Biophysical Journal, Volume *114*

Supplemental Information

Increasing the Time Resolution of Single-Molecule Experiments with Bayesian Inference

Colin D. Kinz-Thompson and Ruben L. Gonzalez Jr.

Supporting Information: Increasing the time resolution of single-molecule experiments with Bayesian inference

Colin D. Kinz-Thompson and Ruben L. Gonzalez, Jr.

Department of Chemistry, Columbia University, New York, NY 10027, USA

1 Distributions of Fractional Occupancies

Two-state System

Consider a single-molecule whose dynamics are governed by the stochastic, two-state system,

$$1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} 2.$$

For a memoryless two-state system, the individual lifetimes that the single-molecule dwells in states 1 and 2 are exponentially distributed with rate constants k_1 and k_2, respectively. In this case, the equilibrium probabilities of finding states 1 and 2 occupied are,

$$\begin{aligned} p_1 &= \frac{k_2}{k_1 + k_2}, \text{ and} \\ p_2 = 1 - p_1 &= \frac{k_1}{k_2 + k_2}, \end{aligned} \tag{1}$$

respectively. However, during a period of observation, $t = 0$ to τ, a single-molecule will begin in one of the two states, and then will processively transition between the two states a random number of times, which is governed by k_1, k_2, and τ. Along these lines, during the observation period, the time spent in the ith state (*i.e.*, the sojourn time), T_i, is also governed k_1, k_2, and τ. For T_1 and $T_2 > 0$ sec, the probability distribution of T_1 is known to be (1, 2),

$$\begin{aligned} P^{\dagger}\left(T_1|k_1,k_2,\tau\right) = e^{-k_1T_1-k_2(\tau-T_1)} \cdot \Bigg[& \left(p_1k_1 + p_2k_2\right) \cdot I_0\left(2(k_1k_2T_1(\tau-T_1))^{1/2}\right) \\ & + \left(k_1k_2\right)^{1/2}\left(p_1\left(\frac{T_1}{\tau-T_1}\right)^{1/2} + p_2\left(\frac{\tau-T_1}{T_1}\right)^{1/2}\right) \cdot I_1\left(2(k_1k_2T_1(\tau-T_1))^{1/2}\right)\Bigg], \end{aligned} \tag{2}$$

where p_1 and p_2 are the probability of finding the single-molecule is states 1 and 2, respectively, (*i.e.,* at equilibrium, $p_1 = \frac{k_2}{k_1+k_2}$ and $p_2 = \frac{k_1}{k_1+k_2}$), and I_0 and I_1 are modified Bessel functions of the first kind. This distribution is not normalized because it lacks density at $T_1 = 0$ and $T_1 = \tau$. These conditions represent the situations when the single-molecule is exclusively in state 1 or in state state 2, respectively, for the entire duration of the observation period. *Ad hoc*, at equilibrium, the contributions to the probability distribution at these points should be the equilibrium population-weighted probability that the single-molecule does not undergo a transition out of the ith state during the observation period; this probability is the survival probability (*i.e.,* 1 - the cumulative distribution function) of the exponential distribution, which is $e^{-k_i\tau}$. Therefore, considering these contributions, the entire probability distribution of time spent in state 1 during the observation period is,

$$P\left(T_1|k_1,k_2,\tau\right) = P^{\dagger}\left(T_1|k_1,k_2,\tau\right) + p_1e^{-k_1\tau} \cdot \delta\left(\tau - T_1\right) + p_2e^{-k_2\tau} \cdot \delta\left(T_1\right), \tag{3}$$

where δ is the Dirac delta function, which ensures that these terms only contribute when the single-molecule is exclusively in one of the two states. The probability distribution of T_2 during an observation period can be obtained by interchanging the rate constants. If the observation period, τ, is known, then the probability distribution of the total time spent in state 1 during an observation period given in Eqn. (3), can be transformed into the probability distribution of the fraction of time spent in state 1 during the observation period, f. This transformation is,

$$P\left(f|k_1,k_2,\tau\right) = P(T_1 = f\cdot\tau|k_1,k_2,\tau)\cdot|J|^{-1}\text{; where } f \equiv \frac{T_1}{\tau}\text{, so } \frac{\partial f}{\partial T_1} = \frac{1}{\tau} \text{ and } |J|^{-1} = \tau, \tag{4}$$

where J is the Jacobian. Therefore, by plugging Equation (3) into Equation (4) and making the substitution $T_1 = f\tau$,

$$\begin{aligned} P\left(f|k_1,k_2,\tau\right) = {} & \frac{k_2}{k_1+k_2}\tau e^{-k_1\tau}\cdot\delta\left(\tau-(f\tau)\right) + \frac{k_1}{k_1+k_2}\tau e^{-k_2\tau}\cdot\delta\left((f\tau)\right) \\ & + 2\frac{k_1k_2}{k_1+k_2}\tau e^{-k_1(f\tau)-k_2(\tau-(f\tau))}\cdot\left[I_0\left(y\right) + \frac{k_2\left(f\tau\right)+k_1\left(\tau-(f\tau)\right)}{y}\cdot I_1\left(y\right)\right], \end{aligned} \tag{5}$$

where $y \equiv 2\sqrt{k_1k_2\left(f\tau\right)\left(\tau-(f\tau)\right)} = 2\tau\sqrt{k_1k_2f\left(1-f\right)}$.
Noting the identity,

$$\delta\left(ax\right) = \frac{1}{|a|}\cdot\delta\left(x\right), \tag{6}$$

we can simplify Equation (5) to yield,

$$\begin{aligned} P\left(f|k_1,k_2,\tau\right) = {} & \frac{k_2}{k_1+k_2}e^{-k_1\tau}\cdot\delta\left(1-f\right) + \frac{k_1}{k_1+k_2}e^{-k_2\tau}\cdot\delta\left(f\right) \\ & + 2\frac{k_1k_2}{k_1+k_2}\tau e^{-(k_1f+k_2(1-f))\tau}\cdot\left[I_0\left(y\right) + \frac{\left(k_2f+k_1(1-f)\right)\tau}{y}\cdot I_1\left(y\right)\right]. \end{aligned} \tag{7}$$

Above, we have chosen to simplify Equation (7) so that it is apparent that it is equivalent to the expression of Berezhkovskii and coworkers, and that used by Gopich and coworkers (3, 4).

2 Bayesian Inference

Overview

Bayesian inference allows for the parameters describing an initial hypothesis, θ, to be modified to account for new data, D. Mathematically, this process can be written using Bayes' rule,

$$p(\theta|D) = \frac{p(D|\theta)\cdot p(\theta)}{p\left(D\right)} = \frac{p(D|\theta)\cdot p(\theta)}{\sum\left(p(D|\theta)\cdot p(\theta)\right)} \propto p(D|\theta)\cdot p(\theta) \tag{8}$$

which is analogous to saying that the probability of the hypothesis after having seen the data (the posterior probability, $p(\theta|D)$) is proportional to the product of the probability of the data given the hypothesis (the likelihood, $p(D|\theta)$) and the initial probability of the hypothesis itself (the prior probability, $p(\theta)$). For an introduction to Bayesian inference, see Ref. (5). According to Eqn. (8), one can use Bayes' rule in order to calculate the probability of a hypothesis describing a biomolecular system after obtaining a series of single-molecule observations if one has a model for the data (*i.e.*, an expression for likelihood). Unfortunately, direct enumeration of the posterior distribution (*i.e.*, all possible hypotheses) can be computationally expensive, especially when there are large number of parameters in the model for the data. As an alternative, methods such as Markov chain Monte Carlo (MCMC) can be used to explore the expansive hyper-volume of the posterior distribution in a much more economical manner. Sufficient sampling of the posterior probability distribution then provides random samples, which can be used to statistically describe the posterior probability distribution (see below).

Likelihood Function

A general model for the signal originating from a single-molecule is that the signal from each state is distributed according to a normal distribution. In such a situation, if the single-molecule occupies multiple states during a signal acquisition period, that datapoint will distributed according to a convolution of the corresponding normal distributions for the different states; such a convolution of normal distributions is also a normal distribution. Additional detection noise or baseline contributions can be accounted for by additional convolutions. If every distribution in the convolution is a normal distribution, then the resulting distribution is also a normal distribution with a mean, μ, and variance, σ^2, that are linear-combinations of the means and variances weighted by the time spent in each particular state. For instance, in a two-state system, if a molecule is in state 1 for the entire i$^{\text{th}}$ period of observation, the observed emission signal, d_i, would be centered at ϵ_1 with variance σ_1^2. Similarly, emission from only state 2 would be centered at ϵ_2 with variance σ_2^2. For an observation period with a fractional occupation of these states, the signal will be centered at the linear combination $\epsilon_1 f + \epsilon_2 (1-f)$, and the variance will be $\sigma_1^2 f + \sigma_2^2 (1-f)$. For this two-state system,

$$P\left(d_i|\epsilon_1,\epsilon_2,\sigma_1,\sigma_2,f\right) = \frac{1}{\sqrt{2\pi\left(\sigma_1^2 f + \sigma_2^2(1-f)\right)}} \cdot e^{\frac{-1}{2}\frac{\left(d_i-(\epsilon_1 f+\epsilon_2(1-f))\right)^2}{\left(\sigma_1^2 f+\sigma_2^2(1-f)\right)}}. \tag{9}$$

Unfortunately, the exact fractional occupation during the observation period, f, is not a parameter that is observed during an experiment; so, Equation (9) is not useful for inferring model parameters. However, one can derive the probability distribution of the fractional occupation; the exact form for the memoryless two-state model, $P(f|k_1,k_2,\tau)$, was given in Section 1. Knowing this probability distribution allows the dependence upon f to be removed from Equation (9); this process is known as marginalization (Fig. S1A). For the two-state system, the resulting, marginalized probability distribution using the analytical expression for the probability distribution given in Equation (7) is,

$$\begin{aligned} P\left(d_i \mid \epsilon_1,\epsilon_2,\sigma_1,\sigma_2,k_1,k_2,\tau\right) &= \int_0^1 df \cdot P\left(d_i|\epsilon_1,\epsilon_2,\sigma_1,\sigma_2,f\right) \cdot P\left(f|k_1,k_2,\tau\right) \\ &= \frac{1}{\sqrt{2\pi}} \cdot \left[\frac{1}{\sigma_1}\frac{k_2}{k_1+k_2} e^{-\frac{1}{2}\left(\frac{d_i-\epsilon_1}{\sigma_1}\right)^2 - k_1\tau} + \frac{1}{\sigma_2}\frac{k_1}{k_1+k_2} e^{-\frac{1}{2}\left(\frac{d_i-\epsilon_2}{\sigma_2}\right)^2 - k_2\tau} \right. \\ &\quad \left. + \frac{2k_1k_2\tau}{k_1+k_2}\int_0^1 df \cdot \frac{1}{\sqrt{\sigma_1^2 f+\sigma_2^2(1-f)}} e^{-\frac{1}{2}\frac{\left(d_i-(\epsilon_1 f+\epsilon_2(1-f))\right)^2}{\left(\sigma_1^2 f+\sigma_2^2(1-f)\right)} - (k_1 f + k_2(1-f))\tau} \cdot \left(I_0(y) + \frac{(k_2 f + k_1(1-f))\tau}{y} \cdot I_1(y) \right) \right] \end{aligned} \tag{10}$$

In this form, the integral can be computed numerically, for instance, with a Gaussian quadrature method. Having experimentally observed a set of signals from a single-molecule, $\{d\} = \{d_1, \cdots, d_N\}$, we can calculate the likelihood of observing this particular set, $L \equiv P\left(D|\Theta\right)$; this expression is the product of the probability of the observed signal during each measurement period, $P\left(d_i|\epsilon_1,\epsilon_2,\sigma_1,\sigma_2,k_1,k_2,\tau\right)$, which is

$$L = P\left(\{d\}|\Theta\right) = P\left(\{d\}|\{\epsilon\},\{\sigma\},\{k\},\tau\right) = \prod_{i=1}^{N} P\left(d_i|\{\epsilon\},\{\sigma\},\{k\},\tau\right), \tag{11}$$

where $\{\epsilon\}$, $\{\sigma\}$, and $\{k\}$ are the sets of emission means, emission standard deviations, and rate constants defining the single-molecule system (*e.g.*, for a two-state system $\{\epsilon\} = \{\epsilon_1,\epsilon_2\}$, $\{\sigma\} = \{\sigma_1,\sigma_2\}$, and $\{k\} = \{k_1,k_2\}$).

This assumes that the datapoints are independent and identically distributed, and that the system is at equilibrium. That assumption is reasonable for the case of fast dynamics (*i.e.*, $k_1\tau$ or $k_2\tau > 1$), where subsequent signal measurements are made during a time when the single-molecule occupies several different states. For the case of slow dynamics, if there are a sufficient number of measurements in the signal *versus* time trajectory, then the single-molecule can be considered at equilibrium (*i.e.*, $\langle f \rangle \approx k_2/(k_1+k_2)$). Thus, we believe that it is a reasonable approximation. As we show in the main text, the likelihood function performs well in this regime under this assumption.

Finally, given a large number of measurements in $\{d\}$, L can easily underflow on a computer; thus, it is better computed on a

log scale. The resulting log-likelihood, $\ln(L)$, is

$$\ln(L) = \ln\left(P\left(\{d\}|\Theta\right)\right) = \sum_{i}^{N} \ln\left(P\left(d_i|\{\epsilon\},\{\sigma\},\{k\},\tau\right)\right). \tag{12}$$

Prior Distributions

Prior probability distributions are subjective (*i.e.*, conditional), but not unsystematic. Choosing a prior probability distribution is equivalent to systematically encoding knowledge of initial conditions. For instance, if a particular parameter describes a magnitude, then it must be a greater than or equal to zero. Objectivity is ingrained in this process since, given the same initial conditioning, two rational individuals produce the same prior probability distribution (5).

Certain parameters are more conveniently represented by certain probability distributions. Just considering the allowed values of particular parameters and the support of particular distributions, probabilities and ratios are conveniently represented with beta distributions, magnitudes and rates are conveniently represented with gamma distributions, and positions and signals are conveniently represented by normal distributions. These distributions are:

$$\text{Beta, } P(x|\alpha,\beta) = \frac{x^{\alpha-1}\cdot(1-x)^{\beta-1}}{B(\alpha,\beta)}, \text{ for } 0 \leq x \leq 1, \tag{13}$$

$$\text{Gamma, } P(x|\alpha,\beta) = \frac{\beta^{\alpha}\cdot x^{\alpha-1}\cdot e^{-\beta x}}{\Gamma(\alpha)}, \text{ for } x \geq 0, \tag{14}$$

$$\text{Normal, } P(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \text{ and} \tag{15}$$

$$\text{Uniform, } P(x|\alpha,\beta) = \frac{1}{\beta-\alpha}, \text{ for } \alpha \leq x \leq \beta. \tag{16}$$

When using BIASD to analyze single-molecule fluorescence resonance energy transfer (smFRET) experiments, we typically employ beta distributions for the ϵ, because they are ratios of distances or photon counts. We often use gamma distributions for the σ, because it represents the magnitude of the noise. Finally, for the k, we often also use gamma distributions, because these rate constants represent a number per time and therefore cannot be negative.

Markov Chain Monte Carlo Sampling

Markov chain Monte Carlo (MCMC) allows for the efficient sampling of high dimensional space (such as a posterior probability distribution) by taking a random walk along the distribution (see Ref. (6) for an introduction). Briefly, random steps are drawn from a proposal distribution, and these steps are either accepted or rejected based upon an acceptance criteria. The sequence of steps forms a Markov chain that is ideally able to traverse the relevant (*i.e.*, higher probability) hyper-volume of the N-dimensional space; thus eliminating the need to enumerate the entire N-dimensional space. To do so effectively, the acceptance rate of steps must not be too large, because then the Markov chain will not diffuse far from its initial location. Likewise, the acceptance rate must not be too small, because then the Markov chain will not diffuse along the probability distribution at all. One method to facilitate this exploration is to choose a proposal step based upon the positions of other Markov chains, which presumably are already located in relevant regions of the N-dimensional space. Such approaches amount to simultaneously moving an ensemble of MCMC chains (called 'walkers'), and carefully choosing proposal moves based upon the positions of these walkers that maintain detailed balance (7, 8).

3 Analysis using BIASD

Synthetic Titration

To begin our analysis, prior probability distributions, as described in Section 2, were chosen for the five parameters in the model, ϵ_1, ϵ_2, $\sigma \equiv \sigma_1 = \sigma_2$, k_1, and k_2. For the synthetic titration of a ligand to its receptor, the prior probability distributions for the BIASD

parameters were chosen to be

$$\begin{aligned}
\epsilon_1 &\sim \text{Normal}\left(\mu_{\epsilon_1} = 0.0, \sigma_{\epsilon_1} = 0.01\right) \\
\epsilon_2 &\sim \text{Normal}\left(\mu_{\epsilon_2} = 1.0, \sigma_{\epsilon_2} = 0.01\right) \\
\sigma &\sim \text{Gamma}\left(\alpha_\sigma = \zeta, \beta_\sigma = \zeta/0.04\right) \\
k_1 &\sim \text{Gamma}\left(\alpha_{k_1} = \zeta, \beta_{k_1} = \zeta/\left(10.0 \cdot [L]\right)\right) \\
k_2 &\sim \text{Gamma}\left(\alpha_{k_2} = \zeta, \beta_{k_2} = \zeta/10.0\right),
\end{aligned}$$

where ζ was chosen to be 2.0 to create broad prior distributions. The relatively narrower standard deviations for the prior probability distributions of ϵ_1 and ϵ_2 are justified by inspecting the signal trajectories at the lowest and highest $[L]$. In an experimental situation, this could also be learned from previous experiments performed on mutant receptors, with ligand analogues, or in the presence of a drug that stabilizes state 1 or state 2. Similarly, the parameters used to describe the prior probability distribution of σ could have been previously characterized for a particular instrument (*e.g.*, a particular microscope). The prior probability distributions for σ, k_1, and k_2 were chosen to be Gamma-distributed because of they are supported between 0 and ∞. We note that all of these prior probability distributions were constructed such that the mean of the distribution is centered at the parameter value used in simulated the signal trajectory that was analyzed. Finally, the BIASD posterior probability distributions for simulated signal trajectories were sampled using ensemble, affine-invariant MCMC as described in the Materials and Methods (7, 8). The posterior probability distributions for k_1 and k_2, as plotted in Fig. 2 in the main text, are summarized below in Table 3 containing the rate constant used in the simulation (truth), the marginalized mean of the MCMC samples (inferred mean), and the 95% credible intervals (95% C.I.) of the MCMC samples.

	k_1 (s^{-1})			k_2 (s^{-1})		
Entry	**Truth**	**Inferred Mean**	**95% C.I.**	**Truth**	**Inferred Mean**	**95% C.I.**
1	1.00×10^{-2}	1.21×10^{-2}	$(2.93 \times 10^{-03}, 2.84 \times 10^{-2})$	$1.00 \times 10^{+1}$	$8.24 \times 10^{+0}$	$(1.95 \times 10^{+0}, 2.02 \times 10^{+1})$
2	1.58×10^{-2}	2.15×10^{-2}	$(6.39 \times 10^{-03}, 4.61 \times 10^{-2})$	$1.00 \times 10^{+1}$	$1.03 \times 10^{+1}$	$(3.27 \times 10^{+0}, 2.29 \times 10^{+1})$
3	2.51×10^{-2}	3.54×10^{-2}	$(1.23 \times 10^{-2}, 7.16 \times 10^{-2})$	$1.00 \times 10^{+1}$	$1.27 \times 10^{+1}$	$(4.37 \times 10^{+0}, 2.58 \times 10^{+1})$
4	3.98×10^{-2}	4.37×10^{-2}	$(1.40 \times 10^{-2}, 8.94 \times 10^{-2})$	$1.00 \times 10^{+1}$	$1.03 \times 10^{+1}$	$(3.65 \times 10^{+0}, 2.08 \times 10^{+1})$
5	6.31×10^{-2}	8.05×10^{-2}	$(2.27 \times 10^{-2}, 1.84 \times 10^{-1})$	$1.00 \times 10^{+1}$	$1.21 \times 10^{+1}$	$(4.79 \times 10^{+0}, 2.26 \times 10^{+1})$
6	1.00×10^{-1}	1.10×10^{-1}	$(2.90 \times 10^{-2}, 2.41 \times 10^{-1})$	$1.00 \times 10^{+1}$	$1.46 \times 10^{+1}$	$(6.31 \times 10^{+0}, 2.75 \times 10^{+1})$
7	1.58×10^{-1}	1.22×10^{-1}	$(3.98 \times 10^{-2}, 2.48 \times 10^{-1})$	$1.00 \times 10^{+1}$	$1.51 \times 10^{+1}$	$(6.67 \times 10^{+0}, 2.82 \times 10^{+1})$
8	2.51×10^{-1}	1.95×10^{-1}	$(8.53 \times 10^{-2}, 3.63 \times 10^{-1})$	$1.00 \times 10^{+1}$	$1.51 \times 10^{+1}$	$(7.50 \times 10^{+0}, 2.52 \times 10^{+1})$
9	3.98×10^{-1}	3.01×10^{-1}	$(1.72 \times 10^{-1}, 4.79 \times 10^{-1})$	$1.00 \times 10^{+1}$	$1.24 \times 10^{+1}$	$(6.00 \times 10^{+0}, 2.24 \times 10^{+1})$
10	6.31×10^{-1}	4.93×10^{-1}	$(2.66 \times 10^{-1}, 8.26 \times 10^{-1})$	$1.00 \times 10^{+1}$	$1.11 \times 10^{+1}$	$(7.31 \times 10^{+0}, 1.58 \times 10^{+1})$
11	$1.00 \times 10^{+0}$	9.47×10^{-1}	$(6.81 \times 10^{-1}, 1.27 \times 10^{+0})$	$1.00 \times 10^{+1}$	$1.32 \times 10^{+1}$	$(9.37 \times 10^{+0}, 1.75 \times 10^{+1})$
12	$1.58 \times 10^{+0}$	$1.46 \times 10^{+0}$	$(1.12 \times 10^{+0}, 1.85 \times 10^{+0})$	$1.00 \times 10^{+1}$	$1.16 \times 10^{+1}$	$(8.69 \times 10^{+0}, 1.45 \times 10^{+1})$
13	$2.51 \times 10^{+0}$	$2.22 \times 10^{+0}$	$(1.74 \times 10^{+0}, 2.76 \times 10^{+0})$	$1.00 \times 10^{+1}$	$1.05 \times 10^{+1}$	$(8.17 \times 10^{+0}, 1.29 \times 10^{+1})$
14	$3.98 \times 10^{+0}$	$4.05 \times 10^{+0}$	$(3.34 \times 10^{+0}, 4.81 \times 10^{+0})$	$1.00 \times 10^{+1}$	$1.18 \times 10^{+1}$	$(9.69 \times 10^{+0}, 1.40 \times 10^{+1})$
15	$6.31 \times 10^{+0}$	$6.86 \times 10^{+0}$	$(5.80 \times 10^{+0}, 8.00 \times 10^{+0})$	$1.00 \times 10^{+1}$	$9.58 \times 10^{+0}$	$(8.05 \times 10^{+0}, 1.12 \times 10^{+1})$
16	$1.00 \times 10^{+1}$	$9.17 \times 10^{+0}$	$(7.81 \times 10^{+0}, 1.06 \times 10^{+1})$	$1.00 \times 10^{+1}$	$1.00 \times 10^{+1}$	$(8.37 \times 10^{+0}, 1.16 \times 10^{+1})$
17	$1.58 \times 10^{+1}$	$1.74 \times 10^{+1}$	$(1.49 \times 10^{+1}, 2.00 \times 10^{+1})$	$1.00 \times 10^{+1}$	$1.02 \times 10^{+1}$	$(8.66 \times 10^{+0}, 1.17 \times 10^{+1})$
18	$2.51 \times 10^{+1}$	$2.59 \times 10^{+1}$	$(2.17 \times 10^{+1}, 3.01 \times 10^{+1})$	$1.00 \times 10^{+1}$	$9.37 \times 10^{+0}$	$(7.96 \times 10^{+0}, 1.09 \times 10^{+1})$
19	$3.98 \times 10^{+1}$	$4.09 \times 10^{+1}$	$(3.43 \times 10^{+1}, 4.77 \times 10^{+1})$	$1.00 \times 10^{+1}$	$9.38 \times 10^{+0}$	$(7.84 \times 10^{+0}, 1.11 \times 10^{+1})$
20	$6.31 \times 10^{+1}$	$7.00 \times 10^{+1}$	$(5.83 \times 10^{+1}, 8.26 \times 10^{+1})$	$1.00 \times 10^{+1}$	$1.06 \times 10^{+1}$	$(8.49 \times 10^{+0}, 1.31 \times 10^{+1})$
21	$1.00 \times 10^{+2}$	$1.01 \times 10^{+2}$	$(8.14 \times 10^{+1}, 1.22 \times 10^{+2})$	$1.00 \times 10^{+1}$	$9.08 \times 10^{+0}$	$(6.68 \times 10^{+0}, 1.18 \times 10^{+1})$
22	$1.58 \times 10^{+2}$	$1.50 \times 10^{+2}$	$(1.16 \times 10^{+2}, 1.87 \times 10^{+2})$	$1.00 \times 10^{+1}$	$8.44 \times 10^{+0}$	$(5.62 \times 10^{+0}, 1.19 \times 10^{+1})$
23	$2.51 \times 10^{+2}$	$2.67 \times 10^{+2}$	$(1.72 \times 10^{+2}, 3.69 \times 10^{+2})$	$1.00 \times 10^{+1}$	$8.67 \times 10^{+0}$	$(3.62 \times 10^{+0}, 1.60 \times 10^{+1})$
24	$3.98 \times 10^{+2}$	$4.67 \times 10^{+2}$	$(2.61 \times 10^{+2}, 6.77 \times 10^{+2})$	$1.00 \times 10^{+1}$	$8.46 \times 10^{+0}$	$(1.68 \times 10^{+0}, 1.89 \times 10^{+1})$
25	$6.31 \times 10^{+2}$	$8.65 \times 10^{+2}$	$(3.85 \times 10^{+2}, 1.73 \times 10^{+3})$	$1.00 \times 10^{+1}$	$8.63 \times 10^{+0}$	$(1.37 \times 10^{+0}, 2.08 \times 10^{+1})$
26	$1.00 \times 10^{+3}$	$1.34 \times 10^{+3}$	$(5.45 \times 10^{+2}, 2.96 \times 10^{+3})$	$1.00 \times 10^{+1}$	$8.87 \times 10^{+0}$	$(1.27 \times 10^{+0}, 2.34 \times 10^{+1})$
27	$1.58 \times 10^{+3}$	$1.83 \times 10^{+3}$	$(6.42 \times 10^{+2}, 4.44 \times 10^{+3})$	$1.00 \times 10^{+1}$	$9.64 \times 10^{+0}$	$(1.31 \times 10^{+0}, 2.54 \times 10^{+1})$
28	$2.51 \times 10^{+3}$	$2.91 \times 10^{+3}$	$(8.11 \times 10^{+2}, 7.32 \times 10^{+3})$	$1.00 \times 10^{+1}$	$9.03 \times 10^{+0}$	$(1.05 \times 10^{+0}, 2.45 \times 10^{+1})$
29	$3.98 \times 10^{+3}$	$4.08 \times 10^{+3}$	$(8.03 \times 10^{+2}, 1.12 \times 10^{+4})$	$1.00 \times 10^{+1}$	$9.36 \times 10^{+0}$	$(1.36 \times 10^{+0}, 2.48 \times 10^{+1})$
30	$6.31 \times 10^{+3}$	$6.36 \times 10^{+3}$	$(8.73 \times 10^{+2}, 1.76 \times 10^{+4})$	$1.00 \times 10^{+1}$	$9.33 \times 10^{+0}$	$(1.25 \times 10^{+0}, 2.64 \times 10^{+1})$
31	$1.00 \times 10^{+4}$	$9.94 \times 10^{+3}$	$(1.43 \times 10^{+3}, 2.86 \times 10^{+4})$	$1.00 \times 10^{+1}$	$9.57 \times 10^{+0}$	$(1.18 \times 10^{+0}, 2.63 \times 10^{+1})$

Table 1: Summary of Rate Constants Inferred Using BIASD for Synthetic Titration Simulation

Temperature-dependent Pre-translocation Complexes

The temperature-dependent datasets of smFRET trajectories from Wang and coworkers were truncating at the first photophysical anomaly, such as photobleaching, in order to ensure two-state behavior (Figure S2) (9). Next, all of the data points in the datasets of smFRET trajectories were assumed to be independent and identically distributed (*i.e.*, there is only one type of complex in the ensembles). As such, we computed the joint likelihood of all of the data points from the smFRET trajectories at a particular temperature to perform Bayesian inference.

For the prior probability distributions for ϵ_{GS1}, ϵ_{GS2}, $\sigma \equiv \sigma_{GS1} = \sigma_{GS2}$, k_{GS1}, and k_{GS2}., we used uniform distributions, because we assumed no previous analysis of these experiments. Specifically,

$$\begin{aligned}
\epsilon_{GS1} &\sim \text{Uniform}\,(\alpha = -0.5, \beta = 0.5) \\
\epsilon_{GS2} &\sim \text{Uniform}\,(\alpha = 0.5, \beta = 1.5) \\
\sigma &\sim \text{Uniform}\,(\alpha = 0.02, \beta = 0.20) \\
k_{GS1} &\sim \text{Uniform}\,(\alpha = 0.01, \beta = 100) \\
k_{GS2} &\sim \text{Uniform}\,(\alpha = 0.01, \beta = 100)\,.
\end{aligned}$$

The limits of these uniform distributions were chosen due to experimental considerations. For the ϵ, the FRET efficiency is limited between 0 and 1, but can exceed these limits due to background corrections; therefore, we chose to limit the uniform distributions between -0.5 and 1.5, while enforcing $\epsilon_{GS1} < \epsilon_{GS2}$. The limits chosen for σ reflect maximum signal to noise ratios of $1/0.2 = 5$ to $1/0.02 = 50$, which is reasonable for our total internal reflection fluorescence microscope. Finally, the limits for the k reflect the length of the experiment on one end, (60 s)$^{-1}$ =0.016 s^{-1}, and on the other end, five times faster than the acquisition rate, $\tau^{-1} = 20$ s^{-1}, which is reasonable upon inspection of individual trajectories.

Hierarchical Mixtures of and Transitions between Sub-populations

In order to consider multiple different sub-populations with BIASD, we will show how to combine BIASD with a mixture model and an HMM. Consider denoting that the i$^{\text{th}}$ data point from the j$^{\text{th}}$ single-molecule trajectory belongs to the m$^{\text{th}}$ classes of K different classes by using an indicator variable, z_{ijm}, which is one if the data point belongs to the assigned class, and zero otherwise. We can then write the likelihood of observing this data point using the BIASD model, but conditioned upon z_{ijm} (*i.e.*, belonging to the m$^{\text{th}}$ class). This would be the same expression in Eqn. (10), but written to denote the conditional probability as

$$P(d_{ij}|\Theta_m, \tau, z_{ijm}) = P(d_{ij}|\Theta_m, \tau) \cdot z_{ijm}, \tag{17}$$

where Θ_m denotes the BIASD model parameters $\Theta = \{\epsilon_1, \epsilon_2, \sigma, k_1, k_2\}$ for the m$^{\text{th}}$ sub-population. Note that z_{ijm} functions like a delta function in this case. Given a multitude of classes, we can write the probability of d_{ij} conditioned upon all of the possible K subsystems as,

$$P(d_{ij}|\vec{\Theta}, \tau, \vec{z}_{ij}) = \sum_{m=1}^{K} P(d_{ij}|\Theta_m, \tau, z_{ijm}), \tag{18}$$

where $\vec{z}_{ij} = \{z_{ij1}, \ldots, z_{ijK}\}$ and $\vec{\Theta} = \{\Theta_1, \ldots, \Theta_m\}$ and it is important to note that only one of the z_{ijm} will be one, because d_{ij} belongs to only one class.

As mentioned in the main text, inferring the values of $\vec{z_{ij}}$ simultaneously with the values of the model parameters $\vec{\Theta}$ is difficult, but we also do not particularly care about the values of $\vec{z_{ij}}$. Therefore, we will marginalize out the $\vec{z_{ij}}$. To perform this marginalization, we need an expression for the probability of z_{ijm}, $P(z_{ijm})$. In the case of a mixture model (*e.g.*, without any transitions between data points), we will use π_m as this probability, and enforce that $\sum_{m=1}^{K} \pi_m = 1$. Note that this means in the set of π_m, $\vec{\pi} = \{\pi_1, \ldots, \pi_K\} = \{\pi_1, \ldots, \pi_{K-1}, 1 - \sum_{m=1}^{K-1}\}$, so that there are only $K - 1$ independent parameters. Now we can write the

marginalized probability as

$$P(d_{ij}|\vec{\Theta},\tau,\vec{\pi}) = \sum_{m=1}^{K} P(d_{ij}|\Theta_m,\tau,z_{ijm}) \cdot P(z_{ijm}|\pi_m), \tag{19}$$

and additionally note that for the case of a mixture model, the probability of all of the data points from the j[th] molecule is

$$P(d_j|\vec{\Theta},\tau,\vec{\pi}) = \prod_i P(d_{ij}|\vec{\Theta},\tau,\vec{\pi}). \tag{20}$$

In the case of an HMM (Fig. S1B), this marginalization process is more involved, since the $P(z_{ijm})$ depends upon $z_{i-1,jm}$, but this can be performed with the forward-backward algorithm (6, 10). Briefly, given a rate matrix $\tilde{\mathbf{Q}}$, where rate constants for transitions between states are on the off-diagonals, and the diagonals are the negative sum of the rows in order to conserve mass, a transition probability matrix, $\tilde{\mathbf{A}}_{\mathbf{ij}}$, from the i-1[th] to the i[th] data point from the j[th] single-molecule signal trajectory can be calculated as

$$\tilde{\mathbf{A}}_{\mathbf{ij}} = \exp\left(\tilde{\mathbf{Q}} \cdot \mathbf{t}_{\mathbf{ij}}\right), \tag{21}$$

where t_{ij} is the amount of time that has passed between the i-1[th] and the i[th] data point. Note t_{ij} is not necessarily τ, nor do all of the t_{ij} have to be the same. This transition matrix provides the $P(z_{ijm}|z_{i-1,jm})$ necessary to marginalize out the $\vec{z}_{ij}$. As a result, the probability distribution of a signal trajectory can be written as

$$P(d_{1j}\ldots d_{Nj}|\vec{\Theta},\tau,\tilde{\mathbf{Q}},\{t_{1j}\ldots,t_{Nj}\}) = \prod_i \sum_{m=1}^{K} P(d_{ij}|\Theta_m,\tau,z_{ijm}) \cdot P(z_{ijm}|z_{i-1,jm},\tilde{\mathbf{A}}_{ij}), \tag{22}$$

where $P(z_{0jm})$ is the steady-state probability of state m, which can be obtained by solving differential equations, by using the diagram method, or from Eqn. (21) when $t_{ij} \to \infty$. Again, this marginalization can be efficiently calculated by using the forward-backward algorithm (6, 10).

To illustrate the ability to couple BIASD with an HMM, we simulated the five-state system in Fig. 4A of the main text. States 1 and 2 had signal values of ϵ_1 = 1000 (a.u.), and ϵ_2 = 500 (a.u.), respectively, while all other states had a signal value of 0. Since, these states were intended to correspond to a photoblinked, or photobleached fluorophore, we did not infer these values, but fixed them at 0 in the likelihood expression. We set $\sigma \equiv \sigma_0 = \sigma_1 = \sigma_2$ = 10 (a.u.), and simulated the system for 2000 data points with τ = 0.1 s, using the rate constants given in Table 2.

Initial State	Final State	Initial Fluorophore	Final Fluorophore	$\langle \mathbf{t} \rangle = \mathbf{k^{-1}}$
1	2	on	on	5 sec
1	2	off	off	5 sec
2	1	on	on	15 sec
2	1	of	off	15 sec
1	1	on	off	50 msec
1	1	off	on	15 msec
2	2	on	off	40 msec
2	2	off	on	20 msec
1	1	on	bleach	800 sec
2	2	on	bleach	800 sec

Table 2: Parameters Used in Photoblinking Trajectory Simulation.

We then created prior probability distributions where σ, and all of the rate constants were distributed according to the Gamma distribution with a mean of the simulation parameters and $\alpha = 100.0$, while ϵ_1, and ϵ_2 were distributed according to the Normal distribution with a mean of the simulation parameters and a standard deviation of 10.0. The resulting posterior probability distribution was sampled using MCMC, and is shown in Figure S3.

4 Dependence of BIASD Performance on Parameter Values

To investigate how BIASD performs as the underlying molecular parameters and several important experimental parameters vary through the range of possible values that they could take, we have analyzed the analytical equations that were presented in Section 2, which are used to infer those molecular parameters from the observed data. Consequentially, below we derive equations that allow an experimentalist to determine whether signal trajectories can be appropriately analyzed using BIASD, as well as to optimally design experiments that they will analyze using BIASD in the future. Specifically, the equations and corresponding analyses provided below allow one to determine the signal-to-noise ratio (SNR) of the signal trajectory and the number of observed datapoints in the signal trajectory that are required in order for BIASD to provide meaningful results.

We begin by considering the asymptotic limit of the likelihood function (Eqn. (7)) where $k\tau \equiv (k_1 + k_2)\tau \gg 1$, and specifically $\min(\mathrm{k_1}, \mathrm{k_2}) \cdot \tau > 5$. As described by Berezhkovskii and coworkers (3), in this asymptotic limit, Equation (7) can be written as

$$P^{\ddagger}(f|\Theta) = \mathcal{N}(f|p_1, \sigma_b^2 \equiv \frac{2p_1p_2}{k\tau}), \tag{23}$$

where $\ddagger$ denotes the expression is in the asymptotic limit, $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution as described in Section 2 on prior probability distributions with mean μ and variance σ^2, and Θ is the collection of relevant model parameters. In this asymptotic limit, and assuming $\sigma_n \equiv \sigma_1 = \sigma_2$ for simplicity, the two-state BIASD likelihood function given in Equation (11) simplifies to yield

$$\begin{aligned} P^{\ddagger}(d|\Theta) &= \int_0^1 df \cdot P(d|f, \Theta) \cdot P^{\ddagger}(f|\Theta) \\ &\approx \int_{-\infty}^{\infty} df \cdot \mathcal{N}(d|\varepsilon_1 f + \varepsilon_2(1-f), \sigma_n^2) \cdot \mathcal{N}(f|p_1, \sigma_b^2) \\ &= \mathcal{N}(d|\langle\mu\rangle, \sigma_B^2), \end{aligned} \tag{24}$$

where $\langle\mu\rangle \equiv (\varepsilon_1 p_1 + \varepsilon_2(1-p_1))^2$, and $\sigma_B^2 \equiv \sigma_n^2 + \sigma_b^2 \cdot (\varepsilon_1 - \varepsilon_2)^2$. The extension of the integral to positive and negative infinity is only valid in the asymptotic limit. This expression shows that, in this asymptotic limit, the likelihood function is a normal distribution centered at the ensemble-average signal value (*i.e.*,$\langle\mu\rangle$), and has a minimum width that is the instrument noise (*i.e.*, σ_n^2), but which is broadened by a term that depends upon the rate constants that govern the kinetics of the underlying molecular system (*i.e.*, $\sigma_b^2(\varepsilon_1 - \varepsilon_2)^2$·); we will refer to this latter contribution as "kinetic broadening". Figure S4 shows the exact likelihood function (black, dashed), and the likelihood function in the asymptotic limit (blue, solid) for an arbitrarily chosen Θ; the residual between the two functions is also shown (blue, dashed).

Given a collection of observed datapoints from a particular signal trajectory, $\{d\} \equiv \{d_1, \cdots d_i, \cdots d_N\}$, there are two interesting models to consider in the context of Bayesian inference that might describe this $\{d\}$. Model one, M_1, is the asymptotic limit of a two-state system described above with the kinetic broadening expression in Equation (24), or rather

$$p(\{d\}|\Theta, M_1) = \prod_{i=1}^{N} \mathcal{N}(d_i|\langle\mu\rangle, \sigma_B^2). \tag{25}$$

Model two, M_2, is just instrument noise coincidentally centered at the ensemble-average signal value, such that

$$p(\{d\}|\Theta, M_2) = \prod_{i=1}^{N} \mathcal{N}(d_i|\langle\mu\rangle, \sigma_n^2). \tag{26}$$

Essentially, M_1 is a kinetically broadened two-state system centered at $\langle\mu\rangle$, while M_2 is a false positive where some noisy, one-state signal happens to have an average value that is the same as $\langle\mu\rangle$ (Fig. S4, red).

The question of how BIASD performs in the analysis of a particular experimentally recorded signal trajectory (*i.e.*, a particular $\{d\}$) can be evaluated by asking if the signal trajectory contains enough information about the kinetic broadening to make M_1 significantly more probable than M_2. Is there evidence for a two-state system in the signal trajectory, or is it more likely that the signal trajectory is just a false positive? This question is a traditional model-selection problem, which can therefore be answered using Bayesian inference.

In order to establish a minimum baseline for the performance of BIASD, we will proceed without invoking any underlying knowledge of the experimental system, and assume equal prior probabilities for M_1 and M_2 (*i.e.*, $P(M_1) = P(M_2)$). Using knowledge of a previous experiment, or even just knowing that the observed signal trajectory truly originated from a *bona fide* two-state system, for instance by first observing a slowly interchanging two-state system and then altering the experimental conditions to increase the rate of transitions and produce a signal trajectory with kinetic broadening to be analyzed with BIASD, will improve the performance of BIASD. With equal prior probabilities for both models, for one particular collection of N observed datapoints in a particular signal trajectory, $\{d\}$, the model-selection question posed above is

$$\begin{aligned} P(M_1|\Theta,\{d\}) &= \frac{P(\{d\}|\Theta,M_1)\cdot P(M_1)}{P(\{d\}|\Theta,M_1)\cdot P(M_1)+P(\{d\}|\Theta,M_2)\cdot P(M_2)} \\ &= \left(1+\left(\frac{\sigma_B^2}{\sigma_n^2}\right)^N \cdot e^{-\frac{1}{2}\left(\frac{1}{\sigma_n^2}-\frac{1}{\sigma_B^2}\right)\cdot\sum_{i=1}^N (d_i-\langle\mu\rangle)^2}\right)^{-1}. \end{aligned} \tag{27}$$

Notably, because $\sigma_n^2 < \sigma_B^2$, regardless of the parameter values underlying the signal trajectory, this equation states that $P(M_1|\Theta,\{d\}) > 0.5$. This means that there is *always* some information gained by using BIASD, and that every additional datapoint that is observed contributes more and more information. As a result, this equation states that one way to increase the performance of BIASD is to simply collect and analyze more datapoints - a conclusion that agrees well with intuition.

To understand how BIASD performs as the values of the experimental and molecular parameters that underlie that signal trajectory are changed, consider the expectation value of $P(M_1|\Theta,\{d\})$ given the different possible collections of $\{d\}$ that might be observed in a signal trajectory. These possible collections can be described with the exact two-state likelihood distribution of a collection of observed datapoints in a particular signal trajectory given in Equation (11). In the asymptotic limit given in Equation (25), this expectation value is thus

$$\mathbb{E}\left[P(M_1|\Theta,\{d\})\right] = \int dd_1 \cdots \int dd_N \cdot P(M_1|\Theta,\{d\}) \cdot \prod_{i=1}^N P(d_i), \tag{28}$$

where $P(d_i) = \mathcal{N}(d_i|\langle\mu\rangle,\sigma_B^2)$, and $\mathbb{E}$ denotes an expectation value. Using Jensen's inequality ($\mathbb{E}\left[f(x)\right] \geq f(\mathbb{E}\left[x\right])$), a lower bound for this expectation value in the asymptotic limit can be calculated as

$$\begin{aligned} \mathbb{E}[P] \equiv \mathbb{E}\left[P(M_1|\Theta,\{d\})\right] &\geq P(M_1|\Theta,\mathbb{E}\left[\{d\}\right]) \\ &\geq \left(1+\left(\frac{\sigma_B^2}{\sigma_n^2}\right)^{N/2} \cdot e^{\frac{-1}{2}\left(\frac{1}{\sigma_n^2}-\frac{1}{\sigma_B^2}\right)\cdot\sum_{i=1}^N \mathbb{E}\left[(d_i-\langle\mu\rangle)^2\right]}\right)^{-1} \\ &\geq \left(1+\left(\frac{\sigma_B^2}{\sigma_n^2}\right)^{N/2} \cdot e^{\frac{-N}{2}\left(\frac{\sigma_B^2}{\sigma_n^2}-1\right)}\right)^{-1} \\ \mathbb{E}[P] &\geq \left(1+\left(1+\phi^2\right)^{N/2} \cdot e^{\frac{-N}{2}\cdot\phi^2}\right)^{-1}, \end{aligned} \tag{29}$$

where $\phi^2 \equiv \frac{\sigma_b^2(\varepsilon_1-\varepsilon_2)^2}{\sigma_n^2} = \frac{2p_1p_2}{k\tau}\cdot\frac{(\varepsilon_1-\varepsilon_2)^2}{\sigma_n^2} \equiv \frac{2p_1p_2}{k\tau}\cdot \mathrm{SNR}^2$. Thus, this equation for the expected probability of M_1 relates the SNR of a particular signal trajectory and the number of observed datapoints that comprise the signal trajectory, N, to the lower bound-performance of BIASD at a particular value of k_1, k_2, and τ. Additionally, rearranging Equation (29) and solving for N, yields the upper bound for the number of observed datapoints that one would be required to collect in order to have a particular expected probability of M_1. This is

$$N \leq \frac{2\cdot \ln\left(\frac{1}{\mathbb{E}[P]}-1\right)}{\ln\left(1+\phi^2\right)-\phi^2}. \tag{30}$$

Notably, Equations (29) and (30) connect the SNR and the number of observed datapoints in a particular signal trajectory to the performance of BIASD in the analysis of that signal trajectory. Thus, if a researcher can estimate the kinetics ($k = k_1 + k_2$) and thermodynamics ($K_{\mathsf{eq}} = \frac{p_2}{p_1} = \frac{k_1}{k_2}$) of their system, or, in the worst-case scenario, the bounds that they might expect on the kinetics

and thermodynamics of their system, then these equations can then be used to plan time-resolved, single-molecule biophysical experiments of their system. For instance, one might calculate the expected probability that M_1 rather than M_2 is best described by their signal trajectory, or one can calculate the number of datapoints that one would be required to collect in order for BIASD to be able to analyze the kinetic broadening present in a particular signal trajectory. In particular, while any value of $\mathbb{E}[P] > 0.5$ provides information on the underlying experimental and molecular parameters, we would suggest that using a conservative $\mathbb{E}[P] > 0.95$ is a reasonable target to choose when planning an experiment. Figure S5 contains a series of heat maps of $\mathbb{E}[P]$ at various k with $\tau = 1$ as a function of SNR and number of analyzed datapoints; the $\mathbb{E}[P] = 0.95$ contour is marked as a white, dashed line. As the heat maps demonstrate, increasing the SNR and/or the number of analyzed datapoints improves the performance of BIASD. This is an intuitive and expected result, as increasing the SNR an/or the number of analyzed datapoints will increase the evidence of kinetic broadening present that is present in the signal trajectory and that can be analyzed using BIASD.

There are several caveats to this analysis. First, it has been performed in the asymptotic limit described above. As a result, the equations, analyses, and associated conclusions are only valid in this limit (*i.e.*, both $k_1\tau \gg 1$ and $k_2\tau \gg 1$). Additionally, in the asymptotic limit, the likelihood function is a normal distribution, which is symmetric about the first moment; the exact two-state likelihood function used for BIASD is asymmetric, as can be seen in the residual between the asymptotic limit and exact likelihood functions shown in Figure S4. Accounting for this asymmetry in M_1 would otherwise have provided additional evidence to distinguish between M_1 and M_2. Thus, the equations derived for Equations (29) and (30) are under- and over-estimates, respectively. This is in addition to the fact that Equations (29) and (30) are already lower and upper bounds, respectively, because they were derived with Jensen's inequality. Finally, while the collection of observed datapoints in a particular signal trajectory might not contain much information regarding the experimental and molecular parameters underlying that signal trajectory and causing the kinetic broadening, even if $k\tau \gg 1$, the use of Bayesian inference in BIASD ensures that the corresponding analysis will *always* be accurate to within the precision contained within the data, as long as the results of the BIASD analysis are properly reported with credible intervals.

References

1. Good, I. The Frequency Count of a Markov Chain and the Transition to Continuous Time. *Ann. Math. Stat.* **32,** 41–48 (1961).
2. Dobrushin, R. Limit theorems for a markov chain of two states. *Izv. Ross. Akad. Nauk. USSR Seriya Mat.* **17,** 291–330 (1953).
3. Berezhkovskii, A. M., Szabo, A. & Weiss, G. H. Theory of single-molecule fluorescence spectroscopy of two-state systems. *J. Chem. Phys.* **110,** 9145 (1999).
4. Gopich, I. V. & Szabo, A. Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proc. Natl. Acad. Sci.* **109,** 7747–52 (May 2012).
5. Sivia, D. S. & Skilling, J. *Data Analysis: A Bayesian Tutorial* 1–259 (Oxford University Press, Oxford, 2006).
6. Bishop, C. M. *Pattern Recognition and Machine Learning* 461–652 (Springer, New York, 2006).
7. Goodman, J. & Weare, J. Ensemble samplers with affine invariance. *Commun. Appl. Math. Comput. Sci.* **5,** 65–80 (Jan. 2010).
8. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee : The MCMC Hammer. *Publ. Astron. Soc. Pacific* **125,** 306–312 (2013).
9. Wang, B., Ho, J., Fei, J., Gonzalez, R. L. & Lin, Q. A microfluidic approach for investigating the temperature dependence of biomolecular activity with single-molecule resolution. *Lab Chip* **11,** 274–81 (Jan. 2011).
10. Rabiner, L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE* **77,** 257–286 (1989).

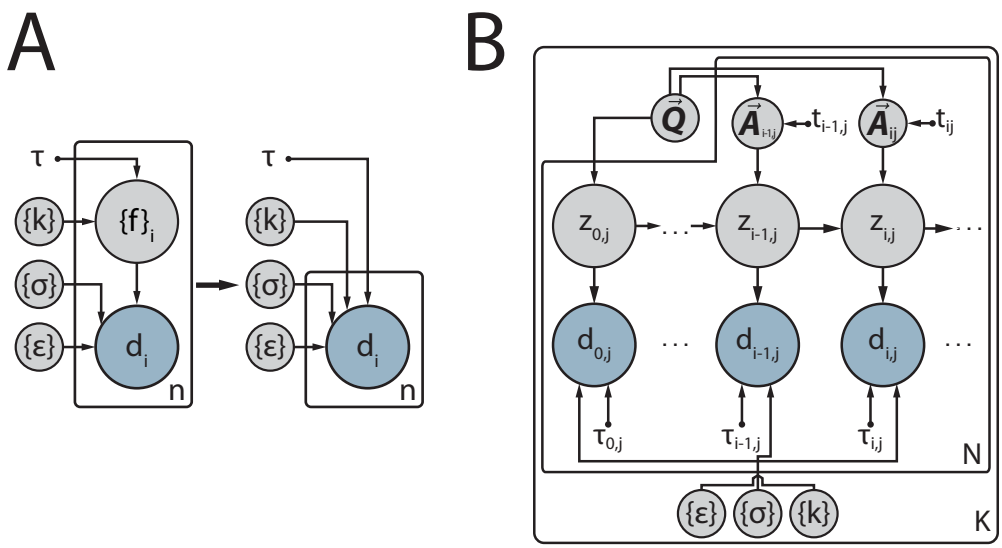

Figure S1: Graphical models for BIASD. (A) In BIASD, the dependence of the observed data point, d_i, upon the set of fractional occupancies, $\{f\}$, is marginalized to yield the graph on right, which depends upon sets of stochastic parameters $\{\varepsilon\}$, $\{\sigma\}$, and $\{k\}$, as well as upon the deterministic value of τ. (B) A directed graph describing hierarchical, Markovian transitions between hidden sub-temporal-resolution sub-populations is shown. The rate matrix, **Q**, dictates the steady-state probabilities, as well as transition probabilities between sub-temporal-resolution sub-populations for neighboring data points. The plates denote N signal trajectories and K sub-temporal-resolution sub-populations.

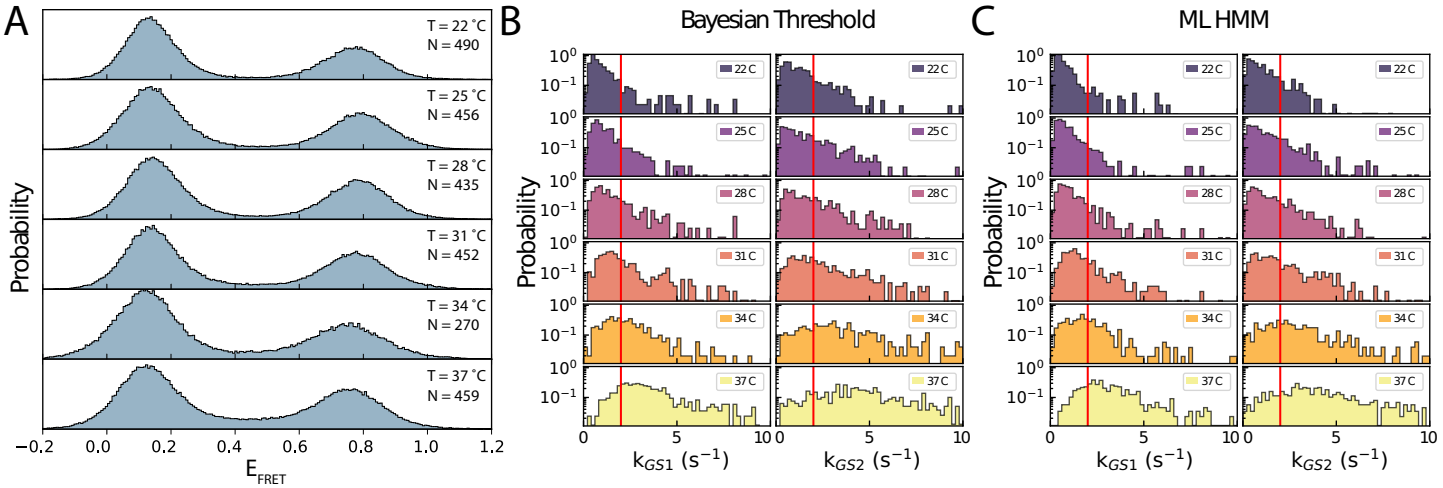

Figure S2: Temperature Dependence of PRE^{-A} complexes. (A) Histogram of $\mathrm{E_{FRET}}$. GS1 corresponds to the peak at $\mathrm{E_{FRET}} \sim 0.15$, and GS2 corresponds to the peak at $\mathrm{E_{FRET}} \sim 0.75$. Increasing rate constants with increasing temperatures yield more time-averaging, which results in $\mathrm{E_{FRET}}$ data points in the region between ϵ_{GS1} and ϵ_{GS2}. (B,C) Histograms of the rate constants estimated for each $\mathrm{E_{FRET}}$ trajectory in the set of $\mathrm{E_{FRET}}$ trajectories collected at each temperature obtained using: (B) Bayesian Thresholding-based idealization with the threshold set at an $\mathrm{E_{FRET}}$ of 0.45 and (C) Maximum Likelihood Hidden Markov Model (ML HMM)-based idealization. $\mathrm{E_{FRET}}$ trajectories that did not converge were not included in the histograms.

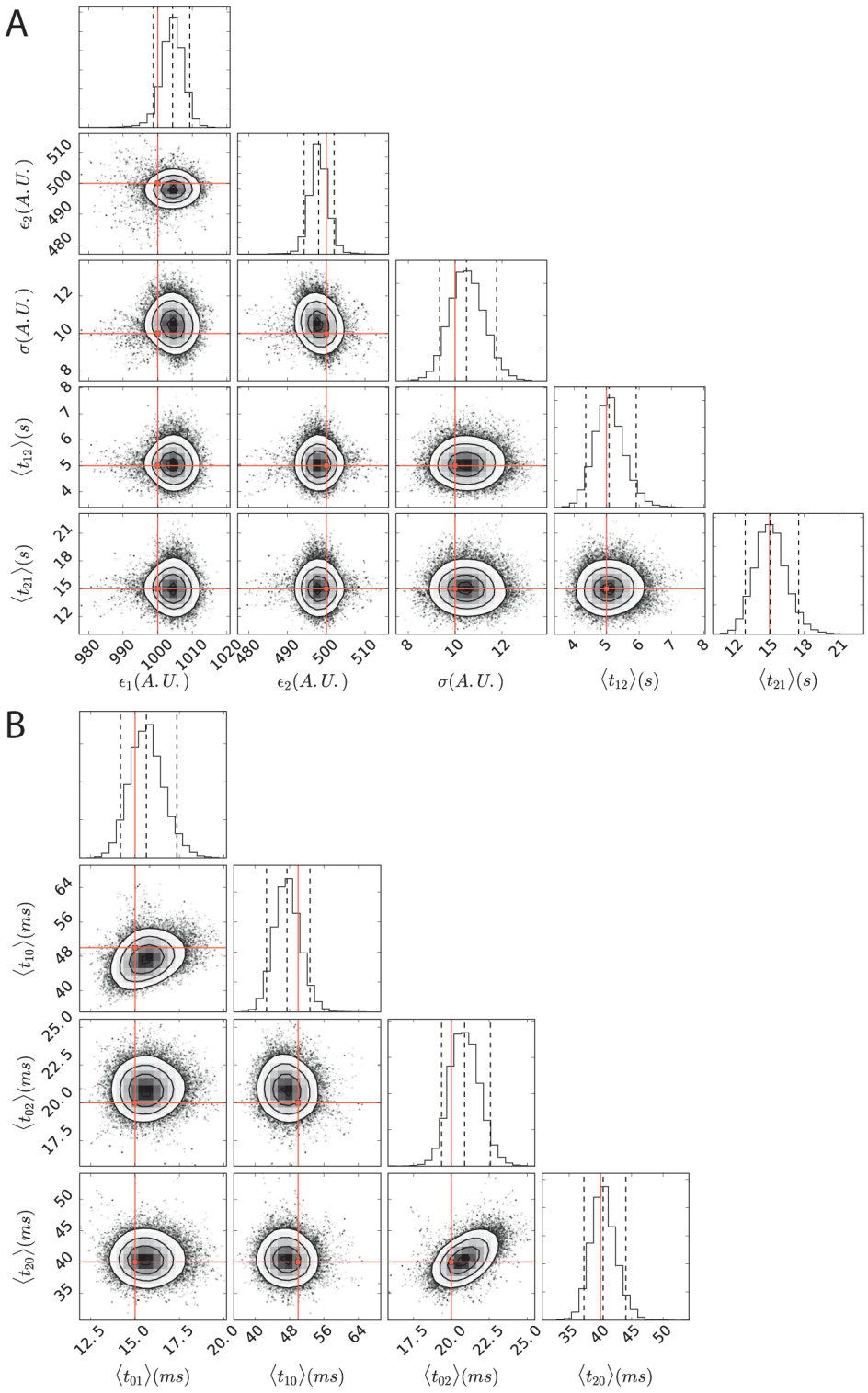

Figure S3: Corner Plot of Posterior Probability Distribution for Hierarchical Model with Sub-temporal-resolution Photoblinking Dynamics. Corner plot for (A) ϵ_1, ϵ_2, σ, $\langle t_{12}\rangle = 1/k_{12}$, and $\langle t_{21}\rangle = 1/k_{21}$, and (B) $\langle t_{01}\rangle = 1/k_{01}$, $\langle t_{10}\rangle = 1/k_{10}$, $\langle t_{02}\rangle = 1/k_{20}$, and $\langle t_{20}\rangle = 1/k_{20}$. Cross-sections between the parameters shown in (A) and in (B) are not shown. Cross-sections of the 9D-posterior probability distribution are shown as 2D heat-maps of the MCMC samples. Marginalized 1D-histograms for each parameter shown the 2.5%, 50%, and 97.5% probability levels. Parameter values from the simulation are shown in red.

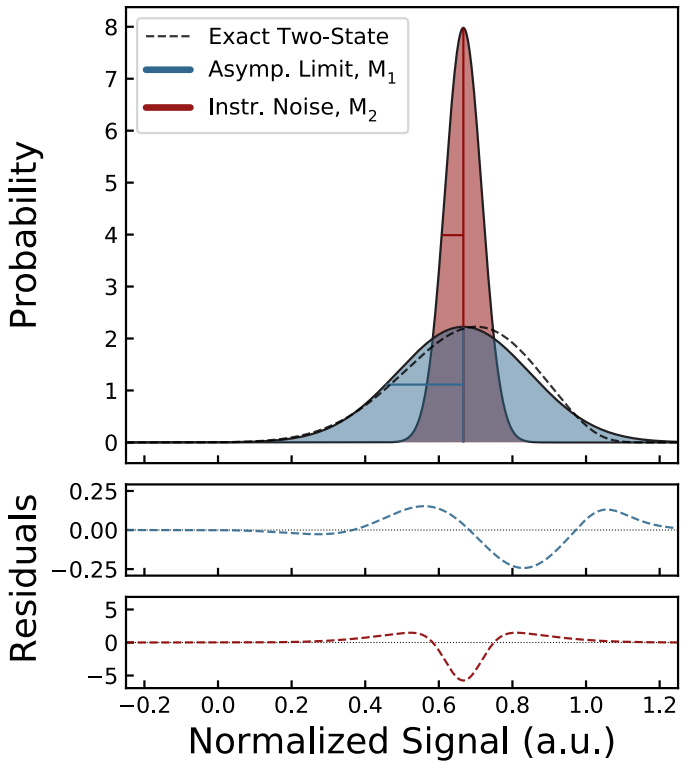

Figure S4: Example likelihood functions for a single datapoint. The exact two-state likelihood function for one datapoint (Equation (11)) (black, dashed), and the two-state likelihood function for one datapoint in the asymptotic limit (M_1, Equation (25)) (blue, solid), are shown with parameters $\varepsilon = 0$, $\varepsilon = 1$, $\sigma_1 = \sigma_2 = 0.05$, $k_1 = 10s^{-1}$, $k_2 = 5s^{-1}$, and $\tau = 1s$. The signal values ε_1 and ε_2 are normalized to 0 and 1, and therefore generalizable to any linear signal by shifting and rescaling. The likelihood function for a one state system centered at $\langle\mu\rangle$ and broadened only by instrument noise $\sigma = 0.05$ (M_2, Equation (26)) is shown in red. The means, and half widths at half maximum are shown for both M_1 and M_2. Residual differences between M_1 and the exact two-state likelihood function (blue, dashed), and between M_2 and the exact two-state likelihood function (red, dashed) are shown below; these have vastly different scales. The ability of BIASD to infer the experimental and molecular parameters underlying a signal trajectory composed of many datapoints depends upon discriminating between M_1 and M_2.

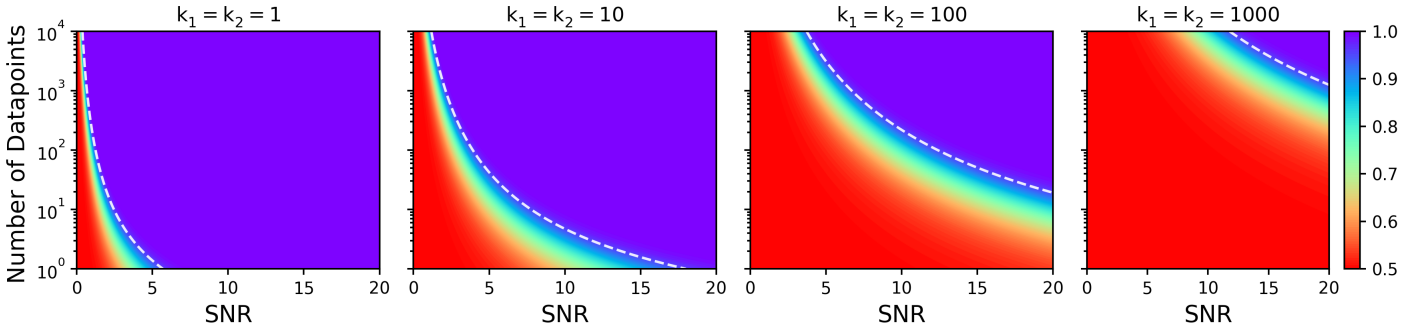

Figure S5: The SNR and number of datapoints dependence of the lower bound expected probability of M_1. The heat maps show the lower limit of $\mathbb{E}[P]$ obtained in the asymptotic limit (Equation (29)) as a function of N and SNR at various k_1 and k_2 when $\tau = 1$. The white dashed line is the 95% contour, which can be used to as a mark to surpass when designing experiments. The performance of BIASD increases with increasing SNR and increasing number of datapoints that are analyzed, and is better than shown in this lower limit.