Single Photon smFRET. II. Application to Continuous Illumination

Ayush Saurabh[1], Matthew Safar[1,2], Mohamadreza Fazel[1], Ioannis Sgouralis[3], and Steve Pressé[1,4]

[1]Center for Biological Physics, Department of Physics, Arizona State University, Tempe, AZ, USA
[2]Department of Mathematics and Statistical Science, Arizona State University, Tempe, AZ, USA
[3]Department of Mathematics, University of Tennessee Knoxville, Knoxville, TN, USA
[4]School of Molecular Sciences, Arizona State University, Tempe, AZ, USA

July 21, 2022

Contents

Abstract

Single molecule Förster resonance energy transfer (smFRET) plays a crucial role in revealing biomolecular dynamics over a wide range of timescales. Currently available tools that analyze smFRET data, however, remain limited in their scope as they are unable to simultaneously estimate: 1) the number of system states visited by a biomolecular complex; 2) the associated kinetic rates; and 3) uncertainties over the estimated parameters, while including features such as crosstalk, instrument response function (IRF) and background in the model. In this paper, we adapt the Bayesian nonparametrics (BNP) framework presented in the first paper to analyze dynamics from single photon smFRET traces generated under continuous illumination. Using this method, we learn the lifetimes/escape rates and the number of system states given a trace of photons. We benchmark our method by analyzing a range of synthetic and experimental data. Particularly, we apply our method to simultaneously learn the the number of system states and the corresponding dynamics for intrinsically disordered proteins (IDPs) using two-color FRET under varying chemical conditions. Moreover, using synthetic data, we show that our method can deduce the number of system states even when dynamics occur at timescales of interphoton intervals.

1 Terminology Convention

To be consistent throughout our three part manuscript, we precisely define some terms as follows:

1. a biomolecular complex labeled with a FRET dye pair is always referred to as a *system*,
2. the spatial configurations through which a system transitions are termed *system states*,
3. FRET dyes undergo quantum mechanical transitions between *photophysical states*,
4. a system-FRET combination is always referred to as a *composite*, and
5. a composite undergoes transitions among its *superstates*.

2 Introduction

Single molecule Förster Resonance Energy Transfer (smFRET) experiments are widely used [1] to study molecular dynamics across timescales on both stationary [2, 3] and freely diffusing molecules [4]. These timescales include faster events, below the micro- to millisecond timescales, including some domain rotations, configurational dynamics of disordered proteins, protein folding and protein-protein interactions, all the way to slower events, such as present misfolding and refolding events, occurring on minute and even hour long timescales [5].

In a typical experiment we consider herein, a continuous wave (CW) laser illuminates a sample with a beam of constant intensity and power over a period of time. CW sources

are common as they both are typically cheaper and technically simpler to implement in an experimental setup than their pulsed counterparts [6]. However, as compared to pulsed sources, a disadvantage lies in the increased photon flux through the sample which can accelerate photodamage [7].

While pulsed illumination can significantly reduce sample photobleaching and phototoxicity [8], in practice it is restricted to analyzing one (time-stamped) photon per interpulse period. This in turn limits the data acquisition rate and sets a bound on the temporal resolution of the dynamics we may deduce from pulsed single photon arrival. In principle, we may generalize the likelihood to consider not just the first photon arising after each pulse but, rather, all photons recorded following each pulse. While this is in principle possible, it is seldom done due to added computational cost.

By contrast, continuous illumination avoids this problem, by allowing a larger number of photons to be detected in the time that would normally be considered an interpulse period in pulsed illumination [9]. The cost then comes at the loss of direct knowledge of excited state lifetime which can, with difficulty and high uncertainty, then be decoded from photon-antibunching statistics if required [10].

It is common practice to analyze photon arrival data under continuous illumination by binning the data and subsequently using hidden Markov models (HMMs) [11–14]. It remains conceptually preferred, though more computationally costly, to use photon-by-photon methods [9, 10, 15–19]. Indeed, photon-by-photon methods can be used to learn both photophysical and system transition rates directly from the detected photon colors and interphoton arrival times. Additionally, this has the benefit of avoiding averaging dynamics that may occur when binning data [12].

Currently available methods to analyze smFRET data in a photon-by-photon manner [9, 15] rely on the foundational works of Gopich and Szabo [9, 10, 20], where the likelihood is taken as the product of as many generator matrix exponentials as there are photons in a FRET trace. Such a generator matrix constitutes transition rates encoding the kinetics of the system-FRET composite [21].

When analyzing smFRET data, of particular interest is the dimensionality of this generator matrix which is determined by the number of system states. In all existing analyses, the dimensionality is fixed by hand *a priori* and the transition rates are then learned as point estimates using maximum likelihood methods.

Yet point estimates can be biased. In fact, limited data, lack of temporal resolution to estimate very fast kinetics, and noise all contribute to bias [22] in addition to a likelihood being flat or multimodal [23, 24]. This motivates why we wish to operate in a Bayesian setting to learn distributions over the number of system states and transition rates, while incorporating unavoidable noise sources such as detector electronics and background.

For this reason, we developed a complete Bayesian nonparametric (BNP) framework in the first paper of the current series [21]. This framework incorporates many key complexities of a typical smFRET experimental setup, including background emissions, fluorophore photophysics (blinking, photobleaching, and direct acceptor excitation), instrument response function (IRF), detector dead time, and crosstalk.

Here, we delve deeper into this framework for the case of continuous illumination by exploring its utility in cases where the number of system states is unknown.

We first test the robustness of our nonparametric method and its software implementa-

tion BNP-FRET [25] by analyzing synthetically generated data for dynamics varying from very slow to as fast as the interphoton arrival times. Then apply our method to experimental smFRET data capturing interactions between intrinsically disordered protein (IDP) fragments [26, 27] relevant to signaling and regulation of cellular activities. IDPs are of particular interest to nonparametric analyses as IDP lack of order and stability results in a broader spectrum of dominant FRET pair distances sensitive to their chemical environment.

In particular, we study interactions between the nuclear-coactivator binding domain (NCBD) of a CBP/p300 transcription coactivator and the activation domain of SRC -3 (ACTR) under varying chemical conditions affecting their coupled folding and binding reaction rates [26, 27]. We use a single FRET pair under continuous illumination to observe the possible physical configurations (system states) of the NCBD-ACTR complex. Further, we report new bound/transient system states for the NCBD P20A mutation in buffer (crowder absent), not observed using previous point estimation techniques [27].

3 Forward Model and Inference Strategy

For the sake of completeness, we begin with relevant aspects of the methods presented in the first companion paper [21], including the likelihood needed in Bayesian inference, and our parametric and nonparametric Markov Chain Monte Carlo (MCMC) samplers.

An smFRET experiment involves at least two single photon detectors collecting information on stochastic arrival times. We denote these arrival times with

$$\{T_{start}, T_1, T_2, T_3, \ldots, T_K, T_{end}\},$$

in detection channels

$$\{c_1, c_2, c_3, \ldots, c_K\},$$

for a total number of K photons. In this representation above, T_{start} and T_{end} are the experiment's start and end times, respectively.

Using this stochastic dataset, we would like to infer parameters governing a system's kinetics, that is, the number of system states M_σ and the associated transition rates $\lambda_{\sigma_i \to \sigma_j}$, as well as M_ψ photophysical transition rates $\lambda^{(i)}_{\psi_l \to \psi_m}$ corresponding to each system state σ_i. Here, $\sigma_i \in \{\sigma_1, \ldots, \sigma_{M_\sigma}\}$ and $\psi_l \in \{\psi_1, \ldots, \psi_{M_\psi}\}$ are the system states and photophysical states, respectively. These rates populate a generator matrix $\mathbf{G}$ of dimension $M_\phi = M_\psi \times M_\sigma$ now representing transitions among composite superstates, $\phi_i \equiv (\sigma_j, \psi_k)$ where $i = (k-1)M_\psi + j$ (see the first companion manuscript for details [21] on the structure of such a matrix). This matrix governs the evolution of the system-FRET composite via the master equation

$$\frac{d\boldsymbol{\rho}(t)}{dt} = \boldsymbol{\rho}(t)\mathbf{G}, \tag{1}$$

as described in Sec. 2.2 of the first companion manuscript [21].

In estimation of these parameters, we must account for all the sources of uncertainty present in the experiment, such as shot noise and detector electronics. Therefore, we naturally work within the Bayesian paradigm where the parameters are learned by sampling from a probability distribution termed the posterior over these parameters. This posterior is

proportional to the product of the likelihood, which is the probability of the collected data given the physical model, and prior distributions over the parameters as follows:

$$p(\mathbf{G}|\boldsymbol{w}) \propto L(\boldsymbol{w}|\mathbf{G})p(\mathbf{G}), \tag{2}$$

where $\boldsymbol{w}$ constitutes the set of all observations, including photon arrival times and detection channels.

To construct the posterior, we begin with the likelihood

$$L(\boldsymbol{w}|\mathbf{G}) \propto \boldsymbol{\rho}_{start}\ \boldsymbol{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \boldsymbol{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \ldots\ \boldsymbol{\Pi}_{K-1}^{non}\mathbf{G}_{K-1}^{rad}\ \boldsymbol{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \boldsymbol{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{norm}^T, \tag{3}$$

derived in Sec. 2.2 in the first manuscript. Here, $\boldsymbol{\Pi}_k^{non}$ and $\mathbf{G}_k^{rad}$ are the non-radiative and radiative propagators, respectively. Furthermore, $\boldsymbol{\rho}_{start}$ is computed by solving the master equation assuming the system was at equilibrium immediately preceding the time at which the experiment began. That is, we solve

$$\boldsymbol{\rho}_{start}\mathbf{G} = 0.$$

Next, assuming that the transition rates are independent of each other, we can write the associated prior as

$$p(\mathbf{G}) = \prod_{i,j} p(\lambda_{\phi_i \to \phi_j}),$$

where we choose Gamma prior distributions over individual rates, that is,

$$p(\lambda_{\phi_i \to \phi_j}) = \mathbf{Gamma}\left(\lambda_{\phi_i \to \phi_j}; \alpha, \frac{\lambda_{ref}}{\alpha}\right),$$

to guarantee positive values. Here, ϕ_i represents one of the M_ϕ superstates of the system-FRET composite collecting both the system and photophysical states as described in Sec. 2.2. Furthermore, α and λ_{ref} are the parameters of the Gamma prior.

In what follows, we first assume that the number of system states are known and will describe an inverse strategy that uses the posterior above to only learn the transition rates. Next, we generalize our model to a nonparametric case accommodating more practical situations with unknown number of system states. We do so by assuming an infinite dimensional system state space and making the existence of each system state itself a random variable.

3.1 Inference Procedure: Parametric Sampler

Now, with the posterior defined, here we prescribe a sampling scheme to learn distributions over all the parameters of interest, namely, transitions rates of $\mathbf{G}$ and the number of states. However, our posterior in Eq. 2 does not assume a form amenable to analytical calculations. Therefore, we employ Markov Chain Monte Carlo (MCMC) techniques to draw numerical samples.

Particularly convenient here is the Gibbs algorithm that sequentially and separately generate samples for individual transition rates in each MCMC iteration. This requires us to first write the posterior in Eq. 2 using the chain rule as follows:

$$p(\mathbf{G}|\boldsymbol{w}) = p(\lambda_{\phi_i \to \phi_j}|\mathbf{G}\backslash\lambda_{\phi_i \to \phi_j}, \boldsymbol{w})p(\mathbf{G}\backslash\lambda_{\phi_i \to \phi_j}|\boldsymbol{w}), \tag{4}$$

where the backslash after $\mathbf{G}$ indicates exclusion of the following rate parameter. Furthermore, the first term on the right hand side is the conditional posterior for the individual rate $\lambda_{\phi_i \to \phi_j}$. The second term in the product is a constant in the corresponding Gibbs step as it is independent of $\lambda_{\phi_i \to \phi_j}$. Similarly, the priors $p(\mathbf{G} \backslash \lambda_{\phi_i \to \phi_j})$ for the rest of the rate parameters on the right hand side of Eq. 2 are also considered constant. Equating the right hand sides of Eqs. 2 & 4 then allows us to write the following conditional posterior for $\lambda_{\phi_i \to \phi_j}$ as

$$p(\lambda_{\phi_i \to \phi_j} | \mathbf{G} \backslash \lambda_{\phi_i \to \phi_j}, \boldsymbol{w}) \propto L(\boldsymbol{w}|\mathbf{G}) \;\; \mathbf{Gamma}\left(\lambda_{\phi_i \to \phi_j}; \alpha, \frac{\lambda_{ref}}{\alpha}\right). \tag{5}$$

Since the conditional posterior above does not take a closed form that allows direct sampling, we use the Metropolis-Hastings (MH) step [28–34], where new samples are drawn from a proposal distribution q and accepted with the probability

$$\alpha(\lambda^*_{\phi_i \to \phi_j}, \lambda_{\phi_i \to \phi_j}) = \mathbf{min}\left\{1, \frac{p(\lambda^*_{\phi_i \to \phi_j} | \boldsymbol{w}, \mathbf{G} \backslash \lambda_{\phi_i \to \phi_j}) \; q(\lambda_{\phi_i \to \phi_j} | \lambda^*_{\phi_i \to \phi_j})}{p(\lambda_{\phi_i \to \phi_j} | \boldsymbol{w}, \mathbf{G} \backslash \lambda_{\phi_i \to \phi_j}) \; q(\lambda^*_{\phi_i \to \phi_j} | \lambda_{\phi_i \to \phi_j})}\right\}, \tag{6}$$

where the asterisk represents the proposed rate values from the proposal distribution q.

Now, to generate an MCMC chain of samples, we first initialize the chains for all the transition rates $\lambda_{\phi_i \to \phi_j}$, by randomly draw values from their corresponding prior distributions. We then one-by-one go through each transition rate in each new MCMC iteration and draw new samples from the corresponding conditional posterior using the MH step.

In the MH step, a convenient choice for the proposal is a Normal distribution leading to a simpler formula for acceptance probability in Eq. 6. This is due to its symmetry resulting in $q(\lambda_{\phi_i \to \phi_j} | \lambda^*_{\phi_i \to \phi_j}) = q(\lambda^*_{\phi_i \to \phi_j} | \lambda_{\phi_i \to \phi_j})$. However, a normal proposal distribution would allow forbidden negative transition rates, leading to automatic rejection in the MH step and thus inefficient sampling. Therefore, it is more convenient to propose new samples using a Normal distribution in the logarithmic space to allow exploration along the full real line as follows:

$$\log(\lambda^*_{\phi_i \to \phi_j}/\kappa) \;\big|\; \log(\lambda_{\phi_i \to \phi_j}/\kappa), \sigma^2 \sim \mathbf{Normal}\left(\log(\lambda_{\phi_i \to \phi_j}/\kappa), \sigma^2\right),$$

where $\kappa = 1$ is an auxiliary parameter in the same units as $\lambda_{\phi_i \to \phi_j}$ introduced to obtain a dimensionless quantity within the logarithm.

The transformation above requires introduction of Jacobian factors in the acceptance probability as follows:

$$\alpha(\lambda^*_{\phi_i \to \phi_j}, \lambda_{\phi_i \to \phi_j}) = \mathbf{min}\left\{1, \frac{p(\lambda^*_{\phi_i \to \phi_j} | \boldsymbol{w}, \mathbf{G} \backslash \lambda_{\phi_i \to \phi_j})}{p(\lambda_{\phi_i \to \phi_j} | \boldsymbol{w}, \mathbf{G} \backslash \lambda_{\phi_i \to \phi_j})} \frac{(\partial \log(\lambda_{\phi_i \to \phi_j}/\kappa)/\partial \lambda_{\phi_i \to \phi_j})}{(\partial \log(\lambda_{\phi_i \to \phi_j}/\kappa)/\partial \lambda_{\phi_i \to \phi_j})^*}\right\}.$$

where the derivatives represent the Jacobian and the proposal distributions are canceled by virtue of using a Normal distribution.

The acceptance probability above depends on the difference of the current and proposed values for a given transition rate. This difference is determined by the covariance of the Normal proposal distribution σ^2 which needs to be tuned for each rate individually to achieve an optimum performance of the BNP-FRET sampler, or equivalently approximately one-third acceptance rate for the proposals [35].

In our case where the smFRET traces analyzed contain about 10^5, we found it prudent to make the sampler alternate between two sets of variances, $\{\sigma^2_{ex} = 10^{-5}, \sigma^2_{FRET} = 0.01, \sigma^2_{sys} = 0.1\}$ and $\{\sigma^2_{ex} = 10^{-5}, \sigma^2_{FRET} = 0.5, \sigma^2_{sys} = 5.0\}$, for the excitation rates, FRET rates, and system transition rates. This ensure that sampler is quickly able to explore values at different orders of magnitude.

Intuitively, these covariance values in the proposal distributions above represent the relative widths of the conditional posteriors for these parameters (in log-space). Since a posterior width depends on the amount of data used, an increase or decrease in the number of photons used for analysis will require a corresponding decrease or increase in the covariances, respectively.

3.2 Inference Procedure: Nonparametric BNP-FRET Sampler

In realistic situations, the dimensionality of the system state space is usually unknown as molecules under study may exhibit complex and unexpected behaviors across conditions and timescales. Consequently, the dimensionality M_ϕ of the generator matrix $\mathbf{G}$ is also unknown, and must be determined by adopting a BNP framework.

In such a framework, we assume an infinite set of system states and place a binary weight, termed load, on each system state such that if it is warranted by the data, the value of the load is realized to one. Put differently, we must place a Bernoulli prior on each candidate state (of which there are formally an infinite number) [36–38]. In practice, we learn distributions over Bernoulli random variables b_i that activate/deactivate different portions of the full generator matrix as (see Sec. 3.2.1 of the first companion manuscript)

$$\mathbf{G} = \begin{bmatrix} * & b_1^2\lambda_{\psi_1\to\psi_2} & b_1^2\lambda_{\psi_1\to\psi_3} & b_1b_2\lambda_{\sigma_1\to\sigma_2} & 0 & 0 & \dots \\ b_1^2\lambda_{\psi_2\to\psi_1} & * & b_1^2\lambda^{(1)}_{\psi_2\to\psi_3} & 0 & b_1b_2\lambda_{\sigma_1\to\sigma_2} & 0 & \dots \\ b_1^2\lambda_{\psi_3\to\psi_1} & b_1^2\lambda_{\psi_3\to\psi_2} & * & 0 & 0 & b_1b_2\lambda_{\sigma_1\to\sigma_2} & \dots \\ b_1b_2\lambda_{\sigma_2\to\sigma_1} & 0 & 0 & * & b_2^2\lambda_{\psi_1\to\psi_2} & b_2^2\lambda_{\psi_1\to\psi_3} & \dots \\ 0 & b_1b_2\lambda_{\sigma_2\to\sigma_1} & 0 & b_2^2\lambda_{\psi_2\to\psi_1} & * & b_2^2\lambda^{(3)}_{\psi_2\to\psi_3} & \dots \\ 0 & 0 & b_1b_2\lambda_{\sigma_2\to\sigma_1} & b_2^2\lambda_{\psi_3\to\psi_1} & b_2^2\lambda_{\psi_3\to\psi_2} & * & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where active loads are set to 1, while inactive loads are set to 0. Furthermore, $*$ represents negative row-sums. Finally, the number of active loads provides an estimate of the number of system states warranted by a given dataset.

As we have introduced new variables we wish to learn, we upgrade the posterior of Eq. 2 to incorporate the full set of loads, $\mathbf{b} = \{b_1, b_2, \dots, b_\infty\}$, as follows:

$$p(\mathbf{b}, \mathbf{G}|\boldsymbol{w}) \propto L(\boldsymbol{w}|\mathbf{b}, \mathbf{G})\, p(\mathbf{b})p(\mathbf{G}),$$

where we assume that all the parameters of interest are independent of each other.

As in the parametric sampler presented in the previous subsection, we generate samples from the nonparametric posterior above using Gibbs algorithm. That is, we first initialize the MCMC chains for loads and rates by drawing random samples from their priors. Next, to construct the chains, we iteratively draw samples from the posterior in two steps: 1)

sequentially sample all the rates using the MH procedure; then 2) loads by direct sampling, from their corresponding conditional posteriors. Since step (1) is similar to the parametric case, we only focus on the second step in what follows.

To generates samples for load b_i, the corresponding conditional posterior is given by [37]

$$p(b_i|\mathbf{b}\backslash b_i, \mathbf{G}, \boldsymbol{w}) \propto L(\boldsymbol{w}|\mathbf{b}, \mathbf{G})\ \mathbf{Bernoulli}\left(b_i; \frac{1}{1+\frac{M_\sigma^{max}-1}{\gamma}}\right),$$

where the backslash after $\mathbf{b}$ indicates exclusion of the following load. We may set the hyper-parameters M_σ^{max}, the maximum allowed number of system states used in computations, and γ, the expected number of system states based on simple visual inspection of the smFRET traces.

Now, the conditional posterior in the equation above is discrete and describes the probability for the load to be either active or inactive, that is, it is itself a Bernoulli distribution as follows:

$$p(b_i|\mathbf{b}\backslash b_i, \mathbf{G}, \boldsymbol{w}) = \mathbf{Bernoulli}(b_i; q_i),$$

where

$$q_i = \frac{L(\boldsymbol{w}|b_i = 1, \mathbf{b}\backslash b_i, \mathbf{G}, \boldsymbol{\rho}_{start})}{L(\boldsymbol{w}|b_i = 1, \mathbf{b}\backslash b_i, \mathbf{G}) + L(\boldsymbol{w}|b_i = 0, \mathbf{b}\backslash b_i, \mathbf{G})}.$$

The simple form of this posterior is amenable to direct sampling. In the end, the chain of generated samples can be used for subsequent statistical analysis.

4 Results

In this section, we first demonstrate the robustness of our BNP-FRET sampler by investigating the effects of excitation rate on the distributions for the transitions rates and the number of system states. Once we have illustrated the BNP-FRET sampler's performance on synthetic data, we apply it to estimate the number of system states along with the associated escape rates from publicly available experimental data for a complex involving ACTR-NCBD instrinsically disorder proteins. We compare our results with reported literature values.

4.1 Resolution of Timescales Given Excitation Rate: Nonparametrics

To demonstrate the performance of our BNP-FRET sampler over a range of timescales given an excitation rate, we follow the same approach as presented in the first companion manuscript (see Sec. 4.1). That is, we generate four synthetic smFRET traces containing $K = 2$ million photons each for a biomolecular complex with three system states, $\{\sigma_1, \sigma_2, \sigma_3\}$. To synthesize these datasets, we fix the excitation rate to be $\lambda_{ex} = 10$ ms^{-1} and FRET efficiencies ε_{FRET} to be 0.09, 0.5, and 0.9 for the three system states, respectively, motivated by experiments in [27].

The remaining parameters are the system transition rates $\lambda_{\sigma_i \to \sigma_j}$, which are varied across datasets to test our BNP-FRET sampler over a wide spectrum of timescales ranging from

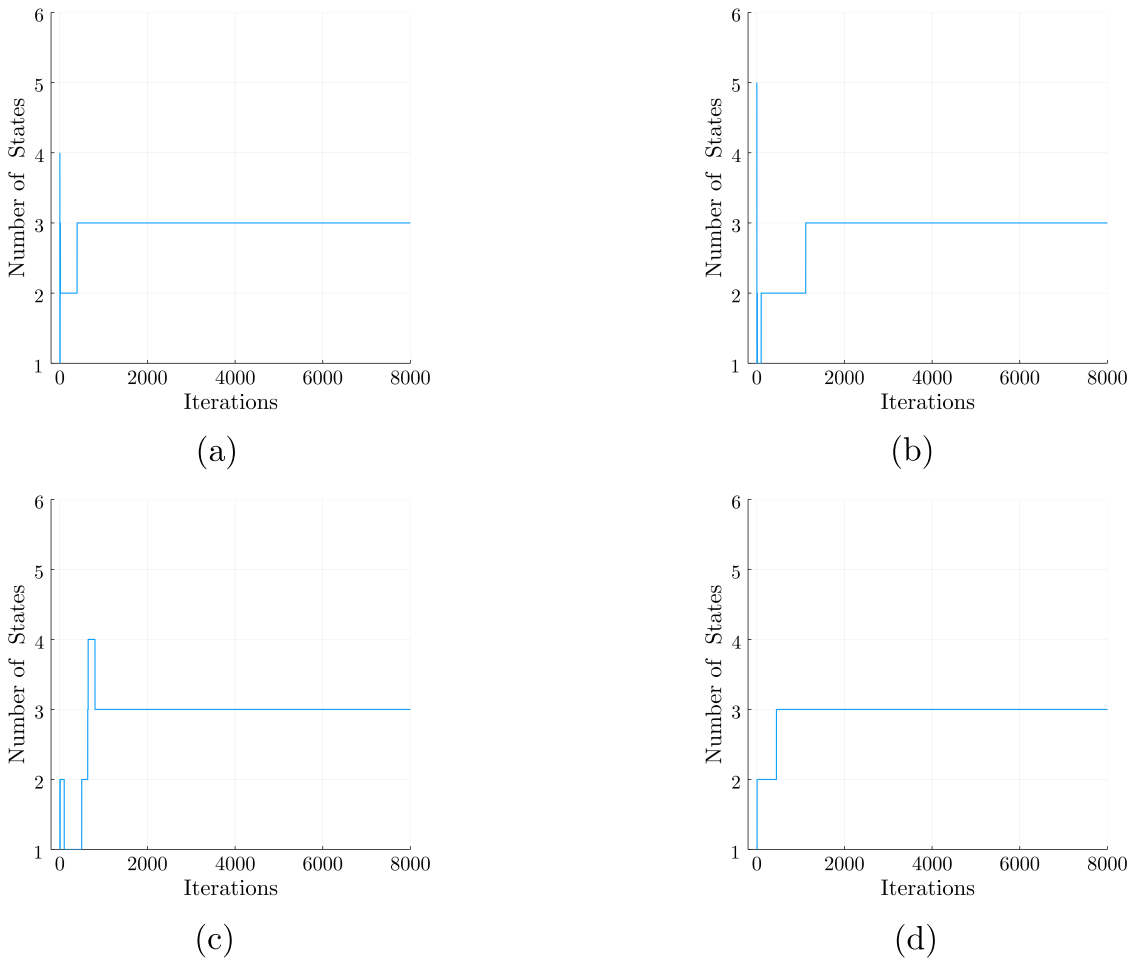


Figure 1: **MCMC chains generated by the BNP-FRET sampler for the number of system states**. The synthetic smFRET datasets used to generate these chains assume uniform excitation rate of 10 ms^{-1} and FRET efficiencies of 0.09, 0.5, and 0.9 for the three system states. However, molecular dynamics become faster by a factor of 10 as we move from panel (a) to (d). In the slowest case, we use escape rates of 0.01, 0.02, and 0.03 ms^{-1} for the three system states, while in the fastest case dynamics are as fast as the excitation rate itself. Our method converges to the correct number of system states for each dataset.

a thousand times longer than the average interphoton arrival time ($1/\lambda_{ex}$) to as short as the average interphoton arrival time itself representing an extreme case. We do not probe dynamics any faster because the excitation rate does not provide enough temporal resolution for resolving system transitions in this regime, as demonstrated in the first manuscript (see Sec. 4.1).

We start the analysis by applying our BNP-FRET sampler to learn the number of system states for the case with slowest escape rates, *i.e.* the sum of all the transition rates out of a given system state. These escape rates are $\lambda_{esc} = 0.01$, 0.02, and $0.03 ms^{-1}$. We show that our BNP-FRET sampler can correctly learn the number of system states; see Fig. 1(a).

Next, we one-by-one analyze datasets generated using escape rates that are 10 times faster in each subsequent dataset. Our nonparametric smFRET algorithm deduces the cor-

rect number of system states in all case; see Fig. 1(b)-(c). Specifically, we show that even for escape rate as fast as the excitation rate, our algorithm infers the true number of system states; see Fig. 1(d). However, the estimates for escape rates in that case have large uncertainties due to lack of temporal resolution (see panel (d) of Fig. 2).

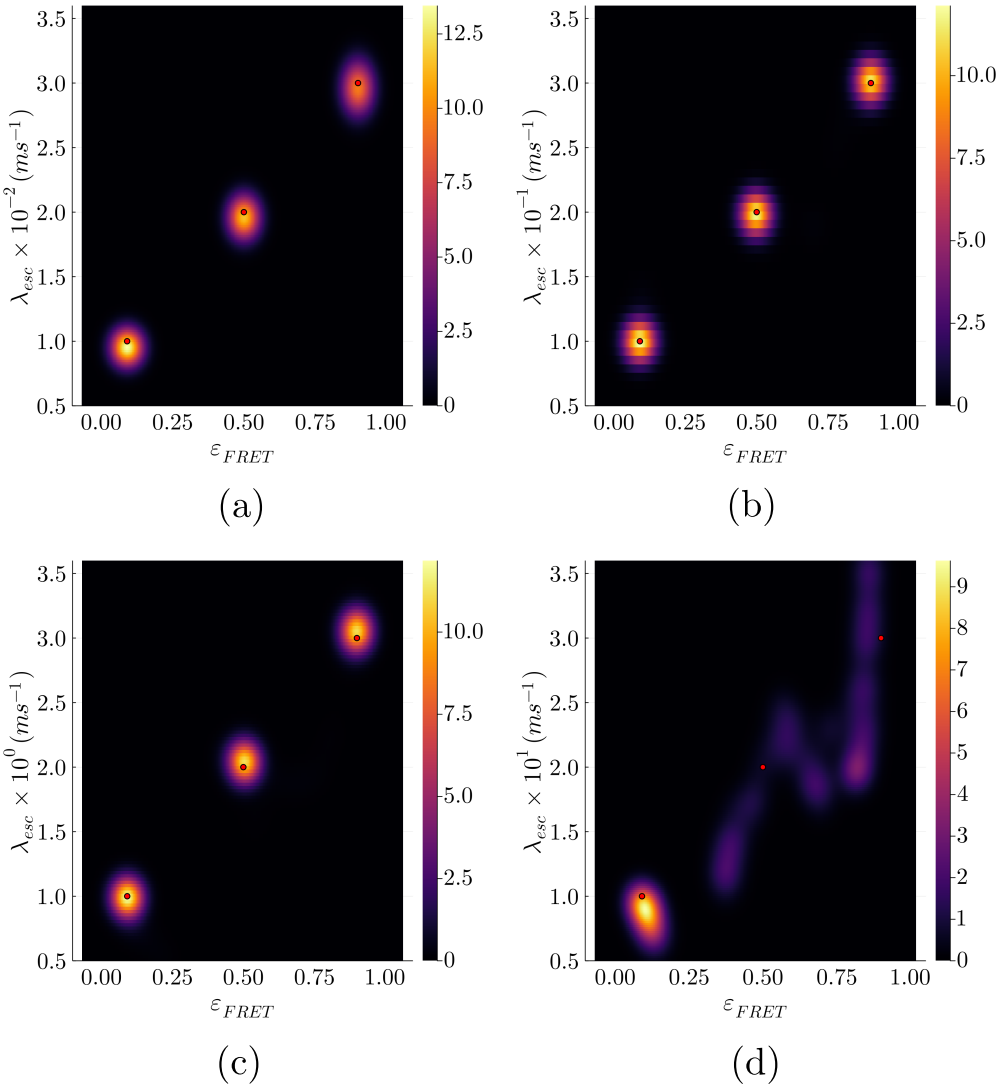


Figure 2: **Learned bivariate posterior for the escape rates λ_{esc} and FRET efficiencies ε_{FRET} from synthetic data also used in Fig. 1**. Going from panels (a) to (d), we speed up the dynamics (escape rates) by a factor of 10 each time leading to a gradual loss of temporal resolution needed to identify system transitions. Eventually, as seen in panel (d), the excitation rate does not provide enough temporal resolution to resolve system transitions occurring at the timescales of interphoton arrival times, resulting in large errors in the parameter estimates. The ground truth is shown with the red dots. We have smoothed the posterior distributions here using KDE technique available in Julia plots package.

4.2 Analysis of Experimental Data: NCBD-ACTR Interactions

Here, we apply our BNP-FRET sampler to two experimental datasets that were originally collected to study interactions between partner IDPs, NCBD and ACTR, under different environmental conditions [26, 27]. Precise knowledge of binding and unbinding reactions of such proteins is of fundamental importance toward understanding how they regulate expression of their target genes and other protein-protein reactions.

Methods that have been used in the past [26, 27] to analyze smFRET traces from experiments observing NCBD-ACTR interplay assume a fixed number of system states to obtain maximum likelihood estimates for transition rates. In addition, these methods bin photons to reduce computational expense. However, given the inherently unstructured and flexible nature of IDPs, fixing the dimensionality of the model *a priori* can be limiting and, as we will see, may bias analysis. Therefore, our nonparametric method, which puts no constraints on the number of system states while incorporating all the major sources of noise, is naturally suited for studying IDPs.

In the following subsections, we first analyze data for a system where an immobilized ACTR labeled with a Cy3B donor dye interacts with an NCBD labeled with a CF680R acceptor dye in the presence of ethylene glycol (EG), 36% by volume, to more closely mimic cellular viscosity [26]. Here, the binding of NCBD to ACTR is monitored in smFRET experiments using a confocal microscope setup. Next, we analyze data for a system with relatively faster dynamics, where an immobilized ACTR interacts with a freely-diffusing mutated NCBD (P20A) in a buffer without EG under a similar experimental setup [27].

To acquire these two experimental FRET datasets containing about 200000 photons each, laser powers of 0.5 μW and 0.3 μW were used, respectively, leading to excitation rates varying from 3000 to 11000 s^{-1} in the confocal region.

Moreover, we are provided a route correction matrix (RCM) by the authors of [26, 27] to account for spectral crosstalk, and detection efficiencies of donor and acceptor channels. We defined such an RCM in the first companion paper [21] and specify it here as

$$\mathbf{RCM}' = \frac{\mathbf{RCM}}{\mathbf{RCM}_{1,1}} = \begin{bmatrix} 1 & -\frac{\phi_{21}}{\phi_{22}} \\ -\frac{\phi_{12}}{\phi_{22}} & \frac{\phi_{11}}{\phi_{22}} \end{bmatrix} = \begin{bmatrix} 1.0 & -0.22 \\ 0.0 & 1.02 \end{bmatrix},$$

where channels 1 & 2 are, respectively, designed to receive acceptor and donor photons, and ϕ_{ij} denotes probability of photons intended for channel i to enter channel j. Since all the probabilities ϕ_{ij} appear as ratios in $\mathbf{RCM}'$, we can normalize it further such that all of the emitted donor photons are detected in one of the two channels. This also amounts to setting an effective excitation rate in a manner that none of the emitted donor photons go undetected. Simple algebraic manipulations then give the following values for the effective crosstalk factors ϕ'_{ij} as

$$\phi'_{11} = 0.84,\ \phi'_{12} = 0.0,\ \phi'_{21} = 0.18, \text{ and } \phi'_{22} = 0.82.$$

As such, these values imply that approximately 18% of the emitted donor photons are detected in the acceptor channel due to crosstalk. Furthermore, only 84% of emitted acceptor photons are detected in the acceptor channel, and acceptor photons do not suffer any crosstalk.

Now with all the experimental details at hand, we proceed to analyze the experimental data using our BNP-FRET sampler.

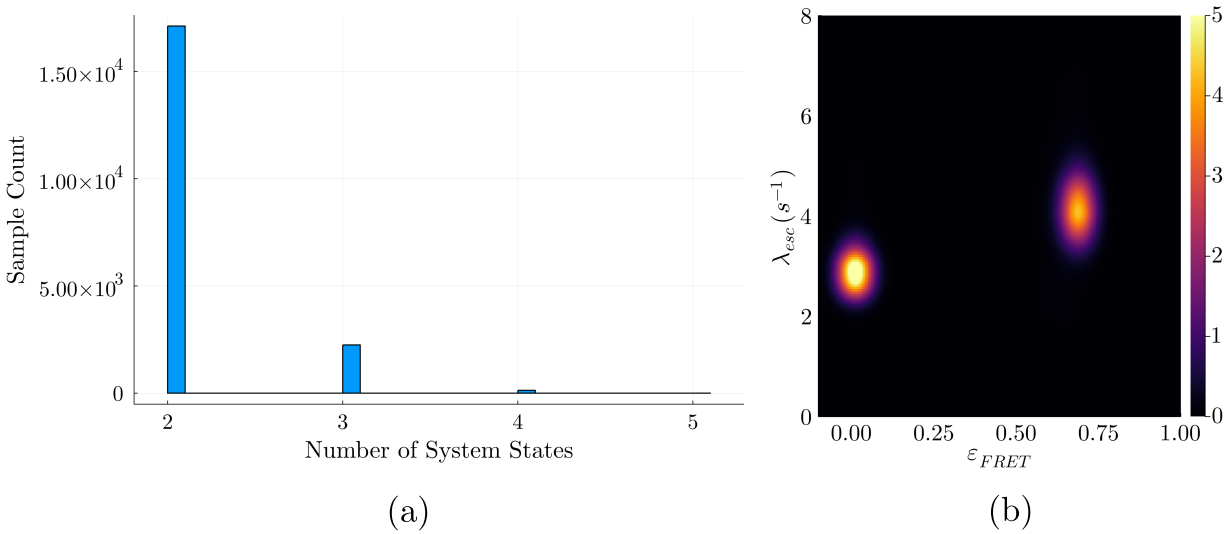


Figure 3: **Results for NCBD-ACTR interactions in the presence of ethylene glycol (EG)**. In panel (a), we show a histogram for the number of system states produced by the BNP-FRET sampler. The sampler spends a majority of its time in a model with two system states with a small possibility ascribed to more states. In the posterior distribution for the escape rates and FRET efficiencies in panel (b), two distinct FRET efficiencies are evident with values of about 0.0 (unbound) and 0.7 (bound), and corresponding escape rates of about 2.9 and 4.1 s^{-1}.

4.2.1 Immobilized ACTR in 36% EG

Binding of NCBD to ACTR leads to the formation of a stable and ordered complex in the presence of EG. In addition, when two fluorescent dyes labeling the IDPs come in close proximity, we expect FRET interactions. Therefore, bound and unbound system states of the NCBD-ACTR complex correspond to high and low FRET efficiency signals, respectively.

For the analysis of the collected smFRET data from such a complex, we must take into account all sources of noise. We have already discussed the noise due to detector effects such as crosstalk at the beginning of this section. We must also incorporate the precalibrated background rates for the donor and acceptor channels given as 0.283 s^{-1} and 0.467 s^{-1}, respectively [26].

With all such corrections applied, our BNP-FRET sampler now predicts two system states (see Fig. 3). The system state with the lowest FRET efficiency of 0.0 corresponds to the unbound NCBD. The remaining system state with a higher FRET efficiency of about 0.7 coincides with the bound configuration of the NCBD-ACTR complex. These results are in agreement with the results reported in supplementary table S1 of [26].

Furthermore, the reported transition rates learned using maximum likelihood analysis applied to a hidden Markov Model (HMM) with two system states for the case with 36% EG concentration (0.401 g/ml) are reported as $k_{on} = 2.7 \pm 0.2$ s^{-1} and $k_{off} = 3.3 \pm 0.3$ s^{-1}

[26]. The escape rates associated with these transition rates lie within the bounds of the posterior distribution obtained using our method (see Fig. 3(b)).

4.2.2 Immobilized ACTR in Buffer

As before, we expect binding of the ACTR-NCBD complex to lead to high FRET efficiency signals. However, in the absence of EG, the viscosity of the solution is lowered [26], leading to faster system transitions representing a unique analysis challenge.

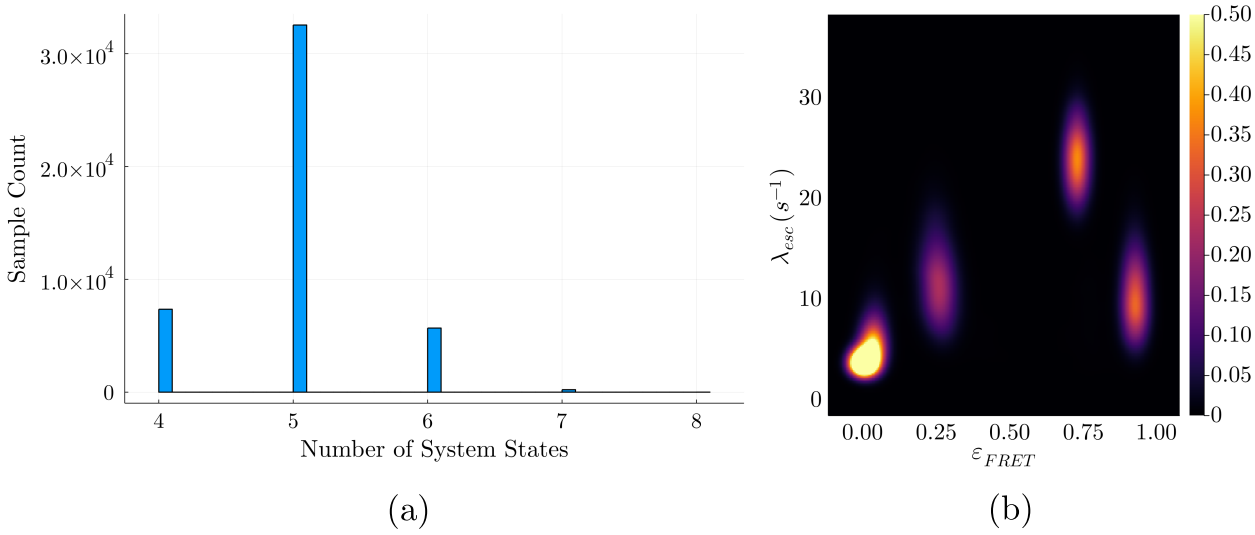
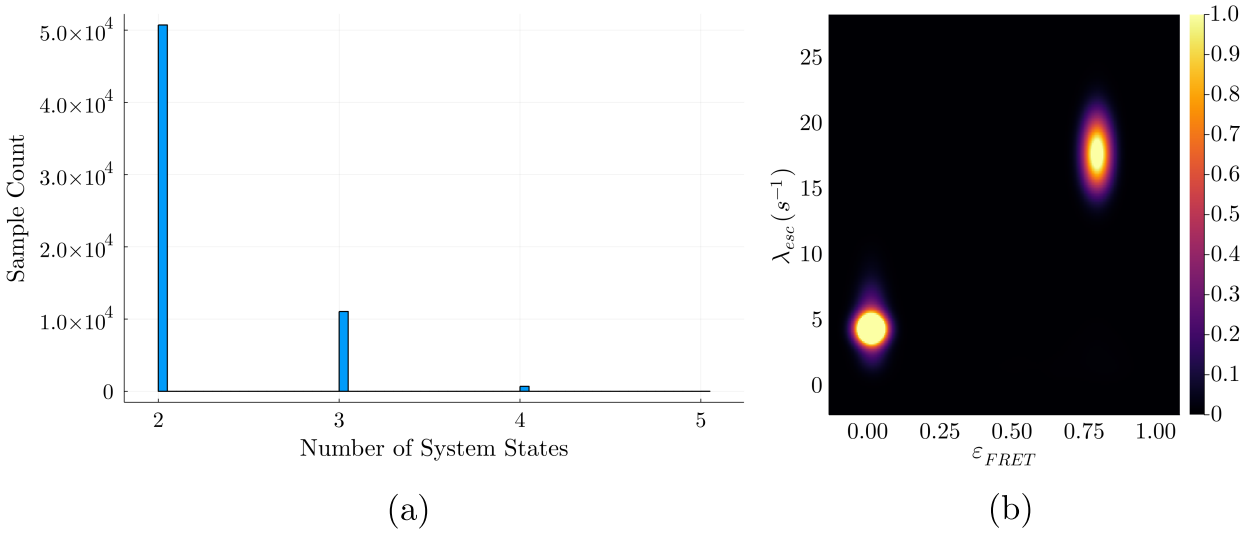


Figure 4: **Results for NCBD-ACTR interactions in buffer without EG**. In panel (a), we show a histogram produced by the BNP-FRET sampler for the number of system states. The most sampled model contains five system states. In the posterior distribution for the escape rates and FRET efficiencies in panel (b), five distinct FRET efficiencies are evident with values of 0.0, 0.03, 0.28, 0.72, and 0.92 with corresponding escape rates of about 4.6, 4.8, 11.0, 24.0, and 9.8 s^{-1}. The first two system states with almost vanishing FRET efficiencies clearly represent the same unbound configuration with the small splitting most likely a result of various sources of noise present in the dataset.

As before, after correcting for background rates of 0.312 s^{-1} and 1.561 s^{-1} for the donor and acceptor channels, respectively, our BNP-FRET sampler now predicts five system states (see Fig. 4(a)&(b)) with FRET efficiencies of 0.0, 0.03, 0.28, 0.72, and 0.92 approximately. Here, the first two system states with vanishingly small estimated FRET efficiencies, namely 0.0 and 0.03, most likely represent the same configuration where NCBD is diffusing freely away from the immobilized ACTR, leading to no FRET interactions. Various sources of noise in the dataset may have resulted in this splitting of the unbound system state. Furthermore, the system state with the FRET efficiency and escape rate of 0.72 and 24.0 s^{-1} coincides with the previously predicted bound configurations [27].

However, the two newly observed system states with FRET efficiencies of 0.28 and 0.92 with corresponding escape rates of 11.0 s^{-1} and 9.8 s^{-1} are bound configurations that were not previously detected [27]. To assure ourselves that these states are not artefactually added by our computational algorithm (overfitting), we performed additional analysis using synthetic



(a) (b)

Figure 5: **Robustness test using synthetic data to confirm no overfitting**. The synthetic data is generated under the same conditions (excitation rate, crosstalk, background, and photon budget) as the experiment and using previously reported transitions rates [27]. In panel (a), we show histogram produced by the BNP-FRET sampler for the number of system states. The most sampled model contains two system states with a small possibility ascribed to higher dimensional models. In the posterior distribution for the escape rates and FRET efficiencies in panel (b), two distinct FRET efficiencies are evident with values of 0.0, 0.8 with corresponding escape rates of about 4.5 and 18.0 s^{-1}. The absence of additional system states confirms that the additional system states encountered in the experimental results are not introduced as artefacts of the computational algorithm.

data generated under the same conditions (excitation rate, crosstalk, and background) as the experiment. That is, we simulate two system states with the previously reported escape rates [27]) of 4.3 s^{-1} and 23.0 s^{-1} with corresponding FRET efficiencies of 0.0 and 0.8, and the same photon budget of 200000 photons. The results for the analysis of this synthetic dataset in Fig. 5(a)&(b) show no additional system states introduced by our method under this parameter regime suggesting the robustness of our findings for the experimental data.

The observation of new system states of an NCBD-ACTR complex in the absence of EG suggests that effects of EG on the dynamics of IDPs are not straightforward to predict. More concretely, reduction in EG concentration leads to reduced solution viscosity and perhaps the ability to visit more transient states in a shorter period of time while under observation [39, 40]. That is, the system is more ergodic. Furthermore, IDPs interact in a complex manner with high possibility for residual secondary structures [41], which may explain the possibility of observing transient states in NCBD-ACTR complexes in lower viscosity environments otherwise undetectable to existing methods of analysis thus far. This is by contrast, to our agreement with other methods for slower dynamics in more viscous environments.

5 Discussion

FRET techniques have been an essential tool in life sciences to investigate molecular interactions on nanometer scales. For instance, FRET has been utilized to directly detect SARS-COV2 virus [42, 43]. Yet, the quantitative interpretation of smFRET data suffers from several issues including difficulties in estimating the number of system states, dealing with fast transition rates and providing uncertainties over estimations particularly uncertainties over the number of system states [44, 45].

In this paper, we implemented the general nonparametric smFRET data analysis framework presented in [21] to address the issues associated with smFRET data analysis acquired under continuous illumination. The developed framework is capable of learning posterior distributions over the number of system states as well as the corresponding dynamics ranging from slow to as fast as photon emission rates. Our method does so by propagating uncertainty from all the existing sources of noise such as instrument and background.

We benchmarked our method starting from synthetic data with three system states with a range of different timescales. We challenged our method by simulating data with dynamics as fast as the interphoton arrival times and were able to correctly deduce the number of system states even under such extreme conditions. We further assessed our method using experimental data acquired observing NCBD interacting with ACTR under different ethylene glycol (EG) concentrations that may govern the timescales at which the binding/unbinding reactions occur. From the previous point estimate analyses [26, 27], two system states are expected, respectively, for 0 and 36% EG concentrations. However, our nonparametric method predicts additional two system states in the absence of EG (fast dynamics). This observation can be directly tied to the inherently unstable nature of the two IDPs under investigations [39].

A careful treatment of how experimental noise propagates into uncertainties over the number of system states and rates does come with associated computational cost. Other methods have managed to mitigate these costs by making approximations including: 1) assuming molecular dynamics much slower compared to the fluorophore excitation and relaxation rates [9, 10]; 2) assuming dyes' fast photophysics is completely irrelevant to the system transition rate and that FRET efficiency sufficiently identifies transitions between system states [10]; 3) ignoring detector effects such as IRF and relegating other noise sources, such as background, to post-processing steps [9]; and 4) binning data [38]. In the general case without such approximations, however, the primary computation–the likelihood–remains expensive due to the required evaluation of many matrix exponentials. Rather, we must adopt other strategies to reduce computational time, such as computing likelihoods for several data traces in parallel, and minimization of the required photon budget via spatial and temporal variation of excitation rate.

The method described in this paper was developed for cases with discrete system state spaces. However, it can be extended for continuous system state spaces by modifying the likelihood in Eq. 3. Furthermore, our method can be modified to deal with extremely short lifetimes by incorporating the IRF into the likelihood model. Moreover, our framework is flexible to be adopted for different illumination modalities such as alternating laser excitation (ALEX) [46] to directly excited both donor and acceptor dyes. This would facilitate the accurate estimation of the crosstalk factors as well as detection efficiencies and quantum

yield of the dyes.

6 Acknowledgments

We thank Weiqing Xu and Dr Zeliha Kilic for regular feedback and help, especially during the development of the nonparametrics samplers. We also thank Dr Benjamin Schuler, Dr Daniel Nettels, and Oliver Stach for regular feedback and providing the necessary experimental data. S. P. acknowledges support from the NIH NIGMS (R01GM130745) for supporting early efforts in nonparametrics and NIH NIGMS (R01GM134426) for supporting single-photon efforts.

Bibliography

[1] Eitan Lerner, Anders Barth, Jelle Hendrix, Benjamin Ambrose, Victoria Birkedal, Scott C Blanchard, Richard Börner, Hoi Sung Chung, Thorben Cordes, Timothy D Craggs, Ashok A Deniz, Jiajie Diao, Jingyi Fei, Ruben L Gonzalez, Irina V Gopich, Taekjip Ha, Christian A Hanke, Gilad Haran, Nikos S Hatzakis, Sungchul Hohng, Seok-Cheol Hong, Thorsten Hugel, Antonino Ingargiola, Chirlmin Joo, Achillefs N Kapanidis, Harold D Kim, Ted Laurence, Nam Ki Lee, Tae-Hee Lee, Edward A Lemke, Emmanuel Margeat, Jens Michaelis, Xavier Michalet, Sua Myong, Daniel Nettels, Thomas-Otavio Peulen, Evelyn Ploetz, Yair Razvag, Nicole C Robb, Benjamin Schuler, Hamid Soleimaninejad, Chun Tang, Reza Vafabakhsh, Don C Lamb, Claus AM Seidel, and Shimon Weiss. Fret-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *eLife*, 10:e60416, 2021.

[2] James J McCann, Ucheor B Choi, Liqiang Zheng, Keith Weninger, and Mark E Bowen. Optimizing methods to recover absolute fret efficiency from immobilized single molecules. *Biophysical journal*, 99(3):961–970, 2010.

[3] Ioannis Sgouralis, Shreya Madaan, Franky Djutanta, Rachael Kha, Rizal F Hariadi, and Steve Pressé. A bayesian nonparametric approach to single molecule forster resonance energy transfer. *The Journal of Physical Chemistry B*, 123(3):675–688, 2018.

[4] Jae-Yeol Kim, Cheolhee Kim, and Nam Ki Lee. Real-time submillisecond single-molecule fret dynamics of freely diffusing molecules with liposome tethering. *Nature Communications*, 6(1):6992, 2015.

[5] Benjamin Schuler and Hagen Hofmann. Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Current Opinion in Structural Biology*, 23(1):36–47, 2013.

[6] Y. Wu, X. Wu, R. Lu, M. Li, L. Toro, and E. Stefani. Super-resolution light microscopy: Stimulated emission depletion and ground-state depletion. In Ralph A. Bradshaw and Philip D. Stahl, editors, *Encyclopedia of Cell Biology*, pages 76–85. Academic Press, Waltham, 2016.

[7] Christian Eggeling, Jerker Widengren, Leif Brand, Jörg Schaffer, Suren Felekyan, and Claus A. M. Seidel. Analysis of photobleaching in single-molecule multicolor excitation and förster resonance energy transfer measurements. *The Journal of Physical Chemistry A*, 110(9):2979–2995, 2006.

[8] Colton Boudreau, Tse-Luen (Erika) Wee, Yan-Rung (Silvia) Duh, Melissa P. Couto, Kimya H. Ardakani, and Claire M. Brown. Excitation light dose engineering to reduce photo-bleaching and photo-toxicity. *Scientific Reports*, 6(1):30892, 2016.

[9] Irina V. Gopich and Attila Szabo. Decoding the Pattern of Photon Colors in Single-Molecule FRET. *The Journal of Physical Chemistry B*, 113(31):10965–10973, 2009.

[10] Irina Gopich and Attila Szabo. Theory of photon statistics in single-molecule förster resonance energy transfer. *The Journal of Chemical Physics*, 122(1):014707, 2005.

[11] Sean A. McKinney, Chirlmin Joo, and Taekjip Ha. Analysis of single-molecule fret trajectories using hidden markov modeling. *Biophysical journal*, 91(5):1941–1951, 2006.

[12] Zeliha Kilic, Ioannis Sgouralis, Wooseok Heo, Kunihiko Ishii, Tahei Tahara, and Steve Pressé. Extraction of rapid kinetics from smfret measurements using integrative detectors. *Cell Reports Physical Science*, 2(5):100409, 2021.

[13] Zeliha Kilic, Ioannis Sgouralis, and Steve Pressé. Generalizing hmms to continuous time for fast kinetics: Hidden markov jump processes. *Biophysical Journal*, 120(3):409–423, 2021.

[14] Zeliha Kilic, Ioannis Sgouralis, Wooseok Heo, Kunihiko Ishii, Tahei Tahara, and Steve Pressé. A continuous time representation of smfret for the extraction of rapid kinetics. *Biophysical Journal*, 120(3):186a, 2021.

[15] Menahem Pirchi, Roman Tsukanov, Rashid Khamis, Toma E. Tomov, Yaron Berger, Dinesh C. Khara, Hadas Volkov, Gilad Haran, and Eyal Nir. Photon-by-photon hidden markov model analysis for microsecond single-molecule fret kinetics. *The Journal of Physical Chemistry B*, 120(51):13065–13075, 2016.

[16] Daniel Nettels, Irina V. Gopich, Armin Hoffmann, and Benjamin Schuler. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proceedings of the National Academy of Sciences*, 104(8):2655–2660, 2007.

[17] Irina V. Gopich and Attila Szabo. Single-molecule fret with diffusion and conformational dynamics. *The Journal of Physical Chemistry B*, 111(44):12925–12932, 2007.

[18] Janghyun Yoo, Jae-Yeol Kim, John M. Louis, Irina V. Gopich, and Hoi Sung Chung. Fast three-color single-molecule fret using statistical inference. *Nature Communications*, 11(1):3336, 2020.

[19] Janghyun Yoo, John M. Louis, Irina V. Gopich, and Hoi Sung Chung. Three-color single-molecule fret and fluorescence lifetime analysis of fast protein folding. *The Journal of Physical Chemistry B*, 122(49):11702–11720, 2018.

[20] Irina V. Gopich and Attila Szabo. Single-Macromolecule Fluorescence Resonance Energy Transfer and Free-Energy Profiles. *The Journal of Physical Chemistry B*, 107(21):5058–5063, 2003.

[21] Ayush Saurabh, Matthew Safar, Ioannis Sgouralis, Mohamadreza Fazel, and Steve Pressé. Single photon smFRET. I. theory and conceptual basis. *In preparation.*

[22] K.V. Mardia, H.R. Southworth, and C.C. Taylor. On bias in maximum likelihood estimators. *Journal of Statistical Planning and Inference*, 76(1):31–39, 1999.

[23] Rolf Sundberg. Flat and multimodal likelihoods and model lack of fit in curved exponential families. *Scandinavian Journal of Statistics*, 37(4):632–643, 2010.

[24] Malcolm Sambridge. A Parallel Tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International*, 196(1):357–374, 2013.

[25] Ayush Saurabh, Matthew Safar, and Steve Pressé. BNP-FRET: A software suite to analyze smFRET data using bayesian nonparametrics. *In preparation. Github link to appear.*

[26] Franziska Zosel, Andrea Soranno, Karin J. Buholzer, Daniel Nettels, and Benjamin Schuler. Depletion interactions modulate the binding between disordered proteins in crowded environments. *Proceedings of the National Academy of Sciences*, 117(24):13480–13489, 2020.

[27] Franziska Zosel, Davide Mercadante, Daniel Nettels, and Benjamin Schuler. A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nature Communications*, 9(1):3332, 2018.

[28] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[29] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

[30] Christopher M. Bishop. Pattern recognition. *Machine learning*, 128(9), 2006.

[31] J Shepard Bryan IV, Ioannis Sgouralis, and Steve Pressé. Diffraction-limited molecular cluster quantification with bayesian nonparametrics. *Nature Computational Science*, 2(2):102–111, 2022.

[32] Mohamadreza Fazel, Michael J Wester, Hanieh Mazloom-Farsibaf, Marjolein Meddens, Alexandra S Eklund, Thomas Schlichthaerle, Florian Schueder, Ralf Jungmann, and Keith A Lidke. Bayesian multiple emitter fitting using reversible jump markov chain monte carlo. *Scientific reports*, 9(1):1–10, 2019.

[33] Sina Jazani, Ioannis Sgouralis, Omer M Shafraz, Marcia Levitus, Sanjeevi Sivasankar, and Steve Pressé. An alternative framework for fluorescence correlation spectroscopy. *Nature communications*, 10(1):1–10, 2019.

[34] Mohamadreza Fazel, Sina Jazani, Lorenzo Scipioni, Alexander Vallmitjana, Enrico Gratton, Michelle A Digman, and Steve Pressé. High resolution fluorescence lifetime maps from minimal photon counts. *ACS Photonics*, 9(3):1015–1025, 2022.

[35] Andrew Gelman, Walter R Gilks, and Gareth O Roberts. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120, 1997.

[36] Ioannis Sgouralis, Shreya Madaan, Franky Djutanta, Rachael Kha, Rizal F. Hariadi, and Steve Pressé. A Bayesian Nonparametric Approach to Single Molecule Förster Resonance Energy Transfer. *The Journal of Physical Chemistry B*, 123(3):675–688, 2019.

[37] Ioannis Sgouralis and Steve Pressé. An introduction to infinite hmms for single-molecule data analysis. *Biophysical Journal*, 112(10):2021–2029, 2017.

[38] Zeliha Kilic, Ioannis Sgouralis, and Steve Pressé. Generalizing HMMs to continuous time for fast kinetics: Hidden Markov jump processes. *Biophysical Journal*, 120(3):409–423, 2021.

[39] Andrea Soranno, Iwo Koenig, Madeleine B Borgia, Hagen Hofmann, Franziska Zosel, Daniel Nettels, and Benjamin Schuler. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proceedings of the National Academy of Sciences*, 111(13):4874–4879, 2014.

[40] Daniel Johansen, Cy MJ Jeffries, Boualem Hammouda, Jill Trewhella, and David P Goldenberg. Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching. *Biophysical journal*, 100(4):1120–1128, 2011.

[41] Intrinsically disordered protein exhibits both compaction and expansion under macromolecular crowding. *Biophysical Journal*, 114(5):1067–1079, 2018.

[42] Maolin Lu. Single-molecule fret imaging of virus spike–host interactions. *Viruses*, 13(2):332, 2021.

[43] Byunghoon Kang, Youngjin Lee, Jaewoo Lim, Dongeun Yong, Young Ki Choi, Sun Woo Yoon, Seungbeom Seo, Soojin Jang, Seong Uk Son, Taejoon Kang, et al. Fret-based hace2 receptor mimic peptide conjugated nanoprobe for simple detection of sars-cov-2. *Chemical Engineering Journal*, 442:136143, 2022.

[44] Sean A. McKinney, Chirlmin Joo, and Taekjip Ha. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophysical Journal*, 91(5):1941–1951, 2006.

[45] Jonathan E. Bronson, Jingyi Fei, Jake M. Hofman, Ruben L. Gonzalez, Jr., and Chris H. Wiggins. Learning rates and states from biophysical time series: A Bayesian approach to model selection and single-molecule FRET data. *Biophysical Journal*, 97(12):3196–3205, 2009.

[46] Achillefs N Kapanidis, Ted A Laurence, Nam Ki Lee, Emmanuel Margeat, Xiangxu Kong, and Shimon Weiss. Alternating-laser excitation of single molecules. *Accounts of chemical research*, 38(7):523–533, 2005.