

Bayesian Inference of Binding Kinetics from Fluorescence Time Series

J. Shepard Bryan IV^{*1}, Stanimir Asenov Tashev^{*2,3,4}, Mohamadreza Fazel¹, Michael Scheckenbach⁵, Philip Tinnefeld⁵, Dirk-Peter Herten^{2,**}, and Steve Pressé^{1,6,**}

¹Department of Physics, Arizona State University

²College of Medical and Dental Sciences, University of Birmingham

³School of Chemistry, University of Birmingham

⁴Centre of Membrane Proteins and Receptors (COMPARE), Universities of Birmingham and Nottingham

⁵Faculty of Chemistry and Pharmacy, Ludwig-Maximilians-University Munich

⁶School of Molecular Sciences, Arizona State University

*These authors made equal contributions

**d.herten@bham.ac.uk, spresse@asu.edu

February 3, 2025

Abstract

The study of binding kinetics via the analysis of fluorescence time traces is often confounded by measurement noise and photophysics. Although photoblinking can be mitigated by using labels less likely to photoswitch, photobleaching generally cannot be eliminated. Current methods for measuring binding and unbinding rates are therefore limited by concurrent photobleaching events. Here, we propose a method to infer binding and unbinding rates alongside photobleaching rates using fluorescence intensity traces. Our approach is a two-stage process involving analyzing individual regions of interest (ROIs) with a Hidden Markov Model to infer the fluorescence intensity levels of each trace. We then use the inferred intensity level state trajectory from all ROIs to infer kinetic rates. Our method has several advantages, including the ability to analyze noisy traces, account for the presence of photobleaching events, and provide uncertainties associated with the inferred binding kinetics. We demonstrate the effectiveness and reliability of our method through simulations and data from DNA origami binding experiments.

1 Introduction

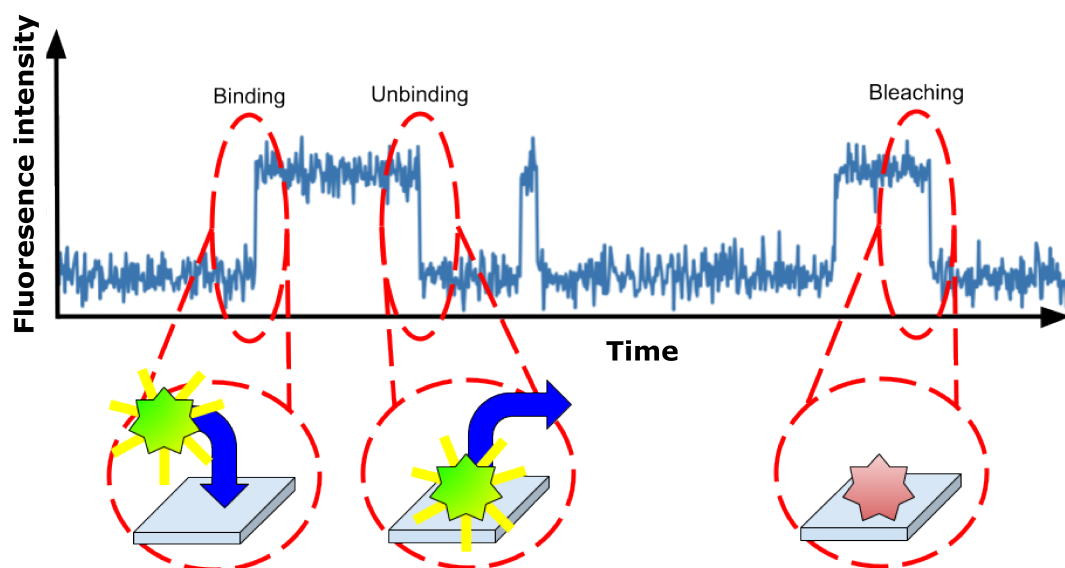


Figure 1: Illustration of a fluorescence intensity time trace. A hypothetical fluorescence intensity trace from a fluorescent ligand binding experiment. Circles are drawn around the three key regions indicating a binding event, an unbinding event, and a photobleaching event, respectively. Of note is the inability to distinguish unbinding from bleaching by eye.

The study of binding kinetics is critical to understanding the behavior of biological systems and chemical reactions at the molecular level. For instance, accurate determination of binding kinetics can provide valuable insights into the mechanisms of protein self-assembly [1, 2, 3, 4], enzymatic and protein-protein interactions [5, 6, 7, 8, 9, 10], DNA-protein interactions and gene expression [11, 12, 13], nuclear breakdown during mitosis [14], nuclear pore complex (NPC) assembly [15, 16, 17], and the signaling mechanism and cell response via receptor clustering and protein recruitment [18, 19] among others. Likewise, binding rates are essential to understand and control reactions in chemistry, including but not limited to nanocatalysis [20, 21], polymerization catalysis [22] and fundamental reaction mechanisms [23, 24, 25]. Currently, several techniques exist for measuring binding kinetics, including surface plasmon resonance [26, 27], and fluorescence correlation spectroscopy [28, 29, 30, 31], among others [32]. However, these methods have limitations, such as the difficulty in analyzing data in the low signal-to-noise ratio (SNR) regime or their inability to account for photophysical events including photobleaching, or simply termed bleaching, which may be confounded with unbinding.

Quantifying the stoichiometric fluctuations of clusters [33, 34, 35, 36, 37, 38, 39] is another important motivation for studying binding kinetics. Here, super-resolution methods such as PALM, dSTORM, and qPAINT provide means to estimate the stoichiometry of proteins clusters [40, 41, 42, 43] albeit mostly limited to static structures, *e.g.*, fixed cells. Hidden Markov models [44] can be used to infer transition probabilities, but in the presence of photophysics, these do not directly translate to rates which becomes relevant when studying faster events. Moreover, current methods for inferring binding kinetics are prone to deviations, such as over-

estimating unbinding probabilities, in their estimates due to binding kinetics and photophysical effects, such as photobleaching (figure 1).

Here, we propose a method to achieve accurate binding rates while considering photobleaching. We focus on inferring rates instead of transition probabilities because rates allow us to disentangle the kinetic contributions of binding from bleaching [45, 46]. Concretely, to model the stoichiometry of assemblies, we assume two photophysical states (bright and photobleached) and two binding states (bound and unbound) for each monomer. An illustration of our system can be found in figure 1. Our method determines the stoichiometry of assemblies by inferring binding rates alongside the photobleaching rates from intensity fluctuations over time. To achieve this, our method analyzes the sum of intensities of pixels around individual regions of interest (ROI) to infer state trajectory of the system, including both photobleaching and binding states for each ROI alongside all the rates. Such intensity traces are often given as a convolution of noisy measurement of photon counts over time due to photon shot noise and noises intrinsic to the detector devices such as readout noise [47]. Moreover, the kinetics of binding and fluorophore photobleaching are inherently random processes leading to additional layers of noise. We are able to separate the unbinding rates from photobleaching rates by utilizing information about the laser intensity of each trace. As such, we adopt a Bayesian framework to account for all the sources of uncertainty inherent to the problem. In particular, we leverage the Hidden Markov model (HMM) within a Bayesian framework to deal with the input noisy intensity traces [48, 49, 44, 50].

For the sake of computational efficiency, our framework consists of two parts or modules: in the first part, we learn the state trajectories using the noisy input data; in the second part, we use the inferred state trajectories to learn the rates. An important advantage of this two-step framework is that it allows us to simultaneously analyze different ROIs with different intensity statistics, drastically speeding up computation through parallelization. In this work, we illustrate the two-stage mathematical framework of our method and construct the likelihood and posterior. After which, we focus on the inverse model and benchmark our method using synthetic data, and *in vitro* experimental data obtained from DNA origami experiments.

2 Methods

2.1 Computational Methods

Our computational method is divided into two modules: State Inference and Rate Inference. In State Inference, for each individual ROI we determine how many fluorophores are present in the bright state at each time point given the intensity trace of the ROI. We then consolidate and pass these results to our Rate Inference step, where we use the state trajectories to infer the binding rates and photobleaching rates for the system. While in principle we could combine these two steps into a single analysis, inputting raw intensity traces to output rates, we break them up to drastically speed up computation via parallelization. The loss in accuracy due to splitting the analysis into two stages will be minimal, so long as data are in a regime where the brightness steps are clear in the intensity traces. On the other hand, if the SNR in the data is too low, then our analysis may fail, for example, if the standard deviation of the noise is much larger than the size of a step.

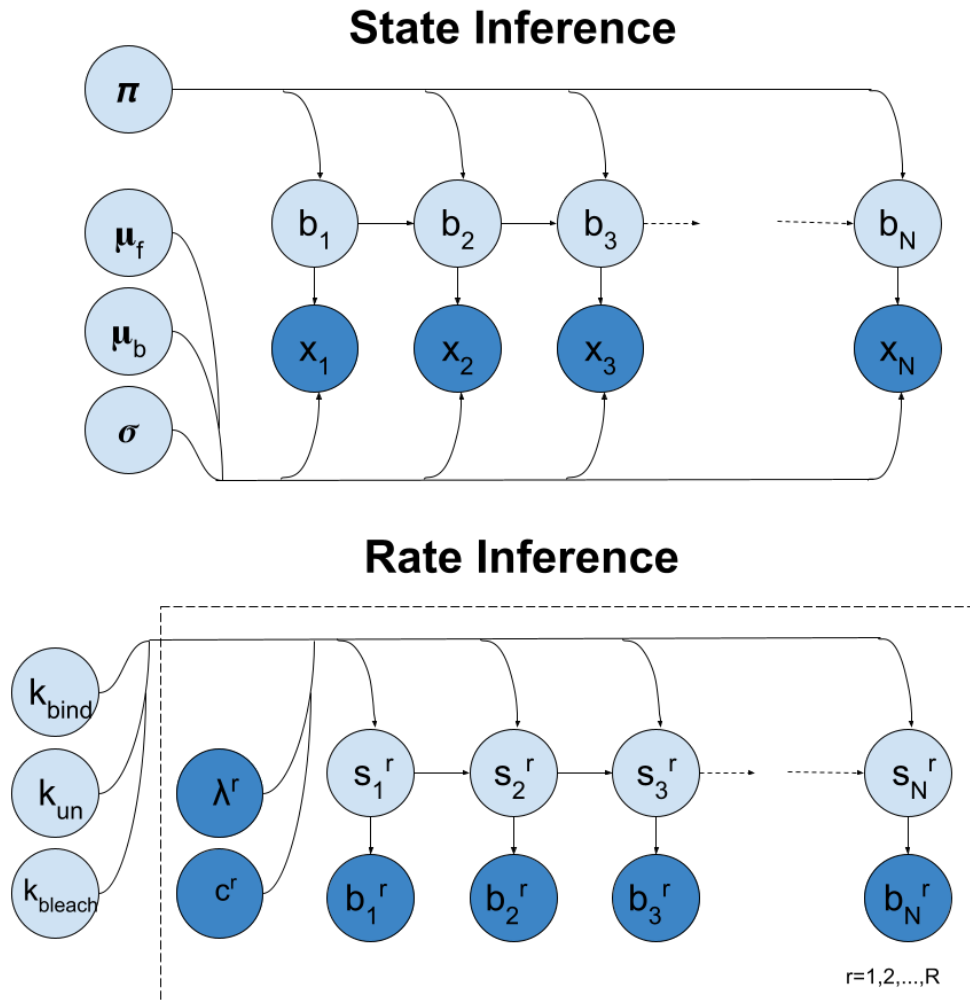


Figure 2: Graphical model of our analysis. Nodes (circles) represent random variables. Dark nodes represent measurements. Arrows between nodes represent conditional dependence. Plates (dashed rectangles) around nodes represent groupings of variables, which are repeated along the index in the lower right corner. In the top of this figure we present our graphical model for State Inference. In the bottom of this figure we present our graphical model for Rate Inference. All variables are defined in the main body.

2.1.1 State Inference

In the State Inference step of our analysis, we take in an intensity trace and output the state trajectory, which is the number of bright fluorophores at each time point. We define $\mathbf{x}_{1:N}$, the “brightness trace”, to be the ROI’s intensity at each time level with x_n the intensity at time n . We also define $\mathbf{b}_{1:N}$, which we will refer to as the “state trace”, to represent the number of bright fluorophores across the time points of the measurement where b_n is the number of bright

fluorophores in the ROI at time n . Our goal in the State inference step is to find the most probable state trace, $\mathbf{b}_{1:N}$, given a brightness trace, $\mathbf{x}_{1:N}$.

In order to find the most probable state trace, $\mathbf{b}_{1:N}$, given a brightness trace, $\mathbf{x}_{1:N}$, we use Bayes' theorem

$$\mathcal{P}(\mathbf{b}_{1:N}, \Theta | \mathbf{x}_{1:N}) \propto \mathcal{P}(\mathbf{x}_{1:N} | \mathbf{b}_{1:N}, \Theta) \mathcal{P}(\mathbf{b}_{1:N} | \Theta) \mathcal{P}(\Theta) \quad (1)$$

where Θ are model parameters and we split the posterior, $\mathcal{P}(\mathbf{b}_{1:N}, \Theta | \mathbf{x}_{1:N})$, into a likelihood, $\mathcal{P}(\mathbf{x}_{1:N} | \mathbf{b}_{1:N}, \Theta)$, a prior over brightness trace, $\mathcal{P}(\mathbf{b}_{1:N} | \Theta)$, and a prior over model parameters defined shortly, $\mathcal{P}(\Theta)$. For the likelihood, we assume each measurement independent and normally distributed around a mean given by the background plus the number of bright fluorophores,

$$\mathcal{P}(\mathbf{x}_{1:N} | \mathbf{b}_{1:N}, \Theta) = \mathcal{P}(x_{1:N} | b_{1:N}, \mu_f, \mu_b, \sigma^2) \quad (2)$$

$$= \prod_{n=1}^N \mathcal{P}(x_n | b_n, \mu_f, \mu_b, \sigma^2) \quad (3)$$

$$= \prod_{n=1}^N \mathbf{Normal}(x_n; \mu_b + \mu_f b_n, \sigma^2) \quad (4)$$

where we have introduced three variables: the fluorophore brightness, μ_f , the background brightness, μ_b , and the noise σ^2 . We will model the prior over fluorescence intensity trace as a Hidden Markov Model in which each state is categorically distributed according to the state at the previous time level [44],

$$\mathcal{P}(\mathbf{b}_{1:N} | \Theta) = \mathcal{P}(\mathbf{b}_{1:N} | \boldsymbol{\pi}, \boldsymbol{\pi}_0) \quad (5)$$

$$= \mathcal{P}(b_1 | \boldsymbol{\pi}_0) \prod_{n=2}^N \mathcal{P}(b_n | b_{n-1}, \boldsymbol{\pi}) \quad (6)$$

$$= \mathbf{Categorical}(b_1; \boldsymbol{\pi}_0) \prod_{n=2}^N \mathbf{Categorical}(b_n; \boldsymbol{\pi}_{b_{n-1}}) \quad (7)$$

where we have introduced two variables: the state transition matrix, $\boldsymbol{\pi}$, and an array of starting probabilities $\boldsymbol{\pi}_0$. The quantity $\boldsymbol{\pi}_{b_{n-1}}$ is interpreted as the ‘‘row of $\boldsymbol{\pi}$ corresponding to the state at time level $n - 1$ ’’. Each element, π_{ij} , of $\boldsymbol{\pi}$ represents the probability that a time step will have j fluorophores given that the previous time step had i fluorophores. While in principle, there could be an unlimited number of fluorophores in the ROI at a given time step, in practice we will cap the possible number of fluorophores to M , meaning that $\boldsymbol{\pi}$ is a $M + 1$ by $M + 1$ matrix (since we include 0 fluorophores as a possibility). The full set of parameters can now be defined to include $\Theta = \{\mu_f, \mu_b, \sigma^2, \boldsymbol{\pi}_0, \boldsymbol{\pi}\}$.

Working in the Bayesian paradigm, we must assign priors over the variables: μ_f , μ_b , σ^2 , $\boldsymbol{\pi}_0$, and $\boldsymbol{\pi}$. We choose a Gaussian over μ_f and μ_b , an inverse gamma for σ^2 , and Dirichlet distributions for $\boldsymbol{\pi}_0$ and each row of $\boldsymbol{\pi}$, since they are the suitable conjugate priors of their respective likelihood distributions [44],

$$\begin{aligned} \mathcal{P}(\Theta) &= \mathcal{P}(\mu_f, \mu_b, \sigma^2, \boldsymbol{\pi}_0, \boldsymbol{\pi}) \\ &= \mathcal{P}(\mu_f) \mathcal{P}(\mu_b) \mathcal{P}(\sigma^2) \mathcal{P}(\boldsymbol{\pi}_0) \mathcal{P}(\boldsymbol{\pi}) \end{aligned} \quad (8)$$

where the individual priors are given by

$$\mathcal{P}(\mu_f) = \mathbf{Normal}(\mu_f; \alpha_{\mu_f}, \beta_{\mu_f}) \quad (9)$$

$$\mathcal{P}(\mu_b) = \mathbf{Normal}(\mu_b; \alpha_{\mu_b}, \beta_{\mu_b}) \quad (10)$$

$$\mathcal{P}(\sigma^2) = \mathbf{InvGamma}(\sigma^2; \alpha_{\sigma^2}, \beta_{\sigma^2}) \quad (11)$$

$$\mathcal{P}(\boldsymbol{\pi}_0) = \mathbf{Dirichlet}(\boldsymbol{\pi}_0; \boldsymbol{\alpha}_0) \quad (12)$$

$$\mathcal{P}(\boldsymbol{\pi}_i) = \mathbf{Dirichlet}(\boldsymbol{\pi}_i; \boldsymbol{\alpha}_i) \quad \text{for } m \text{ in } 1, 2, \dots, M \quad (13)$$

and where the α 's and β 's are hyperparameters. Altogether, the posterior over $\mathbf{b}_{1:N}$ and Θ now reads

$$\begin{aligned} \mathcal{P}(\mathbf{b}_{1:N}, \Theta | \mathbf{x}_{1:N}) &\propto \mathcal{P}(\mathbf{x}_{1:N} | \mathbf{b}_{1:N}, \mu_f, \mu_b, \sigma^2) \mathcal{P}(\mathbf{b}_{1:N} | \boldsymbol{\pi}, \boldsymbol{\pi}_0) \mathcal{P}(\mu_f, \mu_b, \sigma^2, \boldsymbol{\pi}_0, \boldsymbol{\pi}) \quad (14) \\ &= \mathbf{Normal}(x_1; \mu_b + \mu_f b_1, \sigma^2) \mathbf{Categorical}(b_1; \boldsymbol{\pi}_0) \\ &\quad \times \left(\prod_{n=2}^N \mathbf{Categorical}(b_n; \boldsymbol{\pi}_{b_{n-1}}) \mathbf{Normal}(x_n; \mu_b + \mu_f b_n, \sigma^2) \right) \\ &\quad \times \mathbf{Normal}(\mu_f; \alpha_{\mu_f}, \beta_{\mu_f}) \mathbf{Normal}(\mu_b; \alpha_{\mu_b}, \beta_{\mu_b}) \\ &\quad \times \mathbf{InvGamma}(\sigma^2; \alpha_{\sigma^2}, \beta_{\sigma^2}) \\ &\quad \times \left(\prod_{m=0}^M \mathbf{Dirichlet}(\boldsymbol{\pi}_i; \boldsymbol{\alpha}_i) \right). \quad (15) \end{aligned}$$

In the top of figure 2, we present our posterior as a graphical model [44]. In the graphical model, nodes (circles) represent random variables, and arrows represent conditional dependence.

We wish to find the value of $\mathbf{b}_{1:N}$ that maximize equation (15), which we will refer to as the maximum a posteriori (MAP) sample. We will do so using Gibbs sampling [44], where we start with an initial guess then iteratively propose new values for variables, accepting and rejecting according to a ratio of the probabilities. We use Gibbs sampling, as opposed to optimization, to avoid local minima in our posterior. The algorithm is as follows:

- Start with initial set of variables, $\Theta^{(0)} = \{\mathbf{b}_{1:N}^{(0)}, \mu_f^{(0)}, \mu_b^{(0)}, \sigma^{2(0)}, \boldsymbol{\pi}_0^{(0)}, \boldsymbol{\pi}^{(0)}\}$.
- Set the MAP sample to this value, $\Theta_{MAP} = \Theta^{(0)}$.
- For many iterations i :
 - Propose a new set of parameters Θ' based on a proposal distribution $q(\Theta' | \Theta^{(i-1)})$.
 - Calculate the acceptance ratio a , defined as:

$$a = \min \left(1, \frac{P(\Theta') q(\Theta^{(i-1)} | \Theta')}{P(\Theta^{(i-1)}) q(\Theta' | \Theta^{(i-1)})} \right) \quad (16)$$

where $P(\Theta)$ represents the posterior probability of the parameters and $q(\Theta_a | \Theta_b)$ represents the proposal distribution.

- Accept or reject the new parameter set Θ' with probability a .
- Compare $\Theta^{(i)}$ to Θ_{MAP} . If $P(\Theta^{(i)}) > P(\Theta_{MAP})$, update $\Theta_{MAP} = \Theta^{(i)}$.
- The final Θ_{MAP} is the estimate for the maximum a posteriori value.

The final $\mathbf{b}_{1:N}^{MAP}$ is the state trace. All the state traces are then fed into the next step, Rate Inference.

2.1.2 Rate Inference

Once we have state traces for each ROI, we are ready to estimate the binding and photobleaching rates governing transitions. Here we assume there are R ROIs, with state traces $\mathbf{b}_{1:N}^{1:R}$ where $\mathbf{b}_{1:N}^r$ is the state trace of ROI r and b_n^r is the number of fluorophores in ROI r at time level n . We wish to infer the binding rate, k_{bind} , the unbinding rate, k_{un} , and the photobleaching rate, k_{bleach} given the state traces, $\mathbf{b}_{1:N}^{1:R}$, as well as the laser intensity of each trace, $\boldsymbol{\lambda}^{1:R}$, and the concentration of binding agent at each trace, $\mathbf{c}^{1:R}$.

We again invoke Bayes' theorem to decompose the probability of our target rates, k_{bind} , k_{un} , and k_{bleach} , given the state traces, $\mathbf{b}_{1:N}^{1:R}$ into a likelihood and prior

$$\mathcal{P}(\mathbf{k} | \mathbf{b}_{1:N}^{1:R}, \boldsymbol{\lambda}^{1:R}, \mathbf{c}^{1:R}) \propto \mathcal{P}(\mathbf{b}_{1:N}^{1:R} | \mathbf{k}, \boldsymbol{\lambda}^{1:R}, \mathbf{c}^{1:R}) \mathcal{P}(\mathbf{k}) \quad (17)$$

$$= \mathcal{P}(\mathbf{k}) \prod_{r=1}^R \mathcal{P}(\mathbf{b}_{1:N}^r | \mathbf{k}, \boldsymbol{\lambda}^{1:R}, \mathbf{c}^{1:R}) \quad (18)$$

where \mathbf{k} is the grouping of all rates (k_{bind} , k_{un} , and k_{bleach}) and we have assumed that the likelihoods governing different ROIs are independent.

We now determine the likelihood of an ROI's state trace $\mathbf{b}_{1:N}^r$. The state at each time point, b_n^r , represents the number of fluorophores that are bright at the time level n . However, there are many different configurations of fluorophore binding and photophysics that can give rise to a single b_n^r . For example, even if we see only one bright fluorophore there could be other fluorophores bound in a photobleached state. We call b_n^r a "macrostate" and denote all fluorophore configurations, s_n^r , a "microstate", where a fluorophore configuration is the number of bright and photobleached fluorophores bound at the time level. We let $\text{Bright}(s_n^r)$ and $\text{Bleached}(s_n^r)$ be two functions selecting the number of bright and photobleached fluorophores in the microstate. We can then enumerate the possible microstates as follows

$$\begin{aligned} \Sigma_1 &= \{0 \text{ Bright}, 0 \text{ Photobleached}\} \\ \Sigma_2 &= \{1 \text{ Bright}, 0 \text{ Photobleached}\} \\ \Sigma_3 &= \{0 \text{ Bright}, 1 \text{ Photobleached}\} \\ \Sigma_4 &= \{2 \text{ Bright}, 0 \text{ Photobleached}\} \\ \Sigma_5 &= \{1 \text{ Bright}, 1 \text{ Photobleached}\} \\ &\vdots \quad \quad \quad \vdots \end{aligned}$$

where Σ_i is the i th macrostate, and we enumerate up until we exhaust all M available fluorophores. The likelihood of a state trace, $\mathbf{b}_{1:N}^r$ is the probability of a microstate trace that gives rise to the observed macrostate trace, marginalized over all possible microstate traces. Assuming that microstates follow a Hidden Markov Model this becomes,

$$\mathcal{P}(\mathbf{b}_{1:N}^r | \mathbf{k}) = \sum_{\mathbf{s}_{1:N}^r} \mathcal{P}(\mathbf{b}_{1:N}^r | \mathbf{s}_{1:N}^r, \mathbf{k}, \boldsymbol{\lambda}^r, \mathbf{c}^r) \quad (19)$$

$$= \sum_{\mathbf{s}_{1:N}^r} \mathcal{P}(b_1^r | s_1^r) \mathcal{P}(s_1^r | \mathbf{k}, \boldsymbol{\lambda}^r, \mathbf{c}^r) \prod_{n=2}^N \mathcal{P}(b_n^r | s_n^r) \mathcal{P}(s_n^r | s_{n-1}^r, \mathbf{k}, \boldsymbol{\lambda}^r, \mathbf{c}^r) \quad (20)$$

where $\mathbf{s}_{1:N}^r$ represents the full microstate trace. In the bottom of figure 2, we present our posterior as a graphical model [44]. In the graphical model, nodes (circles) represent random variables, arrows represent conditional dependence, and the plate (rectangle) represents groupings of random variables repeated over an index.

Equation (20) requires a form for the probability of microstate given a macrostate. The probability of the microstate given a macrostate is simply an indicator function, or delta function, equal to 1 when the number of bright fluorophores in the microstate is equal to the macrostate and 0 otherwise

$$\mathcal{P}(b_n|s_n^r) = \delta_{b_n^r, \text{Bright}(s_n^r)} \quad (21)$$

where δ_{ij} is the Kronecker delta. The microstate transition probability is calculated using the transition rates, k_{bind} , k_{un} , k_{bleach} . Assuming one event per time level per ROI there are only four possible events: no transition, bright fluorophore binding, bright fluorophore unbinding, and photobleached fluorophore unbinding. We exclude the possibility of photobleached fluorophores binding, which is assumed to be rare. Letting $\text{Bright}(s_n^r)$ and $\text{Bleached}(s_n^r)$ be functions which count the number of bright and photobleached fluorophores in state s_n^r respectively, we can calculate the probability of each event as follows

$$k_{\text{tot}}(s_n^r, \lambda^r, c^r) = (\text{Bright}(s_n^r) + \text{Bleached}(s_n^r))k_{\text{un}} + c^r k_{\text{bind}} + \lambda^r \text{Bright}(s_n^r)k_{\text{bleach}} \quad (22)$$

$$\mathcal{P}(\text{No Event}|s_n^r, \mathbf{k}) = 1 - \int_0^{\Delta t} dT k_{\text{tot}}(s_n^r, \lambda^r, c^r) e^{-k_{\text{tot}}(s_n^r, \lambda^r, c^r)T} \quad (23)$$

$$= k_{\text{tot}}(s_n^r, \lambda^r, c^r) \Delta t e^{k_{\text{tot}}(s_n^r, \lambda^r, c^r) \Delta t} \quad (24)$$

$$= P_0(s_n^r) \quad (25)$$

$$\mathcal{P}(\text{Bright binding}|s_n^r, \mathbf{k}, \lambda^r, c^r) = (1 - P_0(s_n^r)) \frac{k_{\text{bind}}}{k_{\text{tot}}(s_n^r, \lambda^r, c^r)} \quad (26)$$

$$\mathcal{P}(\text{Bright unbinding}|s_n^r, \mathbf{k}, \lambda^r, c^r) = (1 - P_0(s_n^r)) \frac{\text{Bright}(s_n^r)k_{\text{un}}}{k_{\text{tot}}(s_n^r, \lambda^r, c^r)} \quad (27)$$

$$\mathcal{P}(\text{Bleached unbinding}|s_n^r, \mathbf{k}, \lambda^r, c^r) = (1 - P_0(s_n^r)) \frac{\text{Bleached}(s_n^r)k_{\text{un}}}{k_{\text{tot}}(s_n^r, \lambda^r, c^r)} \quad (28)$$

$$\mathcal{P}(\text{Photobleaching}|s_n^r, \mathbf{k}, \lambda^r, c^r) = (1 - P_0(s_n^r)) \frac{\text{Bright}(s_n^r)k_{\text{bleach}}}{k_{\text{tot}}(s_n^r, \lambda^r, c^r)} \quad (29)$$

where $k_{\text{tot}}(s_n^r, \lambda^r, c^r)$ is the total rate, $P_0(s_n^r)$ is the probability of a self transition (calculated as 1 minus the probability of an exponentially distributed event), and the remaining events are calculated according to their relative rates [51]. We note that the total binding rate is assumed to scale linearly with the concentration and the total photobleaching rate is assumed to scale linearly with laser intensity. These rates can be used to construct a master transition probability matrix, $\mathbf{\Pi}(k_{\text{bind}}, k_{\text{un}}, k_{\text{bleach}})$, where each element can be calculated by looking at changes in state populations between the initial state and final state for the 5 possible transitions: 1) No transition, 2) Binding, 3) Bright Unbinding, 4) Bleached unbinding, and 5) Photobleaching

$$\mathbf{\Pi}(\mathbf{k})_{ij} = \begin{cases} P_0(\Sigma_i) & \text{if } \Delta \text{Bright}(\Sigma_j, \Sigma_i) = 0 \text{ and } \Delta \text{Bleached}(\Sigma_j, \Sigma_i) = 0 \\ (1 - P_0(\Sigma_i)) \frac{k_{\text{bind}}}{k_{\text{tot}}(\Sigma_i)} & \text{if } \Delta \text{Bright}(\Sigma_j, \Sigma_i) = 1 \text{ and } \Delta \text{Bleached}(\Sigma_j, \Sigma_i) = 0 \\ (1 - P_0(\Sigma_i)) \frac{\text{Bright}(\Sigma_i)k_{\text{un}}}{k_{\text{tot}}(\Sigma_i)} & \text{if } \Delta \text{Bright}(\Sigma_j, \Sigma_i) = -1 \text{ and } \Delta \text{Bleached}(\Sigma_j, \Sigma_i) = 0 \\ (1 - P_0(\Sigma_i)) \frac{\text{Bleached}(\Sigma_i)k_{\text{un}}}{k_{\text{tot}}(\Sigma_i)} & \text{if } \Delta \text{Bright}(\Sigma_j, \Sigma_i) = 0 \text{ and } \Delta \text{Bleached}(\Sigma_j, \Sigma_i) = -1 \\ (1 - P_0(\Sigma_i)) \frac{\text{Bright}(\Sigma_i)k_{\text{bleach}}}{k_{\text{tot}}(\Sigma_i)} & \text{if } \Delta \text{Bright}(\Sigma_j, \Sigma_i) = -1 \text{ and } \Delta \text{Bleached}(\Sigma_j, \Sigma_i) = 1 \end{cases} \quad (30)$$

where

$$\Delta\text{Bright}(\Sigma_i, \Sigma_j) = \text{Bright}(\Sigma_j) - \text{Bright}(\Sigma_i) \quad (31)$$

$$\Delta\text{Bleached}(\Sigma_i, \Sigma_j) = \text{Bleached}(\Sigma_j) - \text{Bleached}(\Sigma_i) \quad (32)$$

calculate the change in bright and bleached populations. We use the matrix presented in equation (30) to marginalize over the trajectories in equation (20) using the Forward-Backward algorithm [52], thus our likelihood, $\mathcal{P}(\mathbf{b}_{1:N}^r | \mathbf{k})$, can be calculated exactly [52].

In order to find the set of rates, \mathbf{k} , that maximizes the posterior, equation (18), we again use a Gibbs sampling scheme, similar to that laid out in the previous section, but this time saving all samples, not just the MAP. The algorithm is as follows,

- Start with initial set of variables, $\mathbf{k}^{(0)} = \{k_{\text{bind}}^{(0)}, k_{\text{un}}^{(0)}, k_{\text{bleach}}^{(0)}\}$.
- For many iterations i :
 - Propose a new set of parameters \mathbf{k}' based on a proposal distribution $q(\mathbf{k}' | \mathbf{k}^{(i-1)})$.
 - Calculate the acceptance ratio a , defined as:

$$a = \min \left(1, \frac{P(\mathbf{k}')q(\mathbf{k}^{(i-1)} | \mathbf{k}')}{P(\mathbf{k}^{(i-1)})q(\mathbf{k}' | \mathbf{k}^{(i-1)})} \right) \quad (33)$$

where $P(\mathbf{k})$ represents the posterior probability of the parameters and $q(\Theta_a | \Theta_b)$ represents the proposal distribution.

- Accept or reject the new parameter set \mathbf{k}' with probability a .
- Save the sample, $\mathbf{k}^{(i)}$.

After drawing many samples, we may histogram the results to visualize the mean and variance of the distribution.

2.2 Experimental Methods

2.2.1 DNA Origami

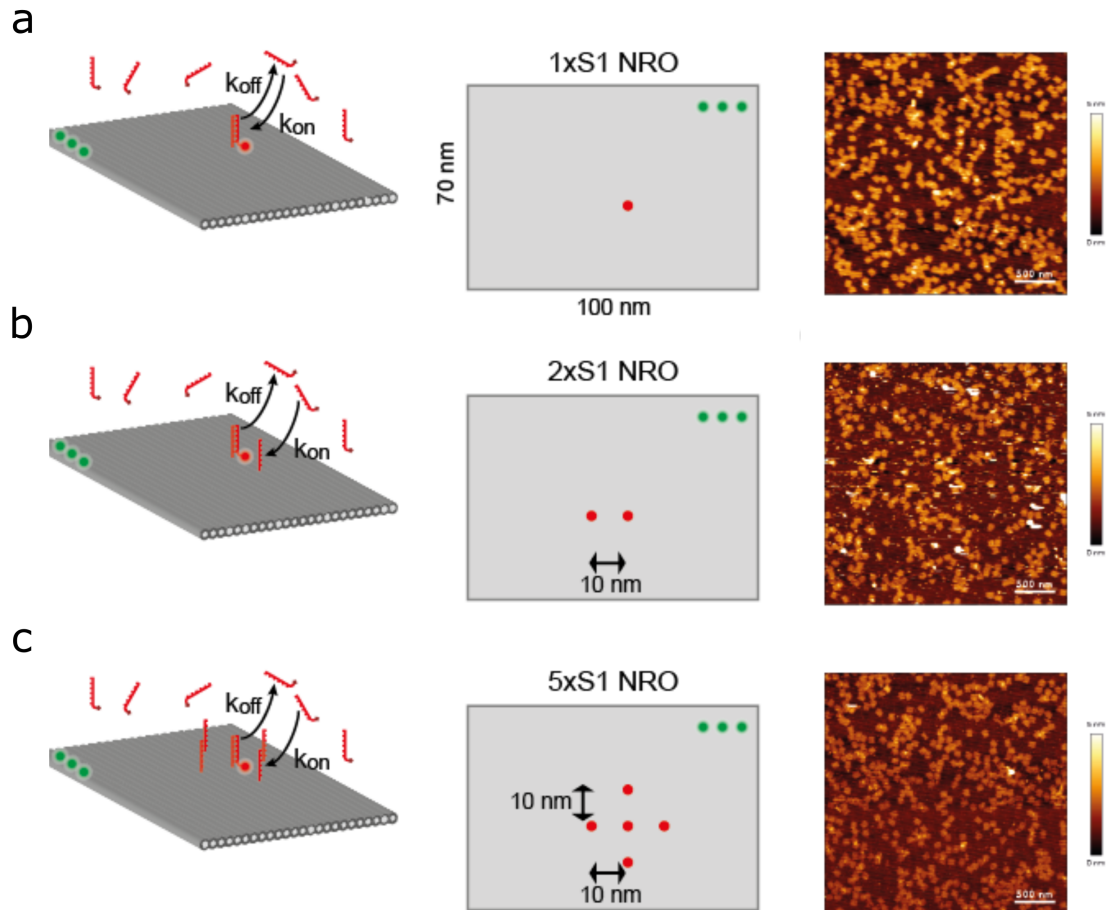


Figure 3: Illustration of the new rectangular DNA origami (NRO) structure and its binding site arrangement. Each row provides information about DNA origami with a set number of S1 docking sites- (a) one binding site, (b) two binding sites, (c) five binding sites. The left column shows cartoon of the DNA origami with the imager strands (red) binding to the S1 docking sites and a stable reference label on three S2 docking sites (green). The middle column shows a diagram of the layout of binding sites and reference labels on the DNA origami. The right column shows atomic force microscopy images of the DNA origami immobilised on a mica surface.

In order to test our method of inferring binding rates for different numbers of docking sites within a ROI, we designed a DNA Origami with varying numbers of external docking sites featuring the S1 sequence (see figure 3 and table SI1). The docking sites were added to the 3'-ends of selected staple strands positioned at 10 nm distances. For colocalization, the new rectangular origami (NRO) was equipped with three S2 docking sites (21 nt) for permanent external labeling of the structure with a Atto542-labeled complementary S2 strand (see table SI2). For immobilization of the DNA origami to a glass surface functionalized with Biotin-Streptavidin, six staple strands

of the NRO were labeled with biotin on their 3'-end. Exact strand sequence can be found in the SI in table SI1, table SI2, and table SI3.

The DNA origami was folded using 20 nM of the scaffold p7249, 200 nM of each unmodified staple strand and 600 nM of each modified staple strand (biotinylated and external S1 labeling strands) in buffer containing 10 mM Tris, 1 mM EDTA at pH 8 and 12.5 mM MgCl₂. The mixture was heated to 70 °C and then cooled down to 25 °C with a linear thermal annealing ramp of 1°C/min [53].

Sample purification was performed by filtration using Amicon Ultra filters (100 kDa molecular weight cut-off, Merck, Germany). The filter was first centrifuged with folding buffer for 7 min at 6000 g. The sample solution was then loaded into the filter and centrifuged for 15 min at 6000 g. Then 300 µL of folding buffer was loaded into the filter and centrifuged for 15 min at 6000 g, which was repeated once. After the three washing steps, the filter was inverted and placed into a new collection tube. The purified sample could then be collected by centrifugation for 2 min at 1000 g. The successful folding of structures was confirmed by AFM analysis. DNA origami solutions were stored at -20 °C until used.

To determine correct folding, AFM scans in aqueous solution (AFM buffer = 40 mM Tris, 2 mM EDTA, 12.5 mM Mg(OAc)₂·4 H₂O) were performed on a NanoWizard® 3 ultra AFM (JPK Instruments AG) (see figure 3). For sample immobilization, a freshly cleaved mica surface (Quality V1, Plano GmbH) was incubated with 10 mM solution of NiCl₂ for 3 min. The mica was washed three times with ultra-pure water to remove unbound Ni²⁺ ions and air-dried. The dried mica surface was incubated with 1 nM sample solution for 3 min and washed with AFM buffer three times. Measurements were performed in AC mode on a scan area of 3 x 3 µm with a BioLeverMini cantilever ($\nu_{res} = 110$ kHz air / 25 kHz fluid, $k_{spring} = 0.1$ N/m, Bruker AFM Probes).

2.2.2 Sample preparation for fluorescence microscopy

Samples were prepared in chambered coverslips (LabTek) which were cleaned twice with 0.1 M HF with two water rinses after each cleaning step. The chambers were then incubated for 5 min in magnesium-free phosphate buffered saline (PBS). The coverslips were then coated with 1 mg/ml biotin-BSA overnight before being washed three times in PBS. Next, the coverslips were incubated for 30 min with 0.2 mg/ml streptavidin before they were washed three times with Red Base composed of 2x PBS with additional 0.5 M NaCl and 0.05% Tween20. 10 fmol of rectangular DNA origami was then incubated for 30 min on the surface of the coverslip, before removing the solution and washing three times with Red Base. Each DNA origami construct was pre-labeled with Atto542 as a reference marker and had a defined number of S1 docking sites (see Supplement). The samples were inspected for sufficient DNA origami density and in cases of low coverage additional DNA origami was added to achieve sufficient coverage. Each chamber of the coverslips was then filled with imaging buffer and encapsulated with self-sealing Parafilm(M). The imaging buffer contained 2.1 mM ascorbic acid (AA), 1 mM methyl viologen (MV), 2.5 mM protocatechuic acid (PCA) and 50 nM protocatechuate-3,4-dioxygenase (PCD) in Red Base. Different concentrations of Cy5-conjugated oligonucleotides were used in each experiment (1 nM, 2 nM, 5 nM, 10 nM). The imager strand concentration was adjusted by substituting Red Base with 1-10 µl of 1 µM imager strand diluted in Red Base.

2.2.3 Data Acquisition

The samples were imaged on a custom microscope setup using an inverted microscope (Nikon Eclipse Ti, Nikon, Japan) and a fibre-coupled multilaser engine (MLE-LFA, TOPTICA Photonics, Germany). The excitation was coupled to the Nikon manual TIRF unit and then directed

to 100 \times , 1.49 NA oil immersion objective (Apo TIRF, Nikon, Japan) via a four-band dichroic mirror (405, 488, 532, 640 nm). Emission was then filtered by 525/50 nm, 605/70 nm, and 690/70 nm bandpass filters (all AHF Analysetechnik, Germany) mounted on a motorized filter wheel (FW212C, Thorlabs, Germany) and collected by an EMCCD (Electron Multiplying Charge-Coupled Device) camera (iXon Ultra 897, Andor, UK) with a 105.6 nm pixel size. Initially, the fields of view were imaged for five frames in the 561 nm channel to attain reference data for localisations of the origami. Afterwards, the same positions are imaged at 640 nm under TIRF illumination with 21.7 mW and 43.8 mW measured before the objective with a power meter kit (PM120D, Thorlabs, Germany). These correspond to 98 W/cm² and 198 W/cm² at the focal plane, respectively. To gather kinetic data 8000 frames were recorded at 100 ms exposure (124.6 ms frame time) over a total duration of 16.5 min and saved as an image stack.

2.2.4 Post Acquisition Data Processing

For each image stack, a custom Fiji script [54] determines the brightest values of each pixel, creating a maximum intensity projection, after which ThunderSTORM [55] is used to localise the coordinates for each region of interest (ROI), i.e. the location of DNA origami, in both channels. Then a Python script, which utilizes the trace extraction function of the quickPBSA library [37], is used to determine the intensity trace inside each ROI and subtract the background around that ROI. Then the maximum intensity projections of the two channels were registered using the MultiStackReg plugin [56] in Fiji giving a correction matrix used to transform the coordinates in the reference channel via a custom Fiji script. Finally, the ROIs further than 200 nm of the reference channel are considered unspecific binding and filtered out.

3 Results

The main goal of our algorithm is to learn the binding (k_{bind}), unbinding (k_{un}), photobleaching (k_{bleach}) rates and the state trajectories given a set of noisy intensity traces $s_{1:N}^r$ for every given ROI r . To do so, our algorithm takes a chain of numerical samples from the posterior using Monte Carlo methods as detailed in section 2. As such, our results are in the form of histograms of the numerical samples taken from the posterior for every individual unknown parameter, where the uncertainty of rate inference given a microstate trace over each parameter is reflected in the width of the corresponding histograms.

3.1 Simulated data

We first benchmarked our method using data simulated under different regimes data using the forward model described in the Methods (section 2). Starting with a base set of parameters chosen to mimic the experimental data, we ran inference while varying one parameter at a time to probe the robustness of our method with respect to different data paradigms. The base set of parameters chosen to qualitatively match the data is as follows: $k_{\text{bind}} = 1$ mHz/nM, $k_{\text{un}} = 25$ mHz, $k_{\text{bleach}} = .01$ mHz, $\mu_f = 400$ ADU, $\mu_B = 0$ ADU, $\sigma = 2500$ ADU, $dt = 124$ ms, and 2000 ROIs with traces simulated with either $\lambda = 50$ or $\lambda = 100$ in even proportion and $c = 1$ nM, $c = 2$ nM, or $c = 5$ nM in even proportion. We simulated data varying k_{bind} , k_{un} , k_{bleach} , and the number of ROIs summarized in table SI4.

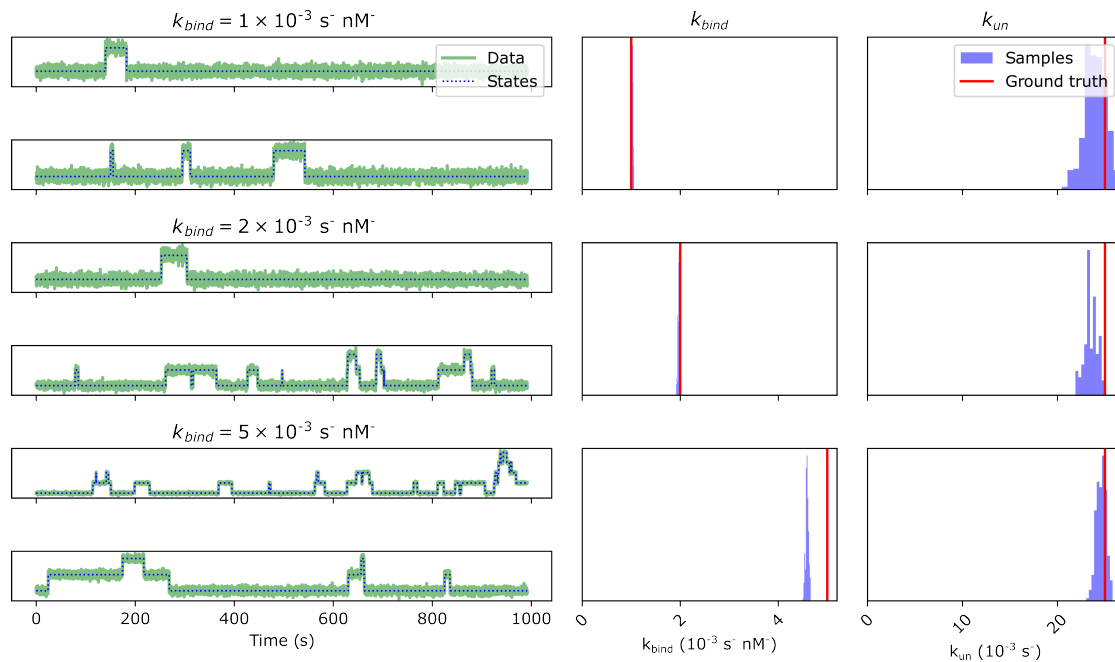


Figure 4: Simulated data with different k_{bind} binding rates. The top two rows of the left column show example traces from data simulated with $k_{\text{bind}} = 1 \text{ mHz/nM}$, the middle two rows of the left column show example traces from data simulated with $k_{\text{bind}} = 2 \text{ mHz/nM}$, and the bottom two rows of the left column show traces from data simulated with $k_{\text{bind}} = 5 \text{ mHz/nM}$. The middle column shows inferred posteriors over binding rates, with blue indicating the probability distribution and a red line indicating the ground truth used for simulation. The right column shows inferred posteriors over unbinding rates, with blue indicating the probability distribution and red lines indicating ground truth values used for simulation. Dataset included 200 traces per condition. The parameters used in simulation are specified in main text and table SI4.

Initially, we probed the model's robustness with respect to binding rate, k_{bind} , by comparing analysis of data simulated with $k_{\text{bind}} = 1 \text{ mHz/nM}$, $k_{\text{bind}} = 2 \text{ mHz/nM}$, and $k_{\text{bind}} = 5 \text{ mHz/nM}$. Figure 4 shows results from the data simulated with these different k_{bind} rates. Figure 4 is broken into three columns where the left column shows representative data traces, the middle column shows the inferred k_{bind} binding rate, and the right column shows the inferred k_{un} unbinding rate. For the plots showing data traces, we show two representative traces taken from the dataset, stacked on top of each other. For the plots showing inferred rates, the blue indicates the probability distribution (histogram of samples from the MCMC), and the red line indicates the ground truth value used in the simulation. As seen in the middle column of figure 4, our method accurately (within 10%) infers the correct k_{bind} rate with the analysis returning 1.031 ± 0.012 , 1.971 ± 0.018 and $4.588 \pm 0.030 \text{ mHz/nM}$ for the k_{bind} value. k_{un} results remain also constant only slightly lower than the set 25 mHz.

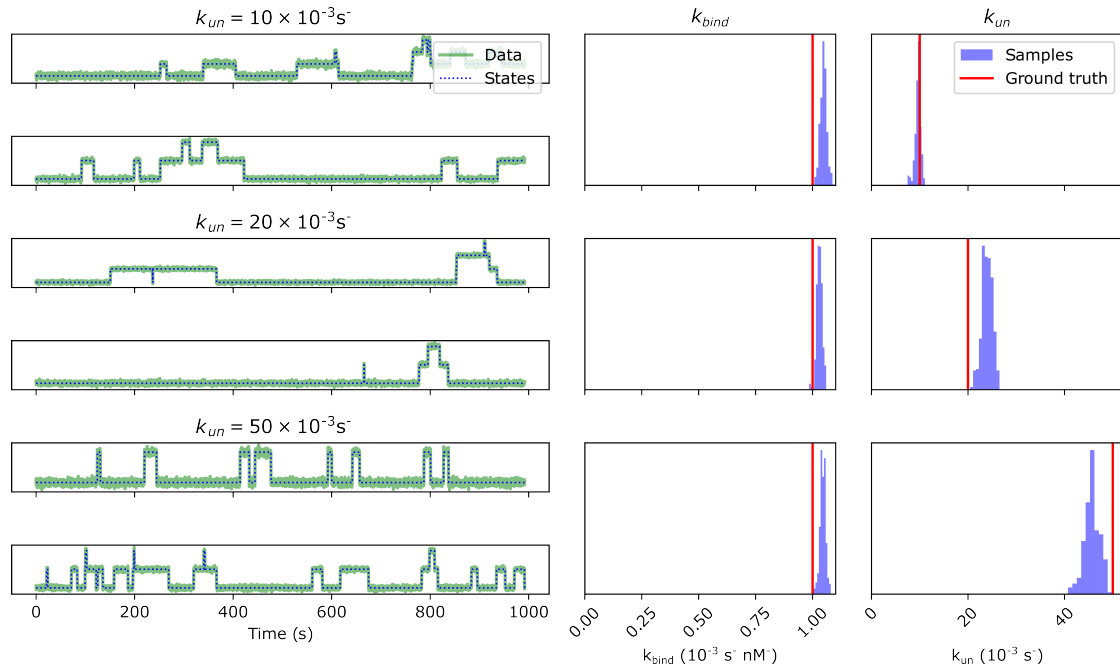


Figure 5: Simulated data with different k_{un} unbinding rates. The top two rows of the left column show example traces from data simulated with $k_{\text{un}} = 10$ mHz, the middle two rows of the left column show example traces from data simulated with $k_{\text{un}} = 20$ mHz, and the bottom two rows of the left column show traces from data simulated with $k_{\text{un}} = 50$ mHz. The middle column shows inferred posteriors over binding rates, with blue indicating the probability distribution and a red line indicating the ground truth used for simulation. The right column shows inferred posteriors over unbinding rates, with blue indicating the probability distribution and red lines indicating ground truth values used for simulation. Dataset included 200 traces per condition. The parameters used in simulation are specified in main text.

We next probed the model's robustness with respect to unbinding rate, k_{un} , by comparing analysis of data simulated with $k_{\text{un}} = 10$ mHz, $k_{\text{un}} = 20$ mHz, and $k_{\text{un}} = 50$ mHz. Figure 5 shows the results on data simulated with these different k_{un} rates. As seen in the far right column of figure 5, our model accurately infers the correctly the change of k_{un} rate with 9.623 ± 0.612 mHz, 23.950 ± 1.086 mHz, 45.650 ± 1.586 mHz for the 10 mHz, 20 mHz and 50 mHz, respectively. In line with the input, the k_{bind} results remained around within 0.1 of the 1 mHz/nM input for all simulations.

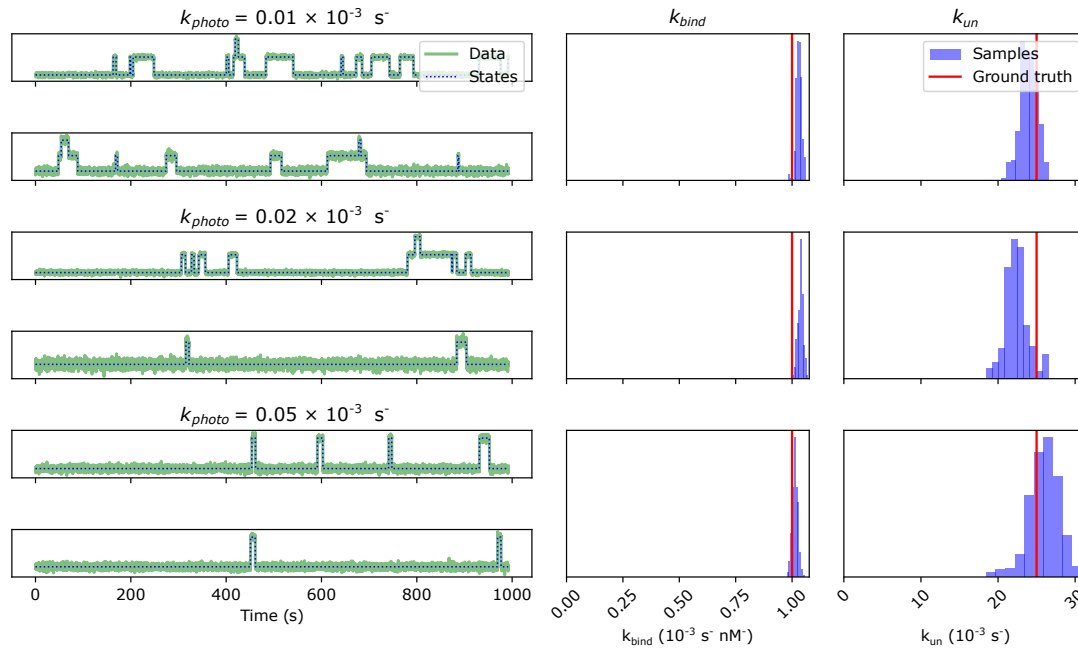


Figure 6: Simulated data with different k_{bleach} photobleaching rates. The top two rows of the left column show example traces from data simulated with $k_{\text{bleach}} = .01$ mHz, the middle two rows of the left column show example traces from data simulated with $k_{\text{bleach}} = .02$ mHz, and the bottom two rows of the left column show traces from data simulated with $k_{\text{bleach}} = .05$ mHz. The middle column shows inferred posteriors over binding rates, with blue indicating the probability distribution and a red line indicating the ground truth used for simulation. The right column shows inferred posteriors over unbinding rates, with blue indicating the probability distribution and red lines indicating ground truth values used for simulation. Dataset included 200 traces per condition. The parameters used in simulation are specified in main text and table SI4.

After that, we probed the model's robustness with respect to photobleaching rate, k_{bleach} , by comparing analysis of data simulated with $k_{\text{bleach}} = 0.01$ mHz, $k_{\text{bleach}} = 0.02$ mHz, and $k_{\text{bleach}} = 0.05$ mHz. Figure 6 shows the results on data simulated with these different k_{bleach} rates. As seen in figure 6, we are still able to infer accurate k_{bind} and k_{un} rates for all values of k_{bleach} .

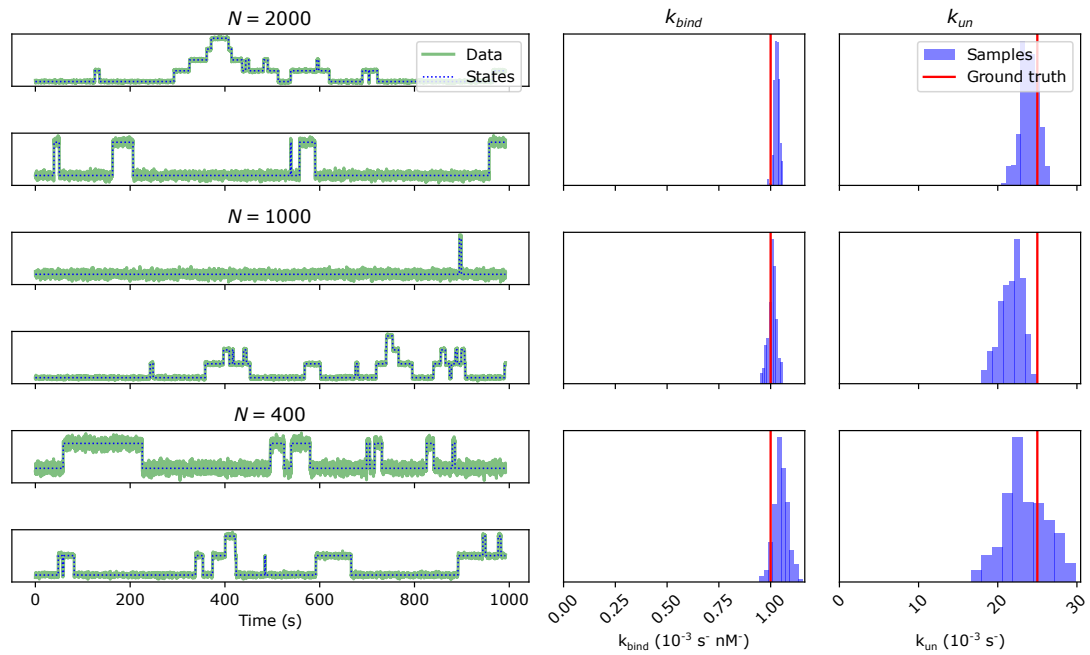


Figure 7: Simulated data with different number of ROIs. The top two rows of the left column show example traces from data simulated with 2000 ROIs, the middle two rows of the left column show example traces from data with 1000 ROIs, and the bottom two rows of the left column show traces from data with 400 ROIs. The middle column shows inferred posteriors over binding rates, with blue indicating the probability distribution and a red line indicating the ground truth used for simulation. The right column shows inferred posteriors over unbinding rates, with blue indicating the probability distribution and red lines indicating ground truth values used for simulation. The parameters used in simulation are specified in main text and table SI4.

Finally, we probed the model's robustness with respect to number of ROIs, by comparing data simulated with 2000 ROIs, 1000 ROIs, and 400 ROIs. Figure 7 shows the results on the datasets simulated with these different numbers of ROIs. Despite the small increase in uncertainty, marked by the spread in probability, our method was able to accurately learn k_{bind} and k_{un} from even a small amount of data.

3.2 Experimental data

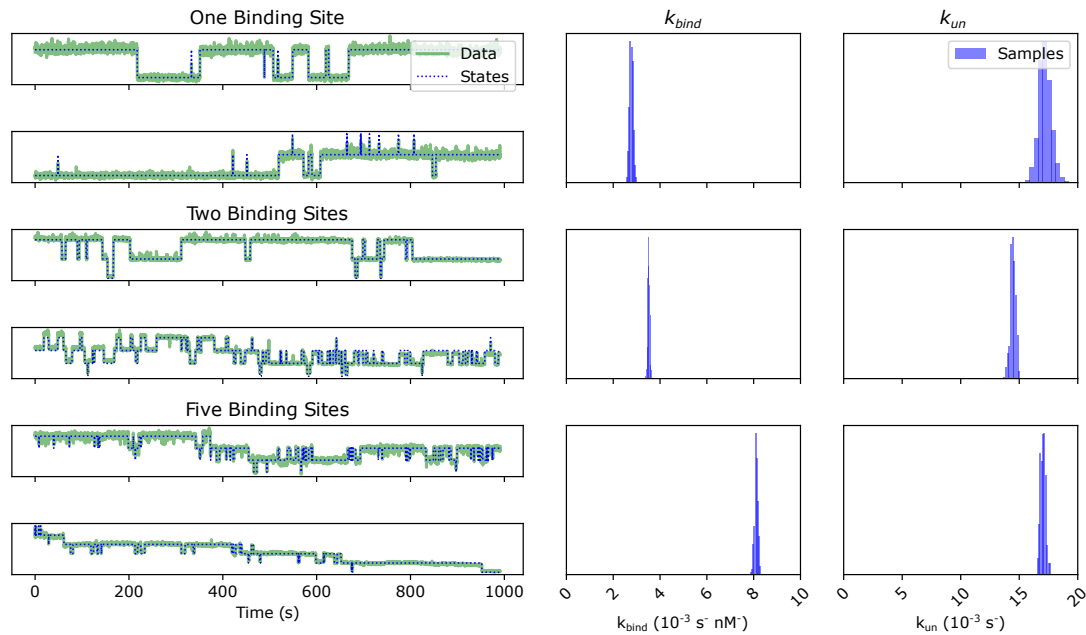


Figure 8: Data from DNA origami binding experiments. The left column shows example traces from our experiments. The top two rows of the left column show example traces from data with one binding site, the middle two rows of the left column show example traces from data with two binding sites, and the bottom two rows of the left column show traces from data with five binding sites. The middle column shows inferred posteriors over binding rates. The right column shows inferred posteriors over unbinding rates. Data from 126, 393 and 353 traces were analyzed for the one, two and five binding site samples, respectively.

We applied our method to data taken from highly controlled DNA origami experiments in which fluorescently labeled particles were allowed to bind and unbind to origami with 1, 2, and 5 binding sites. Results for inference on experimental data are shown in figure 8. Each row of figure 8 represents an experiment with a different number of binding sites with the inferred rates of binding and unbinding shown as a histogram.

We cannot compare our inferred rates directly to a ground truth, as the ground truth is unknown. However, we can evaluate the accuracy of our model by checking to see if the output behavior matches expectations. For example, the binding rate should scale with the number of available binding sites, but the unbinding rate should be unaffected by the number of binding sites. As seen in figure 8, inferred rates qualitatively increase with binding sites. Our inferred binding rates are the 2.77 ± 0.07 mHz/nM (*mean* \pm *SD*) for 1 binding site, 3.52 ± 0.04 mHz/nM for 2 binding sites, and 8.1 ± 0.07 mHz for 5 binding sites, which increases with concentration. These values also match the values seen in literature when it comes to a 10 nucleotide imager strands [57]. Meanwhile, the unbinding rates are 17.2 ± 0.5 mHz, 14.4 ± 0.2 mHz, and 17.0 ± 0.2 mHz, effectively remaining constant. The experimental data is intrinsically noisy, containing many artifacts, such as spikes (see figure 8), which complicate analysis. These can be attributed to

unbound strands moving over the ROI or the molecule escaping a prolonged dark state. The bleaching rates show a small values not correlated to the binding site number (see table SI5).

4 Discussion

Quantifying binding kinetics while overcoming noise and photobleaching artifacts is important as it provides a deeper and more accurate understanding of molecular interactions and behaviors in chemical and biological systems. This enhanced understanding is crucial in dissecting the intricacies of protein self-assembly [1, 2, 3], enzymatic and protein-protein interactions [5, 6, 7, 8, 9], DNA-protein interactions [11, 12, 13], and broader cellular processes, such as nuclear dynamics during mitosis and [15, 16, 17].

Here, we presented a Bayesian inference scheme for inferring binding rates separately from photobleaching rates, by directly monitoring fluorescence intensity over time. We prioritized inferring rates over transition probabilities, as it enables us to disentangle the contributions of binding and photobleaching. Our method is broken into two modules: a State Inference algorithm that infers the number of fluorophores at each time level directly from the data, and secondly a Rate Inference step which infers the binding rates from the output fluorescence trace. We benchmarked our method on simulated data, inferring reliable results. We showed that when run on experimental data we inferred expected behavior of increasing binding rates with constant unbinding rates for increasing concentrations.

A significant way to enhance this work would be by incorporating spatial information. Currently, the method works by analyzing intensity traces from isolated ROIs. By incorporating spatial information we would be able to take into account other potential noise generating artifacts such as uneven illumination of ROIs [58].

Another way to improve this work would be to incorporate additional photophysical states. For example, in the case of photo-blinking the number of brightness steps is not equal to the number of fluorophores, making analysis with traditional HMMs impossible [39]. Our method, on the other hand, is generalizable to an arbitrary number of dark and bright states. However, while a state inference analysis can be effective when state traces are constrained, as in the case of counting by photobleaching where the fluorophores are assumed to start bound and end bleached with no unbinding [39], without such a constraint the number of possible state trajectories becomes combinatorically larger. As a result it may be difficult, for fundamental reasons, to perform scalable analysis with a higher number of photophysical states.

Data Availability

Data used in this work can be found at labpresse.com.

Code Availability

Code used in this work be found at labpresse.com.

Acknowledgements

SP acknowledge support from an NIH MIRA award (NIGMS R35GM148237). DPH acknowledges the support from the Centre of Membrane Proteins and Receptors (COMPARE) funded by the Universities of Nottingham and Birmingham as well as support from the Academy of Medical Sciences (Grant APR2\1013).

Author Contributions

SP and DPH conceived of the project. MS designed and created the DNA origami. SAT performed the sample preparation, imaging and trace extraction for the experimental data. JSB carried out the coding and development. JSB and MF performed the analyses and wrote the manuscript.

Competing Interests

SP and JSB acknowledge a competing interest with their affiliation with Saguaro Solutions.

Correspondence

Correspondence and requests for materials should be addressed to Steve Pressé. (email: [spresso@asu.edu](mailto:spresse@asu.edu)).

References

- [1] Sibali Debnath and Krishnan Raghavachari. “Investigating the Stacking Interactions Responsible for Driving 3D Self-Association of Tricarb Macrocycles”. In: *The Journal of Physical Chemistry A* 127.39 (2023), pp. 8110–8116.
- [2] Helen W German, Sahin Uyaver, and Ulrich HE Hansmann. “Self-assembly of phenylalanine-based molecules”. In: *The Journal of Physical Chemistry A* 119.9 (2014), pp. 1609–1615.
- [3] Janet M Petroski, Travis C Green, and Mostafa A El-Sayed. “Self-assembly of platinum nanoparticles of various size and shape”. In: *The Journal of Physical Chemistry A* 105.23 (2001), pp. 5542–5547.
- [4] Guangzhong Ma et al. “Quantifying ligand–protein binding kinetics with self-assembled nano-oscillators”. In: *Analytical chemistry* 91.21 (2019), pp. 14149–14156.
- [5] Kinshuk Banerjee, Anatoly B Kolomeisky, and Oleg A Igoshin. “Accuracy of substrate selection by enzymes is controlled by kinetic discrimination”. In: *The journal of physical chemistry letters* 8.7 (2017), pp. 1552–1556.
- [6] Aleksandra M Kusova et al. “Protein translational diffusion and intermolecular interactions of globular and intrinsically unstructured proteins”. In: *The Journal of Physical Chemistry A* 123.46 (2019), pp. 10190–10196.
- [7] Gideon Schreiber. “Kinetic studies of protein–protein interactions”. In: *Current opinion in structural biology* 12.1 (2002), pp. 41–47.
- [8] Gideon Schreiber, Gilad Haran, and H-X Zhou. “Fundamental aspects of protein- protein association kinetics”. In: *Chemical reviews* 109.3 (2009), pp. 839–860.
- [9] Yibing Shan et al. “How does a small molecule bind at a cryptic binding site?” In: *PLoS computational biology* 18.3 (2022), e1009817.
- [10] Mengmeng Zhang, Bin Wang, and Bingqian Xu. “Mapping single molecular binding kinetics of carbohydrate-binding module with crystalline cellulose by atomic force microscopy recognition imaging”. In: *The Journal of Physical Chemistry B* 118.24 (2014), pp. 6714–6720.
- [11] Tom Misteli. “Protein dynamics: implications for nuclear architecture and gene expression”. In: *Science* 291.5505 (2001), pp. 843–847.

- [12] Stephen E Halford and John F Marko. “How do site-specific DNA-binding proteins find their targets?” In: *Nucleic acids research* 32.10 (2004), pp. 3040–3052.
- [13] Sebastian Strauss et al. “Principles of RNA recruitment to viral ribonucleoprotein condensates in a segmented dsRNA virus”. In: *Elife* 12 (2023), e68670.
- [14] Anthony Kar Lun Leung et al. “Quantitative kinetic analysis of nucleolar breakdown and reassembly during mitosis in live human cells”. In: *The Journal of cell biology* 166.6 (2004), pp. 787–800.
- [15] Elisa Dultz et al. “Systematic kinetic analysis of mitotic dis- and reassembly of the nuclear pore in living cells”. In: *The Journal of cell biology* 180.5 (2008), pp. 857–865.
- [16] Elisa Dultz and Jan Ellenberg. “Live imaging of single nuclear pores reveals unique assembly kinetics and mechanism in interphase”. In: *Journal of Cell Biology* 191.1 (2010), pp. 15–22.
- [17] Bernhard Hampoelz et al. “Structure and assembly of the nuclear pore complex”. In: *Annual review of biophysics* 48 (2019), pp. 515–536.
- [18] Arup K Chakraborty and Arthur Weiss. “Insights into the initiation of TCR signaling”. In: *Nature immunology* 15.9 (2014), pp. 798–807.
- [19] William K Kanagy et al. “Docking of Syk to FcεRI is enhanced by Lyn but limited in duration by SHIP1”. In: *Molecular Biology of the Cell* 33.10 (2022), ar89.
- [20] Bin Dong et al. “In situ quantitative single-molecule study of dynamic catalytic processes in nanoconfinement”. In: *Nature Catalysis* 1.2 (2018), pp. 135–140. ISSN: 2520-1158. DOI: 10.1038/s41929-017-0021-1. URL: <https://doi.org/10.1038/s41929-017-0021-1>.
- [21] Weilin Xu et al. “Single-molecule nanocatalysis reveals heterogeneous reaction pathways and catalytic dynamics”. In: *Nature Materials* 7.12 (2008), pp. 992–996. ISSN: 1476-1122. DOI: 10.1038/nmat2319. URL: <https://dx.doi.org/10.1038/nmat2319>.
- [22] Thorben Cordes and Suzanne A. Blum. “Opportunities and challenges in single-molecule and single-particle fluorescence microscopy for mechanistic studies of chemical reactions”. In: *Nature Chemistry* 5.12 (2013), pp. 993–999. ISSN: 1755-4330. DOI: 10.1038/nchem.1800. URL: <https://dx.doi.org/10.1038/nchem.1800>.
- [23] A. Rybina et al. “Distinguishing alternative reaction pathways by single-molecule fluorescence spectroscopy”. In: *Angewandte Chemie-International Edition* 52.24 (2013), pp. 6322–5. ISSN: 1521-3773 (Electronic) 1433-7851 (Linking). DOI: 10.1002/anie.201300100. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23630097>.
- [24] M. Jäger et al. “Analysis of single-molecule fluorescence spectroscopic data with a Markov-modulated Poisson process”. In: *Chemphyschem* 10.14 (2009), pp. 2486–95. ISSN: 1439-7641 (Electronic) 1439-4235 (Linking). DOI: 10.1002/cphc.200900331. URL: <https://www.ncbi.nlm.nih.gov/pubmed/19644999> <https://chemistry-europe.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/cphc.200900331?download=true>.
- [25] A. Kiel et al. “Direct monitoring of formation and dissociation of individual metal complexes by single-molecule fluorescence spectroscopy”. In: *Angewandte Chemie-International Edition* 46.18 (2007), pp. 3363–6. ISSN: 1433-7851 (Print) 1433-7851 (Linking). DOI: 10.1002/anie.200604965. URL: <https://www.ncbi.nlm.nih.gov/pubmed/17397126> <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/anie.200604965?download=true>.

- [26] Robert Karlsson, Anne Michaelsson, and Lars Mattsson. “Kinetic analysis of monoclonal antibody-antigen interactions with a new biosensor based analytical system”. In: *Journal of immunological methods* 145.1-2 (1991), pp. 229–240.
- [27] Ray Bakhtiar. “Surface plasmon resonance spectroscopy: a versatile technique in a biochemist’s toolbox”. In: *Journal of chemical education* 90.2 (2013), pp. 203–209.
- [28] Douglas Magde, Elliot Elson, and Watt W Webb. “Thermodynamic fluctuations in a reacting system—measurement by fluorescence correlation spectroscopy”. In: *Physical review letters* 29.11 (1972), p. 705.
- [29] Laura Rusu et al. “Fluorescence correlation spectroscopy studies of peptide and protein binding to phospholipid vesicles”. In: *Biophysical journal* 87.2 (2004), pp. 1044–1053.
- [30] Jonas Mücke et al. “Quantifying reversible surface binding via surface-integrated fluorescence correlation spectroscopy”. In: *Nano Letters* 18.5 (2018), pp. 3185–3192.
- [31] Lan Yu et al. “A comprehensive review of fluorescence correlation spectroscopy”. In: *Frontiers in physics* 9 (2021), p. 644450.
- [32] Carlos S Smith et al. “An automated Bayesian pipeline for rapid analysis of single-molecule binding data”. In: *Nature communications* 10.1 (2019), p. 272.
- [33] Maximilian H Ulbrich and Ehud Y Isacoff. “Subunit counting in membrane-bound proteins”. In: *Nature methods* 4.4 (2007), pp. 319–321.
- [34] Jia Lin et al. “Nanoscale cell-wall architecture of an immunogenic ligand in *Candida albicans* during antifungal drug treatment”. In: *Molecular biology of the cell* 27.6 (2016), pp. 1002–1014.
- [35] KS Grubmayer, Klaus Yserentant, and Dirk-Peter Herten. “Photons in-numbers out: perspectives in quantitative fluorescence microscopy for in situ protein counting”. In: *Methods and applications in fluorescence* 7.1 (2019), p. 012003.
- [36] Johan Hummert, Stanimir Asenov Tashev, and Dirk-Peter Herten. “An update on molecular counting in fluorescence microscopy”. In: *The International Journal of Biochemistry & Cell Biology* 135 (2021), p. 105978.
- [37] Johan Hummert et al. “Photobleaching step analysis for robust determination of protein complex stoichiometries”. In: *Molecular biology of the cell* 32.21 (2021), ar35.
- [38] Mohamadreza Fazel and Michael J Wester. “Analysis of super-resolution single molecule localization microscopy data: A tutorial”. In: *AIP advances* 12.1 (2022).
- [39] J Shepard Bryan IV, Ioannis Sgouralis, and Steve Pressé. “Diffraction-limited molecular cluster quantification with Bayesian nonparametrics”. In: *Nature computational science* 2.2 (2022), pp. 102–111.
- [40] Prabuddha Sengupta et al. “Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis”. In: *Nature methods* 8.11 (2011), pp. 969–975.
- [41] Fuwei Yang et al. “dSTORM-Based Single-Cell Protein Quantitative Analysis Can Effectively Evaluate the Degradation Ability of PROTACs”. In: *ChemBioChem* 24.4 (2023), e202200680.
- [42] Ralf Jungmann et al. “Quantitative super-resolution imaging with qPAINT”. In: *Nature methods* 13.5 (2016), pp. 439–442.
- [43] Geoffrey C Rollins et al. “Stochastic approach to the molecular counting problem in super-resolution microscopy”. In: *Proceedings of the National Academy of Sciences* 112.2 (2015), E110–E118.

- [44] Steve Pressé and Ioannis Sgouralis. *Data Modeling for the Sciences: Applications, Basics, Computations*. Cambridge University Press, 2023.
- [45] Zeliha Kilic et al. “Extraction of rapid kinetics from smFRET measurements using integrative detectors”. In: *Cell Reports Physical Science* 2.5 (2021).
- [46] Zeliha Kilic, Ioannis Sgouralis, and Steve Pressé. “Generalizing HMMs to continuous time for fast kinetics: Hidden Markov jump processes”. In: *Biophysical journal* 120.3 (2021), pp. 409–423.
- [47] Mohamadreza Fazel et al. “Fluorescence Microscopy: a statistics-optics perspective”. In: *Reviews of Modern Physics* 96.2 (2024), p. 025003.
- [48] Ioannis Sgouralis and Steve Pressé. “An introduction to infinite HMMs for single-molecule data analysis”. In: *Biophysical journal* 112.10 (2017), pp. 2021–2029.
- [49] Ayush Saurabh et al. “Single-photon smFRET. I: Theory and conceptual basis”. In: *Biophysical Reports* 3.1 (2023).
- [50] Mark Jäger et al. “Analysis of Single-Molecule Fluorescence Spectroscopic Data with a Markov-Modulated Poisson Process”. In: *ChemPhysChem* 10.14 (2009), pp. 2486–2495.
- [51] J Shepard Bryan and Steve Pressé. “Learning continuous potentials from smfret”. In: *Biophysical journal* 122.2 (2023), pp. 433–441.
- [52] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*. Vol. 4. 4. Springer, 2006.
- [53] Sungwook Woo and Paul WK Rothemund. “Programmable molecular recognition based on the geometry of DNA nanostructures”. In: *Nature chemistry* 3.8 (2011), pp. 620–627.
- [54] Johannes Schindelin et al. “Fiji: an open-source platform for biological-image analysis”. In: *Nature methods* 9.7 (2012), pp. 676–682.
- [55] Martin Ovesný et al. “ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging”. In: *Bioinformatics* 30.16 (2014), pp. 2389–2390.
- [56] Philippe Thevenaz, Urs E Ruttimann, and Michael Unser. “A pyramid approach to subpixel registration based on intensity”. In: *IEEE transactions on image processing* 7.1 (1998), pp. 27–41.
- [57] Ralf Jungmann et al. “Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami”. In: *Nano letters* 10.11 (2010), pp. 4756–4761.
- [58] Ayush Saurabh et al. “Quantitative image reconstruction for microscopy with correct sub-pixel photon statistics”. In: *Biophysical journal* 122.3 (2023), 132a.

Supplementary Information

S1 Docking site staple strands	5'-end	3'-end	1xS1 NRO	2xS1 NRO	5xS1 NRO
AATTGAGAAATTCGTCCAGACGACTAAACCAATTTCCCTCTACCCACTACATCAC	14[207]	12[208]			X
CTGAGCAA AAATTAATTACATTTTGGGTTA TTTCCCTCTACCCACTACATCAC	18[175]	16[176]			X
CCAATAGCTCATCGTAGGAATCATGGCATCAA TTTCCCTCTACCCACTACATCAC	11[160]	12[144]	X	X	X
TTAACGTCTAACATAAAAACAGGTAACGGATTTCCCTCTACCCACTACATCAC	10[175]	8[176]			X
CATGTAATAGAAATATAAAGTACCAAGCCGTTTCCCTCTACCCACTACATCAC	14[175]	12[176]		X	X

Table S11: External S1 docking site staple strands of the NRO DNA Origami. Sequences are denoted from 5'- to 3'-end. The S1 docking site staple strands exhibit an over 20nt long docking site, marked in red, on the 3'-end. The numbers for the 5'- end 3'-end of the staples represent the helix number in the corresponding caDNA file. Number in brackets represent the starting and ending position of the staple in the corresponding helix.

S2 Docking site staple strands	5'-end	3'-end
GCAAGGCCTCACCAGTAGCACCATGGGCTTGATAAACATTCCCTAACTTCTCATA	5[160]	6[144]
TTGACAGGCCACCACCAGAGCCGCGATTGTATAAACATTCCCTAACTTCTCATA	3[160]	4[144]
GATGGTTTGAACGAGTAGTAAATTTACCATTATAAACATTCCCTAACTTCTCATA	6[143]	5[159]

Table SI2: Modified staple strands of the NRO DNA Origami. Sequences are denoted from 5'- to 3'-end. The three S2 docking site staple strands exhibit an over 20nt long docking site, marked in green for colocalization, on the 3'-end. For immobilization, the biotinylated staple strands are modified with biotin on the 3'-end. The numbers for the 5'- end 3'-end of the staples represent the helix number in the corresponding caDNAno file. Number in brackets represent the starting and ending position of the staple in the corresponding helix.

Table SI3: Unmodified staple strands of NRO DNA origami. Sequences are denoted from 5'- to 3'-end. The numbers for the 5'-end 3'-end of the staples represent the helix number in the corresponding caDNA file. Number in brackets represent the starting and ending position of the staple in the corresponding helix.

Unmodified staple strands	5'-end	3'-end
CATAAATCTTTGAATACCAAGTGTTAGAAC	17[224]	19[223]
AAGCCTGGTACGAGCCGGAAGCATAGATGATG	20[143]	19[159]
TCATTCAGATGCGATTTTAAAGAACAGGCATAG	5[96]	7[95]
GCCATCAAGCTCATTTTTTAAACCACAAATCCA	16[143]	15[159]
TATAACTAACAAAGAACGCGAGAACGCCAA	16[175]	14[176]
TTGCTCCTTTCAAATATCGCGTTTGAGGGGGT	10[111]	8[112]
GTATAGCAAACAGTTAATGCCCAATCCTCA	1[224]	3[223]
AAAGTCACAAAATAAACAGCCAGCGTTTTA	9[224]	11[223]
GGCCTTGAAGAGCCACCACCCTCAGAAACCAT	3[192]	5[191]
TTAACGTCTAACATAAAAAACAGGTAACGGA	10[175]	8[176]
AGTATAAAGTTCAGCTAATGCAGATGTCTTTC	14[239]	12[240]
GATGTGCTTCAGGAAGATCGCACAAATGTGA	18[79]	16[80]
TCAAATATAACCTCCGGCTTAGGTAACAATTT	15[192]	17[191]
TTTCGGAAGTGCCGTCGAGAGGGTGAGTTTCG	2[207]	0[208]
GAGGGTAGGATTCAAAAGGGTGAGACATCCAA	14[111]	12[112]
TATTAAGAAGCGGGGTTTTGCTCGTAGCAT	2[175]	0[176]
GCCCTTCAGAGTCCACTATTAAGGGTGCCGT	21[64]	23[63]
ATGCAGATACATAACGGGAATCGTCATAAATAAAGCAAAG	7[56]	9[63]
AGCCAGCAATTGAGGAAGGTTATCATCATTTT	22[207]	20[208]
TAAATGAATTTTCTGTATGGGATTAATTTCTT	0[111]	1[95]
AAACAGCTTTTTGCGGGATCGTCAACACTAAA	1[96]	3[95]
CGGATTCTGACGACAGTATCGGCCGCAAGGCGATTAAGTT	16[63]	18[56]
GCAATTCACATATTCCTGATTATCAAAGTGTA	19[160]	20[144]
GCGCAGACAAGAGGCAAAAGAATCCCTCAG	4[79]	2[80]
AGAGAGAAAAAATGAAAATAGCAAGCAAACCT	9[160]	10[144]
GACAAAAGGTAAAGTAATCGCCATATTTAACAAAACCTTTT	13[184]	15[191]
ACACTCATCCATGTTACTTAGCCGAAAGCTGC	3[96]	5[95]
CTACCATAGTTTGAGTAACATTTAAAATAT	19[224]	21[223]
TATATTTTGTTCATTGCCTGAGAGTGGAAGATTGTATAAGC	13[64]	15[71]
CGGATTGCAGAGCTTAATTGCTGAAACGAGTA	9[64]	11[63]
TAAATCATATAACCTGTTTAGCTAACCTTTAA	12[111]	10[112]
GTACCGCAATTCTAAGAACGCGAGTATTATTT	12[207]	10[208]
TCTTCGCTGCACCGCTTCTGGTGCGGCCTTCC	18[111]	16[112]
GATTTAGTCAATAAAGCCTCAGAGAACCCTCA	11[64]	13[63]
GCAAGGCCTCACCAGTAGCACCATGGGCTTGA	5[160]	6[144]
ATTACCTTTGAATAAGGCTTGCCCAAATCCGC	6[111]	4[112]
CTTATCATTTCCCGACTTGCGGGAGCCTAATTT	12[239]	10[240]
TTATACCACCAAATCAACGTAACGAACGAG	6[79]	4[80]
GTAATAAGTTAGGCAGAGGCATTTATGATATT	13[160]	14[144]
CAACCGTTTCAAATCACCATCAATTCGAGCCA	14[143]	13[159]
GATGGTTTGAACGAGTAGTAAATTTACCATTA	6[143]	5[159]
GCACAGACAATATTTTTGAATGGGGTCAGTA	23[224]	22[240]

AGCAAGCGTAGGGTTGAGTGGTGTAGGGAGCC	21[96]	23[95]
TCCACAGACAGCCCTCATAGTTAGCGTAACGA	0[175]	0[144]
TCACCAGTACAAACTACAACGCCTAGTACCAG	0[207]	1[191]
ATTATACTAAGAAACCACCAGAAGTCAACAGT	19[192]	21[191]
TAAGAGCAAATGTTTAGACTGGATAGGAAGCC	7[96]	9[95]
ATACATACCGAGGAAACGCAATAAGAAGCGCATTAGACGG	7[192]	9[199]
CAACTGTTGCGCCATTGCGCATTCAAACATCA	18[143]	17[159]
GATGGCTTATCAAAAAGATTAAGAGCGTCC	10[79]	8[80]
TAGGTAAACTATTTTTGAGATCAAACGTTA	13[96]	15[95]
AGGCAAAGGGAAGGGCGATCGGCAATTCCA	17[128]	19[127]
ATTATCATTCAATATAATCCTGACAATTAC	20[175]	18[176]
GAAATTATTGCCTTTAGCGTCAGACCGGAACC	6[239]	4[240]
AATGGTCAACAGGCAAGGCAAAAGAGTAATGTG	11[96]	13[95]
CCAATAGCTCATCGTAGGAATCATGGCATCAA	11[160]	12[144]
ATACCCAACAGTATGTTAGCAAATTAGAGC	8[175]	6[176]
ATAAGGGAACCGGATATTCATTACGTCAGGACGTTGGGAA	4[63]	6[56]
CACCAGAAAGGTTGAGGCAGGTCATGAAAG	4[175]	2[176]
ATCCCAATGAGAATTAACCTGAACAGTTACCAG	10[207]	8[208]
CATGTAATAGAATATAAAGTACCAAGCCGT	14[175]	12[176]
CCAACAGGAGCGAACCAGACCGGAGCCTTTAC	10[143]	9[159]
GCTATCAGAAATGCAATGCCTGAATTAGCA	14[79]	12[80]
GACCTGCTCTTTGACCCCGAGCGAGGGAGTTA	4[111]	2[112]
AGGAACCCATGTACCGTAACACTTGATATAA	0[239]	1[223]
CAGCGAAACTTGCTTTCGAGGTGTTGCTAA	2[79]	0[80]
GCTTTCGATTACGCCAGCTGGCGGCTGTTTC	17[96]	19[95]
ACAACCTTCAACAGTTTCAGCGGATGTATCGG	0[79]	1[63]
CAGCAAAGGAAACGTCACCAATGAGCCGC	6[175]	4[176]
ACCTTTTTATTTTAGTTAATTTTCATAGGGCTT	16[207]	14[208]
CGATAGCATTGAGCCATTTGGGAACGTAGAAA	5[192]	7[191]
GCCCGAGAGTCCACGCTGGTTTGCAGCTAACT	22[111]	20[112]
ATTTTAAATCAAAATTATTTGCACGGATTCG	20[239]	18[240]
ACCTTGCTTGGTCAGTTGGCAAAGAGCGGA	22[175]	20[176]
CTGAGCAAAAATTAATTACATTTTGGGTTA	18[175]	16[176]
CCTGATTGCAATATATGTGAGTGATCAATAGT	18[239]	16[240]
TCAATATCGAACCTCAAATATCAATTCCGAAA	21[160]	22[144]
AAAGGCCGGAGACAGCTAGCTGATAAATTAATTTTGT	13[120]	15[127]
CTTTAGGGCCTGCAACAGTGCCAATACGTG	21[224]	23[223]
AATAGTAAACACTATCATAACCCTCATTGTGA	8[111]	6[112]
TCACCGACGCACCGTAATCAGTAGCAGAACCG	6[207]	4[208]
GCCCGTATCCGGAATAGGTGTATCAGCCCAAT	2[239]	0[240]
TGTAGCCATTAATAATTCGCATTAATGCCGGA	16[111]	14[112]
TCGGCAAATCCTGTTTGATGGTGGACCTCAA	22[143]	21[159]
TGACAACTCGCTGAGGCTTGCATTATACCA	1[128]	3[127]
CCACCCTCTATTCACAAACAAATACCTGCCTA	4[207]	2[208]
CCCGATTTAGAGCTTGACGGGAAAAAGAATA	23[96]	22[112]
AAGTAAGCAGACACCACGGAATAATATTGACG	8[239]	6[240]
AAATTAAGTTGACCATTAGATACTTTTGCG	12[79]	10[80]
CACATTAATAATTGTTATCCGCTCATGCGGGCC	20[111]	18[112]
TTAAAGCCAGAGCCGCCACCCTCGACAGAA	3[224]	5[223]

ATATTCGGAACCATCGCCACGCAGAGAAGGA	2[143]	1[159]
TTCTACTACGCGAGCTGAAAAGGTTACCGCGC	12[143]	11[159]
AACGTGGCGAGAAAGGAAGGAAACCAGTAA	23[128]	23[159]
GAATTTATTTAATGGTTTGAAATATTCTTACC	16[239]	14[240]
AGCGCGATGATAAATTGTGTCGTGACGAGA	3[128]	5[127]
AACGCAAAGATAGCCGAACAAACCCTGAAC	7[224]	9[223]
AGAAAACAAAGAAGATGATGAAACAGGCTGCG	17[160]	18[144]
CGCGCAGATTACCTTTTTTAATGGGAGAGACT	18[207]	16[208]
CACAACAGGTGCCTAATGATGCCAGCAG	19[128]	21[127]
GCGGAACATCTGAATAATGGAAGGTACAAAAT	20[207]	18[208]
TAAAAGGGACATTCTGGCCAACAAAGCATC	23[160]	22[176]
AATTGAGAATTCTGTCCAGACGACTAAACCAA	14[207]	12[208]
GCGAAAAATCCCTTATAAATCAAGCCGGCG	21[128]	23[127]
AACACCAAATTTCAACTTTAATCGTTTACC	5[128]	7[127]
TAAATCAAATAAATTCGCGTCTCGGAAACC	15[128]	17[127]
GAAACGATAGAAGGCTTATCCGGTCTCATCGAGAACAAGC	10[191]	12[184]
GCCTCCCTCAGAATGGAAAGCGCAGTAACAGT	4[239]	2[240]
GCGAACCTCCAAGAACGGGTATGACAATAA	11[224]	13[223]
TTAGGATTGGCTGAGACTCCTCAATAACCGAT	1[160]	2[144]
ATCGCAAGTATGTAAATGCTGATGATAGGAAC	15[160]	16[144]
GCGGATAACCTATTATTCTGAAACAGACGATT	1[192]	3[191]
AAGGAAACATAAAGGTGGCAACATTATCACCG	8[207]	6[208]
ACCCTTCTGACCTGAAAGCGTAAGACGCTGAG	23[192]	22[208]
ATATTTTGCTTTCATCAACATTATCCAGCCA	15[96]	17[95]
TCAAGTTTCATTAAAGGTGAATATAAAAAGA	5[224]	7[223]
TCTAAAGTTTTGTGCTCTTCCAGCCGACAA	0[143]	1[127]
TTCCAGTCGTAATCATGGTCATAAAAGGGG	20[79]	18[80]
AAAGCACTAAATCGGAACCCTAATCCAGTT	23[64]	22[80]
AATACTGCCCAAAGGAATTACGTGGCTCA	8[79]	6[80]
TTTATCAGGACAGCATCGGAACGACACCAACCTAAAACGA	1[64]	3[71]
TTGACAGGCCACCACCAGAGCCGCGATTTGTA	3[160]	4[144]
CTGTGTGATTGCGTTGCGCTCACTAGAGTTGC	19[96]	21[95]
GCGAGTAAAAATATTTAAATTGTTACAAAG	16[79]	14[80]
TAGAGAGTTATTTTCATTTGGGATAGTAGCATT	10[127]	12[120]
CGAAAGACTTTGATAAGAGGTCATATTTGCA	9[96]	11[95]
TCATCGCCAACAAAGTACAACGGACGCCAGCA	4[143]	3[159]
TTAACACCAGCACTAACAACCTAATCGTTATTA	22[239]	20[240]
TTATTACGAAGAACTGGCATGATTGCGAGAGG	7[160]	8[144]
GCCAGTTAGAGGGTAATTGAGCGCTTTAAGAA	10[239]	8[240]
ACAACATGCCAACGCTCAACAGTCTTCTGA	13[224]	15[223]
CATTTGAAGGCGAATTATTCATTTTTGTTTGG	17[192]	19[191]
TGAAAGGAGCAAATGAAAAATCTAGAGATAGA	21[192]	23[191]
TGGAACAACCCGCTGGCCCTGAGGCCCGCT	22[79]	20[80]
TACCGAGCTCGAATTCGGGAAACCTGTCGTGCAGCTGATT	19[56]	21[63]
GTTTATTTTGTGTCACAATCTTACCGAAGCCCTTTAATATCA	7[248]	9[255]
ACAAACGGAAAAGCCCCAAAAACACTGGAGCA	16[47]	14[48]
GTTTATCAATATGCGTTATACAAACCGACCGTGTGATAAA	13[256]	15[263]
ACGGCTACAAAAGGAGCCTTTAATGTGAGAAT	2[47]	0[48]
GACCAACTAATGCCACTACGAAGGGGGTAGCA	4[47]	2[48]

AAGGCCGCTGATACCGATAGTTGCGACGTTAG	2[111]	0[112]
CTCCAACGCAGTGAGACGGGCAACCAGCTGCA	22[47]	20[48]
ACCGATTGTCGGCATTTCGGTCATAATCA	6[271]	4[272]
CAGAAGATTAGATAATACATTTGTCGACAA	22[271]	20[272]
TGCATCTTTCCCAGTCACGACGGCCTGCAG	17[32]	19[31]
TTAGTATCACAATAGATAAGTCCACGAGCA	14[271]	12[272]
GTTTTAACTTAGTACCGCCACCCAGGCCA	2[271]	0[272]
TTAATGAACTAGAGGATCCCCGGGGGTAACG	20[47]	18[48]
CTTTTACAAAATCGTCGCTATTAGCGATAG	18[271]	16[272]
ATCCCCCTATACCACATTCAACTAGAAAAATC	8[47]	6[48]
AGAAAGGAACAACATAAAGGAATTCAAAAAAA	0[47]	1[31]
TTTTATTTAAGCAAATCAGATATTTTTTGT	12[175]	10[176]
AGCCACCCTGTAGCGCTTTTCAAGGGAGGGAAGGTAAA	4[255]	6[248]
AACAAGAGGGATAAAAATTTTTAGCATAAAGC	14[47]	12[48]
GCCGTCAAAAAACAGAGGTGAGGCCTATTAGT	21[256]	23[255]
TGTAGAAATCAAGATTAGTTGCTCTTACCA	12[271]	10[272]
GAGAGATAGAGCGTCTTTCCAGAGTTTTGAA	9[256]	11[255]
CCACCCTCATTTTCAGGGATAGCAACCGTACT	0[271]	1[255]
CTTTAATGCGCGAACTGATAGCCCCACCAG	23[256]	22[272]
CCAGGGTTGCCAGTTTGAGGGACCCGTGGGA	18[47]	16[48]
CAAATCAAGTTTTTTGGGGTCGAAACGTGGA	23[32]	22[48]
ACGCTAACACCCACAAGAATTGAAAATAGC	10[271]	8[272]
CTTTTGCAGATAAAAACCAAAAATAAAGACTCC	8[143]	7[159]
TACGTTAAAGTAATCTTGACAAGAACCGAACT	6[47]	4[48]
TAATCAGCGGATTGACCGTAATCGTAACCG	15[32]	17[31]
TTTTCACTCAAAGGGCGAAAAACCATCACC	21[32]	23[31]
GCCTTAAACCAATCAATAATCGGCACGCGCCT	11[256]	13[255]
AATAGCTATCAATAGAAAATTCAACATTCA	8[271]	6[272]
CATCAAGTAAAACGAACATAACGAGTTGAGA	5[32]	7[31]
CAGGAGGTGGGGTCAGTGCCTTGAGTCTCTGAATTTACCG	1[256]	3[263]
AAATCACCTTCCAGTAAGCGTCAGTAATAA	4[271]	2[272]
CTCGTATTAGAAATTGCGTAGATACAGTAC	20[271]	18[272]
TTTACCCCAACATGTTTTAAATTTCCATAT	9[32]	11[31]
CCTAAATCAAAATCATAGGTCTAAACAGTA	15[224]	17[223]
GTCGACTTCGGCCAACGCGCGGGTTTTTC	19[32]	21[31]
CGTAAACAGAAATAAAAATCCTTTGCCGAAAGATTAGA	19[248]	21[255]
AGGCTCCAGAGGCTTTGAGGACACGGGTAA	1[32]	3[31]
GAGAAGAGATAACCTTGCTTCTGTTTCGGGAGAAACAATAA	16[255]	18[248]
TTTAGGACAAATGCTTTAAACAATCAGGTC	7[32]	9[31]
AATACGTTTGAAGAGGACAGACTGACCTT	3[32]	5[31]
CTTAGATTTAAGGCGTTAAATAAAGCCTGT	16[271]	14[272]
TAAATCGGGATTCCCAATTCTGCGATATAATG	12[47]	10[48]
AACAGTTTTGTACCAAAAACATTTTATTTTC	11[32]	13[31]
CTGTAGCTTGACTATTATAGTCAGTTCATTGA	10[47]	8[48]
AGACGACAAAGAAGTTTTGCCATAATTCGAGCTTCAA	7[128]	9[135]
AACGCAAAATCGATGAACGGTACCGGTTGA	13[32]	15[31]

Figure	$k_{\text{bind}}(s^{-}nM^{-})$	$k_{\text{un}}(s^{-})$	$k_{\text{bleach}}(s^{-})$	Number of states
4	1×10^{-3}	20×10^{-3}	10×10^{-5}	2
4	2×10^{-3}	20×10^{-3}	10×10^{-5}	2
4	5×10^{-3}	20×10^{-3}	10×10^{-5}	2
5	1×10^{-3}	10×10^{-3}	10×10^{-5}	2
5	1×10^{-3}	20×10^{-3}	10×10^{-5}	2
5	1×10^{-3}	50×10^{-3}	10×10^{-5}	2
6	1×10^{-3}	20×10^{-3}	10×10^{-5}	2
6	1×10^{-3}	20×10^{-3}	20×10^{-5}	2
6	1×10^{-3}	20×10^{-3}	50×10^{-5}	2
7	1×10^{-3}	20×10^{-3}	10×10^{-5}	2

Table SI4: Parameters used to create the simulated data. The three values were used in the forward model described in section 2. All data for figure 7 shares the same parameters with variation only in the number of data points.

Figure	$k_{\text{bind}}(10^{-3} s^{-}nM^{-})$	$k_{\text{un}}(10^{-3} s^{-})$	$k_{\text{bleach}}(10^{-5} s^{-})$
4	1.031 ± 0.012	23.950 ± 1.086	8.800 ± 1.418
4	1.971 ± 0.018	23.500 ± 0.680	10.260 ± 0.913
4	4.588 ± 0.030	24.620 ± 0.504	8.446 ± 0.674
5	1.048 ± 0.014	9.623 ± 0.612	9.661 ± 0.848
5	1.031 ± 0.012	23.950 ± 1.086	8.800 ± 1.418
5	1.046 ± 0.012	45.650 ± 1.586	10.650 ± 1.890
6	1.031 ± 0.012	23.950 ± 1.086	8.800 ± 1.418
6	1.038 ± 0.011	22.470 ± 1.455	20.780 ± 1.974
6	1.014 ± 0.012	25.980 ± 1.984	43.150 ± 2.852
7	1.031 ± 0.012	23.950 ± 1.086	8.800 ± 1.418
7	1.008 ± 0.021	21.720 ± 1.400	11.940 ± 1.937
7	1.050 ± 0.033	23.480 ± 2.720	8.741 ± 3.470
8	2.774 ± 0.077	17.200 ± 0.538	0.725 ± 0.355
8	3.515 ± 0.044	14.460 ± 0.234	0.201 ± 0.161
8	8.110 ± 0.067	17.030 ± 0.220	0.419 ± 0.207

Table SI5: Results from the Bayesian inference scheme. Summary of the results shown in all figures. The values are shown as *mean* \pm *SD*.