Single molecule imaging to characterize protein interactions with the environment

Megan Julia Armstrong

Submitted in partial fulfillement of the

requirements for the degree of

Doctor of Philosophy

in the Graduate School of Arts and Sciences

Columbia University

2019

ABSTRACT

Single molecule imaging to characterize protein interactions with the environment

Megan Julia Armstrong

In the past decade, single molecule imaging has advanced our understanding of processes at the molecular scale. Total internal reflection fluorescence (TIRF) microscopy is one implementation in particular that has been extensively applied in the study of protein adsorption to surfaces. The spatial and temporal resolution provided by TIRF has enabled dynamic measurements of individual proteins in solution, where previously only bulk measurements or static electron microscopy observations were possible. The ability to study individual proteins has revealed and sometimes clarified the complex interactions at their interfaces. Here, the utility of TIRF is expanded to introduce a new model of protein adsorption to the suface and to study the protein interface in contact with solution.

Protein adsorption to surfaces has implications in surface biocompatibility, protein separation, and pharmaceutical nanoparticle development. For this reason, the phenomenon has been quantitatively by a variety of techniques, including single molecule imaging. The key data are the protein lifetimes on the surface, which have been shown to be broadly distributed and well-approximated by the sum of several exponential functions. The determined desorption rate constants are thought to reflect different interaction types between surface and protein, but the rates are not typically linked to a specific physical interaction. In the first part of this thesis, we establish appropriate imaging conditions and analysis methods

for TIRF. A robust survival analysis technique is applied to capture the range of protein adsorption kinetics. In the second part, we utilize single molecule lifetime data from the adsorption of fibrinogen and bovine serum albumin (BSA) to glass surfaces and discover a heavy-tailed distribution: a very small fraction of proteins adsorbs effectively permanently, while the majority of proteins adsorb for a very short time. We then demonstrate that this characteristic power law behavior is well described by a model with a novel interpretation of the complex protein adsorption process.

The second half of the thesis extends TIRF to study the solution-facing interface of the protein as opposed to the surface facing interface by establishing the parameters for a super-resolution imaging technique. Point accumulation for imaging nanoscale topography (PAINT) generates high-resolution images of the sample of interest through the positional tracking of many temporally-distinct instances of a fluorescent probe binding to the sample. Previously, this technique has been applied in the mapping of DNA nanostructures. Here, in the third part, we apply PAINT to the study of proteins. First, a workstream is established for a model system of Nile red and BSA. The kinetic parameters for the system are established to allow rational design of PAINT experiments with this system. The on-rate and off-rate for Nile red are determined. Additionally, the binding model between the two components is tested by studying how the presence of an inhibitor effects the parameters.

In the final part, TIRF is used to study the protein-solution interface to examine the glycosylation of immunoglobulin A 1 (IgA1). Over 50% of eukaryotic proteins are glycosylated, and the glycan sequence is simultaneously difficult to study and crucial in the many functional roles proteins play. The glycosylation of IgA1, for example, plays a key role in the pathophysiology of IgA1 nephropathy. Lectins are proteins that bind to specifc glycan sequences and are often used to isolate glycosylated proteins. In this study, the appropriate surface conditions are established to allow specific binding between lectins and IgA1 glycans. The association and dissociation rate between lectins specific for the glycans on IgA1 are measured and affinity constants calculated. These efforts will help to rationally design

experiments in the future to elucidate unknown glycan sequences on proteins.

Contents

List of Figures

Acknowledgements

I am immensely grateful to the friends, family, and colleagues that have been with me throughout this process!

Of course, I'd like to express my gratitude to my advisor, Henry. Looking back, it seems that you took quite a risk hiring me, which I'm sure was very calculated, but I'm so grateful that you did! I would not trade this experience for anything. Thank you for the debates, jokes, career support, copious amounts of caffeine, and of course, the academic mentoring.

I'd also like to thank my family who established early on that they cared far more about my character than my accomplishments. That has given me so much freedom to engage with the world in whatever way I see fit. Thank you.

Thank you to my best friend Claire who made the effort to stay friends and woke up early to have 7:00 am Skype calls while living in Cambodia. That was one of the toughest years of my life and you made it so much better. Thanks for just being my friend, thanks for putting in the effort, thanks for debugging my python code via Messenger while you were in a meeting, thanks for validating all my feelings from the completely crushed to the incredibly proud, thanks for receiving my bad puns, thanks for sometimes updating your apps so we can communicate, and thanks for graduating within a month of me so that we can be free together afterwards!!

Thank you to Brad, who filled my life with so much love and growth in a time that I expected to be isolating and fatiguing!

Finally, I'd like to acknowledge the community of people that have made grad school a

Dedication

I dedicate this to my family without whom none of this would have been possible.

Chapter 1

Introduction

1.1   Single molecule imaging for protein studies

In the past decade, single molecule imaging has advanced our understanding of processes at the molecular scale. The ability to follow the dynamic behavior of individual molecules has elucidated the details of numerous processes: from the mechanism of molecular motors [1] to the intricate behaviors of cell membrane receptors [2] to enzyme kinetics [3]. Total internal reflection fluorescence (TIRF) microscopy is one implementation in particular that has been extensively applied in the study of protein adsorption to surfaces. The spatial and temporal resolution provided by TIRF has enabled dynamic measurements of individual proteins in solution, where previously only bulk measurements or static electron microscopy observations were possible. The ability to study individual proteins has revealed and sometimes clarified the complex interactions at their interfaces. Here, the utility of TIRF is expanded to introduce a new model of protein adsorption to the surface and to study the protein interface in contact with solution.

Protein adsorption to surfaces has implications in surface biocompatibility, protein separation, and pharmaceutical nanoparticle development. For this reason, the phenomenon has been quantitatively by a variety of techniques, including single molecule imaging. The

key data are the protein lifetimes on the surface, which have been shown to be broadly distributed and well-approximated by the sum of several exponential functions. The determined desorption rate constants are thought to reflect different interaction types between surface and protein, but the rates are not typically linked to a specific physical interaction.

1.2 Organization of Dissertation

In the first part of this thesis, we establish appropriate imaging conditions and analysis methods for TIRF. The materials preparation and imaging parameters are described. Image processing and data analysis developed for single molecule TIRF experiments are described in details. A robust survival analysis technique is developed to capture the range of protein adsorption kinetics.

In the second part, we utilize single molecule lifetime data from the adsorption of fibrinogen and bovine serum albumin (BSA) to glass surfaces to introduce a new mechanistic model for protein adsorption. The model describes the data well and is linked to physical meanings in the adsorption process.

The second half of the thesis extends TIRF to study the solution-facing interface of the protein as opposed to the surface facing interface by establishing the parameters for a super-resolution imaging technique. Point accumulation for imaging nanoscale topography (PAINT) generates high-resolution images of the sample of interest through the positional tracking of many temporally-distinct instances of a fluorescent probe binding to the sample. Previously, this technique has been applied in the mapping of DNA nanostructures.

In the third part, we begin to apply PAINT to the study of proteins. First, a workstream is established for a model system of Nile red and BSA. The kinetic parameters for the system are established to allow rational design of PAINT experiments with this system. The on-rate and off-rate for Nile red are determined. Additionally, the binding model between the two components is tested by studying how the presence of an inhibitor effects the parameters.

In the final part, TIRF is used to study the protein-solution interface to examine the glycosylation of immunoglobulin A 1 (IgA1). Over 50% of eukaryotic proteins are glycosylated, and the glycan sequence is simultaneously difficult to study and crucial in the many functional roles proteins play. The glycosylation of IgA1, for example, plays a key role in the pathophysiology of IgA1 nephropathy. Lectins are proteins that bind to specifc glycan sequences and are often used to isolate glycosylated proteins. In this study, the appropriate surface conditions are established to allow specific binding between lectins and IgA1 glycans. The association and dissociation rate between lectins specific for the glycans on IgA1 are measured and affinity constants calculated. These efforts will help to rationally design experiments in the future to elucidate unknown glycan sequences on proteins.

Chapter 2

Single molecule imaging and analysis system

2.1 Experimental

2.1.1 Flow cell preparation

Long, 24x60 mm (VWR #1), and short, 22x22 mm (Fisherbrand, #1), glass coverslips were prepared for the flow cell. The coverslips were rinsed three times in ultrapure water (10 M-cm, Barnstead) then sonicated for five minutes and rinsed three times again in ultrapure water. They were then sonicated for 20 minutes in acetone, rinsed three times in ultrapure water, sonicated for 20 minutes in 1N KOH and rinsed three more times in ultrapure water before being oven dried. Dry slides were treated with a UV/ozone cleaner (Bioforce Nanoscience) for 20 minutes on each side before being rinsed three times in ultrapure water, sonicated for 20 minutes, rinsed three more times, and finally oven dried.

This extensive procedure was adopted because the residual contaminants on the flow cell surface have the appearance and kinetics of the objects of interest. The factors that were found to contribute most significantly to the presence of fluorescent contaminants were 1) the choice of solvent, 2) the power and number of sonication steps, 3) the rinsing repetition,

and 4) the time between cleaning and experiment. Regarding the solvents, ethanol not rated for its purity introduced about significant contamination, on the order of 100 spots per 1 s frame with no sign of diminishing in number after 20 minutes of acquisition. Methanol (Fisher), dimethylsulfoxide (DMSO, Sigma-Aldrich) and dimethylformamide (DMF, Sigma) produced contaminant-free videos in the absence of other contamination. An additional sonication step after ozone treatment reduced the contamination by half but continued to appear throughout a 20-minute acquisition. Rinsing three times after ozone treatment and sonication finally reduced the number of fluorescent species per frame down to about 3 in the initial 5 seconds, and none for the remainder of the video. These observations were all made within 3 days of the final rinsing. Measurements made after this time decrease in quality and increase in the number of fluorescent contaminants and their persistence. After 3 days, the coverslips would be recleaned before use. Clean coverslips were stored covered in the same jars that they were cleaned in at room temperature ($23 \pm 2°C$).

Flow cells were constructed by attaching a short cover slip to the middle of a long coverslip with double-sided tape on each long edge of the longer coverslip in order to create one or two channel between the strips of tape, as shown in Fig. 2.1.
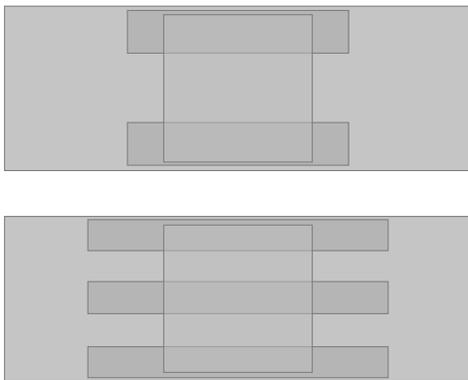
Figure 2.1: Single channel (top) and double channel (bottom) flow cell schematic from the top view.

2.1.2   Image Acquisition

Videos, such as the one in Fig. 2.2 were acquired with a Nikon Ti-Eclipse Inverted microscope (Morrell Instrument Company) with total internal reflection fluorescence (TIRF) module using a $100\times$ 1.49 NA oil objective and a back-illuminated EMCCD camera (Andor iXon). Samples were illuminated during the exposure time only with various lasers: 642 nm or 488 nm LuxX diode lasers (Omicron Laserprodukte GmbH) or 561 nm Sapphire™laser (Coherent). The exposure time for each image was determined based on the experimental needs, balancing photobleaching, photoblinking, and time resolution.
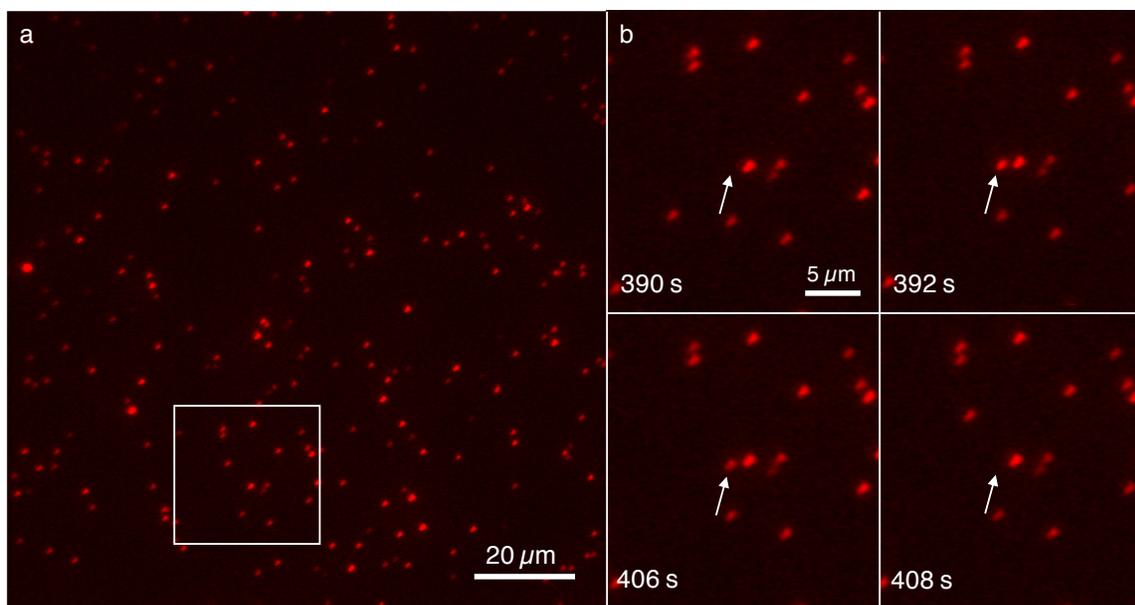
Figure 2.2: A single frame from a single molecule imaging experiment with 40 pM Alexa 647-labeled fibrinogen at 5% laser power (a). In (b), a region of interest outlined in (a) shows the sequence of a protein, indicated by the arrow, adsorbing at 392 s then desorbing 16 seconds later at 408 s.

2.2 Analysis

2.2.1 Image Analysis

Image Processing

Videos were processed with the Matlab package developed by R. Parthasaraty [4] to detect, localize, and preliminarily link objects in the field of view. The point-spread function is much larger than the object itself, thus it depends on the analysis to detect the true sub-pixel position of the object below the diffraction limit. This analysis package uses a radial symmetry method to accomplish this. The end output of the program is trajectories of object positions on each frame, as well as the integrated intensity of the object in the image.

Drift Correction

The alignment of protein localizations on different frames generates the trajectories through time that represent the path and surface residence time of individual proteins. Stage drift affects this process by artificially introducing motion between frames, which increases the likelihood that the paths of two proteins very close to each may be linked, or that linking fails altogether because of the distance between localizations on different frames. These effects lead to a decrease in localization accuracy over the course of the entire trajectory as the points are spread out along the path of the stage drift, rather than clustered in the expected distribution of a point spread function. In order to correct this, an algorithm was developed to detect and remove the stage drift from the trajectories, the code for which is in Appendix, .1 and available on Github at https://github.com/meganarms/SSA. The principle of the algorithm developed was to use a cross-correlation method to determine the lateral shift between frames. It is similar to methods described by Peterson, Wang, and Mlodzianoski[5, 6, 7], but with no *a priori* structure to which to compare the drift, the drift path is determined by comparing the positions of proteins with long residence times across different frames.

First, trajectories were created with a software package developed by R. Parthasarathy[8]. The package has several user-defined parameters. The bandpass filter, set to 7 pixels, sets the maximum radius for an object so that large variations in intensity, i.e. due to uneven illumination, may be eliminated. The standard deviation intensity threshold determines the number of standard deviations above the mean intensity a spot must be in order to be detected as a location for further analysis. This was set to a range of values (1-7) depending on the signal to noise ratio of the image. Some videos were analyzed in intervals in order to set different thresholds to account for changing contrast levels over the duration of the video. A higher contrast image led to a higher threshold, and the reverse. The localization method was set to "radial to take advantage of the speed and accuracy of the radial sub-pixel localization method described in the reference[8]. The maximum step size and frames allowed between two objects to be connected into an individual trajectory were set to 4 square pixels and 1 frame, respectively.

This initial connection was necessary to determine which localizations were suitable to track the drift between frames. Additionally, based on visual inspection, the drift between frames is less than one pixel, which permits the first, rough estimate of trajectories. Then, a density map S_i was created from object localizations. An individual localization is subject to significant noise, so localization coordinates were averaged over a time window t_{bin} into spatial bins of 160 nm, s_{bin}, to obtain better signal in a density map. Trajectories with residence time less than t_{bin} were excluded so as not to skew the cross-correlation calculation by comparing localizations with no true physical correlation. The cross-correlation C_{ij} between two density maps S_i and S_j was calculated in Matlab and the output image cropped near the origin for further fitting. The cross-correlation image was fit to a 2D Gaussian surface to determine the distance of peak correlation from the origin, $\mathbf{r_{ij}}$. The cross-correlation was determined between density maps Δt frames apart on a sliding window. The overall drift path $\mathbf{D}$ was sampled initially at Δt intervals. An additional sample of the drift path was acquired with time windows one-tenth t_{bin} from the start of the video up to $t_{bin}/2$ in order to capture drift in

9

the early frames, as this time frame is frequently the most subject to drift. The early-frame drift path was concatenated to the overall drift path. The drift path was interpolated to determine the drift within the Δt intervals and subtracted from coordinates on individual frames to bring the coordinates into registration.

The appropriate values for the drift correction parameters to take was determined by quantifying the improvement in eccentricity and spatial localization precision of trajectories with a long residence time ($\tau > t_{bin}$) for a range of parameters. The three parameters varied were t_{bin}, s_{bin}, and Δt. A spatial bin of 160 nm (1 pixel) produced the best improvement over the entire parameter space for the other two parameters. The maximum change in eccentricity and spatial localization precision was acquired for t_{bin} of 50 frames and Δt of 25 frames. This parameter set was successful in removing drift for a range of actual drift rates from a slow as 0.09 nm/s up to 1 nm/s (Fig. 2.3.
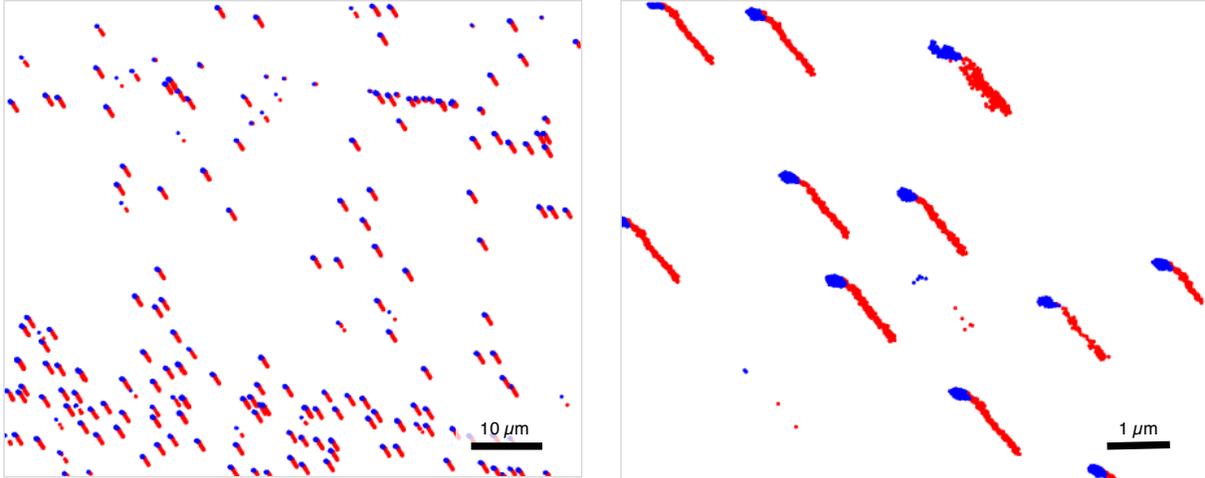
Figure 2.3: Results of drift correction at two levels of zoom. Red points are spot localizations pre-correction; blue points are post correction.

2.2.2   Residence Time Distribution Construction

Protein interaction with the surface leads to highly variable residence times ranging from microseconds to days within a single experiment. This wide range of behaviors cannot be captured in an experiment, therefore an appropriate method for statistical inference must be developed in order to account for the finite length of the experimental time frame. Survival analysis is a set of methods designed to investigate the time until an event of interest occurs, e.g. the time to cancer recurrence in a clinical trial of limited duration, or in the case of protein adsorption, the time to desorption from the surface. There are four types of adsorption events: (1) events with a right censored residence time, in which adsorption is observed, but desorption has not yet been observed by the last frame of the video (Fig. 2.4); (2) fully observed events, whose adsorption and desorption both occur within the duration of the video; (3) events in which adsorption occurs prior to the start of the observation window and desorption occurs within the window; and (4) events in which adsorption and desorption occur within a time interval shorter than a single exposure, that is, left truncated events. Event types (1) and (2) determine the survival function, whose construction will be described below. Type (3) events must be excluded from further analysis because they violate the assumption of noninformative censoring. This means the time at which they are censored, i.e. the time beyond which the protein is unobservable, is dependent on the residence time itself, as shown in Figure 2.5. Type (4) events are unobservable but demand that any survival function estimate be conditional on the residence time being at least the length of a single exposure.

Let there be D unique residence times $t_1 < t_2 < ... < t_D$ and at time t_i there are d_i number of proteins with residence time t_i, i.e. the number of proteins that are observed as a type (2) event. For the j-th protein, let T_{min} be the shortest observable time, which is the time between frames, and let T_j be its residence time, either fully observed or censored. For left-truncated, right-censored data, let Y_i be the number of proteins at risk, i.e. with residence time t_i or greater. This includes both fully observed (type (1)) and censored (type

(2)) events. The quantity d_i/Y_i is the conditional probability that a protein with residence time of at least $t_i - \delta$ will have residence time t_i, where $\delta < T_{min} < (t_i - \delta)$. Thus, the probability that a protein has a residence time of t or greater, given survival to time a, is the survival function $S_a(t) = \Pr(X > t | X \geq a)$, which is approximated by the Kaplan-Meier estimator [9, 10]:

$$\hat{S}_a(t) = \prod_{a \leq t_i \leq t} \left(1 - \frac{d_i}{Y_i}\right), t \geq a. \tag{2.1}$$

	t = 0	Observation window	t = T	Censoring	Event type
				1	1
				-	4
				0	2
				0	2
				1	1
				0	2
				0	2
				-	4
				0	2
				0	2
				1	1
				-	4
				1	1
				-	4
				0	2
				-	3
				-	3
				-	3
				-	3

Figure 2.4: This schematic describes the four types of adsorption data observed, labeled as each type (1)-(4) in the far right column. Right-censored residence times, type (1), labeled "1", start after the video begins at $t = 0$ and extend beyond the duration of the video, $t = T$. Fully observed residence times, type (2), labeled "0" for no censoring, both adsorb and desorb within the duration of the video. Residence times that adsorb before the start of the observation window and desorb within the observation window, type (3), and left-truncated residence times, type (4), are both are unlabeled for censoring, "-" as they are not included in the analysis for reasons described in the text.

2.2.3 Adsorption simulation

The accuracy of the Kaplan-Meier estimator was verified by comparing the survival function estimator from simulated adsorption values to the true survival function. One million residence times were randomly drawn from a continuous exponential distribution with an average of 1000 s and assigned adsorption times uniformly distributed over a time period of 24 hours. An observation time frame of 30 minutes was chosen with a randomly selected start point. In order to convert residence times into frame numbers, the observation window start time was subtracted from each adsorption time, and the result was divided by two seconds and rounded up to the nearest frame number. Then, the total number of frames was determined by the residence time. Survival analysis was then applied to the residence times given by the number of frames that the simulated residence time covered, as would be done experimentally.
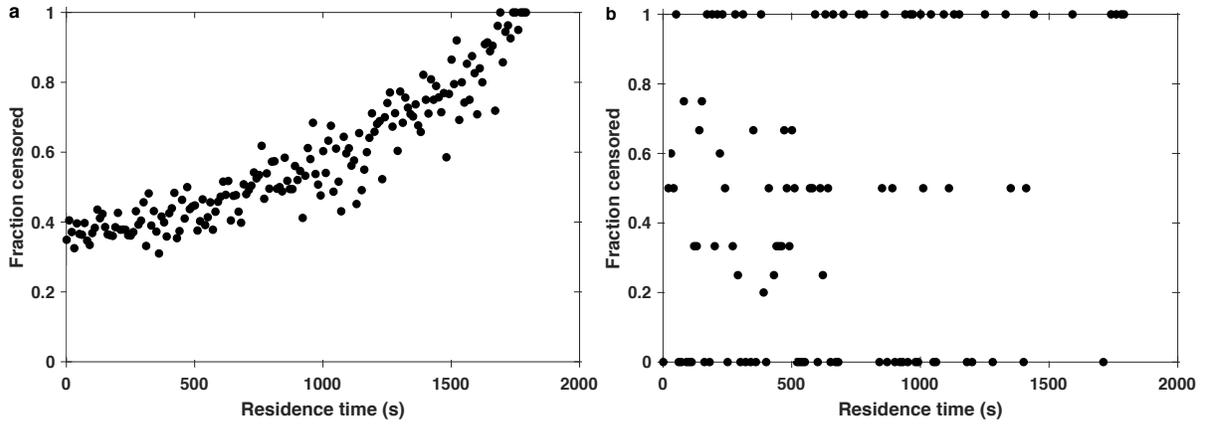
Figure 2.5: The dependence of censoring on the residence time plotted as the fraction censored at a given residence time, versus the residence time. In (a), adsorption events that occur before the start of the observation window and desorb after the start (type (3) events) are included as if they were right censored trajectories. The fraction censored has a linear correlation coefficient of 0.9 with the residence time. In (b), type (3) events are excluded and the correlation coefficient reduces to 0.2. This shows that the assumption of noninformative censoring is violated for type (3) events and therefore must be excluded from the survival analysis.

For the simulation (Fig. 2.6), the Kaplan-Meier estimator (blue line) nearly perfectly overlaps the true distribution (black line). A method used widely in protein adsorption analysis excludes censored data from the analysis. Instead, the frequency of each residence time is weighted with the inverse of the probability to fully observe a residence time of the given length in the observation window. Longer residence times are less likely to be detected in the observation window and therefore given a greater weight than shorter residence times. In comparison to the Kaplan-Meier estimator, this method utilizes only type (2) observations and therefore produces a less accurate estimate of the survival function. We therefore use the Kaplan-Meier estimator to construct the survival functions from our experimental data.
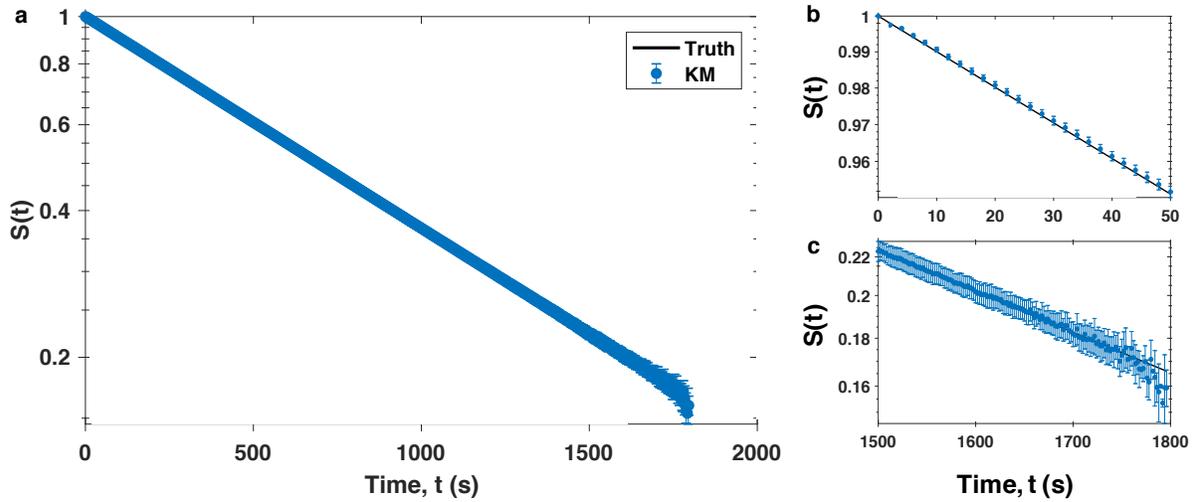
Figure 2.6: The Kaplan-Meier estimate from simulated residence times compared to the true distribution of residence times, (A). One million simulated residence times were drawn from an exponential distribution with average time of 1000 s and assigned adsorption times uniformly over 24 hours. Residence times that fell within the randomly chosen observation window of 1800 seconds were divided into frames as would be the case in a physical experiment. The simulation was run 100 times to get the average value of the Kaplan-Meier estimate at each unique residence time, plotted in blue. The error bars are the standard deviation of the averages. Early (b) and later (c) times in the survival function are expanded to visually demonstrate how closely the Kaplan-Meier method reproduces the true distribution.

Chapter 3

A stochastic model of protein adsorption

3.1 Introduction

The interactions of proteins with surfaces have implications in biocompatibility of a surface [11, 12, 13, 14], protein separation [15, 16, 17, 18, 19], and pharmaceutical nanoparticle development [20]. Protein adsorption effects the ability of both the protein and the surface to fulfill their intended purpose in these capacities [21]. Protein adsorption can be quantitatively studied by a variety of techniques [22, 23, 24, 25, 26, 27, 28] and the data are typically interpreted as having Langmuir-type adsorption kinetics and a desorption kinetics composed of a constant signal from an irreversibly adsorbed protein fraction and Langmuir-type desorption kinetics from a reversibly bound protein fraction [29, 30, 31, 32, 33, 34, 21].

Recent advances in single molecule microscopy have enabled the observation of individual protein-surface interaction events [35, 36, 37]. Protein adsorption studies at the single molecule level have shown that the amount of time proteins of a given type are bound to the surface is broadly distributed and well-approximated by the sum of several exponential functions [38, 18]. The determined desorption rate constants are thought to reflect distinct

surface sites [39], distinct subpopulations of adsorbing proteins [38], or transitions between different binding states [40] (stepwise denaturation). However, an assignment of each rate to a specific type of event is usually not attempted [41, 38].

In the field of economics, Benoit Mandelbrot has sharply criticized the fitting of various data with sums of exponentials on the grounds that the data can often be concisely fit with power laws to which the sum of exponentials often provide only awkward approximations relying on many additional parameters[42]. However, it is often practical from a computational perspective to approximate a power law distribution with a sum of exponentials, and an approach to find the optimal number of exponentials is defined in the literature[43].

This raises the question if the same arguments apply to the interpretation of single molecule data. The "sum of exponentials vs. power laws debate however is not simply about the search for the model which best satisfies an information criterion [44, 45], but about which mechanistic interpretation is appropriate. A number of physical mechanisms giving rise to power laws have been reviewed by Newman [46]. An application is the work by Cichos et al. who studied the distribution of blinking times of nanocrystals and attributed the observed power law behavior to the interplay between exponential dependencies of trapping and recovery rates of dark states [47].

Here, we propose that the observed protein desorption kinetics originates from the sequential and reversible establishment of protein-surface contacts, where each contact adds a small contribution to the overall binding energy. After making the initial contact, the protein performs a slightly biased random walk over a large number of small energy barriers until it makes a maximum number of contacts. This mechanism corresponds to the statistical zippering observed by Penna et al. in molecular simulations of peptide adsorption[48]. In contrast to the desorption kinetics arising from the passage over a single, dominant energy barrier (exponentially distributed residence times), the one-dimensional random walk along the reaction coordinate results in a power law with exponent 1/2 for the residence time distribution for an infinite number of barriers.

20

In this chapter, we first describe the theoretical model and how it is fit to experimental data, and then present measurements of the residence time distributions of fluorescently labeled bovine serum albumin (BSA) and fibrinogen (Fg) on glass surfaces. The experimental data are comparable to previously published data of BSA and fibrinogen adsorption to fused silica and other surfaces by the Schwartz and Landes groups[49, 50, 51, 38, 52, 18]. The survival analysis method established in the previous chapter is utilized to capture the entire range of adsorption kinetics. We interpret the data using the statistical zippering model and discuss the implications.

3.2   Methods

3.2.1   Single Molecule Imaging

Bovine serum albumin conjugated with AlexaFluor 488 dye and fibrinogen (Fg) conjugated with AlexaFluor 488 dye dye were both purchased from Thermo Fisher Scientific and (A13100 and F13191, respectively). These were dissolved from powder in 10 mM phosphate buffered serum (PBS) at 1 mg/mL, stored at 4°C and used without further purification. Flow cells were constructed from 24×60 mm and 22×22 mm glass coverslips from VWR of #1 thickness. Coverslips were cleaned by sonicating in ultrapure ($\geq$ 18 MΩ-cm) water, acetone, the 1N potassium hydroxide for 20 minutes each and rinsed before each step with ultrapure water. After drying completely in a 40°C oven, coverslips were plasma etched in the UV Ozone Cleaner (Procleaner™Plus BioForce Nanosciences, Inc.) before final 20-minute sonication in ultrapure water. Coverslips were then dried again at 40°C and stored in sealed glass jars for less than 24 hours before use. Flow cells were assembled as described in 2.1.1. For adsorption experiments, Fg solution was prepared fresh for each experiment at 45 pM, 450 pM and 6 pM for glass, Pluronic®F-108 (Sigma-Aldrich) and silane experiments, respectively, in 10 mM PBS at pH 7.4. BSA solutions were prepared similarly, at 450 pM, 450 pM, and 9 pM. These concentrations were chosen to allow for a high enough object density to acquire

enough data in a reasonable experimental time, but low enough density to facilitate analysis of spatially distinct spots. An oxygen-scavenging system was added to reduce the effects of photobleaching on the experiment. Glucose oxidase (Sigma-Aldrich G2133), D-glucose (Sigma-Aldrich G8270), catalase (Sigma-Aldrich C40), and DL-dithiothreitol (Sigma-Aldrich D0632) were added in excess to final concentrations of 0.2 mg/mL, 200 mM, 0.08 mg/mL, and 100 mM. Images were acquired with a Nikon Ti-Eclipse Inverted microscope (Morrell Instrument Company) with TIRF module using a 100× 1.45 NA oil objective and a back-illuminated EMCCD camera (Andor iXon DU-567). Samples were illuminated during the exposure time only with a 488 nm LuxX diode laser (Omicron Laserprodukte GmbH) at a low laser intensity of 8% of the maximum laser intensity of 150 mW to reduce the effects of photobleaching. Images were acquired in 200 ms exposures with 1 s between the start of each frame for 60 minutes with 150× gain amplification. The time between exposures minimizes the appearance of photoblinking as it allows the fluorophore to recover out of the dark, triplet state.

3.2.2   Single Molecule Identification and Tracking

Videos of protein solutions were analyzed using the Matlab package developed by R. Parthasarathy[8]. This package detects proteins as higher-intensity spots in the field of view, and connects these spots across frames to create trajectories of position and brightness of the fluorescently-labeled protein in time. The package has several user-defined parameters. The bandpass filter, set to 7 pixels, sets the maximum radius for an object so that large variations in intensity, i.e. due to uneven illumination, may be eliminated. The standard deviation intensity threshold determines the number of standard deviations above the mean intensity a spot must be in order to be detected as a location for further analysis. This was set to a range of values (1-7) depending on the signal to noise ratio of the image. Some videos were analyzed in intervals in order to set different thresholds to account for changing contrast levels over the duration of the video. A higher contrast image led to a higher threshold,

and the reverse. The localization method was set to radial to take advantage of the speed and accuracy of the radial sub-pixel localization method described in Parathasarathy 2012. The maximum step size and frames allowed between two objects to be connected into an individual trajectory were set to 4 square pixels and 1 frame, respectively. The resulting trajectories were further processed to remove the effect of stage drift. Stage drift over the course of a video causes objects on the surface to appear to move along the path that the stage moves. A cross-correlation algorithm was implemented to calculate the drift path and remove it from the path of the adsorbed proteins in order to accurately determine trajectories, as described in Chapter 2, Section 2.2.1. The connected trajectories are then aggregated into survival functions as described in Chapter 2, section 2.2.2.

3.2.3 Photobleaching Assessment

In the experimental data, photobleaching can remove a protein from observation. Elsewhere [53], the solution to this problem is to reduce the excitation laser intensity to a level where the rate of photobleaching is far below the rate of the estimated adsorption kinetics. It is then assumed that one may neglect the effect of photobleaching on the observed adsorption kinetics.

We examined the effect of photobleaching on the adsorption survival function by performing experiments at varying laser intensities (Fig. 3.1b, faded colors). As in bulk experiments of protein adsorption, one can interpret the fast dynamics within the first minute as originating from reversible protein adsorption. The slow dynamics originate from an irreversibly adsorbed protein fraction. These permanently bound proteins are removed from observation by photobleaching. We isolated the rate of removal due to photobleaching by covalently linking the protein to the surface and exciting the surface-bound proteins at the same laser intensities as in the desorption experiments. The resulting survival functions shown in Fig. 3.1(a) are dependent only on the bleaching of fluorophores. An exponential fit to these curves provides the intensity-dependent photobleaching rate.

Overall, the change in the survival function at each unique residence time, t_i, is a function of the rate of removal from observation due to photobleaching and due to desorption:

$$\frac{S_{t_{i+1}} - S_{t_i}}{\Delta t} = -k_{PB}S_{t_i} - k_{off}S_{t_i}, \tag{3.1}$$

where k_{PB} is the apparent rate of removal from observation due to photobleaching and k_{off} is the rate of removal from observation due to desorption. The desorption rate, k_{off}, is then determined from Eq. 3.1 with the known photobleaching rate and observed survival function (Fig. 3.1b). The measured survival functions are corrected for photobleaching by reconstructing them with k_{off} and $S_{t_1} = 1$:

$$S_{t_{i+1}} = S_{t_i} + k_{off}\Delta t S_{t_i}. \tag{3.2}$$

The arrows in Fig. 3.1(b) show the decrease in the slope between the original and photobleaching-corrected curves at different laser intensities. Between 3% and 10%, the change in slope from the original to the corrected curve is small, indicating that photobleaching has little effect on the survival curve and protein adsorption kinetics dominates the results at low intensities. However, at increasing laser intensity - 20% and 50% - the slope of the survival function changes significantly after correction. Therefore, without correction, a higher laser intensity increases the rate of removal from observation due to photobleaching enough to cause photobleaching to be the dominant factor in the observed kinetics. Isolating the effect of photobleaching allowed us to correct for the effect of photokinetics at any laser intensity. Based on these results, a laser intensity of 5% was chosen for further experiments, as the contribution of photobleaching is negligible.
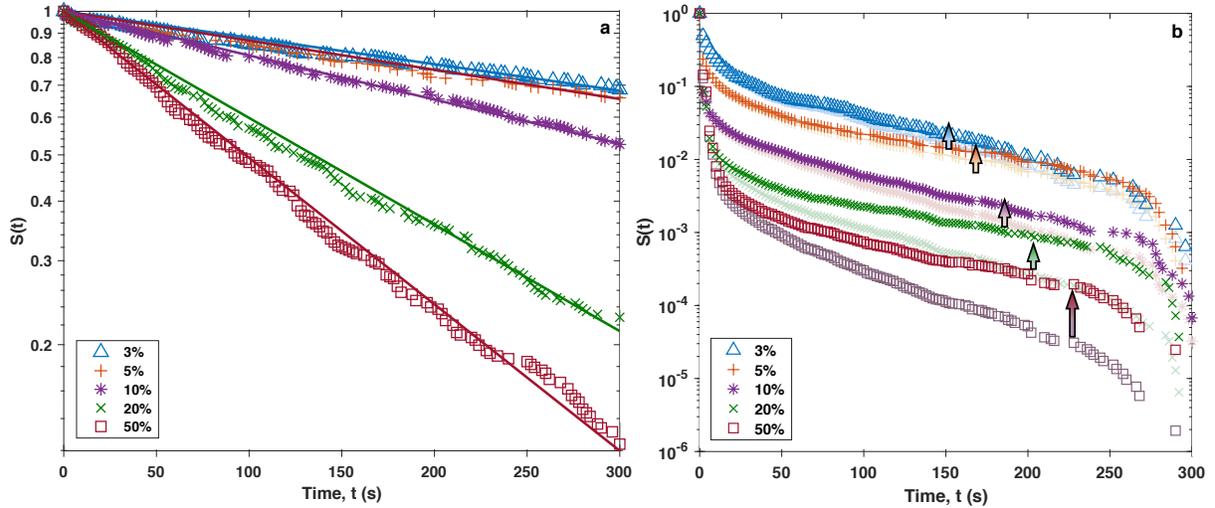
Figure 3.1: The survival functions of 400 pM Alexa 647-labeled fibrinogen covalently linked to glass at 3%, 5%, 10%, 20% and 50% of maximum laser intensity, in (a). Lines are single exponential fits to the curve via a maximum likelihood estimate. The survival functions for 40 pM Alexa 647-labeled fibrinogen on glass at 3%, 5%, 10%, 20% and 50% of maximum laser intensity in (b). The faded color indicates the original survival function and full color indicates the photobleaching-corrected survival function. The decrease in slope after correction demonstrates that photobleaching generated an apparent decrease in the average residence time. The effect is stronger at higher laser intensities.

3.3 Theory

3.3.1 Power-law model

A model was developed by Prof. Parag Katia, San Diego State University to describe the data based on the suggestion from the Mandlebrot that the behavior underlying the sum of exponentials is a power law. A stochastic model to simulate the adsorption and desorption processes. The simulation models adsorption/desorption as a 1-D random walk along a zipping/unzipping process, where each step is a conformational change with a finite probability of moving forward or back. This assumes that 1) conformational changes happen in a sequential manner whose process must be followed in reverse for desorption, and 2) there is a maximum number of conformational states, at which point the protein is irreversibly denatured.

The protein adsorption process was simulated as follows:

1. The simulation initiates with protein attached to the surface with one bond.

2. The time step, t is obtained from an exponential distribution whose average is the timescale of transitions between protein states determined from a Gillespe algorithm.

3. A bond forms at the next timestep with probability $1 - p$ or breaks with probability p.

4. (2) and (3) repeat until the protein leaves the surface (0 bonds) or reaches the maximum number of bonds that can form with the surface (*sites*).

The residence time is the total amount of time spent on the surface. Residence times were generated for 10^6 proteins to obtain the residence time distribution. The parameters fit to the experimental data were t, p, and *sites* with a Monte-Carlo Simulated Annealing search. Additionally, due to the requirement that proteins remain on the surface for at least two frames (i.e. 2 steps, or 2 seconds) to distinguish them from noise, the simulation results were normalized with the same condition.

3.4 Results

3.4.1 Adsorption experiments

TIRF experiments to observe protein adsorption were carried out for fibrinogen and bovine serum albumin. Representative images of 1-hour videos of adsorption experiments are shown for 45 pM Fg on cover glass in Fig. 3.2(a) and 450 pM BSA on cover glass in Fig. 3.2(b). For Fg and BSA, respectively, there were over 780,000 and 160,000 objects detected in the video with an average signal-to-noise ratio of 2 ± 1 and 2.9 ± 0.5 counts. The average object density was 218 and 46 spots per frame, or 0.03 μm^{-2} and 0.007 μm^{-2}. This low density facilitated accurate detection and characterization of the surface adsorbed proteins because there was sufficient space to prevent object spots from overlapping and appearing as one large object. Over 350,000 and 75,000 adsorption events were detected over the course of the video, around 38,000 and 13,000 of these were analyzed further for Fg and BSA, respectively, after eliminating events. Adsorption events were eliminated on the basis of incomplete information trajectories beginning on the first frame have an unknown start time and trajectories lasting less than 2 frames are too easily confounded with noise. There was an average of 10 and 4 adsorption events per frame that were kept for further analysis, giving landing rate of 10 s^{-1} and 4 s^{-1} for Fg and BSA, respectively.

Typically in protein adsorption experiments, the complement cumulative distribution function of the surface lifetimes are plotted on a log-linear plot as shown with the results from these experiments in Fig. 3.8. The shape of these curves are characteristic of protein adsorption experiments performed by other groups [49, 16]. On a log-linear plot, a single exponential would be appear to be linear curve. While the tail of these distributions appear linear at the very end, the rest of the curve does not appear so. This behavior, plus the assumption of exponential rate process has motivated many groups to fit these lifetime distributions to a sum of exponential distributions. However, considering the work by Mandlebrot, the lifetime distributions were plotted on log-log plots where a power law would

27

appear linear (Fig. 3.9. On the log-log plots, the faster kinetics, i.e. the shorter residence times, appear linear and the tail curves down - the exponential tail seen in the log-linear plots.
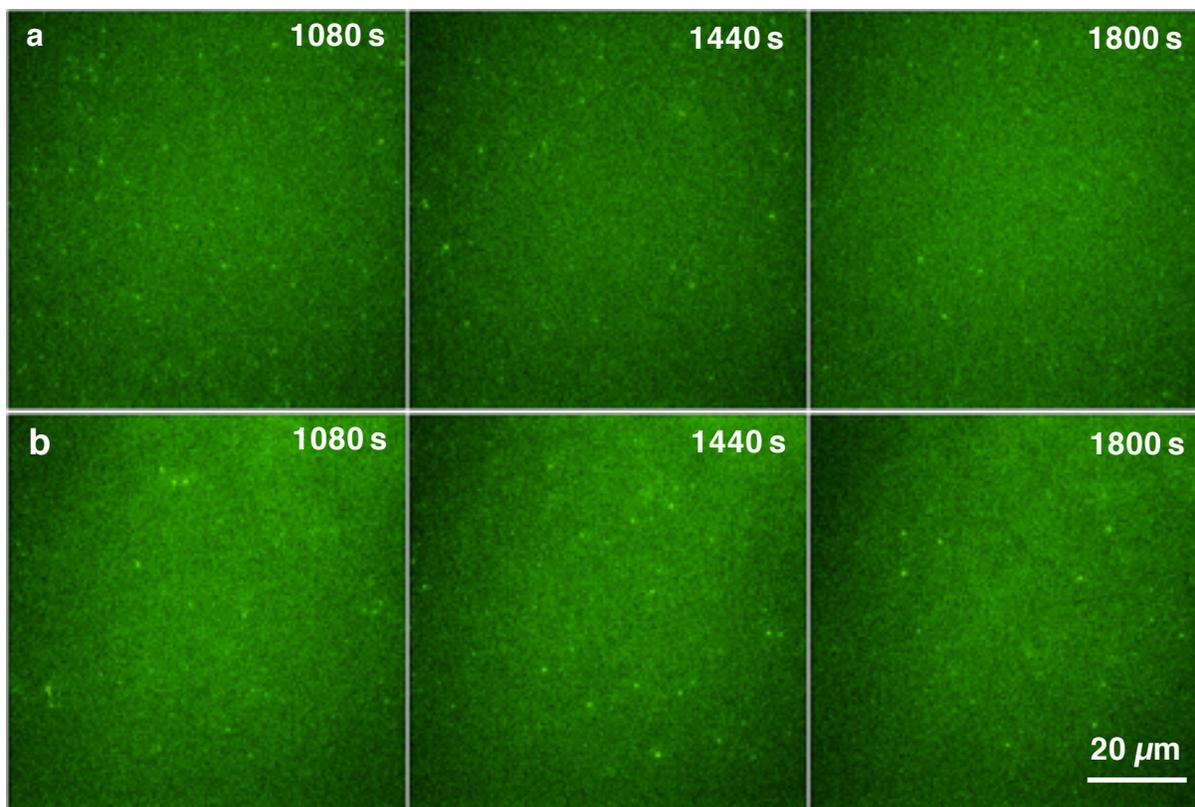
Figure 3.2: Protein adsorption experiments generate complex survival functions not well-described by a single exponential. In (a), representative images of 45 pM Fg-AlexaFluor488 conjugate landing on cover glass, and 450 pM BSA-AlexaFluor488 in (b). These images are from exposures at 18, 24, and 30 minutes into a 60-minute video of adsorption. Each exposure lasted for 200 ms and was followed by an 800 ms dark interval to prevent deleterious photoeffects of photoblinking and photobleaching.
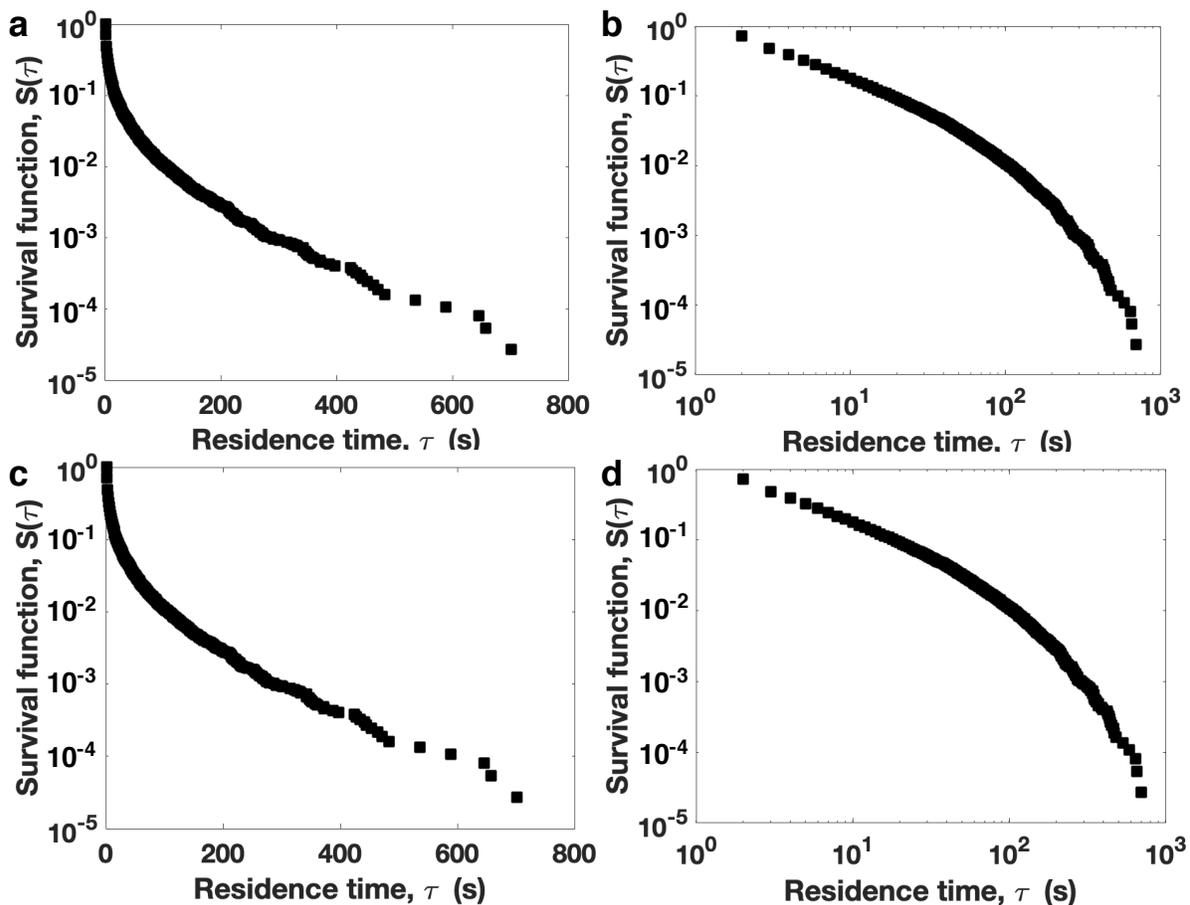
Figure 3.3: The survival functions characterizing the protein adsorption was plotted on the commonly used semilog plot in (a) for 45 pM Fg and (c) for 450 pM BSA both on cover glass. The survival functions were plotted again in (b) and (d) on a logarithmic plot to demonstrate that the early portion of the curve looks like a straight line as a power law would on this plot.

The surface lifetime distributions, i.e. the survival functions, for these experiments are shown in Fig. 3.3. The shape of the distribution in Fig. 3.3(a), (c) is similar to what appears in the foundational work by Kastantin and Schwartz in 2011[38] where the non-exponential nature of the protein adsorption function was first observed. The fundamental model of protein adsorption predicts that the survival function would appear as a line on a log-linear plot and would be fit by one exponential process with a characteristic residence time of the proteins on the surface. Instead, a sum of multiple exponential processes was required to fit the survival function. Kastantin proposed that this arose from multiple aggregation states in the bulk fibrinogen population that uniquely interact with the surface. One test for the presence of multiple aggregation states is in the distribution of intensities that each adsorption event produces. If there are multiple aggregation states leading to different characteristic residence times, then there would be groups of adsorption events with characteristic intensities, corresponding to a single exponential process.

To uncover the presence of these characteristic intensity groups, the same process was followed as in Kastantins paper: the probability intensity distribution was determined for increasing minimum residence time and plotted as a function of that residence time as in Fig. 3.7(a), (c). These surface plots show an initial peak for average intensity around 100 counts above the background, which corresponds to the monomer group, and additional peaks or a broadening of peaks as the residence time cutoff increases, which could correspond to higher aggregation states. The comparison between distributions for low and high residence time is demonstrated Fig. 3.7(b), (d), where it is clear that the monomer population, i.e. the smallest intensity peak, gives way to other aggregation states as most frequent for longer residence times. The presence of multiple peaks for both fibrinogen and BSA suggests that there are multiple aggregation states in the protein solutions and that these correspond to certain residence times. However, it is necessary to determine if these aggregation states cause the distinct exponential processes summed to describe the survival function of the entire population. The characteristic intensity for monomers, dimers, trimers

31

and greater were estimated as one-, two-, three- and four-times the intensity of the initial peak in Fig. 3.7(a), (c). Trajectories were segregated based on their average intensity into these groups using the midpoint between each multiple of the initial peak. This corresponds to intensity counts of ¡ 165, 165 275, 275 385, and ¿ 385. Survival functions were then calculated and fit for each group to determine if each group had a single, unique characteristic lifetime that contributed to the sum of exponential components required to fit the survival function calculated for all trajectories. If the estimate of intensities for each group perfectly matched reality and the aggregation state was the cause for multiple exponential components, then each the survival function for each population would appear as a straight line on a log-linear plot and have one characteristic residence time. However, considering the that the segregation of groups was estimated, the presence of the same characteristic residence time in consecutive intensity groups would not negate the hypothesis given the potential for overlap in the groups.

The data were fit in two stages: first by determining the number of exponential processes necessary with a maximum entropy method[54], then by determining the characteristic lifetime of those exponential processes with non-linear least squares fitting. In Fig. 3.4(a), (b), the survival functions for the intensity-segregated data are plotted on a semilog plot, where it is clear that no one sub-population could be well-described by a single exponential. This is additionally demonstrated, and the effect of group overlap accounted for, in Fig. 3.5, where the characteristic residence time for each group is represented in a bar graph. Multiple exponential processes were required to fit each sub-population, meaning that aggregation state does not give rise to the discrete exponential processes used to fit the whole population of trajectories. This confirms the need for another model to explain the complexity of the protein adsorption survival function and implies that the power law model is a valid explanation.

The different aggregation states do not lead to pure exponential survival functions, but neither do they produce pure power laws. These would appear as a straight line on a log-log

32

plot, but instead curve off in the tail like an exponential. The complexity of the curve may be due not to heterogeneity of the adsorbing proteins, but instead to the heterogeneity of the surface to which they adsorb, causing a mix of power laws. The heterogeneity of the surface can be tested using the location of adsorption events because a homogenous surface would give rise to a Poisson distribution of event frequencies. To test this, one sub-population of the proteins was taken (monomers) and their landing locations studied. The location of each landing event from this subpopulation was recorded, and the number of events at each location tallied. The histogram of the number of events at any location was compared to a theoretical histogram of the number of events at any location in Fig. 3.6(a), (c). In blue, the theoretical distribution perfectly fits a Poisson distribution. In black, the experimental distribution quickly deviates from the shape of the theoretical curve. This indicates that many locations on the surface had far more events occur than could be explained by a random spatial distribution of landing on a homogeneous surface.
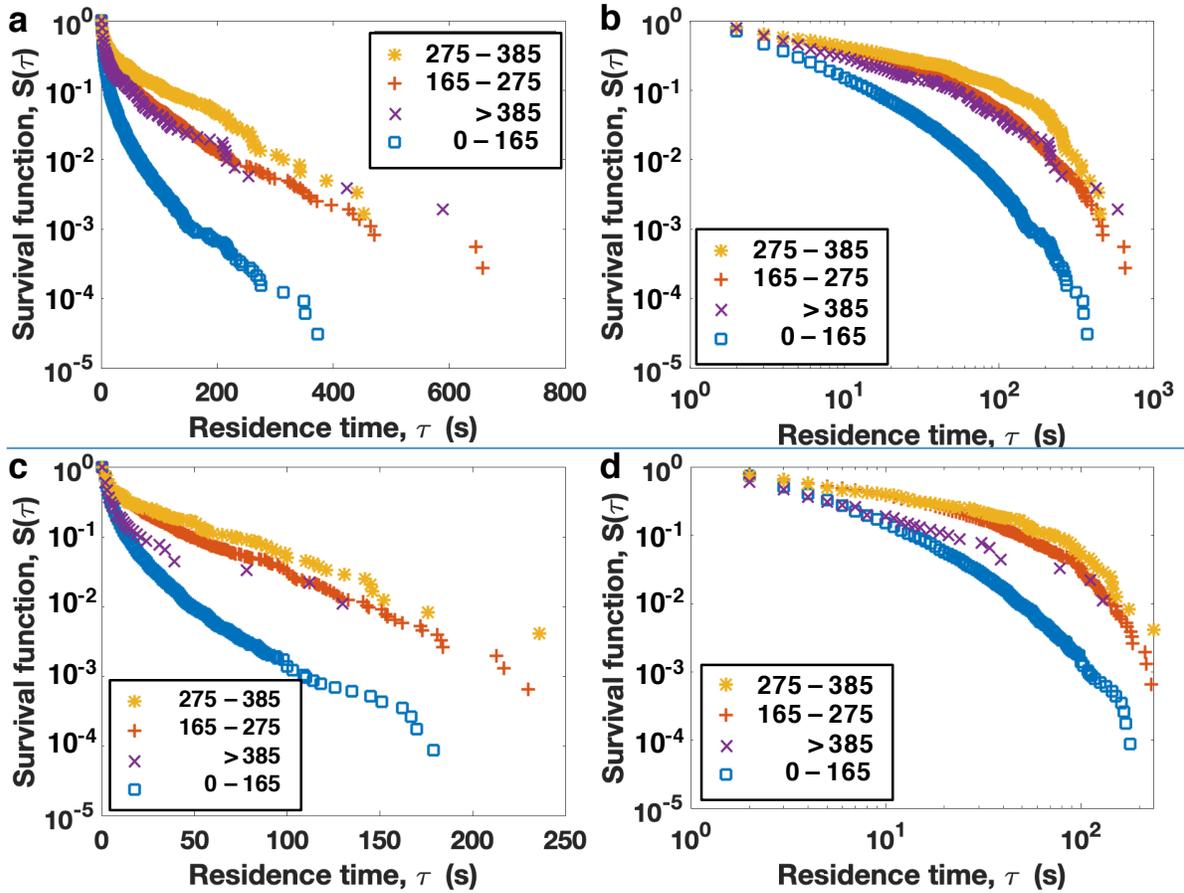
Figure 3.4: The survival functions for groups of trajectories segmented based on their average intensities were plotted to determine if each could be assigned a single exponential. The legend lists ranges of counts used to segment the trajectories that were based on multiples of the first peaks. The segmented survival functions for 45 pM Fg are in (a) as a semilog plot and (b) as a logarithmic plot, and for 450 pM BSA in (c), semilog, and (d), logarithmic. In (a) and (c), none of the curves present the straight line expected of a single exponential process on a semilog plot. In (b) and (d), a significant portion of the curves appear to be straight for groups with over 165 counts.
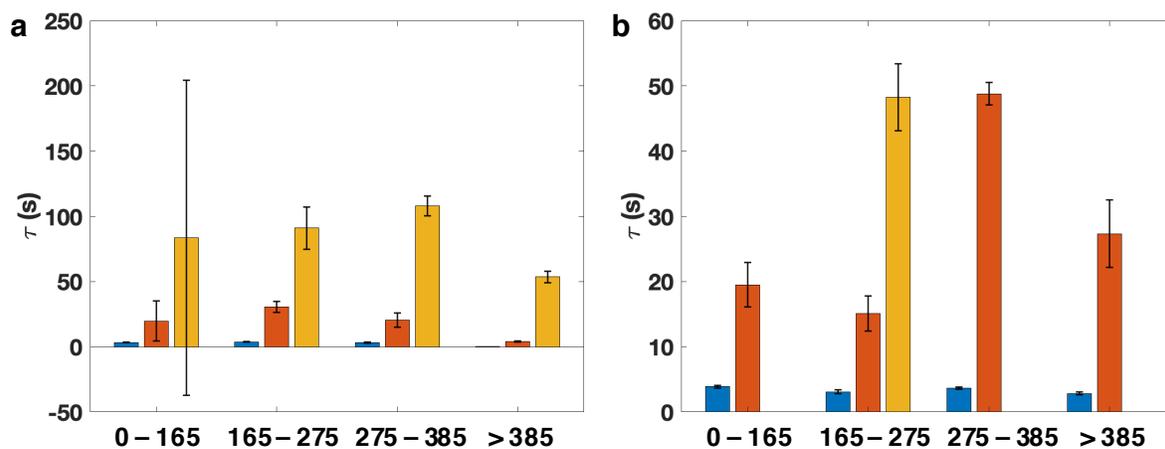
Figure 3.5: The survival functions for sub-populations based on intensity were calculated and fit as described in the text. The results of the fit for 45 pM Fg sub-populations are in (a) and 450 pM BSA are in (b). The range of counts used to define each sub-population are on the x-axis: 0-165 counts for monomers, 165-275 for dimers, 275-385 for trimers and over 385 for tetrameric or greater aggregates. Each bar represents a characteristic lifetime of one exponential process. The presence of multiple bars for a single aggregation state represents the multiple exponential processes summed to fit the survival function of one estimated aggregation state. Error bars are the 95% confidence intervals of the fit parameters.
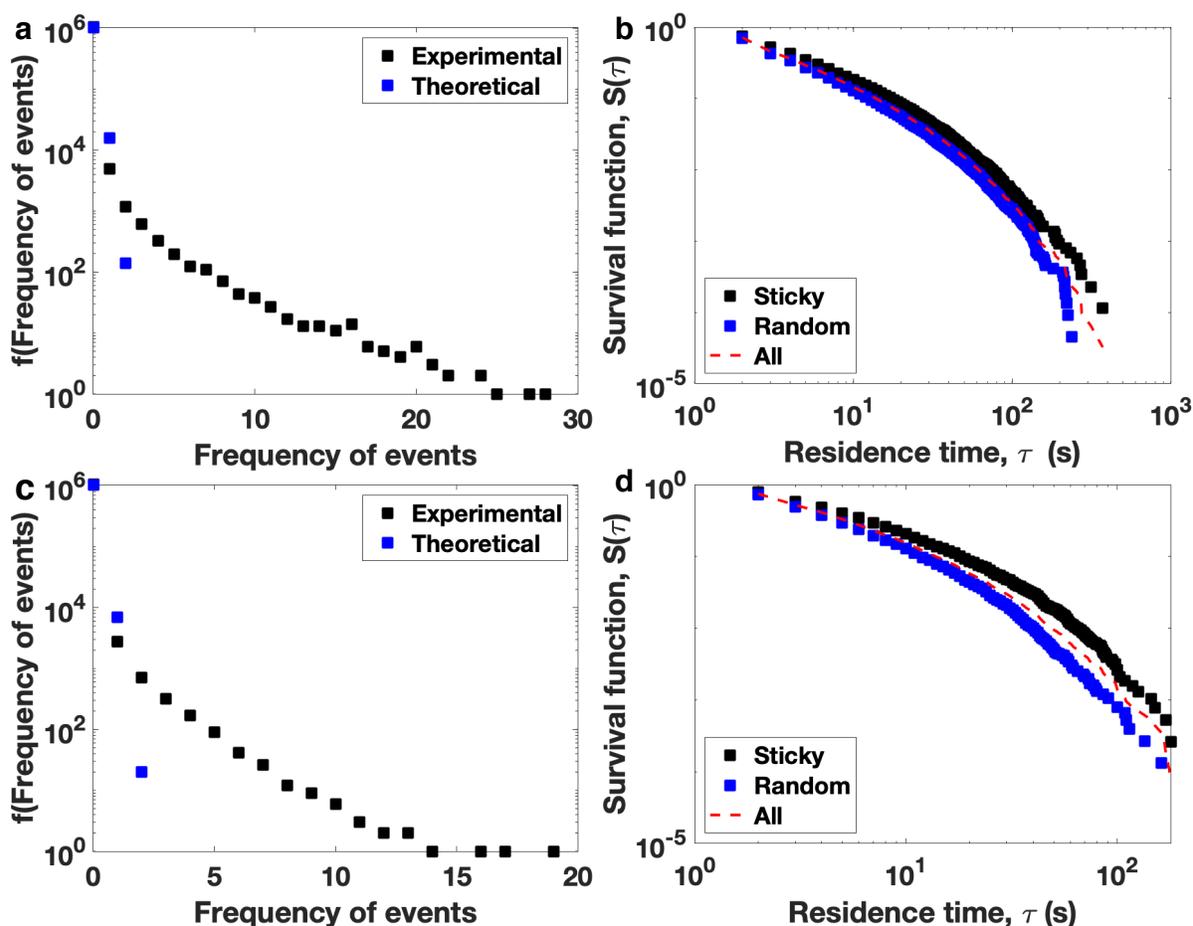
Figure 3.6: The frequency with which surface locations are visited indicates a heterogeneous surface but this does not explain the complex survival curve. A histogram of the number of events that occurred at each location was compared to a theoretical histogram of the same, (a) and (c), for the monomer population of 45 pM Fg and 450 pM BSA, respectively, on cover glass. The long tail of the experimental distribution shows that there are many sites ("sticky sites) that are visited far more than could be explained by a random spatial distribution of landing events ("random sites). The sticky and random sites were segmented, and their survival functions calculated (b) and (d). Since the sticky and random groups are not evenly split along the factors that determine the normalization constant, at any individual point, they may have a greater fractional value than the survival function for the complete data set due to the difference in normalization factor at that point, even though the number of trajectories in the two subpopulations equals the number of trajectories in the entire dataset at a given point.
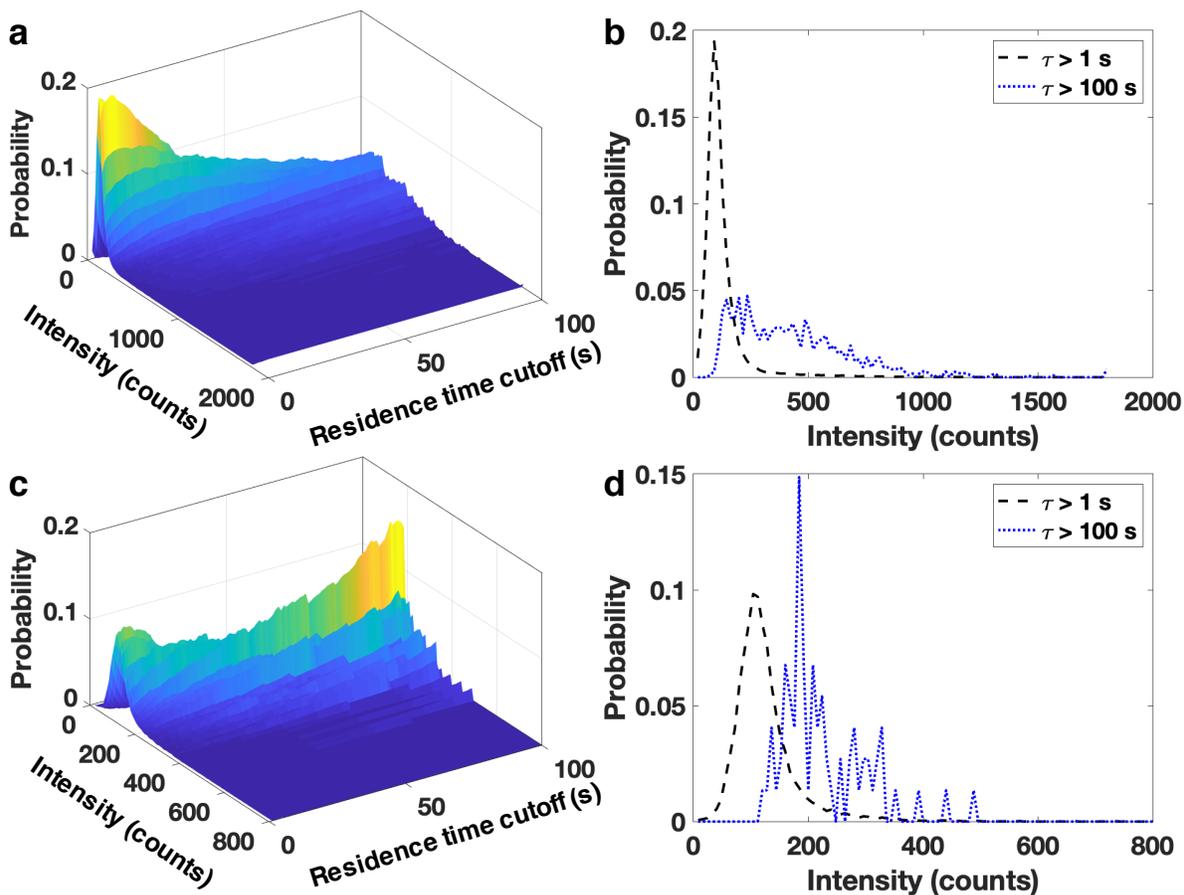
Figure 3.7: In (a) and (c), the probability distributions of the average intensity of trajectories plotted as a function of the minimum residence time for a trajectory included in the probability distribution for 45 pM Fg and 450 pM BSA, respectively. Distinct ridges corresponding to residence time cutoffs exist for BSA, and a broadening of the initial peak appears for Fg. In (c) and (d), the probability distributions for residence time cutoffs (τ) 1 s (black dashes) and 100 s (blue dots) explicitly shows the difference between the dominant intensities of trajectories with these residence times.
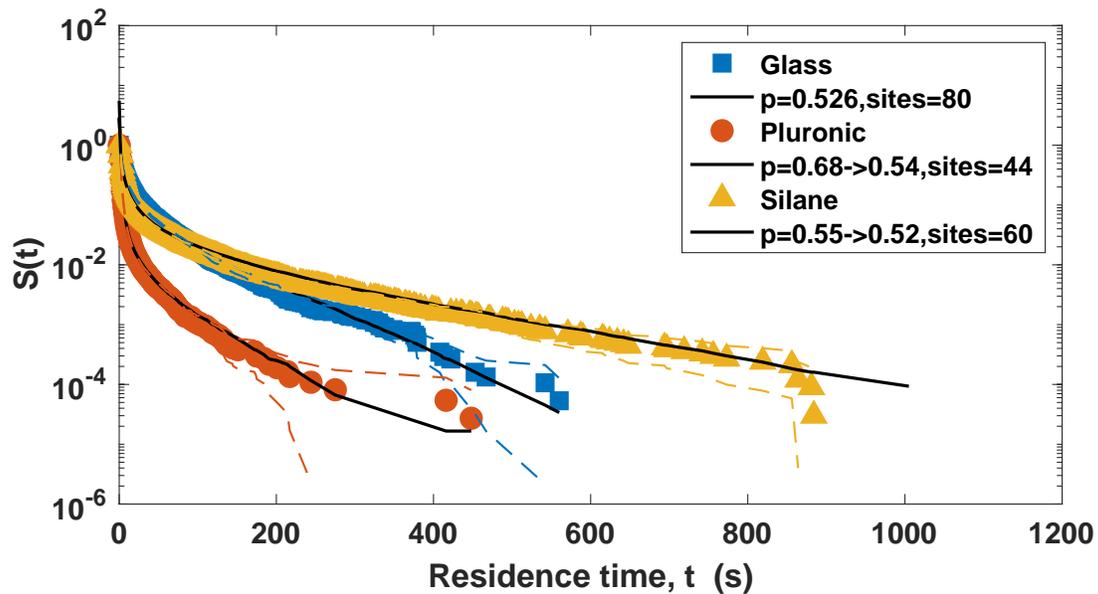
Figure 3.8: The survival functions for BSA on clean glass, Pluronic-coated, and silane-coated surfaces. The upper and lower confidence intervals for each point were calculated according to Greenwood's formula and plotted in the dashed lines. The simulated fits are in black with the results of the fit parameters for p and *sites* in the legend.
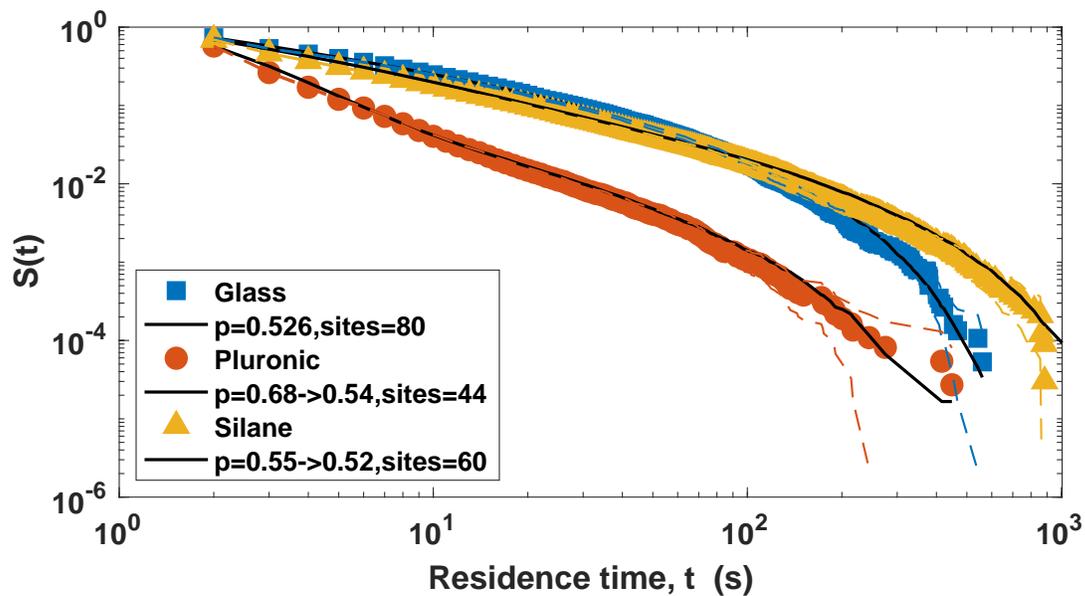
Figure 3.9: The survival functions for BSA on clean glass, Pluronic-coated, and silane-coated surfaces. The upper and lower confidence intervals for each point were calculated according to Greenwood's formula and plotted in the dashed lines. The simulated fits are in black with the results of the fit parameters for p and *sites* in the legend.
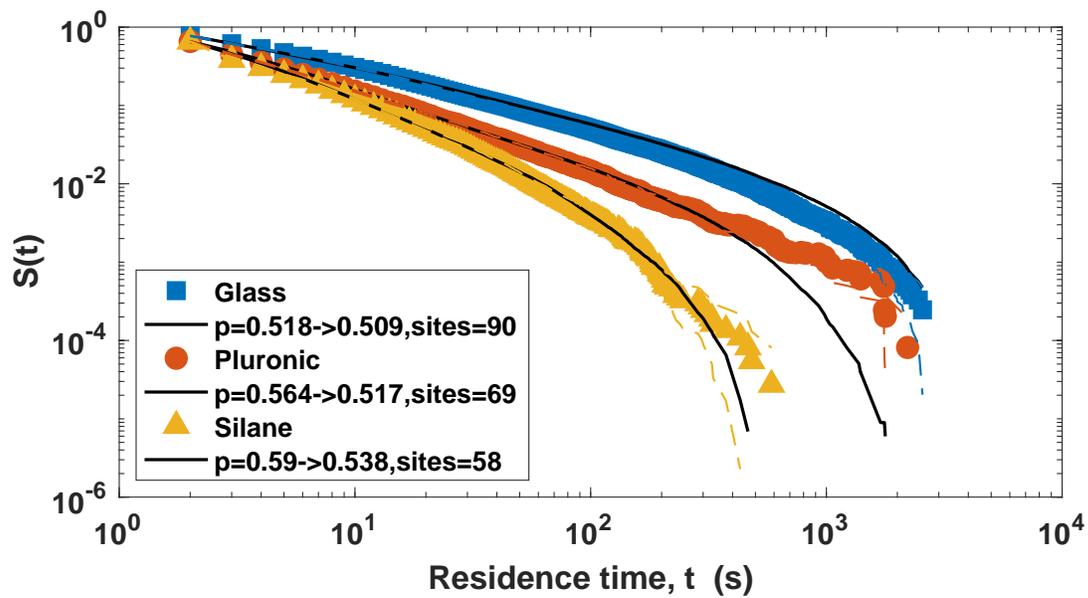
Figure 3.10: The survival functions for fibrinogen on clean glass, Pluronic-coated, and silane-coated surfaces. The upper and lower confidence intervals for each point were calculated according to Greenwood's formula and plotted in the dashed lines. The simulated fits are in black with the results of the fit parameters for p and *sites* in the legend.

3.4.2   Power law-based model

Protein adsorption follows a simple 1-D model for a search in energy space for the deepest bound state by forming or breaking one bond at a time. The results of the simulation to fit experimental data based on this principle are shown in the black curves overlaying survival functions in Fig. 3.9 and Fig. 3.10. The timestep parameter, t was determined to be 0.3 in the first few fits, then held constant for the remainder of fits to reduce the size of the parameter space and the time to simulate. The simulation nicely fit the data for the BSA on glass without any modification of the proposed model.

However, for fibrinogen - a larger globular protein with less internal coherence - and for modified surfaces with multiple layers or complex protein interactions this is not the case. The 1-D zipping/unzipping model does not adequately capture the dynamics when constrained to a single mean likelihood of bond formation, $1 - p$. Instead, a two-stage process is introduced: the protein initially performs the bond forming/breaking walk with a given value of $1 - p$, then at a critical threshold number of bonds formed, continues the walk with a higher likelihood of bond formation, i.e. a deeper state of binding. This modification adds two parameters: n, the critical threshold as a percent of the maximum number of bonds, and $1 - p'$, the new probability of bond formation after the critical point. It fits all of the data well and leads to results that physically make sense. For example, BSA binding to Pluronic initially forms a new bond at each step with probability 0.32. After 10% of bonds are formed, the probability of bond formation is 0.46. Physically, this is intuitive, as the protein binding to a Pluronic-coated surface first must find a site that is not inhibited by the anti-adsorption polymer. After finding that site, or a conformation of the protein that allows it to get between the polymer chains, the probability of bond formation increases, as the interactions with the surface are more favorable. While the numbers are different for each surface coating and protein, the transition from a lower to a higher probability of forming the next bond is a physically intuitive explanation of the protein adsorption process, as it has been demonstrated for many proteins that the conformation must change to enter the

irreversibly bound state.

Interpretation of adsorption results

The maximum number of sites allowed for binding is expected to be higher for stronger interactions, and lower for weaker interactions. Therefore it is expected to be the highest for silane, to which proteins can bond through the hydrophobic interaction. However for BSA and Fg, the highest maximum number of sites is 80 and 90, respectively, for adsorbing to glass. The reason for this is that glass, even with the extensive cleaning procedure, is an imperfect surface. A more pure substrate choice for these experiments would have been fused silica or even quartz. Therefore it is difficult to accurately predict the kind of interaction expected between the protein and the glass. Given the fact that there is no need for BSA to enter the post-critical phase of the adsorption process to describe the experimental data (for all three trials), there is likely some impurity that is very favorable to BSA binding. Fibrinogen however, needs to pass a critical threshold of binding steps first.

Pluronic and silane surfaces both require moving into the second phase of the adsorption process to adequately capture the adsorption process for both proteins. Without the second phase, the longer lifetimes are never captured by the model, suggesting that the there is a certain conformational state that allows the protein access to the more strongly bound state. For silane, that conformation is reached after $n = 12\%$ and $n = 10\%$ of the maximum number of bonds for BSA and Fg, and for Pluronic it is after $n = 10\%$ of maximum number of bonds for both proteins. As the Pluronic is adsorbed to the surface via the hyrdophobic interaction with silane, it is expected that a protein would have a similar or slightly lower (due to steric hinderance) probability of forming a new bond in the second phase a protein adsorbing to silane. Fibrinogen does not fit this expectation for reasons that are unclear. However, the model produces that expected result for BSA, with $1 - p' = 0.46$ for Pluronic and $1 - p' = 0.48$ for silane (Fig. 3.9). The difference in the maximum number of bonds is likely due to the reduced freedom of motion that the protein has in spots between the

Pluronic polymer brushes.

3.5 Discussion

Protein adsorption is a complex phenomenon that continues to drive researchers even after decades of study. The reason for this is that it is a universally important process determining many things including health outcomes. Additionally, there is no universal model of protein adsorption explaining the underlying mechanism of the process. Here, a model has been introduced to provide a novel interpretation of the adsorption of proteins to surfaces with mechanistic insight. The model has five parameters with physically-linked meanings, as opposed to having been indiscriminately added until the ideal fit was achieved. This model was validated with experimental data under six unique conditions: two proteins on three different types of surfaces.

However, it is difficult to experimentally validate the mechanistic interpretations of the results. There is so much heterogeneity that is picked up due to the individual nature of single molecule studies that it becomes difficult to isolate how different factors contribute to the adsorption process. Therefore in future work, it would be important to develop a system with as few sources of experimental noise as possible to better test and confirm the mechanistic interpretation introduced by the model.

Chapter 4

Point accumulation in nanoscale topography of protein surfaces

4.1 Introduction

The conformation of a protein in any given context is the driving factor in its function. The "wrong conformation can lead to dangerous aggregation [55], and set off a chain of immune responses in the blood [56, 57]. The environment surrounding the protein has significant role in determining the actual conformation in a given context [58, 59, 60]. For example, a hydrophobic surface will cause at least partial denaturation in most proteins as they change conformation to adsorb due to the hydrophobic effect [61, 62, 63, 64], or a highly ionic solution will cause a much looser internal coherence [65, 24]. These example environments allow more facile conformational changes, thus affecting, for example, the activity of enzymes [66] and the immune identifiability of proteins [67, 68].

Angstrom-level conformational detail that arises from bulk studies and crystal structures cannot be applied to the flat interface. Optical methods such as single-molecule Frster resonance energy transfer (smFRET) [40] and total internal reflection fluorescence (TIRF) have begun to fill that gap, confirming the conformational changes predicted in bulk experiments

[69, 70]. These conformational changes would be expected to effect the surface of the protein facing solution, but the effect is yet to be qualified and quantified. This is a highly significant interface as it is what interacts with any other species in solution such as coagulating factors in blood or cell receptors. Predictions of charge distributions, hydrophobic and hydrophilic sites on the surface and the average surface charge are useful starting points [71, 72, 73, 74], but the real effect of adsorption on these key factors must be determined in order to better understand the protein-solution interface. However, the requirement that experiments takes place in solution and the nanoscale nature of the system challenges available imaging modalities.
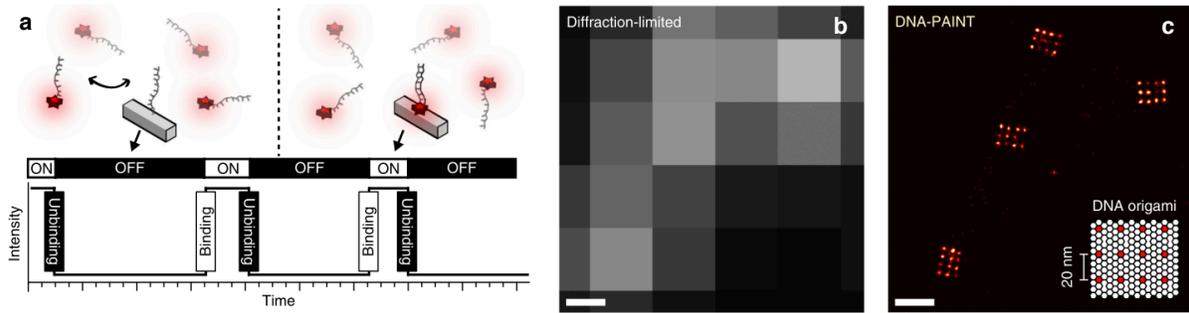
Figure 4.1: A demonstration of the PAINT technique by DNA-PAINT. In (a), the dim fluorescent probes freely diffuse in solution (left) and the target is in the OFF state while there is no attached probe. The target is in the ON state when one of the probes binds to the target, providing one point to localize the target. In (b) and (c) the resolution of regular fluorescence microscopy and the results of PAINT are visually compared. Scale bar is 100 nm. Figure adapted from J. Schnitzbauer et al.[75].

Point accumulation for imaging in nanoscale topography (PAINT) is a powerful technique to measure nanoscale features [76, 77, 78]. Instead of directly labeling the species of interest with a fluorescent dye by covalent conjugation to the protein, PAINT relies on a weak interaction between the label and the target to allow many (100s) of labels to associate and rapidly dissociate from each target site to generate a high resolution picture of the target sites. For example, in Fig. 4.1, DNA origami was mapped out with nanometer precision by the binding and unbinding of fluorescently-labeled marker strands with target strands. In contrast to the covalent labeling of targets, far more fluorophores are able to label each individual site over a longer period of time because the off-rate is faster than the photobleaching rate. Additionally, the stochastic nature of the binding, similar to other super-resolution techniques like STORM [79], means that each label can be more accurately located with subpixel resolution of the PSF due to the lack of nearby, confounding, other point sources.

Here we aim to study the solution-protein interface with PAINT. Bovine serum albumin will be stochastically probed by the binding of the Nile red fluorophore. Nile red is a solvatochromic dye, meaning that the emisson frequency changes with strong dependence on the polarity of its environment [80]. BSA has many hydrophobic sites accessible on the surface in an aqueous solution and in particular two known drug (i.e. small hydrophobic moleulce) binding sites per subunit [81]. When bound to these sites, Nile red fluoresces when excited at 561 nm and otherwise would not fluoresce in the polar buffer[82, 83]. This combination fulfills the components necessary to setup a PAINT system. Beyond its utility as a model PAINT system, the interaction of serum albumin with hydrophobic species in solution is an important factor in human health and this study could provide valuable insight into the kinetic range of those interactions.

In this chapter, the PAINT technique will be established to study the solution interface of adsorbed proteins. The kinetic parameters are determined for a model system of BSA to establish the expected affinity between Nile red and BSA in the single molecule imaging

set up. BSA is non-specifically adsorbed to a glass surface and Nile red is free in solution and allowed to bind to the BSA. Total internal reflection fluorescence (TIRF) microscopy is used to generate high signal to noise ratio for the detection of Nile red binding events. The improvement in positional accuracy due to PAINT is quantified for determining position below the diffraction limit. The dissociation rates are obtained from the duration of binding events, and association rates are obtained from the time between binding events. The effect of acetaminophen on Nile red access to hydrophobic is assessed by determining the drug's effect on kinetics. Acetaminophen, a drug with two hydrophobic binding sites on BSA, is added in increasing concentration to asses Nile red's interaction sites on the BSA surface. With this application of single molecule imaging, the availability of hydrophobic surfaces sites to the solution face is inferred from the effect of acetaminophen binding on Nile red kinetics.
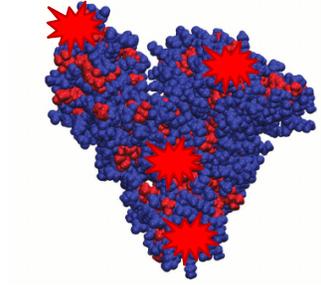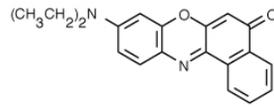
Figure 4.2: The structure of Nile red (a) and the hydrophobic sites on the surface of BSA, in red (b), as determined by the crystal structure [84]. Not to scale.

4.2 Methods

4.2.1 Single Molecule Imaging

Detailed methods are provided in Chapter 2, but briefly, flow cells made with cleaned coverslips were incubated with an excess of fatty acid-free bovine serum albumin (> 30 μM) for five minutes in 10 mM Tris buffered saline (TBS). Unadsorbed BSA was rinsed out with 50×volume of TBS. Nile red was stored in dimethyl sulfoxide (DMSO) and diluted just prior to use into TBS to a final concentration of 1, 2, 3, and 5 nM in one channel of the flow cell. The final solution was never more than 2% v/v DMSO as it effects the viscosity, interactions between all components in the flow cell, and the refractive index of the solution. The last effect makes it impossible to maintain the focus in the Nikon Ti-E microscope using the Perfect Focus System, thereby introducing image drift in the Z-axis. The second channel was treated in the same manner as the first, but with no sample in any solution, i.e. the same incubation and rinse cycles, in order to assess with the microscope the quality of the surface cleaning. If the second channel had significant signal compared to the channel with sample, the flow cell was discarded. For experiments with acetaminophen, the Nile red and acetaminophen were diluted in TBS into the same final solution and added to the flow cell at the same time at 1 nM and 1, 5, or 10 μM, respectively. The second channel was just Nile red at 1 nM in TBS for comparison. Images were acquired with a Nikon Ti-Eclipse as described in 2.1.2. The 561 nm laser was used at 100% laser power for an exposure time of 200 ms with no delay between frames for the experiments with just Nile red and BSA and with 1 s between frames for experiments with acetaminophen. Videos were recorded for an hour each. Images were analyzed as described in § 2.2.1.

4.2.2 Trajectory Localizing

Positions from the output of the drift correction were re-linked across frames using a kd-tree algorithm to connect trajectories that land in the same location. The code is available on

Github at https://github.com/meganarms/SSA and in Appendix A.2. First, the algorithm segments out the lifetimes longer than a single frame for further analysis. Lifetimes less than two frames could be noise above the intensity threshold, whereas lifetimes greater than one frame have a much lower probability to be due to noise. From the segmented lifetimes, the locations of the proteins on the last frame of each lifetime are chosen as seed locations for a kd-tree algorithm. The locations of proteins on the first frame of the lifetimes are searched for those within a two-pixel radius and infinite number of frames after the frames that the seed appears on. When there are multiple trajectories that meet those criteria, the closest protein is picked, and its time and position on the last fame of the trajectory update the seeds last frame time and position to create a new, connected lifetime. If a protein is close to two seeds on the same frame, it is similarly connected to the closer seed and removed from the possible connections for the future seeds. This process is executed on each frame, over all of the last frames in the lifetimes, including the repeats when the a later trajectory had previously been connected.

4.2.3   Binding kinetics

Trajectories obtained from the output of the image analysis were further analyzed to determine the kinetics. Two different off-rates were determined: one based on the interaction of Nile red with BSA at the population level, and one designed to reveal heterogeneities in binding sites. The first type was calculated from the survival function of all residence times with no processing of trajectories further than described above. For the second type, residence times were first identified as coming from one location. The kd-tree algorithm described in   4.2.2 was applied with an extended time allowance so that all events in one location were connected into a single pseudo trajectory. The averages over the connected lifetimes, the "bright times" at an individual location were combined into a cumulative distribution function over all locations then fit to a model of binding.

The times between each connected lifetime in a pseudotrajectory, the "dark times" were

averaged and combined into a CDF over all locations. For a single binding behavior, a single exponential distribution would be expected. However, BSA has the freedom to take different orientations and aggregate when bound to the surface in a monolayer, so multiple characteristic lifetimes are expected. Both the dark time and bright time CDFs were fit in two ways: first with a continuous distribution of exponential lifetimes, and second with a sum of discrete exponential distributions. The inverse of the characteristic lifetimes from the exponential distributions provides the average off-rates, $k_{off,i}$ and on-rates, $k_{on,i}$. The effect on Nile red binding kinetics is examined by adding increasing concentrations of acetaminophen from which the conformation of the solution-facing interface can be inferred

4.3   Results

4.3.1   Point accumulation

Videos of Nile red adsorbing to BSA were acquired for 10 minutes to an hour and processed to determine the binding kinetics between Nile red and bovine serum albumin. The success of detecting multiple binding events at a single binding location was assessed by studying the pseudotrajectories of binding events at a monolayer of BSA and a disperse surface density of BSA. The latter ensures that neighboring proteins are not mistaken for the same binding location by creating an surface density with spacing much greater than that of objective's resolution. For the monolayer, 32% of detected locations experienced multiple events, and for the disperse layer, 20% did at 1 nM Nile red. It is expected that the monolayer has a higher percent of locations with repeating events, due to the possibility of overlapping locations appearing as one location. However, the result with the disperse layer implies that as many as 2/3 of the locations in the monolayer are indeed single locations experiencing multiple events.

A representative pseudotrajecotry (1 of > 5000) is Fig. 4.3 shows that the typical kinetics qualitatively exhibit ideal PAINT characteristics. The peaks represent the time that

a Nile red fluorophore is bound to a BSA, and the rest is the background with no signal. Narrow peaks indicate short binding lifetimes, which means that the site is free for many additional binding events. Wider peaks indicate a longer lifetime and could be subject to photobleaching. A short lifetime ensures that the off-rate detected is not confounded with photobleaching. A short lifetime could also mean fewer frames upon which the location appears, reducing the positional accuracy, however, multiple peaks as in Fig. 4.3 indicate that the photons from multiple events are collected to finally determine the location of the surface protein with high accuracy.
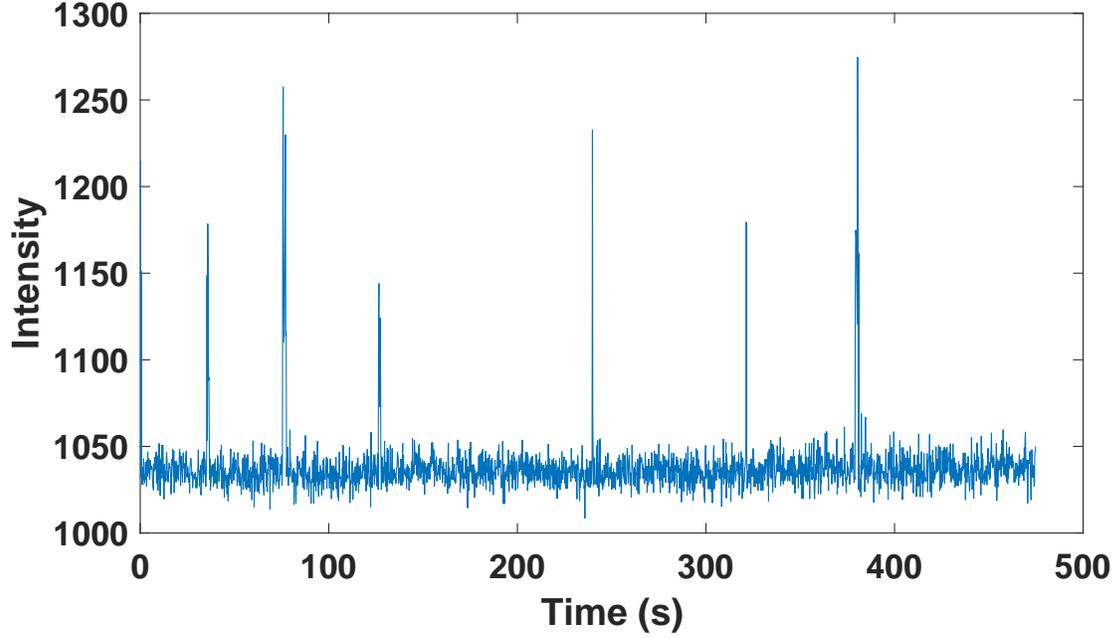
Figure 4.3: A representative intensity versus time trace to visualize a pseudotrajectory. Peaks indicate binding events, the duration of which is the width of the peak, $\tau_{B,i,j}$, also called the "bright time". The time between the end of one peak and the start of the next is the "dark time"", $\tau_{D,i,j}$. For the population-based analysis, each peak is considered to be its own event, unrelated to other events in any given pseudotrajectory. For the location-based analysis, a single location, i, is represented by an average bright time, $\tau_{B,i}$ and an average dark time, $\tau_{D,i}$.

4.3.2   Equilibrium binding

The equilibrium number of objects on the surface was measured at various concentrations to compare the single molecule imaging method to the bulk affinity studies. To determine the equilibrium surface coverage, the cumulative sum in the change in number of objects on the surface was plotted versus time. As each concentration of Nile red is assumed to have the same number of BSA sites to bind to, and the number bound was assumed to be consistent across the surface of the flow cell, so these equilibrium numbers were left as absolute counts from a single field of view. These plots were fit with to a linear curve and had near-zero slopes, so the average surface object across the duration of the video was taken as the equilibrium number of objects on the surface. In Fig. 4.4, it is clear that there is no change in surface object counts across the concentrations tested. Considering that bulk affinity studies are typically accomplished over a concentration range of several orders of magnitude, this result is reasonable. However, single molecule studies are constrained to a much smaller range of concentrations to achieve the optimal imaging and analysis conditions. These data demonstrate that there is not enough resolution within the range of concentrations suitable for single molecule imaging to measure affinity in a manner analogous to bulk affinity studies.
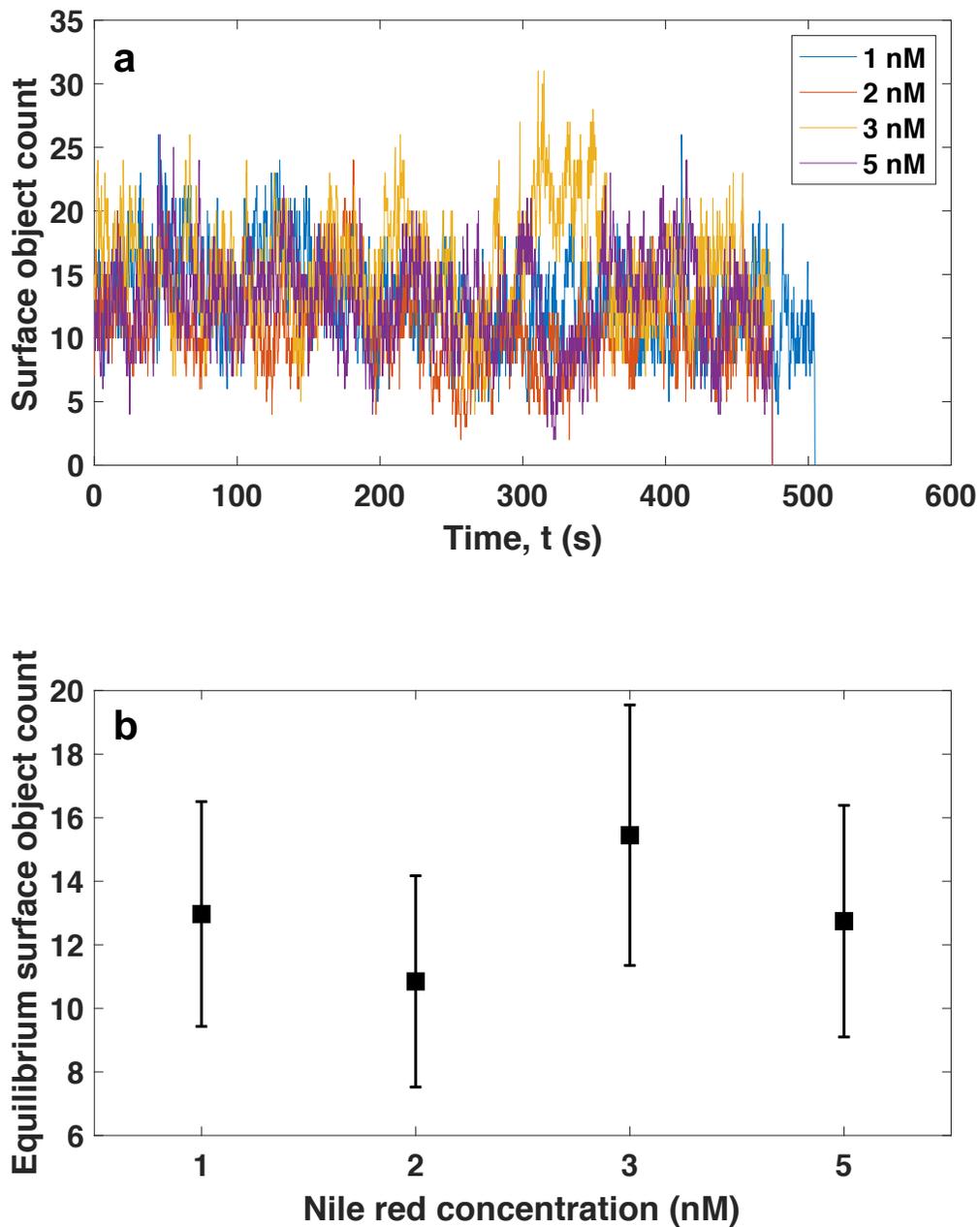
Figure 4.4: In (a), the number of objects on the surface at any given time is plotted. For all concentrations, there is no significant increase or decrease in the number of objects on the surface with time, confirming that the adsorption and desorption rates are at equilibrium. In (b), there is also no significant increasing trend in the equilibrium number of objects with increasing concentration as would be expected.

4.3.3 Binding kinetics

The off-rates were calculated assuming homogeneity over the whole population. The survival function was constructed to obtain the average lifetime that gives the off-rates. The duration of a single lifetime, τ_B, is the width of an individual peak in a trace such as in Fig. 4.3, but is not followed by any other events as "locations" were not considered under this assumption. The individual events were collated into a survival function (§ 2.2.2) to determine the off-rate without regard to location. The MEMExp fit to the survival function (Fig. 4.5(a)) produced two peaks, as shown in the continuous description of the data in Fig. 4.5(b). This suggests that that are two dominant behaviors with dissociation constants $k_{off,1} = 0.7659s^{-1}$ and $k_{off,2} = 0.4203s^{-1}$. Under the assumption of a homogeneous surface, this suggests that at each site there are two binding sites for Nile red with different dissociation constants. As Nile red is expected to bind to the same known binding sites that small hydrophobic molecules bind to, this explanation is justified.
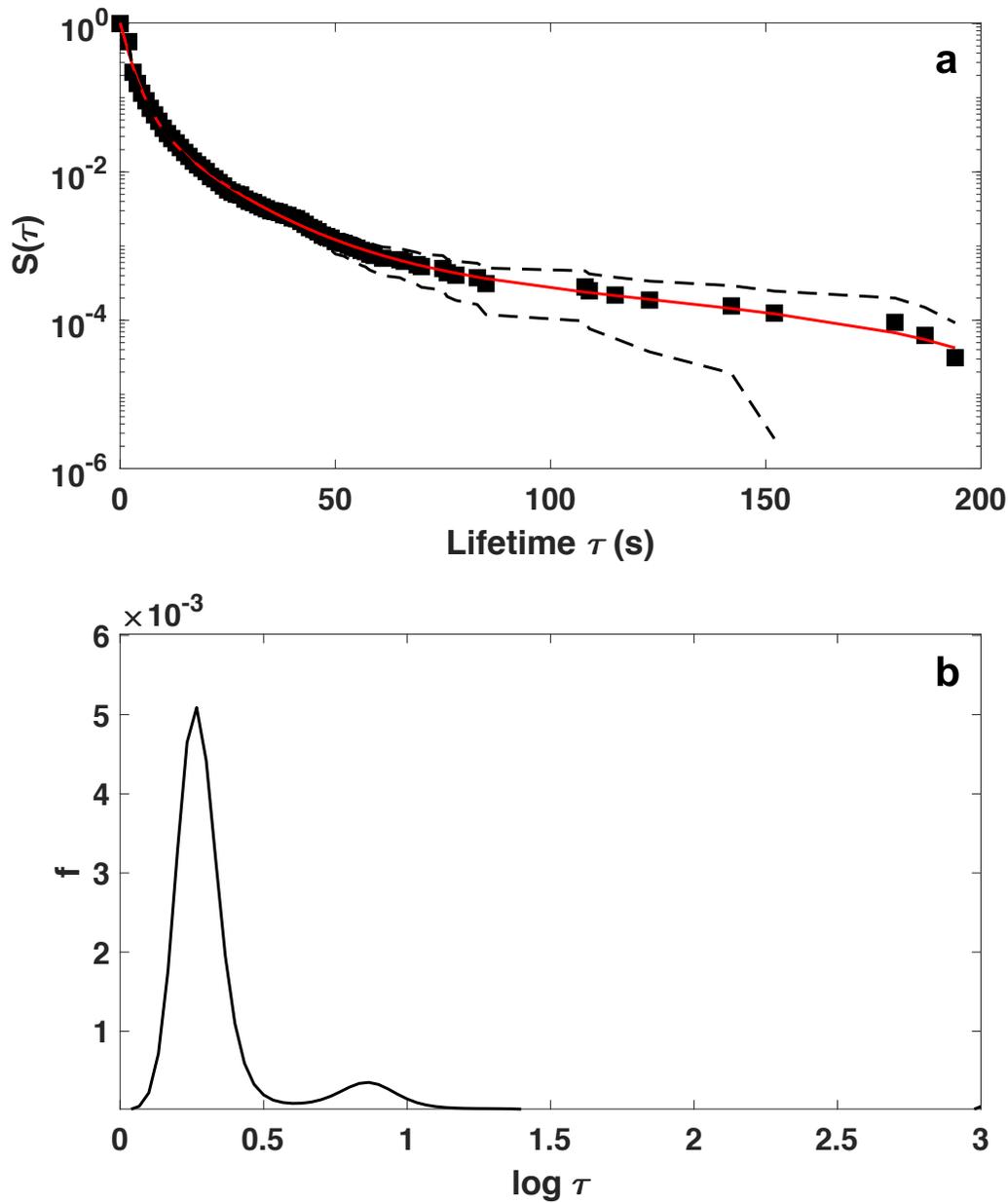
Figure 4.5: In (a), the survival function constructed from the population of all binding events is plotted on a log-linear plot. The upper and lower confidence intervals of the survival function (squares) are plotted in the dashed lines. The fit result from MEMExp is plotted in red and the distribution of exponential kinetic parameters used to fit the data are in (b). There are two peaks, suggesting that the discrete sum of exponential parameters would be composed of two terms.

The alternative assumption is that the binding sites on the surface are not homogeneous. Kinetics under this assumption were calculated from pseudotrajectories, in which the lifetimes of all binding events, $\tau_{B,i,j}$ at a location were averaged. This averaged lifetime, $\tau_{B,i}$ represents the binding kinetics at this location. The CDF was calculated from the location average lifetimes, "bright times", and fit to determine the off-rate, $1/\tau B$, from the parameters of a multi-exponential model. Under this assumption, as each location was treated individually, the time between events could also be analyzed. These "dark times", $\tau_{D,i,j}$, seen in Fig. 4.3, were collated and fit in the same manner as the brighas each

The continuous description of the fit and discrete fits are in Figs. 4.6(a) and 4.7(a) for 1, 2, 3, and 5 nM Nile red binding to a BSA monolayer. The continuous description is interpreted as follows: the x-axis displays the logarithm of the average lifetime of exponential distributions. The value along the y-axis is the relative contribution of that average lifetime to the fit of the data. The parameters for sum of discrete exponential distributions to describe the data were fit separately. The inverse of these discrete parameters are the calculated rate constants, shown in the bar plots in Figs. 4.6(b) and 4.7(b).

The dissociation constants were determined for multiple concentrations and expected to remain constant. In Fig. 4.6(a), this was the case as the continuous descriptions of the data show a peak around 0.45 for all concentrations. This is matched in (b) with the discrete fits, so it can be assumed that this is descriptive of the dominant binding behavior between Nile red and BSA. The additional peaks and exponential terms in (a) and (b), respectively, are likely due to heterogeneity in the surface that changes the way that BSA adsorbs, and therefore what is accessible for Nile red binding.

Compared to the population analysis of kinetics, the location-based analysis seems to reveal more heterogeneity in the off-rates. The implication from the population-based analysis was not rejected by the location-based analysis. The existence of two distinct binding sites on the surface of a homogeneous population of BSA would be averaged into one rate per location. Additionally, heterogeneity of the spots may be blurred by the dominance of

a single off-rate. This is possible because the population-based method gives every lifetime the same weight even if it is repeated several times at the same location. On the other hand, the location-based method gives every location the same weight, even if there are two events at one location, and 20 at another. By using the complementary methods, the average bulk behavior can be determined, while also revealing anomalous behaviors that would be otherwise hidden.

The association constant was also determined for multiple concentrations. The continuous distribution that fits data in Fig. 4.7(a), shows there are similar peaks at all concentrations tested. There is good resolution between the peaks, meaning that the discrete description of the data is a valid interpretation. The concentration-dependent relationship between on-rate and concentration is evident in Fig. 4.7(b), where the overall trend is a positive correlation between the variables. This is seen in particular for the two fastest on-rates that are present for all concentrations. The existence of multiple peaks in this location-based analysis can be interpreted through an understanding of what drives the on-rate. It is a product of the mass transport to the surface, i.e. the collision rate, and the sticking probability, i.e. the successful binding rate. The collision rate is concentration dependent, and the success rate depends on the accessibility of the of the binding sites. Therefore, the characteristic on-rates depend on how the BSA is oriented on the surface. It also depends on the approach of the Nile red, but the Nile red has the freedom to diffuse and rotate in solution, making the BSA orientation the limiting factor in binding. The two fastest on-rates then (blue and orange bars), are representative of the two most favorable orientations for Nile red to access and bind to the BSA surface. Because this is a location-based analysis, it also means these orientations are the most frequent out of the orientations which allow binding. Of course, it does not necessarily indicate that there only two distinct orientations contributing to these type of binding events because multiple orientations can lead to indistinguishable binding kinetics. For the sake of understanding the BSA-solution interface however, there are effectively two orientation types that are consistently favorable to

binding about every 2 seconds and about every 10 seconds. The additional peaks that are inconsistent across concentrations indicate that there other adsorption orientations possible. However, they are likely dependent on the presence of conditions external to the designed system, such as imperfections in the surface that allow for different conformations of adsorbed protein and therefore different interactions with the Nile red. The slower rate means that the hydrophobic sites are less accessible to components of the solution than the conformations producing faster on-rates, suggesting either a more closed conformation, or more likely, that these proteins are bound via the hydrophobic effect to a hydrophobic patch on the surface and the hydrophobic residues are turned more towards the surface.
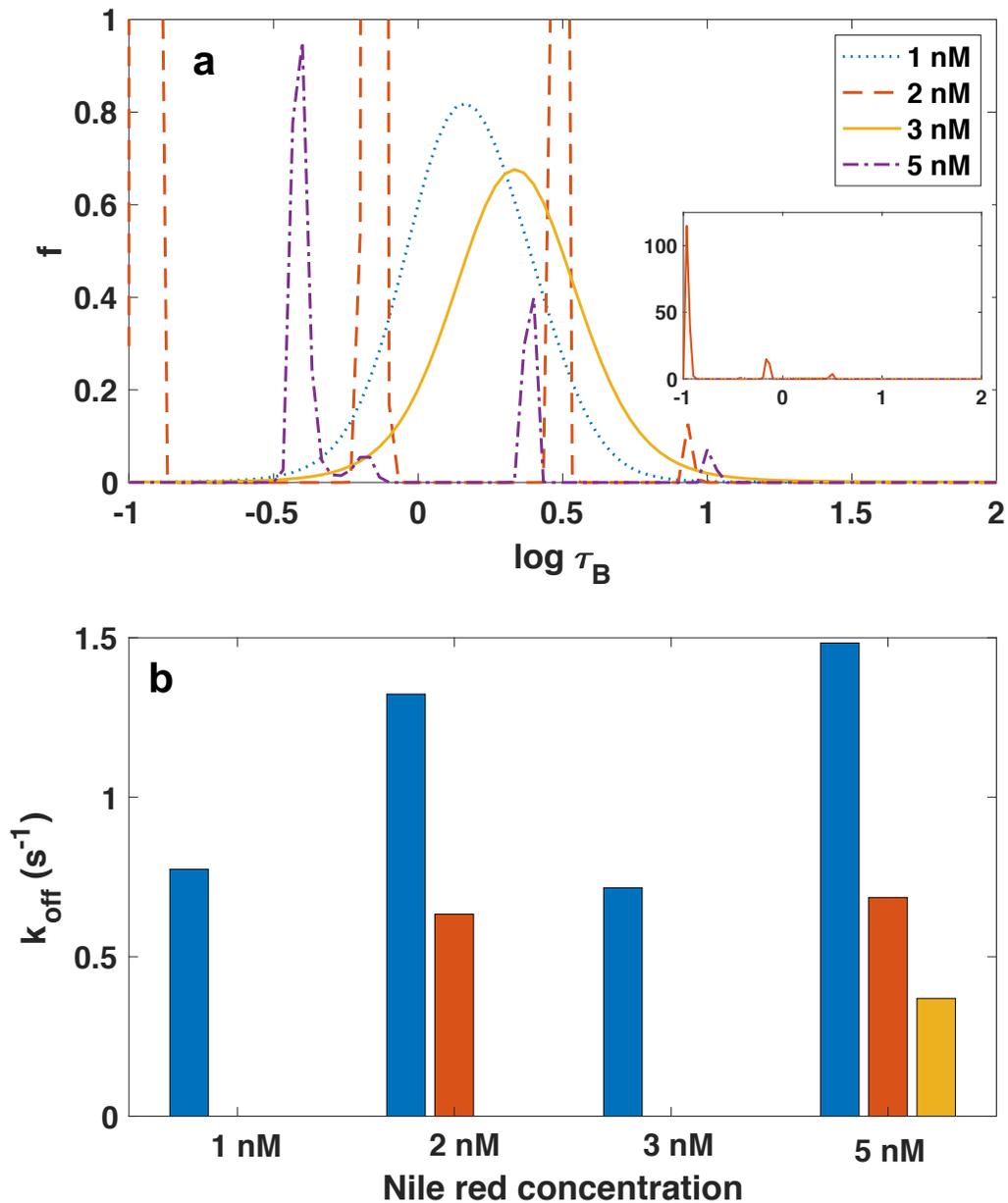
Figure 4.6: In (a), plotted is the distribution of the logarithm of the kinetic parameters, $\log \tau_B$, that fit the average bright time CDF from pseudotrajectories. Each concentration was fit separately, and the peaks describe the most frequent exponential distribution parameters. All four concentrations overlap at one peak around 0.4 that is matched in the discrete fits (b). This represents a consistent off-rate expected for changing concentration. Other peaks and discrete fit parameters are likely due heterogeneity in the surface itself and therefore different BSA binding leading to access to different sites for binding. Inset is the full range of the peaks cut off for 2 nM.
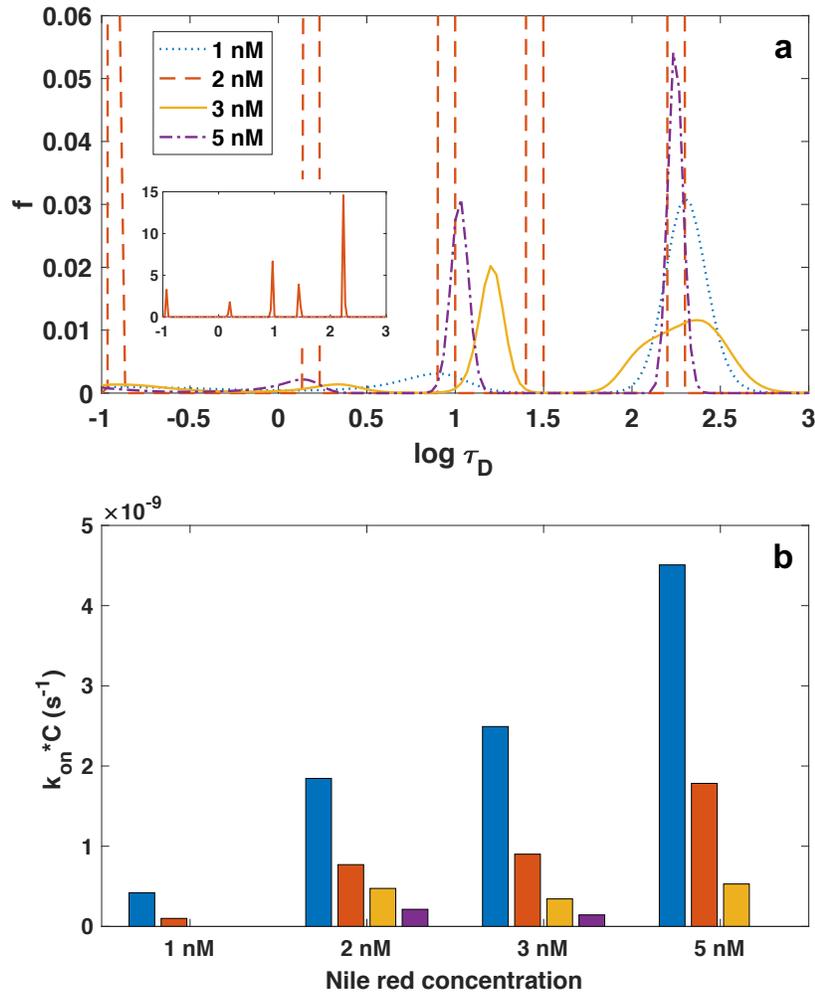
Figure 4.7: In (a), plotted is the distribution of kinetic parameters, $\log \tau_D$, that fit the average dark time CDF from pseudotrajectories. Each concentration was fit separately, and the peaks describe the most frequent exponential distribution parameters. There are consistent peak locations across all four concentrations, with 2 nM having an additional fifth peak that does not correspond to the peaks in the distributions for 1, 3, and 5 nM. Due to the effects of fit normalization, the top of the peaks are cutoff for 2 nM and the fastest two peaks (< 0.5) are difficult to discern for the other three concentrations. The inset in (a) shows the amplitude of the cutoff peaks for the 2 nM data. The overlapping peaks in (a) suggest that there is a consistent concentration-independent on-rate across the concentration trials. In (b), the fastest two concentration dependent on-rates increase proportionally with concentration. The existence of multiple on-rates at each concentration suggests different orientations of BSA adsorbed to the surface may lead to different pathways to the binding pockets with correspondingly different success rates.

4.3.4 Binding inhibition

Acetaminophen was added in increasing concentrations to determine the effect of another small hydrophobic molecule on Nile red binding kinetics. Acetaminophen is expected to inhibit the binding of Nile red, as both are expected to bind the solution-hydrophobic sites. The interactions between acetaminophen and BSA have been studied: there are two expected sites of interaction, where hydrophobic residues allow for the hydrophobic effect to bind the drug to BSA. Nile red is expected to behave in a similar manner. Therefore, the on-rate is expected to decrease with increasing drug concentration as there will be fewer successful collision due to blocked binding sites. The off-rate is expected to increase, as there will be competition for the binding sites with the Nile red. On a monolayer of BSA, 1 nM Nile red was added with 1, 5, or 10 μM acetaminophen in the same solution. The discrete off- and on-rates were calculated in the same manner as described above.

The results (Fig. 4.8) show that Nile red has a single on-rate of $0.4 M^{-1} s^{-1}$ that is exhibited at all concentrations of acetaminophen. The on-rate changed in the expected manner, however apparently independent of the inhibitor concentration. All experiments were done at the same concentration of Nile red, so multiplying by the concentration would not effect the relative amplitude of the on-rates. In Fig. 4.7, the two fast on-rates consistent across concentration indicated that there are two common binding pathways. With the addition of acetaminophen, there is only one consistent consistent on-rate. This implies that acetaminophen has a much more favorable interaction with one pathway type than the Nile red, greatly reducing the chance of Nile red binding in that manner. Additionally, the presence of the drug reduces the rate of any binding, as $0.4 M^{-1} s^{-1}$ is slower than any on-rate without it. As before, the on-rate that is not seen at other concentrations is likely due to heterogeneity on the binding surface for that particular experiment that was not in other experiments.

In Fig. 4.8(b), there is essentially a single Nile red off-rate in the presence of acetaminophen. The off-rate was expected to increase with increasing concentration of acetaminophen

due to competition of the binding site, but instead it decreased from 0.8 to $0.6s^{-1}$. This is likely due to a stronger binding site. The acetaminophen, by making other binding sites inaccessible, forces the Nile red binds to a site that, once bound, is more tightly held. That is, the Nile red binds to a site with a higher energy barrier to binding and to unbinding. However, it then would be expected that there would be a fraction of binding to this site seen in experiments without the inhibitor in Fig. 4.8(b). Understanding temperature dependence of these on- and off-rates would help to test this implication. Alternatively, the acetaminophen could introduce a stronger binding interaction by cooperatively binding with the Nile red. Experiments over a larger range of concentrations of acetaminophen (as discussed above) to produce an isotherm could help elucidate the source of the slower off-rate. Finally, it is tempting to think that at 10 μM there is an increase in the off-rate, but the presence of the more characteristic off-rate and the presence of binding heterogeneity in other experiments suggests that the rate of interest is in the orange bar. Increased concentrations of acetaminophen would be necessary to confirm that.
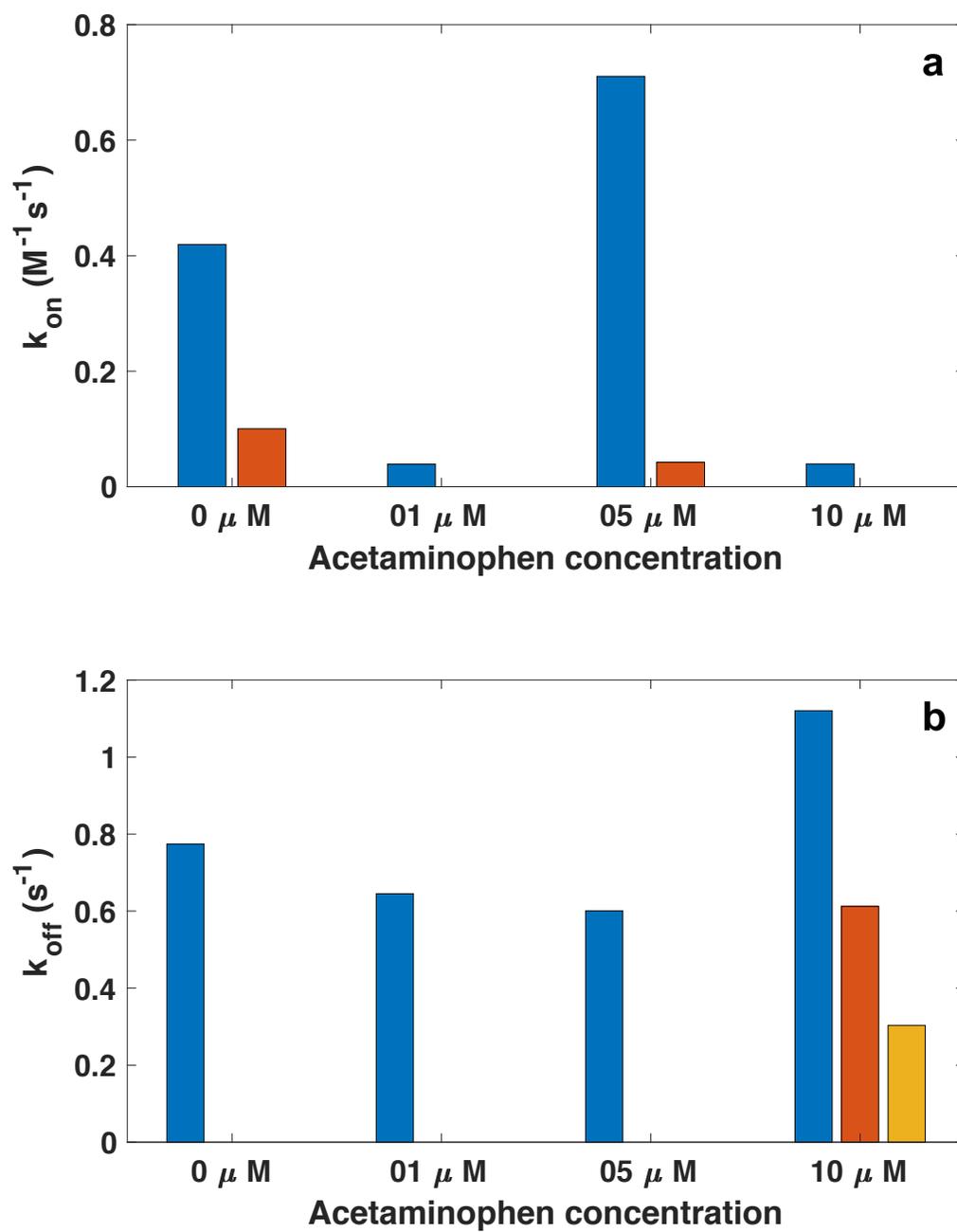
Figure 4.8: The on-rates and off-rates of Nile red in the presence of increasing inhibitor concentration.

4.4 Discussion

Nile red was allowed to bind Bovine serum albumin was adsorbed to a glass surface in a monolayer and at disperse concentrations to demonstrate the multiple events can be accumulated in a single location as required for the super-resolution imaging technique "point accumulation for imaging nanoscale topography", PAINT. The range of Nile red concentrations suitable for single molecule imaging was established to be in the single nanomolar range. The kinetic parameters for these concentrations were established, allowing the rational design of future PAINT experiments (e.g. probe/targe concentration, experiment duration for desired number of events, etc.). The assumed interaction between Nile red and BSA was tested by introducing an inhibitor. The on-rate of Nile red was slowed with the inhibitor, but the off-rate also decreased, suggesting a more complex interaction between inhibitor and Nile red than was expected.

Future work includes increased ranges of the experiments already performed and some complementary experiments in order to confirm the results shown here. Some limitations of single molecule imaging were discussed in the text, but complementary work is ongoing to confirm the results presented here. The initial ground work has been laid to use super-resolution PAINT imaging to better understand the solution-protein interface. Nile red and BSA has a quick off-rate and a slow enough on-rate to distinguish between binding events, and also accumulate significant points over longer experiments. This method need not be limited these two PAINT partners. Once any protein is adsorbed to a biomaterial, it may present a different interface to the solution. This novel technique to determine the mix of kinetic parameters through complementary population- and location-based analysis could be applied to potentially a multitude of protein surface interactions in solution. Eventually, it could give rise to a method for super-resolution imaging of the protein-solution interface, enhancing our ability to design and predict the reaction to introducing protein-presenting biomaterials into biological environments.

Chapter 5

Characterization of IgA1 glycosylation by lectin binding

5.1   Introduction

Over 50% of human proteins are glycosylated, diversifying the function of the 40,000 proteins for which the genome encodes. Glycosylation plays a role in various contexts including development, cell recognition, inflammation, fertilization and cell adhesion. Extensive research has been invested in determining the role and structure of glycans in order to glean from them their biological information. Studying glycans difficult because it is a post-translational process with a vast array of combinations and arrangements of component parts. A single protein will have many glycoforms with separate functions that are influenced by the physiological state of the organism. Despite these difficulties, work thus far has elucidated the role of particular glycans in cancer, rheumatoid arthritis and nephropathy, among other diseases.

Many techniques have been established for glycomic analysis: frontal affinity chromatography (FAC) [85], capillary electrophoresis (CE) [86], nuclear magnetic resonance (NMR), various types of liquid chromatography [87], and mass spectrometry (MS) [88]. These have detailed structural information for a vast array of glycoproteins, but struggle or are unable

to resolve structural isomers. In an effort to directly identify structural information, a series of tandem MS techniques have been developed utilizing different MS techniques in series, or combining with chromatography separation techniques [89, 90, 91, 92]. However, these techniques require extensive and costly sample preparation. The high throughput nature of recently-developed microarrays of lectins or glycans in electro- and optical- biosensor form have led to the rapid increase in understanding in glycobiology [93, 94, 95, 96, 97, 98]. However, microarrays still require the release of the glycans from the protein like the previous generation of techniques and must artificially create enough interactions between glycan and lectin to create enough signal to detect. Additionally, the heterogeneity of glycoforms within a single population can only be detected at a coarse-grained level as this is still a bulk method. Therefore, glycobiology would benefit from the development of a technique probing individual and *in situ* glycan sites with limited experimental preparation.

Here, a protocol is introduced utilizing single molecule imaging to directly probe the glycan structure of glycoproteins. The method would use fluorescently-labeled aptamers to designed to bind to specific saccharides. Aptamers are short, easily tunable sequences of nucleic acid with demonstrated specificity for many biological components, including mono- and disaccharides[99, 100]. The aptamer-glycan affinity is suitable to accumulate a large number of photons from a single binding site through repeated binding. Therefore the PAINT technique introduced in the previous chapter will be implemented in the determination of the structure glycan sites on glycoproteins as an auxiallary structure and sequence determining method. First however, the kinetic parameters must be quantified at a single molecule imaging level for glycans. Correlations must be established between different standard glycan structures and kinetic parameters in order to provide an expected range of output for PAINT. A lectin-glycan system provides an ideal model system to establish these parameters as there are determined kinetic parameters to compare against higher affinities than for a single aptamer-monosaccharide interaction.

In this chapter, the kinetic parameters are established for a model system of IgA1 and

two lectins: jacalin and peanut agglutinin (PNA). The glycosylation of IgA1s is important in particular because those that lack the terminal sialic acid form complexes with IgG that accumulate in the kidney: this is the leading cause of nephropathy[101, 102]. Jacalin and PNA have reported affinity for the unsialylated Galβ 1-3GalNAc sugar epitope in the IgA1 hinge region[103, 104]. Here, IgA1 is fixed to a glass surface and the jacalin and PNA are fluorescently-labeled and free to diffuse in solution and bind to the IgA1 on the surface. Total internal reflection fluorescence (TIRF) microscopy is used to high signal-to-noise ratio images sequences of the binding events. The ability of the surface to prevent non-specific interactions is evaluated. The specificity of the interactions is tested by comparing lectin binding to bovine serum albumin, a non-glycosylated protein, fixed to the surface. The dissociation rates (k_{off}) are obtained from the lifetime of binding events, and the association rates (k_{on}) are obtained from the lifetime between binding events. Finally, affinity constants are calculated as the ratio of these two parameters and compared to previously determined values.

5.2   Methods

The basic setup for single molecule experiments is described in Chapter 2. Briefly, flow cells were prepared and treated with heterobifunctionalized PEG, MW5000. The PEG was functionalized with silane on one end to attach to the surface, and on the other side N-hydroxysuccinimide ester (NHS) to link proteins to the PEG through hydrolysis with surface lysine groups. After assembling the flow cell, silane-PEG-NHS was flowed in at 10 mg/mL in dimethylsulfoxide (DMSO) and allowed to incubate for 30 minutes. This was rinsed out with 10x the volume of the channel of DMSO, followed by 15x the volume of the channel of ultrapure water (> 18 MΩ-cm). Next, the IgA, or other target protein, was added to the channel at about 0.3 mg/mL in 10 mM HEPES buffer with 10 mM CaCl$_2$, pH 8.5, and allowed to incubate for 2 hours at room temperature (22$\pm$1°C) in a humidity chamber.

Unreacted protein was removed with 15x channel volume of HEPES buffer. The fluorescent probe dilution was then prepared for best accuracy of the concentration and added to the channel at the desired concentration - 10, 25, 75, or 100 pM. Finally, the ends of the channel were sealed with Dow Corning high vacuum grease to prevent evaporation and reduce the amount of dissolved oxygen leading to photobleaching.

Samples were imaged on a Nikon Ti-Eclipse inverted microscope with TIRF module on the Andor iXon DU-897 X-1755 camera with a Nikon $100\times$ NA1.45 Oil objective. Images were acquired in the Nikon Elements Advanced Research software. The gain multiplier was 150 and readout speed was 10 MHz with 14 bits per pixel. Images were acquired every second for 200 ms each frame and 1 hour of imaging total per experiment.

Images were analyzed in Matlab using a combination of tracking software from R. Parthasarathy and in-house code, described in detail in the setup chapter, Chapter 2. Additional analysis required for repeated binding events was also described in Chapter 4.2.3. Briefly, videos from individual experiments were segmented to allow for different intensity thresholds required due to an initial depletion of fluorescence intensity in the frame. Tracked objects were then linked across frames using a nearest-neighbors algorithm trajectories of the objects in time, space, and intensity. Linked trajectories were then further assessed for their proximity to other trajectories and pseudotrajectories were created as in Ch.4. The fitting and determination of parameters also follows the same procedure as in Ch.4. The inverse of τ_Bs and τ_Ds determined from the MEMExp fit provided the off- and on-rates, respectively, k_{off} and k_{on}. These parameters are used to calculate the dissociation constant, $K_D = k_{off} * C/k_{on}$, where C is the concentration of the probe protein. Off-rates were also fit for a population-based analysis of the binding events. The calculated parameters were compared to known values for the affinity between the probes and targets in order to assess the validity of the experimental and analysis methods.

5.3 Results

5.3.1 Surface Quality

Multiple surface treatments were tested to determine the best quality system to identify binding events of interest. The best quality surface would exhibit 2 key attributes: 1) resistant protein adsorption in the absence of specific binding sites, 2) covalently links the target protein so that binding kinetics between target and surface do not confound the kinetics of interactions between probe and target. To achieve this, a heterobifunctionalized PEG was utilized. On one end, the silane group binds the PEG to the surface, thus providing a background surface that is resistant to protein adsorption. The other end is N-hydroxysuccinimide ester (NHS), which reacts with the amine group presented by protein surface lysine to covalently attach the protein to the surface. Both sides of the functionalized PEG are necessary for this experiment to work, so both were tested. If the non-fouling surface does not prevent adsorption as expected, then there could be too many non-specific adsorption events to detect the probe-target interactions. Therefore, the attribute (1), the efficacy of the silane-surface attachment to create a PEGylated surface, was first tested.

It has been suggested that improved PEG coating could be achieved with multiple rounds of PEGylation - first with a higher molecular weight, and second with a lower molecular weight[105, 106]. To test this and the non-fouling ability of the silane-PEG-NHS, the silane was allowed to react with the surface for 30 minutes in DMSO and then rinsed off with DMSO. In one channel, another react and rinse step was carried out with MW2000 PEG, after which both channels were rinsed with molecular biology grade water. These were then incubated with HEPES buffer pH 8.5 as if the target protein were reacting with the NHS groups, but instead the NHS groups all just hyrolyzed. When incubated with protein, the carboiimde reaction is competing with the NHS hydrolysis, so even without protein, the groups will be inactivated by spontaneous hydolysis. The NHS ester has a half life of 10 minutes under these conditions, so after two hours, the NHS groups are sufficiently quenched

so that the surface is simply presenting non-fouling PEG[107]. Bovine serum albumin labeled with AlexaFluor 488 (BSA488) was then added after rinsing to image any adsorption to either surface, thereby determining the success of the silane attachment of PEG to the surface.

In Fig. 5.1, (a) shows the protein adsorption (bright green dots) to the 2xPEG surface, and (b) shows the results with a 1xPEG surface. It appears that the additional PEGylation step yields no improvement over a single step. This may perhaps be due to the small difference between the molecular weight of the two polymers used. The ability of a second polymer to adsorb depends on the grafting density of the first polymer and the radius of gyration of both polymers. Unfortunately, PEG2000 was the smallest degree of polymerization available from the supplier. However, a future researcher may investigate the improvement in adsorption resistance by a shorter polymer chain without the NHS functional group, which may be available from other suppliers. This would not improve the density of available sites to which the target protein could attach, but may improve the non-fouling capacity, which is in fact the point of this additional step.

While the number of incubation steps did not seem to improve the non-fouling quality of the surface, compared to the uncoated surface, the result is as expected. In Fig. 5.1(c), BSA488 is allowed to adsorb at the same concentration as in (a) and (b). The 10000-fold decrease in adsorption from the uncoated to the coated surface provides assurance that the non-fouling coating is present and functional.
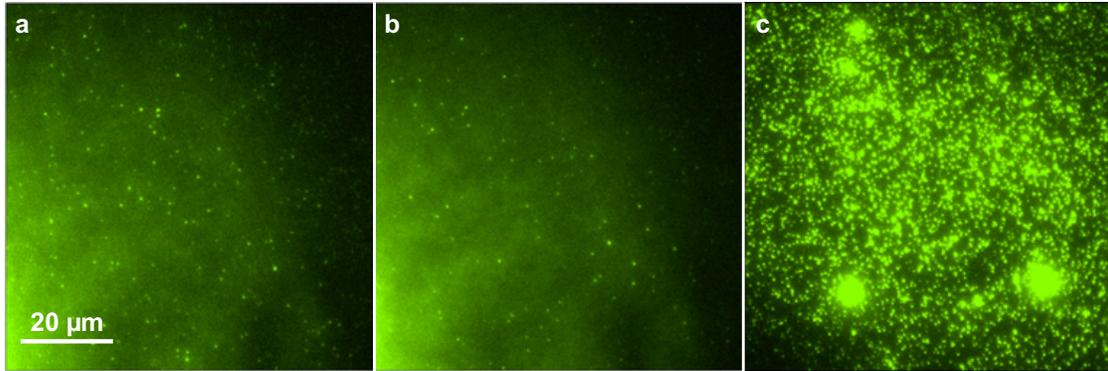
Figure 5.1: Representative images of different surface treatments demonstrate how the PEGylation does or does not prevent non-specific adsorption. In (a), multiple rounds of PEGylation with first MW5000 PEG then MW2000 PEG does not improve the resistance to adsorption compared to a single round of PEGylation with MW5000 (b). For comparison, (c) shows the same concentration of protein adsorbing to a surface treated exactly the same just excluding PEG.

To test attribute (2), i.e. the success of the covalent-bond formation between the PEG and target protein, BSA488 was incubated with the NHS as would be the target protein. The flow cells were rinsed of unreacted protein and imaged to detect the linked protein. The surface was completely coated to the point of not being able to resolve distinct spots and remained this way for the duration of the video (Fig. 5.2). Uneven illumination causes the bottom left of the image (Fig. 5.2(a)) to appear more densely coated in earlier frames and the upper right to appear more densely coated in later frames (b).
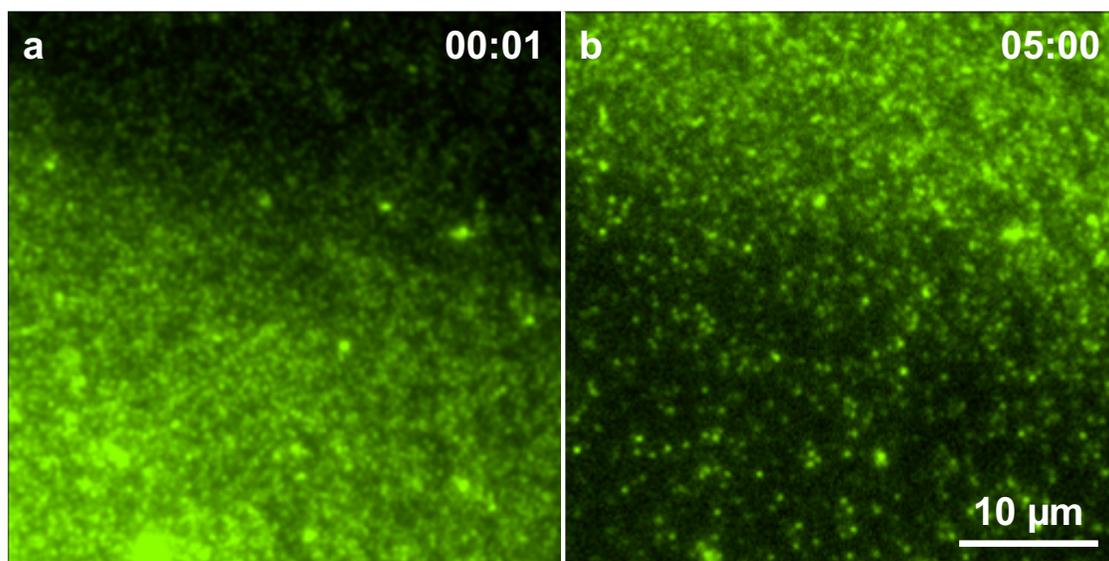
Figure 5.2: Representative images from video of BSA488 covalently-linked to the surface after 1 s (a) and 300 s (b). At both time points there is enough protein linked that individual proteins are unresolvable, demonstrating the success of the NHS-linkage.

5.3.2 Specificity

IgA1 myeloma protein presents 4-6 oligosaccharides in its hinge region that have known specificity with glycan binding protein (GBP or lectins) jacalin [cite papers that use jacalin to concentrate IgA1]. Jacalin has specificity for Galβ1-3GalNAc, one of the terminal sugars on the hinge-region IgA1 glycans. The K_D is expected to be about 10^{-6}. Peanut agglutinin may also bind to Galβ1-3GalNAc, but with lower expected affinity. While the PNA has specificity for that sugar, it does not demonstrate such high specificity when the sugar is linked to the protein. This is likely due to interactions with the surrounding protein structure that interrupt binding between PNA and the target sugar.

In order to test that specific binding occurs between the lectins and the IgA, the lectins were allowed to probe a non-glycosylated protein surface. Bovine serum albumin was covalently linked to the surface and unreacted protein was removed. In separate channels, Jac488 and PNA488 were added at 100 pM and imaged. As is seen in Fig. 5.3, both Jac and PNA had similar binding affinity to the BSA surface, and provided a baseline for what non-specific binding looks like that that concentration. Baseline statistics were also acquired for the sake of comparison to the statistics of specific binding experiments at both 10 pM and 100 pM.
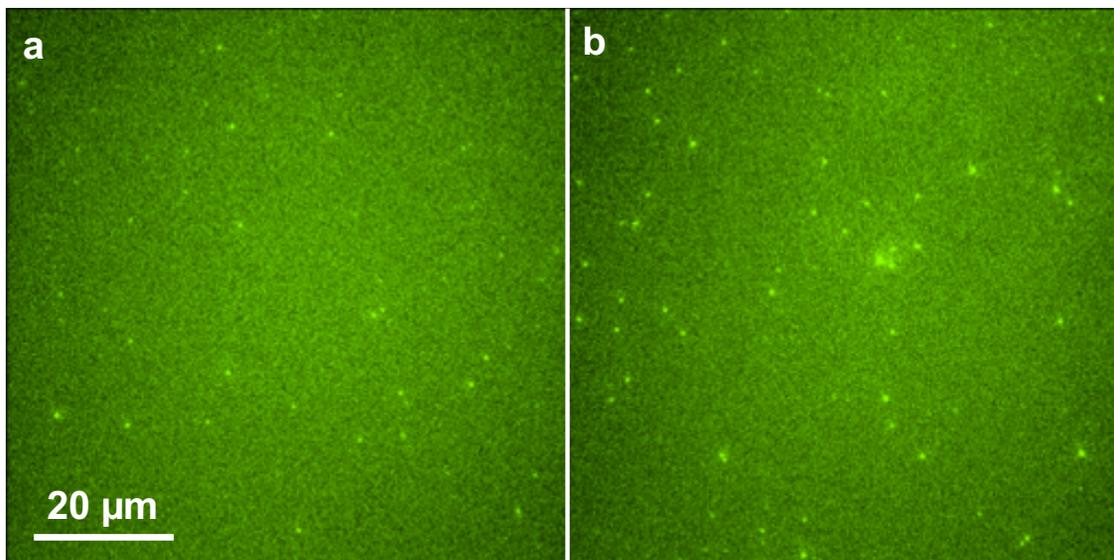
Figure 5.3: Representative frames from the same time point in acquisition to compare binding of probes jacalin (a) and peanut agglutinin (b) to non-specific protein (bovine serum albumin) covalently linked to the surface. The number of events is expected to be the same across each trial, as this value depends on the concentration and sticking probability. For a non-fouling surface with all samples at the same concentration, the number of events should therefore on average be the same, as is seen here.

5.3.3 Affinity Constants

Distribution calculation method

Single molecule imaging experiments generally rely on residence time distributions for the calculation of statistics due to the high number of samples that are acquired and the limited data that can be acquired for a single landing event [38]. This method is generally also effective to detect heterogeneity in the system through the fitting of models to the residence time distributions, as was discussed in detail in Chapter 3. For these reasons, it was expected that the affinity constants could be calculated from a ratio of off-rates and on-rates determined from the residence time distributions.

To construct the residence time distribution, the intensity vs. time data was calculated from the individual binding locations as described in Ch. 4. For each location, an average was determined for the time a lectin would stay on that location, "bright time", and the time between bright events, "dark time". The averages for each type of event - bright or dark - were collected into cumulative distribution functions over all locations. CDF bins were 50 ms for the 200 ms exposure time and 30 ms for the 30 ms exposure time. The complement CDFs were fit with the MEMExp fitting algorithm[108]. The algorithm determines a continuous distribution of lifetime parameters for a sum of exponentials to describe the data. A representative example of the results from the MEMExp program is shown in Figs. 5.4 and 5.5.
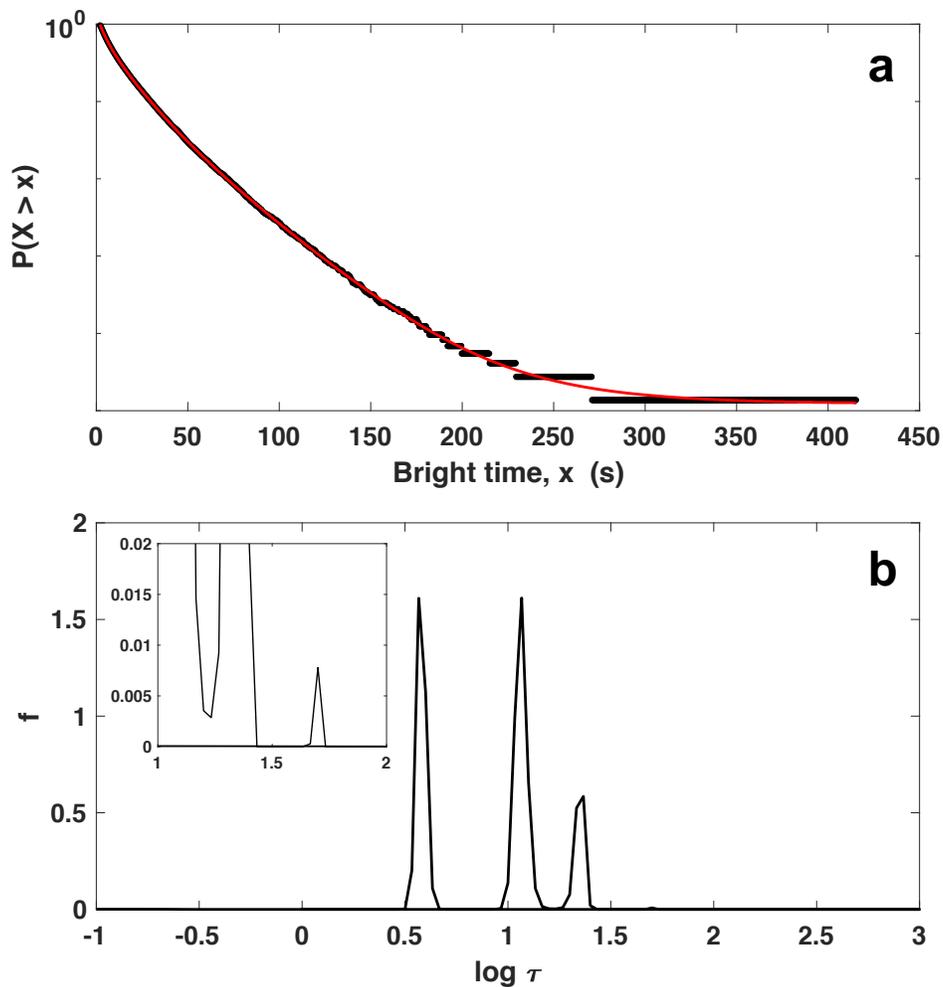
Figure 5.4: In (a), a plot of the MEMExp fit to the bright time averages from individual spots is shown. The experiment was run for 100 pM jacalin on IgA for 1 hour. In (b), a sample plot of the frequency of exponential decay parameters (τ) versus the log of τ generated from the same data. The full distribution was used to generate the fit (red line, (a)), and its maxima determine the kinetic parameters to describe the experimental results. Inset shows a fourth maximum by rescaling the axes.
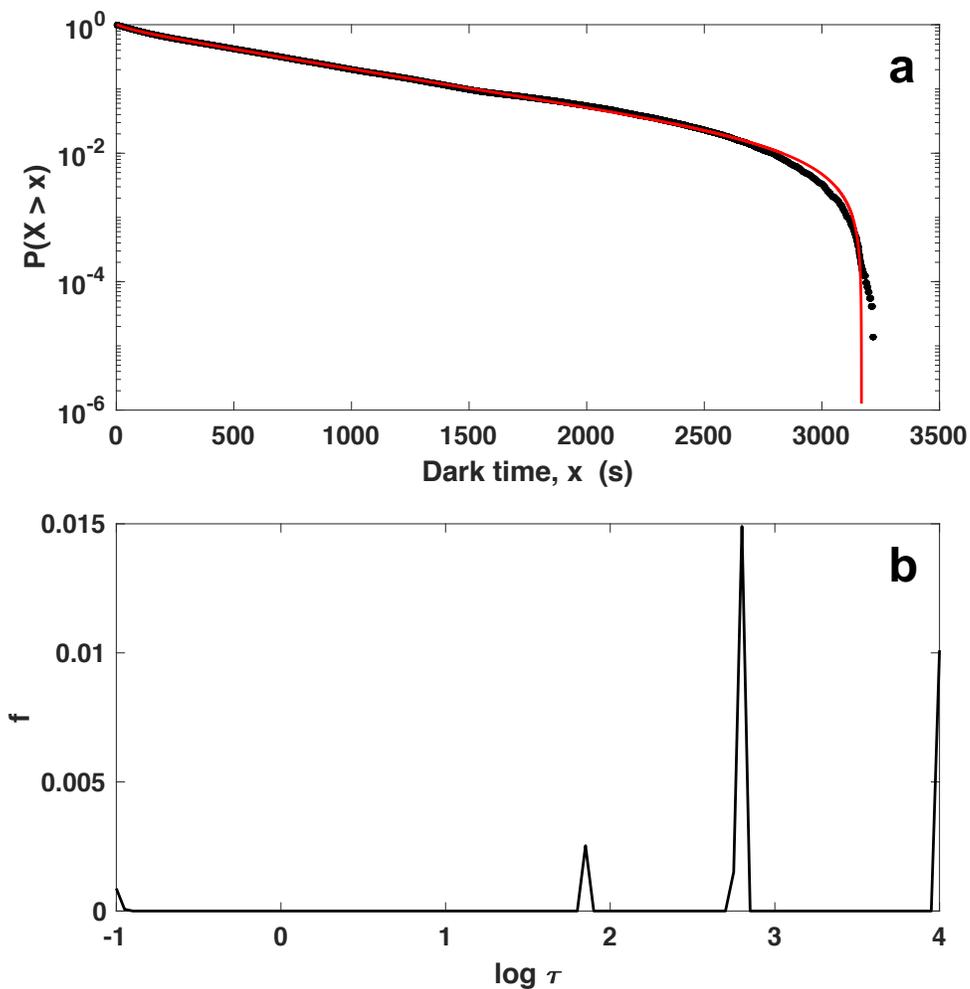
Figure 5.5: The MEMExp method was used to generate a fit and kinetic parameters, as in Fig. 5.4, but here for the dark time averages of individual spots rather than the bright time. The on-average higher decay rates seen in these two plots is representative of the fits for the experiments not shown.

The peaks in the continuous distribution provide the off-rates characteristic of the given CDF. That is, the exponent of the location of the peak on the x-axis provide the logarithm of the average lifetime, τ of the exponentials used to fit the CDF.

$$\tau_i = \exp \log \tau_i \tag{5.1}$$

The continuous description of the data for the IgA-binding experiments was compared to that of the BSA-binding experiments to detect any kinetics that may be due to non-specific binding during the experiment. In Fig. 5.6(a), both PNA and Jac have overlapping peaks with BSA in the continuous description of the data. As some peaks are hidden due to the scale of the y-axis, Fig. 5.8(b) presents the $k_{off,i}$ values to be clear which number peaks overlap. The rates were calculated from the continuous distribution for the location of each maximum along the x-axis, $\log \tau_i$:

$$k_{off,i} = \frac{1}{\exp \log \tau_i} \tag{5.2}$$

Following the same logic with the bright-data, i.e. that which aggregates the average time between events, non-specific k_{on} rates were also eliminated. The $K_{D,i}$'s were then calculated as:

$$K_{D,i} = \frac{k_{off,i}}{k_{on,i}} \dot{C} \tag{5.3}$$

where C is the concentration of the lectin. The resulting affinity constants are shown in Fig. 5.7 for lectins jacalin and PNA binding to IgA. The expected dissociation constants are on the order of 10^{-6} M and 10^{-4} M for jacalin [109] and PNA [110, 111], respectively, but the calculated results are on the order of 10^{-9} or 10^{-10} M.
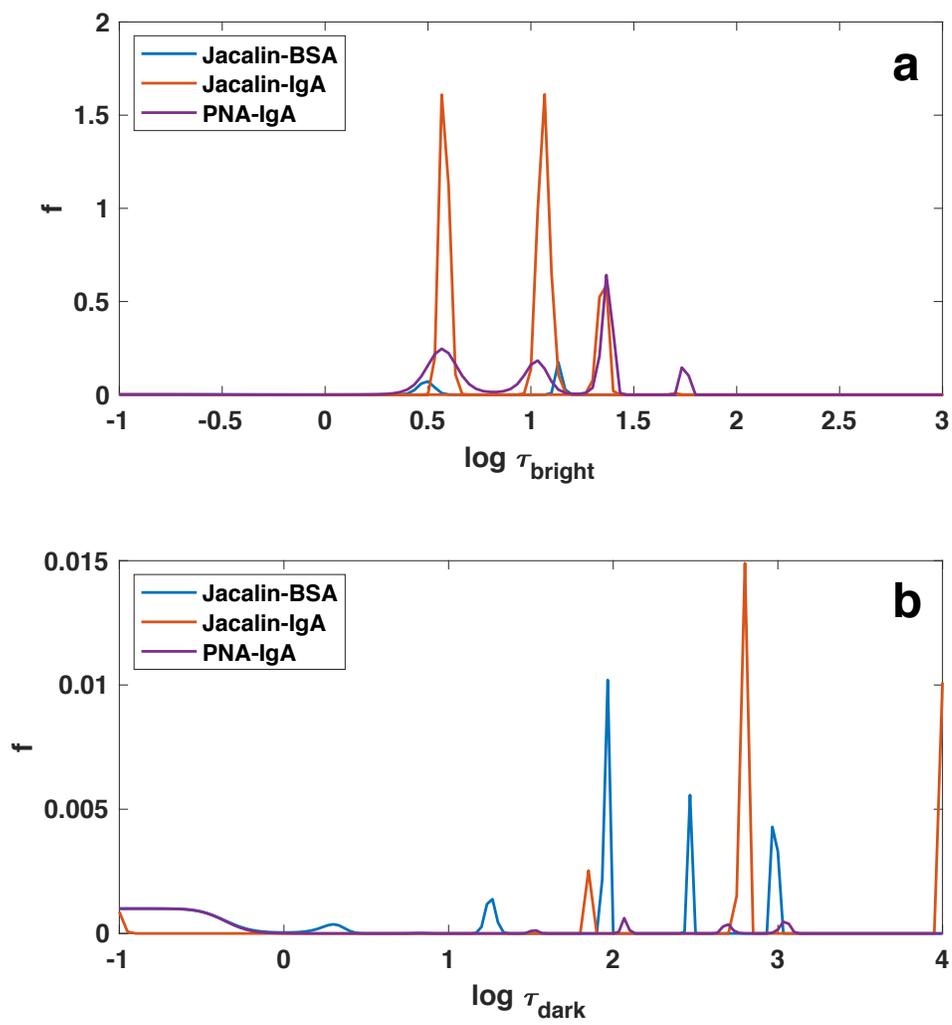
Figure 5.6: Continuous description of lectin-IgA binding kinetics data
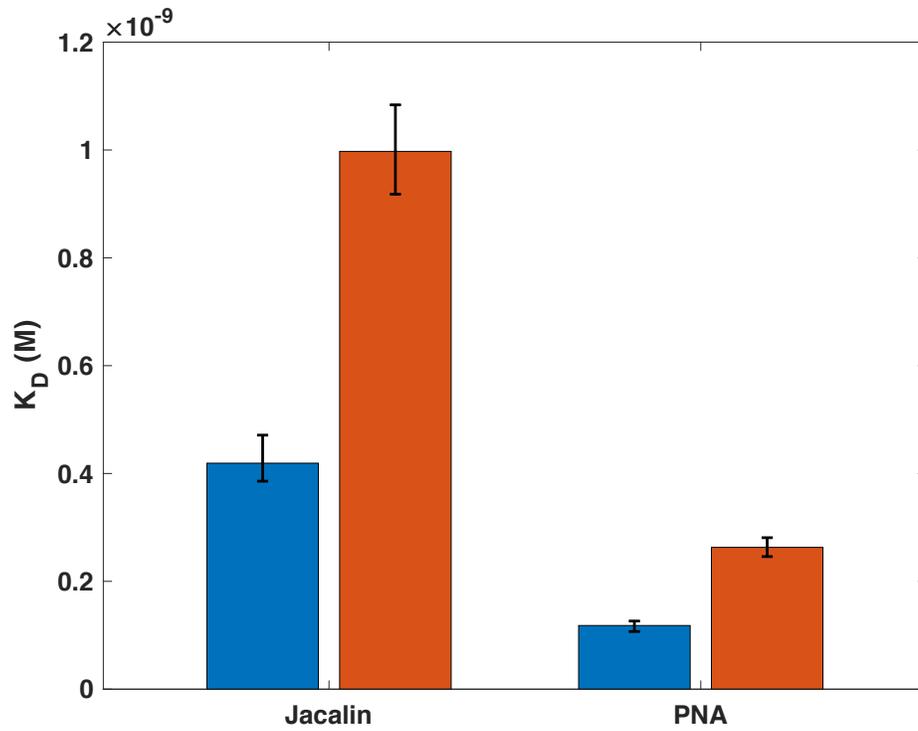
Figure 5.7: Lectin-IgA affinity constants. The two color bars represent the two affinity constants accessible to the binding interaction. The error bars are upper and lower bounds of the K_D based on the with of the peaks in the continuous description of the bright time and dark time data.
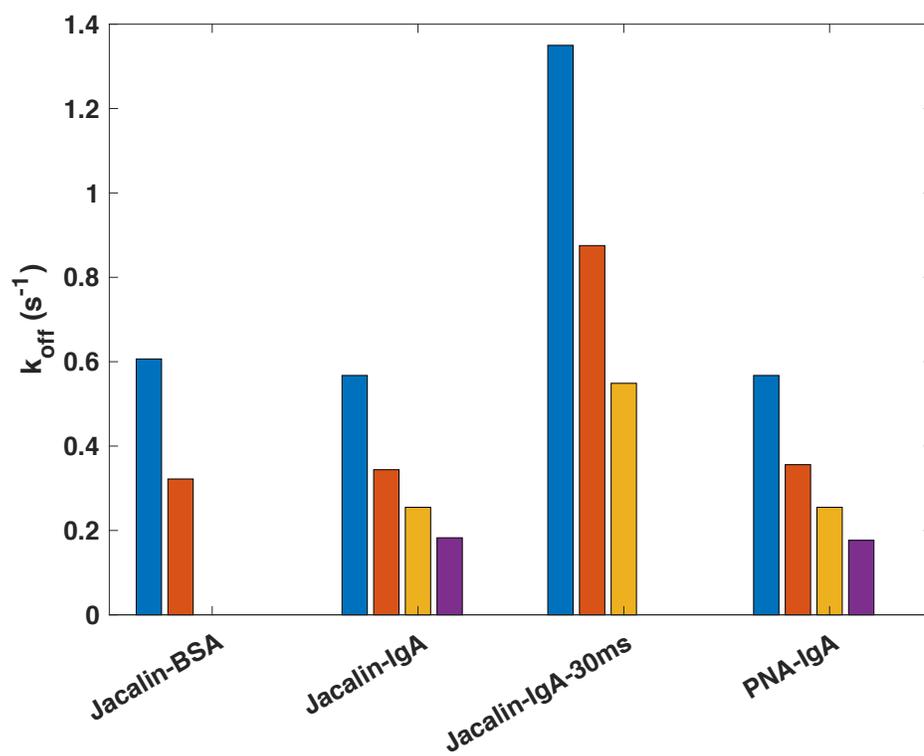
Figure 5.8:   Discrete off-rates of lectin-IgA binding kinetics

The cause for this discrepancy was likely due to one or a combination of factors. The first is seen when examining the off- and on-rates that generate the $K_{D,i}$ calculated. In Fig. 5.9, the on-rates are on the order of $1\ s^{-1}$. This is very slow compared to the on-rate determined in, for example, Loontiens' work [111] in which the on-rate was $7.1 \pm 0.2 \times 10^3\ M^{-1}s^{-1}$. The factors that contribute to the determination of the on-rate are acquisition parameters in the experimental setup and the time and distance tolerance for connecting two events together on different frames of a video in the analysis. Regarding the first factor, the on-rate cannot be faster than the inverse of time between frames, which is $0.5s^{-1}$. It is evident based on this that experimental limit has been reached, even exceeded by extrapolation of fitting. However, even with that extrapolation, some events are likely being missed by the long lower limit on time between two events. Therefore, experiments were conducted at a shorter exposure time and frame interval in an attempt to detect faster on-rates. Unfortunately, an exposure time of 30 ms, which is the limit of the equipment, and interval time of 60 ms generated a single K_D of $6.9 \pm 2 \times 10^{-10}$ M. The on-rate, shown in Fig. 5.9, was $0.2\ M^{-1}s^{-1}$, that is, even slower than before.
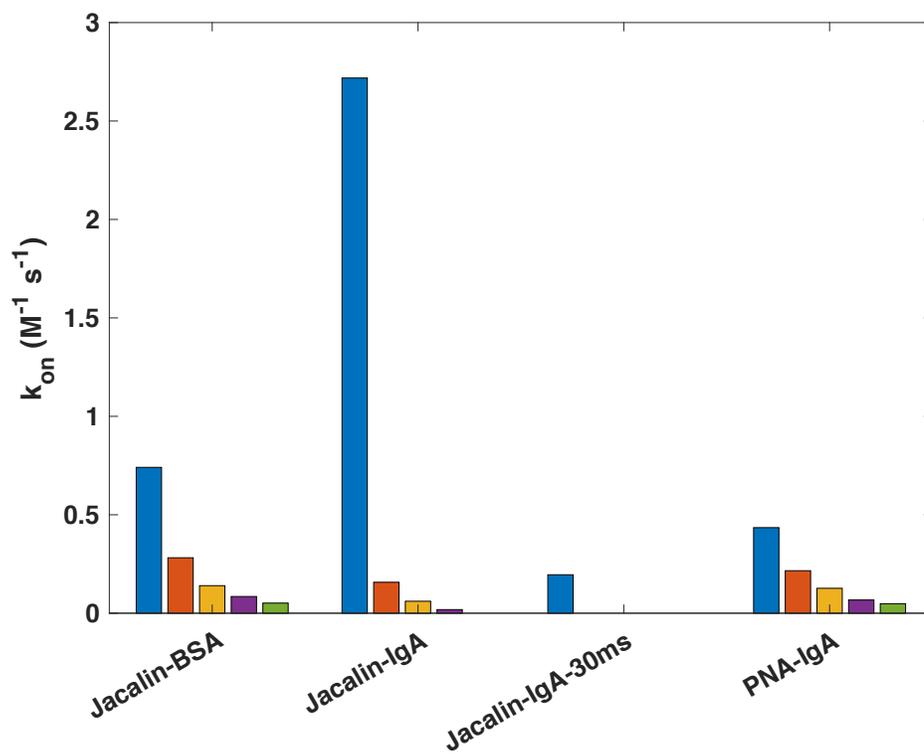
Figure 5.9: Discrete on-rates of lectin-IgA binding kinetics

Regarding the second factor, the analysis, previous inspection (described in Chapter 3) has shown that the selected parameters are the most appropriate consistent results from the analysis. Spurious events a long time from the actual landing events could have generated outliers in the calculation of an average dark time at a single location. However, after eliminating these from the average times, no change was seen in the dark time distribution from which the on-rates were calculated.

The third potential cause of the discrepancy is cross-reactivity between other components of the solution: the dye and the protein may increase affinity due to hydrophobic interactions between the dye and hydrophobic patches on the solution-facing surface of the protein. The fourth is that due to the near-monolayer arrangement of IgA1 on the surface, the lectins may potentially bind more than one IgA1 at a time. It is well-known from studies with individual oligosaccharides that the very low affinity between lectin and glycan is overcome in binding studies by increasing the number of sugars that a lectin can bind to in one location, thereby increasing the avidity between lectin and glycoprotein. Additionally, the tetrameric nature of both jacalin [103] and PNA [104] support the hypothesis that the lectins may be binding to more than one IgA1 at a time, given that they each have four identical binding pockets.

5.4   Discussion

It is widely understood that glycans have an important role in human health. Therefore significant effort continues to be invested into understanding how glycosylation affects the function of the proteome. The challenge remains to directly determine the complete glycan profile of healthy and pathogenic glycoproteins to produce biomarkers and therapeutic targets. Here, the groundwork has been laid for a aptamer-based single molecule imaging sequencing method. The kinetic parameters were determined for two lectins binding to IgA1 in single molecule imaging experiments. The parameters measured were several orders of magnitude different than those determined through other techniques. At the single molecule

level, the outcomes are highly sensitive to the experimental set-up, as any noise or extrane-ous effect such as cross-interactions will be easily detected. This result could be improved by further tuning the system to reduce the proximity and heterogeneity of IgA1 binding sites on the surface. With the prospect of directly probing the structure of the glycan cluster in place, the effort to improve the system seems a worthy endeavor.

Chapter 6

Conclusion

A single molecule imaging protocol was established for protein adsorption and kinetics measurements in a TIRF system. A new outlook was established for a classic single molecule imaging experiment, and the kinetic parameters were measured to lay the ground work for future super-resolution experiments.

A mechanistic model of protein adsorption was presented and validated against experimental data for six different conditions. By modeling the process of adsorption as a process of protein unzipping/zipping on to the surface, the heterogeneous behaviors of protein adsorption are unified under one physical model. This is opposed to, but also complementary to, the customary sum of exponential distributions used to discretely describe adsorption kinetics. The conditions that the model was tested against demonstrated the same behavior as seen in many other protein adsorption experiments, that is, the distribution of surface residence times to describe the dissociation rates could be described by a sum of exponential processes. Therefore, this model could be used in conjunction with a classic description to link the observed behaviors to a physical meaning for systems as yet untested.

The adsorption of Nile red to the solution interface of bovine serum albumin (BSA) was studied and the association and dissociation rates determined. The interaction between serum albumin and species in solution is very important to drug transport and other phys-

iological processes as it is the most common serum protein. Nile red is a very hydrophobic fluorophore that lights up when bound to the surface of BSA but not otherwise when in solution. This system therefore served two purposes - one to study the a model of interactions between serum albumin and a drug-like hydrophobic model, and two to establish the protocol and kinetic parameters that would allow for the rational design of a super-resolution experiment to study in detail the interface that a adsorbed protein presents to the solution. The kinetic parameters for this system were established and the details of the accessibility to hydrophobic binding sites were superficially probed by introducing an inhibitor. This is the first time, apparently, that this non-specific interaction has been capitalized upon to probe the location and characteristics of individual proteins.

Similar to the Nile red and BSA system, the kinetic parameters were determined for glycan binding proteins and glycoprotein immunglobulin A1. The presence of sugars on the IgA1 is crucial to its role in the immune system and leads to diseased states for some variations off of the normal profile. Many methods have been developed in recent years to study protein glycosylation, but it is still a challenge to directly profile the sugar sequence. However, establishing these parameters in a single molecule system for lectin-glycoprotein pairs with known affinity was groundwork for eventually developing a super-resolution technique to detect the presence of specific glycans. Even without the development of the super-resolution method, this work is applicable to any glycoprotein-lectin pair. If there is one lectin with established affinity for the normal glycan sequence on the glycoprotein of interest, glycoprotein from unhealthy states could be probed to detect any suspected changes in the glycan sequence. In either future system, the work done here will allow for the rational design of glycan detecting experiments.

Bibliography

[1] A. Yildiz, M. Tomishige, and R. D. Vale, "Kinesin Walks Hand-Over-Hand," *Science*, vol. 303, no. January, pp. 676–678, 2004.

[2] K. G. Suzuki, R. S. Kasai, T. K. Fujiwara, and A. Kusumi, "Single-Molecule Imaging of ReceptorReceptor Interactions," *Methods in Cell Biology*, vol. 117, pp. 373–390, 1 2013.

[3] R. Roy, S. Hohng, and T. Ha, "A practical guide to single-molecule FRET.," *Nature Methods*, vol. 5, pp. 507–16, 6 2008.

[4] R. Parthasarathy, "Rapid, accurate particle tracking by calculation of radial symmetry centers," *Nature Methods*, vol. 9, no. 7, pp. 724–726, 2012.

[5] E. M. Peterson, M. W. Manhart, and J. M. Harris, "Single-Molecule Fluorescence Imaging of Interfacial DNA Hybridization Kinetics at Selective Capture Surfaces," *Analytical Chemistry*, vol. 88, no. 2, pp. 1345–1354, 2016.

[6] Y. Wang, J. Schnitzbauer, Z. Hu, X. Li, Y. Cheng, Z.-L. Huang, and B. Huang, "Localization events-based sample drift correction for localization microscopy with redundant cross-correlation algorithm.," *Optics Express*, vol. 22, no. 13, pp. 15982–15991, 2014.

[7] M. J. Mlodzianoski, J. M. Schreiner, S. P. Callahan, K. Smolková, A. Dlasková, J. Šantorová, P. Ježek, J. Bewersdorf, and P. Je, "Sample drift correction in 3D flu-

orescence photoactivation localization microscopy," *Optics Express*, vol. 19, no. 2006, pp. 15009–15019, 2011.

[8] R. Parthasarathy, "Rapid, accurate particle tracking by calculation of radial symmetry centers," *Nature Methods*, no. July, pp. 2–7, 2012.

[9] J. P. Klein and M. L. Moeschberger, "Nonparametric Estimation of Basic Quantities for Right-Censored and Left-Truncated Data," in *Survival Analysis*, pp. 83–124, Springer-Verlag New York, 2 ed., 1997.

[10] E. L. Kaplan and P. Meier, "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457–481, 1958.

[11] I. Lynch, T. Cedervall, M. Lundqvist, C. Cabaleiro-Lago, S. Linse, and K. A. Dawson, "The nanoparticle-protein complex as a biological entity; a complex fluids and surface science challenge for the 21st century," *Advances in Colloid and Interface Science*, vol. 134-135, pp. 167–174, 2007.

[12] I. Lynch, A. Salvati, and K. A. Dawson, "Protein-nanoparticle interactions: What does the cell see?," *Nature Nanotechnology*, vol. 4, no. 9, pp. 546–547, 2009.

[13] M. P. Monopoli, F. B. Bombelli, and K. A. Dawson, "Nanobiotechnology: Nanoparticle coronas take shape," *Nature Nanotechnology*, vol. 6, no. 1, pp. 11–12, 2011.

[14] M. P. Monopoli, D. Walczyk, A. Campbell, G. Elia, I. Lynch, F. Baldelli Bombelli, and K. A. Dawson, "Physical-Chemical aspects of protein corona: Relevance to in vitro and in vivo biological impacts of nanoparticles," *Journal of the American Chemical Society*, vol. 133, no. 8, pp. 2525–2534, 2011.

[15] L. Kisley and C. F. Landes, "Molecular approaches to chromatography using single molecule spectroscopy," *Analytical Chemistry*, vol. 87, no. 1, pp. 83–98, 2015.

[16] L. Kisley, J. Chen, A. P. Mansur, S. Dominguez-Medina, E. Kulla, M. K. Kang, B. Shuang, K. Kourentzi, M. V. Poongavanam, S. Dhamane, R. C. Willson, and C. F. Landes, "High ionic strength narrows the population of sites participating in protein ion-exchange adsorption: A single-molecule study," *Journal of Chromatography A*, vol. 1343, pp. 135–142, 2014.

[17] L. Kisley, J. Chen, A. P. Mansur, B. Shuang, K. Kourentzi, M.-V. Poongavanam, W.-H. Chen, S. Dhamane, R. C. Willson, and C. F. Landes, "Unified superresolution experiments and stochastic theory provide mechanistic insight into protein ion-exchange adsorptive separations," *Proceedings of the National Academy of Sciences*, vol. 111, no. 6, pp. 2075–2080, 2014.

[18] L. Kisley, U. Patil, S. Dhamane, K. Kourentzi, L. J. Tauzin, R. C. Willson, and C. F. Landes, "Competitive multicomponent anion exchange adsorption of proteins at the single molecule level," *Analyst*, vol. 142, no. 17, pp. 3127–3131, 2017.

[19] J. N. Mabry, M. J. Skaug, and D. K. Schwartz, "Single-Molecule Insights into Retention at a Reversed-Phase Chromatographic Interface - SI," *Analytical Chemistry*, vol. 86, pp. 9451–9458, 2014.

[20] C. F. Wertz and N. P. Ryde, "NANOPARTICULATE COMPOSITIONS HAVING LYSOZYME AS A SURFACE STABILIZER," 2013.

[21] S. Devineau, L. Zargarian, J. P. Renault, and S. Pin, "Structure and Function of Adsorbed Hemoglobin on Silica Nanoparticles: Relationship between the Adsorption Process and the Oxygen Binding Properties," *Langmuir*, vol. 33, no. 13, pp. 3241–3252, 2017.

[22] B. M. C. Chan and J. L. Brash, "Adsorption of fibrinogen on glass: reversibility aspects," *Journal of Colloid And Interface Science*, vol. 82, no. 1, pp. 217–225, 1981.

[23] R. M. A. Azzam and N. M. Bashara, *Ellipsometry and polarized light.* North-Holland. sole distributors for the USA and Canada, Elsevier Science Publishing Co., Inc., 1987.

[24] W. Norde and A. C. Anusiem, "Adsorption, desorption and re-adsorption of proteins on solid surfaces," *Colloids and Surfaces*, vol. 66, no. 1, pp. 73–80, 1992.

[25] W. Norde and J. P. Favier, "Structure of adsorbed and desorbed proteins," *Colloids and Surfaces*, vol. 64, no. 1, pp. 87–93, 1992.

[26] L. Vroman and A. Lukosevicius, "Ellipsometer Recordings of Changes in Optical Thickness of Adsorbed Films associated with Surface Activation of Blood Clotting," *Nature*, vol. 204, no. 4959, pp. 701–703, 1964.

[27] A. Rothen, "The Ellipsometer, an Apparatus to Measure Thickness of Thin Surface Films," *The Review of Scientific Instruments*, vol. 16, no. 2, pp. 26–30, 1945.

[28] M. Mrksich, G. B. Sigal, and G. M. Whitesides, "Surface Plasmon Resonance Permits in Situ Measurement of Protein Adsorption on Self-Assembled Monolayers of Alkanethiolates on Gold," *Langmuir*, vol. 11, no. 11, pp. 4383–4385, 1995.

[29] C. Kößlinger, E. Uttenthaler, S. Drost, F. Aberl, H. Wolf, G. Brink, A. Stanglmaier, and E. Sackmann, "Comparison of the QCM and the SPR method for surface studies and immunological applications," *Sensors and Actuators: B. Chemical*, vol. 24, no. 1-3, pp. 107–112, 1995.

[30] F. Caruso, D. Furlong, and P. Kingshott, "Characterization of Ferritin Adsorption onto Gold," *Journal of Colloid and Interface Science*, vol. 186, no. 1, pp. 129–140, 1997.

[31] F. Höök, J. Vörös, M. Rodahl, R. Kurrat, P. Böni, J. Ramsden, M. Textor, N. Spencer, P. Tengvall, J. Gold, and B. Kasemo, "A comparative study of protein adsorption on titanium oxide surfaces using in situ ellipsometry, optical waveguide lightmode

spectroscopy, and quartz crystal microbalance/dissipation," *Colloids and Surfaces B: Biointerfaces*, vol. 24, no. 2, pp. 155–170, 2002.

[32] F. Höök, M. Rodahl, B. Kasemo, and P. Brzezinski, "Structural changes in hemoglobin during adsorption to solid surfaces: Effects of pH, ionic strength, and ligand binding," *Proceedings of the National Academy of Sciences*, vol. 95, no. 21, pp. 12271–12276, 1998.

[33] M. Thompson, C. L. Arthur, and G. K. Dhaliwal, "Liquid-Phase Piezoelectric and Acoustic Transmission Studies of Interfacial Immunochemistry," *Analytical Chemistry*, vol. 58, no. 6, pp. 1206–1209, 1986.

[34] S. Devineau, J. M. Zanotti, C. Loupiac, L. Zargarian, F. Neiers, S. Pin, and J. P. Renault, "Myoglobin on silica: A case study of the impact of adsorption on protein structure and dynamics," *Langmuir*, vol. 29, no. 44, pp. 13465–13472, 2013.

[35] L. Shen and X.-Y. Zhu, "Evidence of a Mobile Precursor State in Nonspecific Protein Adsorption," *Langmuir*, vol. 27, no. 11, pp. 7059–7064, 2011.

[36] S. Dominguez-Medina, L. Kisley, L. J. Tauzin, A. Hoggard, B. Shuang, A. S. D. S. Indrasekara, S. Chen, L. Y. Wang, P. J. Derry, A. Liopo, E. R. Zubarev, C. F. Landes, and S. Link, "Adsorption and Unfolding of a Single Protein Triggers Nanoparticle Aggregation," *ACS Nano*, vol. 10, no. 2, pp. 2103–2112, 2016.

[37] L. Shen and J. Zhu, "Heterogeneous surfaces to repel proteins," *Advances in Colloid and Interface Science*, vol. 228, pp. 40–54, 2016.

[38] M. Kastantin, B. B. Langdon, E. L. Chang, and D. K. Schwartz, "Single-molecule resolution of interfacial fibrinogen behavior: Effects of oligomer populations and surface chemistry," *Journal of the American Chemical Society*, vol. 133, pp. 4975–4983, 2011.

[39] J. N. Mabry, M. Kastantin, and D. K. Schwartz, "Capturing Conformation-Dependent Molecule-Surface Interactions When Surface Chemistry Is Heterogeneous," *ACS Nano*, vol. 9, no. 7, pp. 7237–7247, 2015.

[40] J. S. Weltz, D. K. Schwartz, and J. L. Kaar, "Surface-mediated protein unfolding as a search process for denaturing sites," *ACS Nano*, vol. 10, no. 1, pp. 730–738, 2016.

[41] B. B. Langdon, M. Kastantin, and D. K. Schwartz, "Apparent activation energies associated with protein dynamics on hydrophobic and hydrophilic surfaces," *Biophysical Journal*, vol. 102, no. 11, pp. 2625–2633, 2012.

[42] B. B. Mandelbrot, "Stochastic volatility, power laws and long memory," *Quantitative Finance*, vol. 1, pp. 558–559, 6 2001.

[43] T. Bochud and D. Challet, "Optimal approximations of power laws with exponentials: application to volatility models with long memory," *Quantitative Finance*, vol. 7, pp. 585–589, 12 2007.

[44] B. D. Ripley, "Selecting Amongst Large Classes of Models," *Methods and Models in Statistics*, pp. 155–170, 2004.

[45] K. P. Burnham and D. R. Anderson, "Multimodel Inference: Understanding AIC and BIC in Model Selection," *Sociological Methods & Research*, vol. 33, pp. 261–304, 11 2004.

[46] M. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005.

[47] F. Cichos, C. von Borczyskowski, and M. Orrit, "Power-law intermittency of single emitters," *Current Opinion in Colloid and Interface Science*, vol. 12, no. 6, pp. 272–284, 2007.

[48] M. J. Penna, M. Mijajlovic, and M. J. Biggs, "Molecular-Level Understanding of Protein Adsorption at the Interface between Water and a Strongly Interacting Uncharged Solid Surface," *Journal of the American Chemical Society*, vol. 136, no. 14, pp. 5323–5331, 2014.

[49] M. Kastantin and D. K. Schwartz, "Distinguishing Positional Uncertainty from True Mobility in Single-Molecule Trajectories That Exhibit Multiple Diffusive Modes," *Microscopy and microanalysis : the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*, vol. 18, pp. 793–797, 2012.

[50] M. Kastantin, R. Walder, and D. K. Schwartz, "Identifying mechanisms of interfacial dynamics using single-molecule tracking," *Langmuir*, vol. 28, no. 34, pp. 12443–12456, 2012.

[51] M. Kastantin and D. K. Schwartz, "Identifying multiple populations from single-molecule lifetime distributions.," *Chemphyschem : a European journal of chemical physics and physical chemistry*, vol. 14, pp. 374–380, 2 2013.

[52] M. Kastantin, B. B. Langdon, and D. K. Schwartz, "A bottom-up approach to understanding protein layer formation at solid-liquid interfaces," *Advances in Colloid and Interface Science*, vol. 207, no. 1, pp. 240–252, 2014.

[53] M. Kastantin, B. B. Langdon, E. L. Chang, and D. K. Schwartz, "Single-molecule resolution of interfacial fibrinogen behavior: Effects of oligomer populations and surface chemistry," *Journal of the American Chemical Society*, vol. 133, no. 13, pp. 4975–4983, 2011.

[54] P. J. Steinbach, R. Ionescu, and C. R. Matthews, "Analysis of kinetics using a hybrid maximum-entropy/nonlinear-least-squares method: application to protein folding.," *Biophysical Journal*, vol. 82, no. 4, pp. 2244–2255, 2002.

[55] C. a. Ross and M. a. Poirier, "Protein aggregation and neurodegenerative disease.," *Nature Medicine*, vol. 10 Suppl, no. July, pp. S10–S17, 2004.

[56] B. D. Ratner, A. S. Hoffman, F. J. Schoen, and J. E. Lemons, *Biomaterials Science - An Introduction to Materials in Medicine.* Elsevier, 3rd ed., 2013.

[57] D. G. Castner and B. D. Ratner, "Biomedical surface science: Foundations to frontiers," *Surface Science*, vol. 500, pp. 28–60, 3 2002.

[58] E. Y. Chi, S. Krishnan, T. W. Randolph, and J. F. Carpenter, "Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation," *Pharmaceutical Research*, vol. 20, no. 9, pp. 1325–1336, 2003.

[59] A. Balasubramanian, V. Durairajpandian, S. Elumalai, N. Mathivanan, A. K. Munirajan, and K. Ponnuraj, "Structural and functional studies on urease from pigeon pea (Cajanus cajan)," *International Journal of Biological Macromolecules*, vol. 58, no. April, pp. 301–309, 2013.

[60] Y. F. Yano, "Kinetics of protein unfolding at interfaces," *Journal of Physics Condensed Matter*, vol. 24, no. 50, 2012.

[61] C. F. Wertz and M. M. Santore, "Adsorption and reorientation kinetics of lysozyme on hydrophobic surfaces," *Langmuir*, vol. 18, no. 4, pp. 1190–1199, 2002.

[62] C. F. Wertz and M. M. Santore, "Fibrinogen adsorption on hydrophilic and hydrophobic surfaces: Geometrical and energetic aspects of interfacial relaxations," *Langmuir*, vol. 18, no. 3, pp. 706–715, 2002.

[63] M. M. Santore and C. F. Wertz, "Protein spreading kinetics at liquid-solid interfaces via an adsorption probe method," *Langmuir*, vol. 21, no. 22, pp. 10172–10178, 2005.

[64] C. F. Wertz and M. M. Santore, "Adsorption and relaxation kinetics of albumin and fibrinogen on hydrophobic surfaces: Single-species and competitive behavior," *Langmuir*, vol. 15, no. 26, pp. 8884–8894, 1999.

[65] W. Norde and C. A. Haynes, "Reversibility and the Mechanism of Protein Adsorption," in *Proteins at Interfaces II* (T. Horbett and J. L. Brash, eds.), (Washington, D.C.), pp. 26–40, American Chemical Society, 1995.

[66] R. Lencki, J. Arul, and R. Neufeld, "Effect of Subunit Dissociation, Denaturation, Aggregation, Coagulation, and Decomposition on Enzyme Inactivation Kinetics: II. Biphasic and Grace Period Behavior," *Biotechnology and Bioengineering*, vol. 40, no. 11, pp. 1427–1434, 1992.

[67] S.-Y. Seong and P. Matzinger, "Opinion: Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses," *Nature Reviews Immunology*, vol. 4, no. 6, pp. 469–478, 2004.

[68] E. Herczenik and M. F. B. G. Gebbink, "Molecular and cellular aspects of protein misfolding and disease," *The FASEB Journal*, vol. 22, no. 7, pp. 2115–2133, 2008.

[69] C. Mathé, S. Devineau, J. C. Aude, G. Lagniel, S. Chédin, V. Legros, M. H. Mathon, J. P. Renault, S. Pin, Y. Boulard, and J. Labarre, "Structural determinants for protein adsorption/non-adsorption to silica surface," *PLoS ONE*, vol. 8, no. 11, pp. 1–13, 2013.

[70] P. Pereira, S. M. Kelly, P. R. Gellert, and C. F. van der Walle, "Interdomain mobility and conformational stability of type III fibronectin domain pairs control surface adsorption, desorption and unfolding," *Colloids and Surfaces B: Biointerfaces*, vol. 64, no. 1, pp. 1–9, 2008.

[71] P. R. Burney, E. M. Nordwald, K. Hickman, J. L. Kaar, and J. Pfaendtner, "Molecular dynamics investigation of the ionic liquid/enzyme interface: Application to engineering

enzyme surface charge," *Proteins: Structure, Function and Bioinformatics*, vol. 83, no. 4, pp. 670–680, 2015.

[72] W. M. De Vos, P. M. Biesheuvel, A. De Keizer, J. M. Kleijn, and M. A. Stuart, "Adsorption of the protein bovine serum albumin in a planar poly(acrylic acid) brush layer as measured by optical reflectometry," *Langmuir*, vol. 24, no. 13, pp. 6575–6584, 2008.

[73] F. Turci, E. Ghibaudi, M. Colonna, B. Boscolo, I. Fenoglio, and B. Fubini, "An integrated approach to the study of the interaction between proteins and nanoparticles," *Langmuir*, vol. 26, no. 11, pp. 8336–8346, 2010.

[74] A. Marucco, F. Turci, L. O'Neill, H. J. Byrne, B. Fubini, and I. Fenoglio, "Hydroxyl density affects the interaction of fibrinogen with silica nanoparticles at physiological concentration," *Journal of Colloid and Interface Science*, vol. 419, pp. 86–94, 2014.

[75] J. Schnitzbauer, M. T. Strauss, T. Schlichthaerle, F. Schueder, and R. Jungmann, "Super-resolution microscopy with DNA-PAINT," *Nature Publishing Group*, vol. 12, no. 6, pp. 1198–1228, 2017.

[76] A. Sharonov and R. M. Hochstrasser, "Wide-field subdiffraction imaging by accumulated binding of diffusing probes.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 50, pp. 18911–18916, 2006.

[77] F. Schueder, M. T. Strauss, D. Hoerl, J. Schnitzbauer, T. Schlichthaerle, S. Strauss, P. Yin, H. Harz, H. Leonhardt, and R. Jungmann, "Super-Resolution Microscopy Universal Super-Resolution Multiplexing by DNA Exchange," *Angewandte Chemie International Edition*, vol. 56, pp. 4052–4055, 2017.

[78] J. Chen, A. Bremauntz, L. Kisley, B. Shuang, and C. F. Landes, "Super-resolution mbPAINT for optical localization of single-stranded DNA," *ACS Applied Materials and Interfaces*, vol. 5, no. 19, pp. 9338–9343, 2013.

[79] S. W. Hell, S. J. Sahl, M. Bates, X. Zhuang, R. Heintzmann, M. J. Booth, J. Bewersdorf, G. Shtengel, H. Hess, P. Tinnefeld, A. Honigmann, S. Jakobs, I. Testa, L. Cognet, B. Lounis, H. Ewers, S. J. Davis, C. Eggeling, D. Klenerman, K. I. Willig, G. Vicidomini, M. Castello, A. Diaspro, and T. Cordes, "The 2015 super-resolution microscopy roadmap," *Journal of Physics D: Applied Physics*, vol. 48, no. 44, p. 443001, 2015.

[80] C. Reichardt, *Solvent Effects on the Absorption Spectra of Organic Compounds*. Weinheim: WILEY-VCH verlag GmbH & Co. KGaA, 3rd ed., 2003.

[81] A. Sułkowska, B. Bojko, J. Równicka, and W. W. Sułkowski, "Paracetamol and cytarabine binding competition in high affinity binding sites of transporting protein," *Journal of Molecular Structure*, vol. 792-793, pp. 249–256, 2006.

[82] P. Greenspan and S. D. Fowler, "Spectrofluorometric studies of the lipid probe, nile red," *Journal of Lipid Research*, vol. 26, pp. 781–789, 1985.

[83] J. F. Deye, T. A. Berger, and A. G. Anderson, "Nile Red as a Solvatochromic Dye for Measuring Solvent Strength in Normal Liquids and Mixtures of Normal Liquids with Supercritical and Near Critical Fluids," *Analytical Chemistry*, vol. 62, no. 6, pp. 615–622, 1990.

[84] K. A. Majorek, P. J. Porebski, A. Dayal, M. D. Zimmerman, K. Jablonska, A. J. Stewart, M. Chruszcz, and W. Minor, "Structural and immunologic characterization of bovine, horse, and rabbit serum albumins," *Molecular Immunology*, vol. 52, pp. 174–182, 10 2012.

[85] J. Hirabayashi, T. Hashidate, Y. Arata, N. Nishi, T. Nakamura, M. Hirashima, T. Urashima, T. Oka, M. Futai, W. E. G. Muller, F. Yagi, and K.-I. Kasai, "Oligosaccharide specificity of galectins: a search by frontal affinity chromatography," *Biochimica et Biophysica Acta*, vol. 1572, pp. 232–254, 2002.

[86] H. Iwase, A. Tanaka, Y. Hiki, T. Kokubo, T. Sano, I. Ishii-Karakasa, K. Hisatani, Y. Kobayashi, and K. Hotta, "Analysis of the microheterogeneity of the IgA1 hinge glycopeptide having multiple O-linked oligosaccharides by capillary electrophoresis," *Analytical Biochemistry*, vol. 288, no. 1, pp. 22–27, 2001.

[87] D. Reusch, M. Haberger, B. Maier, M. Maier, R. Kloseck, B. Zimmermann, M. Hook, Z. Szabo, S. Tep, J. Wegstein, N. Alt, P. Bulau, and M. Wuhrer, "Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 1: Separation-based methods Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 1: Separation-based methods," *mAbs*, vol. 7, no. 1, pp. 167–179, 2014.

[88] D. Reusch, M. Haberger, D. Falck, B. Peter, B. Maier, J. Gassner, M. Hook, K. Wagner, L. Bonnington, P. Bulau, and M. Wuhrer, "Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 2: Mass spectrometric methods Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 2: Mass spectrometric methods," *mAbs*, vol. 7, no. 4, pp. 732–742, 2015.

[89] W. R. Alley, B. F. Mann, and M. V. Novotny, "High-sensitivity Analytical Approaches for the Structural Characterization of Glycoproteins," *Chemical Reviews*, vol. 115, pp. 2668–2732, 2013.

[90] J. Ma and G. W. Hart, "O-GlcNAc profiling: from proteins to proteomes," *Clinical Proteomics*, vol. 11, no. 8, pp. 1–16, 2014.

[91] K. Takahashi, S. B. Wall, H. Suzuki, A. D. Smith, S. Hall, K. Poulsen, M. Kilian, J. A. Mobley, B. A. Julian, J. Mestecky, and J. Novak, "Clustered O-glycans of IgA1: Defining macro-and micro-heterogeneity by use of electron capture/transfer dissociation.," *Molecular & Cellular Proteomics*, p. mcp.M110.001834, 2010.

[92] M. B. Renfrow, C. L. Mackay, M. J. Chalmers, B. A. Julian, J. Mestecky, M. Kilian, K. Poulsen, M. R. Emmett, A. G. Marshall, J. Novak, M. B. Renfrow, C. L. Mackay, M. J. Chalmers, M. R. Emmett, A. G. Marshall, B. A. Julian, J. Mestecky, J. Novak, M. Kilian, and . K. Poulsen, "Analysis of O-glycan heterogeneity in IgA1 myeloma proteins by Fourier transform ion cyclotron resonance mass spectrometry: implications for IgA nephropathy," *Analytical and Bioanalytical Chemistry*, vol. 389, pp. 1397–1407, 2007.

[93] T. Feizi and W. Chai, "Oligosaccharide microarrays to decipher the glyco code," *Nature Reviews Molecular Cell Biology*, vol. 5, pp. 582–588, 7 2004.

[94] T. Feizi, F. Fazio, W. Chai, and C.-H. Wong, "Carbohydrate microarrays-a new set of technologies at the frontiers of glycomics," *Current Opinion in Structural Biology*, vol. 13, pp. 637–645, 2003.

[95] O. Oyelaran, Q. Li, D. Farnsworth, and J. C. Gildersleeve, "Microarrays with Varying Carbohydrate Density Reveal Distinct Subpopulations of Serum Antibodies," *Journal of Proteome Research*, vol. 8, no. 7, pp. 3529–3538, 2009.

[96] J. Hirabayashi, M. Yamada, A. Kuno, and H. Tateno, "Lectin microarrays: concept, principle and applications," *Chemical Society Reviews*, vol. 42, p. 4443, 2013.

[97] A. Kuno, N. Uchiyama, S. Koseki-Kuno, Y. Ebe, S. Takashima, M. Yamada, and J. Hirabayashi, "Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling," *Nature Methods*, vol. 2, pp. 851–856, 11 2005.

[98] O. Blixt, S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C. Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D. J. Stevens, J. J. Skehel, I. van Die, D. R. Burton, I. A. Wilson, R. Cummings, N. Bovin, C.-H. Wong, and J. C. Paulson, "Printed covalent glycan array for ligand profiling of diverse glycan binding proteins.,"

Proceedings of the National Academy of Sciences of the United States of America, vol. 101, pp. 17033–8, 12 2004.

[99] X. Chen, Y. Wang, Y. Zhang, Z. Chen, Y. Liu, Z. Li, and J. Li, "Sensitive Electrochemical Aptamer Biosensor for Dynamic Cell Surface N-Glycan Evaluation Featuring Multivalent Recognition and Signal Amplification on a DendrimerGraphene Electrode Interface," *Anal. Chem*, vol. 86, p. 18, 2014.

[100] Y. He, J. Li, and Y. Liu, "Reusable and Dual-Potential Responses Electrogenerated Chemiluminescence Biosensor for Synchronously Cytosensing and Dynamic Cell Surface N-Glycan Evaluation," *Analytical Chemistry*, vol. 87, pp. 9777–9785, 2015.

[101] R. J. Wyatt and B. A. Julian, "IgA Nephropathy," *New England Journal of Medicine*, vol. 368, pp. 2402–2414, 2013.

[102] H. Suzuki, K. Kiryluk, J. Novak, Z. Moldoveanu, A. B. Herr, M. B. Renfrow, R. J. Wyatt, F. Scolari, J. Mestecky, A. G. Gharavi, and B. A. Julian, "The Pathophysiology of IgA Nephropathy," *Journal of the American Society of Nephrology*, vol. 22, pp. 1795–1803, 2011.

[103] Y. Bourne, C. Houle, R. Zamboni, W. J. Peumans, L. Menu-bouaouiche, E. J. M. Van Damme, A. Barre, and P. Rouge, "Structural basis for the unusual carbohydrate-binding specificity of jacalin towards galactose and mannose," *Biochemical Journal*, vol. 364, pp. 173–180, 2002.

[104] R. Banerjee, S. C. Mande, V. Ganesh, K. Das, V. Dhanaraj, S. K. Mahanta, K. Suguna, A. Surolia, and M. Vuiayan, "Crystal structure of peanut lectin, a protein with an unusual quaternary structure," *Biochemistry*, vol. 91, pp. 227–231, 1994.

[105] P. Katira, A. Agarwal, and H. Hess, "A random sequential adsorption model for protein adsorption to surfaces functionalized with poly(ethylene oxide)," *Advanced Materials*, vol. 21, pp. 1599–1604, 4 2009.

[106] S. D. Chandradoss, A. C. Haagsma, Y. Kwang Lee, J.-H. Hwang, J.-M. Nam, C. Joo, and C. Surface, "Surface Passivation for Single-molecule Protein Studies," *Journal of Visual Experiments*, vol. 1, no. 86, p. 50549, 2014.

[107] G. T. Hermanson, "Zero-Length Crosslinkers," *Bioconjugate Techniques*, pp. 259–273, 2013.

[108] P. J. Steinbach, R. Ionescu, and C. R. Matthews, "Analysis of kinetics using a hybrid maximum-entropy/nonlinear-least-squares method: application to protein folding.," *Biophysical Journal*, vol. 82, pp. 2244–55, 4 2002.

[109] M. M. Pedroso, N. C. Pesquero, S. M. Thomaz, M. C. Roque-Barreira, R. C. Faria, and P. R. Bueno, "Jacalin interaction with human immunoglobulin A1 and bovine immunoglobulin G1: Affinity constant determined by piezoelectric biosensoring," *Glycobiology*, vol. 22, no. 3, pp. 326–331, 2012.

[110] K. J. Neurohr, D. R. Bundle, N. M. Young, and H. H. Mantsch, "Binding of Disaccharides by Peanut Agglutinin as Studied by Ultraviolet Difference Spectroscopy," *European Journal of Biochemistry*, vol. 123, no. 2, pp. 305–310, 1982.

[111] F. G. Loontiens, "Slow dissociation of 4-methylumbelliferyl β-D-galactosyl-(13)-N-acetyl β-D-galactosaminide from its complex with peanut agglutinin," *FEBS Letters*, vol. 162, pp. 193–196, 10 1983.

Chapter 7

Appendix

.1   Drift Correlation Function

```matlab
function [iq_full, jq_full, Ccorr, Ccorrmat] = driftcorr5(s_bin,t_bin,pixel_size,Nframes,C,er,et,↙
int,B,F)
% Drift Correction 5 - Correlation Matrix Shift with Rolling Window
% (MA-coded dark event bridging)

% INPUT
%   S_BIN: Spatial bin size (nm)
%   T_BIN: Temporal bin size (even number of frames)
%   PIXEL_SIZE: Width of pixel (nm)
%   TRACKS: Tracks and frame numbers (output of collate_woll)
%   NFRAMES: Number of frames in stack.
%   C: Output of COLLATE that gives the trajectories in pixel-based
%   coordinates
%   ER: Maximum allowable calculated drift. Used to remove outliers. Set to
%   infinity to not use this feature.
%   ET: Exposure time, i.e. the time between the start of each exposure.
%   INT: Interval between the middle frame in a time bin. Drift will be
%   interpolated through the intervals.
%   B: Cell array of brightnesses of trajectories.
%   F: Cell array of frames of trajectories.

%% Capture majority of drift over all frames

% Check input t_bin
if rem(t_bin,2) ~= 0
    disp('Time window must be even')
    iq_full = []; jq_full = []; Ccorr = []; Ccorrmat = [];
    return
end

% Use Frame to collect the coordinates of the particles on the frames in
% the interval of interest, want to minimize the for loops
m = 512;
slices = Nframes - t_bin;
corrmax = zeros(slices-1,2);
s_bin = s_bin/pixel_size;

% Keep long trajectories spanning across consecutive slices only
K = C(cellfun(@(x)size(x,2),C)>=t_bin,:);
FF = F(cellfun(@(x)size(x,2),C)>=t_bin,:);
FF = cellfun(@(x)x',FF,'UniformOutput',false);
K = cellfun(@(x)x',K,'UniformOutput',false);


Xedges = s_bin/2:s_bin:m+s_bin/2;
Yedges = Xedges;
% N = zeros(length(Xedges)-1,length(Yedges)-1,slices);
frames_used = zeros(length(t_bin/2:Nframes-t_bin/2),1);
partperint = frames_used;
h = waitbar(0,'Calculating drift path...');
for k = t_bin/2:int:Nframes-t_bin/2
    startframe = k - t_bin/2 + 1;
    endframe = k + t_bin/2;

    % Find the tracks that span the interval
    touse = false(length(K),1);
    x = cell(length(touse),1);
    y = cell(length(touse),1);
    for l = 1:length(touse)
        if sum(ismember([startframe,endframe],FF{l})) == 2
            touse(l) = true;
        end
        interval = find(FF{l}==startframe):find(FF{l}==endframe);
        x{l} = K{l}(interval,1);
        y{l} = K{l}(interval,2);
    end
```

Figure 1:   Drift correction code, page 1.

```matlab
        if sum(touse) == 0
            continue
        else
            frames_used(k) = k;
        end
%      spanning_tracks = K(touse,:);


        % Spatially bin the tracks that span the interval
        X = cat(1,x{:});
        Y = cat(1,y{:});
        if k == t_bin/2 || k == sum(frames_used)
            N = histcounts2(X,Y,Xedges,Yedges);
            partperint(k) = sum(sum(N));
            continue
        else
            Nprev = N;
            N = histcounts2(X,Y,Xedges,Yedges);
            partperint(k) = sum(sum(N));
        end


        % Dilate N for better SNR
        % Ndil(:,:,k) = imdilate(N(:,:,k),strel('diamond',1));

        % Perform cross-correlation between each of the projections
        n = size(N,1);
        % tmp = xcorr2(N1,N);
        tmp = xcorr2(Nprev,N);
%      O = size(tmp,1);
%      gaussfit = fit([1:size(tmp,1),1:size(tmp,2)],tmp,'gauss2');
        D = tmp(n-9:n+10,n-9:n+10);
%      D = tmp(n/2+1:n+n/2,n/2+1:n+n/2);
        % Get location of maximum correlation
%      x = (1:size(D,1))';
%      y = (1:size(D,2))';
%      [xg,yg] = meshgrid(x,y);
%      [xout,yout,zout] = prepareSurfaceData(x,y,D);
%      gaussfit = fit([xg,yg],D,'gauss2');
        [f,~] = fit2D(D,pixel_size);
        corrmax(k-1,:) = [f.x0,f.y0]/pixel_size;
%          % Find first maximum in the image
%          corrmax(k-1) = find(D == max(max(D,[],1),[],2),1);

        waitbar(k/length(t_bin/2:Nframes-t_bin/2))
end
frames_used = frames_used(frames_used ~= 0);
corrmax = corrmax(corrmax(:,1) ~= 0,:);
if isempty(corrmax)
    iq_full = []; jq_full = []; Ccorr = []; Ccorrmat = [];
    disp('No trajectories span time interval. Choose smaller interval')
    close(h)
    return
end
close(h)
% clear tmp D

% Find the difference and interpolate difference per frame
% [I,J] = ind2sub([n n],corrmax);
I = corrmax(:,1); J = corrmax(:,2);
dI = I - size(D,1)/2; dJ = J - size(D,1)/2;
% dI = I-n/2; dJ = J-n/2;
dI(abs(dI) >= er) = 0; dJ(abs(dJ) >= er) = 0;
Ivect = cumsum(dI); Jvect = cumsum(dJ);

% See how things are going
```

Figure 2:   Drift correction code, page 2.

109

```matlab
% figure,plot(dI),hold on,plot(dJ),hold off
% figure,plot(Ivect),hold on,plot(Jvect),hold off

ints = et*frames_used(1:end-1);
% % ints = round(linspace(et*T,et*Nframes,slices-1));
% timeVec = t_bin/2*et:et:(Nframes-t_bin/2)*et;
% iq = interp1(ints,Ivect,timeVec,'makima');
% jq = interp1(ints,Jvect,timeVec,'makima');
% iq_full = padarray(iq,[0 t_bin/2-1],iq(1),'pre');
% iq_full = padarray(iq_full,[0 t_bin/2],iq(end),'post');
% jq_full = padarray(jq,[0 t_bin/2-1],jq(1),'pre');
% jq_full = padarray(jq_full,[0 t_bin/2],jq(end),'post');

%% Capture the drift on earliest frames
NframesE = t_bin/2;
t_binE = round(NframesE/10);
if rem(t_binE,2) ~= 0
    t_binE = t_binE + 1;
end
slices = NframesE - t_binE;
corrmax = zeros(slices-1,2);

% Keep long trajectories spanning across consecutive slices only
K = C(cellfun(@(x)size(x,2),C)>=t_binE,:);
FF = F(cellfun(@(x)size(x,2),C)>=t_binE,:);
FF = cellfun(@(x)x',FF,'UniformOutput',false);
K = cellfun(@(x)x',K,'UniformOutput',false);


Xedges = s_bin/2:s_bin:m+s_bin/2;
Yedges = Xedges;
% N = zeros(length(Xedges)-1,length(Yedges)-1,slices);
frames_used = zeros(length(t_binE/2:NframesE-t_binE/2),1);
partperint = frames_used;
h = waitbar(0,'Calculating drift path...');
for k = t_binE/2:int:NframesE-t_binE/2
    startframe = k - t_binE/2 + 1;
    endframe = k + t_binE/2;

    % Find the tracks that span the interval
    touse = false(length(K),1);
    x = cell(length(touse),1);
    y = cell(length(touse),1);
    for l = 1:length(touse)
        if sum(ismember([startframe,endframe],FF{l})) == 2
            touse(l) = true;
            interval = find(FF{l}==startframe):find(FF{l}==endframe);
            x{l} = K{l}(interval,1);
            y{l} = K{l}(interval,2);
        end

    end
    if sum(touse) == 0
        continue
    else
        frames_used(k) = k;
    end
%     spanning_tracks = K(touse,:);


    % Spatially bin the tracks that span the interval
    X = cat(1,x{:});
    Y = cat(1,y{:});
    if k == t_binE/2
        N = histcounts2(X,Y,Xedges,Yedges);
        partperint(k) = sum(sum(N));
```

Figure 3:   Drift correction code, page 3.

```matlab
            continue
        else
            Nprev = N;
            N = histcounts2(X,Y,Xedges,Yedges);
            partperint(k) = sum(sum(N));
        end


        % Dilate N for better SNR
        % Ndil(:,:,k) = imdilate(N(:,:,k),strel('diamond',1));

        % Perform cross-correlation between each of the projections
        n = size(N,1);
        % tmp = xcorr2(N1,N);
        tmp = xcorr2(Nprev,N);
%       O = size(tmp,1);
%       gaussfit = fit([1:size(tmp,1),1:size(tmp,2)],tmp,'gauss2');
        D = tmp(n-9:n+10,n-9:n+10);
%       D = tmp(n/2+1:n+n/2,n/2+1:n+n/2);
        % Get location of maximum correlation
%       x = (1:size(D,1))';
%       y = (1:size(D,2))';
%       [xg,yg] = meshgrid(x,y);
%       [xout,yout,zout] = prepareSurfaceData(x,y,D);
%       gaussfit = fit([xg,yg],D,'gauss2');
        [f,~] = fit2D(D,pixel_size);
        corrmax(k-1,:) = [f.x0,f.y0]/pixel_size;
%       % Find first maximum in the image
%       corrmax(k-1) = find(D == max(max(D,[],1),[],2),1);

        waitbar(k/length(t_bin/2:Nframes-t_bin/2))
end
frames_used = frames_used(frames_used ~= 0);
corrmax = corrmax(corrmax(:,1) ~= 0,:);
if isempty(corrmax)
%       iq_full = []; jq_full = []; Ccorr = []; Ccorrmat = [];
    disp('No discernable drift in early frames')
    ints_early = [];
    IvectE = [];
    JvectE = [];
    earlyused = false;
else
    % Find the difference and interpolate difference per frame
    I = corrmax(:,1); J = corrmax(:,2);
    dIE = I - size(D,1)/2; dJE = J - size(D,1)/2;
    dIE(abs(dIE) >= er) = 0; dJE(abs(dJE) >= er) = 0;
    IvectE = cumsum(dIE); JvectE = cumsum(dJE);
    ints_early = et*frames_used(1:end-1);
    earlyused = true;
end
close(h)
% clear tmp D


%% Interpolate entire drift path
if earlyused
    tV = et:et:(Nframes-t_bin/2)*et;
    iq = interp1([ints_early;ints],[IvectE;Ivect+IvectE(end)],tV,'makima');
    iq_full = padarray(iq,[0 t_bin/2],iq(end),'post');
    tV = et:et:(Nframes-t_bin/2)*et;
    jq = interp1([ints_early;ints],[JvectE;Jvect+JvectE(end)],tV,'makima');
    jq_full = padarray(jq,[0 t_bin/2],jq(end),'post');
else
    tV = et:et:(Nframes-t_bin/2)*et;
    iq = interp1(ints,Ivect,tV,'makima');
```

Figure 4:   Drift correction code, page 4.

```matlab
    iq_full = padarray(iq,[0 t_bin/2],iq(end),'post');
    tV = t_bin/2*et:et:(Nframes-t_bin/2)*et;
    jq = interp1(ints,Jvect,tV,'makima');
    jq_full = padarray(jq,[0 t_bin/2],jq(end),'post');
    jq_full = padarray(jq_full,[0 t_bin/2-1],jq(1),'pre');
end


if ~sum(diff(iq_full)) && ~sum(diff(jq_full))
    disp(['No discernable drift. Try decreasing number of slices or '...
        'reducing object density.'])
    Ccorr = []; Ccorrmat = [];
    return
end


h = waitbar(0,'Correcting...');
Ccorr = cell(length(C),1);
for k = 1:length(C)
    frames = F{k}';
    bt = B{k};
    coords = C{k}';
    i_corr = coords(:,1) + iq_full(frames)';
    j_corr = coords(:,2) + jq_full(frames)';
    tokeep = i_corr <= m & i_corr > 0 & j_corr <= n & j_corr > 0;
    if ~tokeep
        waitbar(k/length(C))
        continue
    end
    frames = frames(tokeep);
    bt = bt(tokeep);
    i_corr = i_corr(tokeep);
    j_corr = j_corr(tokeep);
    Ccorr{k}(:,1) = i_corr;
    Ccorr{k}(:,2) = j_corr;
    Ccorr{k}(:,3) = frames;
    Ccorr{k}(:,4) = bt;
%     Ccorr{k,1} = i_corr;
%     Ccorr{k,2} = j_corr;
%     Ccorr{k,3} = frames;
    waitbar(k/length(C))
end
close(h); clear h

%
% % OLD METHOD, CREATES ARRAYS TOO LARGE BUT FASTER-------------------------
% % Apply correction to all particles
% % h = waitbar(0,'Analyzing...');
% % trkIDall = objs_link(6,:); numAllTrajs = max(trkIDall);
% % trksAll = cell(numAllTrajs,1); frmsAll = trksAll;
% % Xall = NaN(numAllTrajs,Nframes); Yall = Xall; bbAll = Xall;
% %
% % for k = 1:numAllTrajs
% %     trksAll{k} = objs_link(1:2,trkIDall==k);
% %     frmsAll{k} = objs_link(5,trkIDall==k);
% %     Xall(k,frmsAll{k}) = objs_link(1,trkIDall==k);
% %     Yall(k,frmsAll{k}) = objs_link(2,trkIDall==k);
% %     bbAll(k,frmsAll{k}) = objs_link(3,trkIDall==k);
% %     waitbar(k/numAllTrajs);
% % end
% % close(h); clear h
% % XCall = Xall + iq; YCall = Yall + jq;
% % XCall(XCall<=0) = -1; XCall(XCall>m) = -1;
% % YCall(YCall<=0) = -1; YCall(YCall>n) = -1;
% %
% % Ccorr = cell(size(trksAll));
```

Figure 5:   Drift correction code, page 5.

112

```matlab
% % for k = 1:length(Ccorr)
% %     xs = XCall(k,:); ys = YCall(k,:);
% %     removedx = xs==-1; removedy = ys==-1;
% %     todeletex = isnan(xs)|removedx;
% %     todeletey = isnan(ys)|removedy;
% %     xs(todeletex|todeletey)=[];ys(todeletex|todeletey) = [];
% %     if ~isempty(xs)
% %         frames = frmsAll{k};
% %         remFrames = ismember(frames,find(removedx|removedy));
% %         frames(remFrames) = [];
% %         Ccorr{k} = [xs',ys',frames'];
% %     end
% % end
Ccorrmat = cell2mat(Ccorr);

% For debugging set to true
showcorrection = true;
if showcorrection
    figure,plot(et:et:Nframes*et,iq_full,'-',et:et:Nframes*et,jq_full,'--')
    title(['t_bin = ',num2str(t_bin),' s, s_bin = ',num2str(s_bin*pixel_size),' nm'])

%     orig = cell2mat(C);
    orig = cat(2,C{:})';

    figure,plot(orig(:,1),512-orig(:,2),'.')
    hold on
    plot(Ccorrmat(:,1),512-Ccorrmat(:,2),'.')
    h = gca; h.XLim = [1 512]; h.YLim = [1 512];
    title(['t_{bin} = ',num2str(t_bin),' s, s_{bin} = ',num2str(s_bin*pixel_size),' nm, interval ↙
= ',num2str(int),' frames'])
%     orig = [orig,cell2mat(cellfun(@(x)x',F,'UniformOutput',0))];

    % Display corrected points
%     figure, h = gca; h.XLim = [1 512]; h.YLim = [1 512]; hold on
%     for k = 1:Nframes
%         origcenters = orig(orig(:,3) == k,1:2);
%         plot(origcenters(:,1),512-origcenters(:,2),'r.')
%         centers = Ccorrmat(Ccorrmat(:,3) == k,1:2);
%         plot(centers(:,1),512-centers(:,2),'b.')
%         pause(0.025)
%     end
end
```

Figure 6:   Drift correction code, page 6.

113

.2 Trajectory localizing algorithm

```matlab
function [spotEvents, eventfreq] = eventlinks01(tol,ttol,et,C,M,varargin)

% INPUT
%   TOL: Distance tolerance for connection in units of pixels
%   TTOL: Time tolerance for connection in units of seconds
%   ET: Frame interval in units of seconds
%   C: Cell array with each cell containing the coordinates of the objects
%      in the trajectory. If output of COLLATE, include F as an argument.
%      If output of DRIFTCORR, the frame number is included in the array.
%   M: Cell array with each cell containing the molecule index number into
%      objs_link.
%   VARARGIN: Optional input of F (1st) and B (2nd), cell
%      array containing the frame number of each object in the trajectory,
%      and cell array with each cell containing the brightness of the
%      objects in the trajectory, respectively.

% Organize inputs
if ~isempty(varargin)
    F = varargin{1};
    B = varargin{2};
    C = cellfun(@(x)x',C,'UniformOutput',false);
else
    F = cell(length(C),1); B = F;
    for k = 1:length(C)
        if ~isempty(C{k})
            F{k} = C{k}(:,3);
            B{k} = C{k}(:,4);
        else
            F{k} = [];
            B{k} = [];
        end
    end
end

%% Create searchable array
longCoords = C;
clengths = zeros(length(C),1);

% Columns of tosearch: Molecule number, last frame, i_last, j_last, first
% frame, i_first, j_first
tosearch = nan(length(longCoords),7);
for k = 1:length(longCoords)
    clengths(k) = size(longCoords{k},1);
    if clengths(k) <= 1
        longCoords{k} = [];
    else
        tosearch(k,1) = k;
        tosearch(k,2) = F{k}(end);
        tosearch(k,3) = C{k}(end,1);
        tosearch(k,4) = C{k}(end,2);
        tosearch(k,5) = F{k}(1);
        tosearch(k,6) = C{k}(1,1);
        tosearch(k,7) = C{k}(1,2);
    end
end

%% Find connectable trajectories

ftol = ttol/et; % Convert time tolerance to frame tolerance.
spotEvents = struct;
eventfreq = zeros(length(longCoords),1);
lastframes = unique(tosearch(:,2));
lastframes = lastframes(~isnan(lastframes));
toconnect = cell(length(longCoords),1);
for k = 1:length(longCoords)
    toconnect{k} = k;
```

Figure 7:   Trajectory localizing algorithm, page 1.

```matlab
end

h = waitbar(0,'Finding trajectories to connect');
for k = 1:length(lastframes)
    currlastframe = lastframes(k);
    rows_with_lastframe = tosearch(:,2) == currlastframe;
    seeds = tosearch(rows_with_lastframe,3:4);
    longCoords_inds = tosearch(rows_with_lastframe,1);
    in_tol_inds = find(tosearch(:,5) - currlastframe < ftol & tosearch(:,5) - currlastframe > 1 &↙
~rows_with_lastframe);

    [idx,D] = rangesearch(tosearch(in_tol_inds,6:7),seeds,tol);

    closest_traj = zeros(length(idx),1);
    closest_dist = zeros(length(idx),1);
    l = 1;
    while l <= length(idx)
        if ~isempty(idx{l})
            closest_traj(l) = idx{l}(1);
            closest_dist(l) = D{l}(1);
            l = l + 1;
        else
            longCoords_inds(l) = [];
            idx(l) = []; % It's already empty, why are you making it more empty?
            D(l) = [];
        end
    end
    closest_dist = closest_dist(closest_dist ~= 0); % This is okay because longCoords_inds was↙
updated
    closest_traj = closest_traj(closest_traj ~= 0);

    for l = 1:length(idx)
        if ~isempty(idx{l})
%            toconnect{longCoords_inds(l)} = [toconnect{longCoords_inds(l)},in_tol_inds↙
(closest_traj(l))];
%            reveal_repeats = closest_traj(:) == closest_traj(l);
            s = 1;
            while s <= length(idx{l})
                closest = idx{l}(s);
                find_matches = closest_traj(:) == closest;
                if sum(find_matches) > 1
                    [~,min_ind] = min(closest_dist(find_matches));
                    A = find(find_matches);
                    idx_ind_of_traj = A(min_ind); % Index of trajectory to be extended in IDX
                    ind_of_traj = in_tol_inds(idx{idx_ind_of_traj}(1)); % Index of extending↙
trajectory in LONGCOORDS

                    % Remove other mathcing trajectories
                    find_matches(idx_ind_of_traj) = false;
                    A = find(find_matches);
                    for p = 1:length(A)
                        if length(idx{A(p)}) > 1
                            where_other_close_trajs = idx{A(p)} ~= closest_traj(A(p));
                            other_close_trajs = idx{A(p)}(where_other_close_trajs);
                            closest_traj(A(p)) = other_close_trajs(1);
                            other_close_dist = D{A(p)}(where_other_close_trajs);
                            closest_dist(A(p)) = other_close_dist(1);
                        else
                            closest_dist(A) = NaN;
                            closest_traj(A) = NaN;
                            longCoords_inds(A) = NaN;
                            idx{A(p)} = [];
                        end
                    end

                    A = longCoords_inds(idx_ind_of_traj);
```

Figure 8:   Trajectory localizing algorithm, page 2.

116

```matlab
                        toconnect{A} = [toconnect{A},ind_of_traj];
                        tosearch(A,2:4) = tosearch(ind_of_traj,2:4); % Maybe this should be tosearch↙
(ismember(tosearch(:,1),ind_of_traj),2:4) because TOSEARCH and LONGCOORDS don't have same↙
indices? They do actually.
                        tosearch(ind_of_traj,:) = nan(1,7);
                        s = s + 1;
                    else
                        A = longCoords_inds(l);
                        I = in_tol_inds(closest);
                        toconnect{A} = [toconnect{A},I];
                        tosearch(A,2:4) = tosearch(I,2:4);
                        tosearch(I,:) = nan(1,7);
                        break
                    end
                end
            end
        end
    waitbar(k/length(lastframes),h);
end
close(h)
clear h
%% Connect connectable trajectories

p = 1;
for k = 1:length(toconnect)
    if ~isempty(longCoords{k})
        traj2connect = toconnect{k};
        if length(traj2connect) > 1
            tmptraj = M{traj2connect(1)};
            for l = 1:length(traj2connect)-1
                currlastframe = F{traj2connect(l)}(end);
                nextfirstframe = F{traj2connect(l+1)}(1);
                spacer = nan(1,nextfirstframe - currlastframe - 1);
                tmptraj = [tmptraj,spacer,M{traj2connect(l+1)}];
            end
            spotEvents(p).trajectory = tmptraj;
        else
            spotEvents(p).trajectory = M{traj2connect};
        end
        spotEvents(p).coordinates = cat(1,longCoords{traj2connect});
        spotEvents(p).brightness = cat(1,B{traj2connect});
        spotEvents(p).frames = cat(1,F{traj2connect});
        spotEvents(p).eventfreq = length(traj2connect);
        spotEvents(p).trajConnected = traj2connect;
        eventfreq(p) = length(traj2connect);
        p = p + 1;
    end
end
eventfreq = eventfreq(eventfreq~=0);

end
```

Figure 9:   Trajectory localizing algorithm, page 3.