 \mu\text{s}$, but others could be adapted to different experimental situations (see Supporting Information S1 for further illustration on this choice of parameters). As we are using an ALEX scheme, the burst search is performed on all collected photons (from both the donor and acceptor channels and from both laser excitation periods). We call this simple definition the all-photon-burst-search (APBS) method.

As mentioned, this method was originally developed for single-molecule diffusion experiments using a single laser excitation, using both the donor and acceptor photons to determine the beginning and the end of a burst. With ALEX, additional information on the molecule diffusing through the excitation spot can be obtained. In particular, it is possible to determine whether and when the burst consists of photons coming from a single donor dye or from a single acceptor dye. Figure 1 illustrates how this information can be used to better define the beginning and end of a burst in cases where the APBS method would result in skewed FRET estimation.

The first case (Figure 1B) is encountered when one of the dyes bleaches during the molecule's transit through the confocal spot. If the donor bleaches (case not shown), the remaining acceptor-only signal (visible during the acceptor excitation periods) will skew the FRET efficiency toward a larger value. Note that, if one uses a standard single-excitation scheme, this case is eliminated altogether, since donor bleaching will simply truncate the burst. If the acceptor bleaches, the remaining donor-only signal will reduce the calculated FRET efficiency value compared to its true value. Both cases will contribute lower or larger values of the FRET efficiency, widening the dispersion of this observable. A different but

related phenomenon, blinking, is illustrated in Figure 1C for the acceptor. Here, the APBS method would define the burst as the period between the appearance of both donor and acceptor signals and their simultaneous disappearance. The time interval during which the acceptor blinks will therefore result in a calculated FRET efficiency value lower than the true value (which should be calculated only for the time windows when both dyes are fluorescent). Finally, in the case when the diffusion of two molecules through the excitation volume overlaps in time (Figure 1D), it is important to be able to separate the contributions of two molecules that could exhibit very different FRET efficiencies. The APBS method obviously does not fulfill this requirement. To handle the issues raised by these different cases, we propose a new burst search method based on the added information provided by ALEX, which we dub the dual-channel-burst-search (DCBS) method. A first step consists of a burst search including the donor and acceptor photons detected during the donor excitation periods. The second step consists of a separate burst search including only the acceptor photons detected during the acceptor excitation periods. When these two separate burst searches are combined, it is possible to define time intervals during which both the donor and the acceptor are active. These intervals define the burst durations. Typical values used for each individual burst search used in the DCBS methods are $L = 25$, $M = 15$, and $T = 500$ μs (see Appendix A for a justification of these values).

The remainder of this article will use the DCBS method, unless stated otherwise.

Definition of the Proximity Ratio

Ideally, one would like to measure the FRET efficiency value E for each detected molecule (burst) and study its distribution within the sampled population. To do this, one would have to use the standard definition of E

$$E = \frac{F_D^A}{F_D^A + \gamma F_D^D} \quad (3)$$

where the factor γ is the ratio of the donor and acceptor products of the detection efficiency (η) and quantum yield (Φ)

$$\gamma = \frac{\eta_D \Phi_D}{\eta_A \Phi_A} \quad (4)$$

Although we will not directly discuss distance measurements in this article, it is worth recalling the relation between E and the distance R between donor and acceptor molecules

$$E = \left[1 + \left(\frac{R}{R_0} \right)^6 \right]^{-1} \quad (5)$$

where R_0 is the so-called Förster radius, characteristic of the dye pair.

Extracting the quantities intervening in eq 3 from the raw detector data F_D^D and F_D^A can be difficult and lead to undesired uncertainties. Therefore, it is a common practice to compute

not the FRET efficiency value E (eq 3) but instead the so-called proximity ratio PR defined as

$$PR = \frac{F_D^A}{F_D^A + F_D^D} \quad (6)$$

The proximity ratio PR reduces to the FRET efficiency E when there is no leakage, background, and direct excitation of the acceptor, and $\gamma = 1$. Obviously, this is rarely the case, so that PR values cannot be used, in general, to extract real distances. However, qualitatively, PR follows the trend of E upon changes in donor–acceptor distance. In other words, a decrease of PR value can be interpreted as a decrease in E , or equivalently an increase in distance between donor and acceptor, and reciprocally. Moreover, in many experimental cases, even though there is leakage, background, and direct excitation of the acceptor, their contribution is minimal on the calculated histogram, as will be shown in a later section. Finally, it is always possible to artificially achieve a γ ratio of 1. For instance, one can slightly misalign one detector as described in ref 14. However, this can result in a large offset between the donor and acceptor detection volumes, which has a detrimental effect on the PRH, as will be discussed in a later section. Another way of adjusting the detection efficiency of one channel is by discarding a constant fraction of the detected photons by this channel. These reasons justify considering the PRH as we will now do. An added benefit of the PRH is that, being calculated from raw counting data (eq 6), it involves only integer numbers, which greatly facilitates the theoretical analysis.

To simplify notations, we will replace the quantities F_D^D and F_D^A by D and A , the number of photons detected by the donor and acceptor channels, respectively, and will call their sum $S = A + D$, such that

$$PR = \frac{A}{A+D} = \frac{A}{S} \quad (7)$$

Note that, in some figures, the symbol S will be used for the ALEX stoichiometry ratio introduced in refs 6, 14, 13, and 35 without possibility of confusion. For the sake of completeness, we recall here the definition of this ratio, used in alternating laser configurations

$$S = (F_D^D + F^{\text{FRET}}) / (F_D^D + F^{\text{FRET}} + F_A^A) \quad (8)$$

In this work, F^{FRET} will be replaced in eq 8 by F_D^A (eq 2) for simplicity.

The Proximity Ratio Histogram

As defined in eq 7, the set of PR values obtained from the multiple bursts detected in a single experiment are rational numbers between 0 and 1. Although rational numbers are dense in the interval $[0, 1]$, the distribution of numbers obtained as a ratio of two integer numbers is not uniform in this interval, as illustrated by the comb distribution of Figure 2.

Antonik et al. made a similar remark concerning the ratio $F_D^D/F_D^A = 1/(1 + PR)$ they studied in ref 32, although the range of this ratio is not limited to $[0, 1]$, contrary to the PR.

Citing ref 32, “the discrete nature of the distribution leads to features that introduce binning artifacts in the generation of histograms: probability ‘spikes’ at certain values and probability ‘voids’ around the spikes”. Although histograms were generated in logarithmic scale in ref 32, the same question of how to determine the optimal bin size to generate histograms (in our case, PRH’s) occurs in these two approaches.

One possible choice consists of using a fixed bin size for all experimental situations, which is small enough (say, 100 bins over the [0, 1] interval) to reveal potential fine features in the PR distribution (e.g., close multiple peaks). The problem with this is that the PRH will exhibit spikes and voids, even for a smooth underlying FRET efficiency distribution.

An alternative consists of using a statistically optimal bin size, such as that recently proposed by Knuth.³⁶ This approach adapts the bin size to the underlying distribution, without prior knowledge of its actual shape. As an example of the outcome of this approach, the optimal bin number calculated for the combed distribution shown in Figure 2B, is 1. This number is perfectly adapted to the theoretically flat distribution of rational numbers in the interval [0, 1] as shown in Figure S1. Unfortunately, the Knuth algorithm breaks down for most “spikes” and “voids” types of histograms. We therefore modified the Knuth algorithm to handle this particularity of PRH, as described in Appendix A. Briefly, we resort to dithering, a standard pretreatment of discrete data sets, by adding a small random number to each PR value, before applying our modified Knuth bin size optimization. Once this is done, this optimal bin number obtained for a “smoothed” data set is used for the raw data, resulting in histograms mostly devoid of spikes and voids.

In the following, we will present either histograms with 100 bins or with the optimal bin number defined previously. The 100 bin histograms will serve to illustrate the striking ability of our analysis (described in the next section) to account for the distribution of measured PR, whereas the optimal histogram will represent data without spikes and voids.

3. Shot-Noise Broadening of the PRH

Sources of Broadening of the PRH

Experimental smFRET histogram are usually well-fitted by Gaussian functions^{19,25} and/or log-normal functions³¹ or sum thereof providing a simple way to obtain average FRET or average PR values. It has been debated for some time as to what to extract from the second moment (standard deviation or width) of these fitted functions. Quoting from ref 25, “one must exercise great caution in interpreting the width of FRET distribution”, due to the interplay of many factors, including shot noise. Nevertheless, this and other works have gone beyond this cautionary statement and extracted distance distributions and free energy landscapes from the FRET histogram (see, for instance, refs 25, 37, and 38).

Shot-Noise Contribution to the PRH

We now formalize the concepts introduced previously. Our goal is to obtain an estimation of the shot-noise contribution to the width of the PRH. Our starting point will be the burst size distribution and the hypothesis that this information is sufficient to predict the shape of the PRH histogram, knowing the average weights of acceptor and donor channels in the sample.

The experimental burst size distribution BSD is a histogram of measured burst sizes ($S = A + D$), a direct output of the experiment. Note that, although we will in practice use a “filtered” burst size distribution, relying on the ability of the ALEX method to reject donor-only and acceptor-only bursts, the following discussion is independent of the exact distribution that is used. For instance, the analysis can be performed on the PRH generated from all bursts coming out of the burst search algorithm or post-ALEX filtering or after thresholding, and

so forth. In each case, our analysis will generate a shot-noise limited histogram that best fit the experimental PRH, enabling their comparison and the evaluation of additional contributions to the PRH.

The only hypothesis in the analysis is the nature of the distribution of counts in the acceptor and donor channels in each burst and an associated free parameter. We will now argue that this distribution is binomial and therefore fully characterized by the probability ε (to be determined) that a photon will be detected by, say, the acceptor channel.

That the photons emitted by the acceptor dye are distributed according to a binomial distribution follows from the definition of FRET: for each donor excitation followed by an emission event, the probability that the photon will be emitted by the acceptor is by definition the FRET efficiency E . In this respect, if we were in the ideal situation where there would be no leakage, no direct excitation of the acceptor, no background, and 100% detection efficiency in both channels, counting the acceptor photons would constitute what the statistical literature calls a Bernoulli trial, with probability of success E . Considering all bursts (D_i, A_i) of size S , we would find that the probability of obtaining A counts in the acceptor channel, given the total number S of detected photons, is given by the binomial distribution

$$P_E(A|S) = \binom{S}{A} E^A (1-E)^{S-A} = \frac{S!}{A!(S-A)!} E^A (1-E)^{S-A} \quad (9)$$

In a nonideal case, the detection efficiencies for each channel will be different from 1, and there will be leakage and direct excitation issues. Nevertheless, the probability of detecting an acceptor photon rather than a donor photon should still be characterized by a constant. The contribution of background in both channels has a different effect, since it depends only on the burst duration, not on the excited molecule trajectory. In the following discussion, we will assume that no background is present and will postpone the discussion of the influence of background (and other effects) on our results to a later section.

Under these conditions, the same property of Bernoulli trials used to compute the probability of detecting A acceptor photons among S emitted photons in the ideal case can be used in the nonideal case to compute the probability of counting A acceptor photons when a total of S photons are detected. Namely, the probability of this to occur is given by the binomial law, with the replacement of E in eq 9 by an unknown parameter ε

$$P_\varepsilon(A|S) = \binom{S}{A} \varepsilon^A (1-\varepsilon)^{S-A} \quad (10)$$

The value of the parameter ε will be used as an adjustable parameter in the following, but could in principle be calculated from the knowledge of all relevant experimental parameters (leakage, direct excitation, γ factor) as discussed by Antonik et al.³² and in Appendix C.

With the hypotheses recalled before, eq 10 gives the probability of observing A acceptor photons in a burst of S photons. This is therefore also the probability of observing a PR value equal to $x = A/S$ for bursts of S photons

$$p_{\varepsilon}(x=A/S|S)=P_{\varepsilon}(A|S)\delta(x-A/S) \quad (11)$$

Limiting ourselves to rational values x , this can be rewritten

$$p_{\varepsilon}(\text{PR}=x|S)=P_{\varepsilon}(xS|S)\chi_{\text{N}}(xS) \quad (12)$$

where we define $\chi_{\text{N}}(t)$ as the characteristic function of natural numbers

$$\chi_{\text{N}}(t)=\begin{cases} 1 & \text{if } t \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

The total probability to observe the PR value x requires summing eq 12 over all burst sizes S , weighted by their probability, given by the burst size distribution

$$p_{\varepsilon}(x)=\begin{cases} \frac{1}{B} \sum_{S=S_{\min}}^{S_{\max}} \text{BSD}(S) \times P_{\varepsilon}(xS|S)\chi_{\text{N}}(xS) & \text{if } x \in \mathbb{Q} \cap [0, 1] \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

where B is the total number of bursts in the BSD, and S_{\min} (respectively, S_{\max}) the smallest (respectively, largest) burst size taken into consideration. Introducing the characteristic function of rational numbers in $[0, 1]$

$$\chi_{\mathbb{Q} \cap [0, 1]}(x)=\begin{cases} 1 & \text{if } x \in \mathbb{Q} \cap [0, 1] \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

one obtains

$$p_{\varepsilon}(x)=\chi_{\mathbb{Q} \cap [0, 1]}(x) \times \frac{1}{B} \sum_{S=S_{\min}}^{S_{\max}} \text{BSD}(S) \times P_{\varepsilon}(xS|S)\chi_{\text{N}}(xS) \quad (16)$$

Using eq 10, we end up with our final prediction of the probability distribution of PR values (in the absence of background)

$$p_{\varepsilon}(x)=\chi_{\mathbb{Q} \cap [0, 1]}(x) \times \frac{1}{B} \sum_{S=S_{\min}}^{S_{\max}} \text{BSD}(S) \times \binom{S}{xS} \varepsilon^{xS} (1-\varepsilon)^{(1-x)S} \chi_{\text{N}}(xS) \quad (17)$$

To obtain the predicted PRH, a natural way could be to compute the number of predicted occurrences of each PR value x by multiplying eq 17 by the total number of bursts B and histogram these values with the same bin numbers as the PRH. In practice, however, this

means calculating eq 17 for each possible rational value x accessible from bursts in the BSD, which number is close to (we neglect that, as shown in Figure 2, identical values can be obtained for several different burst sizes)

$$N = \sum_{S=S_{\min}}^{S_{\max}} (S+1) = \frac{(S_{\max} + S_{\min} + 2)(S_{\max} - S_{\min} + 1)}{2} \approx \frac{S_{\max}^2}{2} \quad (18)$$

Since the calculations in eq 17 involve $S_{\max} - S_{\min} + 1$ terms, each one requiring a test of rationality of xS and evaluation of a CPU intensive expression, this approach may take quite some time for BSD with very large bursts. In addition, eq 17 contains the free parameter ε that remains to be evaluated, which means that the whole procedure would need to be repeated within a minimization algorithm. It turns out that a simpler alternative algorithm can be used, resulting in an excellent approximation of the predicted PRH.

Computation of the Best-Fit Shot-Noise Limited PRH

The derivation of eq 14 provides the basis of our algorithm. The probability to observe A acceptor photons in bursts of size S is given by the binomial law $P_{\varepsilon}(A|S)$. To obtain the number of occurrences of the PR value $x = A/S$, this probability needs to be multiplied by the number of bursts of size S , $\text{BSD}(S)$. We now claim that a good approximation of the quantity $\text{BSD}(S) \times P_{\varepsilon}(A|S)$ can be obtained as the number of occurrences of A acceptor photons when one draws $\text{BSD}(S)$ numbers from a binomial distribution characterized by (ε, S) . This follows directly from the definition of a binomial distribution, which is obtained asymptotically for large numbers of trials. The number of occurrences of the value A estimated in this manner may depart from the theoretical value, but the difference will be significant only for burst sizes which are rarely represented (because one does not perform enough trials to uniformly sample the distribution). But for this same reason, the discrepancy will have little influence on the histogram, as only a few bursts (i.e., PR values) will be concerned. To compensate for this lack of sampling of the binomial distribution for under-represented burst sizes, it is also possible to systematically oversample all binomial distributions by an arbitrary factor N (e.g., $N = 10$), i.e., replace each $\text{BSD}(S)$ value by $N \times \text{BSD}(S)$, perform as many trials and divide the final histogram by N . In practice, for reasons that are described in Appendix D, we simply go through the collected set of bursts and draw N numbers A distributed according to a binomial distribution of parameters (ε, S) for each burst (of size S). The final PRH is then divided by N . In summary, our algorithm to compute the shot-noise limited PRH for an experimental distribution of bursts is the following:

- i. Choose an oversampling factor N and a realistic initial value ε .
- ii. For each burst i of size S_i , draw a number A from the binomial distribution $P_{\varepsilon}(A|S_i)$ of size S_i and probability ε .
- iii. Add the value A/S_i to the PRH.
- iv. Repeat (ii) – (iii) N times.
- v. Repeat (ii) to (iv) for each collected burst.
- vi. Divide the final PRH by N .
- vii. Improve on ε , for instance, using a dichotomic search and least-squares minimization.

As random number generators distributed according to the binomial law are available in most scientific calculation software (LabView, *Matlab*, <http://www.gnu.org/software/gsl/>,

etc.) or can easily be programmed using published algorithms,³⁹ implementing the above calculation is straightforward. The minimization step needed to compute the best ε value is performed using the least-squares method and proceeds by dichotomy, starting from a reasonable estimate of the mean of the PRH. The resulting histogram is the best shot-noise limited fit of the experimental PRH.

In comparison to a direct calculation of the PRH using eq 17, which involves a number of computations scaling as a power law of the maximum burst size, the previous algorithm scales linearly with the number of bursts. To appreciate the importance of this difference, consider two data sets differing by a single burst with a size 10 times larger than the others. These two data sets will take an identical time to analyze with the proposed algorithm, whereas the direct calculation will require 1000 times longer for the data set containing the extra burst. Another advantage of this algorithm, which amounts to a Monte Carlo calculation of eq 17, is that it is easily extended to more complex situations, as will become clear in the remainder of this article.

Influence of Background and other Phenomena on the PRH

Background was neglected in the previous discussion because the number of background counts in a burst depends solely on the average background count rates and the burst duration. Therefore, the ratio of background counts in the acceptor channel to total background counts is a constant, but one which is unrelated to the constant ratio ε defined above. As a consequence, the presence of background counts skews the theoretical ratio of counts in the acceptor channel to total counts (A/S) in a burst-duration dependent manner and complicates the theoretical analysis of the PRH. Similarly, one should expect that phenomena such as bleaching, blinking, and multiple molecule bursts and other effects will affect the prediction of eq 17.

Although the case of background counts can easily be handled analytically using some simple approximations (as will be shown later), most other effects are strongly system- and setup-dependent and therefore best studied using numerical simulations. We therefore postpone their discussion to a later section.

4. Materials and Methods

DNA Sample Preparation

Single-stranded DNA (ssDNA) molecules were purchased (IDT Inc., Coralville, LA) or synthesized with automatic solid-phase synthesis following published protocols.⁴⁰ The common “top” strand sequence was 5'-TAAATCTAAAGTAACATAAGGTAACATAACGTAAGTCCA-3'. It was labeled with a donor molecule (Cy3B, Amersham Biosciences, Piscataway, NJ) at position 5 (from the 5' end) by introducing a C6 dT amino modifier. Five different samples of the complementary “bottom” strand were labeled with an acceptor molecule (ATTO 647-N, ATTO-TECH GmbH, Siegen, Germany) at positions 13, 18, 23, 28, and 35 (from the 3' end) by introducing a C6 dT amino modifier (see Figure 3). ssDNA molecules were purified on reverse-phase C18 HPLC column (Amersham Bioscience, Piscataway, NJ). dsDNA molecules were formed by hybridization of each of the five bottom strands separately with the top strand in 40 mM Tris-HCL pH 8, 500 mM NaCl. The top and bottom samples were heated to 90 °C and then cooled slowly to achieve maximum hybridization. The top- and bottom-strand concentrations were adjusted to ensure a minimum of nonhybridized top or bottom ssDNA molecules.

The DNA hairpin bottom strand sequence was 3'-AAGAAGTGTGGTCAGGTTTGATAGTGGTTGAAT-5' and was labeled with a donor

(Cy3B, Amersham) at position 10 (from the 3' end) by introducing an amino modifier C6 dT. The top strand 5'-TGGTT-(T)₂₁-AACCATTCTTCACAAACCAGTCCAAACTATCACAACTTA-3' was labeled with Alexa Fluor 647 (Molecular Probes-Invitrogen, Carlsbad, CA) at position 1 (from the 5' end). Hybridization and experiments were performed in 10 mM Tris-Base pH 8, 500 mM NaCl.

Single-Molecule FRET Measurements

smFRET μ s-ALEX experiments were carried out as previously described^{14,18} using a custom-made single-molecule fluorescence microscope based on a commercial inverted microscope (IX71, Olympus America, Melville, NY). Briefly, two CW laser lines (532 nm, GCL-100-L, CrystalLaser, Reno, NV, and 635 nm, MVP, Coherent, Santa Clara, CA) were alternated using an acousto-optic modulator (N48058-XX-.55, Neos Technologies, Melbourne, FL) with a 25 μ s period per laser line and coupled into a single-mode fiber. After expansion and collimation, the green and red beams were focused 30 μ m inside the sample solution (50 pM) using a water immersion objective (NA 1.2, 60 \times , Olympus America, Melville, NY). The emitted fluorescence was separated from the excitation light by a dichroic mirror (535–635 TBDR, Omega Optical, Brattleboro, VT) and further split by a second dichroic mirror (DRLP 630, Omega Optical). Donor emission was directed to a single-photon avalanche photodiode (SPAD) (SPCM-AQR-14, Perkin-Elmer Optoelectronics, Fremont, CA) after filtering by a band-pass filter (580AF60, Omega Optical), while acceptor emission was directed to another SPAD after filtering by a long-pass filter (660 ALP, Omega Optical). The two SPAD signals (photon arrival times) were recorded as a function of time using a 12.5 ns resolution counting board (PCI-6602, National Instruments, Austin, TX).

Single-laser excitation smFRET experiments were performed using the same setup and the single CW 532 nm laser without alternation. The emitted light was filtered with either the green filter or the red filter, as mentioned above, and then split by a nonpolarized beam splitter into two detection channels (SPAD's). One of the channels was treated as a dummy "donor" channel, and the other channel was treated as a dummy "acceptor" channel.

PSF Measurements

To estimate the offset between donor and acceptor channel detection volumes, we acquired vertical and horizontal confocal images scan of TetraSpeck beads (diameter 100 nm, Molecular Probes) embedded in a 3% low melting point agarose gel using a custom-made microscope and software.⁴¹ The fluorescence of these beads was excited using the 488 nm line of an argon ion laser, and scans were performed perpendicularly to the optical axis (*XY* scans) as well as parallel to it (*XZ* or *YZ* scans). Signals from the donor and acceptor channels were simultaneously recorded in the condition of alignment similar to those used in the smFRET experiments. Profiles of fluorescence intensity along a line passing through the image maximum (center of the convolved excitation/emission point-spread function) were then extracted for each detector image.

smFRET Experiment Simulations

smFRET experiments are complex to analyze due to the many different experimental parameters involved¹⁴ and the nontrivial nature of both the molecule trajectories (and therefore burst intensities)^{42,43} and any burst search algorithm.⁴³ This renders it impossible to develop a complete analytical description of the experiments, justifying the use of numerical simulations to study the effect of each parameters on the PRH. The main characteristics of our simulations are the following:

- i. *Photophysics.* Donor and acceptor molecules are treated as two four-state systems, including a ground state, an excited state, a triplet state, and a dark (bleached) state. Figure S2 (Supporting Information) depicts the corresponding Jablonski diagrams and transition rates used in the simulations. The excitation rate k_e (transition between the ground state and the excited state) is a function of the absorption cross section (or extinction coefficient) and the laser intensity at the molecule's three-dimensional (3D) position (see below). Energy transfer occurs only when the donor is in the excited state and the acceptor is in its ground state. In most cases, the energy transfer rate was kept constant, but complex energy landscapes can easily be modeled by time-varying energy transfer rates. Actual transitions occur stochastically during the simulation according to their respective rates. No polarization and therefore rotational diffusion effects were considered in these simulations.
- ii. *Diffusion.* Typically, $N=10$ molecules (donor-acceptor pairs) undergo a Brownian diffusion in a periodic boundary condition 3D box using fixed time steps. The size of the box and/or the number of molecules is adjusted to the desired concentration (typically 50 pM). The diffusion constant used in the simulation is $D = 5 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$. The common time step used for diffusion, photophysics, and, when applicable, dynamics is adjusted as a function of the molecule position in the excitation volume, varying from 100 ns at the center to 10 μs at the periphery of the box. This strategy considerably speeds up the computation, while having no effect on the realism of the simulation, as checked.
- iii. *Excitation and Detection Volumes.* The excitation intensity distribution or point spread function (PSF) is modeled with a 3D Gaussian centered in the middle of the diffusion box and with typical XY dimension $\sigma_{XY} = 0.3 \mu\text{m}$ and Z dimension (optical axis) $\sigma_Z = 1.5 \mu\text{m}$. For alternating laser excitation, the two excitation PSF's are periodically reset to zero (every 25 μs) and can also be spatially offset from one another and set to different dimensions to simulate chromatic aberration or alignment issues. The detection PSF's (probability of detecting emitted donor and acceptor photons) are modeled by similar 3D Gaussians.
- iv. *Recording.* Photons emitted during a transition of either the donor or acceptor molecules from their excited to ground state are detected with a probability proportional to the value of the detection PSF at the location of the molecule. If detected, their time stamp and detection channel (donor or acceptor) are stored.
- v. *Background.* A stochastic, constant rate background photon flux can be added to each previous photon list to mimic the experimentally detected background. To simulate the presence of donor-only or acceptor-only molecules, or of fluorescent or scattering impurities, single-dye photophysics (four-state system) was used as described in (i)–(iii).

Data files generated by the simulations were analyzed using the same tools used during actual experiments (burst search and PRH analysis as described in the text).

5. Test of the PRH Fit on Simulated Data

Simple smFRET Simulation

To check the validity of eq 17 for the PRH shape, we first analyzed simulated data, as a comparison with experimental data (presented later) requires the knowledge of several parameters that are difficult to measure. Figure 4 shows PRH's of simulated background-free photon streams corresponding to donor-acceptor pairs characterized by different FRET efficiencies (gray box histograms). Superimposed on the same graphs (black curves) and

almost indistinguishable from them we represented the best-fit shot-noise limited PRH's (eq 17). In all cases, the fitted parameter ε (which in this ideal case is identical to the FRET efficiency E) equals the value put in the simulation.

The excellent agreement between predicted and experimental PRH is preserved for all values of the FRET efficiency E (Figure 4A–C) and burst size range (Figure 4D,E). In particular, the widening of the PRH for experimental conditions that result in small burst sizes (Figure 4D) compared to experiments resulting in small as well as very large burst sizes (Figure 4E) is quantitatively accounted for by this calculation.

We next separately studied other potential contributors to the PRH width that were not taken into account in the derivation of eq 17: background counts in both channels, dye photophysics (bleaching, blinking), coincident detection of several molecules, and a possible mismatch between the donor and acceptor detection volumes.

Effect of Background

There are at least two ways to analyze the effect of background on the previous results of section 3. The first consists of adopting a mathematical formalism similar to that used by Antonik et al. in ref 32. We present such an approach in Appendix D, which results in a slightly more complex expression than eq 17 (eq 36), and which is best evaluated using the simulation approach presented above for the evaluation of eq 17. The second way of dealing with background consists of studying numerically its effect on the shape of the PRH. Figure 5A shows the effect on simulated data of a rather high level of background (6 kHz) in the donor channel on the PRH. Using eq 17, which assumes that there is no background, a best-fit PRH can be obtained, which satisfactorily matches the experimental PRH. Of course, the extracted ε value is distinct from the FRET value used for the simulations. If we now use eq 36 with our knowledge of the donor channel background level to fit the PRH, a best-fit value of ε is obtained, which is indistinguishable from the input FRET efficiency and a fitted PRH which now perfectly matches the simulated one, as illustrated in Figure 5B.

Effect of Bleaching

Bleaching of a dye (donor or acceptor) during the transit time results in bursts with characteristics that can easily be detected by the DCBS algorithm presented above. Figure 6A1 shows the 2D-ALEX histogram corresponding to bursts obtained with simulated data including acceptor bleaching ($k_{bl} = k_e^D/300$), detected using the simpler APBS method. A significant donor-only peak ($PR = 0, S = 1$) is apparent, as well as a trail of bursts connecting this peak and the peak of interest ($PR = 0.5, S = 0.5$). The PRH resulting from this burst search method, shown on top of the 2D histogram, exhibits a correspondingly long tail toward low PR values, which cannot be accounted for by shot-noise only. By using the DCBS algorithm, however, most of these irrelevant bursts are eliminated or relocated in the 2D-ALEX histogram, as illustrated in Figure 6A2. The remaining bursts result in a PRH (Figure 6A3) that is perfectly fitted by eq 17.

Effect of Blinking

In real experiments, intersystem crossing (ISC) of one or both dyes to their triplet state cannot be excluded. The effect of ISC is to put the affected dye in a nonemitting state for a finite period of time, typically microseconds to milliseconds, resulting in a characteristic blinking behavior. If the donor is affected, no signal will be measured, and the sole effect will be to reduce the burst size and therefore increase the PRH width in a predictable manner. However, if the acceptor is affected, no FRET will take place, but the donor will still emit fluorescence, thus shifting the apparent FRET efficiency (and hence PR) toward smaller values. If this shift in value is similar for all bursts, our analysis will be able to

account for the shape of the PRH, with the parameter ε simply taking the new shifted value. More generally, three different limit cases can be envisioned:

- i. Blinking occurs on a time scale comparable to or larger than the burst duration.
- ii. Blinking occurs on a time scale much shorter than the burst duration.
- iii. Blinking occurs at an intermediate time scale.

In the first case, either the donor or the acceptor will be off during the whole duration of the molecule transit through the excitation spot, therefore resulting in a donor-only or acceptor-only burst. In the second case, the effect of blinking will be averaged over the burst duration, and the situation described before (shifted PR value) applies. The last case will result in less obvious effects. Figure 6B1–3 shows the result of a simulation in which acceptor molecules transit to their triplet state every 300 excitations ($k_{TSC} = k_e^D/300$) and stay there on average 300 μ s. When using the APBS algorithm, a tail of bursts connecting the population of interest to the donor-only location appears in the ALEX histogram (Figure 6B1). The DCBS algorithm, on the other hand, successfully identifies the bursts exhibiting blinking and only leaves the relevant bursts in the final ALEX histogram (Figure 6B2). The remaining bursts result in a PRH (Figure 6B3) that is perfectly fitted by eq 17.

Effect of Coincident Detection of Multiple Molecules

Multiple molecule bursts are bursts corresponding to two or more molecules present at the same time in the excitation volume. They can have different effects on the “ideal” PRH, supposed to report only on single-molecule bursts (Figure 6C1):

- i. Bursts consisting of two or more doubly labeled molecules characterized by the same PR will be larger, thus resulting in a reduced influence of the shot noise.
- ii. Bursts consisting of doubly labeled (D–A) molecules with different PR values will result in intermediate PR measurements, thus deforming the PRH.
- iii. D–A bursts contaminated by donor-only molecules will result in lower PR values than expected and, in the case of ALEX measurement, in larger stoichiometry ratio S .
- iv. D–A bursts contaminated by acceptor-only molecules will result in slightly larger PR values (due to direct excitation of the acceptor) and, in the case of ALEX measurement, in smaller S .
- v. Burst containing only donor-only and acceptor-only molecules will result in random PR values.

The last effect cannot be corrected for and, similarly to the second one, will inevitably result in a departure of the PRH from its ideal shape (obtained if only doubly labeled molecules were detected). The other effects can be partially taken care of by the ALEX approach, which provides additional information for each burst, the stoichiometry ratio S (eq 8). As discussed in refs 14 and 18, the stoichiometry ratio S takes values close to 1 for donor-only bursts and values close to 0 for acceptor-only bursts, while doubly labeled molecules exhibit values close to 0.5 (the exact value depending on several experimental parameters). When using a two-dimensional representation (PR, S) for each burst, it is straightforward to isolate bursts corresponding to relevant molecules. The DCBS method almost entirely takes care of these cases, as illustrated in Figure 6C2,3.

In conclusion to this study of the effect of coincident events on the PRH, we shall reiterate an obvious way to minimize it. Since bursts corresponding to multiple molecules excited

simultaneously occur with a probability which is proportional to the concentration, their effect can be minimized by using smaller concentrations.

In summary, bleaching, blinking due to ISC or other mechanisms, or multiple molecular bursts do not significantly affect the PRH when a DCBS algorithm is used to search for bursts in the photon streams.

Detection Volume Mismatch

The final effect we will consider in this work is that of a possible mismatch between detection volumes. The detection volume for a given channel is centered on a point of maximum detection efficiency and can be defined as the domain within which the probability of detection is larger than some arbitrary value $p > 0$. The existence of a mismatch between donor and acceptor detection volume simply means that the ratio between the probability of detecting a donor photon and the probability of detecting an acceptor photon depends on the location of the emission. This can be due to chromatic aberrations of the optics or misalignment of the detectors or pinhole, for instance. The net result of such a situation is that the ratio of donor to acceptor photons (and hence the PR) will depend on where the molecule is located in the focal region of the objective lens, leading to an additional spread of the measured PR values.

Figure 7 illustrates the effect of a relative offset of the donor and acceptor detection volumes along the optical axis. For an offset amounting to 10% (50 nm) of the point-spread-function width, a modest widening of the PRH can be observed in addition to the effect of shot noise (Figure 7A,B). For a 20% offset (100 nm), the simulated effect is much more severe (Figure 7C,D). To have an idea of the severity of detection volume mismatch in actual experiments, we imaged the product of the excitation and emission point-spread functions (PSF) of a single-molecule confocal microscope as described in Material and Methods. Figure 7E shows the intensity map generated by a 100 nm diameter fluorescent bead along the YZ plane of the microscope. The cross sections of this PSF in both donor and acceptor channels are displayed on the right of Figure 7E. The maximum observed offset is smaller than 3% in all directions, suggesting that the detection volume offset has a negligible effect on the PRH.

6. Test of the PRH Fit on Experimental Data

Having studied the theoretical influence of several effects on the PRH shape, we have concluded that, in commonly encountered experimental conditions, the PRH shape of a single D–A distance (single FRET efficiency) sample should be close to shot-noise limited and obtained by either eq 36 or eq 17 if background is negligible.

We now present simple experimental data as a first step to validating the applicability of these calculations.

Artificial PR Data

A simple way to generate experimental data similar to those expected from a FRET sample consists of observing a single dye sample with two identical detection arms differentiated only by their detection efficiencies. The advantage of this approach is that the ratio of the “donor” and “acceptor” channel signals will be constant and equal to the detection efficiency ratio and not affected by any photophysics or dynamics of the molecule. Therefore, the only phenomena that can occur are those that have been studied in the previous section, and a perfect agreement between the observed PRH and the shot-noise limited one is expected.

Figure 8 shows the results obtained with a 50 pM sample of dsDNA labeled with a single Cy3B dye molecule. The emitted fluorescence is split equally between two equivalent

detectors using a nonpolarizing beam splitter cube, and the detection efficiency ratio between the two detection channels is adjusted using neutral density filters. The PRH's are perfectly fitted by the shot-noise limited PRH for all detection efficiency ratio (or artificial FRET efficiency), as expected. Note that, in addition to proving the validity of our approach, this result also confirms the assumptions on which it is based. In particular, it shows that the whole excitation (laser, EOM, optics) and acquisition chain (optics, filters, pinhole, detector, readout electronics) does not introduce any bias.

dsDNA Data

Double-strand DNA is expected to be quite rigid when the length of the molecule is smaller than its persistence length (~150 bp in standard buffer solutions).⁴⁴ dsDNA has therefore been used as a benchmark of smFRET measurements^{15,19,35,45} together with rigid polyproline peptides.^{33,46} Here, we used a 39 bp dsDNA labeled with Cy3B at a fixed position (position 5 from the 5' end) on one strand and at variable positions on the other strand with ATTO-647N (samples 1–5: positions 13, 18, 23, 28, and 35 from the 3' end) resulting in different distances between the dyes (8, 13, 18, 23, and 30 bp, respectively, or 2.7, 4.4, 6.1, 7.8, and 10.2 nm using a bp separation of 0.34 nm).

To eliminate contamination from unlabeled or ssDNA molecules, which are usually present in non-ALEX measurements, we used a DCBS algorithm preferentially to an APBS (Figure 9). The resulting PRH's and the associated best-fit shot-noise limited PRH's are shown in Figure 10A–E (left and center columns). It is apparent in this latter figure that, except for the sample with the smallest expected FRET efficiency (Figure 10E), the shot-noise limited fit does not succeed in matching the experimental PRH width, being systematically narrower than the observed histograms. As we have extensively discussed before, this discrepancy between the theory and the experiment cannot be attributed to photophysical effects such as bleaching or blinking or to experimental artifacts such as multiple coincident detection or detection volume mismatch of the magnitude typically present in experiments.

Many possible phenomena that we have not discussed previously could potentially play a role in this difference, which can all be grouped under the same general denomination of FRET heterogeneity. Up to now, we have assumed that each sample was characterized by a fixed FRET efficiency between the donor and acceptor molecules. Obviously, if this hypothesis does not hold, but we are in the presence of the sample characterized by a distribution of FRET efficiency values, the resulting broadened PRH will be the weighted sum of the PRH's of each respective population of molecules and therefore wider than the PRH of a single-FRET efficiency value sample. Two types of possible experimental sources of FRET efficiency variations can be envisioned: (i) distance distribution and (ii) distribution of photophysical properties.

Distance Distribution

Although dsDNA is fairly rigid, it is still a semiflexible polymer, and some fluctuations of its shape cannot be excluded.³⁵ Likewise, with the dye molecules being attached to the DNA molecule via 6-carbon linkers, a distribution of center-to-center distance between dyes may not be excluded, due for instance to one of the dyes sticking to DNA backbone, as has been demonstrated for some dyes.⁴⁷ Antonik et al. have investigated the amount of distance fluctuation needed to explain the discrepancy between their smFRET measurements and the shot-noise-only theoretical prediction.³² Note that this will affect the measured FRET efficiencies (and thus the PRH) only if these different configurations have lifetimes comparable to or longer than the diffusion time (quasi-static distribution of distances). Fluctuations on significantly shorter time scales (<50 μ s) would indeed result in complete averaging of the measured FRET efficiency within the diffusion time. Assuming such a

(Gaussian) quasi-static distribution of distances (without any claim to its possible physical origin), Figure 10A–E (right column) shows that a perfect match between the experimental result and the shot-noise limited one can be obtained with a Gaussian distribution of distances with an average standard deviation of $1.6 \pm 0.1 \text{ \AA}$.

Distribution of Photophysical Properties

The FRET efficiency not only depends on the distance between the two dyes, but also, via the Förster radius, on the donor quantum yield, the orientation factor, as well as the spectral characteristics of the dyes.⁴⁸ ATTO-647N is, for example, provided as a mixture of two diastereoisomers, but with identical fluorescent properties, according to its manufacturer. Any modification of one or another of these characteristics occurring on a time scale comparable to or longer than the diffusion time could result in a widening of the PRH compared to the shot-noise limited PRH calculated assuming a single distance and a single set of photophysical properties. Unfortunately, it is difficult to check for any fluctuations on these time scales. Single-molecule anisotropy measurements,²⁰ which could possibly reveal some discrepancy with the complete isotropic averaging of the orientation factor κ^2 assumed up to now in this work, only give access to time scales shorter than the dsDNA rotational diffusion time scale, which is on the order of 10–50 ns. On this time scale, indices of DNA distance fluctuations have recently been reported,³⁵ but as we just argued, are irrelevant for the present study. The fluorescence correlation spectroscopy type of measurements, which would allow access to the diffusion time scale of interest, are far from adapted to differentiate between diffusion and dynamics occurring on the same time scale (an order of magnitude between time scales is usually needed to differentiate between processes).²¹

In summary, to our knowledge, there is no simple way to prove or disprove the existence of dye photophysical properties which would be quasi-static on the diffusion time scale relevant to these experiments. As we have done in the previous section for a distance distribution, it would be possible to assume some distribution, say, of κ^2 values and adjust its parameters to obtain a perfect fit to the experimentally observed PRH, but the interpretation of these parameters would remain open to legitimate criticism.

7. Discussion

Comparison with Previous Approaches

As mentioned before, our approach is similar to that recently published by Antonik et al.,³² since both use the experimental BSD, the assumption that donor and acceptor counts are distributed binomially and background counts are approximately Poisson-distributed. However, this work brings several new elements:

- i. We present an analysis of the proximity ratio histogram (PRH), which is more widely used than the ratio studied in ref 32.
- ii. We extend the analysis of the background influence to bursts of arbitrary durations.
- iii. We show that, by using an alternating laser excitation (ALEX) scheme, several sources of contamination of the PRH can be easily eliminated.
- iv. We perform extensive simulations to verify the validity of this approach.
- v. In particular, we study the effect of detection volume mismatch on the PRH.
- vi. We present a practical and rapid algorithm to compute the shot-noise limited histogram.
- vii. This algorithm has the advantage of being easily extendable to more complex situations, such as nontrivial energy landscapes.

viii. In addition, we provide a sound basis for the choice of PRH bin size, which eliminates the spikes and voids observed in most PRH's.

ix. Finally, we present experimental data on several systems (singly labeled dsDNA's and DNA hairpin, presented and discussed later in this section).

As we briefly mentioned in the Introduction, other treatments of shot-noise effects on FRET histograms have been presented before.

Dahan et al. proposed an upper bound on the PRH broadening by shot noise using a simple argument that we will now discuss.^{26,27} This upper bound on the PRH standard deviation is obtained by considering the bursts containing the smallest number of photons, since shot-noise effects are expected to be most dramatic for the smallest counts, even when a ratiometric measurement is considered (eq 7). For this burst size S (equal, for instance, to the threshold S_{\min}), and assuming that the signal is due to molecules characterized by a fixed FRET efficiency E (or to keep in line with the previous discussion, a fixed proximity ratio ε , all corrections being equal to zero and $\gamma = 1$), the following relations will hold:

$$\begin{aligned}\langle A \rangle &= \varepsilon S \\ \langle D \rangle &= (1 - \varepsilon) S\end{aligned}\quad (19)$$

where $\langle A \rangle$ and $\langle D \rangle$ are the average expected signals in the acceptor and donor channels, respectively. The next step consists of assuming that the corresponding acceptor and donor channel signals can be considered as Poisson distributed, with averages given by eq 19. As we have seen, this assumption is incorrect, since the acceptor and donor counts are distributed according to the binomial law (eq 10). In some cases ($\varepsilon \ll 1$), a binomial law of parameters (ε, N) is well-approximated by a Poisson law of identical mean εN , but in general, this is not the case, as can be seen from the different variances of the two distributions

$$\begin{aligned}\text{var}(P_\varepsilon(n|N)) &= N\varepsilon(1 - \varepsilon) \\ \text{var}(\Pi(n|\varepsilon N)) &= N\varepsilon\end{aligned}\quad (20)$$

where $\Pi(n|\mu)$ is the Poisson distribution of mean μ

$$\Pi(n|\mu) = \frac{1}{n!} \mu^n e^{-\mu}\quad (21)$$

If one ignores this limitation, it is possible to proceed formally and note that, for large enough mean values, a Poisson distribution is well-approximated by a γ distribution $\gamma(\nu, \lambda)$, which is in a sense its extension to noninteger values

$$\Pi(n|\mu) \approx \gamma_{\mu+1,1}(n)\quad (22)$$

$$\gamma_{\nu,\lambda}(x) = \frac{\lambda^\nu x^{\nu-1} e^{-\lambda x}}{\Gamma(\nu)}$$

Using a known property of independent γ distributions A and D of means a and d , the random variable $r = A/A + D$ can be shown to be well-approximated by a β distribution

$$\beta_{a,d}(r) = \frac{r^{a-1}(1-r)^{d-1}}{B(a,d)}$$

$$B(a,d) = \int_0^1 u^a(1-u)^d du \quad (23)$$

Using $a = \varepsilon S$ and $d = (1 - \varepsilon)S$ and the properties of β distributions, we obtain that the random variable r has a mean and standard deviation approximately equal to

$$\langle r \rangle = \varepsilon$$

$$\sigma = \sqrt{\frac{\varepsilon(1-\varepsilon)}{S+1}} \quad (24)$$

This latter value is the estimate of the upper bound to the PRH width proposed by Dahan et al. (eq 15 of ref 26, corrected as eq 4 in ref 27). Although we have seen that the premises of this derivation are incorrect, it turns out that the final result (eq 24) is correct, due to the following property of binomial distributions: if A and D are two binomial random variables of parameters (ε, S) and $(1 - \varepsilon, S)$, respectively, the ratio $r = A/A + D$ is also a β distributed random variable with parameters (eq 23) $a = \varepsilon S$ and $b = (1 - \varepsilon)S$, respectively.

The next step toward a better understanding of the contribution of shot noise to the PRH was made by Gopich and Szabo, who first presented a complete analysis of the theoretical shape of the PRH using a fixed time bin approach and simple models of excitation profiles.²⁹ In particular, they obtained theoretical expressions for the standard deviation of the PRH in diverse cases (including two-state model molecules). Using this theoretical value and that of the mean PR, they deduce the best Gaussian and β functions fit to the PRH, a result best applicable to time-trace data for which the notion of an average burst size (number of photons per bin) is meaningful. Extending this result to diffusing molecules resulted in a theoretical formula that could only be calculated numerically and might therefore not be applicable to experimental situations, in which the burst size distribution depends as much from the burst search algorithm parameters than from the experimental settings. The two results of this analysis that can be easily compared to this work, are an upper bound for the standard deviation of the PRH (identical to the result of Dahan et al., eq 24) and an asymptotic standard deviation of the PRH assuming that bursts have a constant duration larger than the diffusion time τ_{dif} and contain a single molecule

$$\sigma \rightarrow \sqrt{\varepsilon(1-\varepsilon)} \left[\sum_{j=0}^{\infty} \frac{1(\bar{n}_0 \tau_{\text{dif}})^j}{j + N_T (1 + \bar{n}_0 \tau_{\text{dif}})^{j+1}} \right]^{1/2} \quad (25)$$

where \bar{n}_0 is the average global count rate averaged over the excitation spot and N_T is the bin content threshold. These results were extended to the case of a two-state molecule. Although these results provided important insights into the relative importance of shot noise, dye photophysics, and molecule dynamics in the case of a two-state molecule, they are difficult to exploit in experiments where the BSD may depart from this ideal situation and strongly depend on the burst search algorithm parameters.

Finally, a recent work by Watkins et al. presented interesting experimental (as well as theoretical) smFRET results on immobilized poly-L-proline molecules taking advantage of the simplification resulting from single average bin content discussed previously. In this case, the shape of the PRH can be calculated as the convolution of the shot-noise effect (resulting in a Gaussian PRH of known standard deviation) to the underlying source of broadening, such as a distance distribution, as we have done at the end of section 6. Their analysis concluded that individual poly-L-prolines are absolutely rigid molecules for a number of prolines varying from 8 to 24, although different molecules exhibit different average distances. This result therefore seems to indicate that previous smFRET experiments on diffusing poly-L-proline presented by Schuler et al.,⁴⁶ which had shown variable amounts of PRH broadening depending on the number of proline residues, should be explainable by a mixture of shot-noise broadening and distance distribution within each population. In this diffusing geometry, the PRH consists indeed of a mixture of molecules that could be characterized by different static donor to acceptor distance values.

Extension to More Complex Systems

The work of Watkins et al. on poly-L-prolines,³³ that of Antonik et al.,³² as well as our analysis of the dsDNA data of Figure 10, show that shot noise is oftentimes the dominating component of the single-molecule PRH width and shape. The underlying best-fit distance distributions can be δ functions³³ or narrow peaks (Figure 10), but could very possibly turn out to be wider in other experimental cases. Equipped with the tools presented in this work, researchers should now be able to make the two analytical steps evoked in the Introduction and extract meaningful information on distance distributions and energy landscapes.

A note of caution is warranted, however: in the poly-L-proline study of Watkins et al.,³³ each immobilized molecule exhibited a single distance, but different molecules exhibited different distances. Despite the carefulness of the authors, it is possible that the molecules were simply stuck on the surface or interacting with it and therefore artificially frozen in a static configuration. In this respect, a re-analysis (using the approach presented here) of the results of Schuler et al.,⁴⁶ obtained in diffusion geometry, would be illuminating. If poly-L-proline molecules are indeed frozen in a rigid conformation for periods of time longer than the diffusion time, the data analysis in terms of a static distribution of distances convolved with the shot-noise effect should recover the distribution of distances reported in ref 33. On the other hand, if the molecules fluctuate rapidly (compared to the diffusion time) between different conformations, a similar distribution would be expected, thus requiring additional dynamic information, for instance, using fluorescence correlation methods. An interesting problem would occur in the intermediate case, where the dynamic of the conformational fluctuations occurs on a time scale comparable to the diffusion time. In certain conditions, the approach presented here can account for this type of situation and, in addition to extracting distance distributions, also allows extracting kinetics parameters of the system, as will be discussed in detail in a future publication.

As an example of these capabilities, we will now briefly discuss the case of a simple two-state system, a DNA hairpin. As illustrated in Figure 11, different kinetic regimes can be observed in hairpins, depending on the energy barriers separating the two configurations, open and closed. As we have seen before (Figure 10), for a rigid dsDNA molecule, the unique energy minimum results in a distribution of distances, which after convolution with the shot noise effect will yield a single-peak PRH (Figure 11A). For hairpins characterized with opening and closing time scales much longer than the typical diffusion time (~ 1 ms), each molecule is virtually frozen in one of the two energy minima characteristic of the system. After convolution with the shot-noise effect, the PRH will therefore exhibit two separate peaks (Figure 11B). For hairpins with a low energy barrier between the two minima, rapid transitions between the two states are expected, and each molecule samples

the two typical conformations several times during each burst, resulting in a broad peak located between the expected position of each characteristics conformation (Figure 11C). Finally, for time constants comparable with the characteristic diffusion time of the molecule, the two separate peak locations will still be visible, but intermediate values will also be observed (Figure 11D). A combined analysis of the influence of shot noise and of a simple two-state kinetic scheme on the PRH should in principle allow extraction of the kinetic parameters of the system.

We designed a DNA hairpin (Figure 3B) with kinetic parameters corresponding to the last case, i.e., with opening and closing times comparable to the diffusion time. As shown in Figure 12, the experimental PRH (gray bar histogram) exhibits two local maxima lacking a clear separation (for instance, this histogram cannot be fitted by a sum of two Gaussian distributions). Using a Monte Carlo sampling approach similar to that described for the computation of eq 17, but including two transition rates k_C and k_O between two populations characterized by distinct PR values (open ε_O and closed ε_C), we were able to account for the observed shape of the PRH (black curve). The extracted rates, $k_C = 1425 \text{ s}^{-1}$ and $k_O = 835 \text{ s}^{-1}$, correspond to time scales (comparable to the diffusion time) of $\tau_{O \rightarrow C} = 702 \mu\text{s}$ and $\tau_{C \rightarrow O} = 1.2 \text{ ms}$.

8. Conclusions

We developed an approach that accurately accounts for the contribution of shot noise to the PRH in smFRET experiments. In our model dsDNA system, we found that the width of the PRH is almost always in excess of shot noise. This excess width could originate from systematic artifacts inherent in the diffusion smFRET method itself or have a true molecular origin. If molecular static and/or dynamic heterogeneities could indeed be unraveled by our method, one should be able to derive distance distributions and ultimately energy landscapes. Also, direct measurement of fluctuations could shed new light on structure–function relationship in proteins.⁴⁹ It is therefore crucial to understand all possible contributions to the excess width.

We thoroughly examined possible artifacts and possible remedies using both simulations and experiments. We have shown how to account for background, dye blinking and bleaching, coincident detection events, and detection volume mismatch. All of these possible culprits indeed contributed to excess width, but some can be avoided with a proper data acquisition scheme and burst search algorithm, and the remainder do not have a magnitude large enough to explain the dsDNA observations.

Excess width could come from dye tethers, dye docking onto DNA (κ^2 fluctuations), or unexpected flexibility of the dsDNA or the B-DNA itself, as recent studies have suggested.³⁵ However, the time scale of these hypothetical fluctuations, which should be similar to the diffusion time of the molecule in order to have an effect on the PRH width, renders them difficult to study. Regardless of the origin of these fluctuations, our method and control experiments suggest that we are indeed able to extract heterogeneities, the source of which will be the subject of further study. Our method is one more addition to an array of recently introduced fluorescence methods to study such heterogeneities, including FCS,^{50,51} fluorescence lifetime correlation,^{52,53} or pulsed interleaved excitation/ns-ALEX,^{35,54} among others.

Last, our method can be easily extended to fit multistate PRH, extract the multiple distances characteristic of these states and the transition rates between them (Nir et al., unpublished).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Marcus Jager and Xiangxu Kong for helpful discussions. X.M. thanks Kevin Knuth for discussions about his optimal histogram binning algorithm. We are grateful to Attila Szabo for critical reading of an earlier version of this manuscript. This work was funded by NIH grant no. 1R01-GM65382 (to S.W.) and the NSF. The Center for Biophotonics, an NSF Science and Technology Center, is managed by the University of California, Davis, under Cooperative Agreement no. PHY0120999. E.N. is supported by the Human Frontier Science Program (HFSP).

References and Notes

1. Stryer L, Haugland RP. *Proc Natl Acad Sci USA*. 1967; 58:719. [PubMed: 5233469]
2. Weiss S. *Science*. 1999; 283:1676. [PubMed: 10073925]
3. Weiss S. *Nat Struct Biol*. 2000; 7:724. [PubMed: 10966638]
4. Ha T. *Biochemistry*. 2004; 43:4055. [PubMed: 15065847]
5. Michalet X, Weiss S, Jäger M. *Chem Rev*. 2006; 106:1785. [PubMed: 16683755]
6. Kapanidis AN, Laurence TA, Lee NK, Margeat E, Kong X, Weiss S. *Acc Chem Res*. 2005; 38:523. [PubMed: 16028886]
7. Koyama-Honda I, Ritchie K, Fujiwara T, Iino R, Murakoshi H, Kasai RS, Kusumi A. *Biophys J*. 2004; 88:2126. [PubMed: 15596511]
8. Sako Y, Minoghchi S, Yanagida T. *Nat Cell Biol*. 2000; 2:168. [PubMed: 10707088]
9. Haran G. *J Phys: Condens Matter*. 2003; 15:R1291.
10. Ishijima A, Yanagida T. *Trends Biochem Sci*. 2001; 26:438. [PubMed: 11440856]
11. Zhuang XW. *Annu Rev Biophys Biomol Struct*. 2005; 34:399. [PubMed: 15869396]
12. Schuler B. *ChemPhysChem*. 2005; 6:1206. [PubMed: 15991265]
13. Myong S, Stevens BC, Ha T. *Structure*. 2006; 14:633. [PubMed: 16615904]
14. Lee NK, Kapanidis AN, Wang Y, Michalet X, Mukhopadhyay J, Ebricht RH, Weiss S. *Biophys J*. 2005; 88:2939. [PubMed: 15653725]
15. Sabanayagam CR, Eid JS, Meller A. *J Chem Phys*. 2005; 122:061103. [PubMed: 15740360]
16. Watkins LP, Yang H. *Biophys J*. 2004; 86:4015. [PubMed: 15189897]
17. Hübner, CG.; Krylov, V.; Renn, A.; Nyffeler, P.; Wild, UP. *Single-Molecule Fluorescence – Each Photon Counts*. In: Rigler, R.; Orrit, M.; Basche, T., editors. *Single Molecule Spectroscopy*. Springer-Verlag; Stockholm: 2001.
18. Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, Weiss S. *Proc Natl Acad Sci USA*. 2004; 101:8936. [PubMed: 15175430]
19. Deniz AA, Dahan M, Grunwell JR, Ha T, Faulhaber AE, Chemla DS, Weiss S, Schultz PG. *Proc Natl Acad Sci USA*. 1999; 96:3670. [PubMed: 10097095]
20. Ha T, Laurence TA, Chemla DS, Weiss S. *J Phys Chem B*. 1999; 103:6839.
21. Krichevsky O, Bonnet G. *Rep Prog Phys*. 2002; 65:251.
22. Eggeling C, Fries JR, Brand L, Günther R, Seidel CA. *Proc Natl Acad Sci USA*. 1998; 95:1556. [PubMed: 9465054]
23. Eggeling C, Berger S, Brand L, Fries JR, Schaffer J, Volkmer A, Seidel CAM. *J Biotechnol*. 2001; 86:163. [PubMed: 11257530]
24. Margittai M, Widengren J, Schweinberger E, Schröder GF, Felekyan S, Hausteiner E, König M, Fasshauer D, Grubmüller H, Jahn R, Seidel CAM. *Proc Natl Acad Sci USA*. 2003; 100:15516. [PubMed: 14668446]
25. Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, Chemla DS, Dawson PE, Schultz PG, Weiss S. *Proc Natl Acad Sci USA*. 2000; 97:5179. [PubMed: 10792044]
26. Dahan M, Deniz AA, Ha T, Chemla DS, Schultz PG, Weiss S. *Chem Phys*. 1999; 247:85.

27. Deniz AA, Laurence TA, Dahan M, Chemla DS, Schultz PG, Weiss S. *Annu Rev Phys Chem.* 2001; 52:233–253. [PubMed: 11326065]
28. Rothwell PJ, Berger S, Kensch O, Felekyan S, Antonik M, Wöhrl BM, Restle T, Goody RS, Seidel CAM. *Proc Natl Acad Sci USA.* 2003; 100:1655. [PubMed: 12578980]
29. Gopich I, Szabo A. *J Chem Phys.* 2005; 122:014707.
30. Slaughter BD, Unruh JR, Price ES, Huynh JL, Urbauer RJB, Johnson CK. *J Am Chem Soc.* 2005; 127:12107. [PubMed: 16117552]
31. Schuler B, Lipman EA, Eaton WA. *Nature (London).* 2002; 419:743. [PubMed: 12384704]
32. Antonik M, Felekyan S, Gaiduk A, Seidel CAM. *J Phys Chem B.* 2006; 110:6970. [PubMed: 16571010]
33. Watkins LP, Chang H, Yang H. *J Phys Chem B.* 2006; 110:5191. [PubMed: 16539447]
34. Zhang K, Yang H. *J Phys Chem B.* 2005; 109:21930. [PubMed: 16853850]
35. Laurence TA, Kong X, Jäger M, Weiss S. *Proc Natl Acad Sci USA.* 2005; 102:17348. [PubMed: 16287971]
36. Knuth, KH. 2006. <http://arxiv.org/abs/physics/0605197>. Unpublished
37. Jia Y, Talaga DS, Lau WL, Lu HSM, DeGrado WF, Hochstrasser RM. *Chem Phys.* 1999; 247:69.
38. Talaga DS, Lau WL, Roder H, Tang J, Jia Y, DeGrado WF, Hochstrasser RM. *Proc Natl Acad Sci USA.* 2000; 97:13021. [PubMed: 11087856]
39. Press, WH.; Flannery, BP.; Teukolsky, SA.; Vetterling, WT. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press; New York: 1992.
40. Kapanidis AN, Ebright YW, Ludescher RD, Chan S, Ebright RH. *J Mol Biol.* 2001; 312:453. [PubMed: 11563909]
41. Lacoste TD, Michalet X, Pinaud F, Chemla DS, Alivisatos AP, Weiss S. *Proc Natl Acad Sci USA.* 2000; 97:9461–9466. [PubMed: 10931959]
42. Enderlein J, Robbins DL, Ambrose WP, Goodwin PM, Keller RA. *J Phys Chem B.* 1997; 101:3626.
43. Fries JR, Brand L, Eggeling C, Köllner M, Seidel CAM. *J Phys Chem A.* 1998; 102:6601.
44. Bustamante C, Bryant Z, Smith SB. *Nature (London).* 2003; 421:423. [PubMed: 12540915]
45. Widengren J, Schweinberger E, Berger S, Seidel CAM. *J Phys Chem A.* 2001; 105:6851.
46. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA. *Proc Natl Acad Sci USA.* 2005; 102:2754. [PubMed: 15699337]
47. Hillisch A, Lorenz M, Diekmann S. *Curr Opin Struct Biol.* 2001; 11:201. [PubMed: 11297928]
48. Lakowicz, JR. *Principles of Fluorescence Spectroscopy.* 2. Plenum; New York: 1999.
49. Welch, GR., editor. *The Fluctuating Enzyme.* John Wiley; New York: 1986.
50. Chattopadhyay K, Saffarian S, Elson EL, Frieden C. *Proc Natl Acad Sci USA.* 2002; 99:14171. [PubMed: 12381795]
51. Neuweiler H, Doose S, Sauer M. *Proc Natl Acad Sci USA.* 2005; 102:16650. [PubMed: 16269542]
52. Yang H, Luo G, Karnchanaphanurach P, Louie TM, Rech I, Cova S, Xun L, Xie XS. *Science.* 2003; 302:262. [PubMed: 14551431]
53. Xie XS. *J Chem Phys.* 2002; 117:11024.
54. Müller BK, Zaychikov E, Bräuchle C, Lamb DC. *Biophys J.* 2005; 89:3508. [PubMed: 16113120]

Appendix A: Burst Search Algorithms

Information theory provides rigorous criteria for defining photon bursts corresponding to single molecules, provided sufficient knowledge of the sample and optical setup can be provided.³⁴ Zhang and Yang have compared an approach based on information theory and a standard binning/thresholding method, showing that the two have comparable performance for shorter time bins (down to 100 μ s).³⁴ Although they did not perform comparisons with the time lag based method originally introduced by the Seidel group,²³ it is expected that a similar result will hold in this case, as both rely on thresholding.

Here, we briefly describe the rationale behind the choice of parameters used for the “all photons burst search” (APBS) and “dual channel burst search” (DCBS) methods used in this work and based on the Seidel group work.²³ In both algorithms, three parameters are needed to define a burst: the averaging time window, T , the minimum number of photons per window, M , and the minimum number of photons per burst, L .

As a reminder of Eggeling et al.’s method,²³ we restate their definition in terms of photon arrival time rather than in terms of time lag, as originally done. The start (respectively, the end) of a potential burst is detected when the number of photons in the averaging window of duration T is larger (respectively, smaller) than the minimum number of photons M . A potential burst is retained if the number of photons it contains is larger than a minimum number L . Clearly, these parameters need to be adjusted for samples with different brightness or background levels.

Shorter time windows T will result in a better time resolution, which may help detecting transitions such as blinking and bleaching steps. However, too short a time window duration might artificially split bursts corresponding to molecules wandering around in the excitation volume in many sub-bursts, as the excitation intensity decreases in the rim of the excitation volume. This will result in a burst size distribution (BSD) with smaller burst sizes.

A related parameter is the minimum number of photons per time window, M . If kept constant while the time window size is reduced, the number of detected bursts can be kept constant. In other words, bursts will not be split when using smaller averaging time windows, if the requested minimum number of photons is reduced proportionally. The ratio M/T can indeed be considered as an instantaneous emission rate. Consequently, the choice of a given M/T ratio will set the minimum burst brightness. A large value will favor bursts that correspond to molecules diffusing through the center of the excitation volume, whereas a smaller value will also retain bursts corresponding to molecules, which never diffuse through the center of the excitation volume.

The minimum number of photon per bursts, L , only has an effect on the BSD but does not affect the burst search process itself. It can easily be varied by the user after the burst search process has been performed. Smaller L values will retain smaller bursts and therefore result in broadening of the PRH. Increasing L will weed out small bursts and therefore result in narrowing of the PRH.

We have empirically found out that, for the bright samples used in the experiments described here, a set of APBS parameter $M = 30$ and $T = 500 \mu\text{s}$ (emission rate 60 kHz) is appropriate. The PRH values shown are obtained by further limiting ourselves to $L = 50$. For the DCBS algorithm, only half of the diffusion time of the molecule is spent undergoing excitation by the donor-excitation laser. The other half is indeed spent undergoing excitation by the acceptor-excitation laser. Accordingly, to preserve the same burst discrimination level, we require half the number of photons per identical averaging time window: $M = 15$, $T = 500 \mu\text{s}$. Similarly, to obtain an equivalent burst size threshold to the APBS, half the value chosen in the APBS case has to be chosen in the DCBS case where each excitation channel is considered separately.

Appendix B: Modified Knuth Algorithm for Optimal Binning of the PRH

Knuth’s algorithm to determine the optimal bin number \hat{M} for a data set $\underline{d} = \{d_i; 1 \leq i \leq N\}$ relies on a Bayesian approach and does not suppose any specific form for the underlying probability distribution of the data points d_i .³⁶ Instead, one looks for the most likely stepwise function that models the data distribution. \hat{M} is obtained as the integer maximizing the posterior probability

$$p(M|d) \propto \left(\frac{M}{V}\right)^N \frac{\Gamma\left(\frac{M}{2}\right) \prod_{k=1}^M \Gamma\left(n_k + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^M \Gamma\left(N + \frac{M}{2}\right)} \quad (26)$$

where V is the interval size and the n_k 's are the bin values for the specified number of bins M . Practically, it is simpler to minimize the logarithm of the posterior probability (LPP)

$$\ln p(M|d) = N \ln M + \ln \Gamma\left(\frac{M}{2}\right) - M \ln \Gamma\left(\frac{1}{2}\right) - \ln \Gamma\left(N + \frac{M}{2}\right) + \sum_{k=1}^M \ln \Gamma\left(n_k + \frac{1}{2}\right) + K \quad (27)$$

where K is an irrelevant constant. As described in ref 36, a simple approach to find the maximum of this quantity consists of a brute force calculation for $M = 1$ to a maximum value (for instance, equal V divided by the minimum separation between the d_i 's).

Reference 36 tests this algorithm with a Gaussian distribution, a four-step function, a flat distribution, and a three-peaked distribution, recovering in each case a sensible number of bins by maximizing $p(M|d)$.

Additionally, the analysis provides the variance of the bin heights as

$$\sigma_k^2 = \left(\frac{M}{V}\right)^2 \frac{\left(n_k + \frac{1}{2}\right) \left[N + \frac{M}{2} - \left(n_k + \frac{1}{2}\right)\right]}{\left(N + \frac{M}{2}\right)^2 \left(N + \frac{M}{2} + 1\right)} \quad (28)$$

Unfortunately, experimental PRH do not necessarily result in posterior probability having a maximum (that is, the resulting LPP can either be monotonically increasing or exhibit only a local maximum). This results from the discrete nature of its rational values. Two additional ingredients are needed to define a meaningful optimal bin number in the case of the PRH.

Dithering the Data

The first step consists of adding a small random (real) number to each value d_i . If these random numbers are small enough, they should not perturb the overall shape of the histogram, and hence, the optimal number of bins obtained for the modified data set should be valid for the unmodified data set. If they are large enough, they should break the pattern of discrete clusters of values resulting from the rational nature of the PR. Experimentally, too small a value does not lead to a maximum of $p(M|d)$. We found that using a flat distribution of random numbers in the interval $[-5\varepsilon, +5\varepsilon]$, where ε is the smallest separation between the individual data point, usually results in a function $p(M|d)$ having a clear maximum, but a Gaussian distribution of standard deviation 5ε works equally well. In practice, using an amplitude/standard deviation of 0.01 should work for most PRH's.

Averaging the LPP

Even though this dithering step results in a $p(M|d)$ function exhibiting a maximum, it turns out that the location of the maximum depends rather sensitively on the actual distribution of

random numbers used to dither the data set. Fortunately, the average function $\langle \ln p(M|d) \rangle$ obtained from many dithered data sets obtained with different random numbers is itself independent of the random numbers used for dithering. Experimentally, we found that an average over $R = 100$ sets of random numbers is sufficient to attain a stable $\langle \ln p(M|d) \rangle$ function and, therefore, unambiguously define the optimal bin number \hat{M} . Note however that $\langle \ln p(M|d) \rangle$ functions (or, for that matter, $\ln p(M|d)$ functions) are far from smooth, so that some care needs to be used in defining the location of the maximum of $\langle \ln p(M|d) \rangle$. We note also that we base our analysis on the average of $\ln p(M|d)$, rather than on the logarithm of the posterior probabilities, $\ln \langle p(M|d) \rangle$, for computational convenience.

As an illustration of this technique, we present three examples of typical real data sets analyzed following the above protocol. One set corresponds to a sample with 1588 PR values (sample 1, Figure S3) centered around 0.64, another to a different sample with 780 PR values (sample 2, Figure S4) centered around 0.35, and the last one to a combination of both (sample 1 + 2, Figure S5). An additional set corresponding to a flat distribution of PR values already discussed in the main text can also be found in Figure S1 (Supporting information).

Sample 1 (Figure S3A) clearly shows the pathology of data sets made of rational numbers: the logarithm of the posterior probability (LPP) keeps increasing for large bin numbers. Dithering the data set with a random number evenly distributed in $[-0.01, 0.01]$ fixes this problem. Figure S3B illustrates the fact that some averaging is needed before a maximum of the LPP can be unambiguously defined. Notice also that, even averaged over 10 000 dithered data sets, the LPP still looks very noisy. Nevertheless, a clear maximum can be found for $\hat{M} = 37$. Figure S3C,D presents the two histograms (without and with dithering), binned with the optimal bin number and with 100 bins. The latter value reveals the spikes and voids (Figure S3C, red), which do not totally disappear upon dithering (Figure S3D, red), whereas the optimal bin number completely eliminates the problem in both cases.

Sample 2 (Figure S4A) exhibits a similarly pathological LPP, with a local maximum around 20 and a continuous increase for large bin numbers (not shown in the figure). This is fixed with a dithering amplitude $\varepsilon = 0.01$, as before, yielding an optimal bin number $\hat{M} = 21$. As in the previous example, the dithered histogram looks smoother than the original histogram (for $M = 100$ bins, Figure S4C,D), but for the optimal number of bins, both the original data set and the dithered one have for the most part lost spike and void features.

By combining the data from both samples, we artificially create a sample containing a mixture of small and large PR values (Figure S5). Here, as before, dithering is necessary to obtain an LPP presenting a clear maximum, obtained for $\hat{M} = 47$. Note that this larger number of bins results in some spikes and voids still visible in the histogram of the nondithered data set (Figure S5D).

Smoothing the LPP

A close look at the shape of the average LPP in all cases (Figure S3–5B) reveals that they themselves contain a lot of “spikes and voids”. This is somewhat expected for a function defined for integers only and for which each new number results in a new binning of the histogram. For histograms such as PRH’s with spikes and voids, changing the boundaries of successive bins will result in abrupt variations of the bin sizes around these spikes, as can be easily verified (data not shown). The LPP may thus have an absolute maximum that may not correspond to that of the underlying smooth function that one may want to mentally draw on top of the noisy LPP. A simple way to render this discussion quantitative is to fit the LPP with a sum of exponentials. This is done for the average LPP’s obtained from 100 dithered

data sets in Figure S3–5E, using sums of 4 exponentials. As can be seen on these graphs, the location of the maximum of the smoothed LPP can be rather different from that obtained from the original function (sample 1, 33 instead of 37; sample 2, 19 instead of 21; sample 1 + 2, 40 instead of 47).

Further Dithering and Asymptotic Behavior of the LPP

As we just discussed, the reason for the noise in the LPP is the presence of spikes in the underlying histograms. A natural idea is thus to look at the LPP for histograms with spikes that are sufficiently smoothed out to result in a smooth LPP. Since we have already dithered the original data set to obtain LPP's exhibiting a clear maximum, the only option left is to increase the dithering amplitude and check the resulting averaged LPP, until a clear-cut maximum can be defined. This analysis is shown in Figure S6 for the three data sets studied previously. Figure S6A,C,E shows the LPP's obtained after averaging over 1000 dithered data sets, with dithering amplitudes ranging from 0.01 to 0.10. In all cases, a smooth average LPP is obtained for a dithering amplitude $\varepsilon \geq 0.05$. The extracted maxima are plotted in Figure S6B,D, and F. A well-defined asymptotic optimal bin number can be extracted in all cases (sample 1, 22; sample 2, 16; sample 1 + 2, 23), which is significantly smaller than both the maximum obtained from the average LPP for $\varepsilon = 0.01$ and that obtained from the smoothed LPP. Interestingly, the optimal bin number obtained for the combined sample is almost identical to the maximum of the two numbers obtained for each separate sample, as expected. This is in marked contrast to the previous result obtained with either the averaged LPP or the smoothed LPP ($\varepsilon = 0.01$). In these cases, the bin number obtained for the combined data sets was systematically larger than those of the individual data sets, pointing to some potential problem in the definition of these numbers.

It is worth having a look at the evolution of the histograms as a function of the dithering amplitude in one particular case (sample 1, Figure S7), to verify the assumption underlying our approach, namely, that dithering results in fewer spikes, which in turns results in smoother LPP's. Figure S7A–I shows two histograms per dithering amplitude (from $\varepsilon = 0$ to $\varepsilon = 0.08$ in 0.01 steps): a 100 bin histogram, exhibiting clear spikes and voids in the original data set, and the histogram obtained using the optimal number of bins as defined by the maximum of the averaged LPP for this dithering amplitude (see Figure S5B). Unsurprisingly, the 100 bin histograms show fewer and fewer spikes and voids as the dithering amplitude increases, and the histograms built with the optimal bin numbers do not exhibit any for all values of the amplitude. Figure S8 illustrates an expected (though limited) effect of dithering, namely, histogram broadening. Figure S8A superimposes the “optimal” histograms obtained for different dithering amplitudes: qualitatively, they look very similar. A Gaussian fit, however, reveals that their width steadily increases with dithering amplitude, as expected (Figure S8B). Note that these histograms are not what we are interested in but are simply a tool to determine the optimal bin number to apply on the original data set.

For all practical matters, it results from this analysis that the optimal number obtained for $\varepsilon \approx 0.10$ should be within a few units of the asymptotic optimal number.

Summary

The previous section has shown that there is a natural way to define an optimum bin number for the PRH. Our suggested algorithm is the following:

- i. Generate 100–1000 dithered data sets with a dithering amplitude 0.05–0.10.
- ii. Compute the average LPP of these data sets.
- iii. The location of maximum is the optimal bin number.

Figure S9 shows the histograms (of the original data sets) built using the optimal bin numbers obtained in the previous section, superimposed on 100 bin histograms. Spikes and voids are gone, while the fine structure of the PR distribution remains clearly visible.

Appendix C: Theoretical Expression of ε in the Absence of Background

Using the notations introduced in section 2, the expected PR ratio has the following expression:

$$\varepsilon = \frac{Lk + \text{Dir} + F^{\text{FRET}}}{Lk + \text{Dir} + F^{\text{FRET}} + F_D^{\text{D}}} \quad (29)$$

Following ref 14, it is easy to show that the leakage contribution to the acceptor counts is proportional to the donor signal

$$Lk = \lambda F_D^{\text{D}} \quad (30)$$

where λ is the donor-leakage coefficient.

The acceptor direct excitation term was expressed in ref 14 in terms of the acceptor signal upon acceptor excitation, which is only available while performing ALEX experiments. Coming back to the origin of direct excitation of the acceptor

$$\text{Dir} = I_D \sigma_D^{\text{A}} \varphi_A \eta_A \quad (31)$$

where the notations of ref 14 are used. I_D is the donor excitation intensity, σ_D^{A} is the absorption cross section of the acceptor at the donor excitation wavelength. We can rewrite

$$\begin{aligned} \text{Dir} &= I_D \sigma_D^{\text{A}} \varphi_D \eta_D \frac{\sigma_D^{\text{A}} \varphi_A \eta_A}{\sigma_D^{\text{D}} \varphi_D \eta_D} = \frac{\delta}{(1-E)\gamma} F_D^{\text{D}} \\ \delta &= \frac{\sigma_D^{\text{A}}}{\sigma_D^{\text{D}}} \end{aligned} \quad (32)$$

Finally, following ref 14

$$F^{\text{FRET}} = E I_D \sigma_D^{\text{D}} \varphi_A \eta_A = \gamma \frac{E}{1-E} F_D^{\text{D}} \quad (33)$$

where σ_D^{D} is the absorption cross section of the donor at the donor excitation wavelength.

Using these expressions in eq 29, we obtain

$$\varepsilon = 1 - \left(1 + \lambda + \gamma \frac{E}{1-E} + \frac{\delta}{\gamma(1-E)} \right)^{-1} \quad (34)$$

which is the expression obtained by Antonik et al.,³² with the addition of the term due to direct excitation of the acceptor not considered by these authors.

Appendix D: Shot-Noise Limited PRH in the Presence of Background

In ref 32, Antonik et al. proposed an expression to compute the probability distribution of D/A ratios that included the effect of background. To obtain their results, they used several simplifications. First, to avoid having to deal with bursts having different durations, they slice the detected bursts into pieces of equal duration (1 ms) and consider each slice as a new effective burst. The main issue with this treatment is that it reduces the burst size (total number of counts in a burst), therefore increasing the effect of shot noise on the calculated histogram (be it the ratio of donor to acceptor count histogram as in ref 32 or the PRH in this work). Also, as burst durations might not be commensurate to the chosen constant slice duration, a lot of photons will be discarded because they correspond to slices of duration less than 1 ms. A more efficient way would be to look at the burst distribution as a bidimensional one, $\text{BSD}(S, \tau)$: each burst is characterized not only by its total number of photons S (discrete, integer variable) but also by its duration τ (continuous, real variable). To take into account the background counts, we need to know their probability distribution for each different burst size S and duration τ : $\beta_{D,A}(N|S, \tau)$, where the index D or A stands for donor background counts or acceptor background counts. In the absence of burst size constraint, the probability to obtain d (respectively, a) background counts in the donor (respectively, acceptor) channel is given by a Poisson law

$$\begin{aligned} \Pi_D(d|\delta\tau) &= \frac{1}{d!} (\delta\tau)^d e^{-\delta\tau} \\ \Pi_A(a|\alpha\tau) &= \frac{1}{a!} (\alpha\tau)^a e^{-\alpha\tau} \end{aligned} \quad (35)$$

where δ (respectively, α) is the background rate in the donor (respectively, acceptor) channel. The distributions $\beta_{D,A}(N|S, \tau)$ should in principle depart from simple Poisson ones, since a constraint on the burst size is imposed. Indeed, as noted by Antonik et al., Poisson distributions have nonzero values for all integer numbers, which cannot possibly be the case for a finite burst size. In practice, however, the probability of having numbers outside the range $[\mu - 5\sqrt{\mu}, \mu + 5\sqrt{\mu}]$, where μ is the mean of the Poisson distribution, is minimal (Poisson distributions are well-approximated by normal distribution for large μ , for which this probability is easily computed using the error function). Thus, if $\delta\tau + 5\sqrt{\delta\tau}$ and $\alpha\tau + 5\sqrt{\alpha\tau}$ are much smaller than S , using the Poisson distribution $\Pi(N|\mu)$ instead of the unknown distribution $\beta_{D,A}(N|S, \tau)$ should not introduce a large error in the computation. This is illustrated by a direct comparison of the Poisson distribution $\Pi(N|\delta\tau)$ and the distribution $\beta_D(N|S, \tau)$ calculated from simulated data and different values of S and τ in Figure S10.

Using the same type of arguments used in ref 32, but including bursts of any duration, one obtains the following expression for the probability to observe a value x of the PR:

$$\tilde{p}_\varepsilon(x) = \chi_{Q \cap [0,1]} \times \frac{1}{B} \sum_{i=1}^n \sum_{S=S_{\min}}^{S_{\max}} \text{BSD}(S, i\Delta) \sum_{\substack{d+a=S \\ d,a \\ d+a=0}} \beta_D(d|S, i\Delta) \times \beta_A(a|S, i\Delta) \times P_\varepsilon(x[S-d-a]|S-d-a) \chi_N(x[S-d-a]) \quad (36)$$

where $\beta_D(N|S, \tau)$ and $\beta_A(N|S, \tau)$ can be safely replaced by $\Pi(N|\delta\tau)$ and $\Pi(N|\alpha\tau)$ introduced in the main text (eqs 1 and 2), and the burst durations have been histogrammed with a time resolution Δ .

This expression is a bit more difficult to evaluate than the corresponding expression without background, but the same type of algorithm that we proposed in the latter case can be employed here too, using Poisson-distributed random numbers as well as binomial-distributed ones. The algorithm presented in the text for the calculation of the shot-noise limited PRH in the absence of background has thus to be modified as follows:

- i.** Choose an oversampling factor N and a putative value ε .
- ii.** For each burst, calculate $\delta\tau$ and $\alpha\tau$, where τ is the burst duration.
- iii.** Draw one number d from the Poisson distribution of mean $\delta\tau$ and one number a from the Poisson distribution of mean $\alpha\tau$.
- iv.** Draw a number A from the binomial distribution of size $S - d - a$ and probability ε , where S is the burst size.
- v.** Add the value $A/(S - d - a)$ to the PRH.
- vi.** Repeat (iii)–(v) N times.
- vii.** Repeat (ii)–(vi) for each collected burst.
- viii.** Divide the final PRH by N .
- ix.** Improve on ε using a dichotomic search and least-squares minimization.

This algorithm avoids calculating the joint histogram of burst size and duration introduced formally to simplify the notations in eq 36. It is naturally simplified into the practical algorithm presented in the main text when no consideration of background is taken into account.

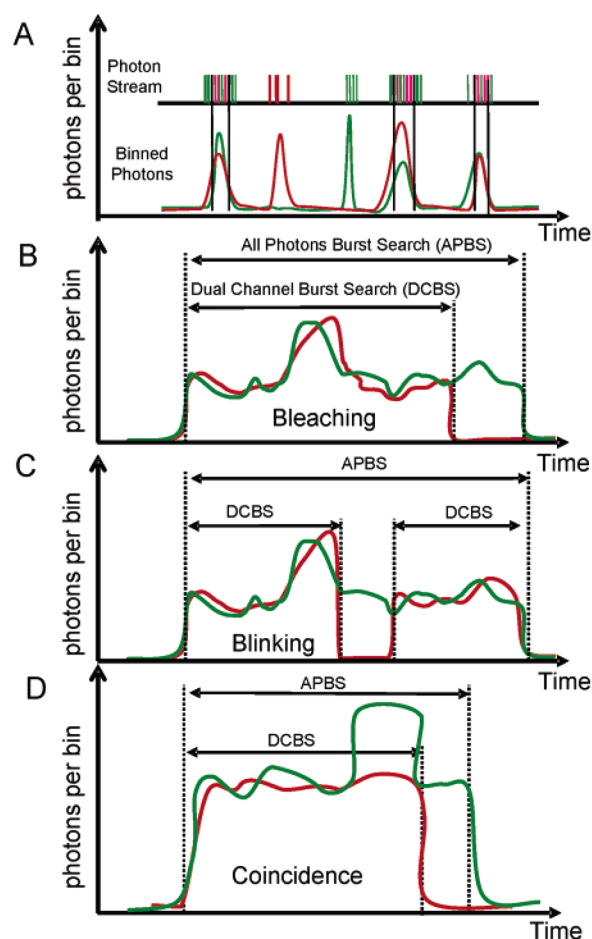


Figure 1.

A schematic comparison of the all-photon-burst-search (APBS) with the dual-channel-burst-search (DCBS) for three possible scenarios. The green line represents the sum of donor photons and acceptor photons detected during the green laser periods, while the red line represents the red photons detected during the red laser periods. A. Schematic relationship between photon streams and binned time traces. B. Bleaching scenario. In the case shown here, the acceptor bleaches toward the end of the burst. The APBS algorithm considers all the detected photons as belonging to the same burst and, hence, results in a calculated PR value that is lower than the true value. On the contrary, the DCBS algorithm interrupts the burst where the acceptor is bleached, resulting in an accurate PR value. C. Blinking scenario. In the case considered here, the acceptor blinks during the burst. As in A, the APBS algorithm results in a PR value that is lower than the true value. The DCBS algorithm results in two successive, smaller bursts with an accurate PR value. D. Coincidence scenario. In the case shown here, the D–A species is mixed with a donor-only molecule toward the end of the burst. The APBS algorithm considers all photons as belonging to a unique burst, resulting in an inaccurate PR value. The DCBS algorithm rejects part of the donor photons, resulting in an improved PR value.

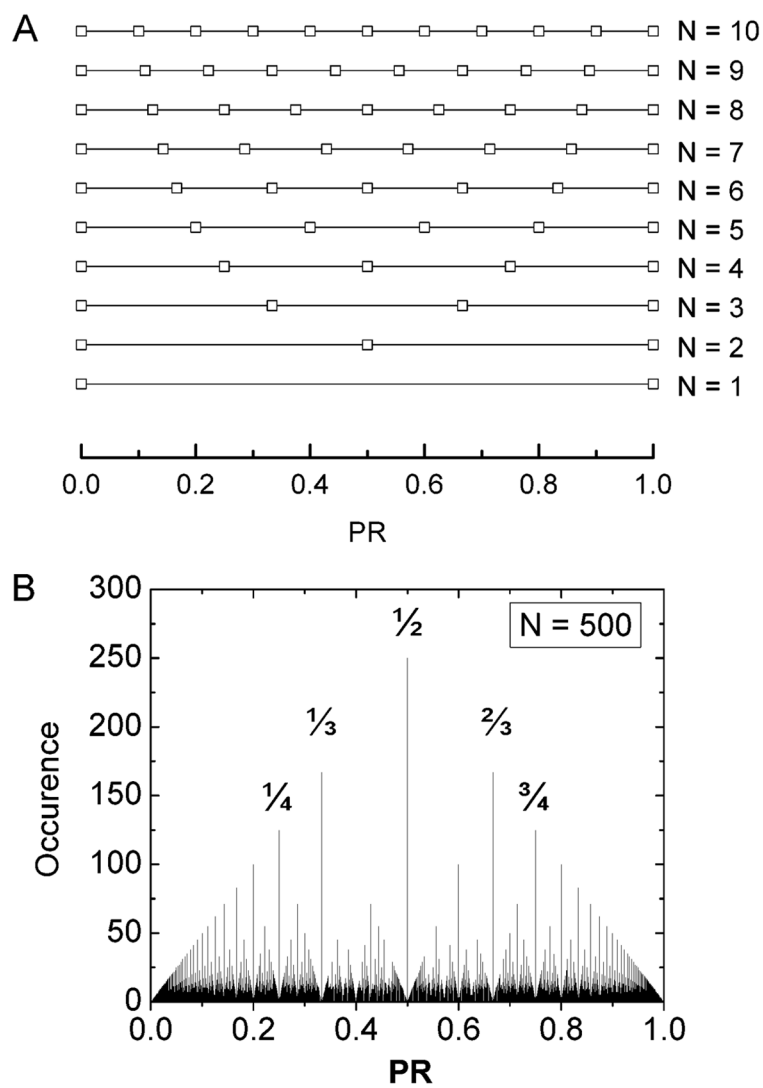
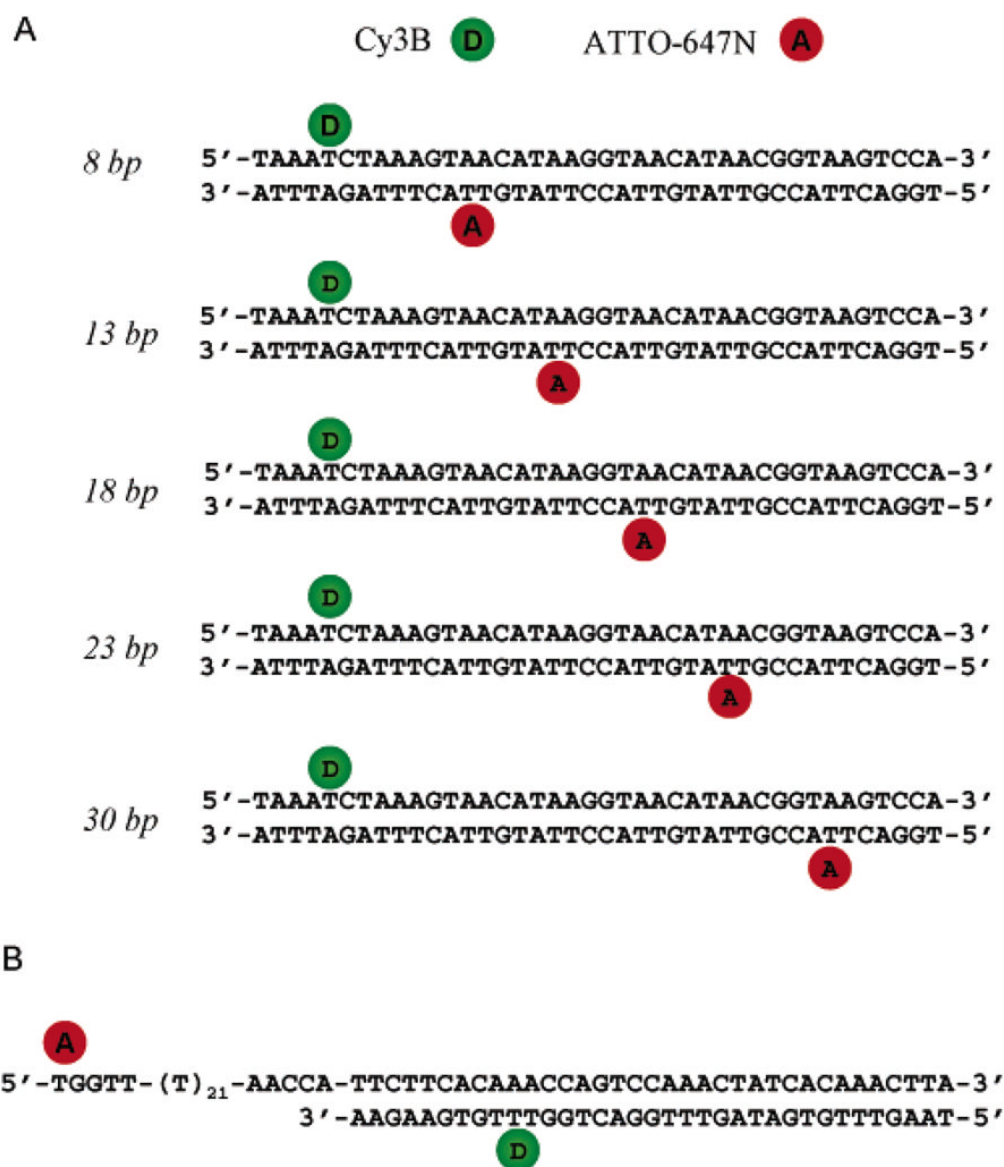


Figure 2. A. Distribution of values $\{i/N; 0 \leq i \leq N\}$ for $N = 1-10$. B. Distribution of values $\{i/N; 0 \leq i \leq N; 1 \leq N \leq 500\}$. The zero-count is not represented ($=500$).

**Figure 3.**

A. 40 base pair dsDNA molecules used in this work. The top strand is labeled at position 5 (from the 5' end) with Cy3B (donor). Five different bottom strands with complementary sequence are labeled with ATTO-647N (acceptor) at position 13, 18, 23, 28, and 35, respectively, from the 3' end. B. DNA hairpin used in this work (Figure 12). The 35 bp dsDNA stem is followed on the top strand by a 5 bp ssDNA stem, a 21 bp poly-T loop, and a complementary sequence to the previous 5 bp. The acceptor dye is attached to the 5' end of the top strand, while the donor dye is attached on the bottom strand at a 10 bp distance from the acceptor dye in the closed conformation.

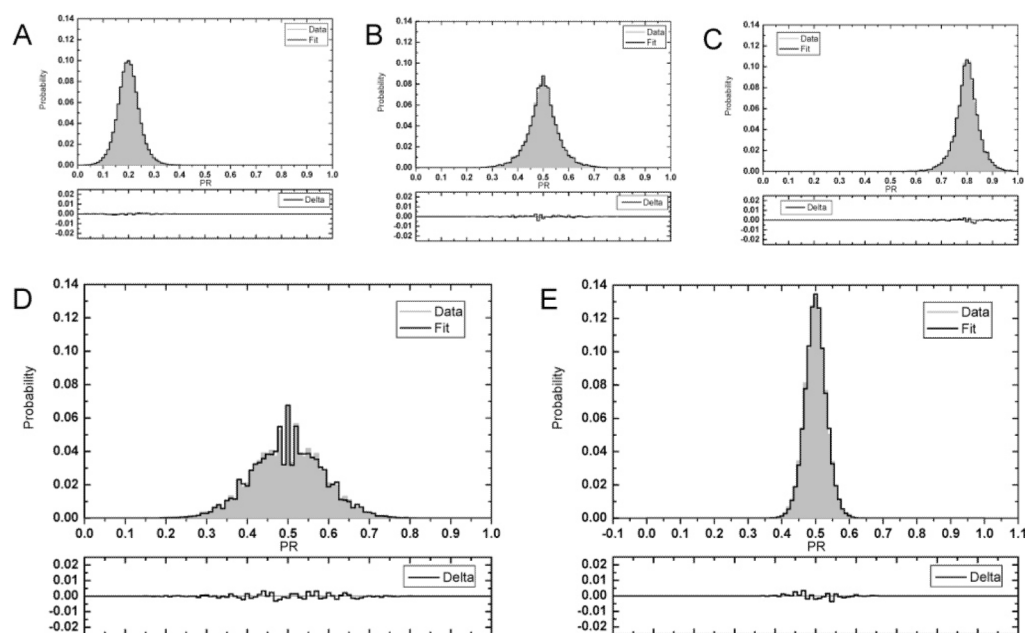


Figure 4.

Simulated PRH (gray box histogram) and best-fit shot-noise limited theoretical PRH (black curve) for different E values. E is the FRET efficiency used in the simulation, and ε is the best-fit parameter. A–C. PRH comparison for $E = \varepsilon = 0.2, 0.5,$ and 0.8 . Bursts were defined using an “all-photon” burst search algorithm with parameters $L = 50$, $M = 30$, and $T = 500 \mu\text{s}$. D,E. PRH for the same data as B, but using different burst size (S) subpopulations (D, $50 < S < 100$; E, $300 < S < 4000$), illustrating the perfect agreement between experimental and theoretical PRH. Histogram bin size: 0.01. Parameters of the simulation: $k_{bl} = k_{ISC} = 0$; $k_e^D = 6 \times 10^5 \text{ s}^{-1}$; $k_r^D = 10^9 \text{ s}^{-1}$; $k_e^A = 6 \times 10^5 \text{ s}^{-1}$; $k_r^A = 10^9 \text{ s}^{-1}$; $\eta_D = \eta_A = 1$ at the center of the confocal spot.

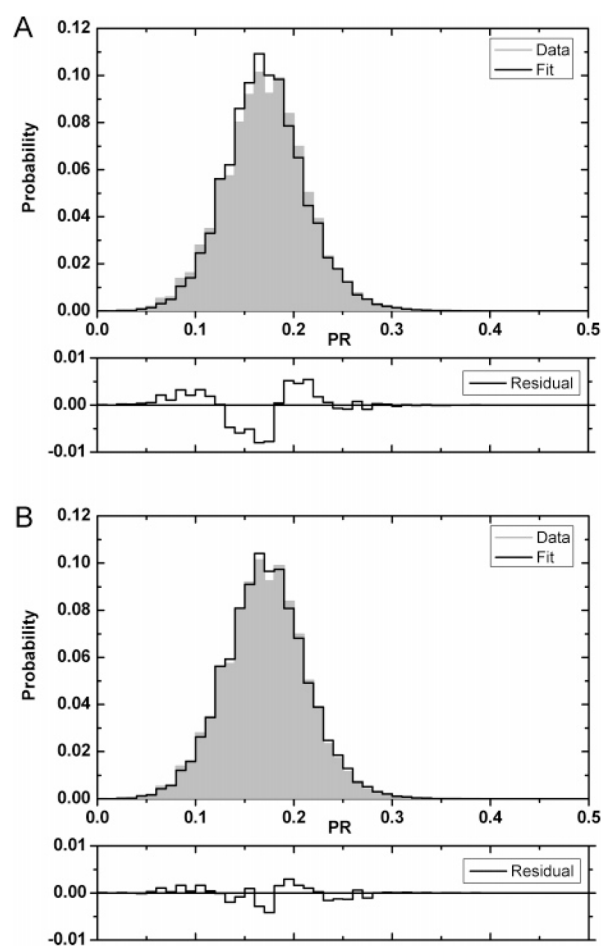


Figure 5.

Influence of background counts on the PRH. Simulated bursts corresponding to molecules with a FRET efficiency of 0.2 were contaminated with 6 kHz of background. A. Using eq 17, a reasonably good fit is obtained for a PR value $\varepsilon = 0.16$. The residual curve shows a slight discrepancy. B. Using the known background level in the donor channel and eq 36, a better fit is obtained, with a PR value $\varepsilon = 0.195$, almost indistinguishable from the exact FRET efficiency. The residual curve shows a much better agreement than in the previous case.

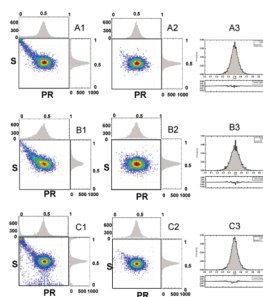


Figure 6.

Effects of bleaching, blinking, and coincident detection on the PRH. Two different burst search algorithms were used to analyze simulated data: the all-photon-burst-search (APBS) algorithm (left column) and the dual-channel-burst-search (DCBS) algorithm (middle and right columns). A. in the case of bleaching, the APBS algorithm (A1) reveals a trail of bursts connecting a donor-only population ($PR = 0, S = 1$) and the population of interest ($PR = S = 0.5$). The DCBS algorithm (A2) eliminates those bursts, resulting in a good fit of the remaining PRH (A3). B. In the case of blinking, the APBS algorithm (B1) reveals a trail of bursts directed toward the location of donor-only bursts and the population of interest. As in A, the DCBS algorithm (B2) eliminates those bursts, resulting in a good fit of the remaining PRH (B3). C. In the case of coincident detection, the APBS algorithm (C1) reveals a trail of bursts directed toward the location of donor-only bursts or the location of acceptor-only bursts ($S = 0$) and the population of interest. Again, the DCBS algorithm (C2) eliminates most of these bursts, resulting in a relatively good fit of the remaining PRH (C3), although an obvious discrepancy exists (the actual PRH is wider than the fitted shot-noise limited PRH).

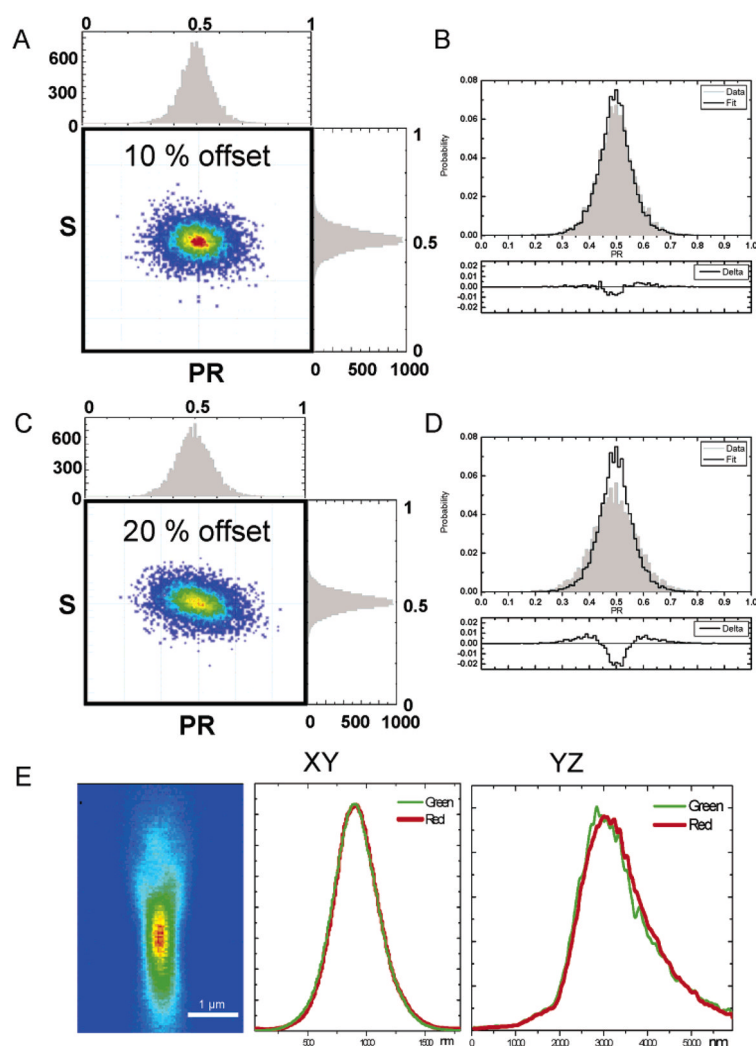


Figure 7. Effect of detection volume mismatch on the PRH. A,B. In the presence of a simulated 10% offset of the acceptor (Gaussian) detection volume with respect to the donor (Gaussian) detection volume, the PRH (B) resulting from the APBS algorithm (A) is reasonably well fitted by a shot-noise limited PRH, although the measured PRH is clearly wider than the fitted one. C,D. For a 20% offset, the PRH (D) resulting from the APBS algorithm (C) is not well fitted by a shot-noise limited PRH. E. In actual experiments, the measured offset of the donor and acceptor detection volume is smaller than 3%, as shown by the XY and YZ cross sections. The size of each pixel is 33.33 nm and the size of the scan is 5 μm on 7 μm.

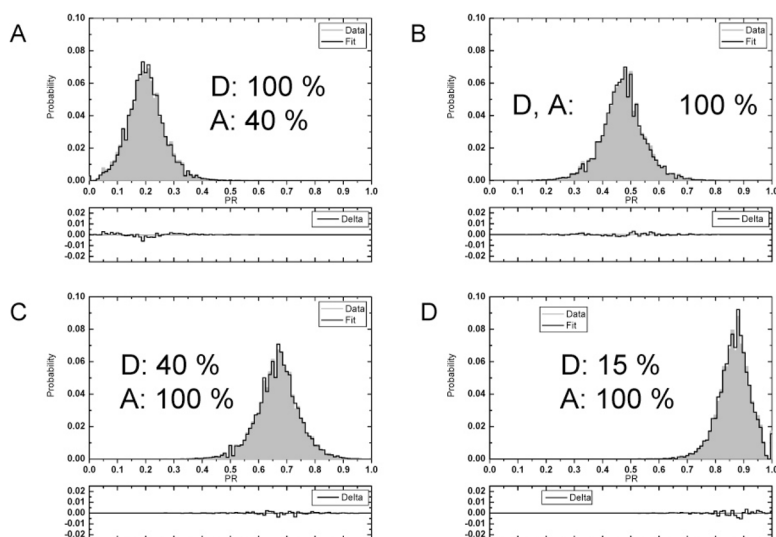


Figure 8. Artificial FRET data. Singly labeled DNA molecules were detected using a standard SMS confocal microscope in which the emitted signal was split equally by a nonpolarizing beam-splitter cube and sent to two SPAD's. Different neutral density filters were used in front of the detectors to modulate their detected signal ratio. A. Experimental PRH (gray box histogram) and the shot-noise limited fit (black curve) for 40% attenuation in front of the acceptor SPAD. B. Same for no attenuation. C. Same with 40% attenuation in front of the donor SPAD. D. Same with 15% attenuation in front of the donor SPAD.

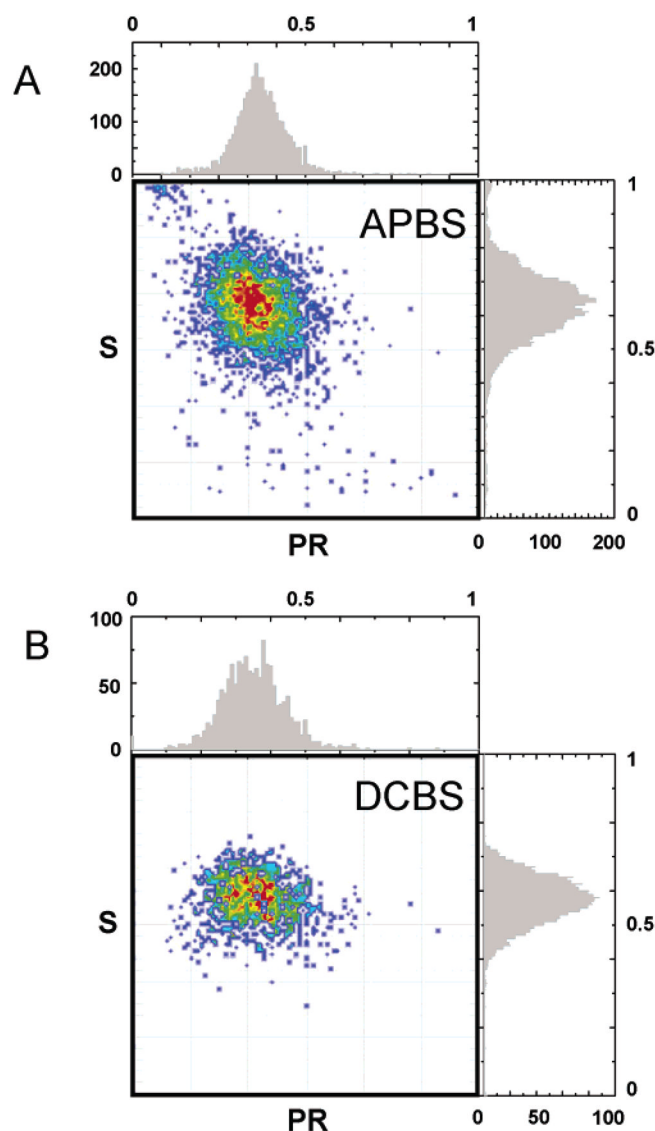


Figure 9. Comparison of APBS and DCBS algorithms for dsDNA data. A 50 pM dsDNA composed of Cy3B-labeled ssDNA molecules hybridized to complementary ATTO-647N-labeled ssDNA molecules was observed using a μ s-ALEX setup. A. Photon stream analyzed with an APBS algorithm ($L = 50$, $M = 30$, $T = 500 \mu\text{s}$), exhibiting a small donor-only ($PR \approx 0$, $S \approx 1$) and acceptor-only ($S \approx 0$) contamination. The corresponding PRH (upper graph) exhibits a small peak for $PR \approx 0$ and small tails in both directions. B. The analysis using a DCBS algorithm ($L = 25$, $M = 15$, $T = 500 \mu\text{s}$) suppresses this contamination and the corresponding artifacts on the PRH.

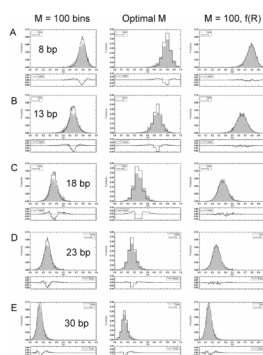
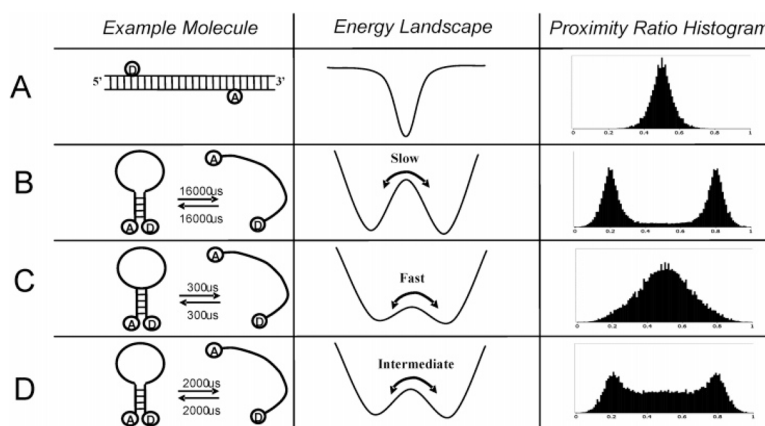


Figure 10.

Experimental PRH of doubly labeled dsDNA samples. Five different samples containing the same Cy3-labeled ssDNA top strand and a complementary ssDNA bottom strand labeled with ATTO-647N at different distance from the top strand label (measured in base pairs) were used. A–E. (left and center column) 8, 13, 18, 23, and 30 bp, respectively. Experimental PRH (gray bar histogram) obtained with a DCBS algorithm ($L = 25$, $M = 15$, $T = 500 \mu\text{s}$) were fitted with eq 17 (black curve). Histograms were binned with either a fixed number of bins $M = 100$ (left column) or the optimal number of bins calculated as described in Appendix B. A–E. (right column) Comparison of the experimental PRH and the shot-noise limited prediction in the presence of a Gaussian distribution of distance of width σ . Fitted values are $\sigma = 1.5$, 1.7, 1.8, 1.6, and 1.5 Å, respectively.

**Figure 11.**

Relation between energy landscape and PRH in a simple two-state system. Left: System studied, with characteristic time scales of opening and closing. Center: Schematic energy landscape; simulated shot-noise limited PRH. A. For a rigid dsDNA molecule, the unique energy minimum results in a distribution of distances, which, after convolution with the shot noise effect, yields a single-peak PRH. B. For hairpins characterized with opening and closing time scales much longer than the typical diffusion time (~ 1 ms), each molecule is virtually frozen in one of the two energy minima characteristic of the system. After convolution with the shot noise effect, the PRH exhibits two separate peaks. C. For hairpins with a low energy barrier between the two minima, rapid transitions between the two states are expected, and each molecule samples the two typical conformations several times during each burst, resulting in a broad peak located between the expected position of each characteristics conformation. D. For intermediate time constants comparable with the characteristic diffusion time of the molecule, the two separate peak locations are still visible, but one also expects that intermediate values will be observed. A combined analysis of the influence of shot noise and of a simple two-state kinetic scheme on the PRH allows extraction of the kinetic parameters of the system.

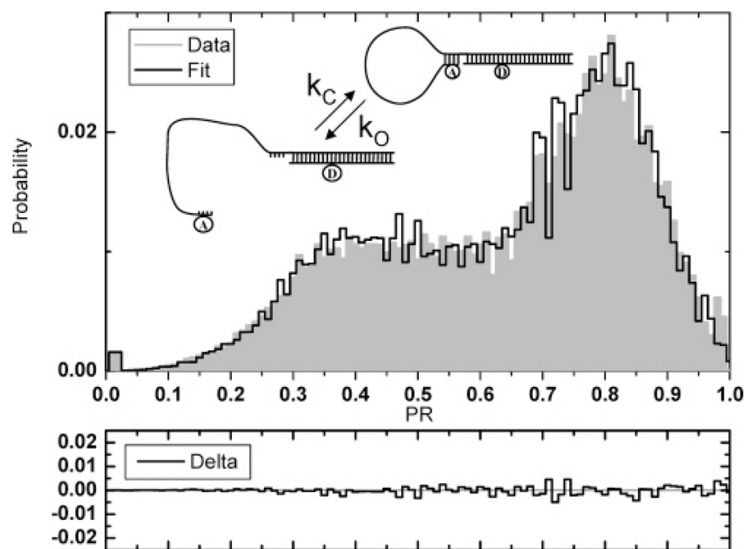


Figure 12. DNA hairpin and two-state model. Experimental PRH (gray columns) and two-state model, shot-noise limited PRH (black curve) calculated with 100 bins. The best fit parameters are as follows: closing rate $k_C = 1425 \text{ s}^{-1}$, opening rate $k_O = 835 \text{ s}^{-1}$, average closed PR $\varepsilon_C = 0.81$, average open PR $\varepsilon_O = 0.36$.