Single-Molecule Studies of Ligand Binding to Hyperpolarization-Activated Cyclic Nucleotide-Gated Ion Channels in Zero-Mode Waveguides

By

David S. White

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Neuroscience)

at the
UNIVERSITY OF WISCONSIN – MADISON
2020

Date of final oral examination: 08/18/20

This dissertation is approved by the following committee:

Baron Chanda, Professor, Neuroscience
Randall Goldsmith, Associate Professor, Chemistry
Mathew Jones, Professor, Neuroscience
James Weisshaar, Professor, Chemistry
Xinyu Zhao, Professor, Neuroscience

Abstract

Hyperpolarization-activated cyclic nucleotide-gated (HCN) ion channels are critical for cellular pacemaking and are functionally augmented by cyclic nucleotide binding. Although this interaction underlies physiological responses such as enhanced neuronal excitability, the mechanism of how cyclic nucleotide binding drives channel modulation remains unclear. Our lack of mechanistic understanding in part stems from a reliance on macroscopic techniques that cannot clearly identify individual binding events to unambiguously dissect channel response in real time. In contrast, single-molecule fluorescence approaches are powerful tools for resolving the often heterogenous behaviors of proteins that are otherwise averaged over in ensemble measurements. However, these sophisticated techniques are restricted by the inherent concentration barrier of focused light and thus cannot resolve ligand associations at physiological concentrations.

Herein, we describe the use of nanophotonic zero-mode waveguides (ZMWs) for dissecting the binding mechanism of cyclic nucleotides to non-activated HCN ion channels at physiological concentrations. In Chapter 2, we devise a novel technique combining Förster resonance energy transfer (FRET) in ZMWs to extend the concentration barrier of single-molecule fluorescence by two orders of magnitude and determine the binding dynamics of isolated binding domains. In Chapter 3, we develop a new time-series idealization algorithm using unsupervised machine learning to speed up data processing by two orders of magnitude, thereby enabling high-throughput analysis of the large single-molecule datasets afforded by ZMWs. Finally, in Chapter 4 we determine a non-cooperative binding mechanism to two HCN isoforms and further resolve the subtle kinetic differences that may underlie their dissimilar responses to cyclic nucleotides. Overall, this work provides key insights into HCN ion channel gating and describes new technologies for studying physiological processes at the single-molecule level.

Acknowledgements

Completing a PhD is a daunting task that can in no way be completed alone. I am fortunate that along the way I have had a terrific group of friends and family that supported me during this journey.

I would first like to thank the Neuroscience Training Program for granting me admission into their terrific program. NTP has provided me with wonderful opportunities to grow as a scientist and engage with the community. In particular, working with the PEOPLE program as a high school neuroscience instructor each summer was a highlight of my graduate career. I would also like to thank the administration of both NTP and of the Department of Neuroscience for providing all the necessary support and guidance to complete my degree.

Next, I would like to thank my co-mentors, Dr. Baron Chanda and Dr. Randall Goldsmith. Together, you two have taught me the importance of collaboration in science and the advantage of merging multiple scientific disciplines to advance fundamental knowledge. From your combined mentorship, I have grown as a researcher and communicator. This project has clearly gone down unexpected avenues and I would not have been able to accomplish such multi-faceted research without your combined expertise and ingenuity. I would also like to thank my committee members Dr. Mathew Jones, Dr. James Weisshaar, and Dr. Xinyu Zhao for their regular discussion and feedback over the years to help me grow as a scientist.

The most unexpected avenue of my PhD in neuroscience was learning how to perform nanofabrication at Oak Ridge National Laboratories. The Center for Nanophase Materials Sciences is staffed with a wonderful group of people that are brilliant and an absolute joy to work with. Dr. Scott T. Retterer was a terrific mentor and I could not have accomplished this project without him. I would also like to thank Bernadeta, Kevin, Dale, Dayrl, and Leslie for all their training and company in the clean room.

Of course, being co-advised also means I was fortunate to have two terrific groups of colleagues across the two labs. In particular, I would like to thank Marcel for taking me under his wing and teaching me not only how to be perform experiments and write software, but also how to strive to maintain a

proper work-life balance. In addition, Sandipan and Ruohan were absolutely terrific scientists to work with and I have learned a lot from each of them. John and Yihao were easily the best people a person could share a lab with. Andrew and Morgan were terrific people to join the lab together with and they always made me feel more welcome as a member of a lab I less frequently visited. Kasie and Ceci deserve an award for the sheer number of cover glasses they cleaned on my behalf for each of my trips to Oak Ridge National Labs. I would also like to thank Lydia for her keen eye pointing out my many (and unique) grammatical and spelling errors over the years.

Finally, to my family. I would like to thank my parents for their support and encouragement when I decided to move to Madison and pursue yet another degree. And, of course, I would like to thank my partner, Sophie. You've encouraged and supported me through all the ups and downs of graduate school and can't express how grateful I truly am. I would not have been able to accomplish this without you.

Table of Contents

List of Figures

List of Tables

Publications

In preparation

1. **White, D.S.**[*], Chowdhury, S.[*], Zhang, R., Retterer, S.T., Goldsmith, R.H., Chanda, B. Single-molecule measurements in photonic waveguides reveal non-cooperative binding dynamics in intact pacemaker channels. *In preparation*
   [*] these authors contributed equally

2. **White, D.S.,** Mackinsey, S.A., Chanda, B., Goldsmith, R.H. Approaches for breaking the single-molecule concentration barrier. *In preparation*
   [*] invited perspective for *Analytical Chemistry*

In press

1. **White, D.S.**, Goldschen-Ohm, M.P., Goldsmith, R.H., Chanda, B. Top-down machine learning approach for high-throughput single-molecule analysis. *Elife* **2020**, *9*.

2. Hinton, D.A.[*], Ng., J.D.[*], Sun, J., Lee, S., Saikin, S.K., Logsdon, J., **White, D.S**., Marquard, A.N., Cavell, A.C., Krasecki, V.K., Knapper, K.A., Lupo, K.M., Wasilewski, M.R., Aspuru-Guzlik, A., Biteen, J.S., Gopalan, P., Goldsmith, R.H., Mapping Forbidden Emission to Structure in Self-Assembled Organic Nanoparticles. *Journal of the American Chemical Society* **2018,** *140* (46), 15827-15841
   [*] these authors contributed equally

3. Goldschen-Ohm, M.P.[*], **White, D.S.**[*], Klenchin, V.A., Chanda, B., Goldsmith, R.H., Observing Single-Molecule Dynamics at Millimolar Concentrations. *Angewandte Chemie-International Edition* **2017,** *56* (9), 2399-2402.
   [*] these authors contributed equally

4. Goldschen-Ohm, M.P.[*], Klenchin, V.A. [*], **White, D.S.**, Cowgill, J.B., Cui, Q., Goldsmith, R.H., Chanda, B. Structure and dynamics underlying elementary ligand binding events in human pacemaking channels. *Elife* **2016,** *5.*
   [*] these authors contributed equally

1 Introduction to HCN Channels and ZMWs

1.1 Hyperpolarization-activated cyclic-nucleotide gated ion channels

1.1.1 Ion channels and electrical signaling

Cellular signaling is one of the most fundamental processes of life. The careful orchestra of small molecules and proteins interacting to enable cell-to-cell communication underpins our very existence, ranging from our basic metabolism to our complex cognition. Within this intricate network of interactions are important classes of proteins that each uniquely facilitate a mode of communication. Ion channels are one such group of proteins that allow cells to connect with the world around us[1]. By regulating the flow of ions into or out of cells, these molecular machines regulate fundamental physiological processes including neuronal signaling, synaptic transmission, and muscle contractions in the heart and skeletal tissue. Thus, every thought you've ever had and every action you've ever taken has been carefully coordinated by a delicate flow of ions through these transmembrane proteins.

Ion channels are a broad class of proteins that are often classified according to their general mode of action[2]. For example, voltage-gated ion channels (VGIC) are tetrameric proteins that open in response to changes in the membrane potential whereas ligand-gated ion channels (LGIC) open and close in response to the binding of small molecules. Each class can be further subdivided by additional unique features including ion conductance like potassium, sodium, calcium, or chloride selective VGICs. The function of each ion channel is correlated with their tissue expression demonstrating how unfathomably complex biology is. For example, LGICs often sit at the dendritic spines of post-synaptic neurons that are activated upon chemical messengers (i.e. neurotransmitters) binding during synaptic transmission[3]. In addition, mechanosensitive channels are found throughout our skin and hair cells to help us to respond to physical stimuli like pressure, touch and sound[4].

Despite our knowledge of the wide-spread diversity of ion channel function, gating, and architecture revealed over the past 70 years of scientific studies[5], biology still surprises us with unexpected yet important channels. Hyperpolarization-activated cyclic nucleotide gated ion channels (HCN) are a prominent example of a truly unique class of ion channel[6-8]. As their name implies, HCN channels activate upon membrane hyperpolarization (negative shift in membrane potential) rather than depolarization (positive shift in membrane potential) in contrast to many other VGICs. HCN channels also display a secondary gating behavior wherein the binding of cyclic nucleotides such as 3',5'-cyclic adenosine monophosphate (cAMP) and 3',5'-cyclic guanosine monophosphate (cGMP) enhances the rate and probability of opening while shifting voltage sensitivity to more depolarizing potentials. Understanding the physical basis and physiological roles of these unique gating mechanisms is of critical interest to biologists. In particular, this dissertation explores the mechanism of ligand binding to HCN channels.

1.1.2 HCN physiology

HCN channels underlie critical physiological processes in the heart and the brain. One such role is their participation in so-called "pacemaker" current. During the hyperpolarization phase of an action potential, HCN channels can open and allow an inward flux of sodium ions into the cell (a typical reversal potential of HCN is -20 mV)[9]. The opening of HCN channels depolarizes the membrane again, thereby closing HCN channels but bringing the membrane potential close the threshold of another action potential. This pattern of activation and self-inhibition of HCN current underlies many processes including the rhythmic and oscillatory behavior of network thalamocortical cells[10]. In particular, during a "fight or flight" response, ß-adrenergic simulation leads to the production of cAMP which in turns binds to HCN channels in the sinoatrial node of the heart, thereby increasing the frequency of pacemaking[11]. HCN channels are further involved in a variety of other roles including dendritic integration, motor learning,

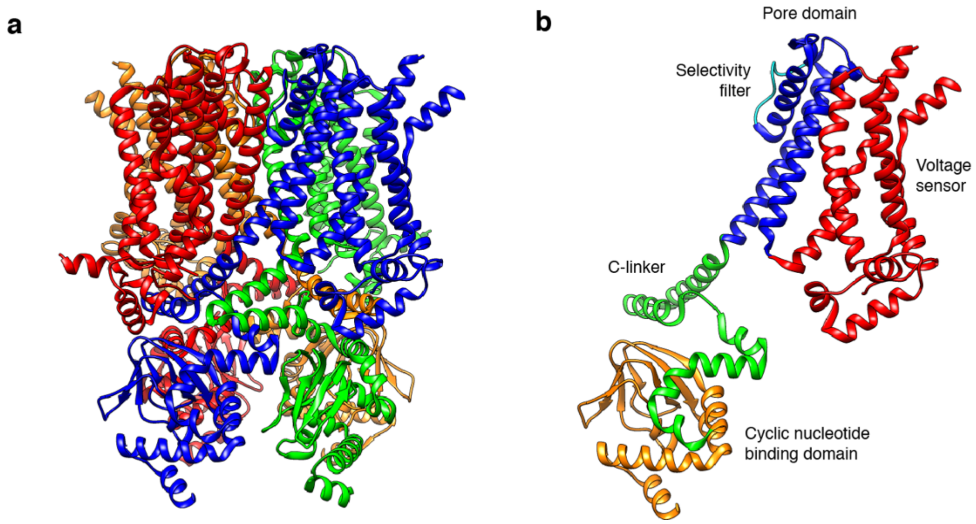

Figure 1-1: Structure of HCN channels. *(**a**) Cryo-EM structure of HCN1 (PDB ID: 5U6P). Each subunit of the tetrameric complex is individually colored. (**b**) A single subunit of HCN1. Primary structural features are colored and include the S1-S4 voltage sensor (red), the S5-S6 pore domain (blue), the selectivity filter (light blue), the C-linker (green) and the cyclic nucleotide binding domain (orange).*

and pain[12-15]. As a result, the dysregulation of HCN channels is associated with a variety of diseases which makes them prime pharmaceutical targets for the treatment of conditions including depression, arrhythmia, chronic pain, and epilepsy[12,16-21]. A robust mechanism describing the coupling between channel activity and ligand binding would help guide future pharmaceutical development.

1.1.3 HCN topology

Four mammalian isoforms of HCN (HCN1-4) have been discovered with 60% sequence similarity[22]. As shown with the recent cryogenic electron microscopy (cryo-EM) structure of HCN1 (**Figure 1-1a**)[23], the transmembrane core of HCN channels resemble typical potassium VGICs. HCN channels can express as either homo- or hetero-tetramers[24]. The S1-S4 helices make up the voltage sensor and the S5-S6 form the pore (**Figure 1-1b)**. HCN channels diverge from potassium channel topology at the C-terminus due to the presence of an intracellular cyclic nucleotide binding domain (CNBD) on each subunit. CNBDs are an ancient protein domain conserved across a variety of proteins including bacterial transcription factors, protein kinases, and cyclic-nucleotide gated (CNG) channels[9,25]. Each CNBD is composed of ~120 amino acids and is connected to the S6 helix of the pore by a C-linker region of ~80 amino acids. The transmembrane core and C-terminal regions exhibit a high degree of sequence homology across the four isoforms (~80-90%) unlike the the N-terminus of each protein which varies in length[26]. Therefore, the static snap-shot of a single HCN isoform captured by cryo-EM is insufficient to provide a broad understanding of the functional and (potentially) structural diversity of the HCN family. Moving forward, we will primarily focus on HCN1 and HCN2 activity for their widespread expression in the heart and nervous system and differing activation kinetics[9].

1.1.4 Modulation by cyclic nucleotides

As discussed previously, HCN channels are activated hyperpolarizing voltages and modulated by cyclic nucleotides. Unlike cyclic nucleotide-gated (CNG) ion channels, cyclic nucleotide binding alone does not open HCN channels but rather enhances the opening kinetics and shifts activation to more depolarizing potentials[25]. Both cAMP and cGMP are effective agonists of HCN channels; however, they exhibit low (high nM to low µM) affinities with a preference for cAMP binding[27]. The activation kinetics and efficacy of cAMP on HCN channels also differ across the isoforms (**Figure 1-2**). For example, the half-maximal voltage for activation ($V_{1/2}$) is more depolarized for HCN1 ($V_{1/2}$= -100 mV) compared to HCN2 ($V_{1/2}$ = -130)[28,29]. Upon binding cAMP, HCN1 exhibits a minor shift in $V_{1/2}$ (+5 mV) compared to larger shift of HCN2 (+15 mV) (**Figure 1-2**)[28,29]. Particularly for HCN2, cAMP binding further increases the maximum open probability (P_O) of the channel. The magnitude of pore modulation from cAMP binding is dependent on the CNBD and C-linker regions and can be altered by mutating a small number of amino acids[28,29].

Despite the high sequence similarity of the C-linker and CNBD domains, it remains unclear how binding to each isoform allosterically and differentially modulates the channel's pore. A prevailing hypothesis in the field arose from the observation that removing the C-terminus of HCN2 mimics the effect

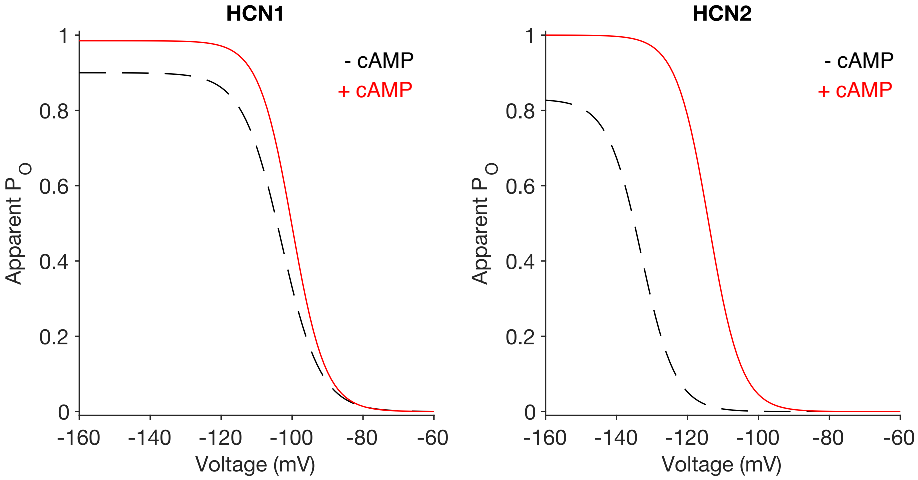

Figure 1-2: Modulation of HCN channels by cAMP. *Simulation of HCN1 (left) and HCN2 (right) activation kinetics in the absence (black dashed) and presence of 10 µM cAMP (red). Values were adapted from Boltzmann fits of normalized tail currents provided in ref 28. The apparent* P_O *was computed by the Boltzmann equation for each value of voltage (x) applied by* $P_O = V_{max}/(1 + e^{1+[(V_{1/2}-x)/m)]})$. *For HCN1,* V_{max} *= 0.90,* $V_{1/2}$*=-103 mV, m =-5.5 mV; for HCN1+cAMP,* V_{max} *= 0.99,* $V_{1/2}$ *= -100 mV, m = -4.7 mV; for HCN2,* V_{max} *= 0.83,* $V_{1/2}$ *= -133 mV, m = -4.8 mV; for HCN2+cAMP,* V_{max} *= 1.0,* $V_{1/2}$ *= -114 mV, m = -4.6 mV.*

of cAMP binding of the wild-type channel[30]. This suggests that the C-linker and CNBD inhibit pore opening 3of HCN2 in the absence of cAMP. Following this observation, we can infer that HCN1 is less endogenously inhibited by its C-terminus given its shifted activation kinetics from HCN2 and minimal modulation upon cAMP binding. However, the molecular mechanism underlying this effect has still not been thoroughly investigated and remains an open question in the field.

1.1.5 Mechanisms of cyclic nucleotide binding to HCN channels

Given the physiological effect of cAMP binding on HCN channel gating, there is much interest in understanding how ligand binding couples to changes in the pore. X-ray crystallography has resolved both liganded and unliganded structures of CNBDs from HCN2 which show a clam-shell like closure upon cAMP binding[31,32]. These results suggest a conformational change could propagate up through the C-Linker to modulate the pore. This hypothesis is further supported by the cryo-EM structure of HCN1 which shows that the unusually long S4 voltage sensing helix extends intracellularly and provides direct contact between the C-Linker and S4-S5 loop[23]. Functionally, the binding of cAMP to isolated CNBDs was shown to induce structural rearranges that could initiate a conformational wave that propagates through the C-helix to the pore domain[33]. Additionally, functional analysis and X-ray crystal structures of tetramerized CNBDs of in the presence of cAMP suggest that ligand binding induces tetramerization of isolated CNBDs from both HCN1 and HCN2[31,34]. These data further suggest an increase in cAMP affinity during the process of CNBD tetramerization[31,34]. This result allows us to reasonably suspect that there maybe be some cooperative interaction between CNBDs during cAMP binding that may allosterically affect the pore.

Investigations into the interactions between subunits upon HCN2 channel activation have primarily originated from the Benndorf group. To observe the interaction of both gating modalities, the group developed a hybrid patch-clamp and confocal fluorometry technique which simultaneously monitors channel activity and tracks the bound fraction of fluorescently-conjugated cyclic nucleotides (i.e. fluorescently-labeled cAMP, fcAMP)[35,36]. Using this technique and global Markov modeling of their data, they have postulated multiple models of ligand binding to HCN2 channel. When the channels were activated (open) at -130 mV, subunits displayed an unusual sequence of positive-negative-positive cooperativity[37], consistent with a dimer-of-dimers hypothesis[38]. When the experiment was repeated with closed channels (-30 mV), a sequence of slightly negative-none-slightly positive cooperativity was determined[39]. The best fitting model of the closed channel further featured a concerted conformation flip of the CNBDs which they speculated to be a tetramerized intermediate state[39]. Overall, these results suggest complex and varying cooperativity of HCN2 subunits during ligand activation.

1.1.6 The macroscopic limitation

Unfortunately, the previous functional studies of isolated CNBDs and the kinetic models of subunit cooperativity are not entirely reliable. This is because these techniques are macroscopic measurements that not only obscure the direct resolution of each individual binding site but average the activity of potentially thousands of individual molecules. In particular, the resulting large number of parameters (i.e. rate constants) from global modeling of HCN2 channels cannot be reliably determined from these experiments due to their low resolution. This issue of so-called *parameter non-identifiability* in biophysical measurements was recently demonstrated by Hines[40]. Using simulations of noiseless ensemble data of ligand binding to calmodulin, Hines showed that analysis of this data could result in multiple cooperative schemes with nearly identical fits of the binding curve data[40]. In particular, the best fitting schemes featured practically no cooperativity between the four binding sites or a sequence of positive-negative-positive cooperativity, essentially the same trend observed in open and closed HCN2 channels[37,39]. Thus, inferences on the modeled rate constants and subunit cooperativity of calmodulin (and therefore HCN2) are potentially meaningless when extracted from binding curves alone[40].

To circumvent the issue of parameter non-identifiability and develop a reliable model of cAMP binding, an experiment with higher resolution is needed. Feynman famously gave us an answer to this problem years ago: “It is very easy to answer many fundamental biological questions; you just look at the thing!”[41]. Coupling this advice with decades of technological advances, directly watching the binding of cyclic nucleotides to HCN channels at the single-molecule level is now possible[42]. This approach allows us to discern binding and unbinding events in real time to directly examine subunit interaction during channel gating[43-45]. The following section will detail how single-molecule analysis can be used to resolve complex kinetics in biological physics and survey the common techniques used to provide high-resolution functional measurements of ion channel gating.

1.2 The single-molecule advantage

1.2.1 Beyond ensemble averaging

Over the past few decades, single-molecule approaches have helped to transform our understanding of complex biological, chemical, and physical processes[42,44,46-48]. In contrast to ensemble (macroscopic) experiments which average the molecular activity of potentially millions of molecules, single-molecule approaches report the behavior of individual molecules and allow quantification of the asynchronous and often heterogenous dynamics not observed in bulk[48,49]. The primary difference between the two experimental modalities is the ability to construct a distribution of a particular parameter at the single-molecule level. While bulk-level experiments that yield an average value of some parameter across large number of (presumably identical) molecules still has tremendous physical value, a distribution of parameter clearly contains more information than the average value alone. For example, the shape of the distribution can be analyzed to determine the presence of multiple peaks which may help reveal potentially more complex or heterogenous dynamics than previously anticipated[49,50]. The applications of single-molecule techniques are broad, ranging from catalysis[51,52], motor protein tracking[53], and observing ion channels open and close[54]. In particular, single-molecule studies of ligand binding mechanisms may provide a powerful avenue to probe the physical basis of binding and allostery[55-59].

In this section, we will discuss the typical statistical analysis of single-molecule data commonly used in in biological physics and throughout **Chapters 2-4**. We will then highlight the primary single-molecule techniques used to study ion channel behavior. Finally, we will motivate the need for new methodologies to directly examine ligand binding dynamics to HCN channels at physiological concentrations. This will by no means be an exhaustive review of all possible analysis methods or techniques amenable to ion channel studies at the single-molecule level. Rather, we hope to introduce the general theory, statistics, and measurements used throughout this dissertation.

1.2.2 General single-molecule analysis

1.2.2.1 A simple single-molecule problem

Let’s begin by considering a general single-molecule experiment: monitoring the reversible dynamics between two discrete states (**Scheme 1-1**).

$$A \underset{k_{BA}}{\overset{k_{AB}}{\rightleftharpoons}} B$$ Scheme 1-1

In the context of ion channels, one could imagine **Scheme 1-1** as a channel fluctuating between open and shut states; however, this is generalizable to a variety of physical processes ranging from protein folding to traffic on a web page. Here, both states A and B are discrete and discernable by single-molecule observation (i.e. different conductance or intensity levels). The transition $A \rightarrow B$ is governed by the transition rate k_{AB} and $B \rightarrow A$ by k_{BA}, both with units of time^{-1}. If we extend **Scheme 1-1** to single-ligand binding dynamics (as is of interest in this thesis)[43], we can imagine an unbound state (A) that transitions into a bound state (B) with a ligand (l) dependent rate constant, $k_{AB}[l]$ and a ligand independent off rate, k_{BA}. For simplicity, we will assume Markovian dynamics, which have been shown to be a powerful model of single-model dynamics[60-64]. This assumes that the process being modeled is memoryless and the probability of transitioning between states is dependent only on the current state. The goal of this experiment is to determine both k_{AB} ands k_{BA}.

It is important to note that this sort of experiment can and has been performed at the ensemble level[65,66]; however, such analysis averages over the behavior of all the molecules without the ability to see their underlying distribution of dynamics. If the underlying dynamics are more complex than the simple process depicted in **Scheme 1-1** (as will be discussed later), mechanistic information could be lost at the bulk level. At the single-molecule level, the obtained distribution of molecular behavior can be analyzed through a process called dwell time analysis[67-69]. Importantly, the visualization and analysis of dwell times provides a more direct avenue to quantitate molecular behavior compared to typical ensemble experiments[49,50]. The following sub-sections will detail the general statistical approaches taken to perform dwell-time analysis of single-molecule trajectories[i].

---

[i] The following discussion is intended to provide the general introduction to single-molecule analysis that I felt was missing from the literature when I first started my dissertation work. For anyone seeking more detail, I encourage reading the work of Hawkes and Colquhoun who provide early and detailed derivations of the fundamental principles of (ion channel) dwell time analysis which are still widely used today.

1.2.2.2 Statistical analysis of dwell time distributions

In general, a single-molecule trajectory provides a sequence of discrete state transitions over time. Given two possible state options (**Scheme 1-1**), we can consider the probability of observing the state transition $A \rightarrow B$ over some discrete observation time (Δt) across a number of equal observation trials (n). Each observation trial is reminiscent of a Bernoulli trial in probability, whereby the success of transitioning between states during Δt is governed by a success rate (p) and failure rate ($1-p$). For context, this is the same distribution underlying tossing a coin wherein there is probability (p = 0.5) of getting "heads" with each discrete coin toss (Δt). If we assume that each trial is independent, we can determine the probability of success on the r^{th} trial ($P(r)$) following $r-1$ failed trails by

$$P(r|p) = p(1-p)^{\,r-1} \quad r = 1, 2, 3 \ldots, \infty \qquad \text{Eq. 1-1}$$

This probability distribution is called a *geometric distribution* which simply means the process being modeled is a sequence of independent trials with only two possible outcomes and an equal probability of success for each trial. The mean (μ) of a geometric distribution (and therefore the mean number of observation trials before a state transition occurs) is

$$\mu = \sum rP(r) = \frac{1}{p} \qquad \text{Eq. 1-2}$$

Relating transition rates to the Bernoulli success rate, the probability of state transition $A \rightarrow B$ during Δt is $p = k_{AB}\Delta t$. It is important to note that a geometric distribution is constrained to discrete space yet proteins operate in continuous space. To account for this, consider the limits of $\Delta t \rightarrow 0$ and $r \rightarrow \infty$. Here, we can imagine being able to watch a protein for infinitely long time with infinitely short observation times. In this limit, $p \rightarrow 0$ for each discrete Δt, which gives rise to

$$\lim_{p \rightarrow 0} (1-p)^{x/p} = e^{-x} \qquad \text{Eq. 1-3}$$

Let us now consider that the probability of a transition from $A \rightarrow B$ is directly related to the time spent in state A before transitioning to state B, a metric called *dwell time*. We can rearrange **Eq. 1-1** to express the probability of remaining in state A for duration t_i before a transition occurs as the exponential probability density function

$$PDF(t_i \mid k_{AB}) = k_{AB} e^{-k_{AB} t_i} \quad \text{where } t_i \geq 0 \qquad \text{Eq. 1-4}$$

Thus, the exponential distribution is simply the continuous analog of the discrete geometric distribution, and therefore also considers the duration of time before an event occurs. Similar to **Eq. 1-2**, the mean of an exponential distribution (commonly notated as τ) is

$$\begin{aligned} \tau_A &= 1/k_{AB} \\ \tau_B &= 1/k_{BA} \end{aligned} \qquad \text{Eq. 1-5}$$

which is the mean dwell time in a given state[ii]. This result gives us a statistical model for evaluating the probability of a state transition given a transition rate.

1.2.2.3 Dwell time analysis: the simple case

Recall that our experimental goal is to determine k_{AB} and k_{BA} from our single-molecule observations of **Scheme 1-1**. For now, let's examine a noiseless case so we can clearly see distinct transitions between the two states over time (**Figure 1-3a**). To perform dwell time analysis, we simply need to collect the duration of time within each state, which are commonly visualized as a histogram (**Figure 1-3b, c**). Under our assumption that **Scheme 1-1** is the underlying process of this data (which may or may not be true), we can fit each dwell time distribution with a single-exponential function to compute the transition rate (**Eq. 1-5**).

But how do we compute best parameters to fit our observed distribution with an exponential? So far, we discussed probability distributions which evaluate the probability of observing the data (x) given a hypothesis (θ). Here, the hypothesis is fixed and the data is variable. However, when we are interested in the hypothesis of the data (i.e. best fit of a distribution) we now assume our data is fixed and the

---

[ii] Outside of rate kinetics, the typical notation is $\tau = 1/\lambda$ where τ is the time constant and λ is the rate parameter.

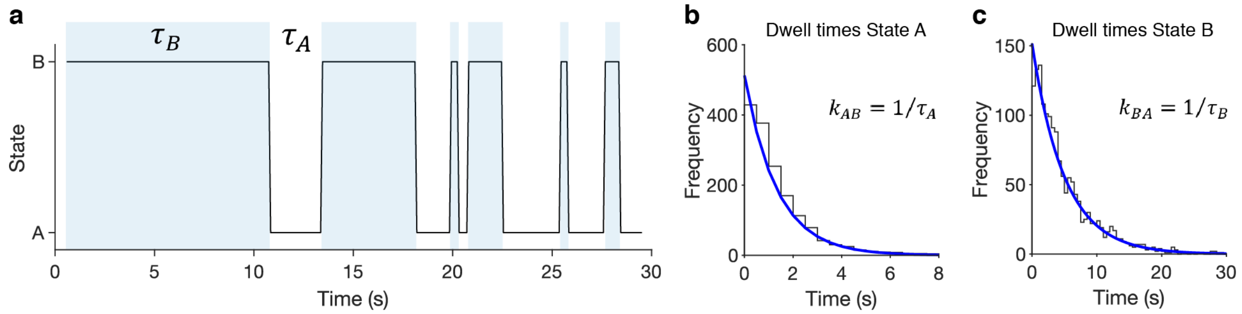

Figure 1-3: Dwell time analysis with two states. *(**a**) Example of a noiseless state vs time single-molecule trajectory. Shaded areas indicate the duration of time spent in state B. Only the first 300 of 10,000 data points are shown. Histograms of dwell times in state A (**b**) and state B (**c**) across 10,000 simulated data points. Dwell times are binned with a bin width of 0.5 s (grey) and overlaid with an exponential distribution determined by MLE (blue). The mean of each dwell time distribution is the reciprocal of the rate transitions.*

hypothesis is variable. This paradigm is regarded as the likelihood of the hypothesis given the data. Mathematically, probabilities are denoted as $P(x|\theta)$ whereas likelihoods are denoted as $L(\theta|x)$. To determine the best hypothesis to fit our dwell times distributions, we can turn to a statistical technique called maximum likelihood estimation (MLE)[68-71]. MLE aims to return the optimal parameters of θ that maximizes the likelihood function over the parameter space. To obtain the best estimated parameters ($\hat{\theta}$), we can take the best value of θ out of the full parameter space which maximizes $L(\theta|x)$

$$\hat{\theta} = \arg\max_{\theta} L(\theta|x) \qquad \text{Eq. 1-6}$$

We can rewrite **Eq. 1-6** to yield the parameters of an exponential distribution ($\hat{\lambda}$) that maximize the likelihood of observing our distribution of N total dwell times ($t_1, \dots t_N$) by

$$\hat{\lambda} = \arg\max_{\lambda} L(\lambda \mid t_1, \dots t_N) \qquad \text{Eq. 1-7}$$

As each dwell time is independent, the likelihood function (L) is simply the product of each density.

$$L(\lambda| \, t_1, \dots t_N) = \prod_{i=1}^{N} \lambda e^{-\lambda t_i} \qquad \text{Eq. 1-8}$$

Since optimizing the product of probabilities is often computationally expensive and can result in numeric underflow[iii], it is common to maximize the log of the likelihood function (LL)

[iii] This is concept confused me in my early graduate studies so I will elaborate. At the level of the CPU, multiplication proceeds via addition; that is $a * \, b = \sum_{i=1}^{b} a_i$. If a and b are probabilities, one may take their log sum of instead of

$$LL(\lambda|\, t_1, \dots t_n) = n \ln{(\lambda)} - \lambda \sum_{i=1}^{n} t_i \qquad \text{Eq. 1-9}$$

For a single exponential distribution, the MLE estimator of λ is simply the reciprocal of the sample mean.

$$\hat{\lambda} = n / \sum_{i=1}^{n} t_i \qquad \text{Eq. 1-10}$$

Overall, this enables to use to compute the transitions rates from our simple simulated scenario easily. Simply, the mean value of the dwell times distribution is the MLE estimated value; therefore, the most likely rate kinetics for the given model are the reciprocal of the average dwell time (**Figure 1-3**).

1.2.2.4 Dwell time analysis: multiple states

A key advantage of a single-molecule experiment is the ability to resolve underlying mechanisms that can be obscured in bulk from the obtained distributions of molecular behavior. In particular, we can imagine a kinetic paradigm more complex than **Scheme 1-1** wherein addition states are present as shown by multiple peaks or skew in the observed dwell time distributions. Although proteins may exist in an infinite landscape of possible conformations, single-ion channel recordings provided early and strong evidence for the direct existence of multiple discrete states of large proteins[61]. Therefore, we can extend our Markovian assumptions and consider a more complex scheme containing a third state. This additional state could be a blocked or inactivated ion channel[72], an alternate form of DNA folding[73], or a conformational exchange of a bound ligand via at "catch-and-hold" mechanism[74]. Overall, these schemes demonstrate the capability of single-molecule measurements to resolve the kinetics of multiple states that may not be observable in bulk experiments. However, there are different analysis considerations depending on whether a new state is *observable*. (**Figure 1-4**).

---

their product because $\log(a * b) = \log(a) + \log{(b)}$. Therefore, we can perform probability computations in log space by $\log \prod p(x_i) = \sum \log{(P(x_i))}$ for increased speed. This arrangement also avoids numeric underflow. Computers use a limited digit floating point representation of a fraction set by the byte size. As multiplying many probabilities will be close to zero, the number of zeros will eventually exceed the storage limit of the byte; therefore, the computer will not have enough precision to represent the true value and can return an error.

In the case of three observable states (i.e. states that are identifiable by unique measurement values, **Figure 1-4a, b**) we can simply extend our Markov assumptions to calculate the underlying transition rates of each state. For states A and C, typical monoexponential fitting of dwell times can be applied. However, as state B has the option of transitioning to two different states, its dwell time is dependent on both k_{BA} and k_{BC} by $\tau_B = 1/(k_{BA} + k_{BC})$. Thus, the same analysis alone cannot resolve these two individual rate constants, but rather their sums. Fortunately, our single-molecule resolution allows us to directly quantitate the relative number of transitions to each state (n_{BA} and n_{BC}) from the total number of transitions (N_B), where N_A=$n_{AB} + n_{BA}$ to discern each individual rate constant.

$$k_{BA} = \frac{n_{BA}}{\tau_B N_B} \qquad\qquad k_{BC} = \frac{n_{BC}}{\tau_B N_B} \qquad \text{Eq. 1-11}$$

In the second case, we can consider a three-state sequential scheme wherein only two emissive states are observed (i.e. A and B); therefore, the third state is not identifiable by experimental observation alone (i.e. B^* in **Figure 1-4f, g**). Dwell time analysis of state A will follow an exponential distribution; however, a single exponential fit of the dwell-time distribution of B leads to a poor fit that clearly does not represent the underlying distribution. This is because the observed dwell time distribution is actually a combination of dwell times of two states with unique time constants ($\tau_B = 1/(k_{BA} + k_{BB^*})$ and $\tau_{B^*} = 1/(k_{B^*B})$ and amplitudes ($a_{BA} + a_{BB^*} = 1$). The PDF for a mixture of n exponential densities is

$$PDF(t) = \sum_{i=1}^{n} a_i \lambda_i e^{-\lambda_i t} \qquad \text{Eq. 1-12}$$

where $\sum a_i = 1$. MLE can again be used to extract the estimated parameters of each exponential distribution ($\widehat{a_1}$, $\widehat{a_2}$, ... $\widehat{a_k}$ and $\widehat{\tau_1}$, $\widehat{\tau_2}$, ... $\widehat{\tau_k}$ for k components). As shown in **Figure 1-4i**, a biexponential fit of the dwell times distribution is clearly more representative of the observed data. While the observation of needing at least a second exponential is easily identifiable at the single-molecule level, , this information could be easily missed at the bulk level without an *a priori* hypothesis.

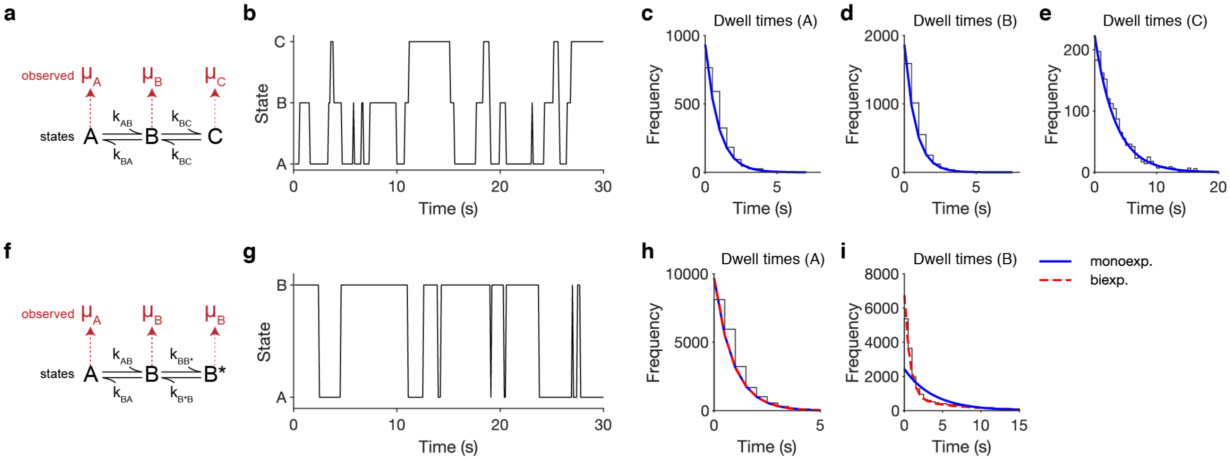

***Figure 1-4: Dwell time analysis with three states.** (**a**) Three state kinetic scheme with each state having a unique observable value denoted by μ. (**b**) Simulated example of a three-state scheme in **a**. The first 300 of 10,000 data points are shown. (**c**, **d**, **e**) Dwell time distributions of states A, B, C overlaid with an exponential distribution from MLE (blue). (**f**) Three state kinetic scheme where only two states have unique observable values. (**g**) Simulated example of a three-state scheme in **f**. The first 300 of 10,000 data points are shown. (**h**, **i**) Dwell time distributions of observable states A and B overlaid with mono- (blue) and bi- (red dashed) exponential distributions from MLE. For state A, a bi-exponential does not yield a better fit than a mono-exponential. For state B, a mono-exponential does not adequately represent the underlying data.*

1.2.2.5 Resolving complex kinetics

In practice, models can often become far more complex than the simple schemes considered here. For example, an activation model of the Shaker potassium channel suggests that each voltage sensor acts independently, leading to 15 possible states during channel opening[75]. Not only can this become computationally complex to assert purely by dwell-time analysis, but real measurements include experimental noise that hides the true signal from the observer. Thus, extracting kinetics from raw data can quickly become a complicated task when both the underlying states and rates are unknown.

A common method to resolve multi-state kinetics in the presence of experimental noise is the hidden Markov model (HMM)[76]. HMMs are probabilistic graphical models that relate a series N of observations in the real world ($X = x_1, \dots, x_N$) to an underlying hidden sequence of K states

($Y = y_1, \dots, y_N$ where $y \in q_1, \dots, q_K$) by parameters $\theta = (S, E, \pi)$[iv]. Here, E is the state transition probability $S = s_{11}, \dots, s_{1K}, \dots, s_{KK}$ where each s_{ij} is the probability of moving from state i to j and $\sum_{j=1}^{K} s_{ij} = 1$. The probability of observation x_n being emitted from the hidden state q_k is $E = e_k(x_n)$. The final parameter π is the probability the model began in each state ($\pi_j = P(q_j = x_1)$).

Given these relationships, one can use HMMs to solve three problems[71]: evaluation, decoding, and learning[76]. For evaluation, given a sequence of observations and HMM parameters, compute $P(X|\theta)$. For decoding, given a sequence of observations and HMM parameters, compute the optimal state sequence $\hat{Y}$. For learning, given a sequence of observations, adjust θ to maximize the $P(X|\theta)$.

HMMs are powerful tools that have been used for a wide range of applications spanning speech recognition to bioinformatics[77]. For single-molecule analysis, the decoding problem can be viewed as time-series idealization (see **1.2.2.7**) wherein we aim to determine the hidden underlying sequence from a known model. The evaluation problem is important to compare models to determine the best representation of the observed data (see **1.2.2.6**). Finally, the learning problem is directly useful for resolving the transition rates and identifying significant observable states in the data. Using the scheme in **Figure 1-4a** as an example, we can imagine the noiseless data is the true underlying Markov chain Y. For this example, we will assume we know three states exist in Y and that their observable values are known (determining number of states in observed data is discussed in **1.2.2.7**). Using HMM learning algorithms like the Baum-Welch approach[76], we can compute the most likely transition probabilities between states.

Within the single-molecule community, HMMs are commonly evaluated using the QuB software[78]. In QuB, transition rates can be globally optimized across many molecules and ligand concentrations for a user-provided model of state connectivity[79]. Further, QuB provides the estimated

---

[iv] Typical notation is A and B[71]; however, to avoid confusion with Scheme 1-1, we will denote them as S and E.

likelihood of the data given the model, allowing users the ability to build and test different models on the same sets of observations to determine the most likely kinetic scheme[62,63]. Within **Chapters 2-4**, the QuB software is used to compute kinetic transitions between states and compare different models to explain best explain sequences of observed states.

1.2.2.6 Model selection

The previous analysis highlights an important question: how many components (i.e. states) are needed to fit the observed data? This concern is generalizable to both HMM and dwell time analysis with multiple exponentials. As more components are used to fit the data, the resulting likelihood of each observation will increase. This leads to the issue of overfitting, wherein the model is now fitting the noise of the data from the inclusion of too many (and possibly unreliable) assumptions. However, if too few components are fit, the underfit model will provide an inaccurate interpretation of the underlying data and fail to generalize to new observations. Thus, model selection is a key step in single-molecule analysis[63]. In general, model-selection follows the notion of Occam's razor's: a simpler explanation is often the best explanation. For a parsimonious model, we need to optimize the trade-off between the fit of the data and the overall complexity (i.e. number of parameters) of the model[v].

Model selection is an extensive topic in statistics and machine learning. For the simple case of comparing two models, such as fitting a distribution with one or two exponentials, a hypothesis test such as likelihood ratio test can be used[61]. In the case of testing multiple models, we can apply information theory. Here, we acknowledge that no model is fully representing the underlying data but aim to minimize the amount of information lost from the simplified parameters. A variety of models are explored and then ranked by the amount of information lost using Akaike Information Criterion (AIC)[80] or Bayesian

---

[v] Also known as the *bias-variance problem* in machine learning.

Information Criterion (BIC)[81]

$$\begin{gathered} AIC = 2k - 2 * LL \\ BIC = k \log(N) - 2 * LL \end{gathered} \qquad \text{Eq. 1-13}$$

where k is the number of components (i.e. states), N is the number of data points, and LL is the loglikelihood of the model. The model with the lowest AIC or BIC value is chosen as the representative model that maximizes the trade-off between goodness of fit and complexity. As shown in **Eq. 1-13**, the primary difference between AIC and BIC is the penalty placed on the number of components, with BIC more likely to favor simpler models than AIC.

1.2.2.7 Trajectory idealization

As mentioned in **1.2.2.5**, real experimental data from single-molecules contains noise relevant to the mode of detection that can obscure the true states and transitions from the observer. Statistically determining the underlying significant states and their transitions is performed in a process commonly called *time-series idealization*. Idealization is a broad class of learning algorithms that essentially merge signal processing and model selection with the aim of uncovering the true sequence of molecular behavior from noisy traces by assigning each data point to a discrete state. The results of idealization are often directly used for dwell time or kinetic analysis; therefore, accurately identifying states and transitions is important and remains an active area of investigation within the community[82].

There are a large number of algorithms and software for single-molecule idealization which can be broadly classified into a few paradigms including unsupervised statistical leaning (i.e. change-point detection and clustering)[83-88], hidden Markov models[89-95], and, recently, neural networks[96-98]. The topic of which software is the best is routinely discussed in the community and is thoroughly highlighted in **Chapter 3**[63,64,82]. Therein, I also detail my own attempts at improving idealization by designing a new algorithm adapting unsupervised machine learning[99].

1.2.3 Single-molecule approaches to ligand gating

1.2.3.1 Conductance

Returning to experimental work with knowledge on how to analyze single-molecule kinetics, we can now discuss common modes of collecting single-molecule data of ion channel behavior. Single-channel conductance measurements have been a powerful tool for resolving complex ion channel behavior since the first measurement of single acetylcholine receptors in 1976[54]. Here, the resolution of current through a channel can be measured with a direct read-out of channel opening, closing, and potential intermediate sub-conductance states. Early analysis of single-channel measurements laid foundational groundwork for the statistical treatment of single-molecule analysis as we know it today (see **1.2**)[61,100]. Although a full review of the applications of single-channel recordings to physiology is beyond the scope of this chapter, there are key examples studying ligand binding mechanisms at the single-channel level worth highlighting. For example, dwell time analysis of single-channel recordings of nicotinic receptors in response to the binding of different partial agonists revealed a “flipped” or “primed” conformational state[101,102]. Mechanistically, this intermediate state follows the binding of the agonist but precedes channel opening which may explain the underlying differences in efficacy of partial agonists on channel opening[101,102]. Additionally, single-channel measurements of acetylcholine receptors directly revealed a “dock, catch and hold” mechanism upon ligand binding reminiscent of early models proposed by Del Castillo and Katz in 1957 from ensemble measurements[74,103]. Combined, these and other studies demonstrate that the seeming simple processes of ligand binding (i.e. **Scheme 1-1**) may involve more complex and intermediate steps which can be directly resolved by single-molecule approaches.

1.2.3.2 Fluorescence

In recent years however, many other single-molecule approaches have gained traction in ion channel studies for their ability to probe electrically silent transitions not observed by conductance measurements. In particular, single-molecule fluorescence has emerged as a powerful and flexible tool for biological physics[48,49,104]. The popularity of these approaches is in part due to the low barrier of entry owing to the standardization of optical set-ups and labeling approaches, reliable analysis software, and high reproducibility between labs[47,62,63,78,104-106].

Single-molecule fluorescence studies of ion channels commonly come in two general paradigms: stoichiometry and dynamics[107]. In stoichiometry experiments, the subunit composition of ion channels can be determined by counting the number of discrete photobleaching steps from a tag on each subunit, providing insight into heterogeneity of molecular assembly[108-110]. Dynamic experiments often resemble single-ion channel recordings and report on the conformational behavior of a protein over time. Typically, intramolecular single-molecule Förster resonance energy transfer (smFRET) is employed to monitor the conformation or interaction of subunits over time. Unlike single-ion channel recordings, smFRET is not limited to a specific class of proteins and therefore has widespread use in biological physics[47].

In a typical intramolecular smFRET experiment, two domains of the protein can be labeled with either a FRET acceptor or FRET donor (i.e. Cy3 and Cy5). Upon excitation at the donor wavelength, the excited donor molecule can non-radiatively transfer energy to the acceptor molecule within a ~10 nm range to excite it.[111] The efficiency of energy transfer (E_{FRET}) is given by

$$E_{FRET} = \left[1 + \left(\frac{R}{R_0}\right)^6\right]^{-1} \qquad \text{Eq. 1-14}$$

where R is the inter-dye distance and R_0 (Förster radius) is the distance at which $E_{FRET} = 0.5$ for the given FRET pair[111]. In practice, smFRET have been mostly standardized to handful of robust FRET pairs such that R_0 is a known quantity[105]. Therefore E_{FRET} is commonly computed as

$$E_{FRET} \approx \frac{I_A}{I_A + I_D} \qquad \text{Eq. 1-15}$$

where I_A and I_D are the acceptor and donor fluorescence intensity at each point in time, respectively[47]. Therefore, an intramolecular smFRET experiment can be performed to measure distances of protein structure or monitor the conformational dynamics of a molecule over time. Recent applications of smFRET experiments to full-length purified ion channels include measuring the pore size of a mechanosensitive channel[112], quantifying domain specific dynamics of potassium channels during gating[113,114], and analyzing the conformational landscape of receptors following agonist binding[115-117].

1.2.4 The binding-gating conundrum

Taken together, single-ion channel recordings and smFRET measurements have fundamentally changed our understanding of ion channel gating. However, these approaches have notable limitations when considering ligand binding studies to HCN channels. While single-ion channel recordings of HCN channels are possible, the very low conductance (~1 pS for HCN2) make these experiments exceptionally difficult to perform[118]. Additionally, both paradigms only report on conformational changes of the channel in response to a ligand; thus, the actual ligand binding event is not observed, but rather, inferred. The need to overcome this so called “binding-gating conundrum” echoes back to Wyman’s classic work on allostery, wherein he postulates that the most rigorous estimate of ligand activation thermodynamics is the direct measurement of the binding of the ligand itself[119]. Therefore, there is a pressing need for single-molecule techniques to directly monitor ligand binding to HCN channels for thorough mechanistic analysis. While this experiment is feasible through intermolecular single-molecule fluorescence paradigms[120], these too present new challenges due to the low affinity of cyclic nucleotide binding.

1.3 ZMWs for high-concentration single-molecule measurements

1.3.1 The concentration barrier of single-molecule fluorescence microscopy

Despite their wide-spread use, single-molecule fluorescence approaches come with an experiential caveat: in order to successfully observe a single-fluorescence species beyond background, a low concentration of fluorescence species is required. To appreciate this so called "concentration barrier", consider the following scheme that single-molecule fluorescence is uniquely suited to address: observing the interaction of two molecules to form a complex. For simplicity, lets name these molecules A and B and their complex AB (**Scheme 1-2**).

$$A + B \rightleftharpoons AB \qquad \text{Scheme 1-2}$$

This is a broad class of experiments encompassing protein-ligand associations, protein-protein interactions, chemical catalysis. Our aim is to learn about the mechanism driving AB formation. At the single-molecule level, the experiment can be performed by labeling A, B, or both A and B with fluorescent tags and directly quantitating the dwell times of each kinetic state (e.g. τ_{on} for the time spent in the associated state and τ_{off} for the time spend dissociated). Typically, one of these species will be immobilized on a surface to extend their observation time, though this is not always necessary[121-123]. Therefore, let us consider the simple case wherein A is bound to the surface without a fluorescence label and a fluorescently-labeled B is in solution; thus, AB formation is reported via fluorescence emission of B[vi]. However, the solution-phase population of fluorescent molecules can provide a potentially overwhelming amount of background light at an emission wavelength that prevents simple wavelength discrimination of single B molecules. Consequently, the magnitude of this background fluorescence is dependent on the concentration of B which in turn bounds the resolution of AB formation.

The following sections 1.3.1 and 1.3.2 are adapted from a manuscript in preparation.

White, D.S.[1,2], Smith, M.A.[1], Chanda, B.[2], Goldsmith, R.H.[2]

[1] Department of Neuroscience, University of Wisconsin-Madison
[2] Department of Chemistry, University of Wisconsin-Madison

All authors contributed to the writing of the manuscript

[vi] If we were to consider the more difficult experiment, we would simply identify the fluorescence signal of B that overlaps with the fluorescence of A- a paradigm often referred to as co-localization microscopy.

To numerically define the concentration barrier of single-molecule fluorescence, we need to consider the Abbe diffraction limit of light. In general, the optimal focus of a spot is diffraction limited of the lateral ($d_{x,y}$) and axial (d_z) planes and is dependent on the wavelength of excitation (λ) and the numerical aperture (NA) of the objective by[124]

$$d_{x,y} = \frac{\lambda}{2\,NA} \qquad\qquad d_z = \frac{2\lambda}{NA^2} \qquad\qquad \text{Eq. 1-16}$$

For typical single-molecule fluorescence, the diffraction limited volume is roughly 250 nm x 250 nm x 600 nm at high NA for visible photons, which is still substantially larger than a single molecule[42,105,125]. Thus, it is the diffraction limit of focused light that is the origin of the single-molecule concentration barrier, where the signal from a single bound molecule cannot be differentiated from the fluorescence of other freely floating molecules within the observation volume of the diffraction limited spot. For a confocal detection scheme, the optimal resolution corresponds to a fL (1 fL = 1 x10^{-15} L) observation volume[125]. A single molecule in this volume has an effective concentration of 2 nM, above which single-molecules cannot be individually resolved. However, single-molecule biophysicists typically deploy total-internal reflection fluorescence microscopy (TIRFM) to further reduce the fluorescence background[48,49]. Here, incident light is totally internally reflected at the glass-water interface which generates an evanescent field with an axial depth of ~100 nm to selectively excite molecules close to the glass surface[126]. TIRFM decreases the observation volume to ~40 aL (1 aL = 1 x 10^{-18} L) which raises the maximum allowable concentration to 40 nM[127]. In practice, fluorescence concentrations of ≤ 10 nM are considered the upper limit of TIRFM to ensure optimal signal-to-noise ratios for single-molecule detection[128]. For the following discussion, we will numerically assert the concentration barrier of single-molecule fluorescence at 10 nM.

Returning to our hypothetical experiment in **Scheme 1-2**, the ability to detect AB is bounded by the concentration of B owing to the inherent concentration barrier (**Figure 1-5**). Measurements of high affinity interactions (e.g. $K_d \leq 10$ nM) can be successfully performed by exciting only specifically associated

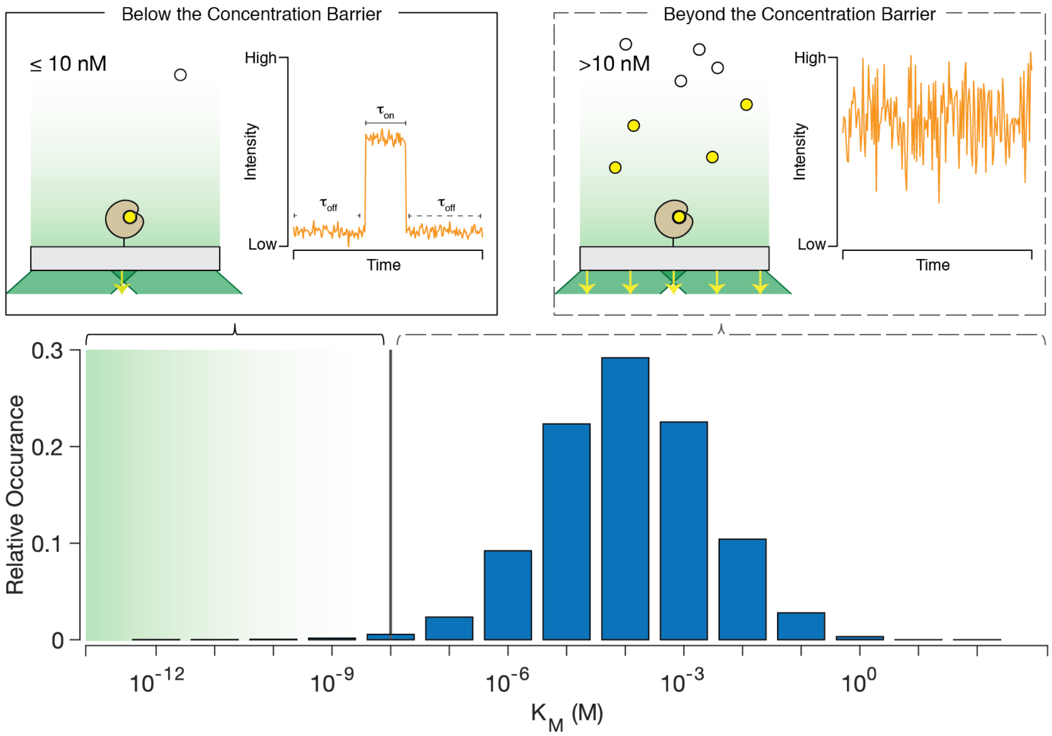

***Figure 1-5: The concentration barrier of single-molecule fluorescence measurements**. (Top left) Cartoon showing a surface tethered receptor binding a fluorescently-labeled ligand. At low ligand concentrations, individual binding events can be resolved by TIRFM. (Top right) At concentrations exceeding the concentration barrier, TIRFM excites freely diffusing fluorescent-ligands in solution, thereby obscuring the resolution of specific binding events. (Bottom) Histogram of over 30,000 enzyme affinities obtained from the BRENDA database[139]. The vertical black line indicates the 10 nM concentration barrier of TIRFM.*

AB complexes without a large background signal from the freely diffusing B molecules in the bulk solution (**Figure 1-5,** top-left). However, weak affinity interactions (e.g. K_d > 10 nM) will require a higher concentration to drive the formation of AB. This inevitably leads to the excitation of freely diffusing B in the observation volume and thus overwhelms our ability to monitor single binding dynamics (**Figure 1-5,** top right). The challenge faced by single-molecule biophysicists is that the majority of physiological interactions occur beyond the concentration limit set by confocal or TIRFM. For example, most enzymes have a Michaelis-Menten constant (K_M) in the micromolar to millimolar range which means TIRFM can only access 0.35% of interactions (**Figure 1-5**, bottom and **Table 1-1**)[139]. The majority other biological interactions are out of reach as well, including protein-protein or DNA-protein interactions[120]. Multi-subunit proteins such as ion channels aggravate these issues as they often feature multiple binding sites

Table 1-1: The concentration limit of single-molecule fluorescence techniques

Technique	Max Concentration Observed (µM)	Enzymes Accessible by K_M (%)
Confocal[49,125]	0.002	0.14
TIRFM[125,128]	0.01	0.35
PhADE[129]	1	4.7
CLIC[130]	2	7.0
Nanovesicle Trapping[131]	3	8.8
smFRET[132]	10	16.6
ZMW[133]	10	16.6
Antenna-in-box[134]	20	22.5
DNA-Origami Nanoantenna[135,136]	25	24.9
sm-PAFRET[137]	30	27.3
ZMW-FRET[138]	1000	70.2

for lower-affinity ligands, modulators, or accessory proteins that all need to be simultaneously resolved. The concentration barrier is therefore not only an experimental compromise, but a wall that prevents studying weak affinity interactions at the single-molecule level.

1.3.2 Techniques for breaking the concentration barrier

As the single-molecule concentration barrier restricts the studies of biological interactions to high-affinity processes, there has been tremendous effort in the community to develop techniques capable of overcoming this limitation. As thorough reviews of this topic exist elsewhere[125,127], we will briefly discuss effective methodologies that have enable single-molecule studies of weak-affinity interactions. These approaches can be broken into distinct classes including fluorescence modulation, restriction of the detection volume, and fluorescence enhancement. While each class is independently successful, they are often combined to achieve even higher single-molecule concentrations. A list of techniques and their experimentally demonstrated concentration limit is shown in **Table 1-1**. This section is not meant to be an exhaustive list of all approaches to overcome the 10 nM limit, but rather a survey to highlight the progress and challenges of single-molecule fluorescence studies at physiologically concentrations.

1.3.2.1 Fluorescence modulation

The first class of high-concentration techniques is fluorescence modulation. Following **Scheme 1-2**, the association of A and B leads to a spectrally separated product, AB. In the optimal scenario of this paradigm, the signal of AB is not contaminated by the background signals of A or B, which thereby removes the concentration dependence of the reactants on the product identification. A classic example of this methodology is seen in the use of fluorogenic reactions, wherein the catalytic activity between a non-fluorescence enzyme A and substrate B can be monitored by the generation fluorescent product[140-143]. For example, English and colleagues revisited Michaelis-Menten kinetics by monitoring enzymatic turn-over of the non-fluorescent substrate resofurin-ß-D-galactopyranoside to the fluorescent resofurin at single ß-galatosidase molecules[144]. Unfortunately, the generalizability of this approach is limited by the number of fluorogenic reagents and chemical modifications available for association reporting[46].

A natural extension of fluorogenic reactions is two-species (intermolecular) smFRET[120]. Following **Scheme 1-2**, molecule A can be labeled with a FRET acceptor (e.g. Cy5) and B with a donor (e.g. Cy3) such that AB formation emits a FRET signal upon excitation at the donor wavelength and collected at the acceptor wavelength. Since the acceptor fluorophore remains mostly dark at the wavelength of the donor fluorophore maximum excitation, the FRET signal is easily detectable above background. Further, as the FRET signal is dependent on the Förster radius (R_0) of the acceptor and donor dyes, a FRET signal is only generated when the two species are in close proximity (2 – 8 nm)[105], which effectively defines a new observation volume with radius R_0. This distance dependence makes FRET less sensitive to diffusing fluorescent molecules or non-specific absorption to a glass surface than conventional TIRFM. Using a quantum-dot as a FRET acceptor to monitor ATP hydrolysis by single myosin V molecules, Sugawa and colleagues reported FRET signal at up to 10 µM of Cy3-ATP, 1,000-fold higher than the concentration limit [132]. As will be further discussed below, FRET is frequently used as a reporting paradigm for other high-concentration modalities.

A final example of fluorescence modulation is the use of photoactivation. Similar to certain types of super-resolution approaches, incorporating photoswitchable fluorophores for isolating signals of interest are a promising avenue for overcoming the concentration barrier[145]. An excellent demonstration of this idea is termed PhADE (PhotoActivation, Diffusion, Excitation), which utilizes the photoconvertible protein mKikGR that fluorescence green upon excitation at 488 nm in its ground state (mKikG) but emits red upon 568 nm excitation in its activated state (mKikR) following a brief illumination pulse at 405 nm[129,146]. In this scheme, a high concentration of a mKikGR-coupled protein of interest (B^{mKikGR}) can bind to a receptive molecule of interest tethered on the surface (A). Following activation at 405 nm, unbound B^{mKikR} will diffuse away leaving only associated complexes (AB^{mKikR}) to be monitored at 568 nm excitation. This technique has been demonstrated at low micromolar concentrations of mKikGR labeled species and can take advantage of a growing number of photoactivatable fluorophores[129]. An extension of this by approach was developed by Peng and co-workers who adapted a photoactivable dye that could be used a FRET donor in its activated form[137]. Compared to PhADE, single-molecule photoactivation FRET (sm-PAFRET) enables faster temporal resolution and was demonstrated at up to 30 µM concentrations.

1.3.2.2 Reduction of the physical observation volume

The next class of successful high-concentration techniques is the reduction of the physical observation volume. In this paradigm, the physical space a single-molecule can occupy is reduced, effectively increasing the concentration of the molecule[147]. Nanovesicle trapping is one such approach whereby molecules are confined in lipid vesicles around 100 nm diameter to yield a physical volume of ~100 zL (1 zL = 1 x 10^{-21} L)[148]. A single molecule contained in a vesicle of this volume has an effective concentration of 3 µM, far exceeding the 10 nM concentration barrier. Importantly, while the actual observation volume is no smaller than typical fluorescence measurements, the reduced physical volume does not include more fluorophores than necessary which minimizes background signal. Using nanovesicle trapping with intermolecular FRET as a reporter, Benítez and colleagues revealed two distinct interaction complexes between the copper chaperone Hah1 and the Wilson disease protein at up to 3 µM[131]. In addition, DNA and RNA sequence recognition was examined as function of mismatch position with K_d values up to 100 µM[149,150].

An alternative approach for reducing the physical volume to overcome the concentration barrier is the use of convex-lens induced confinement (CLIC)[130]. In this crafty approach, a planoconvex lens is pushed against a flat cover glass, creating a lens-cover glass gap on the scale of tens of nm. The reduction in the vertical dimension of this gap decreases the physical volume by 20-fold compared to the TIRFM regime. An advantage of the CLIC device is its ease of implementation[130]: a CLIC device can be assembled from conventional wide-field microscopes and does not require complicated optics, biochemistry, or nanofabrication to achieve high-concentration measurements. The restricted geometry efficiently reduces the background concentration such that a single-molecule can be observed at up to 2 µM. This method also extends the observation period of freely diffusing molecules by up to 10^4-fold compared to TIRFM. Applications of the CLIC technique have been diverse, including measurements of freely diffusing particles and vesicles[130], monitoring the processive motion of myosin V and VI along actin filaments[151], and quantifying the free energy of confinement of semiflexible polymers[152].

1.3.2.3 Fluorescence enhancement

The final class of general techniques for high-concentration single-molecule measurements is fluorescence enhancement. Unlike the previous two techniques which aimed to reduce background by spectrally shifting the desired signal or shrinking the physical observation volume for effective concentration increases, this class of methods instead amplifies the fluorescence signal in a small region of space for detection above the high background fluorescence. In particular, enhancement is often achieved via use of metal nanoantennas. Upon illumination, the conduction electrons of the metal nanoparticle can be driven into collective oscillations called localized surface plasmon resonances. These excitations give rise to a sub-wavelength localization of electromagnetic energy which forms high-intensity “hot-spots” that can enhance the fluorescence of molecules placed near the nanoparticles[153].

An exemplary demonstration of nanoantenna for single-molecule detection was performed by Kinkhabwala and colleagues using gold bowtie nanoantennas with a 10 nm gap[154]. Here, a low quantum yield dye was contained within the zepto-liter volume hotspot wherein its fluorescence was enhanced by a factor of 1,340. A similar approach was taken by Punj and colleagues where they built gold nanoantennas inside of nanofabricated rectangular apertures composed of a 50 nm thick gold film on a glass coverslip[134]. This ‘antenna-in-box’ platform led to 11,000 times enhancement of a low quantum yield dye and could identify single-diffusing molecules in the hot-spot in concentrations of dye up to 20 µM[134]. To overcome the stochastic arrival of a single-molecule through the hot-spot, Acuna and colleagues developed a DNA origami[155] platform with a specific docking site for molecules of interest in optimal proximity to gold nanoparticles[135]. The current generation these self-assembled DNA origami nanoantennas showed a 5,000-fold increase in fluorescence for detection of single-molecules in a 25 µM background[136]. Despite the high concentrations achievable, fluorescence enhancement paradigms are restricted by the small gap-size between gold nanoparticles; therefore, steric hinderance of proteins, viruses, or macromolecules can be prohibitive and may limit widespread adoption[156].

1.3.3 Zero-mode waveguides

As demonstrated with TIRF, decreasing the excitation volume is a powerful approach for increasing the maximum observable concentration. The most popular example of this paradigm is the zero-mode waveguide (ZMW) developed by Levene and coworkers[133]. These nanophotonic devices are typically 50-200 nm dimeter cylindrical holes in a 100 nm film of metal on a glass or quartz surface (**Figure 1-6**). Given that the hole is smaller than the diffraction limit of light, the unique physical properties of

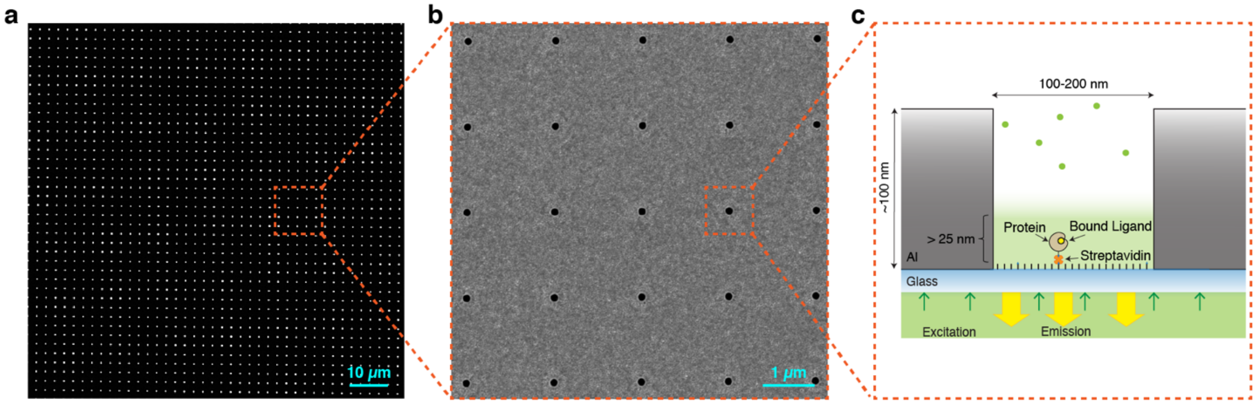

Figure 1-6: Zero-mode waveguides. *(**a**) Representative bright field image of a ZMW array. There are approximately 1,600 ZMWs in the field of view imaged with a 512x512 EMCCD and a 100x objective. Scale bar is 10 µm. (**b**) Representative scanning electron microscopy (SEM) image of a ZMW array. Scale bar is 1 µm. (**c**) Cartoon showing a high concentration of a fluorescently-labeled ligand binding to a surface tethered receptor inside a ZMW. The observation volume decays rapidly from the surface (~25 nm). Only ligands within the small observation volume are observed.*

these devices offer the capability of high-concentration measurements. In addition, fabricated device features thousands of ZMWs, enabling high-through single-molecule measurements.

In the ideal case, a cylindrical nanopore exhibits a cut-off wavelength (λ_c) above which no prorogating modes exist inside the waveguide (i.e., light can no longer propagate). The value λ_c is directly related to the shape and size of the guide. Wavelengths longer than λ_c are evanescent meaning their intensity decays exponentially along the z-axis inside the waveguide (z),

$$I(z) = e^{-\mathrm{z}/\Lambda} \qquad \text{Eq. 1-17}$$

where z is the distance into the pore and Λ is the decay constant defined as

$$\Lambda = \frac{1}{2}\sqrt{\frac{1}{\lambda_c^2} - \frac{1}{\lambda^2}} \qquad \text{Eq. 1-18}$$

Because no propagating modes exist, these cylindrical holes were named “zero-mode” waveguides. The rapid decay of illumination incident to the entry of the waveguide can be as low as z = 30 nm for 50 nm ZMWs[133]. Further, the detection volume is also reduced based on the antenna behavior of the ZMW where dipole emitters are radiatively coupled to free space collection optics over an even smaller total volume (see **1.3.2.3**)[157]. Together, these conditions create an effective observation volume within the waveguide around 20 zL, orders of magnitude lower than TIRF or confocal regimes.

The maximum concentration recorded in ZMWs is 10 µM which is 1,000-fold above the concentration barrier of single-molecule fluoresence[133]. As will be discussed later, we have further combined ZMWs with an intermolecular FRET reporting scheme (ZMW-FRET) to observe protein-ligand associations at up to 1 mM concentration, 100,000-fold above the fluorescence concentration barrier (see **Chapter 2**)[138]. At the time of this dissertation, this remains the highest concentration from which a single molecule has been identified by fluorescence.

1.3.4 ZMW preparation

1.3.4.1 ZMW Fabrication

The fabrication of ZMWs is inherently challenging due to the small size of each aperture. ZMWs were first fabricated using positive-tone electron-beam lithography (**Figure 1-7a**)[133]. Here, ~100 nm of aluminum was thermally evaporated on fused silica substrates, followed by deposition of a positive-tone photoresist via spin-coating. The substrate was patterned using electron-beam lithography and the apertures were formed using a reactive ion etch into the aluminum layer. While this approach can create robust ZMW devices, its difficulty is accentuated from the careful tuning the metal etch step. To overcome this challenge, an alternative approach using a negative-tone photoresist was developed (**Figure 1-7b**)[158]. In this scheme, a photoresist pillar is left standing on the fused silica substrate following electron-beam exposure and development. Metal is then deposited onto the substrate and the photoresist pillar is removed via a metal lift-off step[vii]. The lack of a metal etching step grants greater flexibility in the choice

---

[vii] In my own hands, I found the positive-tone lithography process far more reliable than negative-tone due to issues of photoresist pillars not adequately adhering to the cover glass when fabricating diameters < 200 nm. See **Chapter 7** for a detailed protocol of ZMW fabrication and passivation using positive-tone electron beam lithography.

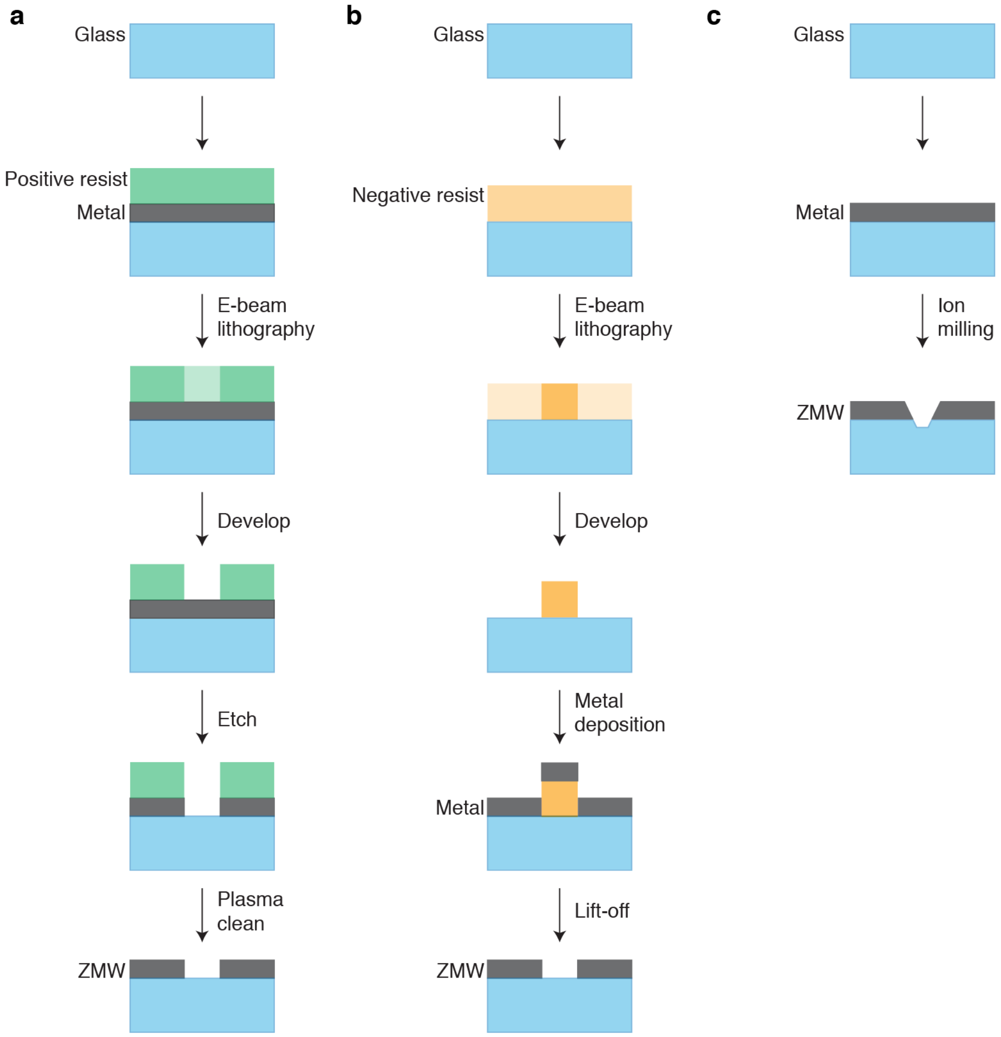

Figure 1-7: Fabrication of ZMWs. *Cartoons showing the general steps to fabricate ZMWs using positive-tone (**a**) and negative-tone (**b**) electron-beam (E-beam) lithography and FIB (**c**).*

of metal cladding. While aluminum is most commonly used, gold ZMWs have shown fluorescence enhancement for red-shifted dyes[159]. In addition, mixed metal ZMWs have shown tunable plasmonic features that can alter photophysical properties of various fluorophores[160,161]. Negative-tone resists are also compatible with deep-UV lithography to increase ZMW chip production to the wafer-scale by circumventing the need of serial electron-beam writing[158].

Various other approaches have been successful for fabricating ZMWs including nanosphere lithography[162], microlithography paired with electrodeposition[163], and chemical etching[164]. In particular, focused ion beam (FIB) is a popular approach that circumvents the use of photoresists (**Figure 1-7c**)[165].

Direct ion milling offers precise control over the feature size in various metallic layers at the cost of dramatically lower fabrication rate from milling each nanoaperture individually[166]. ZMWs fabricated by FIB have a distinct conical shape, as opposed to the standard cylindrical wells, which feature an etch into the silica layer that affects the confinement of molecules and may lead to fluorescence enhancement[167].

1.3.4.2 ZMW passivation

Prior to use, ZMWs require a surface treatment (passivation) to prevent the non-specific absorption of biomolecules and fluorophores onto the metal walls or silica floor[168]. One common passivation method in is the physical absorptions of bovine serum albumin (BSA) which adheres to the untreated surfaces and prevents the non-specific binding of other molecules[133]. Silica surfaces in single-molecule experiments are commonly functionalized with a poly(ethylene glycol) (PEG) layer[105,169,170]. For biophysical experiments in particular, biotinylated-PEG molecules are deposited onto silica surface using silane chemistry, enabling subsequent deposition of streptavidin for single-molecule pull-downs of biotinylated biomolecules[171]. Finally, passivation of the metallic cladding has shown dramatic reductions in non-specific absorption. Aluminum-based ZMWs can be selectively passivated via aluminophosphoate chemistry using poly(vinylphosphonic acid) (PVPA)[172]. Gold-based ZMWs offer greater flexibility in passivation options, including the formation of self-assembled monolayers via use of Au-thiol chemistry[164,173,174].

1.3.5 ZMW applications

Compared to other high-concentration single-molecule techniques, ZMWs are frequently used with broad applications in field[157,166]. The popularity of ZMWs is in part due to the commercial technology of single-molecule real time (SMRT) sequencing from Pacific Biosciences which provide high-throughput and long-read DNA sequencing with direct impact on medical diagnostics [169,175,176]. Beyond sequencing, ZMWs have been used to explore a variety of biophysical and biochemical processes at high nM to low µM concentrations, including translation events at individual ribosomes[177,178], dynamics and stoichiometry of membrane-bound proteins[179,180], protein-protein interactions[167], and protein-ligand interactions[32] (the central theme of this work). The architecture of ZMWs has also been altered for novel applications. For example, ZMWs have been coupled with nanopore technology to generate a hybrid device that reduces the sample requirements of SMRT sequencing[181,182]. ZMWs have been adapted into microfluidic devices[183,184] and have also been combined with plasmonic nanosized antennas to gain the advantages of fluorescence enhancement[134]. Additionally, ion milling into recessed dual ring electrodes enabled repeated oxidation and reduction of a single-molecule between two closely spaced working electrodes for electrochemical studies[185].

1.4 Preliminary single-molecule study of HCN2 CNBD dynamics

In our own hands, we have found ZMWs to be powerful tools for resolving complex binding mechanism of weak affinity interactions[32]. Our initial experiments measuring a fluorescently-labeled cAMP to HCN2 CNBDs were motivated by Dr. Marcel P. Goldschen-Ohm in the Chanda lab with support from the Goldsmith group[viii]. Rather than tackling the complexity of recording from a full-length ion channel at the onset, single CNBDs from HCN2 were first isolated for to explore the mechanism of elementary ligand associations events. To overcome the concentration barrier and monitor the low affinity ($K_d \approx 1$ µM) association events[36], a novel approach utilizing an intermolecular FRET reporter

---

[viii] I was particularly fortunate to join the project just as the optimized experimental conditions to reproducibly collect data from commercially available ZMWs was optimized. I contributed to the completion of this project by collecting a large portion of the single-molecule data.

scheme inside ZMWs was conceived. As both the methodology and conclusions from this study serve as the basis my research hereafter, I will briefly describe our approach and key findings.

To facilitate single-molecule studies in ZMWs, A CNBD mutant was generated to enable the conjugation of a FRET acceptor (DyLight 650) for monitoring the binding of the DyLight 549-labeled cAMP donor molecule (fcAMP[36]). Purified and biotinylated CNBDs were deposited into ZMWs with a streptavidin coated surfaced and bathed with fcAMP (0.1-10 µM). Imaging was performed using an alternating excitation and emission scheme $(\lambda_{ex}/\lambda_{em})$[186] to simultaneously monitor single-molecule photobleaching (640nm/640nm), single-fcAMP diffusion (532nm/532nm), and FRET upon binding (532nm/640nm) (**Figure 1-8a**). This allowed us to find ZMWs containing single proteins and ensure the observed binding events were between the CNBD and fcAMP rather than non-specific absorption to the surface.

Dwell time distributions of unbound and bound events were generated following idealization of each fluorescence time-series[93]. As expected for a biomolecular binding process, unbound dwell times grew shorter with increasing ligand concentration while bound dwell times remained constant (**Figure 1-8b**). MLE of both bound and unbound dwell-time distributions both required two components to adequately fit the observed data. This result suggests underlying dynamic heterogeneity, wherein each CNBD can interconvert between two unique unbound (U1 and U2) and bound (B1 and B2) conformational states. To resolve a complete kinetic mechanism of fcAMP binding to HCN2 CNBD, we compared the likelihood of several hidden Markov models[78,79] using a statistical model selection approach (**see 1.2.2.6**)[80]. We selected a scheme featuring two unbound and two bound states with only one conformation being receptive to ligand binding (**Figure 1-8c**). The most direct interpretation of this result is that although a CNBD can exist in (at least) two different conformations when unliganded, only on one

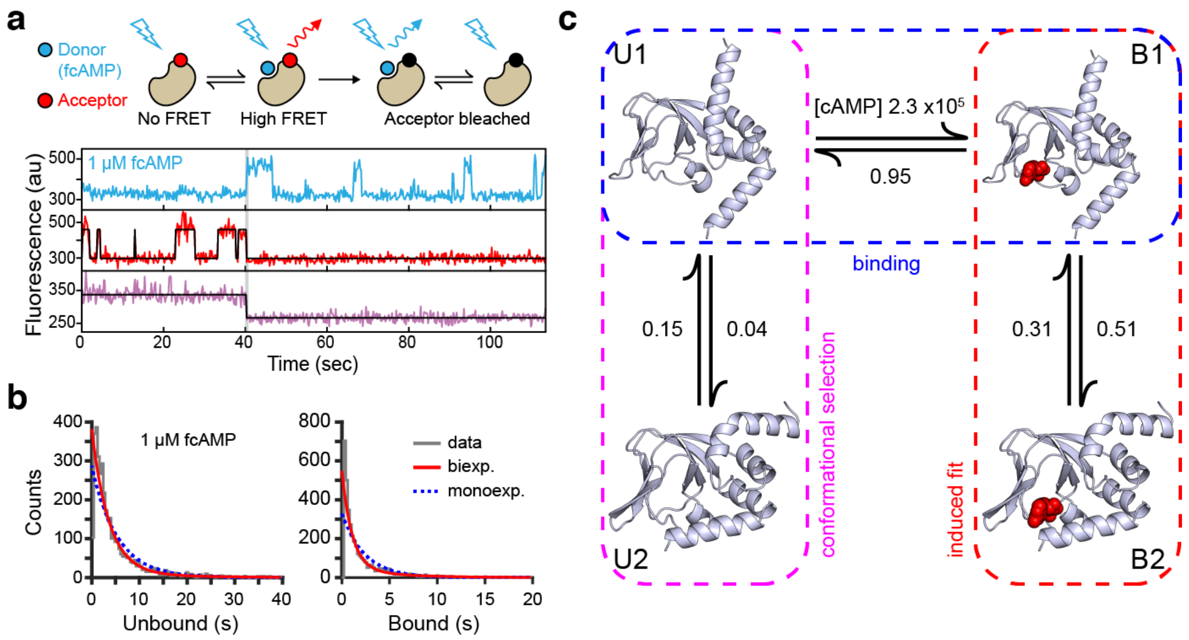

***Figure 1-8: Single-molecule binding dynamics to isolated HCN2 CNBDs**. (**a**, top) Cartoon depicting smFRET during fcAMP binding. Direct excitation of the donor fcAMP results in stimulated emission from the acceptor on the CNBD due to efficient FRET while the donor is bound up until the acceptor bleaches, after which only emission from the donor is observed. (**a**, bottom) Single-molecule fluorescence time series for fcAMP binding to individual acceptor-labeled CNBDs within ZMWs. Simultaneous emission from donor (blue) and acceptor (red) upon donor excitation at 532 nm was interleaved every other frame with emission from acceptor (magenta) upon direct excitation at 640 nm. Acceptor fluorescence for both excitation conditions is overlaid with the idealized time series (black). (**b**) Histograms of unbound and bound single-molecule dwell time distributions for events at 1 µM fcAMP overlaid with maximum likelihood estimates for monoexponential (blue dashed) and biexponential (red) distributions. (**c**) A structural model of cAMP (red spheres) binding dynamics at monomeric CNBDs from HCN2 channels. Rate constants (s^{-1} or $M^{-1}s^{-1}$) were optimized using HMM modeling of idealized single-molecule fcAMP binding time series. This figure has been reproduced and modified from reference 32 under use of a Creative Commons Attribution license. See https://creativecommons.org/licenses/by/4.0/.*

conformation is receptive to binding. Upon binding, the CNBD again can exist in (at least) two possible conformations, reminiscent of a "catch-and-hold" mechanism.

Overall, this study showcased the power of ZMWs for resolving ligand associations in physiologically relevant concentrations and demonstrated that HCN2 CNBDs exhibit complex binding dynamics. However, this study is limited from focusing on a single fragment of a large, tetrameric ion channel; therefore, the extent to which these metastable conformational states impact mechanism of CNBD tetramerization[34] or channel activation[37,39] and whether such behavior is conserved across HCN isoforms cannot be address by this result alone. Fortunately, these and other questions are directly investigated within this thesis.

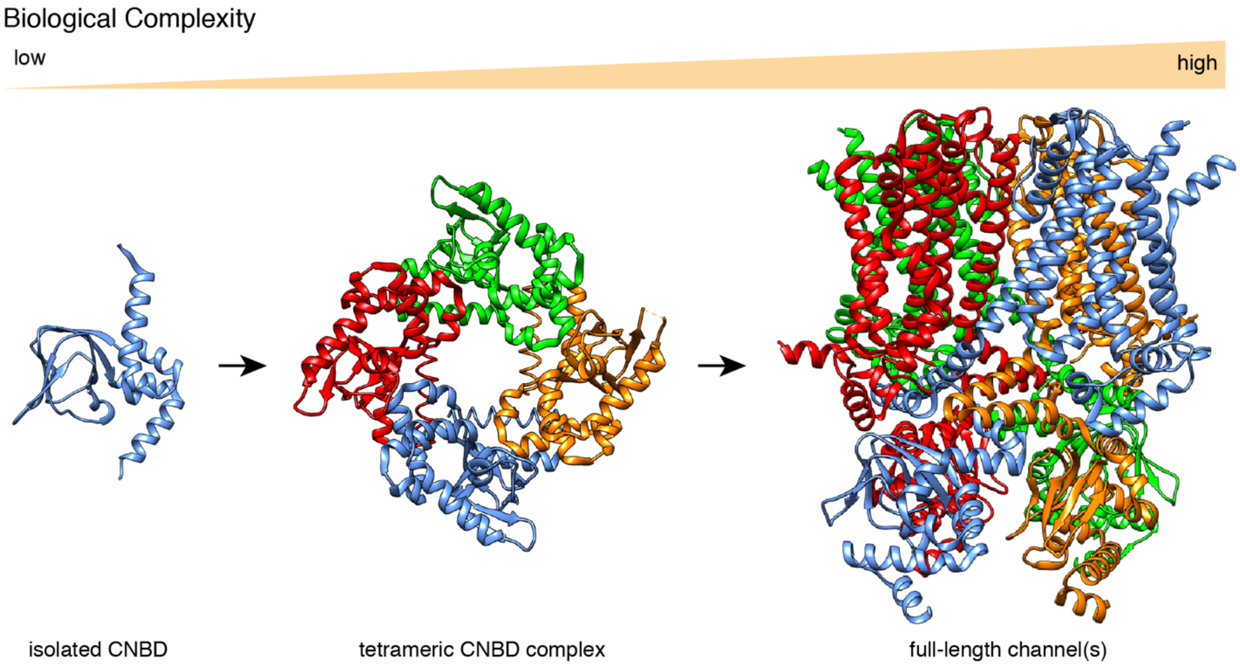

Figure 1-9: A reductionist approach to single-molecule HCN studies. *The general scheme of increasing biological complexity taken to determine binding mechanisms of cyclic nucleotides to (**a**) isolated CNBDs (PDB ID: 5JON), (**b**) tetrameric complex of CNBDs without a transmembrane or pore domains (PDB ID: 3U11) and (**c**) purified full-length and intact HCN channels (PDB ID: 5U6P).*

1.5 Dissertation overview

The research contained herein advances the initial studies of cyclic nucleotide binding to HCN2 CNBDs using single-molecule fluorescence in ZMWs through two parallel avenues: biological insight and technology development. On the biology side, we take a reductionist approach to investigative the elementary binding mechanisms by increasing protein complexity with each study (**Figure 1-9**). First, we continue to examine isolated CNBDs dynamics in response to another cyclic nucleotide (**Chapter 2**), then advance to a tetrameric complex of CNBDs to assess their interactions upon ligand binding (**Chapter 3**). We conclude by determining mechanisms of fcAMP binding to full-length HCN1 and HCN2 channels (**Chapter 4**). On the technology side, we expand the single-molecule concentration barrier by two orders of magnitude using ZMW-FRET (**Chapter 2**), develop a new single-molecule time-series analysis algorithm with faster computational speed and higher accuracy using unsupervised machine learning (**Chapter 3**), and showcase the potential for using a photocleavable DNA Origami construct to position molecules directly into the center of ZMWs (**Chapter 6**). Combined, these paths create new knowledge in the regulation of ion channels by ligands and (hopefully) lower the barrier of entry for new experimentalists to join single-molecule fluorescence world using ZMWs.

1.6 References

1. Bertil, H. & Bertil, H. Ion channels of excitable membranes. *Sunderland, Mass.: Sinauer Associates. 3rd Edition.–2007* (2001).
2. Goldschen-Ohm, M.P. & Chanda, B. SnapShot: Channel Gating Mechanisms. *Cell* **170**, 594-+ (2017).
3. Lemoine, D. et al. Ligand-Gated Ion Channels: New Insights into Neurological Disorders and Ligand Recognition. *Chemical Reviews* **112**, 6285-6318 (2012).
4. Ranade, S.S., Syeda, R. & Patapoutian, A. Mechanically Activated Ion Channels. *Neuron* **87**, 1162-1179 (2015).
5. Isacoff, E.Y., Jan, L.Y. & Minor, D.L. Conduits of Life's Spark: A Perspective on Ion Channel Research since the Birth of Neuron. *Neuron* **80**, 658-674 (2013).
6. Gauss, R., Seifert, R. & Kaupp, U.B. Molecular identification of a hyperpolarization-activated channel in sea urchin sperm. *Nature* **393**, 583-587 (1998).
7. Ludwig, A., Zong, X.G., Jeglitsch, M., Hofmann, F. & Biel, M. A family of hyperpolarization-activated mammalian cation channels. *Nature* **393**, 587-591 (1998).
8. Santoro, B. et al. Identification of a gene encoding a hyperpolarization-activated pacemaker channel of brain. *Cell* **93**, 717-729 (1998).
9. Wahl-Schott, C. & Biel, M. HCN channels: Structure, cellular regulation and physiological function. *Cellular and Molecular Life Sciences* **66**, 470-494 (2009).
10. DiFrancesco, D. The Role of the Funny Current in Pacemaker Activity. *Circulation Research* **106**, 434-446 (2010).
11. Difrancesco, D. & Tortora, P. DIRECT ACTIVATION OF CARDIAC-PACEMAKER CHANNELS BY INTRACELLULAR CYCLIC-AMP. *Nature* **351**, 145-147 (1991).
12. Ramirez, D., Zuniga, R., Concha, G. & Zuniga, L. HCN Channels: New Therapeutic Targets for Pain Treatment. *Molecules* **23**(2018).
13. Nolan, M.F., Dudman, J.T., Dodson, P.D. & Santoro, B. HCN1 channels control resting and active integrative properties of stellate cells from layer II of the entorhinal cortex. *Journal of Neuroscience* **27**, 12440-12451 (2007).
14. Nolan, M.F. et al. A behavioral role for dendritic integration: HCN1 channels constrain spatial inputs to distal dendrites memory and plasticity at of CA1 pyramidal neurons. *Cell* **119**, 719-732 (2004).
15. Nolan, M.F. et al. The hyperpolarization-activated HCN1 channel is important for motor learning and neuronal integration by cerebellar Purkinje cells. *Cell* **115**, 551-564 (2003).
16. Kim, C.S., Brager, D.H. & Johnston, D. Perisomatic changes in h-channels regulate depressive behaviors following chronic unpredictable stress. *Molecular Psychiatry* **23**, 892-903 (2018).

17. DiFrancesco, J.C. et al. Recessive Loss-of-Function Mutation in the Pacemaker HCN2 Channel Causing Increased Neuronal Excitability in a Patient with Idiopathic Generalized Epilepsy. *Journal of Neuroscience* **31**, 17327-17337 (2011).
18. Nava, C. et al. De novo mutations in HCN1 cause early infantile epileptic encephalopathy. *Nature Genetics* **46**, 640-645 (2014).
19. Noh, S. et al. The heart-rate-reducing agent, ivabradine, reduces mechanical allodynia in a rodent model of neuropathic pain. *European Journal of Pain* **18**, 1139-1147 (2014).
20. Young, G.T., Emery, E.C., Mooney, E.R., Tsantoulas, C. & McNaughton, P.A. Inflammatory and neuropathic pain are rapidly suppressed by peripheral block of hyperpolarisation-activated cyclic nucleotide-gated ion channels. *Pain* **155**, 1708-1719 (2014).
21. Postea, O. & Biel, M. Exploring HCN channels as novel drug targets. *Nature Reviews Drug Discovery* **10**, 903-914 (2011).
22. Tanguay, J., Callahan, K.M. & D'Avanzo, N. Characterization of drug binding within the HCN1 channel pore. *Scientific Reports* **9**(2019).
23. Lee, C.H. & MacKinnon, R. Structures of the Human HCN1 Hyperpolarization-Activated Channel. *Cell* **168**, 111-+ (2017).
24. Ulens, C. & Tytgat, J. Functional heteromerization of HCN1 and HCN2 pacemaker channels. *Journal of Biological Chemistry* **276**, 6069-6072 (2001).
25. Craven, K.B. & Zagotta, W.N. CNG and HCN channels: Two peas, one pod. *Annual Review of Physiology* **68**, 375-401 (2006).
26. Baruscotti, M., Bucchi, A. & DiFrancesco, D. Physiology and pharmacology of the cardiac pacemaker ("funny") current. *Pharmacology & Therapeutics* **107**, 59-79 (2005).
27. DeBerg, H.A., Brzovic, P.S., Flynn, G.E., Zagotta, W.N. & Stoll, S. Structure and Energetics of Allosteric Regulation of HCN2 Ion Channels by Cyclic Nucleotides. *Journal of Biological Chemistry* **291**, 371-381 (2016).
28. Alvarez-Baron, C.P., Klenchin, V.A. & Chanda, B. Minimal molecular determinants of isoform-specific differences in efficacy in the HCN channel family. *Journal of General Physiology* **150**, 1203-1213 (2018).
29. Wang, J., Chen, S. & Siegelbaum, S.A. Regulation of hyperpolarization-activated HCN channel gating and cAMP modulation due to interactions of COOH terminus and core transmembrane regions. *Journal of General Physiology* **118**, 237-250 (2001).
30. Wainger, B.J., DeGennaro, M., Santoro, B., Siegelbaum, S.A. & Tibbs, G.R. Molecular mechanism of cAMP modulation of HCN pacemaker channels. *Nature* **411**, 805-810 (2001).
31. Zagotta, W.N. et al. Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* **425**, 200-205 (2003).
32. Goldschen-Ohm, M.P. et al. Structure and dynamics underlying elementary ligand binding events in human pacemaking channels. *Elife* **5**(2016).
33. Taraska, J.W., Puljung, M.C., Olivier, N.B., Flynn, G.E. & Zagotta, W.N. Mapping the structure and conformational movements of proteins with transition metal ion FRET. *Nature Methods* **6**, 532-U94 (2009).
34. Lolicato, M. et al. Tetramerization Dynamics of C-terminal Domain Underlies Isoform-specific cAMP Gating in Hyperpolarization-activated Cyclic Nucleotide-gated Channels. *Journal of Biological Chemistry* **286**, 44811-44820 (2011).
35. Biskup, C. et al. Relating ligand binding to activation gating in CNGA2 channels. *Nature* **446**, 440-443 (2007).
36. Kusch, J. et al. Interdependence of Receptor Activation and Ligand Binding in HCN2 Pacemaker Channels. *Neuron* **67**, 75-85 (2010).

37. Kusch, J. et al. How subunits cooperate in cAMP-induced activation of homotetrameric HCN2 channels. *Nat Chem Biol* **8**, 162-9 (2011).
38. Ulens, C. & Siegelbaum, S.A. Regulation of hyperpolarization-activated HCN channels by cAMP through a gating switch in binding domain symmetry. *Neuron* **40**, 959-970 (2003).
39. Thon, S., Schulz, E., Kusch, J. & Benndorf, K. Conformational Flip of Nonactivated HCN2 Channel Subunits Evoked by Cyclic Nucleotides. *Biophys J* **109**, 2268-76 (2015).
40. Hines, K.E., Middendorf, T.R. & Aldrich, R.W. Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *Journal of General Physiology* **143**, 401-416 (2014).
41. Feynman, R.P. There's plenty of room at the bottom [data storage]. *Journal of microelectromechanical systems* **1**, 60-66 (1992).
42. Moerner, W.E., Shechtman, Y. & Wang, Q. Single-molecule spectroscopy and imaging over the decades. *Faraday Discussions* **184**, 9-36 (2015).
43. Kim, E. et al. A single-molecule dissection of ligand binding to a protein with intrinsic dynamics. *Nature Chemical Biology* **9**, 313-+ (2013).
44. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. & Ha, T. Advances in single-molecule fluorescence methods for molecular biology. *Annual Review of Biochemistry* **77**, 51-76 (2008).
45. Ding, S. & Sachs, F. Evidence for non-independent gating of P2X 2 receptors expressed in Xenopusoocytes. *BMC neuroscience* **3**, 17 (2002).
46. Cordes, T. & Blum, S.A. Opportunities and challenges in single-molecule and single-particle fluorescence microscopy for mechanistic studies of chemical reactions. *Nature Chemistry* **5**, 993-999 (2013).
47. Lerner, E. et al. Toward dynamic structural biology: Two decades of single-molecule Forster resonance energy transfer. *Science* **359**, 288-+ (2018).
48. Shashkova, S. & Leake, M.C. Single-molecule fluorescence microscopy review: shedding new light on old problems. *Bioscience Reports* **37**(2017).
49. Moerner, W.E. & Fromm, D.P. Methods of single-molecule fluorescence spectroscopy and microscopy. *Review of Scientific Instruments* **74**, 3597-3619 (2003).
50. Colquhoun, D. & Hawkes, A.G. The interpretation of single channel recordings. in *Microelectrode Techniques. The Plymouth Workshop Handbook, 2nd Ed. D. Ogden, editor. The Company of Biologists Ltd., Cambridge, UK* 141-188 (1994).
51. Lu, H.P., Xun, L.Y. & Xie, X.S. Single-molecule enzymatic dynamics. *Science* **282**, 1877-1882 (1998).
52. Xu, W.L., Kong, J.S., Yeh, Y.T.E. & Chen, P. Single-molecule nanocatalysis reveals heterogeneous reaction pathways and catalytic dynamics. *Nature Materials* **7**, 992-996 (2008).
53. Yildiz, A. & Selvin, P.R. Fluorescence imaging with one nanometer accuracy: application to molecular motors. *Accounts of chemical research* **38**, 574-582 (2005).
54. Neher, E. & Sakmann, B. SINGLE-CHANNEL CURRENTS RECORDED FROM MEMBRANE OF DENERVATED FROG MUSCLE-FIBERS. *Nature* **260**, 799-802 (1976).
55. Greives, N. & Zhou, H.X. Both protein dynamics and ligand concentration can shift the binding mechanism between conformational selection and induced fit. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 10197-10202 (2014).
56. Liu, D.T., Tibbs, G.R., Paoletti, P. & Siegelbaum, S.A. Constraining ligand-binding site stoichiometry suggests that a cyclic nucleotide-gated channel is composed of two functional dimers. *Neuron* **21**, 235-248 (1998).
57. Ruiz, M. & Karpen, J.W. Single cyclic nucleotide-gated channels locked in different ligand-bound states. *Nature* **389**, 389-392 (1997).
58. Chowdhury, S. & Chanda, B. Free-energy relationships in ion channels activated by voltage and ligand. *Journal of General Physiology* **141**, 11-28 (2013).

59. Sigg, D. & Bezanilla, F. A physical model of potassium channel activation: From energy landscape to gating kinetics. *Biophysical Journal* **84**, 3703-3716 (2003).
60. McManus, O., Spivak, C., Blatz, A., Weiss, D. & Magleby, K. Fractal models, Markov models, and channel kinetics. *Biophysical journal* **55**, 383-385 (1989).
61. Sakmann, B. *Single-channel recording*, (Springer Science & Business Media, 2013).
62. Tavakoli, M., Taylor, J.N., Li, C.B., Komatsuzaki, T. & Presse, S. SINGLE MOLECULE DATA ANALYSIS: AN INTRODUCTION. *Advances in Chemical Physics, Vol 162* **162**, 205-305 (2017).
63. Blanco, M. & Walter, N.G. ANALYSIS OF COMPLEX SINGLE-MOLECULE FRET TIME TRAJECTORIES. in *Methods in Enzymology, Vol 472: Single Molecule Tools, Pt a: Fluorescence Based Approaches*, Vol. 472 (ed. Walter, N.G.) 153-178 (2010).
64. Kinz-Thompson, C.D., Bailey, N.A. & Gonzalez, R.L. Precisely and Accurately Inferring Single-Molecule Rate Constants. in *Single-Molecule Enzymology: Fluorescence-Based and High-Throughput Methods*, Vol. 581 (eds. Spies, M. & Chemla, Y.R.) 187-225 (2016).
65. Maconochie, D.J. & Steinbach, J.H. The channel opening rate of adult- and fetal-type mouse muscle nicotinic receptors activated by acetylcholine. *Journal of Physiology-London* **506**, 53-72 (1998).
66. Buenrostro, J.D. et al. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nature Biotechnology* **32**, 562-+ (2014).
67. Horn, R. & Lange, K. ESTIMATING KINETIC CONSTANTS FROM SINGLE CHANNEL DATA. *Biophysical Journal* **43**, 207-223 (1983).
68. Qin, F. & Li, L. Model-based fitting of single-channel dwell-time distributions. *Biophysical Journal* **87**, 1657-1671 (2004).
69. Colquhoun, D., Hawkes, A.G. & Srodzinski, K. Joint distributions of apparent open and shut times of single-ion channels and maximum likelihood fitting of mechanisms. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences* **354**, 2555-2590 (1996).
70. Milescu, L.S., Yildiz, A., Selvin, P.R. & Sachs, F. Maximum likelihood estimation of molecular motor kinetics from staircase dwell-time sequences. *Biophysical Journal* **91**, 1156-1168 (2006).
71. Kaur, H., Jamalidinan, F., Condon, S.G.F., Senes, A. & Hoskins, A.A. Analysis of spliceosome dynamics by maximum likelihood fitting of dwell time distributions. *Methods* **153**, 13-21 (2019).
72. Armstrong, C.M. Interaction of tetraethylammonium ion derivatives with the potassium channels of giant axons. *The Journal of general physiology* **58**, 413-437 (1971).
73. McKinney, S.A., Declais, A.C., Lilley, D.M.J. & Ha, T. Structural dynamics of individual Holliday junctions. *Nature Structural Biology* **10**, 93-97 (2003).
74. Jadey, S. & Auerbach, A. An integrated catch-and-hold mechanism activates nicotinic acetylcholine receptors. *Journal of General Physiology* **140**, 17-28 (2012).
75. Zagotta, W.N., Hoshi, T. & Aldrich, R.W. SHAKER POTASSIUM CHANNEL GATING .3. EVALUATION OF KINETIC-MODELS FOR ACTIVATION. *Journal of General Physiology* **103**, 321-362 (1994).
76. Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**, 257-286 (1989).
77. Eddy, S.R. What is a hidden Markov model? *Nature Biotechnology* **22**, 1315-1316 (2004).
78. Nicolai, C. & Sachs, F. Solving ion channel kinetics with the QuB software. *Biophysical Reviews and Letters* **8**, 191-211 (2013).
79. Qin, F., Auerbach, A. & Sachs, F. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophysical Journal* **79**, 1915-1927 (2000).
80. Akaike, H. NEW LOOK AT STATISTICAL-MODEL IDENTIFICATION. *Ieee Transactions on Automatic Control* **AC19**, 716-723 (1974).
81. Schwarz, G. ESTIMATING DIMENSION OF A MODEL. *Annals of Statistics* **6**, 461-464 (1978).

82. Hadzic, M.C.A.S., Boerner, R., Koenig, S.L.B., Kowerko, D. & Sigel, R.K.O. Reliable State Identification and State Transition Detection in Fluorescence Intensity-Based Single-Molecule Forster Resonance Energy-Transfer Data. *Journal of Physical Chemistry B* **122**, 6134-6147 (2018).
83. Gnanasambandam, R. et al. Unsupervised Idealization of Ion Channel Recordings by Minimum Description Length: Application to Human PIEZO1-Channels. *Frontiers in Neuroinformatics* **11**(2017).
84. Watkins, L.P. & Yang, H. Detection of intensity change points in time-resolved single-molecule measurements. *Journal of Physical Chemistry B* **109**, 617-628 (2005).
85. Song, N. & Yang, H. Parallelization of Change Point Detection. *Journal of Physical Chemistry A* **121**, 5100-5109 (2017).
86. Li, H. & Yang, H. Statistical Learning of Discrete States in Time Series. *Journal of Physical Chemistry B* **123**, 689-701 (2019).
87. Shuang, B. et al. Fast Step Transition and State Identification (STaSI) for Discrete Single-Molecule Data Analysis. *Journal of Physical Chemistry Letters* **5**, 3157-3161 (2014).
88. Hill, F.R., van Oijen, A.M. & Duderstadt, K.E. Detection of kinetic change points in piece-wise linear single molecule motion. *Journal of Chemical Physics* **148**(2018).
89. McKinney, S.A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical Journal* **91**, 1941-1951 (2006).
90. Sgouralis, I. et al. A Bayesian Nonparametric Approach to Single Molecule Forster Resonance Energy Transfer. *Journal of Physical Chemistry B* **123**, 675-688 (2019).
91. Persson, F., Linden, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nature Methods* **10**, 265-269 (2013).
92. Qin, F. Restoration of single-channel currents using the segmental k-means method based on hidden Markov modeling. *Biophysical Journal* **86**, 1488-1501 (2004).
93. Bronson, J.E., Fei, J., Hofman, J.M., Gonzalez, R.L., Jr. & Wiggins, C.H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophysical Journal* **97**, 3196-3205 (2009).
94. van de Meent, J.W., Bronson, J.E., Wiggins, C.H. & Gonzalez, R.L. Empirical Bayes Methods Enable Advanced Population-Level Analyses of Single-Molecule FRET Experiments. *Biophysical Journal* **106**, 1327-1337 (2014).
95. Greenfeld, M., Pavlichin, D.S., Mabuchi, H. & Herschlag, D. Single Molecule Analysis Research Tool (SMART): An Integrated Approach for Analyzing Single Molecule Data. *Plos One* **7**(2012).
96. Xu, J.C. et al. Automated Stoichiometry Analysis of Single-Molecule Fluorescence Imaging Traces via Deep Learning. *Journal of the American Chemical Society* **141**, 6976-6985 (2019).
97. Celik, N. et al. Deep-Channel uses deep neural networks to detect single-molecule events from patch-clamp data. *Communications Biology* **3**(2020).
98. Misiunas, K., Ermann, N. & Keyser, U.F. QuipuNet: Convolutional Neural Network for Single-Molecule Nanopore Sensing. *Nano Letters* **18**, 4040-4045 (2018).
99. White, D.S., Goldschen-Ohm, M.P., Goldsmith, R.H. & Chanda, B. Top-down machine learning approach for high-throughput single-molecule analysis. *Elife* **9**(2020).
100. Colquhoun, D. What have we learned from single ion channels? *Journal of Physiology-London* **581**, 425-427 (2007).
101. Lape, R., Colquhoun, D. & Sivilotti, L.G. On the nature of partial agonism in the nicotinic receptor superfamily. *Nature* **454**, 722-U56 (2008).
102. Mukhtasimova, N., Lee, W.Y., Wang, H.L. & Sine, S.M. Detection and trapping of intermediate states priming nicotinic receptor channel opening. *Nature* **459**, 451-U172 (2009).

103. Delcastillo, J. & Katz, B. THE IDENTITY OF INTRINSIC AND EXTRINSIC ACETYLCHOLINE RECEPTORS IN THE MOTOR END-PLATE. *Proceedings of the Royal Society Series B-Biological Sciences* **146**, 357-361 (1957).
104. Gust, A. et al. A Starting Point for Fluorescence-Based Single-Molecule Measurements in Biomolecular Research. *Molecules* **19**, 15824-15865 (2014).
105. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nature Methods* **5**, 507-516 (2008).
106. Hellenkamp, B. et al. Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study. *Nature Methods* **15**, 669-+ (2018).
107. Yamamura, H., Suzuki, Y. & Imaizumi, Y. New light on ion channel imaging by total internal reflection fluorescence (TIRF) microscopy. *Journal of Pharmacological Sciences* **128**, 1-7 (2015).
108. Aurousseau, M.R.P., McGuire, H., Blunck, R. & Bowie, D. A Step-by-Step Guide to Single-Subunit Counting of Membrane-Bound Proteins in Mammalian Cells. *Ionotropic Glutamate Receptor Technologies* **106**, 15-30 (2016).
109. Ulbrich, M.H. & Isacoff, E.Y. Subunit counting in membrane-bound proteins. *Nature Methods* **4**, 319-321 (2007).
110. Moonschi, F.H. et al. Cell-Derived Vesicles for Single-Molecule Imaging of Membrane Proteins. *Angewandte Chemie-International Edition* **54**, 481-484 (2015).
111. Lakowicz, J.R. *Principles of fluorescence spectroscopy*, (Springer science & business media, 2013).
112. Wang, Y. et al. Single molecule FRET reveals pore size and opening mechanism of a mechano-sensitive ion channel. *Elife* **3**(2014).
113. Wang, S.Z., Vafabakhsh, R., Borschell, W.F., Ha, T. & Nichols, C.G. Structural dynamics of potassium-channel gating revealed by single-molecule FRET. *Nature Structural & Molecular Biology* **23**, 31-36 (2016).
114. Wang, S.Z. et al. Potassium channel selectivity filter dynamics revealed by single-molecule FRET. *Nature Chemical Biology* **15**, 377-+ (2019).
115. Landes, C.F., Rambhadran, A., Taylor, J.N., Salatan, F. & Jayaraman, V. Structural landscape of isolated agonist-binding domains from single AMPA receptors. *Nature Chemical Biology* **7**, 168-173 (2011).
116. Dolino, D.M. et al. The structure-energy landscape of NMDA receptor gating. *Nature Chemical Biology* **13**, 1232-+ (2017).
117. Akyuz, N. et al. Transport domain unlocking sets the uptake rate of an aspartate transporter. *Nature* **518**, 68-+ (2015).
118. Thon, S., Schmauder, R. & Benndorf, K. Elementary Functional Properties of Single HCN2 Channels. *Biophysical Journal* **105**, 1581-1589 (2013).
119. Wyman, J. The binding potential, a neglected linkage concept. *Journal of molecular biology* **11**, 631-644 (1965).
120. Ha, T. et al. Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6264-6268 (1996).
121. Wang, Q., Goldsmith, R.H., Jiang, Y., Bockenhauer, S.D. & Moerner, W.E. Probing Single Biomolecules in Solution Using the Anti-Brownian Electrokinetic (ABEL) Trap. *Accounts of Chemical Research* **45**, 1955-1964 (2012).
122. Manger, L.H. et al. Revealing Conformational Variants of Solution-Phase Intrinsically Disordered Tau Protein at the Single-Molecule Level. *Angewandte Chemie-International Edition* **56**, 15584-15588 (2017).
123. Wang, Q. & Moerner, W.E. Single-molecule motions enable direct visualization of biomolecular interactions in solution. *Nature Methods* **11**, 555-558 (2014).

124. Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für mikroskopische Anatomie* **9**, 413-468 (1873).
125. Holzmeister, P., Acuna, G.P., Grohmann, D. & Tinnefeld, P. Breaking the concentration limit of optical single-molecule detection. *Chemical Society Reviews* **43**, 1014-1028 (2014).
126. Schneckenburger, H. Total internal reflection fluorescence microscopy: technical innovations and novel applications. *Current Opinion in Biotechnology* **16**, 13-18 (2005).
127. Peng, S.J., Wang, W.J. & Chen, C.L. Breaking the Concentration Barrier for Single-Molecule Fluorescence Measurements. *Chemistry-a European Journal* **24**, 1002-1009 (2018).
128. van Oijen, A.M. Single-molecule approaches to characterizing kinetics of biomolecular interactions. *Current Opinion in Biotechnology* **22**, 75-80 (2011).
129. Loveland, A.B., Habuchi, S., Walter, J.C. & van Oijen, A.M. A general approach to break the concentration barrier in single-molecule imaging. *Nature Methods* **9**, 987-+ (2012).
130. Leslie, S.R., Fields, A.P. & Cohen, A.E. Convex Lens-Induced Confinement for Imaging Single Molecules. *Analytical Chemistry* **82**, 6224-6229 (2010).
131. Benitez, J.J. et al. Probing transient copper Chaperone-Wilson disease protein interactions at the single-molecule level with nanovesicle trapping. *Journal of the American Chemical Society* **130**, 2446-2447 (2008).
132. Sugawa, M., Nishikawa, S., Iwane, A.H., Biju, V. & Yanagida, T. Single-Molecule FRET Imaging for Enzymatic Reactions at High Ligand Concentrations. *Small* **6**, 346-350 (2010).
133. Levene, M.J. et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686 (2003).
134. Punj, D. et al. A plasmonic 'antenna-in-box' platform for enhanced single-molecule analysis at micromolar concentrations. *Nature Nanotechnology* **8**, 512-516 (2013).
135. Acuna, G.P. et al. Fluorescence Enhancement at Docking Sites of DNA-Directed Self-Assembled Nanoantennas. *Science* **338**, 506-510 (2012).
136. Puchkova, A. et al. DNA Origami Nanoantennas with over 5000-fold Fluorescence Enhancement and Single-Molecule Detection at 25 mu M. *Nano Letters* **15**, 8354-8359 (2015).
137. Peng, S.J., Sun, R.R., Wang, W.J. & Chen, C.L. Single-Molecule Photoactivation FRET: A General and Easy-To-Implement Approach To Break the Concentration Barrier. *Angewandte Chemie-International Edition* **56**, 6882-6885 (2017).
138. Goldschen-Ohm, M.P., White, D.S., Klenchin, V.A., Chanda, B. & Goldsmith, R.H. Observing Single-Molecule Dynamics at Millimolar Concentrations. *Angewandte Chemie-International Edition* **56**, 2399-2402 (2017).
139. Scheer, M. et al. BRENDA, the enzyme information system in 2011. *Nucleic Acids Research* **39**, D670-D676 (2011).
140. Rotman, B. MEASUREMENT OF ACTIVITY OF SINGLE MOLECULES OF BETA-D-GALACTOSIDASE. *Proceedings of the National Academy of Sciences of the United States of America* **47**, 1981-& (1961).
141. Smiley, R.D. & Hammes, G.G. Single molecule studies of enzyme mechanisms. *Chemical Reviews* **106**, 3080-3094 (2006).
142. Turunen, P., Rowan, A.E. & Blank, K. Single-enzyme kinetics with fluorogenic substrates: lessons learnt and future directions. *Febs Letters* **588**, 3553-3563 (2014).
143. Sambur, J.B. & Chen, P. Approaches to Single-Nanoparticle Catalysis. in *Annual Review of Physical Chemistry, Vol 65*, Vol. 65 (eds. Johnson, M.A. & Martinez, T.J.) 395-422 (2014).
144. English, B.P. et al. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nature Chemical Biology* **2**, 87-94 (2006).
145. Patterson, G., Davidson, M., Manley, S. & Lippincott-Schwartz, J. Superresolution Imaging using Single-Molecule Localization. *Annual Review of Physical Chemistry, Vol 61* **61**, 345-367 (2010).

146. Habuchi, S., Tsutsui, H., Kochaniak, A.B., Miyawaki, A. & van Oijen, A.M. mKikGR, a Monomeric Photoswitchable Fluorescent Protein. *Plos One* **3**(2008).
147. Laurence, T.A. & Weiss, S. How to detect weak pairs. *Science* **299**, 667-668 (2003).
148. Benitez, J.J., Keller, A.M. & Chen, P. NANOVESICLE TRAPPING FOR STUDYING WEAK PROTEIN INTERACTIONS BY SINGLE-MOLECULE FRET. in *Methods in Enzymology, Vol 472: Single Molecule Tools, Pt a: Fluorescence Based Approaches*, Vol. 472 (ed. Walter, N.G.) 41-60 (2010).
149. Cisse, I., Okumus, B., Joo, C. & Ha, T. Fueling protein-DNA interactions inside porous nanocontainers. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12646-12650 (2007).
150. Cisse, I.I., Kim, H. & Ha, T. A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nature Structural & Molecular Biology* **19**, 623-+ (2012).
151. Elting, M.W. et al. Single-molecule fluorescence imaging of processive myosin with enhanced background suppression using linear zero-mode waveguides (ZMWs) and convex lens induced confinement (CLIC). *Optics Express* **21**, 1189-1202 (2013).
152. Leith, J.S. et al. Free Energy of a Polymer in Slit-like Confinement from the Odijk Regime to the Bulk. *Macromolecules* **49**, 9266-9271 (2016).
153. Giannini, V., Fernandez-Dominguez, A.I., Heck, S.C. & Maier, S.A. Plasmonic Nanoantennas: Fundamentals and Their Use in Controlling the Radiative Properties of Nanoemitters. *Chemical Reviews* **111**, 3888-3912 (2011).
154. Kinkhabwala, A. et al. Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. *Nature Photonics* **3**, 654-657 (2009).
155. Rothemund, P.W.K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302 (2006).
156. Ochmann, S.E. et al. Optical Nanoantenna for Single Molecule-Based Detection of Zika Virus Nucleic Acids without Molecular Multiplication. *Analytical Chemistry* **89**, 13000-13007 (2017).
157. Zhu, P. & Craighead, H.G. Zero-Mode Waveguides for Single-Molecule Analysis. in *Annual Review of Biophysics, Vol 41*, Vol. 41 (ed. Rees, D.C.) 269-293 (2012).
158. Foquet, M. et al. Improved fabrication of zero-mode waveguides for single-molecule detection. *Journal of Applied Physics* **103**(2008).
159. Martin, W.E., Srijanto, B.R., Collier, C.P., Vosch, T. & Richards, C.I. A Comparison of Single-Molecule Emission in Aluminum and Gold Zero-Mode Waveguides. *Journal of Physical Chemistry A* **120**, 6719-6727 (2016).
160. Ponzellini, P. et al. Plasmonic zero mode waveguide for highly confined and enhanced fluorescence emission. *Nanoscale* **10**, 17362-17369 (2018).
161. Al Masud, A. et al. Mixed metal zero-mode guides (ZMWs) for tunable fluorescence enhancement. *Nanoscale Advances* **2**, 1894-1903 (2020).
162. Jamiolkowski, R.M. et al. Nanoaperture fabrication via colloidal lithography for single molecule fluorescence analysis. *Plos One* **14**(2019).
163. Teng, C.H., Lionberger, T.A., Zhang, J., Meyhofer, E. & Ku, P.C. Fabrication of nanoscale zero-mode waveguides using microlithography for single molecule sensing. *Nanotechnology* **23**(2012).
164. Kinz-Thompson, C.D. et al. Robustly Passivated, Gold Nanoaperture Arrays for Single-Molecule Fluorescence Microscopy. *Acs Nano* **7**, 8158-8166 (2013).
165. Wenger, J. et al. Single molecule fluorescence in rectangular nano-apertures. *Optics Express* **13**, 7035-7044 (2005).
166. Crouch, G.M., Han, D. & Bohn, P.W. Zero-mode waveguide nanophotonic structures for single molecule characterization. *Journal of Physics D-Applied Physics* **51**(2018).
167. Miyake, T. et al. Real-time imaging of single-molecule fluorescence with a zero-mode waveguide for the analysis of protein-protein interaction. *Analytical Chemistry* **80**, 6018-6022 (2008).

168. Patra, S., Baibakov, M., Claude, J.-B. & Wenger, J. Surface passivation of zero-mode waveguide nanostructures: benchmarking protocols and fluorescent labels. *Scientific reports* **10**, 1-10 (2020).
169. Eid, J. et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**, 133-138 (2009).
170. Gidi, Y., Bayram, S., Ablenas, C.J., Blum, A.S. & Cosa, G. Efficient One-Step PEG-Silane Passivation of Glass Surfaces for Single-Molecule Fluorescence Studies. *Acs Applied Materials & Interfaces* **10**, 39505-39511 (2018).
171. Jain, A., Liu, R., Xiang, Y.K. & Ha, T. Single-molecule pull-down for studying protein interactions. *Nature Protocols* **7**, 445-452 (2012).
172. Korlach, J. et al. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1176-1181 (2008).
173. Schwendel, D. et al. Temperature dependence of the protein resistance of poly- and oligo(ethylene glycol)-terminated alkanethiolate monolayers. *Langmuir* **17**, 5717-5720 (2001).
174. Zhao, J., Branagan, S.P. & Bohn, P.W. Single-Molecule Enzyme Dynamics of Monomeric Sarcosine Oxidase in a Gold-Based Zero-Mode Waveguide. *Applied Spectroscopy* **66**, 163-169 (2012).
175. Korlach, J. et al. REAL-TIME DNA SEQUENCING FROM SINGLE POLYMERASE MOLECULES. in *Methods in Enzymology, Vol 472: Single Molecule Tools, Pt a: Fluorescence Based Approaches*, Vol. 472 (ed. Walter, N.G.) 431-455 (2010).
176. Ardui, S., Ameur, A., Vermeesch, J.R. & Hestand, M.S. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Research* **46**, 2159-2168 (2018).
177. Uemura, S. et al. Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464**, 1012-U73 (2010).
178. Choi, J.H. et al. Dynamics of the context-specific translation arrest by chloramphenicol and linezolid. *Nature Chemical Biology* **16**, 310-+ (2020).
179. Christensen, S.M. et al. Monitoring the Waiting Time Sequence of Single Ras GTPase Activation Events Using Liposome Functionalized Zero-Mode Waveguides. *Nano Letters* **16**, 2890-2895 (2016).
180. Richards, C.I. et al. Live-Cell Imaging of Single Receptor Composition Using Zero-Mode Waveguide Nanostructures. *Nano Letters* **12**, 3690-3694 (2012).
181. Larkin, J., Foquet, M., Turner, S.W., Korlach, J. & Wanunu, M. Reversible Positioning of Single Molecules inside Zero-Mode Waveguides. *Nano Letters* **14**, 6023-6029 (2014).
182. Larkin, J., Henley, R.Y., Jadhav, V., Korlach, J. & Wanunu, M. Length-independent DNA packing into nanopore zero-mode waveguides for low-input DNA sequencing. *Nature Nanotechnology* **12**, 1169-+ (2017).
183. Martin, W.E. et al. Real-Time Sensing of Single-Ligand Delivery with Nanoaperture-Integrated Microfluidic Devices. *Acs Omega* **2**, 3858-3867 (2017).
184. Zhao, Y.H. et al. Dark-Field Illumination on Zero-Mode Waveguide/Microfluidic Hybrid Chip Reveals T4 Replisomal Protein Interactions. *Nano Letters* **14**, 1952-1960 (2014).
185. Han, D., Crouch, G.M., Fu, K., Zaino, L.P. & Bohn, P.W. Single-molecule spectroelectrochemical cross-correlation during redox cycling in recessed dual ring electrode zero-mode waveguides. *Chemical Science* **8**, 5345-5355 (2017).
186. Santoso, Y., Hwang, L.C., Le Reste, L. & Kapanidis, A.N. Red light, green light: probing single molecules using alternating-laser excitation. *Biochemical Society Transactions* **36**, 738-744 (2008).

2 Observing Single-Molecule Dynamics at Millimolar Concentrations

2.1 Abstract

Single-molecule fluorescence microscopy is a powerful tool for revealing chemical dynamics and molecular association mechanisms, but has been limited to low concentrations of fluorescent species and is only suitable for studying high affinity reactions. Here, we combine nanophotonic zero-mode waveguides (ZMWs) with Förster resonance energy transfer (FRET) to resolve single-molecule association dynamics at up to millimolar concentrations of fluorescent species. This approach extends the resolution of molecular dynamics to >100-fold higher concentrations, enabling observations at concentrations relevant to biological and chemical processes, and thus making single-molecule techniques applicable to a tremendous range of previously inaccessible molecular targets. We deploy this approach to show that the binding of cGMP to pacemaking ion channels is weakened by a slower internal conformational change.

The material in this chapter was originally published as:

Goldschen-Ohm, M.P.[1, †], White, D.S.[1,2, †], Klenchin, V.A.[1], Chanda, B.[3,4], Goldsmith, R.H.[2], Observing Single-Molecule Dynamics at Millimolar Concentrations. *Angewandte Chemie-International Edition* **2017,** *56* (9), 2399-2402.

[1] Department of Neuroscience, University of Wisconsin-Madison
[2] Department of Chemistry, University of Wisconsin-Madison
[3] Department of Biomolecular Chemistry, University of Wisconsin-Madison
[†] These authors contributed equally to this work

M.P.G.O and D.S.W. performed and analyzed all single-molecule experiments. V.A.K. performed molecular biology and protein purification. All authors contributed to conception, experimental design, and writing of the manuscript

2.2 Introduction

Single-molecule fluorescence microscopy reveals details of molecular composition and dynamics otherwise hidden because of averaging in ensemble measurements[1,2]. However, a frequent experimental compromise is the requirement of nM or lower concentrations of fluorescent species. This limitation originates from the diffraction limit of focused light, as the smaller the probe volume can be made, the fewer molecules will contribute to the background signal[3]. For example, confocal detection schemes with diffraction-limited excitation enable observation volumes as small as 0.1–0.2 fL (1 fL=1 x 10^{-15} L)[4]. Thus, in both confocal and total internal reflection (TIRF) modalities, there is an inherent concentration limit of <10 nM to detect the binding of a single fluorescently labeled substrate[5]. This "concentration barrier" is severely debilitating, as many biological and chemical mechanisms require highly concentrated conditions in the µM to mM range to proceed. Metabolites (including ATP), neurotransmitters, and amino acids are frequently present at cellular concentrations of 100 µM and above[3], thus preventing the application of single-molecule microscopy to examine binding in kinases, receptors, translation machinery, and the vast majority of enzymes (**Figure 2-1a**)[6]. Access to high concentrations becomes even more significant for new single-molecule investigations of synthetic catalysts, most of which operate at substrate or ligand concentrations of mM and above[7-9].

Increases in the highest attainable fluorophore concentration can be achieved through reduction in the observation volume to below the diffraction limit. Stimulated emission depletion (STED) in the context of fluorescence correlation spectroscopy (FCS) permits detection volumes near 0.02 fL[10]. Photoactivation of or photobleaching down to sparse subsets of fluorophores transiently creates low concentrations of active fluorescent species from a larger reservoir of fluorophores, thereby enabling concentrations up to 10 µm[11-13]. The use of nanophotonic antennas has enabled detection of fixed individual molecules in up to 25 µM concentrations of quenched fluorescent species[14], with the added benefit of plasmonically enhanced fluorescence[14]. Imaging in the vicinity of the interface between a

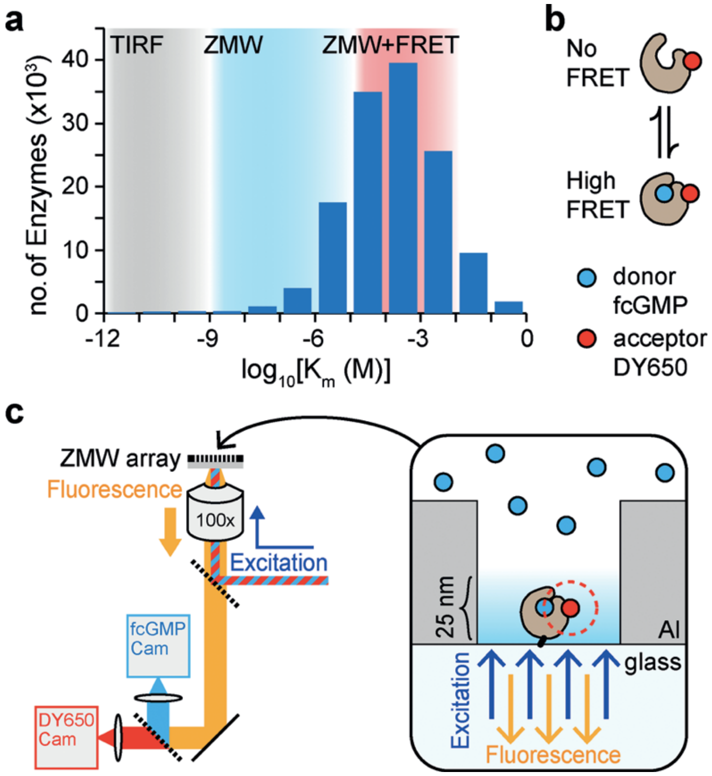

***Figure 2-1: ZMW-FRET imaging**. (**a**) Histogram of over 30,000 enzyme affinities from the BRENDA database. Concentration ranges accessible to single-molecule resolution are indicated for several methods. (**b**) Illustration of FRET between bound donor (fcGMP) and an acceptor on the CNBD. (**c**) Experimental setup for ZMW-FRET microscopy. Fluorescence in the donor and acceptor channels from arrays of ZMWs were simultaneously imaged on two EMCCD cameras. Inset: schematic representation of a single ZMW with an immobilized CNBD. The observation volume decays rapidly within about 25 nm of the surface. The red dashed circle represents the Förster radius, further reducing the effective observation volume. Thus, freely diffusing donors unbound to the CNBD are not observed.*

convex lens and a flat surface enables access to concentrations up to 2 µM[15]. Indeed, a variety of chemical and photonic tools have enabled access to concentrations up to the low µM range[16]. ZMW, nanophotonic arrays of subwavelength holes in a metallic film (**Figure 2-1c**) provide subdiffraction-limited nearfield observation volumes as small as 20 zL (1 zL=1 x 10^{-21} L), far smaller than that achievable with TIRF or STED, such that single fluorophores can be resolved at up to low µM concentrations[2,17]. ZMWs have been successfully used to observe molecular recognition processes at high nM to low µM concentrations including translation events at individual ribosomes[18], dynamics of membrane-bound proteins[19], and single-molecule electrochemistry[20], and have enabled high-throughput single-molecule genomic sequencing[21,22]. They have also been combined with plasmonic nanosized antennas to gain the advantages of fluorescence enhancement[23]. Regardless, association processes that require concentrations upwards of tens to hundreds of µM remain out of reach, thus requiring new single-molecule methods. Here, we show that a combination of ZMWs and single-molecule FRET (smFRET) enables resolution of single-molecule molecular recognition events at mM concentrations. This approach merges ZMW's subdiffraction-limited observation volume with a detection volume defined by the Förster radius of the FRET pair on the order of 1 zL (**Figure 2-1b, c**).

Although smFRET alone enables observation of single-molecule binding dynamics at up to 10 µM[24], access to higher concentrations is limited by nonspecific adsorption and background signals from freely diffusing fluorophores. While both of these interferences exist in our dual ZMW-FRET method, their influence has been sharply reduced by the volume restriction in zero-mode waveguides (ZMWs). Critically, our approach extends the resolution of single fluorophore association by over 100-fold from low µM to low mM concentrations, thus allowing the elucidation of previously inaccessible biological and chemical mechanisms at the level of single molecules.

2.3 Results

2.3.1 Single-molecule binding at millimolar concentrations

As validation, we report time-resolved single-molecule binding events for fluorescently labeled cyclic guanosine monophosphate (fcGMP) to monomeric cyclic nucleotide binding domains (CNBDs) from human hyperpolarization and cyclic nucleotide-activated (HCN) channels[25]. HCN channels are critical for the regulation of heart and brain rhythms, but the mechanism by which cyclic nucleotide binding modifies channel gating remains unclear. Single-molecule binding dynamics report on electrically silent and transient conformations energetically coupled to binding, and inform on the forces by which they interconvert, thus providing a novel window into this process. Here, we used ZMW-FRET to directly observe single binding events of fcGMP. CNBDs were specifically labeled with a FRET and immobilized

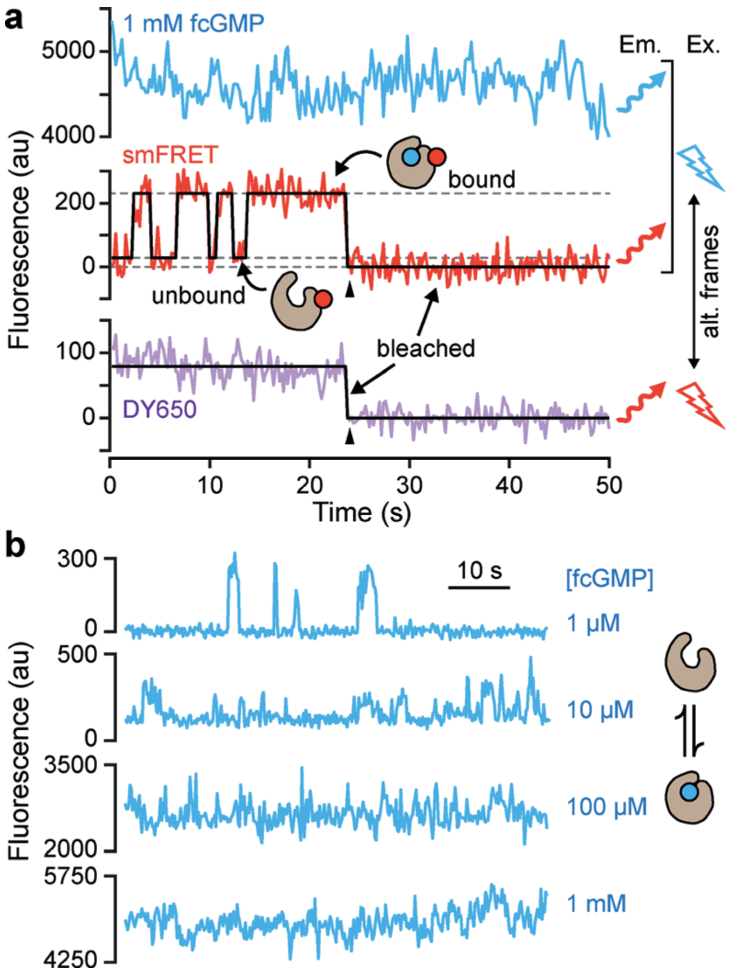

Figure 2-2: Single-molecule ligand binding at mM concentrations with ZMW-FRET. *(**a**) Fluorescence time series for fcGMP binding events at a single CNBD with freely diffusing fcGMP at a concentration of 1 mM. Simultaneous emission from the donor (blue, fcGMP) and acceptor (red) upon interleaved donor (λ=532 nm) and acceptor (λ=640 nm) excitation (see **2.5.2**). The acceptor emission is overlaid with the idealized time series (black). (**b**) Fluorescence time series for fcGMP binding to CNBDs without an acceptor label (donor only). Notably, the background from freely diffusing fcGMP in ZMWs occludes resolution of single binding events at high µM concentrations and above. The fluorescence time series in both (**a**) and (**b**) are background-subtracted, while the smFRET trace in (**a**) additionally underwent crosstalk subtraction and baseline correction by spline (see **2.5.3**)*

within arrays of over 100,000 ZMWs (**Figure 2-1c,** see **2.5.1**). Two EMCCD cameras were used to simultaneously record the donor (fcGMP) and acceptor fluorescence from about 1,000 ZMWs at once, thereby making this approach feasible for high-throughput studies (see **2.5.2**). Excitation alternated between the donor and acceptor pump wavelengths at a frame rate of 10 Hz, which allowed observation on interleaved frames of smFRET arising from donor binding (**Figure 2-2a, middle**) and acceptor stability. Importantly, this method allows quantification of the number of fluorescently labeled proteins in a ZMW

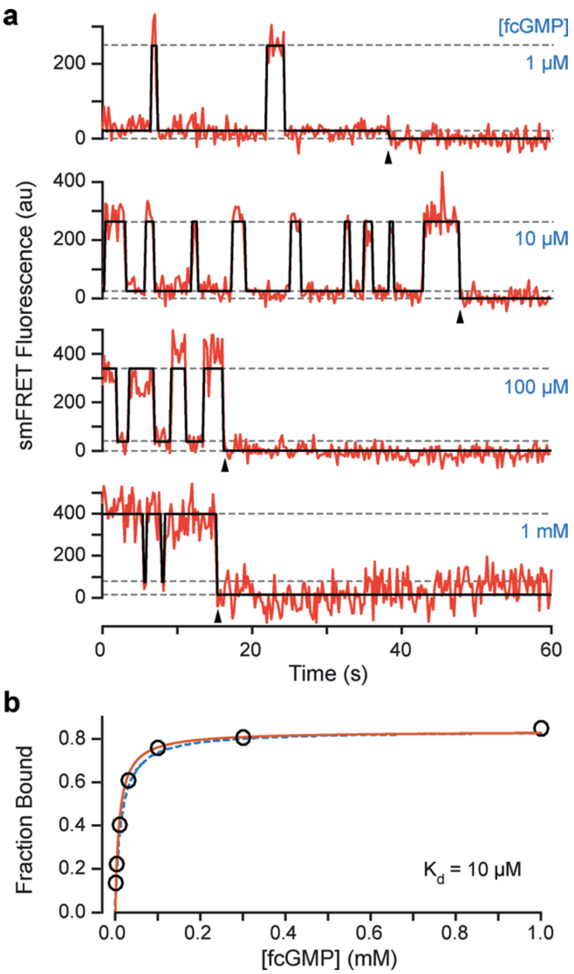

Figure 2-3: Single-molecule association dynamics of fcGMP at HCN2 CNBDs. *(**a**) FRET time series for fcGMP binding to single CNBDs (red) overlaid with idealized traces (black) at various concentrations of fcGMP. Traces shown underwent corrections for the baseline with spline fitting and were both background and crosstalk subtracted (**2.5.3**). Horizontal dashed lines indicate fluorescence levels for bound, unbound, and bleached conditions. Triangles denote the time of the acceptor bleach. (**b**) Bound probability from the total time fraction spent bound for all molecules versus the fcGMP concentration (circles) fit with* $B_{max} = 1 + K_d/[fcGMP]$ *where* $B_{max} = 0.83$ *is the maximal bound probability and* K_d*=10 µM is the apparent dissociation constant (solid line). Prediction from the model in **Figure 2-4b** normalized to* B_{max} *is shown as a dashed line.*

by counting the number of acceptor-photobleaching steps, thus enabling the selection of ZMWs featuring single proteins only (**Figure 2-2a, bottom**). Notably, although individual bound fcGMP molecules could not be resolved at a concentration of 1 mM in the donor channel (**Figure 2-2a, top**), the smFRET signal from single binding events was clearly visible in the acceptor channel prior to bleaching (**Figure 2-2a, middle**). In comparison, direct observation of fcGMP binding without FRET was possible in ZMWs only at low µM concentrations (**Figure 2-2b**). Specific binding at single molecules as reported by smFRET was

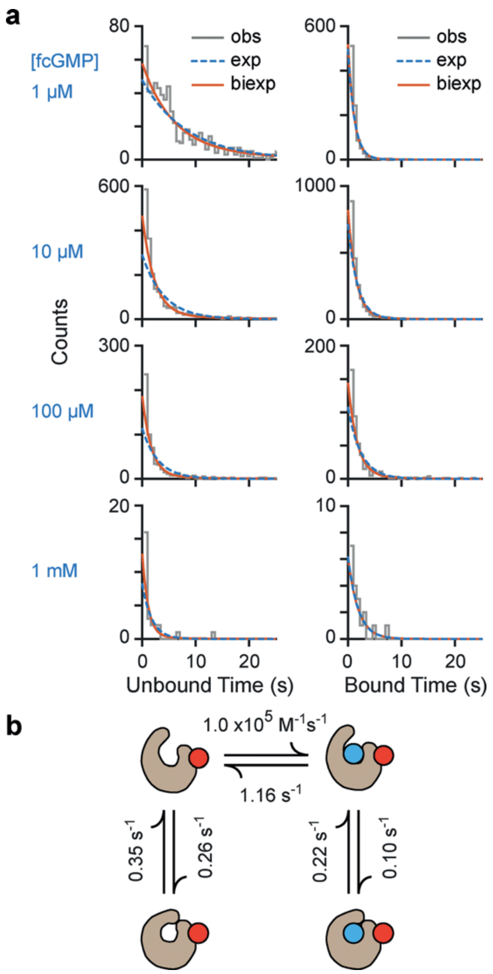

Figure 2-4: A dynamic model of fcGMP association at HCN2 CNBDs*.* (***a***) *Histograms of unbound and bound single-molecule dwell-time distributions (gray) for events from all molecules combined overlaid with monoexponential (blue dashed line) or biexponential (red line) maximum likelihood fits (see* ***2.5.5****). The concentration of fcGMP for each pair of histograms is indicated on the left-most ordinate.* (***b***) *Kinetic model of fcGMP association dynamics.*

recorded at fcGMP concentrations from 1 µM to 1 mM (**Figure 2-3a).** The concentration-dependence of the equilibrium bound probability across all molecules indicates an apparent affinity of about 10 µM (**Figure 2-3b**), similar to previous bulk-averaged measurements[26]. Notably, binding curve saturation required fcGMP at concentrations of hundreds of µM.

2.3.2 Kinetics of fcGMP binding to HCN2 CNBDs

Histograms of bound and unbound dwell times were constructed from idealized pooled data (**Figure 2-4a,** see **2.5.5**). As expected for a binding reaction, unbound lifetimes decreased with increasing fcGMP concentration, whereas bound lifetimes were relatively concentration-independent. Based on previous observations that the CNBD isomerizes between two conformations[27], we compared the likelihood of several kinetic models (see **2.5.4**). The preferred model had two unbound and two bound states, such that isomerization of the CNBD can occur both with and without bound ligand (**Figure 2-4b**). This model is consistent with X-ray crystal structures in which the C-helix caps the bound ligand[27,28] and electron paramagnetic resonance studies that suggest similar capping of the binding site also occurs in the CNBD without a ligand, thereby temporarily blocking access of fcGMP[26]. Furthermore, this scheme is similar to that observed for the binding dynamics of the higher affinity ligand fcAMP[29]. In comparison to fcAMP at monomeric CNBDs, our high-concentration single-molecule studies reveal that the lower apparent affinity of fcGMP compared to fcAMP is due not only to slower binding, but also to a reduction in the probability that bound ligand will induce a stabilizing isomerization of the CNBD that prolongs the total lifetime of the bound state.

2.4 Discussion

Now that several methods allow access to elevated concentrations for single-molecule experiments, thus breaking the “concentration barrier” to varying degrees, it is useful to compare and contrast the advantages of these methods. One issue relevant to the investigation of biological structures is access to the biomolecule. Our ZMW-FRET combination, similar to other ZMW geometries, entails the biomolecule being in a microenvironment with a high surface-to-volume ratio, although the 100–150 nm diameter of the aperture is large compared to the CNBD, and the sidewalls and bottom of the ZMW are passivated to prevent non-specific absorptions[30]. Although this geometry is less restrictive than antenna-based approaches that involve a nanosized aperture[18,31], it is more restrictive than photoactivation and FRET approaches that do not have the same constraints and may be more suitable for in vivo measurements[13,24]. Our approach does not require specialized fluorophores, unlike photoactivation-based approaches[13]. However, the temporal observation window of our approach as well as FRET-based approaches is limited by the photobleaching of the acceptor[24], whereas in ZMWs without FRET[4,17] or other approaches where all fluorophores are continually replenished[2,13,15,17,18,31], the observation window will be longer, although more stable fluorophores can likely extend the observation window in ZMW-FRET[32]. Finally, only ZMW-FRET is capable of reaching the biologically significant range of 100 µM–1 mM.

In conclusion, we demonstrate single-molecule resolution of binding events at up to mM concentration, more than two orders of magnitude higher concentrations than were feasible with other methods. Our combined ZMW-FRET approach enables observation of molecular dynamics at relevant concentrations for the majority of biological and chemical association reactions that were previously inaccessible to single-molecule techniques.

Notably, several of the drawbacks of our ZMW-FRET can also be improved using recently reported strategies. For example, use of more photostable FRET acceptors can greatly augment observation time windows[32]. Use of DNA Origami nanoadapters can significantly enhance the fluorescence signal by placing biomolecules at the center of the ZMW and increasing experimental throughput by streamlining deposition of biomolecules into ZMW's[31]. Thus, even more robust access to single-molecule dynamics at high concentrations is within reach.

2.5 Methods

2.5.1 Protein expression, purification, and labeling

The CNBD from human HCN2 channels was expressed in E. coli, biotinylated and labeled with a maleimide derivative of DyLight 650 (DY650) at an introduced cysteine (E571C) as described in detail previously[27]. For specific labeling at position E571C, two other accessible native cysteine residues were mutated (C508A/C584S), which we have previously shown to have little effect on the CNBDs ability to bind cyclic nucleotide[27].

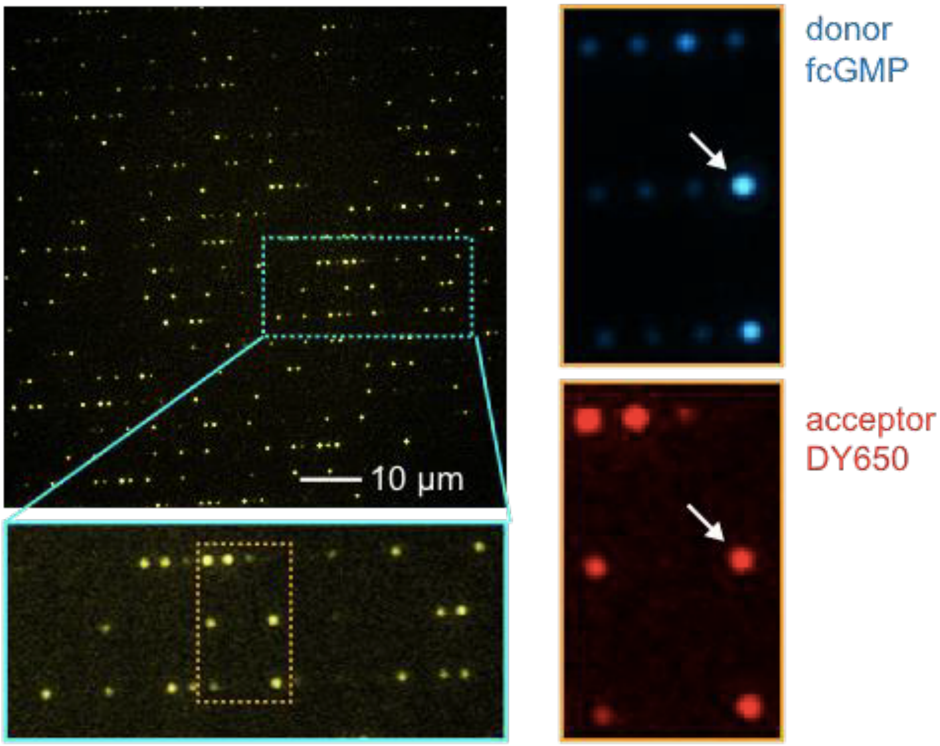

Figure 2-5: High-throughput imaging and FRET channels. *Fluorescence image of ~1,000 ZMWs in the acceptor channel during excitation at the acceptor pump wavelength in a solution of 10 µM fcGMP. Approximately 250 fluorescently labeled CNBDs are identified in this ZMW array. Boxed region is shown on an expanded scale below. Boxed region within expanded region is shown to the right for both donor (top) and acceptor (bottom) channels. Arrow denotes an individual ZMW containing an acceptor-labeled CNBD with bound donor (fcGMP).*

2.5.2 Single-molecule imaging

Single-molecule imaging and analysis of fcGMP (Biolog) binding was performed as described previously for fcAMP[27]. Briefly, CNBDs were deposited in ZMWs with diameters between 100-150 nm (Pacific Biosciences) and imaged on an inverted microscope (Olympus IX-71) with two 512x512 EMCCD cameras (Andor iXon Ultra X-9899) under alternating 532 and 640 nm laser excitation (Coherent) at 60 W/cm2 and 25 W/cm2 at the sample, respectively, and a frame rate of 10 Hz using Metamorph software (Molecular Devices). This set-up enabled simultaneous recording of ~1,000 ZMWs at a time; however, we sparsely deposited the CNBDs onto the array in order to reduce the probability of having more than one CNBD per ZMW (**Figure 2-5**). At our sample loading, we found 25% overall occupancy of the 1,000 wells. Of those ~250 wells, ~10% showed two bleach steps suggesting two proteins per ZMW, and ~1% showed more than three bleach steps suggesting three or more proteins per ZMW. These observed probabilities, P(1) = 0.23, P(2) = 0.03, P(3)<0.003, are close to the probabilities suggested by Poisson statistics with a Poisson rate parameter λ=0.35, P(1) = 0.25, P(2) = 0.04, P(3) = 0.005, where P(k) is the probably of an occupancy of k in a ZMW according to the equation

$$P(k) = \frac{\lambda^k e^{-k}}{k!} \qquad \text{Eq. 2-1}$$

2.5.3 Single-molecule trace selection

Fluorescence time series from single molecules were analyzed with custom software written in MATLAB (The MathWorks, Inc.). A trace was used for analysis if: (1) it featured only a single fluorescently-labeled CNBD in the ZMW, as indicated by a single bleach step in the acceptor channel (Fig. 2a, bottom), and (2) featured at least one event in the FRET channel, indicating fcGMP binding to the CNBD. Traces featuring multiple bleach steps, large amplitude fluctuations in the donor channel likely due to imperfect surface passivation and consequent increased non-specific binding, and traces without any FRET channel fluctuations, indicating no binding events, were all removed prior to analysis. All selected traces were then background subtracted, and crosstalk was also subtracted from the smFRET traces, as described in more detail below (see **2.5.4**). Drifts in the baseline were corrected with spline fitting before the traces were idealized with vbFRET[33]. After idealization, traces were once again visually inspected and single frame events were removed from the idealized records as such events could often be attributed to noise in simulated data. For analysis of dynamics, we selected the subset of molecules with smFRET binding signal-to-noise ratios >2.25 (**Figure 2-7e**). Heterogeneity in signal-to-noise is likely due to both variation in ZMW fabrication and location of CNBDs within each ZMW[4]. Overall, approximately 11,000 ZMWs satisfied the initial screen, and of those, approximately 4,000 ZMWs passed the second signal-to-noise screen. From

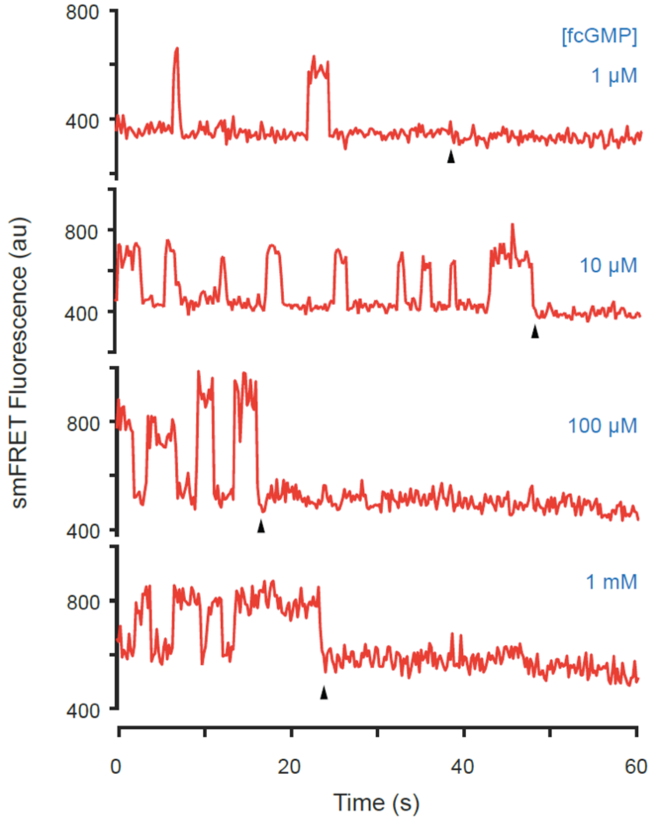

Figure 2-6: Raw ZMW-FRET traces. FRET time series for fcGMP binding to single CNBDs at various concentrations of fcGMP. *Representative traces without baseline correction via spine-fitting, background subtraction, or cross-talk subtraction that were used to produce the analogous figures in* ***Figure 2-2a*** *(1 mM) or* ***Figure 2-3a*** *(1, 10, 100 µM). Triangle denotes time of acceptor bleach.*

these ZMWs, we recorded approximately 15,000 binding/ unbinding events. Hidden Markov modeling (HMM) was performed with QuB using a dead time of 200 ms (**Figure 2-4, 2-8e,** see **2.5.5**)[34,35].

2.5.4 Crosstalk subtraction at high donor concentrations

In general, the signals in the donor (D) and acceptor (A) channels upon excitation at the donor pump wavelength are given by

$$D = D_{bg} + D_{free} + D_{bound} \qquad \text{Eq. 2-2}$$

$$A = A_{bg} + A_{crosstalk} + A_{direct} + FRET_{free} + FRET_{bound} \qquad \text{Eq. 2-3}$$

where $D_{bg}$and A_{bg} are background levels including camera dark counts, and scattered or reflected light, D_{free} and D_{bound} are from directly excited freely diffusing and bound fcGMP, respectively, $A_{crosstalk}$ is crosstalk from the fraction of the donor emission spectrum that falls within the acceptor channel, A_{direct} is due to weak direct excitation of the acceptor at the donor pump wavelength, and $FRET_{free}$ and $FRET_{bound}$ are FRET between freely diffusing and bound donors (fcGMP), respectively, and the acceptor on the CNBD. Crosstalk was described by $A_{crosstalk} = \alpha(D_{free} + D_{bound})$ where $\alpha = 0.08$ was determined from the slope of a linear fit to a scatter plot of $\langle \mathrm{A} \rangle_{bleached}$ versus $\langle \mathrm{D} \rangle_{bleached}$ for each molecule, where $\langle\ \rangle_{bleached}$ denotes a time average after acceptor bleaching (**Figure 2-7a**). This value is slightly higher than the $\alpha = 0.04$ estimated from the relative area of fcGMPs emission spectrum within the donor and acceptor channels, likely due to non-ideal transmission efficiencies along the optical path. Our analysis was relatively insensitive to small changes in α on this order. In the absence of donor and acceptor dyes, the background signals A_{bg} and D_{bg}varied between ZMWs, likely due to heterogeneity in the fabrication process. Empirically, A_{bg} and D_{bg}were found to be similar within a given ZMW under our experimental conditions. Thus, we determined A_{bg} in each individual ZMW by solving $\langle \mathrm{A} \rangle_{bleached} = A_{bg} + \alpha(\langle \mathrm{D} \rangle_{bleached} + \langle \mathrm{D} \rangle_{bg})$with $A_{bg} = D_{bg}$. The distribution of A_{bg} values across ZMWs is shown in **Figure 2-7b**.

Prior to acceptor bleaching, A exhibits two distinct intensity levels $\langle \mathrm{A} \rangle_{unbound}$ and $\langle \mathrm{A} \rangle_{bound}$ corresponding to conditions where the CNBD is unoccupied or occupied by fcGMP (**Figures 2-2a, 2-3a, 2-7c**). Relative to $\langle \mathrm{A} \rangle_{bleached}$, $\langle \mathrm{A} \rangle_{unbound}$ also contains both A_{direct} and $FRET_{free}$ whereas $\langle \mathrm{A} \rangle_{bound}$ includes all of those components plus $FRET_{bound}$. The increase in the per molecule mean $\langle \mathrm{A} \rangle_{unbound}$ (and also $\langle \mathrm{A} \rangle_{bound}$) with increasing concentrations of fcGMP comes from the concentration-dependence of $FRET_{free}$, which arises from freely diffusing fcGMP near the acceptor on the CNBD (**Figure 2-6c, d**). The contribution of $FRET_{free}$ was lower than predicted for a bulk solution of freely diffusing donors and acceptors[36], because only a donor quenched by multiple acceptors was considered, not multiple donors

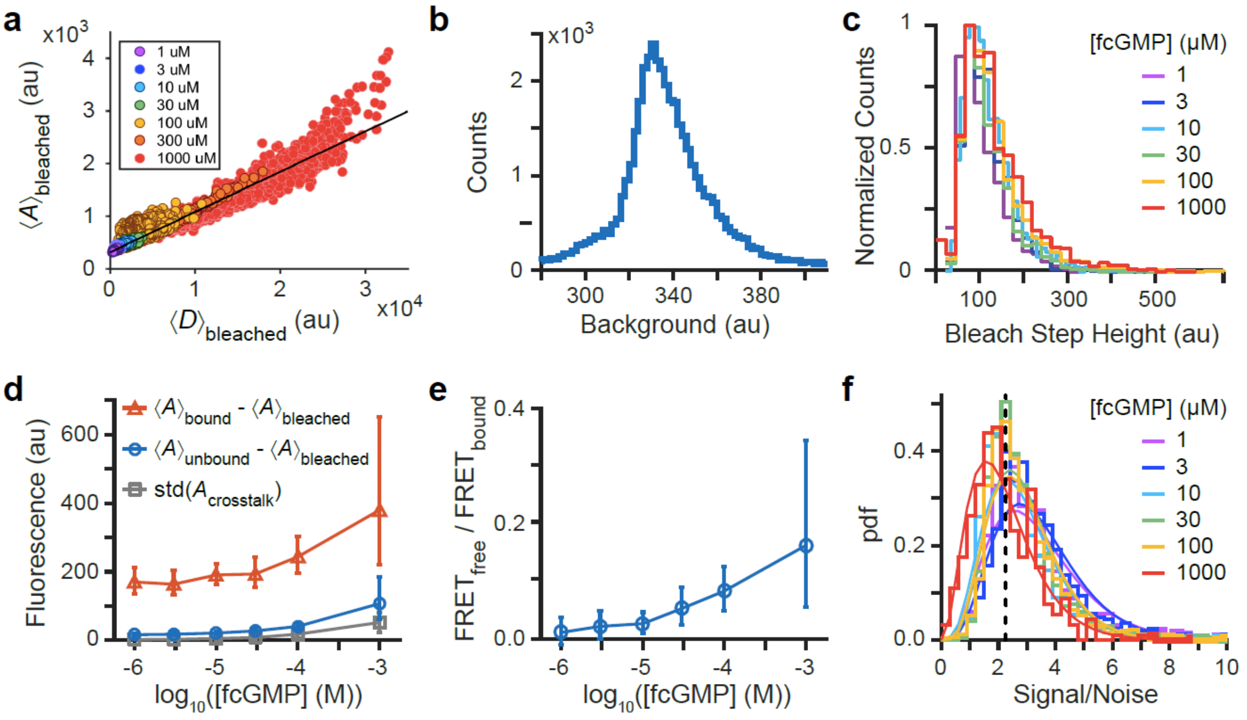

***Figure 2-7: Statistics for single-molecule association dynamics with ZMW-FRET.** (**a**) Correlation between mean fluorescence intensity in donor and acceptor channels after acceptor bleaching at various donor (fcGMP) concentrations overlaid with linear fit (black line). Each circle represents an individual molecule. (**b**) Distribution of camera background intensity levels (A_{bg}) across ZMWs. (**c**) Normalized distribution of the average acceptor bleach step height per concentration of fcGMP, (**d**) Fluorescence intensities as a function of donor (fcGMP) concentration (see **2.5.4**). Symbols are means from a maximum likelihood fit to a log normal distribution, and error bars are the 95% confidence limits in the distribution means. (**e**) The ratio of FRET from freely diffusing or bound donors (fcGMP) as a function of donor concentration. Symbols are means from a maximum likelihood fit to a log normal distribution, and error bars are the 95% confidence limits in the distribution means. (**f**) Histograms of signal-to-noise ratios in the acceptor channel for various donor (fcGMP) concentrations. Histograms at each donor concentration are overlaid with maximum likelihood fits to a gamma distribution. Vertical dashed line indicates cutoff for our analysis.*

transferring to a single acceptor, and because of limited access to the acceptor from regions occupied by the CNBD in our system. In contrast, the mean amplitude of $FRET_{bound} = \langle \mathrm{A} \rangle_{bound}$ - $\langle \mathrm{A} \rangle_{unbound}$ across molecules was relatively independent of fcGMP concentration (**Figure 2-7c**), as expected. At low concentrations, A_{bound} constitutes the dominant source of background in the acceptor channel, whereas at high donor concentrations, $A_{crosstalk}$ becomes a significant contributor, as expected. Critically, the amplitude of $FRET_{bound}$ was large in comparison to both $FRET_{free}$ and the standard deviation of the subtracted crosstalk signal $A_{crosstalk}$ (**Figure 2-7c**) indicating that on average, neither FRET from freely diffusing fcGMP nor subtraction of crosstalk grossly distorted the binding signal even at 1 mM fcGMP.

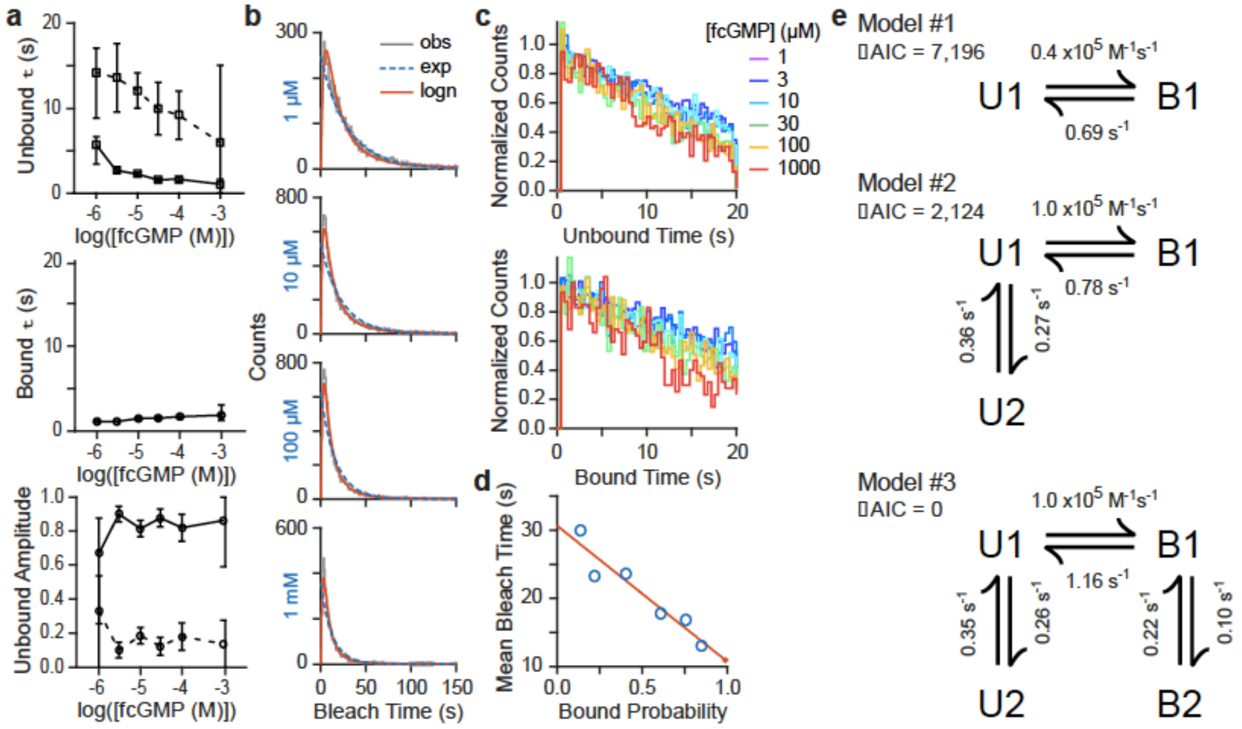

Figure 2-8: Dwell times and HMM modeling of fcGMP association at single molecules. *(**a**) Summary of maximum likelihood exponential fits to unbound (biexponential) and bound (monoexponential) dwell times. Amplitudes for biexponential fast (solid) and slow (dashed) components are shown in the bottom frame. (**b**) Distributions of bleach times at various donor (fcGMP) concentrations overlaid with maximum likelihood fits to exponential (dashed) or log normal (smooth) distributions. Donor concentrations are indicated along the ordinate. (**c**) Simulated dwell times drawn from a uniform distribution using amplitude, noise and bleach times drawn from distributions that describe the observed data. The reduced frequency of longer dwell times reflects truncation of the observation window due to acceptor bleaching. (**d**) Linear correlation between acceptor lifetime and bound probability. (**e**) Kinetic models for transition between unbound (U*) and bound (B*) states explored with HMM. Not shown are irreversible transitions to a bleached state that were allowed from each state (**2.5.5**). The AIC information score for each model is given relative to the most likely model (lower is better).*

2.5.5 Single-molecule kinetic analysis

Dwell time distributions from idealized records were fit with mono or biexponential distributions by maximum likelihood. The conditional probability that a given dwell time would be observed within the associated observation time window defined by acceptor bleaching was accounted for as described by Colquhoun and Sigworth (**Figure 2-4a, 2-8a**)[37]. Simulated data from a uniform distribution of dwell times with amplitudes, noise, and acceptor lifetimes drawn from gamma or log normal distributions describing the observed data (e.g. **Figure 2-8b**), show that although acceptor bleaching limited the observation of longer dwell times, this effect was relatively minor over the majority of the unbound and bound dwell time distributions (**Figure 2-8c**), and was also accounted for in our kinetic model (**Figure 2-4b**) as described below. Acceptor lifetimes were inversely correlated with bound probability, suggesting that excitation due to FRET from bound donor during frames with laser excitation at the donor pump contributed to bleaching (**Figure 2-8d**). To account for this, an irreversible transition to a bleached state was added to all states during HMM model optimization. The bleach rate from either unbound (0.03 s^{-1}) or bound (0.08 s^{-1}) states were determined by extrapolating the correlation between acceptor lifetime and bound probability to bound probabilities of zero or one, respectively. This method ignores the comparatively small concentration-dependence of the bleach rate due to FRET from freely diffusing donors. Models were globally optimized for all molecules and fcGMP concentrations, and ranked by their Akaike information criterion (**Figure 2-8e**)[38].

Consistent with an isomerization of the unliganded CNBD, unbound dwell time distributions were biexponential (**Figure 2-4a**). In contrast, only a single exponential component was resolvable in the bound time distributions. Although we cannot completely rule out the existence of only a single bound state (e.g. model #2), the rate of entry into the state B2 in our preferred model #3 is sufficiently slow that acceptor bleaching limits its observation (**Figure 2-8e**), therefore making it difficult to detect in the distribution of bound times. Regardless, both of the most likely models (#2 and #3) agree that the reduced apparent affinity of fcGMP as compared to fcAMP arises from both slower binding and either reduced or nonexistent probability to isomerize while ligand is bound. Notably, the small number of events in dwell time distributions at 1 mM fcGMP does not reflect the number of observed events at molecules exhibiting binding signals with sufficiently high signal-to-noise ratios, but instead reflects the difficulty in observing fully bracketed events (bound-unbound-bound or unbound-bound-unbound) at a concentration significantly above K_d prior to acceptor bleaching. Thus, protein-ligand complexes with higher K_d values will be even easier to analyze. Finally, significant improvements in data throughput and model analysis could be obtained by utilizing longer-lived acceptor dyes[32].

2.6 References

1. Ha, T. Single-molecule methods leap ahead. *Nature Methods* **11**, 1015-1018 (2014).
2. Moerner, W.E. & Fromm, D.P. Methods of single-molecule fluorescence spectroscopy and microscopy. *Review of Scientific Instruments* **74**, 3597-3619 (2003).
3. Bennett, B.D. et al. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology* **5**, 593-599 (2009).
4. Levene, M.J. et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686 (2003).
5. van Oijen, A.M. Single-molecule approaches to characterizing kinetics of biomolecular interactions. *Current Opinion in Biotechnology* **22**, 75-80 (2011).
6. Scheer, M. et al. BRENDA, the enzyme information system in 2011. *Nucleic Acids Research* **39**, D670-D676 (2011).
7. Cordes, T. & Blum, S.A. Opportunities and challenges in single-molecule and single-particle fluorescence microscopy for mechanistic studies of chemical reactions. *Nature Chemistry* **5**, 993-999 (2013).
8. Ng, J.D. et al. Single-Molecule Investigation of Initiation Dynamics of an Organometallic Catalyst. *Journal of the American Chemical Society* **138**, 3876-3883 (2016).
9. Roeffaers, M.B.J. et al. Single-molecule fluorescence spectroscopy in (bio)catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12603-12609 (2007).
10. Kastrup, L., Blom, H., Eggeling, C. & Hell, S.W. Fluorescence fluctuation spectroscopy in subdiffraction focal volumes. *Physical Review Letters* **94**(2005).
11. Manley, S. et al. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature Methods* **5**, 155-157 (2008).
12. Moertelmaier, M., Brameshuber, M., Linimeier, M., Schutz, G.J. & Stockinger, H. Thinning out clusters while conserving stoichiometry of labeling. *Applied Physics Letters* **87**(2005).
13. Loveland, A.B., Habuchi, S., Walter, J.C. & van Oijen, A.M. A general approach to break the concentration barrier in single-molecule imaging. *Nature Methods* **9**, 987-+ (2012).
14. Kinkhabwala, A. et al. Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. *Nature Photonics* **3**, 654-657 (2009).
15. Leslie, S.R., Fields, A.P. & Cohen, A.E. Convex Lens-Induced Confinement for Imaging Single Molecules. *Analytical Chemistry* **82**, 6224-6229 (2010).
16. Holzmeister, P., Acuna, G.P., Grohmann, D. & Tinnefeld, P. Breaking the concentration limit of optical single-molecule detection. *Chemical Society Reviews* **43**, 1014-1028 (2014).
17. Zhu, P. & Craighead, H.G. Zero-Mode Waveguides for Single-Molecule Analysis. in *Annual Review of Biophysics, Vol 41*, Vol. 41 (ed. Rees, D.C.) 269-293 (2012).
18. Uemura, S. et al. Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464**, 1012-U73 (2010).
19. Christensen, S.M. et al. Monitoring the Waiting Time Sequence of Single Ras GTPase Activation Events Using Liposome Functionalized Zero-Mode Waveguides. *Nano Letters* **16**, 2890-2895 (2016).

20. Zhao, J., Zaino, L.P. & Bohn, P.W. Potential-dependent single molecule blinking dynamics for flavin adenine dinucleotide covalently immobilized in zero-mode waveguide array of working electrodes. *Faraday Discussions* **164**, 57-69 (2013).
21. Eid, J. et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**, 133-138 (2009).
22. Larkin, J., Foquet, M., Turner, S.W., Korlach, J. & Wanunu, M. Reversible Positioning of Single Molecules inside Zero-Mode Waveguides. *Nano Letters* **14**, 6023-6029 (2014).
23. Punj, D. et al. A plasmonic 'antenna-in-box' platform for enhanced single-molecule analysis at micromolar concentrations. *Nature Nanotechnology* **8**, 512-516 (2013).
24. Sugawa, M., Nishikawa, S., Iwane, A.H., Biju, V. & Yanagida, T. Single-Molecule FRET Imaging for Enzymatic Reactions at High Ligand Concentrations. *Small* **6**, 346-350 (2010).
25. Biskup, C. et al. Relating ligand binding to activation gating in CNGA2 channels. *Nature* **446**, 440-443 (2007).
26. DeBerg, H.A., Brzovic, P.S., Flynn, G.E., Zagotta, W.N. & Stoll, S. Structure and Energetics of Allosteric Regulation of HCN2 Ion Channels by Cyclic Nucleotides. *Journal of Biological Chemistry* **291**, 371-381 (2016).
27. Goldschen-Ohm, M.P. et al. Structure and dynamics underlying elementary ligand binding events in human pacemaking channels. *Elife* **5**(2016).
28. Zagotta, W.N. et al. Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* **425**, 200-205 (2003).
29. Goldschen-Ohm, M.P., White, D.S., Klenchin, V.A., Chanda, B. & Goldsmith, R.H. Observing Single-Molecule Dynamics at Millimolar Concentrations. *Angewandte Chemie-International Edition* **56**, 2399-2402 (2017).
30. Korlach, J. et al. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1176-1181 (2008).
31. Puchkova, A. et al. DNA Origami Nanoantennas with over 5000-fold Fluorescence Enhancement and Single-Molecule Detection at 25 mu M. *Nano Letters* **15**, 8354-8359 (2015).
32. Zheng, Q.S. et al. Ultra-stable organic fluorophores for single-molecule research. *Chemical Society Reviews* **43**, 1044-1056 (2014).
33. Bronson, J.E., Fei, J., Hofman, J.M., Gonzalez, R.L., Jr. & Wiggins, C.H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophysical Journal* **97**, 3196-3205 (2009).
34. Nicolai, C. & Sachs, F. Solving ion channel kinetics with the QuB software. *Biophysical Reviews and Letters* **8**, 191-211 (2013).
35. Qin, F., Auerbach, A. & Sachs, F. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophysical Journal* **79**, 1915-1927 (2000).
36. Birks, J.B. & Leite, M. ENERGY TRANSFER IN ORGANIC SYSTEMS .9. EFFECT OF DIFFUSION ON TRANSFER EFFICIENCY. *Journal of Physics Part B Atomic and Molecular Physics* **3**, 513-& (1970).
37. Sakmann, B. *Single-channel recording*, (Springer Science & Business Media, 2013).
38. Akaike, H. NEW LOOK AT STATISTICAL-MODEL IDENTIFICATION. *Ieee Transactions on Automatic Control* **AC19**, 716-723 (1974).

3 Top-Down Machine Learning Approach for High-Throughput Single-Molecule Analysis

3.1 Abstract

Single-molecule approaches provide enormous insight into the dynamics of biomolecules, but adequately sampling distributions of states and events often requires extensive experimentation. Although emerging experimental techniques can generate such large datasets, existing analysis tools are not suitable to process the large volume of data obtained in high-throughput paradigms. Here, we present a new analysis platform (DISC) that accelerates unsupervised analysis of single-molecule trajectories. By merging unsupervised statistical learning with the Viterbi algorithm, DISC idealizes single-molecule trajectories up to three orders of magnitude faster with improved accuracy compared to other commonly used algorithms. Further, we demonstrate the utility of DISC algorithm to probe cooperativity between multiple binding events in the cyclic nucleotide binding domains of HCN pacemaker channel. Given the flexible and efficient nature of DISC, we anticipate it will be a powerful tool for unsupervised processing of high-throughput data across a range of single-molecule experiments.

The material in this chapter was originally published as:

White, D.S.[1,2], Goldschen-Ohm, M.P.[3], Goldsmith, R.H.[2], Chanda, B.[1,4] Top-down machine learning approach for high-throughput single-molecule analysis. *Elife* **2020**, *9*.

[1] Department of Neuroscience, University of Wisconsin-Madison, Madison, WI, USA
[2] Department of Chemistry, University of Wisconsin-Madison, Madison, WI, USA
[3] Department of Neuroscience, University of Texas at Austin, Austin, TX, USA
[4] Department of Biomolecular Chemistry University of Wisconsin-Madison, Madison, WI, USA

D.S.W. conceptualized and wrote the DISC algorithm. D.S.W and M.P.G.O performed and analyzed single-molecule experiments. D.S.W and M.P.G.O wrote scripts for the image processing, simulations, and algorithm quantifications. D.S.W., M.P.G.O., R.H.G, and B.C. contributed to conception, experimental design, and writing of the manuscript.

3.2 Introduction

Single-molecule methods are powerful tools for providing insight into heterogeneous dynamics underlying chemical and biological processes otherwise obscured in bulk-averaged measurements[1]. Use of these techniques has expanded rapidly, with modalities spanning electrophysiology, fluorescence, and force spectroscopy to probe diverse physical phenomena. Generally, single-molecule data are obtained as a time trajectory where molecular behavior is observed as a series of transitions between a set of discrete states obscured by experimental noise. Following the growing realization that molecules involved in physiological and chemical processes exhibit complex kinetics and a diversity of behavior, there is an increasing demand for high-throughput technologies to adequately sample different sub-populations and rare but important events[2]. As a result, there has been tremendous progress in improving both the number of single molecules that can be observed simultaneously and the total observation time of each molecule. For example, the observation window prior to photobleaching in conventional fluorescence paradigms such as single-molecule Förster resonance energy transfer (smFRET) or colocalization single-molecule spectroscopy (CoSMoS) can be dramatically extended with recently developed photostable dyes[3,4]. The current generation of metal-oxide semiconductor (sCMOS) detectors enables simultaneous imaging of 1×10^4 molecules in a total internal fluorescence microscopy (TIRFM) configuration and can be coupled with nanofabricated zero-mode waveguides (ZMWs) to enable access to high concentrations[5-7]. Non-fluorescence-based single-molecule experiments such as plasmon rulers, scattering, magnetic tweezers, and single-molecule centrifugation generate a tremendous amount of data through parallel measurement of hundreds of molecules with orders of magnitude longer recordings than a typical fluorescence experiment[8-12].

Despite these incredible advances in generating statistically robust data sets, standard analysis algorithms impose a computational bottleneck at this scale of data generation[7,13]. This is particularly true when the dynamics and physical states of a system are unknown.

Typical statistical modeling of single-molecule trajectories often adopts one of two approaches. The first is a probabilistic approach that models a molecule's behavior as a Markov chain, wherein the molecule transitions between hidden discrete states whose outputs are measured experimentally (hidden Markov model, HMM). This involves estimating the transition probabilities between a small set of postulated states with defined outputs using methods to maximize the likelihood of the model given the observations or Bayesian inference to estimate model parameter distributions. Numerous software packages have been developed for implementing HMMs, such as QuB[14,15], HaMMy[16], SMART[17], vbFRET[18], ebFRET[19] and SPARTAN[7], each of which utilize a different HMM training method. For example, QuB implements the fast segmental k-means algorithm (SKM) which combines k-means clustering and the Viterbi algorithm to identify transitions between postulated states[20], whereas vbFRET adapts variational Bayesian inference for parameter estimation at faster speeds than traditional HMM training in both smFRET and single-particle tracking experiments[21,22]. Although powerful statistical tools are very useful for single-molecule analysis, HMMs have notable limitations, especially in the context of high-throughput analysis and unknown system dynamics. For example, HMMs are often used in a supervised manner where the user postulates model parameters such as the number of states, their measured outputs, and the allowed transitions between them. As this information is often not known a priori, it is desirable to test multiple models and rank them according to Bayesian probabilistic approaches or objective functions, such as the Bayesian Information Criterion (BIC). This process can dramatically increase the analysis time to ensure the parameters space has been sufficiently explored, which restrict their usefulness in high-throughput single-molecule analysis. Although variants such as infinite HMMs using Bayesian nonparametric inference try to naturally learn the trajectory of the states without the typical parametric model selection, these too are often computationally prohibitive for large datasets[23-25].

The second class of single-molecule analysis approaches idealization as an unsupervised clustering problem from machine learning[26]. Typically, clusters of intensity values (e.g. states) are determined using bottom-up hierarchical agglomerative clustering (HAC) algorithms, which begin by treating each observation of *N* total observations as singleton clusters and perform *N*-1 iterations wherein pairs of clusters are merged until all data-points belong to a single cluster. For each number of possible clusters, an objective function can be minimized to find the optimal trade-off between the complexity and fit. This implementation results in time complexity of $O(N^2)$ owning to the need of computing a *N* x *N* similarity matrix to determine which clusters should be merged at each iteration. In practice, a separate algorithm called change-point (CP) detection precedes HAC to reduce the solution domain of the objective function by identifying statistically significant stepwise changes in signal over time. We denote this combination of algorithms as CP-HAC. At each identified CP, the data are divided into two segments, each described by the mean values of the data-points between sequential change-points. By using the segments as initial clusters rather than all *N* data-points, the HAC computation can be dramatically reduced. The pioneering application of CP-HAC to single-molecule data addressed CP detection and clustering in the presence of Poisson noise[27]. Variants of this framework such as STaSI use other merit functions for Gaussian noise, including the Student's t-test for fast CP detection and minimum description length for state selection[28]. An advantage of CP-HAC methods is that they only require a confidence interval and/or an objective function for the analysis, unlike HMMs which require a model to fit. This makes them very attractive in situations where there is no prior knowledge about the different physical states. In common experimental modalities such as smFRET, CP-HAC methods offer superior computational speed over HMM approaches; however, their quadratic time-complexity renders them inefficient on long trajectories[28]. In addition, simulation studies have suggested CP-HAC algorithms yield lower event detection accuracy than HMM approaches[29].

Despite the utility of HMM and CP-HAC methods, there is an outstanding need for an analysis platform to provide accurate unsupervised idealization with sufficiently high computational performance to keep up with the increasing scale of data generation. Although advances in computing hardware can, to a degree, mitigate these issues[30,31], there remains a pressing need for more computationally efficient algorithms. Here, we present a new algorithm for efficient and accurate idealization of large single-molecule datasets in a model-independent manner. Our method, DISC (DIvisive Segmentation and Clustering), enhances existing statistical learning methods and enables rapid state and event detection. The DISC algorithm draws inspiration from other algorithms like CP-HAC and SKM that rely on unsupervised algorithms, such as k-means and hierarchical clustering. We advance these ideas by adapting divisive clustering algorithms from data mining and information theory to improve the rate and accuracy of identifying signal amplitude clusters (states) in a top-down process as opposed to the typical bottom-up clustering[32-34]. We further couple our unsupervised state detection with the Viterbi algorithm to enable robust event detection on par with HMM methods[35-37]. Overall, DISC is an unsupervised method that combines statistical learning approaches with the high event detection accuracy of HMMs at a fraction of the computational cost, enabling convenient application to large datasets.

3.3 Theory

3.3.1 Motivation

The goal of the DISC algorithm is time series idealization: the hard assignment of data points into discrete states. DISC approaches the problem of idealization as an unsupervised problem in machine learning wherein the number of significant states for a given single-molecule trajectory are not known a priori. This process of learning both the significant states and the transitions between them is accomplished in three phases (**Figure 3-1**): 1) divisive segmentation, 2) HAC, and 3) the Viterbi algorithm. The first two phases use unsupervised statistical learning to identify the intensities of states following an appropriate user-specified objective function. The second phase uses the Viterbi algorithm to decode the

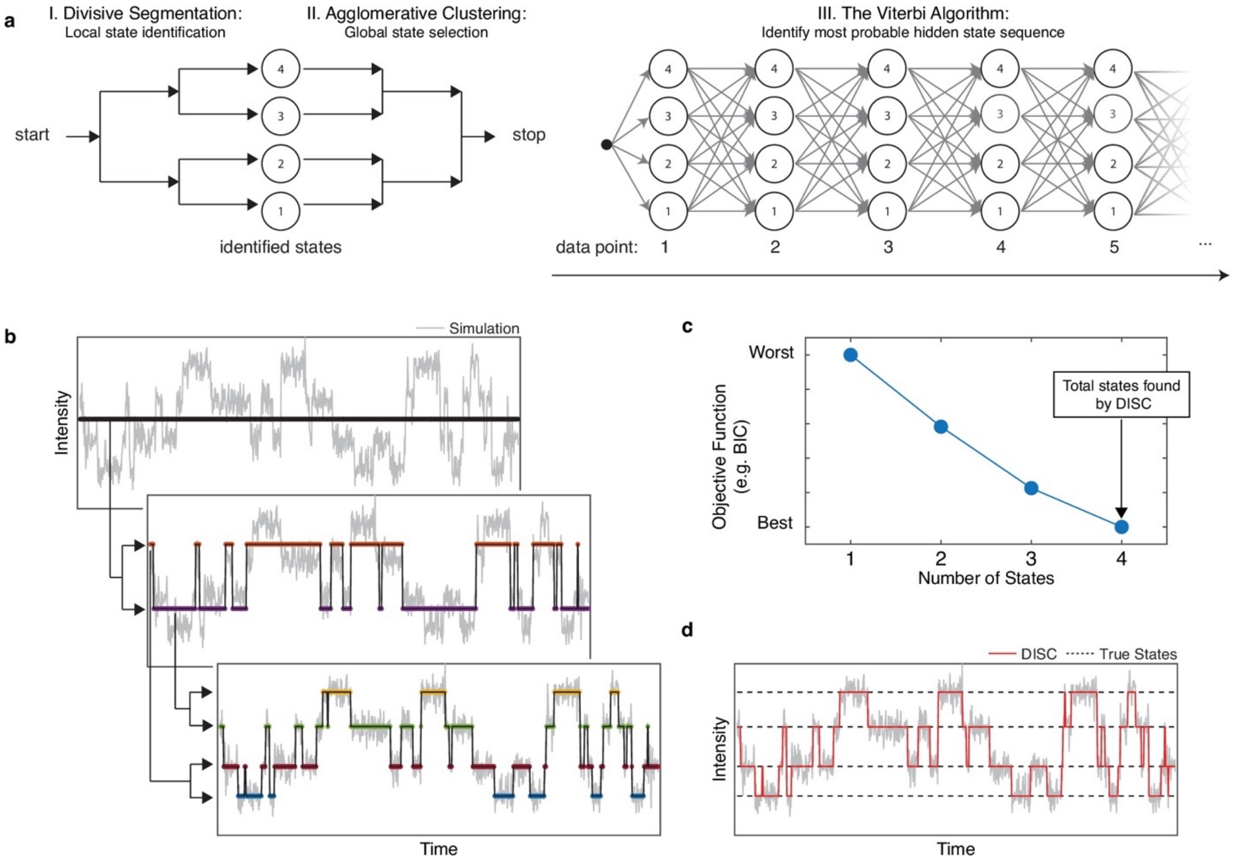

Figure 3-1: Overview of DISC. *(**a**) The major steps of the DISC algorithm combining unsupervised statistical learning with the Viterbi algorithm. (**b**) Stepwise discovery of states locally through divisive segmentation on a simulated trajectory. (**c**) HAC iteratively groups identified states to minimize an objective function for the fit of the whole trajectory to avoid overfitting. (**d**) The Viterbi algorithm is applied to identify the most probable hidden state sequence. The final fit by DISC (red) is overlaid against the true states in the simulation (dashed).*

most probable sequence of transitions between the identified states. A comparison of the overall steps involved in DISC vs common idealization approaches is provided in **Figure 3-S1**.

3.3.2 Divisive segmentation

The first phase of DISC is divisive segmentation. Consider an observed single-molecule trajectory $x = \{x_1, \ldots x_N\}$ where each x_n is the observed intensity value x at time-stamp *n* for *N* total observations contaminated by Gaussian noise. Like CP-HAC, the goal of the divisive segmentation is to identify and allocate each data-point into the optimal number of idealized states denoted by *K*. Following our Gaussian assumption, each state $\phi_j \in \{\phi_1, \ldots, \phi_K\}$ is described by the mean (μ) and standard deviation (σ) of data points allocated to the state $\phi_j = (\mu_j, \sigma_j)$. We denote a series of transitions between states as $y = \{y_1, \dots y_N\}$ where $y_n \in \{\phi_1, \dots, \phi_K\}$ and $1 \leq K \leq N$. Divisive segmentation aims to iteratively yield x by determining whether data-points in a given cluster are better described by one or two states. At the onset of divisive segmentation, it is assumed that x is described by a single idealized state. We will denote this initial fit as $y_0 = \{y_{0_1}, \dots y_{0_N}\}$ where $y_{0_i} \in \{\phi_0\}$ and $\phi_0 = (\mu_0, \sigma_0)$.

Allocating each data-point into two unique states is accomplished in two sequential phases: CP detection and k-means clustering. As opposed to standard divisive algorithms that allocate data points via k-means clustering only, we find CP identification prior to clustering advantageous. Not only does it reduce the solution domain of the objective function, but it also speeds up subsequent clustering while providing a reasonable estimate of state transitions. CP detection is performed with the popular recursive binary segmentation algorithm[27,38]. For each time stamp *n* in x, a hypothesis test is conducted to evaluate the probability that a CP occurred at position *n* via

- H_0: a CP did not occur at position *n*
- H_1: a CP did occur as position *n*

In the context of single-molecule idealization, a CP is the location indicating a significant difference in mean intensity values between the data segments separated at location *n*, where the mean values of each segment are computed by

$$\mu_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \mu_2 = \frac{1}{N-n}\sum_{i=n+1}^{N} x_i \qquad \text{Eq. 3-1}$$

To determine whether there is a statistically significant difference between the two segments, we use a two-way Student’s t-test of unequal sample size but uniform variance to evaluate the differences in mean. This is the same approach used in STaSI[28]. Specifically, a *t*-value is computed for each position *n* by

$$t_k = \frac{|\mu_1 - \mu_2|}{\sigma\sqrt{\frac{1}{n} + \frac{1}{N-n}}} \qquad \text{Eq. 3-2}$$

where⍰⍰⍰is the estimated standard deviation of uniform noise[28]. The most probable CP location *c* corresponds to the maximum t-value t_{max} given by

$$c = \underset{k}{argmax}\, t(k) \qquad \text{Eq. 3-3}$$

$$t_{max} = \max_{k} t(k) \qquad \text{Eq. 3-4}$$

For a user specified confidence interval, a critical value is used to determine whether to accept the change-point. If t_{max} > critical-value, we reject H_0 and the CP is accepted. This in turn segments the data at position *c.* As there are likely multiple CPs in x, the algorithm continues in a recursive manner by searching within each new segment $s_1 = \{x_1, \dots x_c\}$ and $s_2 = \{x_{c+1}, \dots x_N\}$. This process terminates when no significant changes in mean intensity are found within any segment. Importantly, the confidence interval set by the user plays a crucial role by acting as a hard threshold for false positive rate.

Following the completion of CP detection, times-series x can be described as a series of *C*+1 intensity segments where *C* is the total number of CP identified given by an idealized state trajectory where $y_n \in \{\phi_1, \dots, \phi_K\}$ and $1 \leq K \leq C+1$. Like CP-HAC, the next goal is to discover the optimal number of states *K* into which to cluster the *C*+1 intensity segments generated by CP detection. Rather than iteratively merging each segment like bottom-up algorithms, we use divisive segmentation to cluster the data points in a top-down fashion. For divisive segmentation, all identified segments are partitioned into two unique clusters using the k-means algorithm, where the center of each cluster is described by the mean values of the CP-idealized data points within the cluster. For computational efficiency, DISC uses a modified k-means algorithm that is both deterministic and faster than standard implementations through use of triangle inequality for computational reduction[39]. Overall, this results in a series of transitions between two states $y_1 = \{y_{1_1}, \dots y_{1_N}\}$, where $y_{1_n} \in \{\phi_1, \phi_2\}$ and $\phi_j = (\mu_j, \sigma_j)$ the corresponds to a state assignment for each observation in x.

Now that the data-points are allocated to two separate clusters with identified transitions, the goal is to determine if x is better fit with one or two states (y_0 vs y_1). Like CP detection, this decision follows a hypothesis test where

- *H₀*: the data are composed of one unique state
- *H₁*: the data are composed of two unique states

To determine whether one or two states provides a better fit, we use the Bayesian Information Criterion (BIC) which is defined in a general form as

$$BIC = -2\ln(\hat{\mathcal{L}}) + M\, ln\,(N) \qquad \text{Eq. 3-5}$$

where $\hat{\mathcal{L}}$ is the likelihood for the estimated model with *M* free parameters[40]. The likelihood that the observations x arose from a single state (y_0) is simply the product of the probability densities of a Gaussian distribution evaluated for each x_i. For the multi-state fit of y_1, the model extends to a mixture of 1D Gaussians whereby $\hat{\mathcal{L}}$ is computed as a linear combination of each *K* Gaussian components, corresponding to each state $\phi_j \in \{\phi_1, \dots, \phi_K\}$ with $\phi_j = (\mu_j, \sigma_j)$ weighted by a mixing coefficient (π_j)[41].

$$\mathcal{N}(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{\sigma^2}\right) \qquad \text{Eq. 3-6}$$

$$\hat{\mathcal{L}} = \prod_{i=1}^{N} \sum_{j=1}^{K} \pi_j * \mathcal{N}(x_i \mid \mu_j, \sigma_j) \qquad \text{Eq. 3-7}$$

To test the null hypothesis that x is described by one state instead of two, BIC values are computed for x with a fit of a single-state (BIC_1) and fit with two-states from divisive segmentation (BIC_2). If $BIC_2 > BIC_1$, *H₀* is accepted and we believe x is sufficiently described by a single-state. If the $BIC_2 \leq BIC_1$, the *H₀* is rejected and x is split into two states. Assuming two-states are identified on the first iteration, the sequential process of CP detection and bi-partitioning with k-means clustering continues in a recursive fashion within data points belonging to each of the newly identified states (**Figure 3-1b**)[32]. Divisive segmentation continues to identify sub-clusters within each identified cluster until no cluster can be further partitioned. Overall, the recursive bi-partitioning and binary decision making in both CP detection and divisive clustering result in a reduced time complexity on the order of $O(N\log(N))$.

3.3.3 Agglomerative clustering

Self-termination of divisive segmentation results in a series of estimated states and transitions within the trajectory. While this algorithm is exceptionally fast owing to its top-down greedy design, the reliance on local choices for state assignments rather than evaluation of the entire trajectory for an optimal decision can produce sub-optimal results. This error often surfaces as an over-sampling of the number of states and an under-sampling of the kinetic transitions. Over-fitting the number of states results from a downward dissemination of error from early splits: if bisecting a given cluster is suboptimal, two different parent clusters may each produce highly similar and redundant sub-clusters thereby overfitting the number of states.

Fortunately, this over-estimate of the number of states can be corrected using bottom-up clustering. Therefore, the second phase of DISC uses HAC to compute the similarities between all identified states at a global level and assess the fit of the whole trajectory rather than segmented portions. Like CP-HAC schemes, an objective function is used to determine the overall fit and number of states in the trace by merging highly similar clusters whose separation may arise during divisive segmentation (**Figure 3-1b**). For a general application, we continue to use BIC for evaluating fit vs complexity. The similarity between neighboring states ϕ_i and ϕ_j is computed using Ward's minimum variance method[42], which considers the number of data points in each state (*n*) and the Euclidean distance between the means of the states by

$$d(\phi_i, \phi_j) = \sqrt{\frac{2n_{\phi_i}n_{\phi_j}}{n_{\phi_i} + n_{\phi_j}}} \left\| \mu_{\phi_i} - \mu_{\phi_j} \right\|_2 \qquad \text{Eq. 3-8}$$

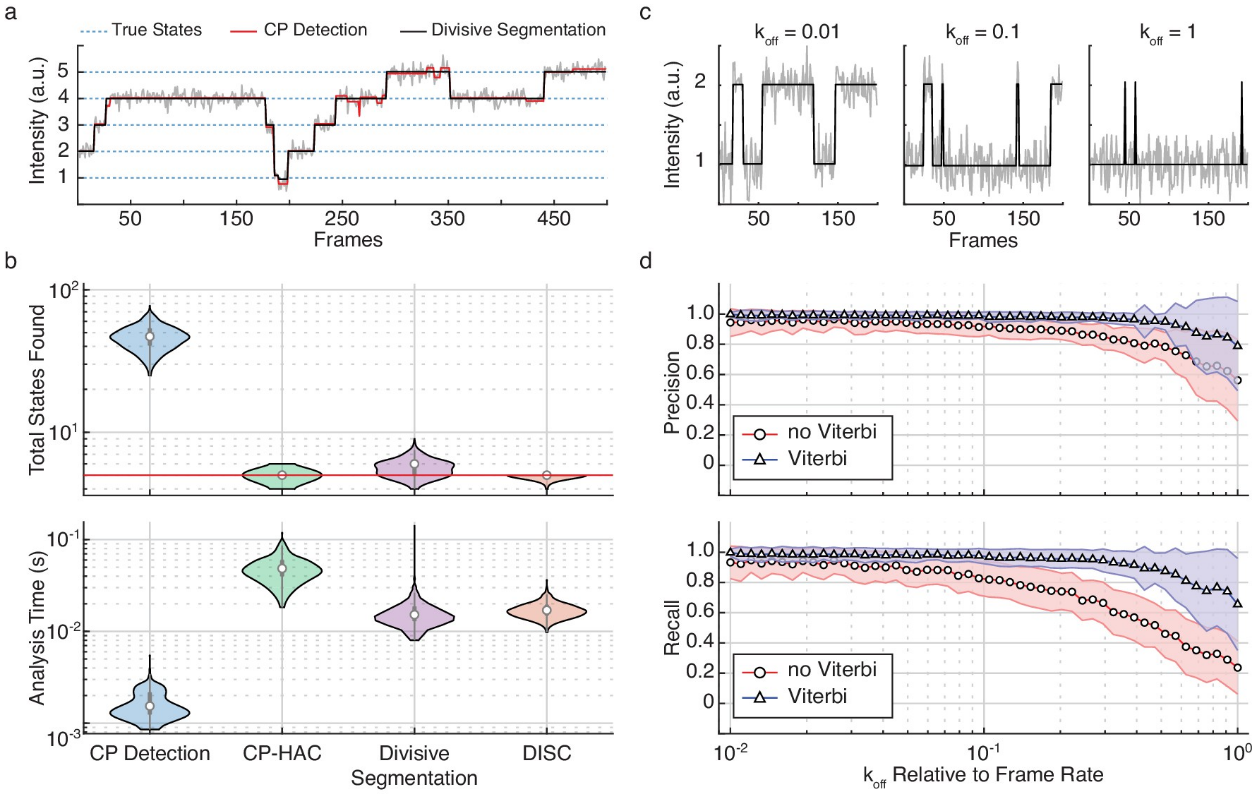

Figure 3-2: Refinement of divisive segmentation. (a) *Example trajectory simulated with 5 states (blue) and a SNR = 5 overlaid with fits obtained from change-point detection (red) and divisive segmentation (black). (**b**) Violin plots showing the number of identified states (top) and analysis time (bottom) of each algorithm across 500 simulated trajectories featuring 5 true states (red line). (**c**) Example simulations of a two-state system with a* k_{on} *= 0.02 frames*$^{-1}$ *and varying* k_{off}*. (**d**) Precision (top) and recall (bottom) values obtained with CP detection (no Viterbi) and Viterbi refinement obtained across 100 trajectories per* k_{off} *(mean ± s.d.).*

The resulting minimal BIC value corresponds to the final number of states within the trajectory (**Figure 3-1c**).

The improvement of HAC on divisive segmentation for state detection is shown in **Figure 3-2**. Although divisive segmentation alone tends to slightly over-estimate the number of states, it provides a more reasonable estimate than CP detection alone (**Figure 3-2a**). The comparative performances in terms of speed and accuracy of these algorithms is further explored in **Figure 3-2b**. While CP detection alone (**Figure 3-2b, blue**) is very fast, it consistently yields a higher number of total states as compared to the ground truth. As CP-HAC frameworks must explore this large state space in its entirety, they can achieve higher accuracy than CP detection alone, but they are much slower algorithms (**Figure 3-2b, green**). In contrast, the use of top-down clustering in divisive segmentation dramatically reduces the total state-space for exploration, resulting in a faster algorithm than CP-HAC with much higher accuracy than CP detection alone (**Figure 3-2b, purple**). Finally, the sequential combination of divisive segmentation and HAC used in DISC lead to the highest state detection accuracy with minimal computational cost (**Figure 3-2b, orange**).

3.3.4 The Viterbi algorithm

Following state refinement with HAC, the trajectory is again described as a series of temporal transitions between identified intensity states. Although the overall states are well estimated at this point, fast transitions are often missed during CP analysis of single-molecule trajectories[29]. To ensure events are accurately detected, the final phase of DISC applies the Viterbi algorithm[37].

The goal of the Viterbi algorithm is to identify the most probable sequence of hidden states through a series of observations. In our scenario, we have K total states and N total observations in our trajectory x. In a naïve manner, determining the most probable sequence of hidden states y could be accomplished by evaluating the likelihood of every possible hidden state sequence and choosing the most probable. However, as there are K^N possible paths though the trajectory, this quickly becomes computationally intractable. A solution to this problem is the Viterbi algorithm, which makes use of dynamics programming to store only the most optimal state sequencing leading up to a given time point[37]. In general, the Viterbi algorithm uses the observation that that the most probable state sequence leading up to data point n can be deduced by examining the most probable path leading up to the previous time point, $n-1$. Dynamic programming is used to keep track of all the optimal state sequences leading to all possible states for a given time point $n-1$ which reduces the amount of required computations. Since there are there are K states at time step $n-1$, the Viterbi algorithm stores K possible state sequences leading up the previous time point $n-1$. At time point n, there are now K^2 paths to consider, given K possible paths leading out of K states. By examining the optimal sequence up to time point n and considering the probability of state transition between time points *n-1* and *n*, the most optimal state sequence up to time point *n* can be constructed. The state assignment of the first data point in the sequence can be determined by a provided initial probability of observing each state. Therefore, the time complexity of idealization with Viterbi is quadratic in the number of states *K* and linear with the number of observations N, $O(K^2N)$, which is dramatically lower than an exhaustive search.

Formally, the Viterbi algorithm is described with a $K\ x\ N$ trellis for states $j\ \in K$ and observations $n\ \in N$ (**Figure 3-1a**). Each cell of trellis $v_n(j)$ represents the probability of being in state *j* after seeing the first *n* observations and passing through the most probable state sequence for the given model parameters, λ. The value $v_n(j)$ is computed by recursively taking the most probable path up to this cell by

$$v_n(j) = \max_{y_1,\cdots y_{n-1}} P(y_1 \ldots y_{n-1},\ x_1 \ldots x_n, y_n = j \mid \lambda) \qquad \text{Eq. 3-9}$$

where λ is a first order Markov process of $\lambda = (\pi, a, b)$. The primary components of a hidden Markov model include the initial probability of observing each state ϕ_j given by π where $\sum_{j=1}^{K} \pi_j = 1$; a transition probability matrix a of size *K x K* where each element a_{ij} is the probability of moving from ϕ_i to ϕ_j , each element a_{ii} is the probability of staying in ϕ_i and $\sum_{j=1}^{K} a_{ij} = 1$; and an emission probability matrix b of size *K x N* where each element $b_j(x_n)$ is the probability of an observation x_n arising from ϕ_j. The values of each component are computed for each trajectory using the fits obtained from sequential steps of divisive segmentation and HAC. Using these parameters, we can compute the most probable path for arriving in ϕ_j at time points *n* by the following recursion

$$v_n(j) = \max_{i \in N}\{v_{n-1}\,(i) a_{ij} b_j(x_n)\} \qquad \text{Eq. 3-10}$$

$$\psi_j(n) = \operatorname*{argmax}_{i \in N}\{v_{n-1}\,(i) a_{ij} b_j(x_n)\} \qquad \text{Eq. 3-11}$$

where $\psi_j(n)$ is a helper function to store the $n-1$ state index i on the highest probability path.

Upon termination, the forward likelihood of the entire state sequence y up to time point N+1 having been produced by the given observations and HMM parameters is

$$P(y \mid x, \lambda) = \max_{i \in N}\{v_T(i)\} \qquad \text{Eq. 3-12}$$

Deducing the most probable hidden state sequence y through observations x can be accomplished in a backtracking step by

$$y_n = \psi_n(y_{n+1}) \qquad N \geq n \geq 2 \qquad \text{Eq. 3-13}$$

To assess the improvement of the Viterbi algorithm for event detection, we simulated a two-state system with a constant k_{on} and varying k_{off} rate (**Figure 3-2c**). As shown previously[29], we found that results from CP detection alone were accurate for slower events, but often failed to identify faster transitions (**Figure 3-2d**). By refining the results of unsupervised clustering with the Viterbi algorithm, we found that event detection accuracy was significantly improved over CP detection and clustering alone across two-orders of magnitude of varying k_{off} (**Figure 3-2d**). Notably, as the changes in rates also affect the change in state occupancy, the high accuracy values returned after Viterbi refinement further highlight the power of DISC for resolving short-lived and rare transitions (**Figure 3-S2**). In general, this improvement was anticipated since we are not the first to apply the Viterbi algorithm to the problem of idealization using unsupervised clustering. Although commonly used in the application of HMMs for hard assignment of data-points into *K* discrete states, the SKM algorithm has shown that the Viterbi algorithm can successfully decode a path sequence following state clustering using k-means as opposed to more rigorous HMM training procedures[20,36]. Therefore, both SKM and DISC can yield the event detection power of standard HMM approaches without the need of rigorous model training. However, unlike SKM, DISC has the added benefit of identifying the states naturally without the need for any user supervision such as initial state specification. This makes DISC a powerful alternative as a computationally efficient unsupervised single-molecule analysis algorithm.

3.3.5 Modular nature of DISC

While the parameters used above are valid for trajectories with Gaussian noise, we do not claim they are optimal for all experimental modalities. We have intentionally developed DISC as a flexible framework for adaptation to different types of data. Although we used the Student's t-test for CP detection in the presence of Gaussian noise, additional merit functions may be more appropriate in different situations[31,43,44]. The same holds true for the use of BIC for state selection as other objective function can be substituted as needed. For example, the harsh penalty for parameters in BIC can lead to underfitting in certain cases; therefore, less stringent Akaike information criterion (AIC) or Hannan-Quinn information criterion (HQC) may be more appropriate depending on the separation of states and noise[45,46]. It is important to note that while DISC performs idealization through unsupervised clustering, obtaining accurate results does require the user to determine the appropriate information criterion and CP detection methods as idealization results heavily depend on these variables. Critically, the central innovation of DISC is to take advantage of the best features of both top-down and bottom-up forms of cluster identification that leads to both fast and accurate state detection.

3.4 Results

3.4.1 Validation of DISC on simulated data

We validate DISC using simulated single-molecule trajectories using kinetic parameters obtained from our recent studies exploring the regulatory mechanisms of cyclic nucleotide binding domains (CNBDs) from hyperpolarization-activated cyclic nucleotide gated ion channels (HCN) which regulate pacemaking in heart and brain cells (see **3.6.1**)[47,48]. In these experiments, isolated CNBDs are tethered into ZMWs whereupon we monitor the binding and unbinding dynamics of fluorescent cyclic nucleotides (e.g.

fcAMP) at physiological concentrations to uncover the elementary dynamics associated with channel gating. While ligand binding has been observed at the single-molecule level via both FRET and CoSMoS (co-localization), we adapt our simulations to the latter case so our dynamics are not limited in time by acceptor photobleaching. Notably, trajectories obtained with CoSMoS exhibit heterogeneous bound intensity values which vary with each binding event (**Figure 3-S3**, described in detail in section **3.6.1**). While we are uncertain as to the exact source of this fluctuation, it is likely caused by shifts of the molecule in the heterogeneous excitation field of the ZMW or dye photodynamics[5,49]. While the excitation field changes particularly sharply in ZMWs, TIRF and confocal microscopy also contain a heterogenous excitation field[50]. Minor changes in apparent dye brightness due to dye conformational or photodynamics (such as in protein-induced fluorescence enhancement, PIFE), shifts of dye orientation, or partial quenching via electron transfer are all commonly observed[51]. Thus, heterogeneous intensity values are a common and inconvenient feature in real life single-molecule fluorescence data. Including this additional noise source in our simulations yields a closer representation of experimentally obtained data.

In total, we simulated 4,000 trajectories composed of 2,000 data points each, totaling 8 x 10^6 data points. Each trajectory is 200 seconds in duration collected at frame rate of 10 Hz. We varied the complexity of the trajectory by simulating one to four independent CNBDs inside a given ZMW (two to five intensity states) and vary the signal to noise ratio (SNR) according to typical values from a ZMW experiment using Gaussian noise (**Figure 3-S3c**)[47,48]. We include the observed heterogenous bound intensities by randomly modulating each binding event according to our fit of the experimentally observed data (**Figure 3-S3, Figure 3-S4**). Given that each simulation features a different number of possible states, the total time spent within each state changes; therefore, these simulations also address the ability to capture states with unequal and even rare observation probabilities (**Figure 3-3**). We benchmark the results of DISC against commonly used HMM and CP-HAC methods: vbFRET and STaSI[18,28]. These algorithms were chosen following the results of a recent comparative study that determined these to be

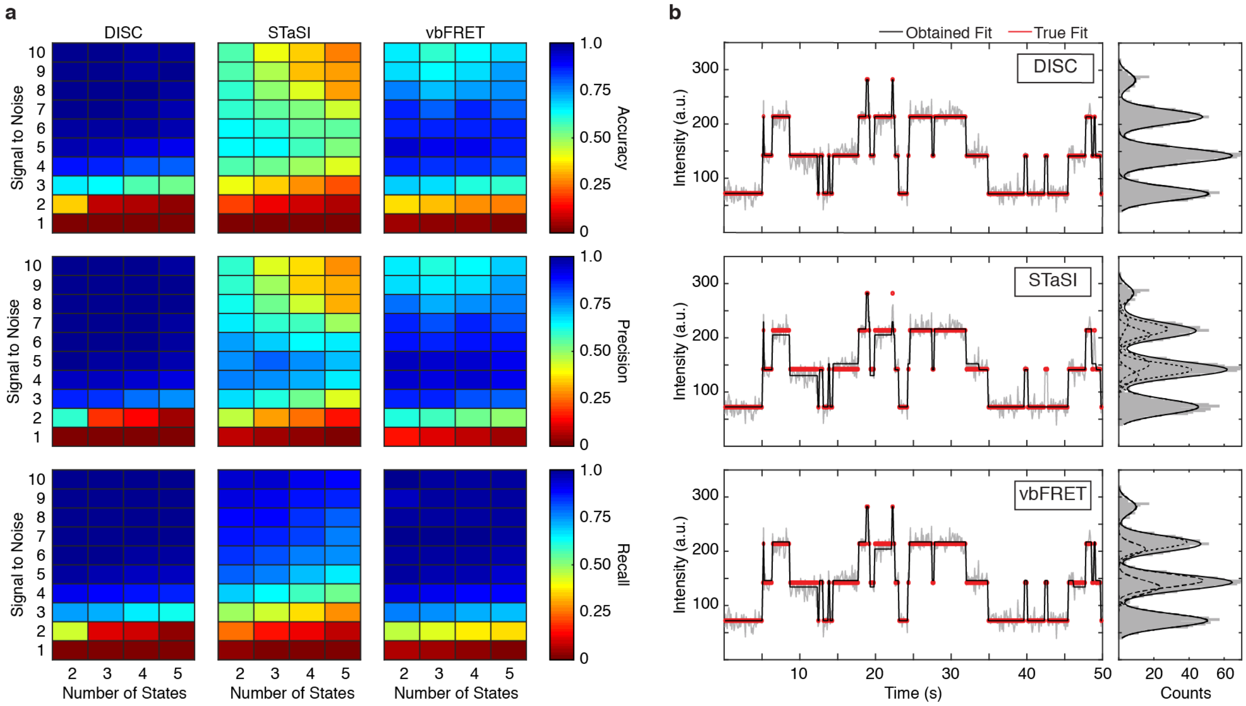

Figure 3-3: Standardizing algorithm performance. *(**a**) Average accuracy (top), precision (middle) and recall values (bottom) computed for DISC, STaSI, and vbFRET across 100 trajectories at the specified signal to noise and number of states (**Methods**). (**b**) Example simulated trajectory with 4 true states (red) fit and added Gaussian noise (grey) to SNR = 6 overlaid with fits (black) from DISC (top), STaSI (middle), or vbFRET (bottom).*

the best performers among their class of analysis methods[29]. In addition, DISC, STaSI and vbFRET all perform trajectory-by-trajectory idealization and are written entirely in MATLAB (MathWorks) which standardizes computational performance (see **3.6.2**).

Across all the simulations, DISC provides the highest average accuracy, precision and recall (**Figure 3-3a,** terms defined in **3.6.2**). While no algorithm can idealize a trajectory in the presence of SNR = 1, DISC returns the lowest accuracy at SNR = 2. We suspect this result is due to using the robust BIC for state detection and accuracy would likely be improved with less penalizing objective functions, such as AIC. While vbFRET performs the best at SNR = 2, the overall accuracy is still quite low: an average accuracy value for each number of simulated states is near chance. This low value demonstrates the inability of many algorithms to analyze data in presence of high noise and reinforces the common practice of discarding noisy data to create a more reliable dataset. For SNR > 3, which accounts for most of our experimentally obtained data (**Figure 3-S4**), DISC performs exceptionally well with highest average accuracy (0.91 ± 0.05) and is robust against false positives (precision = 0.96 ± 0.04) and false negatives (recall = 0.93 ± 0.03) across all simulated conditions (**Figure 3-3a**). While vbFRET matches the recall of DISC in this SNR range (0.94 ± 0.05), the tendency to overfit the number of states at higher SNR lowers precision (0.80 ± 0.18) and overall accuracy (0.76 ± 0.19)**.** We find STaSI returns the lowest overall accuracy (0.47 ± 0.17) likely do to an overfitting the number of states (precision = 0.57 ± 0.2) and a tendency to miss transitions (recall = 0.75 ± 0.10). Notably, DISC is the only method unaffected by inclusion of heterogeneous state intensities of fcAMP likely due to the use of a Gaussian derived BIC for state selection (**Figure 3-S5)**.

Critically, DISC not only returned high accuracy results, DISC was also much faster than the other methods. Idealization of all 4,000 trajectories by DISC was completed in just over two minutes, whereas STaSI took over fifteen minutes and vbFRET took over twelve hours. To thoroughly explore the computational efficiency of DISC, we simulated data with increasing durations per trajectory at a constant SNR and number of states. Remarkably, we find DISC is 400-fold to 1,200-fold faster than vbFRET and 2-fold to 8,700-fold faster than STaSI due to STaSI's quadratic time dependence (**Figure 3-4a**)[28]. For example, a trajectory of 10^6 data points can be analyzed by DISC in 10 seconds compared to 3 hours for vbFRET and 27 hours for STaSI. Thus, DISC can handle the analysis of long trajectories, unlike CP-HAC methods. While fluorescence measurements from a single fluorophore at room-temperature rarely contain this many data points, large trajectory lengths are common in non-fluorescence experiments or fluorescence experiments with replenishing fluorescent labels such as in single-molecule genome sequencing and studies of catalysts via fluorogenic reactions.[52-54]. To evaluate performance on more typical data, we compared the results of each algorithm on simulated smFRET trajectories that are limited in duration by acceptor photobleaching (**Figure 3-S6**). For simulations featuring two or three states FRET, we find DISC 5.5-fold faster than STaSI and 235-fold faster than vbFRET while maintaining the highest accuracy. This

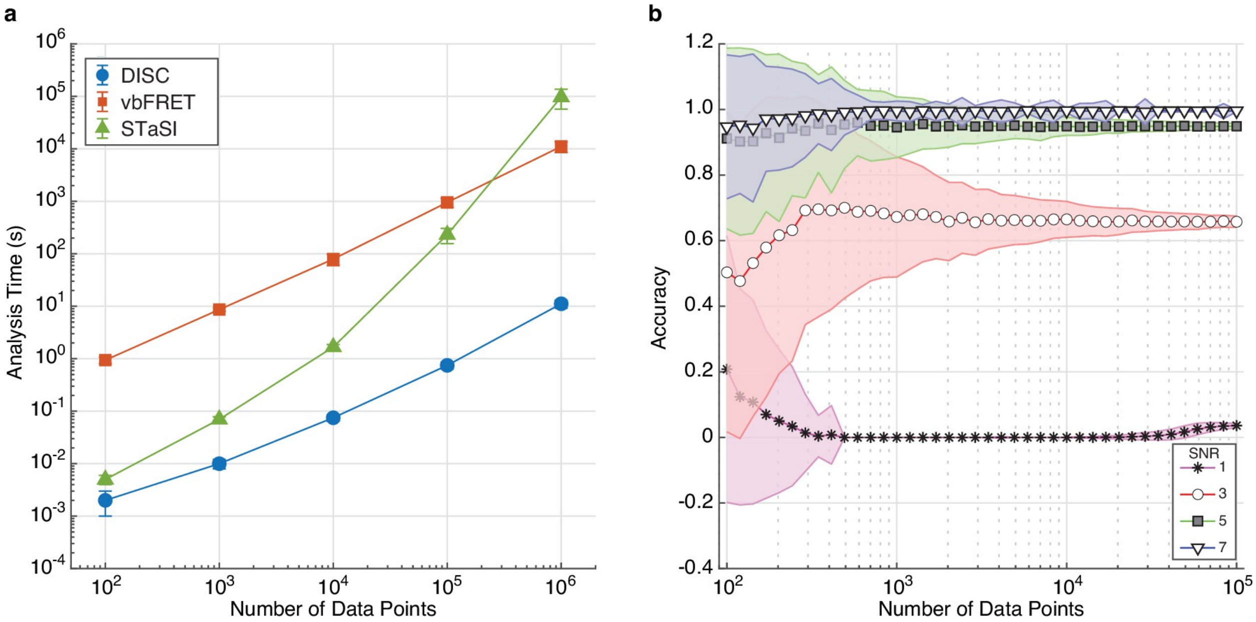

Figure 3-4: The effect of trajectory length on DISC performance. *(**a**) Computational time (mean ± s.d.) of each algorithm for analyzing single trajectories of varying lengths. The test was performed with an Intel Xeon, 3.50 GHz processor running MATLAB 2017a. (**b**) Accuracy (mean ± s.d., N = 5000) of DISC for simulated trajectories of a two-state model with varying SNR and total number of data points.*

result validates the use of DISC for the analysis of large volumes of shorter fluorescence trajectories, especially compared to HMM approaches. This feature is particularly important as advances in hardware such as CMOS cameras and lab-on-chip methods generate larger smFRET data sets[7].

Finally, we evaluate the effect of trajectory duration on DISC accuracy. As expected, we find that the accuracy increases with increasing number of data points per trajectory. This result also indicates the minimum number of data points needed for an accurate idealization for a given SNR (**Figure 3-4b**). Overall, the results of our simulations suggest DISC is more or comparably accurate and critically, is substantially faster than standard idealization approaches, making it an enabling technology for analysis of high-throughput single-molecule experiments.

3.4.2 The binding of cAMP to HCN2 CNBDs is non-cooperative

To verify performance of DISC in an experimental configuration with high volumes of experimental data, we analyzed a large single-molecule data set obtained from ZMWs that explore HCN dynamics (**Figure 3-5**). Previous macroscopic studies of HCN channel gating have revealed that ligand binding to CNBDs exhibits both positive and negative cooperativity depending on the ligation state and the membrane potential[55,56]. Cyclic AMP regulates cardiac pacemaking via HCN channels and, therefore, this unusual allostery has significant physiological implications. However, as this allosteric analysis was based on global fits of ensemble binding data, the reliability of model parameters remains an open question[23]. Evaluating cooperativity is an experiment well-suited for single-molecule investigation given the ability to observe the total time a molecule spends in each liganded state and directly extract state transition probabilities. Therefore, to directly assess the cooperativity between HCN2 CNBDs upon ligand binding, we use single-molecule fluorescence and monitor the binding of individual fcAMP molecules to our previously described tetramerized CNBDs inside ZMWs. (**Figure 3-5a**)[48].

Our initial dataset included 13,670 ZMWs each monitored for 800 seconds at a sampling rate of 10 Hz (see **3.6.3**). All trajectories were obtained in the presence of 1 µM fcAMP which is near the ligand dissociation constant for individual CNBDs[48]. As shown with other high-throughput collection platforms, an essential part of analysis at this scale is the application of stringent criteria to select traces that yield meaningful information about the system[6,7]. Therefore, we first analyzed all trajectories with DISC to find reliable data prior to trace selection. DISC successfully processed this entire data set within 20 minutes using a standard MacBook Air (1.6 GHz Intel Core i5). The same analysis completed with STaSI yielded unphysical results in 4 hours (**Figure 3-S7**). We estimated analysis with vbFRET would take weeks to complete and was therefore not performed. While correcting for non-specific binding is often a necessity in CoSMoS experiments, we find the passivated surfaces within the ZMWs greatly reduce non-specific absorption of fcAMP to either the metallic or glass surfaces, thus minimizing this concern (**Figure 3-S8**, see **3.6.3**)[30,52,57]. Using the idealized fits obtained from DISC, we screened our data to select reliable trajectories for our analysis. Standard cut-offs in state separation, the total number of observed states, and a filter for kinetic activity ensured that each trajectory arose from a ZMW featuring a singly occupied

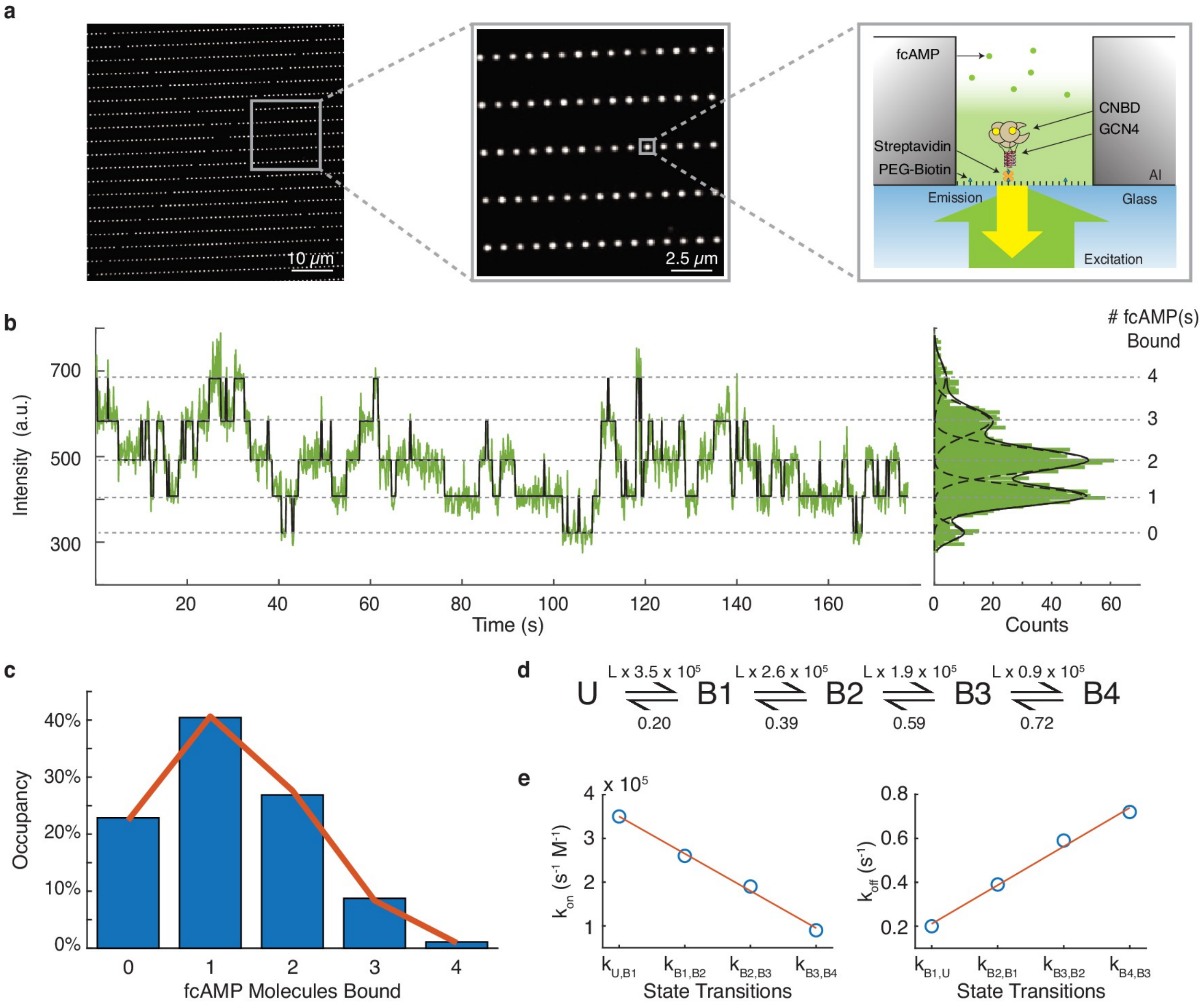

Figure 3-5: DISC analysis of HCN2 CNBDs. *(**a**) Representative ZMW arrays for observing fcAMP binding to tethered tetrameric CNBD. (**b**) Representative time series of 1 μM fcAMP binding to tetrameric CNBD fit with DISC with up to four fcAMP molecules binding simultaneously. (**c**) Observed distribution of fcAMP occupancy fit with a binomial distribution (orange). (**d**) Sequential model of four binding steps and one unbound state with globally optimized rate constants. The rate constants are given as s^{-1} or $s^{-1}M^{-1}$ where L is the ligand concentration in M. e) Linear regression of rate constants k_{on} ($m = -8.5 \times 10^4\ s^{-1}M^{-1}$, $b = 4.35 \times 10^5\ s^{-1}M^{-1}$, $R^2 = 0.99$) and k_{off} ($m = 0.18\ s^{-1}$, $b = 0.035\ s^{-1}$, $R^2 = 0.99$) for each sequential state.*

and functional tetrameric CNBD. Notably, we noticed an asynchronous decay of protein activity over excitation time that may be caused by singlet-oxygen formation and subsequent inhibition of cyclic nucleotide binding to CNBDs (**Figure 3-S9**, see **3.6.3)**[58]. In conclusion, we retained 293 molecules totaling 1.2 x10^5 seconds of combined protein activity across 53,474 events (**Figure 3-S10**).

To determine if the binding of cAMP to CNBDs is cooperative, we first calculated the total time each molecule spends in each of the liganded states (0 to 4 fcAMPs) using the state assignments from the idealized fits (**Figure 3-5b**). The resulting distribution of state occupancies is fit with a binomial distribution to evaluate the independence of each CNBD (**Figure 3-5c,** see **3.6.3**). Our binomial fit matches the distribution well and returns the probability of occupancy at 1 µM fcAMP for a single CNBD in the tetramer as 31% which is similar to our previous monomeric CNBDs studies[48], suggesting a lack of cooperativity between the CNBDs. Notably, our measured state-occupancy distribution is strikingly different that than the unusual cooperativity modeled from either activated or non-activated channels (**Figure 3-S11**)[55,56]. We further explored the underlying dynamics of our data using the idealized single-molecule transitions obtained from DISC. Using QuB, we built a simple HMM of sequential ligand binding across four binding sites that was globally optimized across each molecule's idealized state trajecotry[14,15] (**Figure 3-5d**). As expected for non-cooperative processes, the optimized k_{on} and k_{off} rates for each state transition exhibit a strong linear relationship (**Figure 3-5e**). Combined, these results strongly suggest that CNBD units act independently during ligand binding. We postulate that the macroscopically observed cooperativity is either an artifact of model fitting or that it requires the presence of the transmembrane domains of the HCN channel and is not an intrinsic property of the CNBDs.

3.5 Discussion

We developed a new algorithm for rapid and accurate unsupervised idealization of single-molecule trajectories. Our approach combines unsupervised statistical learning of discrete states with the event detection power of the Viterbi algorithm to quickly identify both significant states and transitions in a model-independent manner. Software implementing the DISC algorithm that includes a graphical user interface is available at https://github.com/ChandaLab/DISC.

Like CP-HAC methods, DISC is not a fully probabilistic approach. While fully probabilistic HMM training approaches are beneficial for providing unbiased estimates of the parameter distributions, their high accuracy comes at the cost of significantly increased computational time. This cost is especially apparent for the recently developed infinite HMM approaches that aim to learn the true number of states from a potentially infinite number of possibilities. However, accomplishing this task costs hundreds of iterations per trace to provide a reproducible fit with some approaches taking days to analyze single trajectories[23-25]. Thus, while exhaustive search algorithms may be desirable in other contexts, they are clearly not suited for large datasets associated with high-throughput experiments. In contrast, by simulating data closely resembling the binding of fcAMP to pacemaker channels, we find that DISC surpasses the accuracy of common CP-HAC and HMM algorithms with a dramatic improvement in computational speed. Therefore, DISC satisfies the need for accuracy and speed in high-throughput analysis. In this regard, DISC is like the SKM algorithm for estimating the parameters of an HMM without direct HMM training. However, unlike SKM which relies on user-specified states, DISC uses unsupervised statistics to learn the states. Therefore, DISC offers the idealization power of SKM with the state-learning capabilities of CP-HAC.

We used DISC to analyze a large dataset obtained from ZMWs to evaluate cooperativity between CNBDs from pacemaker ion channels upon ligand binding. The rapid and robust idealization provided by DISC enabled stringent trace selection to ensure only reliable trajectories were analyzed. These data show that, in contrast to model inferences from ensemble measurements of HCN2 channels, CNBD tetramers do not exhibit cooperative ligand binding. This result suggests that allosteric interactions between binding sites may be coordinated by the channel's transmembrane domains.

Although we have demonstrated the use of DISC on single-molecule fluorescence data, the framework can be easily extended to other data paradigms due to its modular nature. For example, the use of BIC for state determination or the Student's t-test for change-point analysis could be interchanged with other information theoretic approaches or merit functions where appropriate. To allow for easy comparison of a given data set, the provided software and graphical user interface (GUI) allows the user to select from several options the desired parameters such as choice of information criteria. This flexibility makes DISC suitable for a wide array of experimental data provided it can be described as a series of transitions between discrete states, including, for example, single-channel current recordings, force spectroscopy and smFRET. However, there is no inherent knowledge within DISC to consider various sources of experimental noise, such as photo-blinking or baseline drift; therefore, correcting for these noise sources prior to DISC analysis will likely improve idealization accuracy.

Finally, our results show that DISC provides a dramatic improvement in computational speed over current state-of-the-art approaches while either improving or maintaining high accuracy for both state determination and event detection. This increase in speed is directly applicable to analyzing the growing datasets obtained in single-molecule fluorescence paradigms to adequately sample population dynamics. For example, the use of sCMOS camera enables smFRET measurements of tRNA conformational changes during protein translations across thousands of molecules simultaneously with millisecond resolution[7]. Additionally, magnetic tweezers have enabled week-long mechanical measurements of single-protein folding and unfolding, shifting observable dynamics to pathological time-scales and allowing the detection of rare events[11]. Thus, highly computationally efficient and robust algorithms such as DISC may be well suited for analysis of a wide variety of single molecule datasets beyond the standard smFRET data.

3.6 Methods

3.6.1 Single-molecule simulations

Single-molecule trajectories were simulated as a Markov process of transitions between discrete states. All simulations were performed with a frame rate of 10 Hz and featured variable total durations, SNR, and number of states. The primary kinetic scheme used was adapted from our recent studies of fcAMP binding to isolated monomeric CNBDs[48]. This model is a four-state scheme where both the unbound (U) and bound states (B) exhibit conformational changes ($U^* \leftrightharpoons U \leftrightharpoons B \leftrightharpoons B^*$), yet exhibit only two different observable states (e.g., U*/U are indistinguishable via fluorescence intensity, as are B/B*).

fcAMP binding occurs between U and B. The rate constants (s^{-1} or M^{-1} s^{-1}) are: k_{U*U} = 0.15; $k_{U,U*}$ = 0.04; $k_{U,B}$ = 2.3 x 10^{-6} * [fcAMP]; $k_{B,U}$ = 0.95; $k_{B,B*}$ = 0.51; $k_{B*,B}$ = 0.31 at 1 µM fcAMP. To mimic the tetrameric nature of HCN channels with no cooperativity, we extrapolated up to four bound states by summing independent CNBD trajectories prior to the addition of noise. To include realistic SNR, state-intensities, and heterogeneity distribution of bound intensities, we analyzed the direct fcAMP excitation and emission trajectories following acceptor photobleaching from the monomeric CNBD dataset used in our previous work (**Figure 3-S3**)[48]. This dataset consisted of 861 single molecules for a combined acquisition time of 44,090 seconds (4775 total binding events). All trajectories had a SNR > 2 and all events persisted for longer than 2 frames, which resulted in an imbalance in the bound and unbound events. For each simulated trajectory, state intensities were each drawn from log normal distributions fit to monomeric CNBD single-molecule data, with average intensities between subsequent states being uniform. Gaussian noise was applied to trajectories at specified SNR. To quantitate the heterogeneous intensities from fcAMP binding, the mean of individual bound event intensities was taken for each identified event, so long as the event was > 2 frames in duration. Heterogeneity was computed as the absolute percent difference for each event vs the mean bond intensity for the given trajectory by:

$$Percent\ Heterogenity\ = \left| \frac{\langle \mathrm{I} \rangle_{bound} - \langle \mathrm{I} \rangle_{unbound}}{\langle \mathrm{I} \rangle_{unbound}} \right| \times\ 100\% \qquad \text{Eq. 3-14}$$

The heterogeneity of unbound events was minimal and was therefore not included in the simulations. For each simulated event, heterogenous bound intensity emissions were each drawn from an exponential fit monomeric CNBD single-molecule data. Gaussian noise was added to trajectories as specified.

Simulated smFRET data was downloaded from the kinSoftChallenge on June 11th, 2019 (https://sites.google.com/view/kinsoftchallenge/home). Data used came from the provided training data sets titled: “Level 1” and “Level 2” with folder names “sim_190212_194543_level1” and “sim_190212_202530_level2”.

3.6.2 Algorithm performance

DISC, STaSI, and vbFRET are all written entirely in MATLAB (MathWorks). Each algorithm was used outside of their graphical user interfaces (GUIs) to more accurately compare the computational time of native functions within each algorithm. User parameters in DISC include: the confidence interval of CP detection and the objective function for clustering. Unless otherwise stated, a 95% confidence interval was applied for CP detection and BIC was used for all clustering. For analysis with STaSI and vbFRET, we used the recommended default values set by their authors[18,28]. For STaSI, this means a 99.8% confidence interval of CP detection. In vbFRET, users must provide the number of states and fitting attempts per trace (left at the default value of 10). To circumvent providing the number of states, we modified the provided *vbFRET_no_gui.m* script to perform analysis outside of the vbFRET GUI. The modified script begins by fitting the trace to one state and increases the number of states until two more beyond the number of states with the maximum evidence to ensure the maximum fit has been obtained. As no changes were made to native vbFRET functions, implementing this script has no effect on vbFRET's accuracy. We expect changing parameters in both STaSI and vbFRET may lead to different results; however, it was not our goal to optimize the use of these algorithms. Also, as a thorough investigation into the performance of STaSI and vbFRET has been conducted elsewhere, we did not investigate why these algorithms presented lower performance than DISC[29].

All quantifications of computational time were performed using the tic and toc functions in MATLAB. For idealization accuracy, each event returned by a given algorithm is classified as a True Positive (TP), False positive (FP), or False Negative (FN). We define a TP as being in the correct state (± 10 % the correct intensity level's standard deviation) and correct event duration (± 1 frame) for a given simulated event. FPs are either added events or correct events in the wrong state. FNs are missed events. For each trajectory, we computed accuracy, precision, and recall as:

$$Accuracy\ = \frac{TP}{(TP+FP+FN)} \qquad \text{Eq. 3-15}$$

$$Precision = \frac{TP}{(TP+FP)} \qquad \text{Eq. 3-16}$$

$$Recall = \frac{TP}{(TP+FN)} \qquad \text{Eq. 3-17}$$

Accuracy represents the overall performance, whereas precision and recall highlight the false positive error rate (overfitting the data) and false negative rate (underfitting the data), respectively.

3.6.3 Single-molecule florescence microscopy in ZMWs

The expression, purification, biotinylation, and fluorescence labeling of tetrameric CNBDs were performed as previously described[48]. Non-commercial arrays of ZMWs were purchased from Pacific Biosciences. These waveguides featured a polyphosphonate passivation layer on the aluminum walls and a biotinylated polyethylene glycol (PEG) layer on the glass surface to reduce non-specific binding (**Figure 3-S8**)[57,59]. The PEG-Biotin surface was incubated with 0.05 mg/mL streptavidin (Prospec, cat # PRO-791) for 5 minutes in a buffer containing: 40 mM HEPES, 600 mM NaCl, 20% glycerol, 2 mM TCEP, 0.1 % LDAO (Sigma, cat # 40236), 2 mg/mL bovine serum albumin (BSA), 1mM Trolox, 2.5 mM protocatechuic acid (PCA), pH 7.5 (Buffer A). After incubation, the ZMW chip was thoroughly rinsed with Buffer A to remove unbound streptavidin. Next, biotinylated tetrameric-CNBDs were diluted in Buffer A with the addition of the PCA/ PCD oxygen scavenging system by adding 250 nM of protocatechuate 3,4-dioxygenase (PCD) from Pseudomonas sp. (Sigma, cas no. 9029-47-4) to between 100 pM and 2nM for surface immobilization in ZMWs (Buffer B)[60]. This resulted in ≈100 occupied ZMWs out of the total ≈ 1000 ZMWs per field of view identified by fluorescence bleach steps of DY-650 that labels each of the four CNBDs. Fluorescently labeled cAMP (fcAMP; 8-(2-DY-547]-aminoethylthio) adenosine-3',5'-cylic monophosphate) (BioLog, cat # D 109) was added at 1 µM for all single-molecule experiments in Buffer B.

ZMW chips were placed on top of an inverted microscope (Olympus IX-71, 100X, NA 1.49) and imaged under 532 nm (60 W/ cm^2) or 640 nm (25 W/ cm^2) (Coherent) as described previously[47,48]. The only notable difference is that unlike previous experiments we did not use FRET to monitor binding in order to obtain data for extended periods[48]. We excited DY-650 with 640 nm to identify ZMWs featuring DY-650-labeled tetrameric-CNBDs. Next, fcAMP was continuously imaged with 532 nm for 8000 frames at 10 Hz to monitor binding activity. All emission spectra were split with a 650 nm long pass dichroic (Semrock Brightline FF650) and bandpass filtered using pairs of edge filters (532-623.8 nm, 632.9-945 nm; Semrock Cy3/Cy5-A-OMF) and imaged onto two separate EMCCDs (Andor iXon Ultra X-9899) using Metamorph software (Molecular Devices). All data was collected using ZMWs of 150-200 nm diameter which are large enough to accommodate tetrameric CNBD complex (each monomeric CNBD is 4 x 6 x 20 nm)[48].

3.6.4 Single-molecule ligand binding image analysis

All analysis was performed using custom software written in MATLAB (Mathworks) or ImageJ. Single-molecule trajectories of each ZMW were extracted from tiff stacks saved by Metamorph software using MATLAB. Locations of ZMWs were obtained using a threshold mask of the brightfield image of the whole ZMW array. ZMW locations were refined with a 2D Gaussian fit to the local intensity height map. The time-dependent fluorescence at each ZMW was obtained by projecting the average image intensity in a 5-pixel diameter circle onto the ZMW location throughout each image in the stack.

3.6.5 Single-molecule ligand binding time series analysis

A total of 13,670 individual ZMWs (1.1×10^8 data points) were processed from the single-molecule tetrameric CNBD experiments (example traces in **Figure 3-S10**). Each trajectory was idealized with DISC using a 95% confidence interval for CP detection and BIC for state selection. To reflect the ability to cleanly resolve the individual occupation states, we computed the separation of each sequential state vs the noise within a state by:

$$State\ Separation = \frac{1}{K}\sum_{i=2}^{K}\frac{(\mu_i - \mu_{i-1})}{\sigma_{i-1}} \qquad \text{Eq. 3-18}$$

where K is the total number of states, μ is the mean intensity value of a state, and σ is the standard deviation of the data points belonging to a state. This ensures that states are separated well enough to resolve, as would be expected for sequential ligand binding. Traces featuring 4 to 6 identified states with state separation ≥ 3 were retained for further analysis.

To ensure a given trajectory contained a functional tetrameric CNBD, we kept traces that spent less than 50% of the time in the unbound state, resulting in a total of 480 trajectories for visual inspection. The observed asynchronous decay of protein activity was corrected using the CP detection method to identify the most likely point in a given trajectory where protein behavior dramatically changed. This was accomplished using MATLABs findchangepoint function using the change in standard deviation as the statistic. Data points following the identified CP location were discarded from the analysis to include only the frames of consistent fcAMP binding to presumably functional proteins (**Figure 3-S9**).

In total, 293 molecules totaling 1.2 x10^5 seconds (≈34.5 hours) of combined protein activity across 53,474 events was included for the final analysis. Each trajectory exhibited four or five conformational states (3 to 4 fcAMPs bound). Binomial fitting of the total time spent in each state was performed using MATLAB's mle function. HMM modeling of single-molecule binding events was performed with QuB[14,15]. Idealized trajectories from DISC were exported to QuB with the first and last events removed. A sequential model of 0 to 4 ligand binding sites was globally optimized to simultaneously describe the idealized binding trajectories for all molecules.

3.6.6 Code and data availability

DISC is written in MATLAB both with and without a user-friendly GUI. The DISC package and User Manual are available in a GitHub repository at https://github.com/ChandaLab/DISC. Simulated and raw data in addition to analysis scripts are available at https://zenodo.org/record/3727917.

3.7 References

1. Moerner, W.E., Shechtman, Y. & Wang, Q. Single-molecule spectroscopy and imaging over the decades. *Faraday Discussions* **184**, 9-36 (2015).
2. Hill, F.R., Monachino, E. & van Oijen, A.M. The more the merrier: high-throughput single-molecule techniques. *Biochemical Society Transactions* **45**, 759-769 (2017).
3. Grimm, J.B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods* **12**, 244-+ (2015).
4. Altman, R.B. et al. Cyanine fluorophore derivatives with enhanced photostability. *Nature Methods* **9**, 68-U178 (2012).
5. Levene, M.J. et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686 (2003).
6. Chen, J. et al. High-throughput platform for real-time monitoring of biological processes by multicolor single-molecule fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 664-669 (2014).
7. Juette, M.F. et al. Single-molecule imaging of non-equilibrium molecular ensembles on the millisecond timescale. *Nature Methods* **13**, 341-344 (2016).
8. Berghuis, B.A., Köber, M., van Laar, T. & Dekker, N.H. High-throughput, high-force probing of DNA-protein interactions with magnetic tweezers. *Methods* **105**, 90-98 (2016).
9. Ye, W. et al. Conformational Dynamics of a Single Protein Monitored for 24 h at Video Rate. *Nano Letters* **18**, 6633-6637 (2018).
10. Yang, D., Ward, A., Halvorsen, K. & Wong, W.P. Multiplexed single-molecule force spectroscopy using a centrifuge. *Nature Communications* **7**(2016).
11. Popa, I. et al. A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *Journal of the American Chemical Society* **138**, 10546-10553 (2016).
12. Young, G. & Kukura, P. Interferometric Scattering Microscopy. *Annual Review of Physical Chemistry, Vol 70* **70**, 301-322 (2019).
13. Hill, F.R., Monachino, E. & van Oijen, A.M. The more the merrier: high-throughput single-molecule techniques. *Biochemical Society Transactions* **45**, 759 (2017).
14. Nicolai, C. & Sachs, F. Solving ion channel kinetics with the QuB software. *Biophysical Reviews and Letters* **8**, 191-211 (2013).
15. Qin, F., Auerbach, A. & Sachs, F. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys J* **79**, 1915-27 (2000).
16. McKinney, S.A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J* **91**, 1941-51 (2006).

17. Greenfeld, M., Pavlichin, D.S., Mabuchi, H. & Herschlag, D. Single Molecule Analysis Research Tool (SMART): An Integrated Approach for Analyzing Single Molecule Data. *Plos One* **7**(2012).
18. Bronson, J.E., Fei, J., Hofman, J.M., Gonzalez, R.L., Jr. & Wiggins, C.H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophysical Journal* **97**, 3196-3205 (2009).
19. van de Meent, J.W., Bronson, J.E., Wiggins, C.H. & Gonzalez, R.L. Empirical Bayes methods enable advanced population-level analyses of single-molecule FRET experiments. *Biophys J* **106**, 1327-37 (2014).
20. Juang, B.-H. & Rabiner, L.R. The segmental K-means algorithm for estimating parameters of hidden Markov models. *IEEE Transactions on acoustics, speech, and signal Processing* **38**, 1639-1641 (1990).
21. Blanco, M. & Walter, N.G. Chapter 9 - Analysis of Complex Single-Molecule FRET Time Trajectories. in *Methods in Enzymology*, Vol. 472 (ed. Walter, N.G.) 153-178 (Academic Press, 2010).
22. Persson, F., Linden, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nature Methods* **10**, 265-269 (2013).
23. Hines, K.E., Bankston, J.R. & Aldrich, R.W. Analyzing Single-Molecule Time Series via Nonparametric Bayesian Inference. *Biophysical Journal* **108**, 540-556 (2015).
24. Sgouralis, I. & Presse, S. ICON: An Adaptation of Infinite HMMs for Time Traces with Drift. *Biophysical Journal* **112**, 2117-2126 (2017).
25. Sgouralis, I., Whitmore, M., Lapidus, L., Comstock, M.J. & Presse, S. Single molecule force spectroscopy at high data acquisition: A Bayesian nonparametric analysis. *Journal of Chemical Physics* **148**(2018).
26. Li, H. & Yang, H. Statistical Learning of Discrete States in Time Series. *Journal of Physical Chemistry B* **123**, 689-701 (2019).
27. Watkins, L.P. & Yang, H. Detection of intensity change points in time-resolved single-molecule measurements. *Journal of Physical Chemistry B* **109**, 617-628 (2005).
28. Shuang, B. et al. Fast Step Transition and State Identification (STaSI) for Discrete Single-Molecule Data Analysis. *Journal of Physical Chemistry Letters* **5**, 3157-3161 (2014).
29. Hadzic, M.C.A.S., Boerner, R., Koenig, S.L.B., Kowerko, D. & Sigel, R.K.O. Reliable State Identification and State Transition Detection in Fluorescence Intensity-Based Single-Molecule Forster Resonance Energy-Transfer Data. *Journal of Physical Chemistry B* **122**, 6134-6147 (2018).
30. Smith, C.S. et al. An automated Bayesian pipeline for rapid analysis of single-molecule binding data. *Nature Communications* **10**(2019).
31. Song, N. & Yang, H. Parallelization of Change Point Detection. *Journal of Physical Chemistry A* **121**, 5100-5109 (2017).
32. Pelleg, D. & Moore, A.W. X-means: Extending k-means with efficient estimation of the number of clusters. *Icml* **1**, 727-734 (2000).
33. Karypis, M.S.G., Kumar, V. & Steinbach, M. A comparison of document clustering techniques. *TextMining Workshop at KDD2000 (May 2000)* (2000).
34. Hamerly, G. & Elkan, C. Learning the k in k-means. in *Proceedings of the 16th International Conference on Neural Information Processing Systems* 281-288 (MIT Press, Whistler, British Columbia, Canada, 2003).
35. Juang, B.H. & Rabiner, L.R. THE SEGMENTAL K-MEANS ALGORITHM FOR ESTIMATING PARAMETERS OF HIDDEN MARKOV-MODELS. *Ieee Transactions on Acoustics Speech and Signal Processing* **38**, 1639-1641 (1990).
36. Qin, F. Restoration of single-channel currents using the segmental k-means method based on hidden Markov modeling. *Biophysical Journal* **86**, 1488-1501 (2004).

37. Viterbi, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory* **13**, 260-269 (1967).
38. Scott, A.J. & Knott, M.A. *A Cluster Analysis Method for Grouping Means in the Analysis of Variance*, (1974).
39. Elkan, C. Using the triangle inequality to accelerate k-means. *Proceedings of the 20th international conference on machine learning (ICML-03)*, 147-153 (2003).
40. Schwarz, G. Estimating the dimension of a model. *The annals of statistics* **6**, 461-464 (1978).
41. Bishop, C.M. *Pattern recognition and machine learning*, xx, 738 pages (Springer, New York, 2006).
42. Ward Jr, J.H. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* **58**, 236-244 (1963).
43. Watkins, L.P. & Yang, H. Detection of intensity change points in time-resolved single-molecule measurements. *J Phys Chem B* **109**, 617-28 (2005).
44. Li, H. & Yang, H. Statistical Learning of Discrete States in Time Series. *The Journal of Physical Chemistry B* **123**, 689-701 (2019).
45. Akaike, H. NEW LOOK AT STATISTICAL-MODEL IDENTIFICATION. *Ieee Transactions on Automatic Control* **AC19**, 716-723 (1974).
46. Hannan, E.J. & Quinn, B.G. The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)* **41**, 190-195 (1979).
47. Goldschen-Ohm, M.P., White, D.S., Klenchin, V.A., Chanda, B. & Goldsmith, R.H. Observing Single-Molecule Dynamics at Millimolar Concentrations. *Angewandte Chemie-International Edition* **56**, 2399-2402 (2017).
48. Goldschen-Ohm, M.P. et al. Structure and dynamics underlying elementary ligand binding events in human pacemaking channels. *Elife* **5**(2016).
49. Dempsey, G.T. et al. Photoswitching Mechanism of Cyanine Dyes. *Journal of the American Chemical Society* **131**, 18192-+ (2009).
50. Moerner, W.E. & Fromm, D.P. Methods of single-molecule fluorescence spectroscopy and microscopy. *Review of Scientific Instruments* **74**, 3597-3619 (2003).
51. Stennett, E.M.S., Ciuba, M.A., Lin, S. & Levitus, M. Demystifying PIFE: The Photophysics Behind the Protein-Induced Fluorescence Enhancement Phenomenon in Cy3. *Journal of Physical Chemistry Letters* **6**, 1819-1823 (2015).
52. Eid, J. et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**, 133-138 (2009).
53. English, B.P. et al. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nature Chemical Biology* **2**, 87-94 (2006).
54. Sambur, J.B. et al. Sub-particle reaction and photocurrent mapping to optimize catalyst-modified photoanodes. *Nature* **530**, 77-+ (2016).
55. Kusch, J. et al. How subunits cooperate in cAMP-induced activation of homotetrameric HCN2 channels. *Nat Chem Biol* **8**, 162-9 (2011).
56. Thon, S., Schulz, E., Kusch, J. & Benndorf, K. Conformational Flip of Nonactivated HCN2 Channel Subunits Evoked by Cyclic Nucleotides. *Biophys J* **109**, 2268-76 (2015).
57. Foquet, M. et al. Improved fabrication of zero-mode waveguides for single-molecule detection. *Journal of Applied Physics* **103**(2008).
58. Idikuda, V. et al. Singlet oxygen modification abolishes voltage-dependent inactivation of the sea urchin spHCN channel. *Journal of General Physiology* **150**, 1273-1286 (2018).
59. Eid, J. et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**, 133-138 (2009).
60. Aitken, C.E., Marshall, R.A. & Puglisi, J.D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical journal* **94**, 1826-1835 (2008).

3.8 Supplementary information

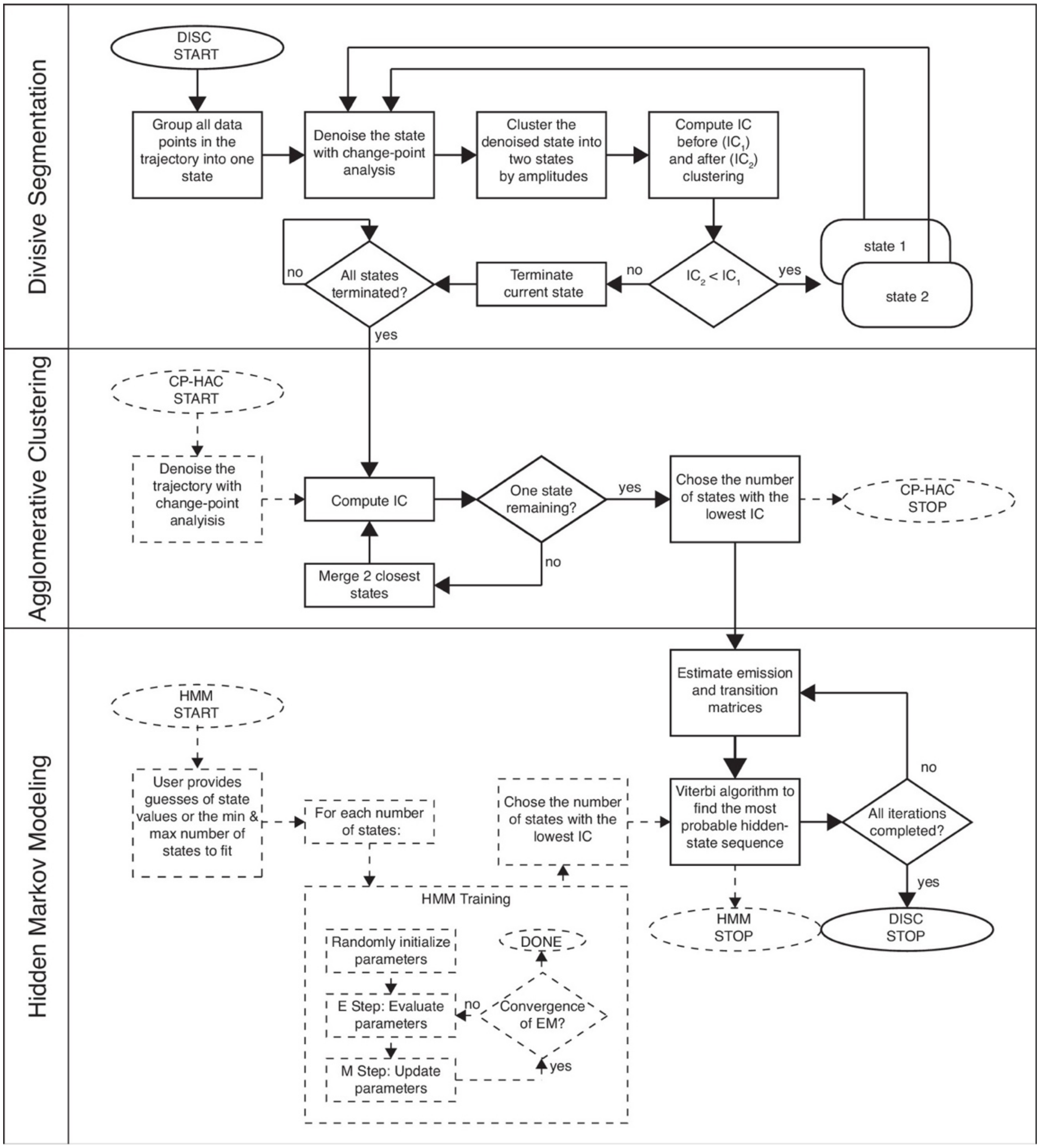

Figure 3-S1: Workflow of DISC. *Solid lines indicate the path DISC takes through divisive segmentation, hierarchical agglomerative clustering, and the Viterbi algorithm. IC = information criterion. General CP-HAC and HMM approaches are shown for comparison. Dotted lines indicate steps that do not overlap with DISC. Ovals indicate start/ stop; rectangles indicate a process or computation; diamonds represent decisions. Note, most current uses of HMMs involve a single user-defined number of states.*

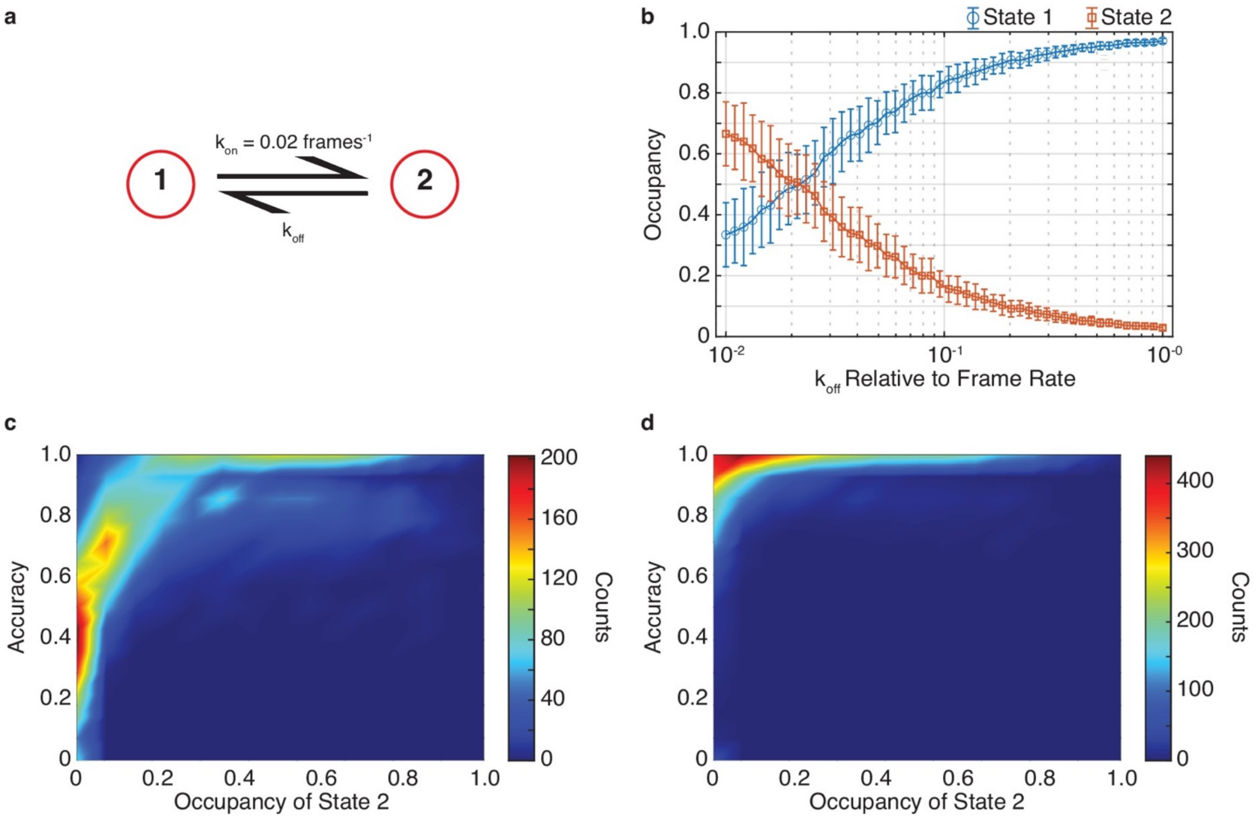

Figure 3-S2: The effect of state occupancy on DISC. *(**a**) Simulated kinetic scheme of **Figure 3-2** featuring two states with a constant k_{on} (k_{on} = 0.02 frames^{-1}) and varying k_{off}. (**b**) Occupancy of state 1 (blue circles) and state 2 (orange squares) for different simulated k_{off} values (mean ± s.d.). (**c, d**) Density plot of idealization accuracy per trajectory vs the occupancy of state 2 across all simulations (N = 5000) for results obtained without (c) and with Viterbi refinement (**d**).*

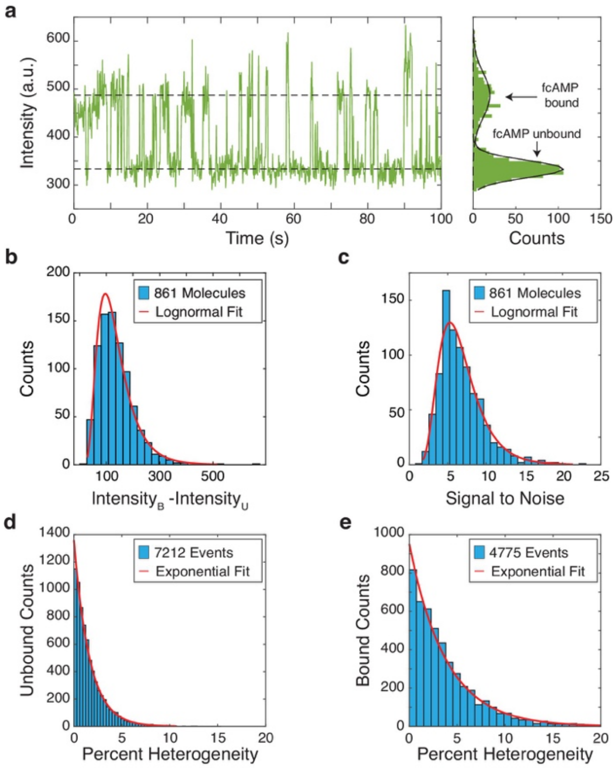

Figure 3-S3: Characterization of fcAMP binding to monomeric CNBDs in ZMWs. *(**a**) Representative trajectory of a fcAMP binding to a monomeric CNBD in a ZMW (**b**) Intensity difference of bound (B) and unbound (U) states per trajectory with a log-normal fit ($\mu = 4.79$, $\sigma = 0.47$, N = 861) (**c**) Signal-to-noise per trajectory with a log-normal fit ($\mu = 1.84$, $\sigma = 0.41$, N = 861) (**c**). Quantification of unbound (**d**) and bound (d) event heterogeneity overlaid with exponential fits (unbound $\mu = 1.62$, bound $\mu = 3.81$).*

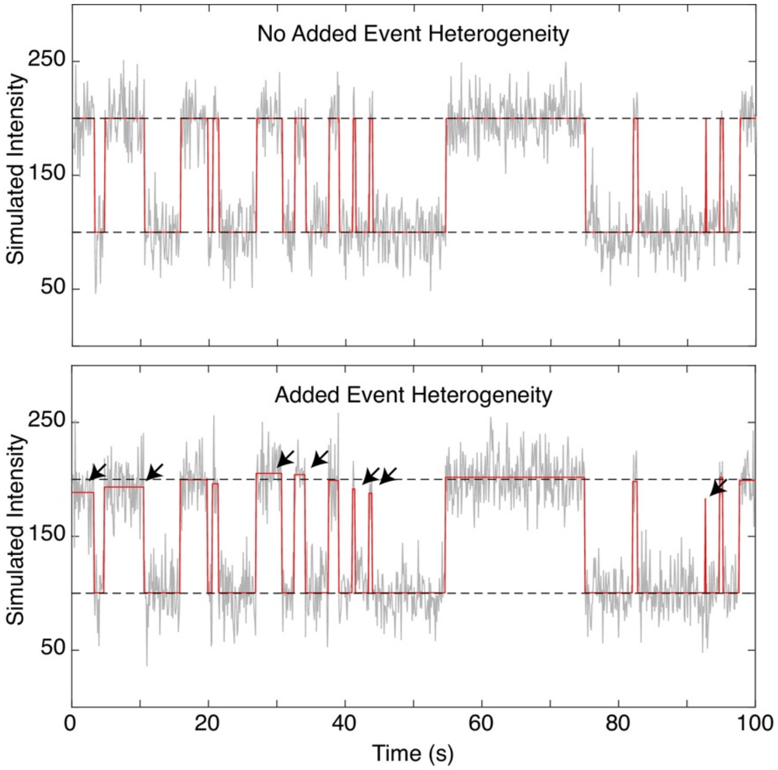

Figure 3-S4: Simulation with heterogenous fcAMP emission. *Representative single-molecule simulation of a two-state system (e.g. unbound and bound) both without (top) and without (bottom) heterogeneous emission of fcAMP upon binding (see* ***Figure 3-S3****, section* ***3.6.1****). Plots show simulated trajectory (red) overlaid with the addition of Gaussian noise (grey) and the average intensity value of each state (dashed black). Arrows indicate events where the heterogeneous intensities show prominent deviations away from the mean state value.*

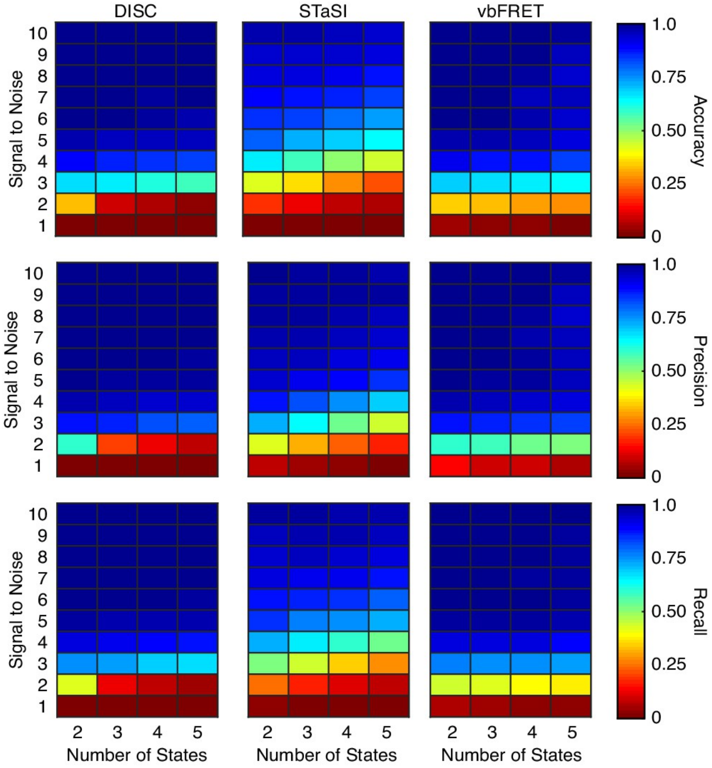

Figure 3-S5: Algorithm performance on simulations without heterogenous fcAMP emission. *Recreation of simulations from **Figure 3-3** of the main text without the inclusion of heterogenous intensities per binding event. Each value is the average accuracy (top), precision (middle), and recall (bottom) across 100 trajectories 200-second-long trajectories collected at 10 Hz at the indicated SNR and number of states.*

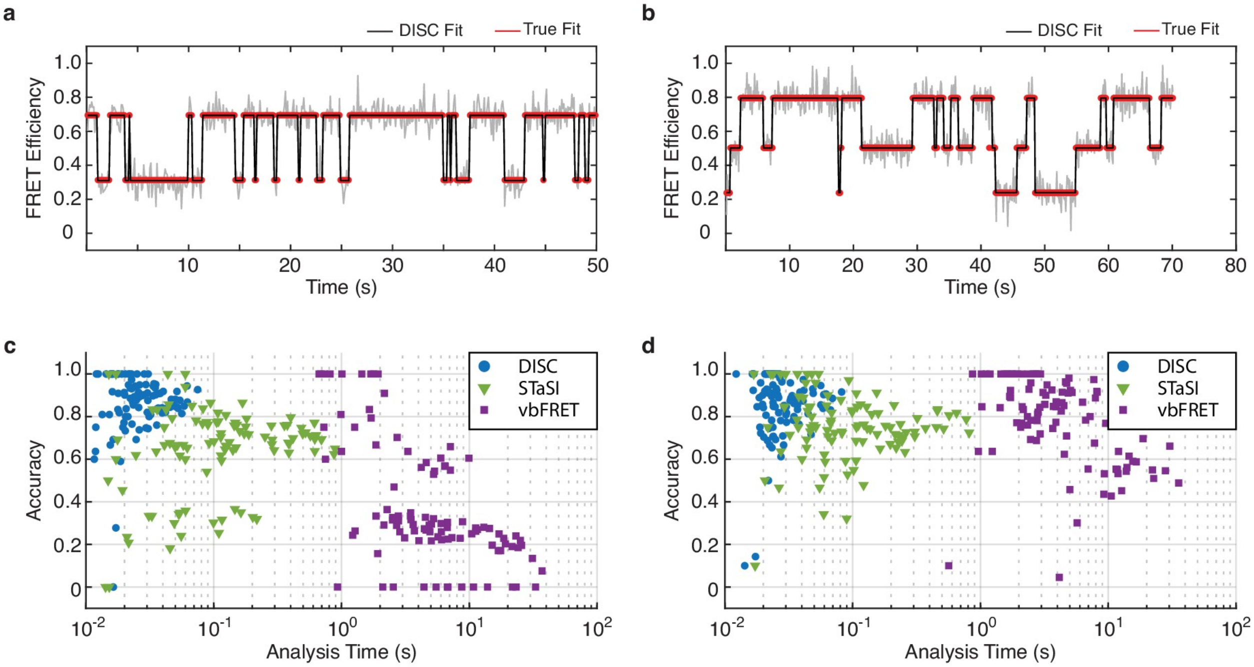

Figure 3-S6: Algorithm performance on simulated smFRET data. *(**a**, **b**) Representative fits of simulated smFRET data with 2 (**a**) or 3 (**b**) states (red) overlaid with the idealized fits from DISC (black). (**c**, **d**) Scatter plot of accuracy vs analysis time of 100 simulated smFRET trajectories featuring 2 (**c**) or 3 (**d**) states analyzed by DISC (blue circles), STaSI (green triangles) and vbFRET (purple squares).*

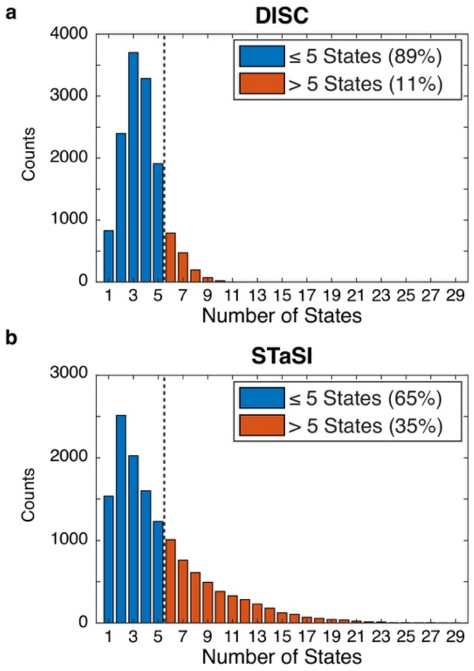

Figure 3-S7: Tetrameric CNBD analysis by DISC and STaSI. *Initial number of states found by DISC (**a**) and STaSI (**b**) when run on the tetrameric CNBD data set across 13,670 trajectories prior to trace selection. The expected distribution is between 1 and 5 states per trajectory to account for empty ZMWs and fully occupied tetrameric CNBDs (4 fcAMP bound states plus 1 unbound states). This analysis was not repeated using vbFRET as we estimated the process would take weeks to complete.*

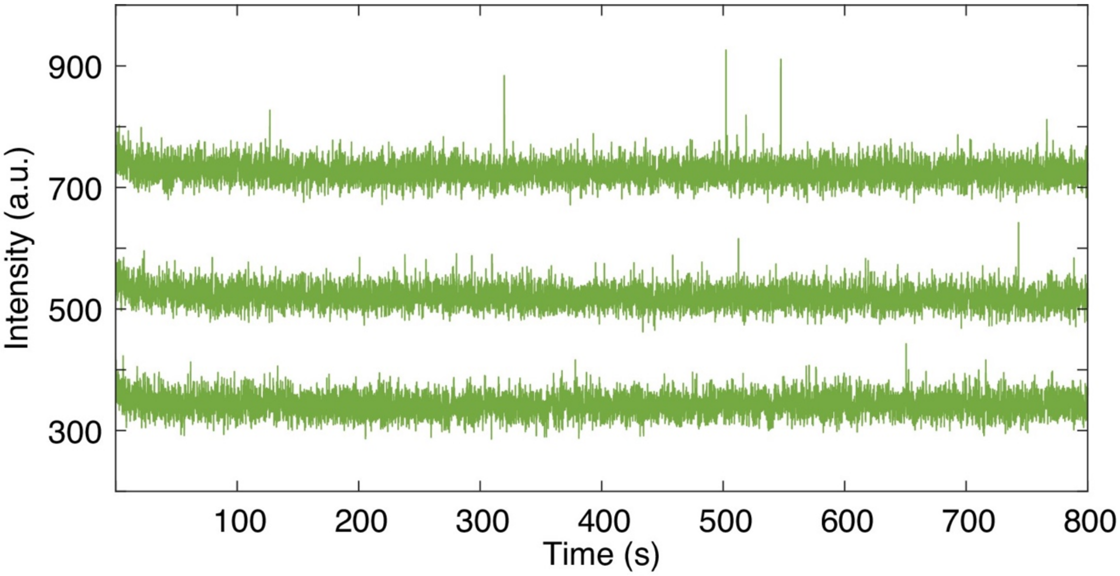

Figure 3-S8: Non-specific fcAMP binding in ZMWs. *Representative trajectories of unoccupied ZMWs in the presence of 1 µM fcAMP. Trajectories are offset by 200 arbitrary units (a.u.) for visualization.*

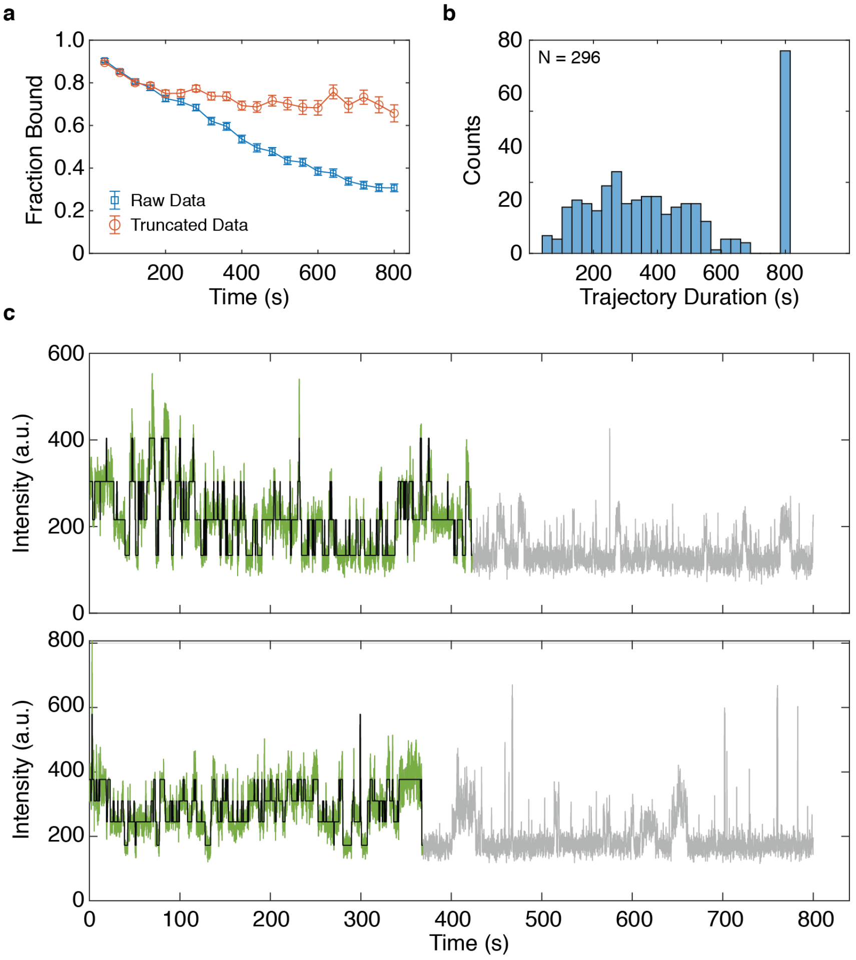

Figure 3-S9: Asynchronous decay of tetrameric CNBD activity over excitation time. *(**a**) Average fraction bound of each molecule over time (binned every 40 seconds) before and after trajectory truncation (mean ± s.e.m., see **3.6.5**). (**b**) Distribution of trajectory durations following truncation (N=296). Each trajectory was initially 800 seconds. (**c**) Representative trajectories showing the truncated data kept for analysis (green) with idealized fit by DISC (black) and the discarded frames (grey).*

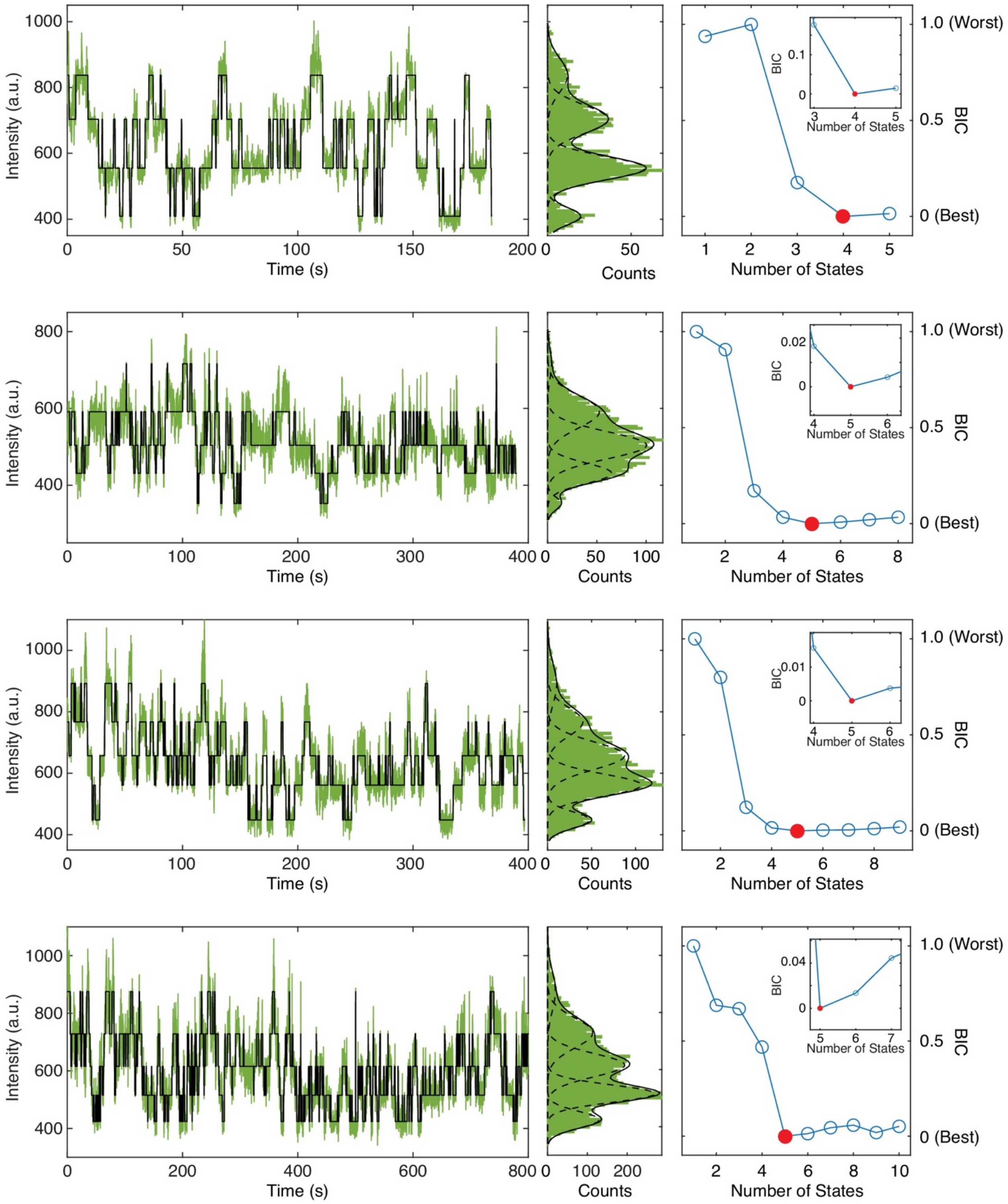

Figure 3-S10: Example trajectories of 1 µM fcAMP binding to tetrameric CNBDs in ZMWs. *Representative trajectories featuring up to 3 or 4 bound fcAMP molecules analyzed by DISC (left) with distribution fits (middle) and BIC curves for optimal state selection (right). Inset shows a zoomed portion of BIC curve to highlight the minimum BIC value identified. The lowest BIC value is indicated in red which corresponds to the final number of states fit to the trajectory. Each BIC plot may feature a different number of total possible states owning to the trajectory-by-trajectory results from divisive segmentation.*

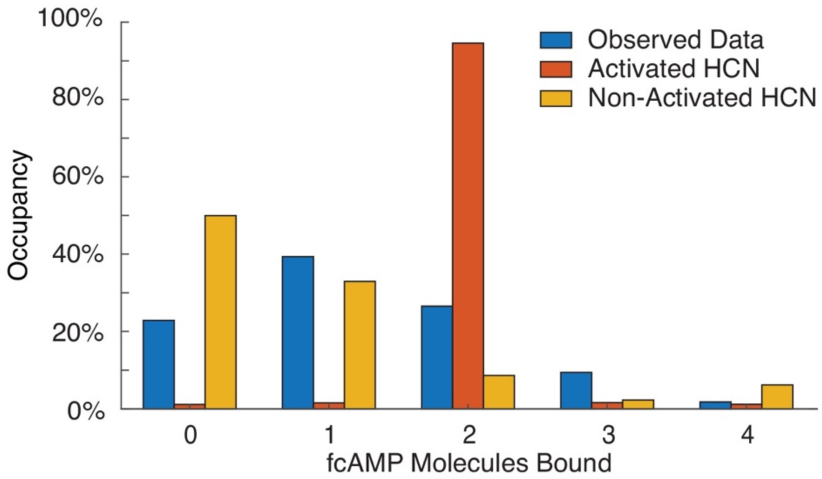

Figure 3-S11: Model comparison of HCN2 CNBDs: *Comparison of obtained state-occupancy distribution from single-molecule experiments with expected values from activated and non-activated HCN channel models.*

4 Single-Molecule Measurements in Photonic Waveguides Reveal Non-Cooperative Binding Dynamics in Intact Pacemaker Ion Channels

4.1 Abstract

Hyperpolarization-activated cyclic nucleotide-gated (HCN) channels are tetrameric ion channels critical for controlling rhythmic activity of specialized cardiomyocytes and neurons. Although primarily activated by hyperpolarization, some members of the HCN family are modulated by the binding of cyclic nucleotides to specialized domains in the C-terminus of each subunit. Electrophysiological and fluorescence studies suggest that this ligand binding is cooperative but the mechanism and nature of the cooperativity remains controversial. Here, in order to directly resolve the dynamics of ligand binding, we use nanophotonic arrays to monitor large number of individual binding events to intact HCN channel isoforms. Our studies, unexpectedly, reveal that the ligand binding in resting state of HCN1 and in HCN2 is non-cooperative. We also find that the differences upon cyclic nucleotide association to individual binding domains is dependent not only on associations rates, but also on a metastable conformational state following binding. These studies shed light on key differences in the physiological mechanism of ligand activation in different pacemaker channels.

This chapter is adapted from a manuscript in preparation

White, D.S.[1,2, †], Chowdhury, S.[1,5 †‡], Zhang, R.[1,2], Retterer, S.T.[3], Goldsmith, R.H.[2], Chanda, B.[1]

[1] Department of Neuroscience, University of Wisconsin-Madison, Madison, WI, USA
[2] Department of Chemistry, University of Wisconsin-Madison, Madison, WI, USA
[3] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[†] These authors contributed equally to this work

D.S.W, S.C., R.H.G, and B.C. designed the studies; S.C. performed molecular biology and protein purification; D.S.W. and R.Z. performed single-molecule experiments; D.S.W. analyzed single-molecule data. D.S.W. fabricated ZMWs under supervision of S.T.R; D.S.W., S.C., R.H.G., and B.C. wrote the manuscript with input from others.

4.2 Introduction

Hyperpolarization-activated cyclic nucleotide-gated (HCN) ion channels[1-3] belong to the voltage-gated ion channel (VGIC) superfamily of membrane proteins and share a common transmembrane architecture. However, two important facets make their gating unique. First, unlike other VGICs which activate upon membrane depolarization, HCN channels open upon membrane hyperpolarization[4,5]. Second, channel opening is acutely sensitive to the binding of cyclic nucleotides, in particular 3',5'-cyclic adenosine monophosphate (cAMP)[6]. These unique gating modalities enable HCN channels to function as pacemakers in the sinoatrial node and perform important roles in the central and peripheral nervous systems such as influencing neuronal plasticity and transmission of action potentials[7]. The cyclic nucleotide sensitivity of these channels governs the “flight-or-fight” behavior in response to ß-adrenergic stimulation and arises from the specialized cyclic nucleotide binding domain (CNBD), a common feature of the cyclic nucleotide-gated ion channel family and a host of other proteins[8,9]. In the HCN/CNG channel family, the C-terminal cytosolic CNBDs are connected to the transmembrane pore via a specialized structural domain called the C-linker which affects the ligand affinity and efficacy[9,10]. This region is thought to play an important role in transducing the conformational change in the CNBD to the channel pore upon ligand binding[6,11].

Despite decades of investigation, the understanding of the mechanism underlying cyclic nucleotide regulation of HCN channels remains limited. Many fluorescence and electrophysiological and studies have shed light on some of the structural lynchpins relaying the conformational change from the CNBD to the channel gate; however, these approaches inherently suffer from what is currently recognized as the “binding gating conundrum”, wherein mechanisms are predominantly inferred by indirectly measuring the “consequence of ligand binding” (i.e. channel opening) instead of directly measuring the ligand binding itself. To some extent, the use of a hybrid patch-clamp and confocal fluorometry technique which simultaneously monitors channel activity and the average residency of fluorescent derivatives of cyclic nucleotides (e.g. fcAMP)[12,13] overcomes this limitation. By building global Markov models of these data, multiple mechanisms of ligand binding to HCN2 channels have been proposed depending on the state of the channel. For instance, open channels displayed an unusual sequence of positive-negative-positive cooperativity[14] consistent with a dimer-of-dimers hypothesis[15]. On the other hand, closed channels showed a much less pronounced pattern of subunit interaction upon ligand binding and demonstrated a negative-no-positive cooperative pattern. Additionally, a concerted conformational "flip" that may be related to the suspected structural rearrangements underlying pore modulation was proposed[16-18].

Unfortunately, such macroscopic studies simply do not have the resolution to reliably resolve the complex multi-parametric kinetic schemes proposed[19]. In part, this parameter non-identifiability arises from the inability to directly resolve each individual binding site; thus, statistical analysis of ensemble averaged ligand occupancies can lead to many dissimilar but equally likely models of cooperativity, including similar schemes obtained from HCN2 as discussed above[19]. In order to reliably assess the interactions of subunits during ligand binding, a more direct approach is needed.

Single-molecule fluorescence microscopy is a powerful tool for resolving details of molecular compositions and dynamics that are otherwise obscured in ensemble measurements[20-23]. The most direct approach to probe the mechanisms of cooperativity is to monitor the individual ligand binding events so that each state visited by a single operational CNBD can be resolved. However, single-molecule binding measurements are challenging because the minimum excitation volume of focused visible light is on the order of a few femtoliters[24-26], resulting in an effective concentration barrier at and above 10 nM of fluorescently labeled ligand where background fluorescence overwhelms the signal from bound label. To circumvent this concentration barrier, we utilize nanophotonic arrays called zero-mode waveguides (ZMWs) which limit observation to sub-diffraction-limited volumes such that single-molecules can be resolved even at micromolar concentrations[30], and even at millimolar concentrations when applied with

FRET[27]. ZMWs have been used to explore a variety of physical processes, including real-time DNA sequencing[28], translation events at individual ribosomes[29,30], dynamics and stoichiometry of membrane-bound proteins[31,32], protein-protein interactions[33], and electrochemisty[34].

Here, we use ZMWs to directly monitor binding of individual fluorescently labeled cAMP to full length HCN1 and HCN2 channels and resolve the transient occupation of all intermediate bound states (fully unbound to fully bound). Our data provides strong evidence that the binding of cAMP to both HCN channels is non-cooperative when the channels are in the closed state. Our measured binding affinities of fcAMP to full length HCN2 channels is similar to the soluble CNBD domains of HCN2, suggesting that cAMP binding relieves the strain on the pore domain imposed by the apo CNBDs. Furthermore, dwell time analysis of ligand binding events highlights that upon cAMP binding, the CNBD isomerizes into a second "bound" state which is much longer lived in the HCN2 than in HCN1 channel. This difference, combined with faster ligand unbinding rates, underlies the dramatically different effects of cAMP on HCN2 vs HCN1.

4.3 Results

4.3.1 Purification of HCN channels for single-molecule experiments

Full-length HCN1 and HCN2 channels were engineered to improve biochemical behavior and purification (see **4.5.1**). In particular, the N-terminus of HCN2 was replaced with the N-terminus of HCN1 to overcome the challenge of the high G/C rich region for expression and purification. Importantly, modifying the N-terminus of HCN2 has shown little impact on cAMP gating[6]. The resultant constructs, hereafter referred to as HCN1SM and HCN2SM, were expressed in suspension cultures of HEK cells and purified in detergent micelles using a combination of affinity and size exclusion chromatography (**Figure 4-S1**). Both isoforms feature an N-terminal eGFP on each monomer for purification and single-molecule localization.

We performed single-molecule photobleaching analysis to confirm the tetrameric nature of isolated complexes attached to surfaces[35]. The N-terminally eGFP tagged HCN proteins were pulled down onto streptavidin coated surfaces using a biotinylated GFP nanobody (GFP-TRAP) and imaged with total internal reflection fluorescence microscopy (TIRFM)[36]. Both HCN1SM and HCN2SM exhibited distributions of photobleaching steps consistent with tetrameric complexes (**Figure 4-S2**). This result suggests that the single-molecule imaging conditions are not perturbing channel assembly or resulting in channel aggregation.

4.3.2 Single-molecule ligand binding inside ZMWs

To assess the mechanism of cAMP modulation of HCN channels, we recorded the binding of fluorescently-tagged cAMP (fcAMP) to purified eGFP tagged HCN channels tethered inside ZMWs (**Figure 4-1a, 4-1b)**. Upon illumination, the resulting exponentially decaying evanescent field inside every ZMW allows resolution of individual bound fluorescent species up to low µM concentrations[37,38]. Arrays of ZMWs with a 150 nm dimeter were fabricated on cover glasses with a 100 nm aluminum layer using positive tone electron-beam lithography and reactive ion etching (see **4.5.2**)[37]. Each ZMW was separated by a 2 µm pitch, allowing up to 1,600 ZMWs to be imaged in our field of view (**Figure 4-S3**). The glass and aluminum layers were passivated with polyethylene glycol (PEG) and poly(vinylphosphonic acid) (PVPA) to prevent non-specific adsorption of both ligand and proteins(**Figure 4-S4**, see **4.5.3**)[39]. Each ZMW was further coated with a streptavidin layer for single-molecule pull-downs of purified GFP-tagged HCN1SM and HCN2SM using a biotinylated GFP nanobody.

eGFP-tagged HCN1SM or HCN2SM molecules were sparsely deposited in ZMWs and bathed in various concentrations of fcAMP (**Figure 4-1b, 4-1c, Figure 4-S5, Figure 4-S6**). A co-localization paradigm was used to identify ZMWs occupied by functional GFP-tagged (see **4.5.4**). First, the array was excited with a 488 nm pump to identify ZMWs containing at least one HCN molecule by exciting and ultimately

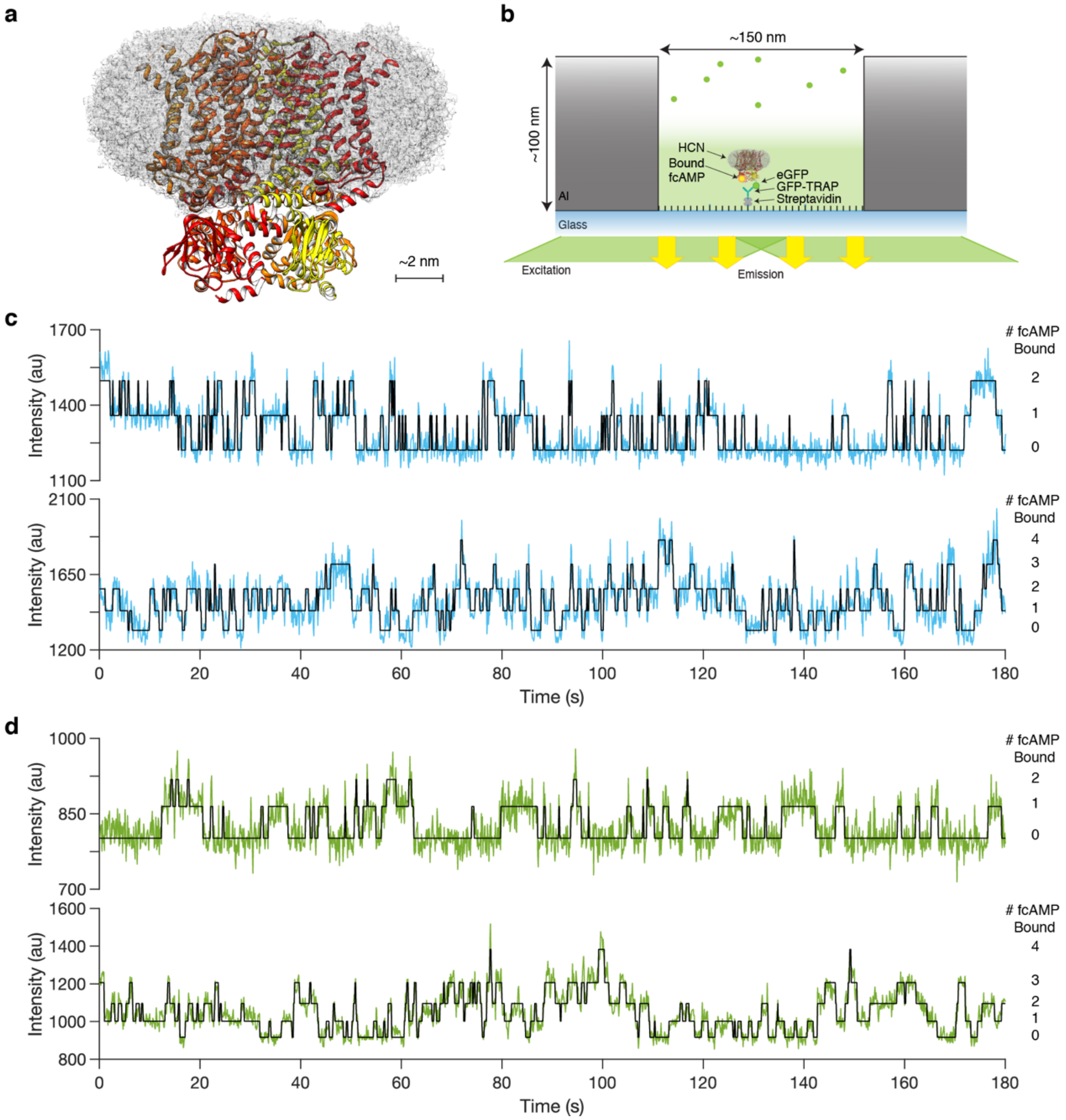

Figure 4-1: Single-molecule ligand binding to HCN1SM and HCN2SM in ZMWs. *(**a**) Structure of HCN1 in detergent micelle (PDB: 6UQG). Scale bar is ~ 2 nm. (**b**) Cartoon of a ZMW with an HCN molecule deposited in the center for fcAMP binding experiments. (**c**, **d**) Representative fluorescence intensity vs time trajectories of fcAMP molecules binding to purified HCN1SM (**c**, blue) and HCN2SM (**d**, green) at 250 nM (top) and 750 nM (bottom). Trajectories are overlaid with idealized fit (black). Only the first 180 seconds of the total 300 seconds observation time are shown for visualization. Additional representative trajectories are provided in **Figure 4-S5, 4-S6**.*

photobleaching the eGFP tags. Next, diffusing fcAMP was continuously excited at 532 nm for at least 300 seconds to monitor binding activity. As this approach is not limited by photobleaching like popular single-molecule Förster resonance energy transfer (smFRET) paradigms because of an effectively limitless fcAMP reservoir, we were able to quickly build a large dataset of binding dynamics across hundreds of single molecules. Images were collected with an exposure time of 100 ms, which is much slower than the expected diffusion of single unbound fcAMP molecules through the ZMW observation volume[40]. Therefore, fluorescence trajectories from our ZMWs report on ligands remaining in the excitation volume for an extended period of time (i.e. binding), but not free-ligand diffusion in the excitation volume. The passivation of the glass and aluminum surfaces ensures non-specific absorption is minimized[34,39].

Discrete changes in fluorescence intensity are indicative of fcAMP molecules binding and unbinding. Unlike intramolecular smFRET, our approach does not return a correlated value across all molecules. Hence, the discrete intensity levels corresponding to numbers of occupied ligands (U, B_1, B_2, B_3, and B_4) needs to be individually assessed for each molecule. The fluorescence intensity values obtained from each ZMW depend on many factors, including the location of the ZWM within the imaging plane, heterogeneity between the ZMWs, and the position of each HCN molecule within a ZMW. To resolve the transitions between liganded sates, we used our recently developed algorithm, Divisive Segmentation and Clustering (DISC), which adapts unsupervised statistical learning to idealize single-molecule trajectories with exceptional speed and accuracy (see **4.5.5**)[41]. In total, our analysis included 2.17×10^5 seconds (60 hours) of HCN1 activity across 739 molecules (1.8×10^5 events) at fcAMP concentrations between 0.1 to 0.9 µM, and 1.26×10^5 seconds (35 hours) of HCN2 activity across 444 molecules (8.2×10^4 events) at fcAMP concentrations between 0.1 to 1.5 µM (**Table 4-S1**). These results are an amalgam of multiple biological replicates from different protein preparations (HCN1SM: N=2, HCN2SM: N=3) and collected across multiple ZMW chips (HCN1SM: N=4, HCN2SM: N=5) each showing consistent binding behavior (**Figure 4-S7**).

4.3.3 Non-cooperative subunit interaction of pacemaker channels upon cAMP binding

Single monomeric CNBDs can exist in one of two discrete functional states: unbound or bound. Therefore, the probability of being in the bound state can be assessed with Bernoulli distribution governed by a binding success rate p. Extending this framework to a complex featuring multiple identical and independent CNBDs allows us to apply the binomial theorem, wherein the probability of k ligands bound to n total binding sites should follow a probability density function with a binding success rate p for each binding site given by

$$p(k \mid n, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{where } k = 1, 2, \dots \qquad \text{Eq. 4-1}$$

As idealized single-molecule trajectories provide a direct measure of the total time spent in each of the liganded states, we can assess whether or not ligand binding exhibits cooperativity between CNBDs for HCN1SM and HCN2SM using the binomial theorem.

For each concentration of fcAMP, the total time spent in each ligand bound state was treated as a binomial distribution and fit using maximum likelihood (**Figure 4-2a, 4-2b, Table 4-S2**). Strikingly, the estimations of each distribution for HCN1SM and HCN2SM are clearly well-modeled by the expected distribution for a non-cooperative and independent binding mechanism. Across all individual molecules, a binomial distribution could account for 94% of the observed state occupancy distribution of HCN1 and 93% for HCN2, as measured by root mean squared error. To account for possible heterogenous behavior between proteins at a given fcAMP concentration, a binomial test was performed on each trajectory which showed only minor deviations away from values expected from a binomial distribution (**Figure 4-S8**)[42], Together, these analyses suggest a binomial distribution is an adequate model of our observed state occupancy distributions, suggesting the possibility of non-cooperative interactions.

A limitation of the binomial test to determine cooperativity is that it ignores kinetic information[42,43]. Therefore, we also examined the transition rates to assess the degree of cooperativity. If the binding mechanism is indeed non-cooperative, one should expect a linear correlation of sequential

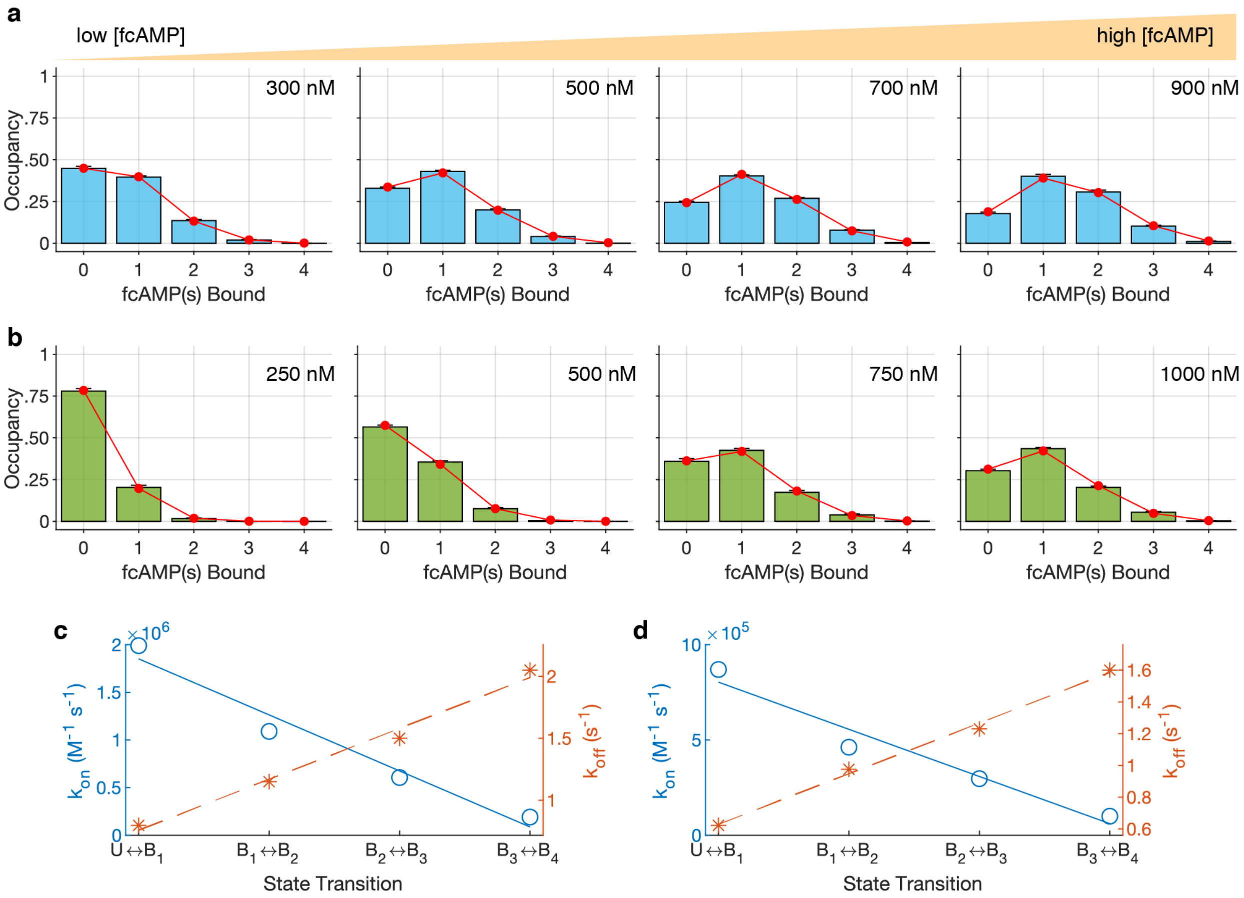

Figure 4-2: The binding of fcAMP to HCN1SM and HCN2SM is non-cooperative. *Total occupancy of each liganded state across all molecules of HCN1SM (**a**) and HCN2SM (**b**) at increasing fcAMP concentrations (mean ± s.e.m) overlaid with maximum likelihood estimations of a binomial distribution (red). Optimized state transition rate constants* k_{on} *(circle) and* k_{off} *(asterisk) for HCN1SM (**c**) and HCN2SM (**d**) overlaid with linear fits (*k_{on} *solid,* k_{off} *dashed). Linear correlations of sequential rate constants were evaluated with the Pearson correlation coefficient (r) for HCN1SM* k_{on}*: r = 0.98,* k_{off}*: r = 1 and for HCN2SM* k_{on}*: r = 0.98,* k_{off}*: r = 1.*

states transitions in time[42]. This is because each binding domain would act independently, thus the sequential binding of four ligands should be governed by a uniform transitions rates (e.g. k_{on} and k_{off}) modified by the occupancy of the complex. Here, the individual associations and dissociation transition rates between each successive ligand occupied states of idealized trajectories were obtained using a Markov chain with all transitions rates left as free parameters[44,45]. Consistent with the non-cooperative hypothesis, we find the globally optimized transition rates display a strong linear relationship between successive steps (**Figure 4-2c, 4-2d**).

Taken together, both observed state occupancies and transitions rates provide strong support for our hypothesis that under our measurement conditions, fcAMP binds to the different subunits of full-length tetrameric channels, HCN1SM and HCN2SM non-cooperatively. This result is consistent with our previous studies of tetrameric HCN2 CNBDs[41], reinforcing the notion that CNBDs are not intrinsically cooperative. This result further demonstrates that CNBD cooperativity is not induced by the presence of transmembrane and pore regions when HCN channels are closed.

4.3.4 HCN1 and HCN2 differ at the CNBD level

Although closed HCN1SM and HCN2SM exhibit non-cooperative binding, there are many possible kinetic mechanisms at the CNBD level that can give rise to this observation, so long as each CNBD is equal and independent. Previously, we performed single-molecule studies on monomeric and isolated CNBDs from HCN2. In the presence of both fluorescently-conjugated cAMP and cGMP (fcAMP and fcGMP), we resolved a second reversible conformational state following ligand binding[27,34], reminiscent of a catch-and-hold mechanism of ligand-gated ion channels[46,47]. We and others hypothesized that the conformationally “flipped” state plays an important role in allosterically modulating the pore through the C-linker[17,34,48]. Similarity, a concerted conformational state was proposed through patch-clamp fluorometry[16]. We therefore aimed to determine if CNBDs of HCN1 and HCN2 from full-length and intact channels also exhibit this complex binding behavior.

Here, our single-molecule fluorescence time trajectories of full-length HCN channels show the cumulative binding activity across four CNBDs. In this paradigm, the binding and unbinding events report the liganded state of a single channel but do not provide any information as to which of the four sites is bound or unbound. However, since our previous analysis strongly supports that each CNBD is equal and independent, we can perform dwell time analysis on the singly-liganded events (B_1) to directly examine individual CNBD behavior.

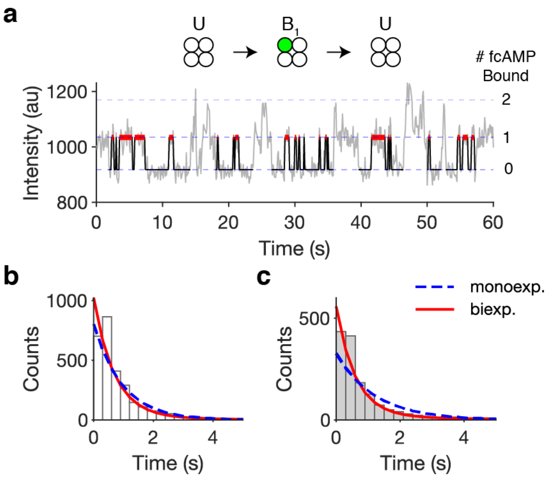

Figure 4-3: CNBD dynamics reveal a reversible conformational exchange upon binding. *(**a**) Example trajectory of fcAMP binding to HCN2SM with isolation of U→B_1→U events (black) to extract singly bound CNBD dwell times (red). Dwell times distribution of isolated-B_1 at 250 nM fcAMP for HCN1SM (**b**) and HCN2SM (**c**) overlaid with maximum likelihood estimates of monoexponential (blue dashed) and biexponential (red) distributions.*

To glean CNBD behavior, we isolated the state transitions of the singly-liganded state that were immediately preceded and followed by the unbound state (U→B_1→U) from idealized trajectories. These dwell times provided the most accurate representation of single CNBDs in our data as they minimize signal truncation from additional ligands binding (**Figure 4-3a**). Maximum likelihood fits of isolated-B_1 dwell time distributions required two exponential components for both HCN1SM and HCN2SM (**Figure 4-3b, 4-3c**), a signature suggesting the presence of a second metastable conformation (i.e. “flipped” state, denoted here as B_1^*). Across all fcAMP concentrations of HCN2, the probability that any isolated-B_1 time distribution was monoexponential as opposed to biexponential was less than 0.001 as determined by twice the difference in their loglikelihood values from maximum likelihood estimation (**Table 4-S3**). The same trend is true for HCN1SM; however, as the average bound dwell time noticeably decreases with increasing fcAMP concentration, the biexponential fits eventually regress to single exponentials (**Figure 4-S9**). This effect arises from a truncation of the true dwell time in isolated-B_1 due to the increased probability of additional fcAMP molecules binding at higher fcAMP concentrations. The kinetic regression is also more apparent for HCN1SM than HCN2SM owing to a smaller ratio of time spent in the isolated-B_1^* than isoalated-B_1 for HCN1SM than HCN2SM.

Overall, these results support the existence of a conformational change of the CNBD following ligand binding in full-length HCN1 and HCN2 channels, despite a lack of observed cooperative binding. This finding agrees with previous structural and functional studies suggesting CNBD flexibility but is in direct contrast with the hypothesis that these conformations underlie pore modulation[49]. Further, this result suggests that the conformational flip occurs at the level of the CNBD and is not concerted across all domains as previously suspected[16].

4.3.5 Kinetics of non-cooperative binding

Finally, we aim to fully describe the binding dynamics of each HCN isoform, including how the conformational change described above affects the overall binding behavior. We built and optimized two hidden Markov models to determine if the conformational flip of single CNBDs can be captured across the whole channel. Model 1 is our null hypothesis and features four equal and identical CNBDs without additional conformational states (**Figure 4-4a**). Model 2 is our alternative hypothesis and features sequential binding with a conformational flip that can occur at any CNBD with no effect on cooperativity (**Figure 4-4a**)[50]. While other models of both open and closed HCN2 channels have been explored using confocal patch clamp fluorometry[14,16], they exhibit cooperative subunit interactions which are incompatible with our non-cooperative observations and were therefore not explored[41].

Both models were first validated by simulating transition rates of monomeric CNBDs from our previous study as a tetrameric complex[34]. We simulated two conditions: (1) CNBD without a conformational flip ($U \rightleftharpoons B$) and (2) with a conformational flip following binding ($U \rightleftharpoons B \rightleftharpoons B^*$). Each simulated CNBD was equal and independent, yielding a non-cooperative tetrameric complex. Both simulations were each fit with Models 1 and 2 to ensure reliability of estimated parameters within our full-length data (**Figure 4-S10, 4-S11, Table 4-S4**).

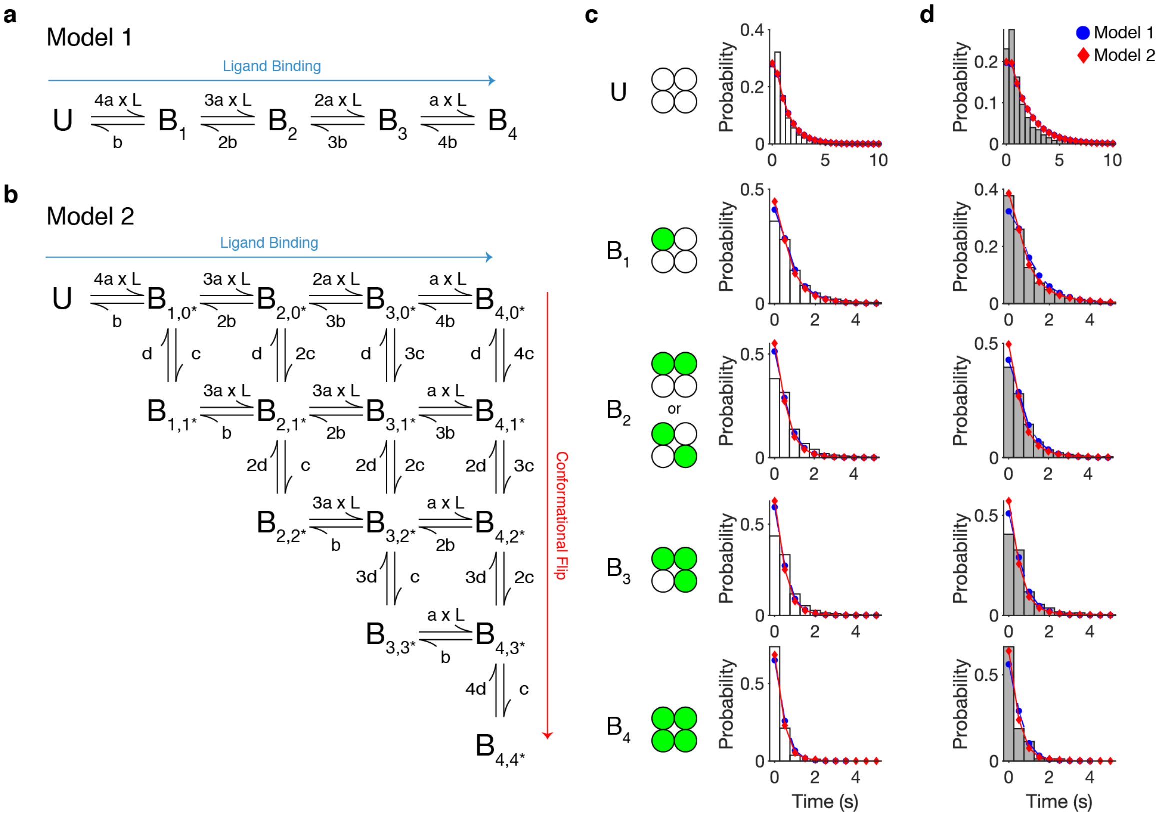

Figure 4-4: Kinetic models of ligand binding to HCN channels. *(**a**) Model 1 of non-cooperative binding without a second conformational bound state. (**b**) Model 2 of non-cooperative binding featuring a second conformational bound state. Each bound state B_{i/j^*} indicates the total number of ligands bound (i) and the number of ligands in a "flipped" state (j^*). Dwell time probability distributions of HCN1SM at 500 nM fcAMP (**c**) and HCN2SM at 750 nM fcAMP (**d**) overlaid with expected dwell time probability distributions of optimized rates for Model 1 (blue circle) and Model 2 (red diamond). At these fcAMP concentrations, HCN1SM and HCN2SM have a similar bound probability (see **Figure 4-S12**).*

Rates of HCN1SM and HCN2SM were globally optimized across all idealized time series from all molecules and fcAMP concentrations. Each model was then ranked by their Bayesian information criterion (BIC) to optimize the trade-off between goodness of fit and model complexity[51]. Consistent with the dwell times estimations of previous monomeric CNBDs[27,34] and the isolated-B_1 dwell times above, Model 2 featuring a conformational flip at each CNBD was deemed more probable for both HCN1SM and HCN2SM (**Table 4-1**). To ensure this result is not arising due to a subpopulation of molecules, we further performed a resampling procedure which showed homogeneity of rates and model selection (see **4.5.6**, **Table 4-S5**). Based on this result, we can confidently reject the null hypothesis of purely sequential binding in favor of

Table 4-1: Optimized rate constants for kinetic models shown in Figure 4-4

Protein	Model	Rates (M^{-1} s^{-1} or s^{-1})				K_d (M^{-1})	ΔBIC
		a	b	c	D		
HCN1	1	4.1×10^5	0.66			1.6×10^{-6}	10,637
	2	4.2×10^5	0.92	0.09	0.23	1.5×10^{-6}	0
HCN2	1	1.8×10^5	0.54			3.0×10^{-6}	11,111
	2	1.9×10^5	0.96	0.15	0.20	2.7×10^{-6}	0

a more complex binding scheme that at least features a reversible conformational flip of ligand occupied CNBDs.

The optimized rates from Model 2 provide key insights into the mechanistic differences between HCN isoforms (**Table 4-1**). Our data shows that the affinity of fcAMP to HCN1 channels ($K_d = 1.5 \times 10^{-6}$ M^{-1}) is 1.8 times higher than to HCN2 channels ($K_d = 2.7 \times 10^{-6}$ M^{-1}) (**Figure 4-S12**). Further, simulations using Model 2 for both HCN1SM and HCN2SM were able to successfully return the observed dwell times distribution across all liganded states suggesting this model can sufficiently account for our observed experimental data (**Figure 4-4c, 4-4d Figure 4-S13, 4-S14**). Between isoforms, the ligand dissociation step converges to similar rates, whereas HCN1SM exhibits a 2.3 times faster binding rate than HCN2SM. A primary difference in the functional dynamics of HCN1 and HCN2 lies in the conformational flip. While the rates out of the bound flipped state are similar (and therefore dwell-times are nearly identical), HCN2 CNBDs enters the flipped state almost twice as fast as HCN1. Therefore, for any bound event of HCN1 CNBDs, only 28% of the time is on average spent in the conformational flipped state, compared to the 43% of time for HCN2 due to its faster entry rate. These results suggest that the difference in cyclic nucleotide modulation of HCN channels is not dependent on ligand association, but may arise from duration of time the CNBD spends a second metastable conformation.

4.4 Discussion

In this study, we performed single-molecule fluorescence microscopy in ZMWs to directly observe a series of elementary ligand binding events to full length HCN channels. Our experimental strategies have enabled us to collect and analyze orders of magnitude more data than comparable single-molecule fluorescence approaches and thereby increases the robustness of our conclusions. Our first, central conclusion, is that binding of fluorescently conjugated cAMP (and thus cAMP) to the tetrameric HCN1 and HCN2 channels is non-cooperative: the binding of cAMP to one subunit does not influence its binding to a neighboring subunit. This conclusion might seem initially seem counter intuitive based on our current understanding of cAMP regulation of HCN channels[14-16,52]; however, in the context of classical MWC-type allosteric models, binding of a ligand to one subunit only influences the other subunits via a quaternary conformational change in the protein. The latter in ion channels, particularly ligand-gated ion channels is usually ascribed to the change in the conformation of the pore[46,53]. Under our conditions, the pore of both HCN1 and HCN2 channels are likely to be in the closed conformation since there is no applied electrochemical gradient. Furthermore, it is known that cAMP binding alone is insufficient to open HCN channels; therefore, over the different ligand concentrations tested here, the channel pore is likely to remain closed. Thus, our single-molecule measurements posit that in both the HCN1 channel, which is only modestly regulated by cAMP, and the HCN2 channel, which is strongly regulated by cAMP, there is no energetic linkage between the CNBDs of the different subunits (or other accessory structural domains of the channel, such as the C-linker domain) when the channel is in the closed state.

We have previously used our single-molecule approach to study elementary ligand binding events to the isolated, soluble, and monomeric CNBD of the HCN2 channel[27,34] and engineered tetramer of the

CNBDs of the HCN2 channel[34,41]. While the single-molecule conditions, reporting scheme, and ZMW tethering strategy differed between all paradigms, we see that our measured fcAMP affinity of our full-length HCN2 channel is within experimental error to the values for the soluble domains (monomeric CNBDs: $K_d = 1.5 \times 10^{-6}$ M[34], tetrameric CNBDs: $K_d = 2.1 \times 10^{-6}$ M[41]). Although our extracted rate constants differ between the soluble forms and the full-length channel, the similar affinities have important implications on our understanding of the mechanism of cyclic nucleotide regulation of the HCN2 channel. Current models of cAMP gating of HCN2 propose that the apo CNBD holds the channel in a closed state and cAMP binding relieves this inhibition and thereby facilitates channel opening[6,54]. In this model, the affinities of the CNBDs for cAMP would be expected to be different between the full-length channel, with the pore closed, and the soluble monomeric CNBDs, which is contrary to our findings. Instead, our data and analyses presented in this study, along with those presented earlier, suggest that CNBDs do not interact with the closed pore of the channel and cAMP modulation is entirely achieved via the interaction with the open pore. Taking into consideration previous electrophysiological studies and our single-molecule data, we propose that the open HCN2 pore is strained by the apo CNBDs and binding of cyclic nucleotide to the CNBDs causes a conformational change which relieves this strain and thereby enhances the stability of the open state.

Comparison of the unitary fcAMP binding events to HCN1SM and HCN2SM channels reveals two important differences between them. First, binding of fcAMP to HCN1 is nearly twice as tight as binding to HCN2 due to a 2.3 times faster binding rate and similar off rates. Second, after the formation of an “initial encounter complex” with cAMP, CNBDs of both isoforms can enter a “flipped” state. While entry into the HCN1 flipped state is relatively short slow, CNBDs of HCN2 enter the second metastable flipped state almost twice as fast as HCN1. For a given bound event, HCN1 CNBDs are 2.5 times more likely to be observed in the initial encounter state than the flipped state, compared to the nearly even likelihood of observation of each conformation for HCN2 CNBDs. The high probability of entering the “flipped” state in

HCN2 might indicate the relatively slower entry of the channel via a conformational change. In HCN1, the faster unbinding rate of cAMP would compete with this transition and thus mitigate the effect of cAMP on gating.

At this point, it is hard to predict the nature and location of such a conformational change. However, since the occupancy of this "flipped" state is also reflected in the single-molecule binding events of the isolated CNBDs, we suggest that it is CNBD specific; however, interactions with the transmembrane domains of the channel, particularly in the open state, may influence the energetics of this transition. Our current views of the structural changes in HCN channels arising from cAMP binding comes from the X-ray crystallographic and NMR structures of the soluble domains[34,52,55] and the cryo-EM structures of full length HCN1 channel[10]. As these approaches are all bulk averaged, they might obscure short lived, functionally relevant structural states of the protein. Moving forward, intramolecular smFRET in addition to cryo-EM structures of full length HCN channels, particularly those which are robustly modulated by cAMP, will be crucial to understand the structural events associated with cAMP induced "flip"-ping[23]. In addition, single-molecule binding studies of HCN channels with conformationally restricted transmembrane domains (such as an open pore or activated voltage-sensor) will be necessary to deconvolve the complexities associated with binding and regulation of HCN channels by cyclic nucleotides.

4.5 Methods

4.5.1 Protein expression and purification

The HCN1 construct used in this study (hereafter referred to as HCN1SM) was the human ortholog of HCN1 featuring a deletion of amino acid residues 636-865 on the C-terminus and on the N-terminus the construct was tagged with eGFP and a Twin strep affinity purification tag. The construct design was based off the recent cryo-EM study of HCN1[10]. cDNA corresponding to HCN1SM, in the modified pEG

BacMam vector[56], extracted from large volumes of bacterial cultures using Endotoxin free Plasmid Purification kits (Qiagen), was transfected into suspension cultures of Freestyle HEK293 cells (Thermo Fisher Scientific) using Trans-IT Pro Transfection reagent (MIRUS) following manufacturer's instructions. Post transfection, cells were grown at 12-14 hours at 37°C, following which sodium butyrate was added to the cultures to a final concentration of 10mM and cultures were grown at 30°C for another 48 hours. Cells were harvested by centrifugation at 3000xg for 20 minutes and washed twice with chilled 150 mM NaCl, 20mM Tris, pH = 8.0. Cell pellets were finally resuspended in lysis buffer (300 mM NaCl, 40 mM Tris, 10 mM DTT, 20% glycerol, 1 mM EDTA, 1% L-MNG, 2mM CHS (Cholesterol Hemi Succinate), pH = 8.0 supplemented with 1x Halt protease inhibitor cocktail (Thermo Fisher Scientific), briefly sonicated on ice and incubator at 4°C with gentle agitation for ~2 hours. The detergent extract was next spun at ~100,000xg for 1.5 hours and the supernatant was purified using Streptactin affinity resin (IBA Life Sciences). Protein bound resin was washed with 10 bed volumes of wash buffer (300 mM NaCl, 20 mM Tris, 10 mM DTT, 5% glycerol, 1 mM EDTA, 0.1% Digitonin (Calbiochem), pH 8.0) and the protein was eluted in wash buffer with 5 mM desthiobiotin. The resultant eluent was concentrated using 100 MWCO centrifugal filters to ~500 µl and further purified using size exclusion chromatography (SEC) on the Superose 6 Increase column at 4 °C. The SEC Buffer used was 300 mM NaCl, 20 mM Tris, 10 mM DTT, 0.1 mM GDN, pH = 8.0. All single-molecule experiments were performed within the peak fraction of the protein (which routinely contained 30-100 nM protein) within 2-6 hours of the SEC step.

The HCN2 construct used in this study (HCN2SM) was the mouse ortholog of HCN2 which was modified in two important ways. First, amino acids corresponding to residues 686-860 on the C-terminus were deleted to improve the biochemical behavior of the purified protein. Second, amino acids corresponding to residues 1-136 on the N-terminus were replaced by residues 1-98 from the human HCN1 ortholog to enable construct cloning which was challenging to perform with the native N terminus due to the preponderance of G/C nucleotides. HCN2SM, like HCN1SM, was tagged on the N-terminus with Twin

Strep tag and eGFP, and was expressed in suspension cultures of Freestyle HEK293 cells as described above for HCN1EM. Purification of HCN2SM was modified from that of HCN1SM in the following ways. The lysis buffer used 1% Digitonin (instead of L-MNG/CHS) and 30% glycerol and the wash/elution buffers for affinity purification included 20% glycerol (instead of 5%). These modifications significantly improved the polydispersity of the SEC profile with HCN2SM, although the profile still exhibited significant aggregation in the affinity purified material. Only the peak SEC fraction (containing 30-50 nM protein) was used for our studies and binding measurements were performed within 2-6 hours of the final protein purification step.

4.5.2 ZMW fabrication

ZMWs were fabricated at the Center for Nanophase Materials Sciences (CNMS) facility at Oak Ridge National Lab using positive-tone electron-beam lithography[37]. Cover glasses (Fisher Scientific Cat. No. 12-548-C) were cleaned by soaking in 5 parts deionized water, 1 part 30% hydrogen peroxide, 1 part 35% ammonium hydroxide for 15 minutes at 75°C. Substrates were rinsed, dried with N_2 gas, and plasma-cleaned with a Harrick PDC-32G for 10 minutes to remove any remaining organic impurities on the surface. The substrates were coated with thermally evaporated aluminum at a rate of 2 Å/second using a JEOL dual source E-beam evaporator to a final thickness of 100 nm. Substrates were spin-coated with the positive-tone electron-beam photoresist ZEP520A (ZEONREX Electronic Chemicals) for 45 seconds at 2,000 rpm followed by baking for 2 minutes at 180°C. ZMW features of 150 nm diameter dots were patterned using JEOL JBX-9300FS E-beam lithography system with a base dose of 450 $\mu C\ cm^{-2}$, 100 kV acceleration voltage, and 2 nA beam current. Following exposure, substrates were developed in xylenes for 30 seconds, rinsed with isopropyl alcohol, and dried with N_2. 100 nm of aluminum was dry etched in an Oxford Plasmalab System 100 Reactive Ion Etcher with a mixture of 30 standard cubic centimeters (sccm) chlorine (Cl_2) and 10 sscm boron trichloride (BCl3) gasses at 50°C for one minute. Following etching, the substrates were plasma cleaned with a Harrick PDC-32G for 15 minutes on a high setting to remove remaining photoresist. This resulted in arrays of round ZMW wells of 150 nm confirmed by scanning electron microscopy (**Figure 4-S3**).

4.5.3 Sample Preparation

Cover glasses intended for photobleaching experiments via TIRFM and not ZMW fabrication were cleaned by successive sonication for 60 minutes in 2% Hellmanex (Hellma), HPLC-grade ethanol (Millipore Sigma) and 1 M KOH, with deionized water rinses between solution exchanges. Both cover glasses and ZMW chips were additionally plasma cleaned for 5 minutes prior to surface functionalization. For ZMWs, the Al layer was passivated by incubation in 2% poly(vinylphonic acid) (PVPA) (Polysciences) for 3 minutes 90°C, followed by rinsing with Milli-Q ultrapure water and drying with Ar gas[39]. A silicone-gasketed chamber was attached to each substrate to hold small volumes and reduce evaporation (Grace Bio-Labs). Both cover glasses and ZMW chips were silanized overnight in 2 mg/mL biotin-PEG-silane (MW = 3,400 g mol^{-1}) and 10 mg/mL mPEG-silane (MW = 2,000 g mol^{-1}) (Laysan Bio Inc.) in HPLC-grade ethanol (Millipore Sigma) with 5% glacial acetic acid. Following incubation, samples were rinsed thoroughly with HPLC-grade ethanol, Milli-Q ultrapure water, and dried with Ar gas. Samples were additionally incubated with 10 mg/ml bovine serum albumin (BSA) in tris buffered saline (TBS: 300 mM NaCl, 20 mM Tris HCl, pH = 7.9) for 30 minutes to ensure robust passivation.

All single-molecule experiments were carried out in TBS supplemented with 1 mg/ml BSA, 100 µM GDN, and 5 mM DTT. Biotinylated cover glasses and ZMWs were sequentially incubated with 1 µM streptavidin (Prospec, cat # PRO-791) and 10 nM biotinylated GFP-TRAP (ChromoTek) for 10 minutes each. GFP tagged HCN1SM/HCN2SM molecules were pulled down to the surface by incubation at either 5 pM (cover glasses) or 250 nM (ZMW) for 10 minutes then thoroughly rinsed to remove freely diffusing GFP-HCN1/2 prior to imaging. For binding experiments, TBS was first bubbled with Argon for 30 minutes (prior to BSA, DTT, or GDN addition), and further supplemented with 2 mM Trolox[57,58], 2.5 mM protocatechuic acid (PCA) (Millipore Sigma), and various concentrations of of 8-(2-[DY-547]-aminoethylthio) adenosine-3',5'-cyclic monophosphate (fcAMP[13]; BioLog). Prior to 532 nm excitation, an additional 250 nM protocatechuate 3,4-dioxygenase from Pseudomonas sp. (PCD) (Millipore Sigma) was added to complete the oxygen scavenging system[59]. All solutions were replenished every 30 minutes to minimize evaporation and ensure the oxygen scavenging system was active. For each chip, passivation was first confirmed via incubation with either GFP tagged HCN1SM/HCN2SM or fcAMP prior to streptavidin and GFP-TRAP addition (**Figure 4-S4**).

4.5.4 Single-molecule imaging

Single-molecule fluorescence imaging was performed on an inverted microscope (Olympus, IX-71) with a high NA oil immersion objective (Olympus, 100x, 1.49 NA) and controlled by Metamorph software (Molecular Devices). Laser excitation at either 488 nm and 532 nm (Coherent, Sapphire LP) was fed into a single AOTF (Laser Launch) and guided into a single-mode fiber (Thorlabs). The beam was collimated with an achromatic lens (Thorlabs), passed through a quarter-wave plate (Thorlabs), and focused on the objective's back aperture with another achromatic lens (Thorlabs). Excitation and emission were filtered using two different dichroic and filter cubes applied separately (Semrock Brightline, LF488-C-000, Cy3/Cy5-A-OMF for fcAMP) and imaged on a 512 × 512 EMCCD (Andor iXon Ultra X-888) at 10 Hz. This set-up enabled simultaneous recording of ~1,600 ZMWs at a time in an approximately 80 x 80 µm field of view; however, proteins were sparsely deposited onto the array in order to reduce the probability of having more than one protein per ZMW. On average, 313 ± 188 ZMWS were occupied per field of view. Considering Poisson statistics for single-molecule deposition[60], the observed deposition rate (λ = 0.22) leads us to only anticipate ~2% of ZMWs per field of view to contain more than one protein.

4.5.5 Single-molecule analysis

Single-molecule fluorescence time trajectories were extracted from tiff stacks saved by Metamorph using MATLAB (Mathworks). For both TIRFM and ZMW experiments, locations of single molecules were identified by GFP emission. For each image stack, an image mask was created by averaging the first 100 images and removing background with a top-hat. Otsu's method was then used to automatically threshold the image to identify potential single-molecule locations in the binary mask[61]. Identified locations with an area greater than 4-pixels and at least 5-pixel separation between all neighboring locations were considered a region of interest (ROI). ROI locations were refined using a 2D Gaussian fit the local intensity height map on the average image. For co-location experiments of fcAMP binding using ZMWs, ROIs identified in the 488 nm channel (GFP photobleaching steps) were linearly transformed to the 532 nm channel (fcAMP) followed by 2D Gaussian refinement. The time-dependent fluorescence at each ROI was obtained by projecting the average image intensity in a 7 x 7-pixel square centered around the ROI for each image of the stack.

All statistical analysis was performed using MATLAB unless otherwise stated. The divisive segmentation and clustering (DISC) algorithm applied to each fcAMP binding trajectory for an unbiased detection of discrete states (number of ligands bound) and transitions[41]. States are identified in DISC using a top-down unsupervised clustering algorithm and transitions are determined using the Viterbi algorithm. All idealized traces were visually inspected following idealization. Traces featuring greater than five discrete states and/or low signal to noise ratios were removed from analysis. In addition, single change-point detection was applied to truncate traces exhibiting an asynchronous decay of activity over time, a phenomenon previously observed in both bulk and single-molecule studies which may be caused by free oxygen radicals modifying CNBDs[41,62]. A summary of the final HCN1SM and HCN2SM data is provided in **Table 4-S1** and representative trajectories are provided in **Figure 4-1**, **4-S5**, and **4-S6**.

4.5.6 HMM analysis

HMM analysis of idealized datasets for HCN1SM and HCN2SM were performed with QuB[44,45]. The first and last event of each trajectory was removed prior to analysis to avoid interpretation of truncated events. Models were globally optimized to simultaneously describe the idealized binding events for all molecules across all fcAMP concentrations. The goodness of fit relative to the number of free parameters of each model was assessed by Bayesian Information Criterion (BIC)[63]

$$BIC = k * \ln(N) - 2 * LL \qquad \text{Eq. 4-2}$$

where k is the number of free parameters in the model, N is the total number data points (frames) across all fcAMP concentrations, and LL is the loglikelihood of the model returned by maximum idealized point (MIP) estimation in QuB. The model with the lower BIC value was considered the better fit.

An additional 3-fold resampling procedure akin to cross-validation was performed to evaluate the homogeneity of the optimized rates within the data. Trajectories were stratified by fcAMP prior resampling with a 33% hold-out and individually used to optimize rates of both models. In this paradigm, each trajectory was included in two of the three samples and left out of one sample. The optimized rates of each model across each fold showed homogeneity in the obtained rates and all showed preference for Model 2 by BIC (**Table 4-S5**).

4.5.7 Data and code availability

All experimental data are available upon reasonable request. The DISC idealization software package is available at https://github.com/ChandaLab/DISC and fully described elsewhere[41]. All additional MATLAB scripts for single-molecule analysis and image processing are available upon reasonable request.

4.6 References

1. Gauss, R., Seifert, R. & Kaupp, U.B. Molecular identification of a hyperpolarization-activated channel in sea urchin sperm. *Nature* **393**, 583-587 (1998).
2. Ludwig, A., Zong, X.G., Jeglitsch, M., Hofmann, F. & Biel, M. A family of hyperpolarization-activated mammalian cation channels. *Nature* **393**, 587-591 (1998).
3. Santoro, B. et al. Identification of a gene encoding a hyperpolarization-activated pacemaker channel of brain. *Cell* **93**, 717-729 (1998).
4. Santoro, B. & Tibbs, G.R. The HCN gene family: Molecular basis of the hyperpolarization-activated pacemaker channels. *Molecular and Functional Diversity of Ion Channels and Receptors* **868**, 741-764 (1999).
5. Difrancesco, D. CHARACTERIZATION OF SINGLE PACEMAKER CHANNELS IN CARDIAC SINOATRIAL NODE CELLS. *Nature* **324**, 470-473 (1986).
6. Wang, J., Chen, S. & Siegelbaum, S.A. Regulation of hyperpolarization-activated HCN channel gating and cAMP modulation due to interactions of COOH terminus and core transmembrane regions. *Journal of General Physiology* **118**, 237-250 (2001).
7. Wahl-Schott, C. & Biel, M. HCN channels: Structure, cellular regulation and physiological function. *Cellular and Molecular Life Sciences* **66**, 470-494 (2009).
8. Robinson, R.B. Hyperpolarization-activated cation currents: From molecules to physiological function. *Annual Review of Physiology* **65**, 453-480 (2003).
9. Craven, K.B. & Zagotta, W.N. CNG and HCN channels: Two peas, one pod. *Annual Review of Physiology* **68**, 375-401 (2006).
10. Lee, C.H. & MacKinnon, R. Structures of the Human HCN1 Hyperpolarization-Activated Channel. *Cell* **168**, 111-+ (2017).
11. Alvarez-Baron, C.P., Klenchin, V.A. & Chanda, B. Minimal molecular determinants of isoform-specific differences in efficacy in the HCN channel family. *Journal of General Physiology* **150**, 1203-1213 (2018).
12. Biskup, C. et al. Relating ligand binding to activation gating in CNGA2 channels. *Nature* **446**, 440-443 (2007).
13. Kusch, J. et al. Interdependence of Receptor Activation and Ligand Binding in HCN2 Pacemaker Channels. *Neuron* **67**, 75-85 (2010).
14. Kusch, J. et al. How subunits cooperate in cAMP-induced activation of homotetrameric HCN2 channels. *Nat Chem Biol* **8**, 162-9 (2011).
15. Ulens, C. & Siegelbaum, S.A. Regulation of hyperpolarization-activated HCN channels by cAMP through a gating switch in binding domain symmetry. *Neuron* **40**, 959-970 (2003).
16. Thon, S., Schulz, E., Kusch, J. & Benndorf, K. Conformational Flip of Nonactivated HCN2 Channel Subunits Evoked by Cyclic Nucleotides. *Biophys J* **109**, 2268-76 (2015).
17. Zagotta, W.N. et al. Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* **425**, 200-5 (2003).
18. Puljung, M.C., DeBerg, H.A., Zagotta, W.N. & Stoll, S. Double electron-electron resonance reveals cAMP-induced conformational change in HCN channels. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9816-9821 (2014).
19. Hines, K.E., Middendorf, T.R. & Aldrich, R.W. Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *Journal of General Physiology* **143**, 401-416 (2014).
20. Ha, T. Single-molecule methods leap ahead. *Nature Methods* **11**, 1015-1018 (2014).

21. Moerner, W.E. & Fromm, D.P. Methods of single-molecule fluorescence spectroscopy and microscopy. *Review of Scientific Instruments* **74**, 3597-3619 (2003).
22. Moerner, W.E., Shechtman, Y. & Wang, Q. Single-molecule spectroscopy and imaging over the decades. *Faraday Discussions* **184**, 9-36 (2015).
23. Lerner, E. et al. Toward dynamic structural biology: Two decades of single-molecule Forster resonance energy transfer. *Science* **359**, 288-+ (2018).
24. Holzmeister, P., Acuna, G.P., Grohmann, D. & Tinnefeld, P. Breaking the concentration limit of optical single-molecule detection. *Chemical Society Reviews* **43**, 1014-1028 (2014).
25. van Oijen, A.M. Single-molecule approaches to characterizing kinetics of biomolecular interactions. *Current Opinion in Biotechnology* **22**, 75-80 (2011).
26. Peng, S.J., Wang, W.J. & Chen, C.L. Breaking the Concentration Barrier for Single-Molecule Fluorescence Measurements. *Chemistry-a European Journal* **24**, 1002-1009 (2018).
27. Goldschen-Ohm, M.P., White, D.S., Klenchin, V.A., Chanda, B. & Goldsmith, R.H. Observing Single-Molecule Dynamics at Millimolar Concentrations. *Angewandte Chemie-International Edition* **56**, 2399-2402 (2017).
28. Eid, J. et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**, 133-138 (2009).
29. Uemura, S. et al. Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464**, 1012-U73 (2010).
30. Choi, J.H. et al. Dynamics of the context-specific translation arrest by chloramphenicol and linezolid. *Nature Chemical Biology* **16**, 310-+ (2020).
31. Christensen, S.M. et al. Monitoring the Waiting Time Sequence of Single Ras GTPase Activation Events Using Liposome Functionalized Zero-Mode Waveguides. *Nano Letters* **16**, 2890-2895 (2016).
32. Richards, C.I. et al. Live-Cell Imaging of Single Receptor Composition Using Zero-Mode Waveguide Nanostructures. *Nano Letters* **12**, 3690-3694 (2012).
33. Miyake, T. et al. Real-time imaging of single-molecule fluorescence with a zero-mode waveguide for the analysis of protein-protein interaction. *Analytical Chemistry* **80**, 6018-6022 (2008).
34. Goldschen-Ohm, M.P. et al. Structure and dynamics underlying elementary ligand binding events in human pacemaking channels. *Elife* **5**(2016).
35. Ulbrich, M.H. & Isacoff, E.Y. Subunit counting in membrane-bound proteins. *Nature Methods* **4**, 319-321 (2007).
36. Jain, A. et al. Probing cellular protein complexes using single-molecule pull-down. *Nature* **473**, 484-U322 (2011).
37. Levene, M.J. et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686 (2003).
38. Zhu, P. & Craighead, H.G. Zero-Mode Waveguides for Single-Molecule Analysis. in *Annual Review of Biophysics, Vol 41*, Vol. 41 (ed. Rees, D.C.) 269-293 (2012).
39. Korlach, J. et al. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1176-1181 (2008).
40. Huang, R.C. & Gillette, R. KINETIC-ANALYSIS OF CAMP-ACTIVATED NA+ CURRENT IN THE MOLLUSCAN NEURON - A DIFFUSION REACTION MODEL. *Journal of General Physiology* **98**, 835-848 (1991).
41. White, D.S., Goldschen-Ohm, M.P., Goldsmith, R.H. & Chanda, B. Top-down machine learning approach for high-throughput single-molecule analysis. *Elife* **9**(2020).
42. Ding, S. & Sachs, F. Evidence for non-independent gating of P2X 2 receptors expressed in Xenopusoocytes. *BMC neuroscience* **3**, 17 (2002).

43. Horn, R. ESTIMATING THE NUMBER OF CHANNELS IN PATCH RECORDINGS. *Biophysical Journal* **60**, 433-439 (1991).
44. Nicolai, C. & Sachs, F. Solving ion channel kinetics with the QuB software. *Biophysical Reviews and Letters* **8**, 191-211 (2013).
45. Qin, F., Auerbach, A. & Sachs, F. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophysical Journal* **79**, 1915-1927 (2000).
46. Jadey, S. & Auerbach, A. An integrated catch-and-hold mechanism activates nicotinic acetylcholine receptors. *Journal of General Physiology* **140**, 17-28 (2012).
47. Purohit, P., Bruhova, I., Gupta, S. & Auerbach, A. Catch-and-Hold Activation of Muscle Acetylcholine Receptors Having Transmitter Binding Site Mutations. *Biophysical Journal* **107**, 88-99 (2014).
48. Lolicato, M. et al. Tetramerization Dynamics of C-terminal Domain Underlies Isoform-specific cAMP Gating in Hyperpolarization-activated Cyclic Nucleotide-gated Channels. *Journal of Biological Chemistry* **286**, 44811-44820 (2011).
49. Taraska, J.W., Puljung, M.C., Olivier, N.B., Flynn, G.E. & Zagotta, W.N. Mapping the structure and conformational movements of proteins with transition metal ion FRET. *Nature Methods* **6**, 532-U94 (2009).
50. Zagotta, W.N., Hoshi, T. & Aldrich, R.W. SHAKER POTASSIUM CHANNEL GATING .3. EVALUATION OF KINETIC-MODELS FOR ACTIVATION. *Journal of General Physiology* **103**, 321-362 (1994).
51. Schwarz, G. ESTIMATING DIMENSION OF A MODEL. *Annals of Statistics* **6**, 461-464 (1978).
52. Lolicato, M. et al. Tetramerization dynamics of C-terminal domain underlies isoform-specific cAMP gating in hyperpolarization-activated cyclic nucleotide-gated channels. *J Biol Chem* **286**, 44811-20 (2011).
53. Lape, R., Colquhoun, D. & Sivilotti, L.G. On the nature of partial agonism in the nicotinic receptor superfamily. *Nature* **454**, 722-U56 (2008).
54. Akimoto, M. et al. A Mechanism for the Auto-inhibition of Hyperpolarization-activated Cyclic Nucleotide-gated (HCN) Channel Opening and Its Relief by cAMP. *Journal of Biological Chemistry* **289**, 22205-22220 (2014).
55. Zagotta, W.N. et al. Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* **425**, 200-205 (2003).
56. Goehring, A. et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nature Protocols* **9**, 2574-2585 (2014).
57. Cordes, T., Vogelsang, J. & Tinnefeld, P. On the Mechanism of Trolox as Antiblinking and Antibleaching Reagent. *Journal of the American Chemical Society* **131**, 5018-+ (2009).
58. Rasnik, I., McKinney, S.A. & Ha, T. Nonblinking and longlasting single-molecule fluorescence imaging. *Nature Methods* **3**, 891-893 (2006).
59. Aitken, C.E., Marshall, R.A. & Puglisi, J.D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical Journal* **94**, 1826-1835 (2008).
60. Pibiri, E., Holzmeister, P., Lalkens, B., Acuna, G.P. & Tinnefeld, P. Single-Molecule Positioning in Zeromode Waveguides by DNA Origami Nanoadapters. *Nano Letters* **14**, 3499-3503 (2014).
61. Otsu, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* **9**, 62-66 (1979).
62. Idikuda, V. et al. Singlet oxygen modification abolishes voltage-dependent inactivation of the sea urchin spHCN channel. *Journal of General Physiology* **150**, 1273-1286 (2018).
63. Schwarz, G. Estimating the dimension of a model. *The annals of statistics* **6**, 461-464 (1978).

4.7 Supplementary information

Table 4-S1: Summary of SMHCN1 and SMHCN2 single-molecule data

Protein	[fcAMP] nM	Number of Molecules	Observation Time (min)	Number of Events	Fraction Bound*
SMHCN1	100	176	865	18,264	0.04 ± 0.02
	250	55	254	10,390	0.13 ± 0.03
	300	135	658	35,338	0.18 ± 0.07
	500	157	761	44,792	0.24 ± 0.06
	700	115	560	40,000	0.30 ± 0.04
	750	54	246	15,119	0.32 ± 0.05
	900	47	265	16,577	0.34 ± 0.04
SMHCN2	100	77	379	6,532	0.04 ± 0.02
	250	43	209	3,772	0.06 ± 0.03
	500	64	308	12,101	0.13 ± 0.04
	750	71	328	15,315	0.22 ± 0.07
	1000	140	650	33,332	0.26 ± 0.07
	1500	49	224	10,978	0.31 ± 0.07

*Fraction bound values given as mean ± s.d.

Table 4-S2: Total observed and expected binomial occupancies of ligand bound states

Protein	fcAMP (nM)	Binding Rate	Number of fcAMP(s) Bound				
			0	1	2	3	4
SMHCN1	100	0.04	0.85	0.15	4.8×10^{-3}	0	0
		0.04	*0.85*	*0.14*	*8.7×10^{-3}*	*2.4×10^{-4}*	*2.5×10^{-6}*
	250	0.13	0.56	0.35	0.08	5.5×10^{-3}	0
		0.13	*0.57*	*0.34*	*0.08*	*7.9×10^{-3}*	*3.0×10^{-4}*
	300	0.18	0.45	0.40	0.14	0.02	1.4×10^{-4}
		0.18	*0.45*	*0.40*	*0.13*	*0.02*	*1.1×10^{-3}*
	500	0.24	0.33	0.43	0.20	0.04	6.8×10^{-4}
		0.24	*0.34*	*0.42*	*0.20*	*0.04*	*3.2×10^{-3}*
	700	0.30	0.25	0.40	0.27	0.08	4.1×10^{-3}
		0.30	*0.24*	*0.41*	*0.26*	*0.07*	*7.9×10^{-3}*
	750	0.32	0.22	0.38	0.30	0.09	0.01
		0.32	*0.21*	*0.40*	*0.29*	*0.09*	*0.01*
	900	0.34	0.18	0.40	0.31	0.1	0.01
		0.34	*0.19*	*0.39*	*0.30*	*0.1*	*0.01*
SMHCN2	100	0.04	0.85	0.15	0.01	0	0
		0.04	0.85	0.14	0.01	*2.4×10^{-4}*	*2.5×10^{-6}*
	250	0.06	0.78	0.20	0.02	7.9×10^{-5}	0
		0.06	0.78	0.20	0.02	*7.3×10^{-3}*	*1.2×10^{-5}*
	500	0.13	0.56	0.35	0.08	4.4×10^{-3}	0
		0.13	0.57	0.34	0.08	*0.01*	*2.8×10^{-4}*
	750	0.22	0.36	0.43	0.17	0.03	1.9×10^{-3}
		0.22	0.36	0.42	0.18	*0.03*	*2.5×10^{-3}*
	1000	0.26	0.30	0.44	0.20	0.05	3.6×10^{-3}
		0.25	0.31	0.42	0.21	*0.05*	*4.1×10^{-3}*
	1500	0.31	0.23	0.40	0.28	0.09	0.01
		0.31	0.22	0.41	0.28	*0.08*	*0.01*

*Expected values italicized.

Table 4-S3: Maximum likelihood estimation of isolated-B_1 dwell times distributions

Protein	fcAMP (nM)	Monoexponential Fit		Biexponential Fit		
		Tau (s)	LL	Tau (s)	Amplitude	LL
SMHCN1	100	0.82 ± 0.02	-6.9×10^3	0.56 ± 0.02	0.86 ± 0.02	-6.3×10^3
				2.4 ± 0.25	1.4 ± 0.02	
	250	0.95 ± 0.03	-2.8×10^3	0.69 ± 0.04	0.90 ± 0.03	-2.6×10^3
				3.3 ± 0.68	0.10 ± 0.03	
	300	0.86 ± 0.02	-6.7×10^3	0.71 ± 0.03	0.93 ± 0.03	-6.5×10^3
				2.7 ± 0.48	0.07 ± 0.03	
	500	0.89 ± 0.02	-6.7×10^3	0.77 ± 0.04	0.91 ± 0.04	-6.6×10^3
				2.2 ± 0.45	0.09 ± 0.04	
	700	0.75 ± 0.02	-3.2×10^3	0.70 ± 0.03	0.97 ± 0.03	-3.1×10^3
				2.3 ± 0.90	0.03 ± 0.03	
	750	0.89 ± 0.05	-1.2×10^3	0.81 ± 0.02	0.94 ± 0.14	-1.2×10^3
				2.0 ± 1.51	0.06 ± 0.14	
	900	1.1 ± 0.07	-1.1×10^3	1.1 ± 0.29	0.51 ± 59.7	-1.1×10^3
				1.1 ± 0.31	0.49 ± 59.7	
SMHCN2	100	0.89 ± 0.03	-2.5×10^3	0.56 ± 0.03	0.89 ± 0.03	-2.2×10^3
				3.6 ± 0.62	0.11 ± 0.03	
	250	1.2 ± 0.06	-1.8×10^3	0.60 ± 0.05	0.82 ± 0.04	-1.6×10^3
				4.0 ± 0.71	0.18 ± 0.04	
	500	0.94 ± 0.03	-3.4×10^3	0.64 ± 0.04	0.86 ± 0.04	-3.2×10^3
				2.8 ± 0.46	0.14 ± 0.04	
	750	0.94 ± 0.03	-2.6×10^3	0.69 ± 0.04	0.93 ± 0.02	-2.4×10^3
				4.5 ± 1.10	0.07 ± 0.02	
	1000	1.0 ± 0.03	-5.1×10^3	0.81 ± 0.04	0.91 ± 0.03	-5.0×10^3
				3.1 ± 0.57	0.09 ± 0.03	
	1500	1.1 ± 0.07	-1.2×10^3	0.85 ± 0.09	0.91 ± 0.06	-1.2×10^3
				4.1 ± 1.56	0.09 ± 0.06	

*Values are mean ± 95% confidence interval.

Table 4-S4: Optimized rate constants for simulations in Figure 4-S10 and Figure 4-S11

Simulation	ID	Rates (M^{-1} s^{-1} or s^{-1})				ΔBIC
		a	b	c	d	
1	True Rates	1.3 x 10^5	0.34			
	Model 1	1.3 x 10^5	0.34			0
	Model 2	1.3 x 10^5	0.34	204	1.5 x 10^3	28.8
2	True Rates	1.4 x 10^5	0.91	0.52	0.31	
	Model 1	1.4 x 10^5	0.33			0
	Model 2	1.4 x 10^5	0.96	0.54	0.29	-1.2 x 10^4

Table 4-S5: Optimized rate constants for re-sampled HCN1SM and HCN2SM data

Protein	Sample	Model	Rates (M^{-1} s^{-1} or s^{-1})				ΔBIC
			a	b	c	d	
HCN1SM	1	1	4.2 x 10^5	0.67			7,251
		2	4.2 x 10^5	0.94	0.01	0.25	0
	2	1	4.1 x 10^5	0.66			
		2	4.2 x 10^5	0.92	0.01	0.25	7,289
	3	1	4.1 x 10^5	0.66			6,613
		2	4.2 x 10^5	0.92	0.09	0.25	0
HCN2SM	1	1	1.8 x 10^5	0.55			7,047
		2	1.9 x 10^5	0.98	0.16	0.22	0
	2	1	1.8 x 10^5	0.53			7,742
		2	1.9 x 10^5	0.96	0.14	0.19	0
	3	1	1.8 x 10^5	0.53			7,404
		2	1.9 x 10^5	0.96	0.14	0.19	0

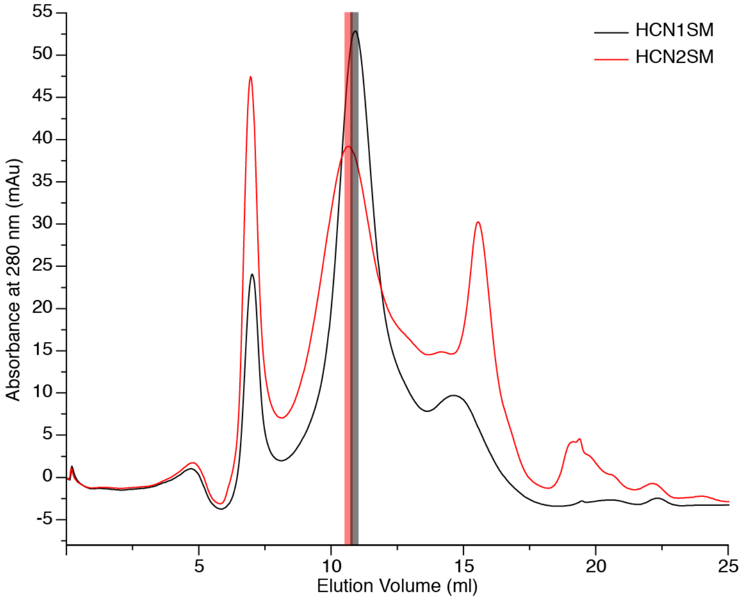

Figure 4-S1: Size exclusion chromatography profiles of purified HCN1SM and HCN2SM. *The SEC profiles of affinity purified HCN1SM and HCN2SM are shown in black and red respectively. In both cases, the running buffer was 300 mM NaCl 20 mM 10 mM DTT 0.1 mM GDN. The peak fraction (~0.3 ml) used for the single-molecule studies are depicted by the shaded boxes. Significantly larger aggregates are seen for HCN2SM than HCN1SM and the peak is relatively much broader for HCN2SM than HCN1SM.*

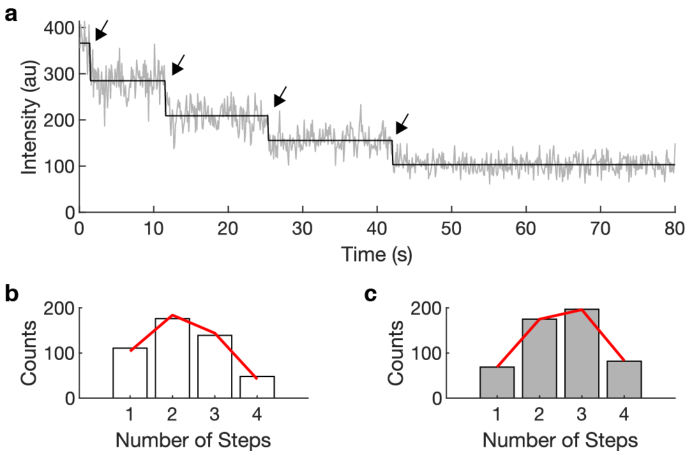

Figure 4-S2: Photobleaching steps of purified HCN1SM and HCN2SM. *(**a**) Example photobleaching trajectory of eGFP-tagged HCN2SM tetramers on a cover glass with four identified photobleaching steps (arrows) overlaid with idealized fit (black). Distributions of HCN1SM (**b**) and HCN2SM (**c**) overlaid with maximum likelihood estimations of a zero-truncated binomial distribution (HCN1SM:* $N = 474$*,* $B(n = 4,\ p = 0.54)$*; HCN2SM:* $N = 523$*,* $B(n = 4,\ p = 0.63)$*).*

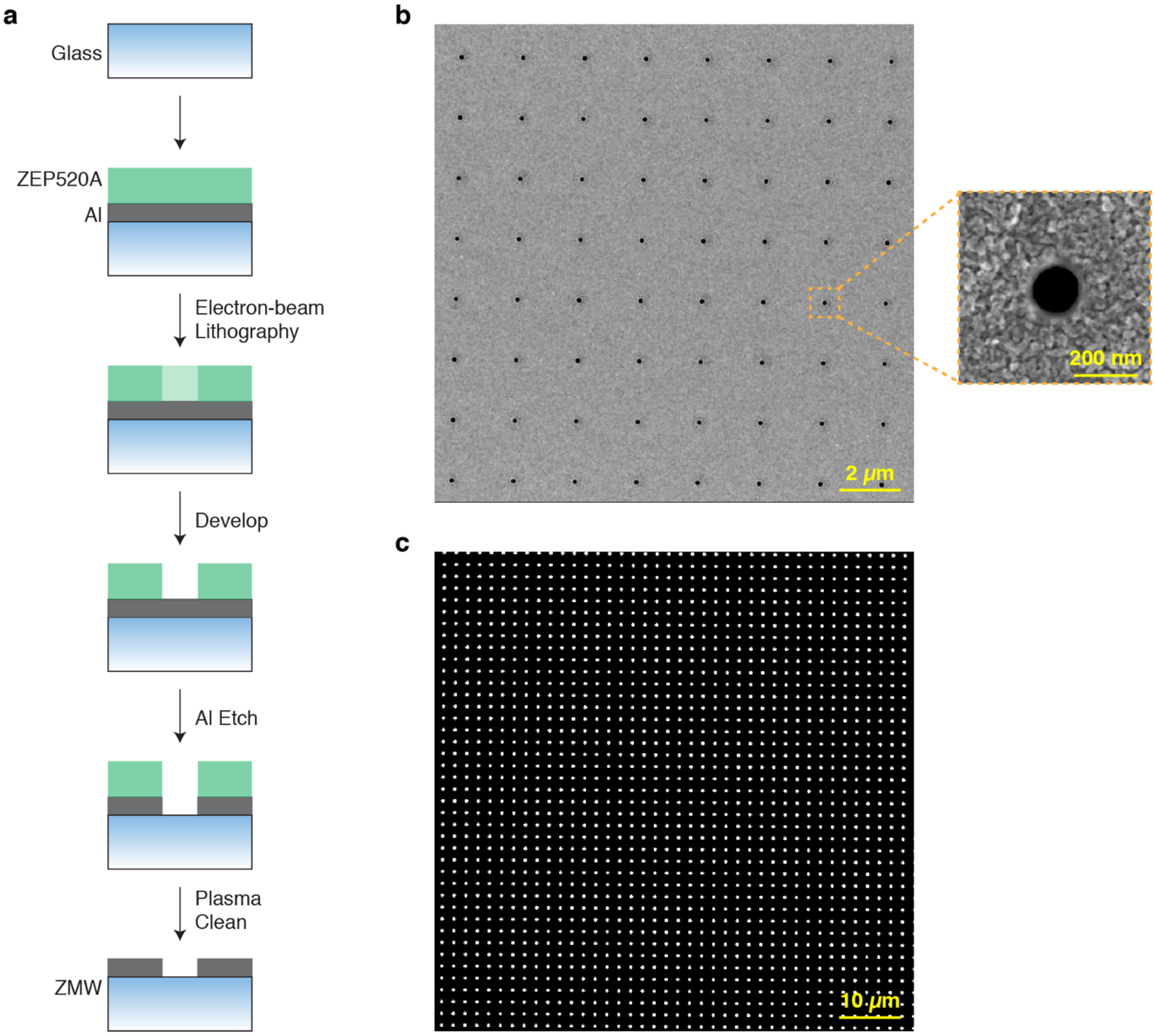

Figure 4-S3: Zero-mode waveguide fabrication. *(**a**) General schematic for the fabrication of a zero-mode waveguide (ZMW) using positive tone electron-beam lithography (see **4.5.2**). (**b**) Scanning electron microcopy image of ZMW array. Inset shows a single ZMW with a dimeter of 150 nm. (**c**) Brightfield image of ZMW array on single-molecule imaging set-up featuring a 512x512 EMCCD and a 100x objective (NA = 1.49).*

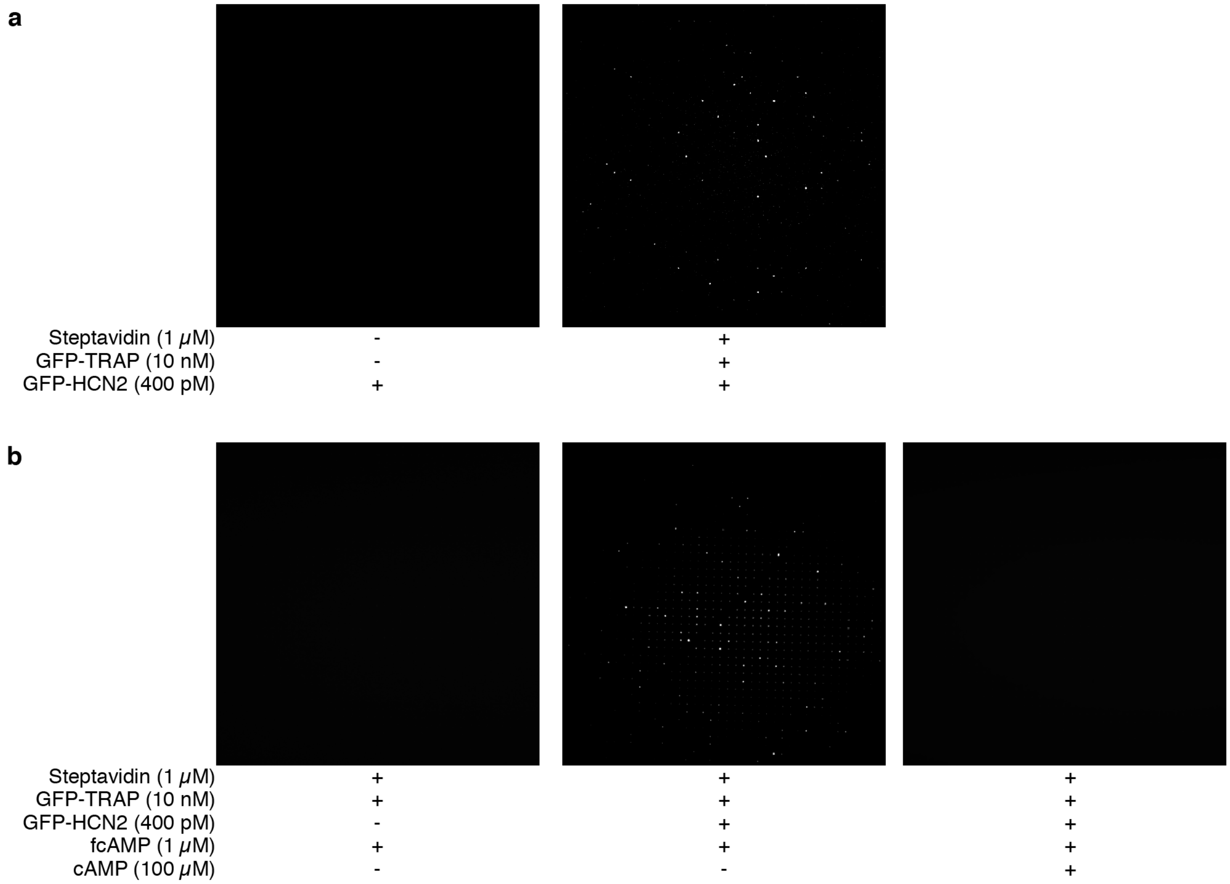

Figure 4-S4: Passivation test of ZMWs. *(**a**) Test of specific binding of eGFP-tagged HCN2SM to ZMWs in the absence and presence of streptavidin and GFP-TRAP. (**b**) Test of specific binding of fcAMP to HCN2SM in ZMWs. fcAMP binding can be competed off by addition of non-fluorescent cAMP. All images shown are averaged over the first 10 frames (1 second), background subtracted, and adjusted for brightness and contrast for clarity.*

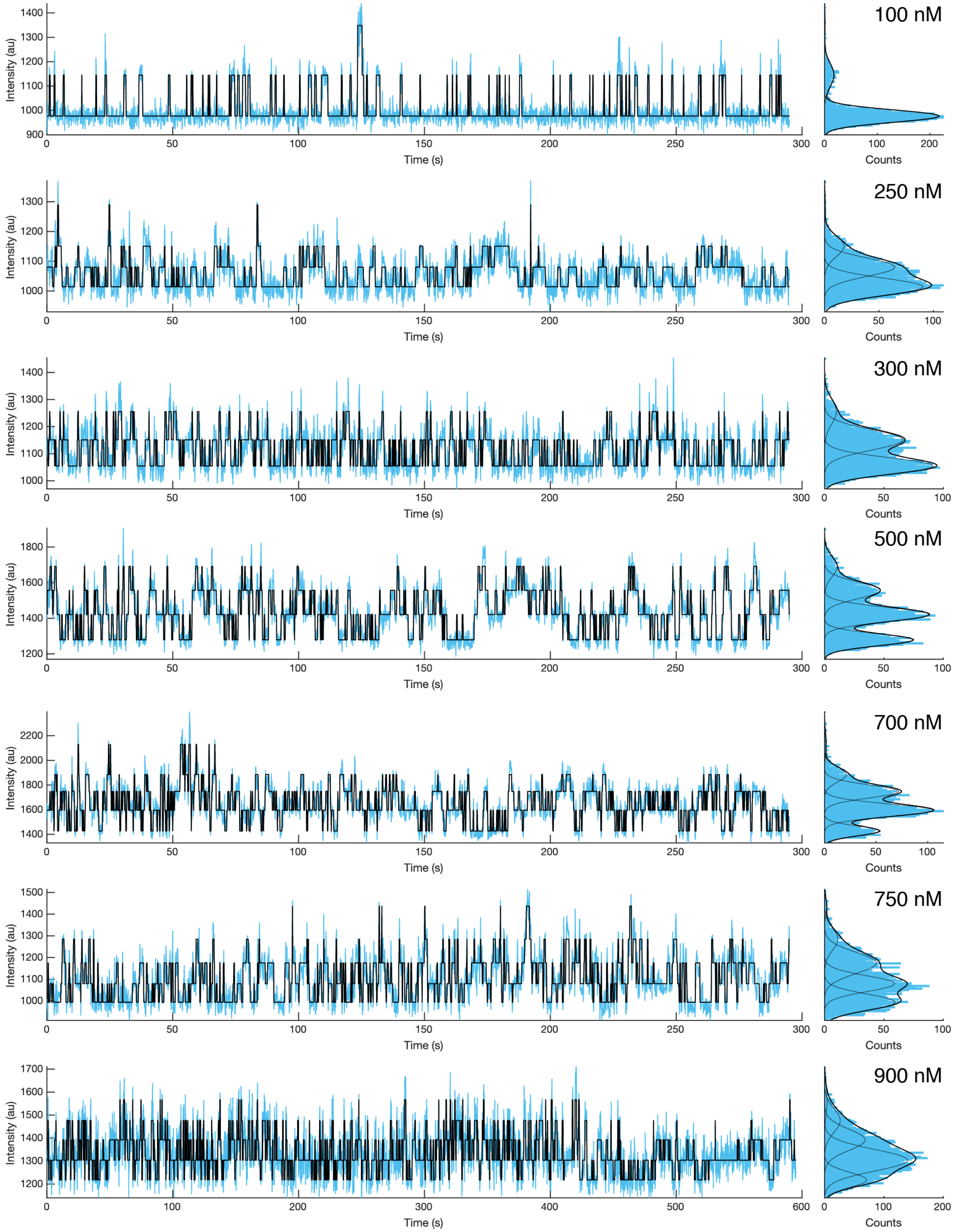

Figure 4-S5: HCN1SM single-molecule binding trajectories. *Representative time fluorescence trajectories of various fcAMP concentrations (100 nM to 900 nM) binding to HCN1SM in ZMWs with idealized fits (black).*

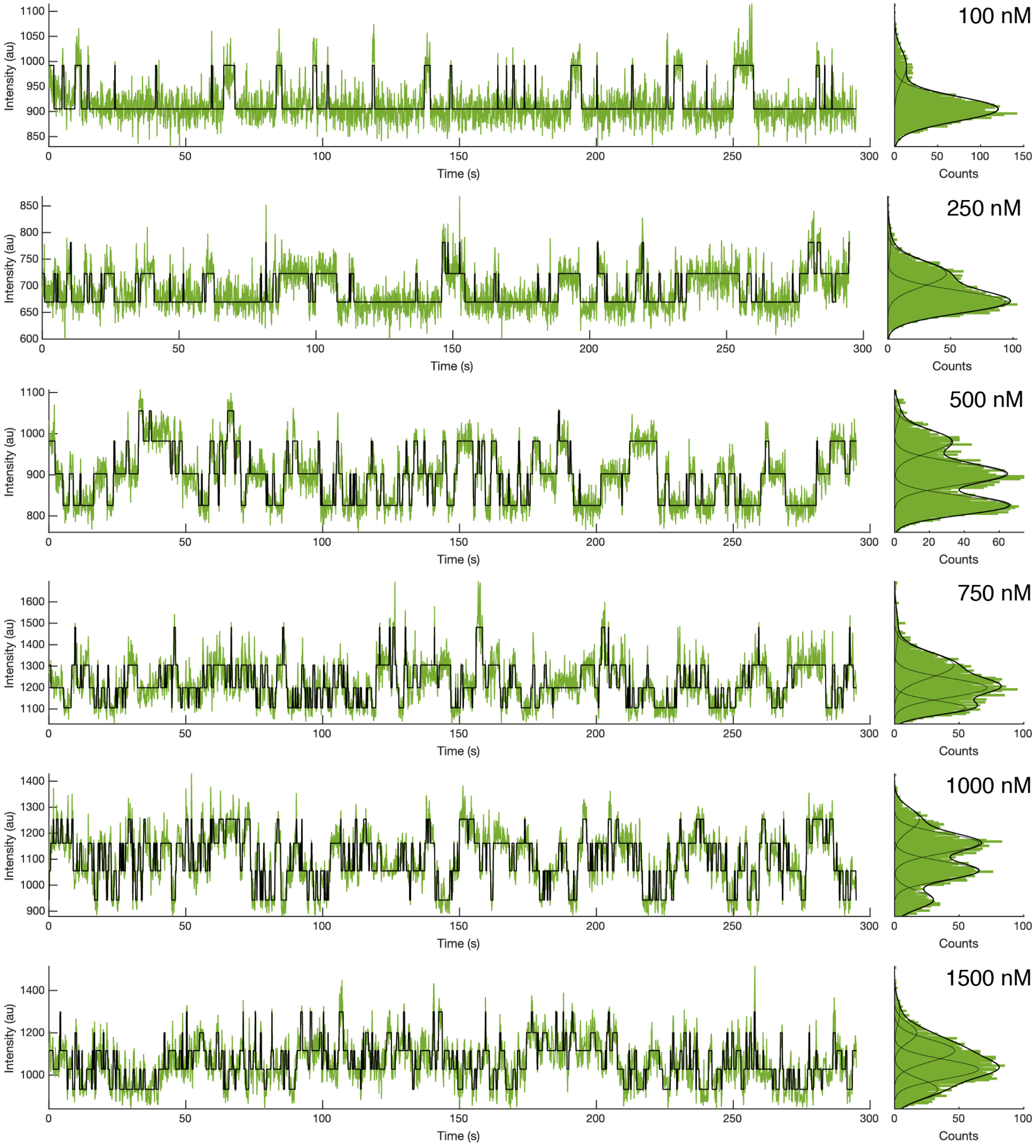

Figure 4-S6: HCN2SM single-molecule binding trajectories. *Representative time fluorescence trajectories of various fcAMP concentrations (100 nM to 1500 nM) binding to HCN2SM in ZMWs with idealized fits (black).*

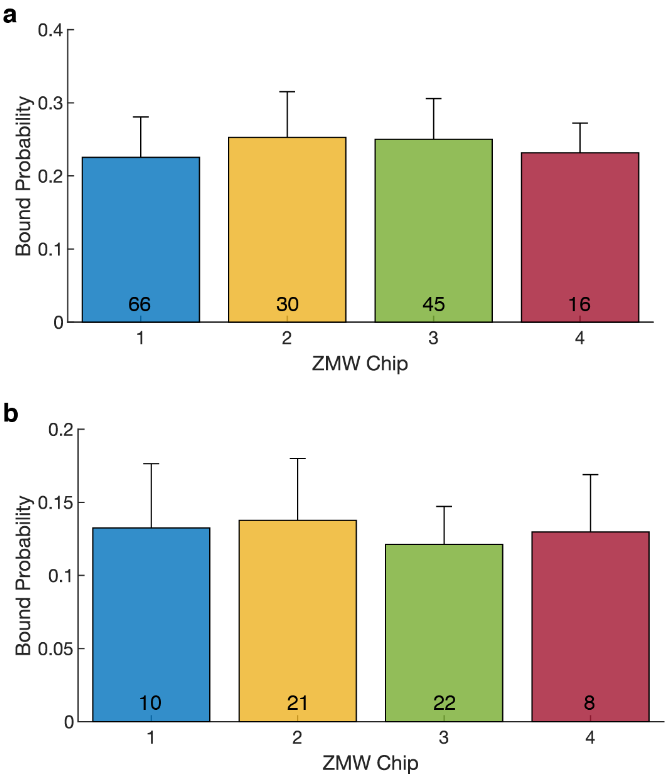

Figure 4-S7: Protein function across ZMW chips is consistent. *Bound probability of across separate ZMW chips with HCN1SM (**a**) and HCN2SM (**b**) at 500 nM fcAMP (mean ± s.d.). Each bar is labeled with the number of proteins included in the analysis.*

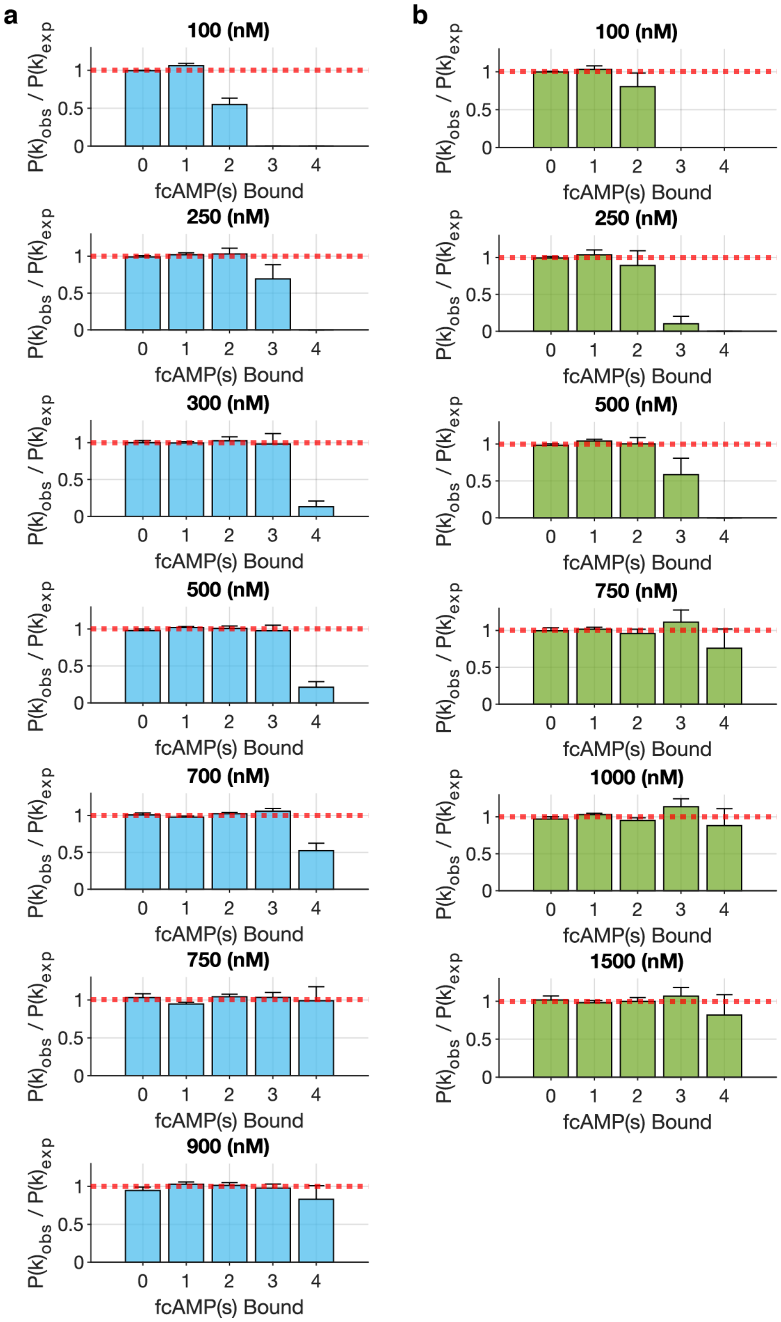

Figure 4-S8: Observed vs expected state occupancies. *Observed values of liganded state occupancies ($P(k)_{obs}$) divided by expectations from binomial fits ($P(k)_{exp}$) across all trajectories for each fcAMP concentration (mean ± s.e.m) of HCN1SM (**a**) and HCN2SM (**b**). Values of 1 indicate a match between observed and expected values (red dashed line).*

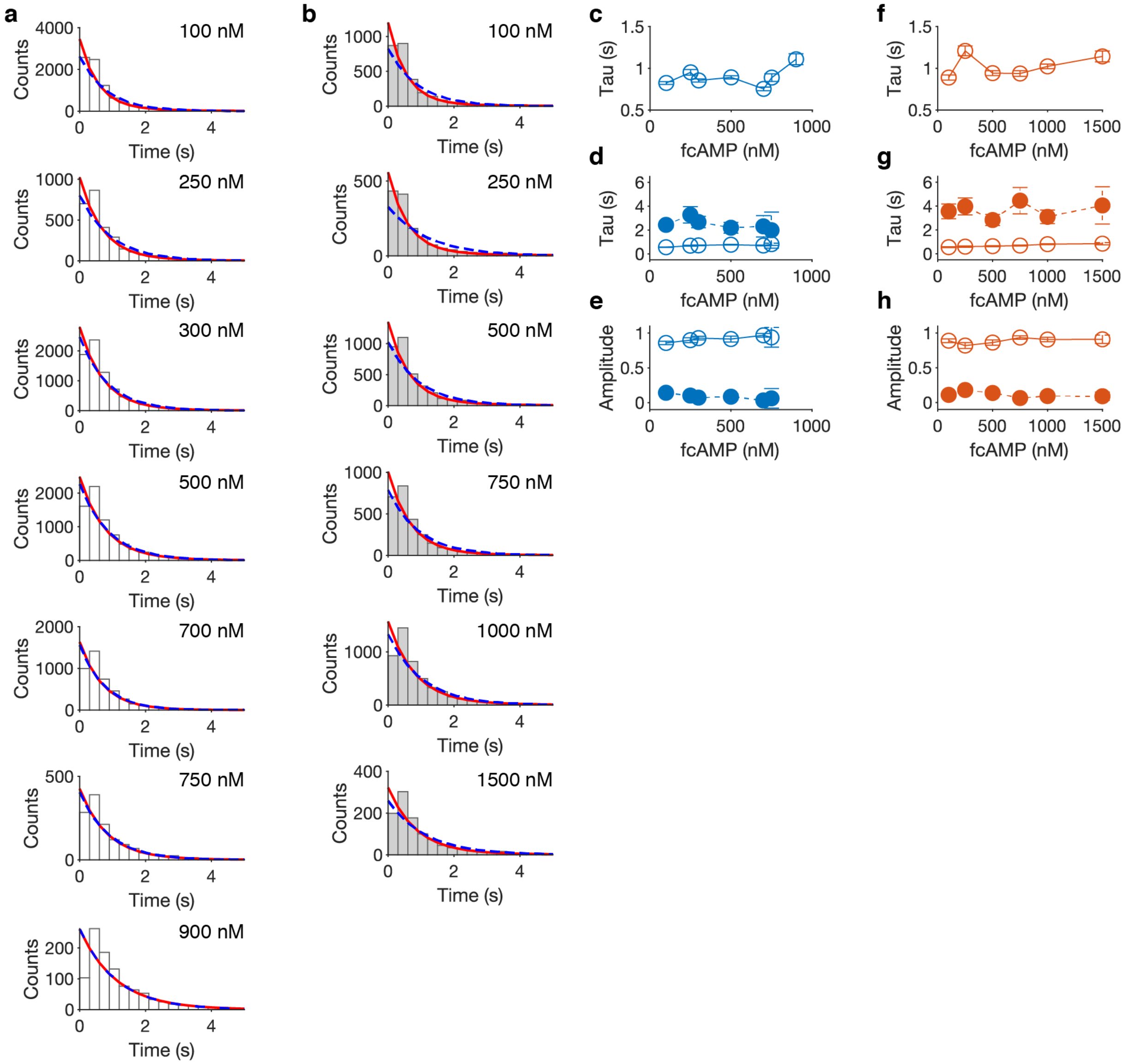

Figure 4-S9: Dwell time distribution analysis of individual CNBDs. *Dwell time distributions of isolated-B$_1$ events for HCN1SM (**a**) and HCN2SM (**b**) at various fcAMP concentrations overlaid with maximum likelihood estimates for monoexponential (blue dashed) and biexponential (red) distributions. Monoexponential rate constants for HCN1SM (**c**) and HCN2SM (**f**) across fcAMP concentrations. Summary of biexponential rates constants and amplitudes for HCN1SM (**d**, **e**) and HCN2SM (**g**, **h**) with corresponding amplitudes. Open circles indicate minor component, filled circles indicate major components. Error bars are 95% confidence intervals. Note, biexponential values of HCN1SM at 900 nM are not shown due to a regression to a monoexponential distribution. All values are additionally provided in **Table 4-S3**.*

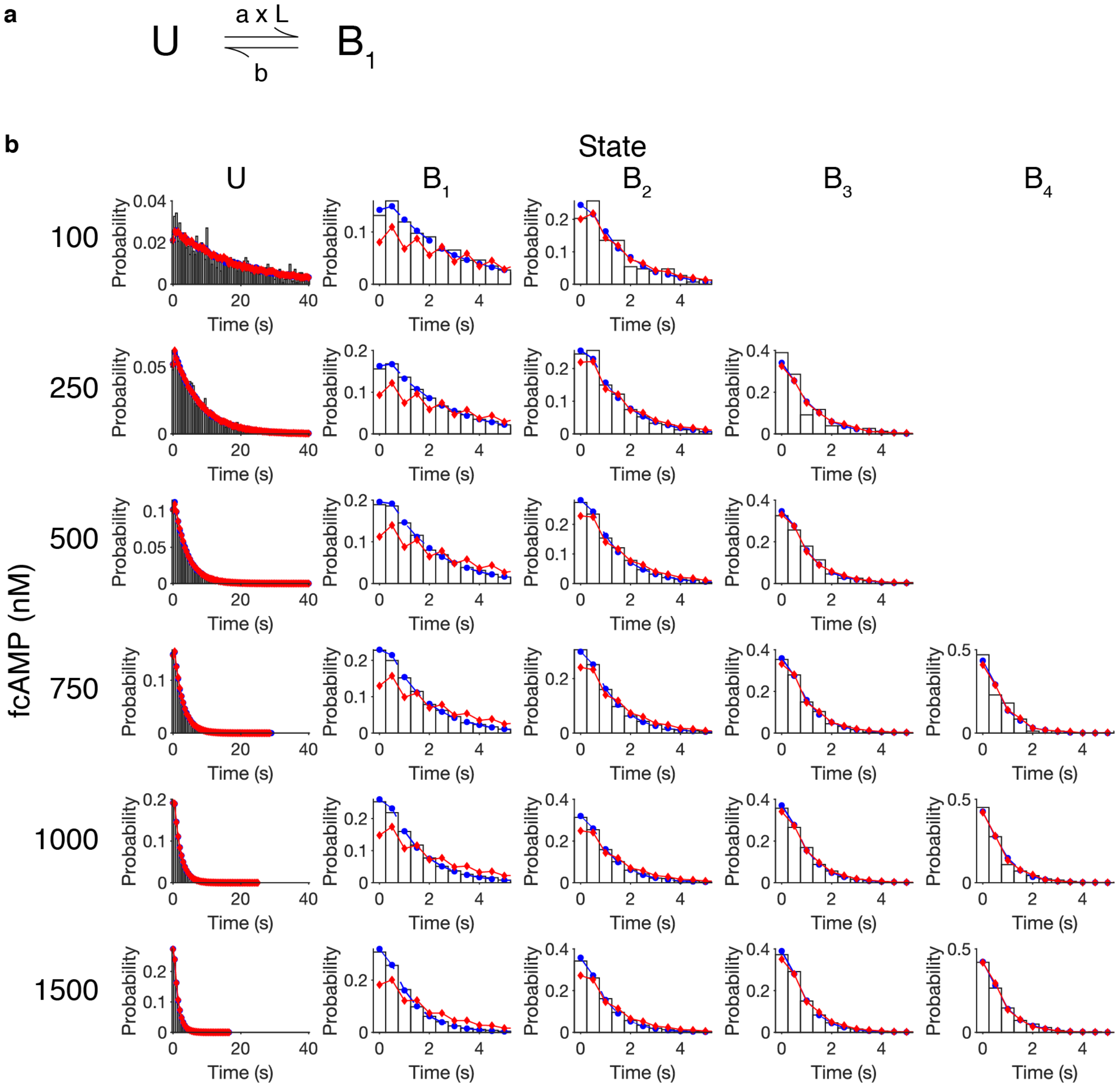

__Figure 4-S10: Simulation of monoexponential bound kinetics.__ (__a__) Model of an individual CNBD with monoexponential bound kinetics used to simulate a tetrameric complex of equal and independent CNBDs. (__b__) Dwell time probability distributions of all liganded states across various fcAMP concentrations overlaid with expectations from optimized rates of Model 1 (circle, blue) and Model 2 (diamond, square). Simulated and optimized rates are located in __Table 4-S4__.

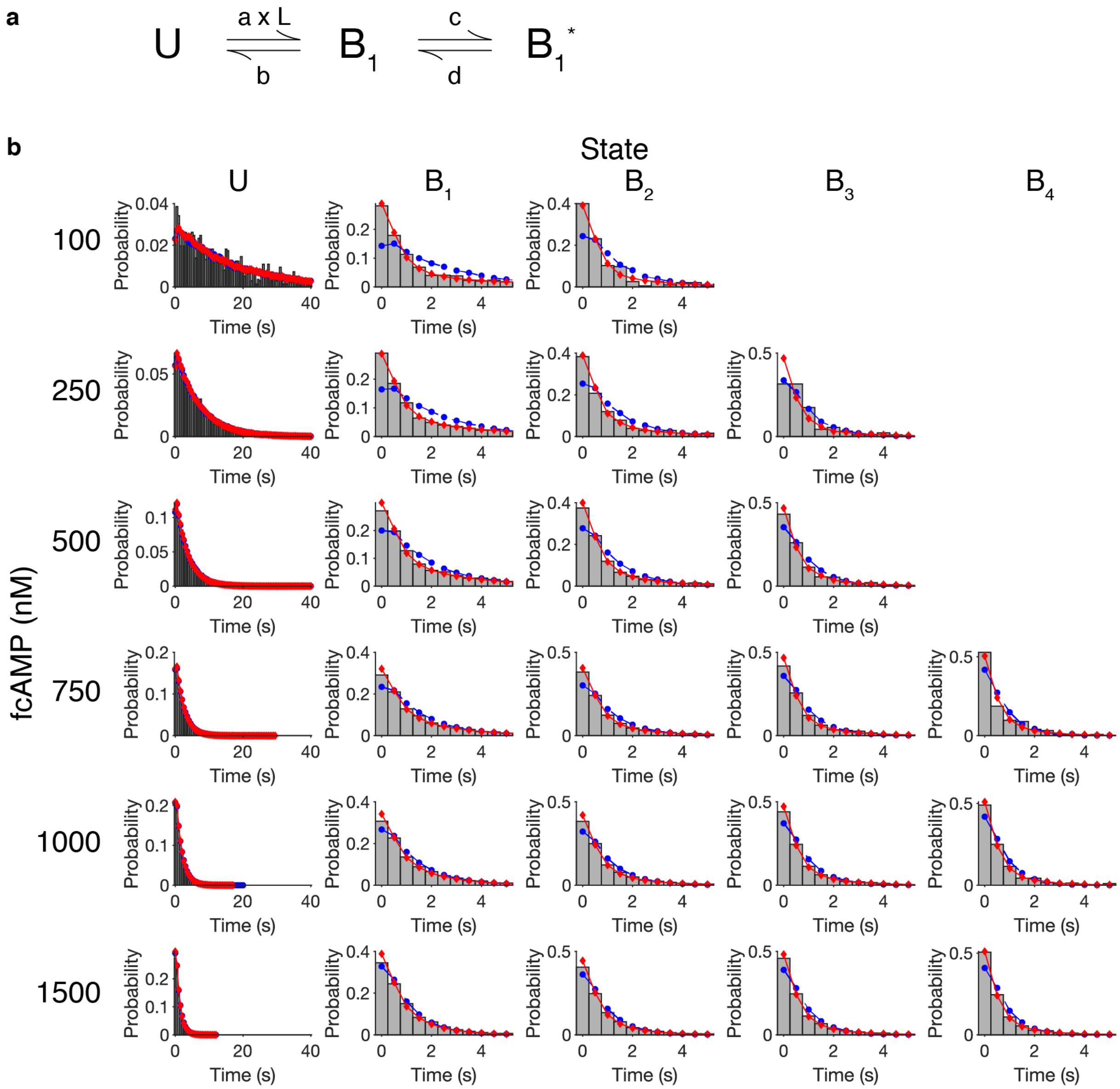

Figure 4-S11: Simulation of biexponential bound kinetics. *(**a**) Model of an individual CNBD with biexponential bound kinetics used to simulate a tetrameric complex of equal and independent CNBDs. (**b**) Dwell time probability distributions of all liganded states across various fcAMP concentrations overlaid with expectations from optimized rates of Model 1 (circle, blue) and Model 2 (diamond, square). Simulated and optimized rates are located in **Table 4-S4**.*

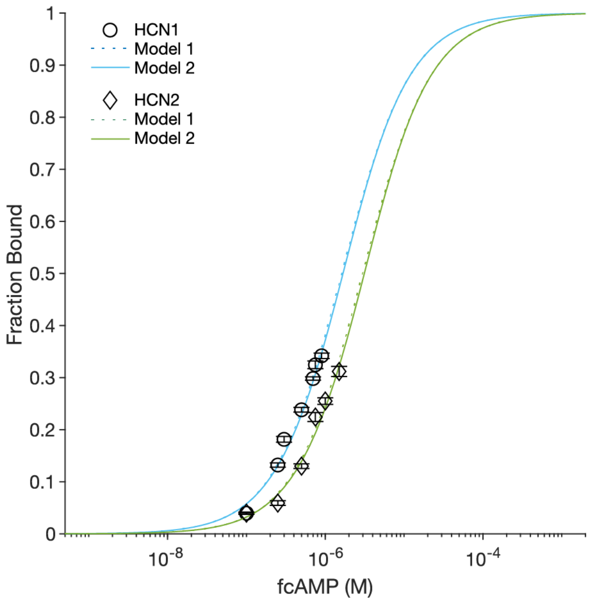

Figure 4-S12: Binding curve of HCN1SM and HCN2SM. *Fraction bound of HCN1SM (circles) and HCN2SM (diamonds) at various fcAMP concentrations overlaid with fits from Model 1 (dashed) and Model 2 (solid). See **Table 4-1**. Values are mean ± s.e.m.*

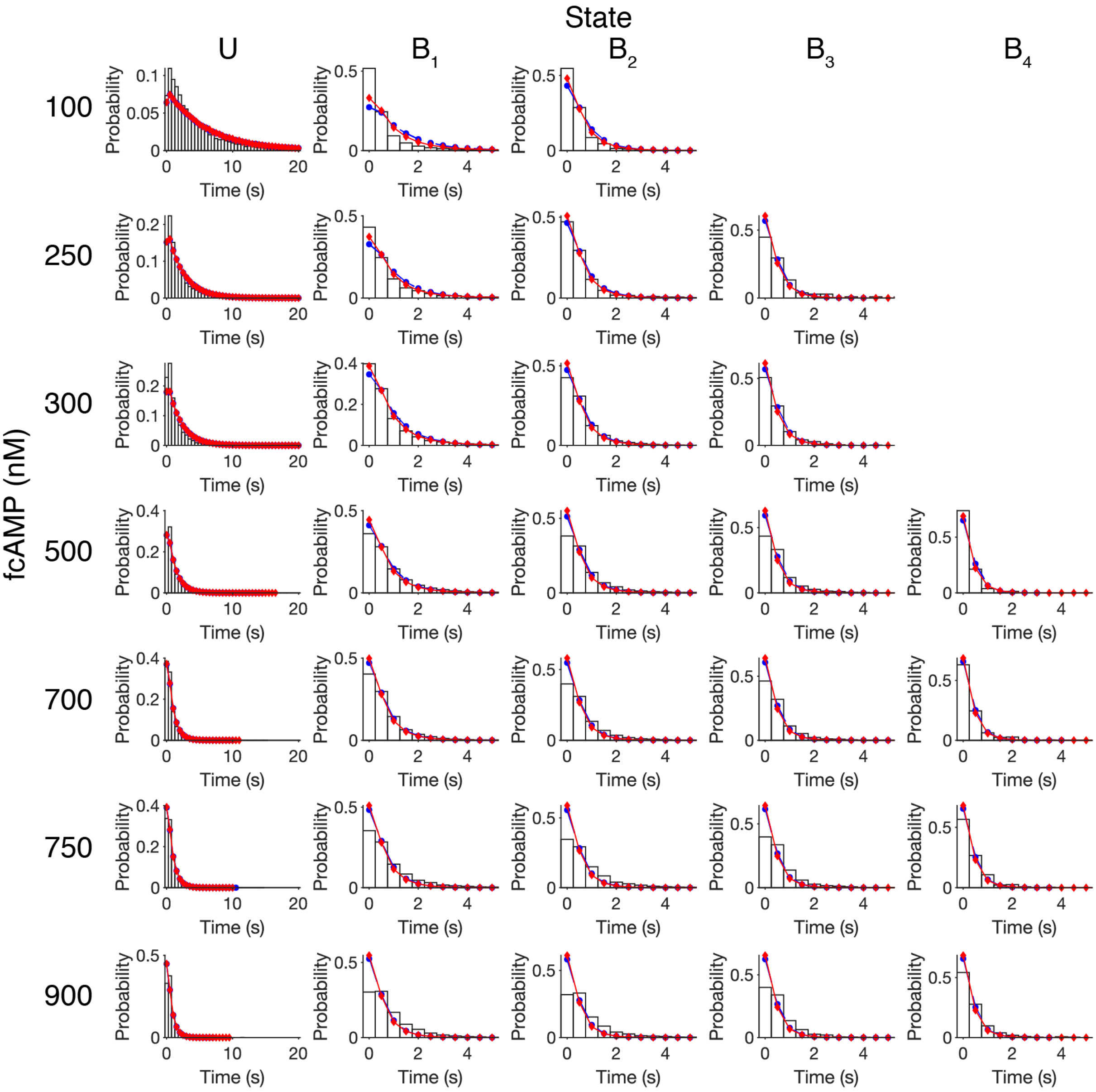

Figure 4-S13: Dwell time distributions of HCN1SM. *Dwell time probability distributions of all liganded states of HCN1SM across various fcAMP concentrations overlaid with expectations from optimized rates of Model 1 (circle, blue) and Model 2 (diamond, red). Optimized rates are located in* ***Table 4-1****.*

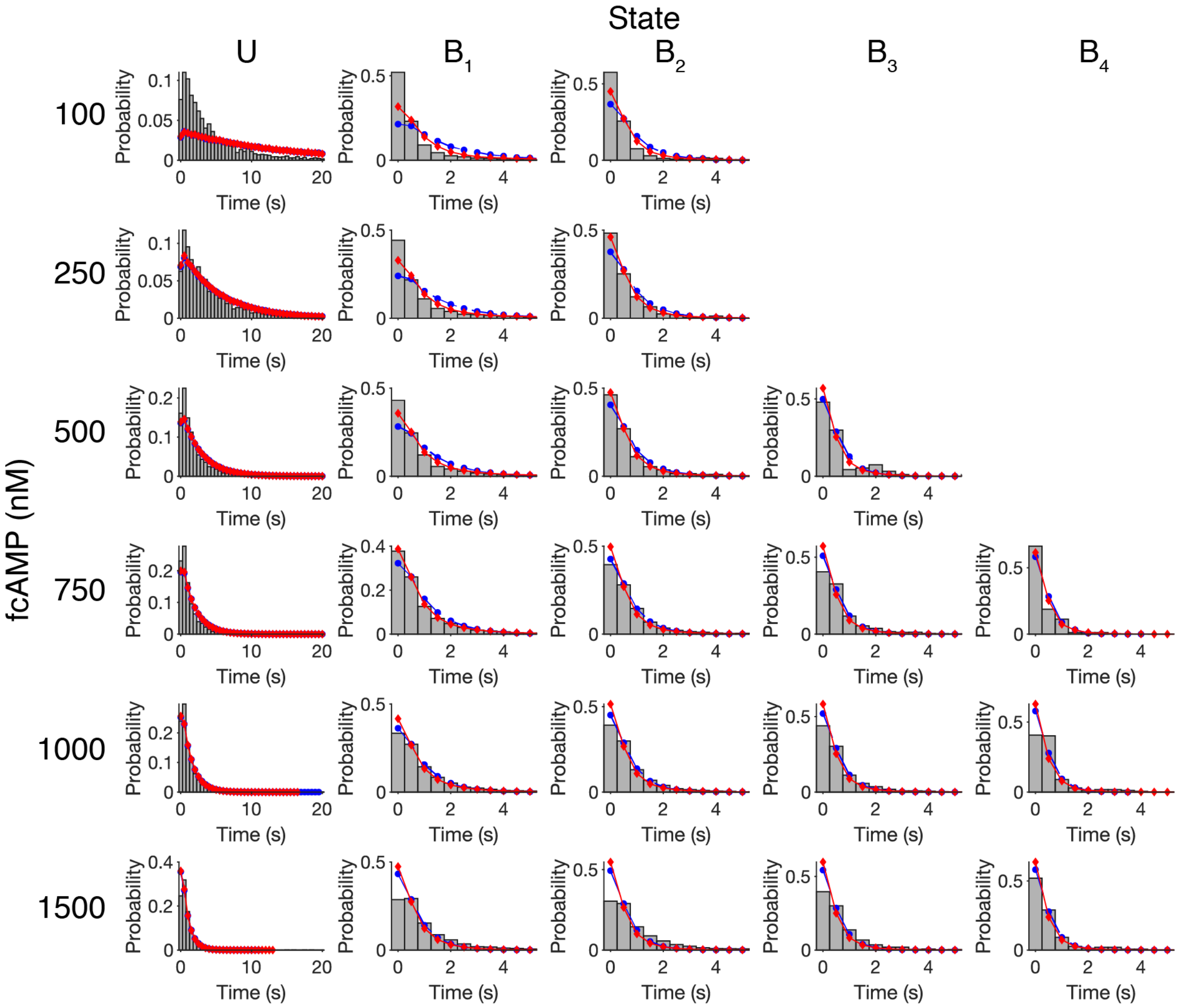

Figure 4-S14: Dwell time distributions of HCN2SM. *Dwell time probability distributions of all liganded states of HCN2SM across various fcAMP concentrations overlaid with expectations from optimized rates of Model 1 (circle, blue) and Model 2 (diamond, red). Optimized rates are located in* ***Table 4-1****.*

5 Conclusions and Future Directions

5.1 Conclusions

5.1.1 Biological insights of HCN gating

A central goal of this dissertation was to uncover the elementary mechanisms of cyclic nucleotide binding to HCN channels. Following a reductionist approach, we increased the complexity of our system with each subsequent experiment (**Figure 1-9**), beginning first with isolated and monomeric CNBDs (**Chapter 2**), progressing to a tetrameric complex of CNBDs without the transmembrane domains (**Chapter 3**), and concluding with two purified and intact HCN channels (HCN1 and HCN2) in detergent micelles (**Chapter 4**). Taken together, our results provide novel insights into the regulation of HCN channels by cyclic nucleotides at the single-molecule level.

At the level of isolated and monomeric CNBDs (**Chapter 2**), we discovered that fcGMP binding can occur through the same binding mechanism as fcAMP, wherein at least two meta-stable unbound conformations exist with one state receptive to ligand binding. Upon binding, the CNBD fluctuates between two possible bound conformations, similar to an induced fit scheme of ligand gated receptors[1]. The notion of a conformational change of the CNBD upon ligand binding is well supported in the field[2-5]. Our result is consistent with observations by DeBerg and others that showed cAMP and cGMP both enhance the activation kinetics of HCN2 and shift its voltage dependence to more depolarizing potentials[4]. Using electron paramagnetic resonance (EPR) and nuclear magnetic resonance (NMR) on isolated CNBDs, DeBerg further revealed the conformational changes of the CNBD following binding which differed across agonists. This result suggested that although the pathway may be different, each ligand allosterically modulates the pore in a similar manner. In our experiment, we find that fcAMP and fcGMP binding can occur through a similar four-step mechanism, albeit with different kinetic rates which underlie their different affinities. While these states are kinetically stable, our approach did not provide structural information about each state. As cAMP and cGMP have been shown to interact in the binding pocket in opposing conformations[4,6], it is likely that the final states structurally differ despite inducing a similar allosteric effect.

In our second study, we examined a tetrameric CNBD complex (**Chapter 3**) and observed non-cooperative CNBD interactions upon cAMP binding in direct contrast to the expectations of the field. For example, both at the level of HCN1 and HCN2 CNBDs, tetramerization was observed in the presence of cAMP and ligand affinity changed as a function of complex assembly[6,7]. However, our single-molecule measurements of a tetrameric complex of HCN2 C-terminal fragments did not provide any evidence of cooperativity either by state occupancy or by transition rates. A clear limitation of this study was the possibility that our artificial construct is tethered together in a manner that could prevent subunit interaction. Therefore, studies in the native architecture at the full-length channel were needed.

In our final study, we aimed to reveal the mechanism of cAMP binding detergent purified and full-length HCN. Following the scheme use for the HCN1 cryo-EM structure[8], we successfully isolated purified HCN1 and HCN2 molecules, albeit with a HCN1 N-terminus on HCN2 to account for the difficulty purifying the large G/C rich region. We again performed single-molecule binding experiments and resolved non-cooperativity binding to each isoform. While this result is perhaps not as surprising for HCN1 since it is likely in a more “primed” state[9], our result disagrees with observations of Kusch and Thon showing unique patterns of subunit cooperativity via patch clamp confocal fluorometry[10,11]. Thus, this advances our previous observation of independent CNBDs (**Chapter 3**) and further suggests CNBDs are not intrinsically cooperative even when in their native (closed) architecture. To determine the mechanism of ligand binding at each CNBD, we isolated singly occupied CNBD transitions (unbound→bound→unbound) and performed dwell time analysis. We found evidence of a “catch and hold” model for the bound well times of both HCN1 and HCN2 CNBDs data in agreement with our previous CNBD studies. This mechanism was further resolved at the full-length channel to extract the transitions rates of ligand association, dissociation, and conformational exchange across the full tetrameric protein. In conclusion, we saw that HCN1 has a higher (~2X) affinity for fcAMP than HCN2. The difference in affinity is primarily due to faster on rates between ligand binding and a ligand-induced conformational exchange that results in a lower transition probability into the in the "flipped" bound state compared to HCN2. This result highlights the prominent hypothesis in the field that conformational changes in the CNBD may allosterically propagate through the C-linker to modulate the pore[5]. In this case, our data and analyses allow us to speculate that the ratio of time spent in the second "flipped" bound state may underpin the magnitude of pore modulation.

5.1.2 Technological advances for single-molecule measurements in ZMWs

The power of ZMWs for exploring dynamic biology at physiological concentrations was well established at the start of this dissertation[,13]. Our contribution to the field was the combination of intermolecular smFRET inside ZMWs (ZMW-FRET) to extend the concentration barrier of single-molecule fluorescence up to 1 mM (**Chapter 2**). This result extends the max resolution of a ZMW (~10 µM [14]) by two-orders of magnitude to enable studies of many more of weak-affinity interactions (**Figure 1-5**, **Table 1-1**). To date, this is still the highest reported concentration of a single-molecule fluorescence measurement[15]. Compared to other high-concentration single-molecule techniques, ZMW-FRET is relatively easy to use since it takes advantage of the robust and reproducible smFRET toolbox and does not require sophisticated optics[16-18]. Of course, obtaining ZMWs will be limiting factor in this scheme, as fabrication can be challenging. However, if this experimentalist can do it, I trust anyone can (see **7.4.1**).

We additionally developed a new time-series idealization algorithm that adapts top-down machine learning called DISC (**Chapter 3**). The motivation for creating DISC was our observation that the standard analysis algorithms were imposing a computational bottleneck when processing the exceptionally large datasets now obtainable by high-throughput single-molecule approaches[19]. For example, smFRET experiments with an sCMOS camera or use of repurposed ZMW-based SMRT sequencers each enabled >10,000 molecules per field of view[20,21]. This bottleneck is particularly limiting when the underlying model of the data is not known *a prior* (as is the more common case in a single-molecule experiment) and thus statistical learning algorithms need to be applied. Our approach DISC is exceptionally fast owning to its use of top-down machine learning[22,23]. Compared to other common idealization approaches, we see DISC is 100-1000x faster with matched or improved accuracy (**Chapter 3**). We believe DISC will be a powerful platform for high-throughput and unsupervised analysis of single-molecule data.

5.2 Future directions

5.2.1 HCN in a lipid environment

Herein, we have demonstrated the power of single-molecule measurements in ZMWs to directly resolve mechanisms of cyclic nucleotide binding to HCN channels. Most recently, we have shown our approach is amenable to intact and purified HCN1/2 channels which opens the door for a variety of new experiments. For example, **in Chapter 4** we used detergents to mimic the lipid environment. While this is a promising start, detergents have many disadvantages as they can disrupt the native fold and function of the channel[24]. Therefore, a next step should be to approach a more native environment of a cell and monitor ligand binding to HCN channels contained in a lipid environment. There are a few possible avenues for accomplishing this which will be discussed below.

The first is the use of nanodiscs which require reconstituting detergent purified channels into a lipid environment encapsulated by a protein scaffold[25,26]. This method would allow the environment to be controlled such that the effect of different lipids could be assessed on channel function. However, the step first involves the solubilization in detergent which is a possible issue for retaining the native channel topology.

An alternative approach is the use of a so-called “native nanodisc” such as the styrene-maleic acid co-polymer (SMA) that uses an organic polymer to extract membrane proteins directly from a cell. This approach was demonstrated with the KcsA ion channel wherein SMA extracted proteins retained their intrinsic function when re-incorporated into a lipid bilayer[27]. SMA approaches have the advantage of retaining the native lipid environment for the cell which further bypasses the need of detergent solubilization. A limitation of the SMA approach is the difficulty controlling the size of the polymer during synthesis which creates nanodiscs of varying sizes which could promote to heterogenous protein behavior[28]. Additionally, large scale purification of membrane proteins using SMA is expensive due to the cost of reagents and the low solubilization efficiency compared to detergent.

A final approach is the use of cell-derived vesicles. Here, cells expressing a protein of interest can be fragmented and formed into membrane vesicles[29]. Like SMA, this bypasses detergent solubilization and retains the endogenous lipid environment of the cell. This technique has proven useful for single-molecule fluorescence experiments of ion channels, including stoichiometric evaluations of nicotinic receptor assembly[29,30] and potassium channel gating dynamics[31,32]. Unfortunately, the average size of these vesicles is ~180 nm which imposes a steric constraint for ZMW experiments[29].

5.2.2 Studying activated HCN channels

Another limitation of our previous experiments of intact HCN channels (**Chapter 3**) is the inability to apply a voltage to change the state of the channel. Therefore, the dynamics at 0 mV we monitored are to that of presumably closed HCN channels. As suggested by Benndorf, the state of the channel may play a role in cooperative interactions of the CNBDs[10,11]. This demonstrates the pressing need to resolve ligand binding to an open state of the channel. This will not be an easy task, but can be accomplished through biochemical modifications engineering a new experimental set-up.

Biochemically, mutations have been reported than promote the opening of the HCN channels even in the absence of hyperpolarizing potentials. As shown by Lee, HCN1 exhibits a closed state a 0 mV[8]. However, they were able to stabilize the open state of the channel by introducing a metal affinity bridge using two site-directed cysteines on the S4 helix and gating charge transfer center (S264C and F188C)[33]. Upon application of 100 µM $Cd^{2+,}$ the channel appeared open[33]. Purification of these mutants for single-molecule experiments would allow us to examine subunit cooperativity on open HCN channels.

On an experimental side, one could build a hybrid electrophysiology-single-molecule experiment. This sort of dual experiment is of interest to the ion channel community[34]; however, applications to HCN channels will be particularly difficult given their low conductance and weak-affinity for cyclic nucleotides[10,35]. If we continue to use ZMWs to monitor fcAMP binding, we can consider the avenues that approach this hybrid experiment. For example, neurons expressing fluorescently labeled nicotinic receptors have been grown on top of ZMW arrays that enabled photobleaching experiments inside the ZMWs[36]. While one could image patching this cell to control the membrane voltage; the issue still remains that the CNBDs are intracellular, and thus monitoring ligand binding becomes more complicated. It is further worth noting that the design of ZMWs is flexible. For example, work in the Bohn group developed an electrochemical ZMW device that features electrodes for monitoring redox reactions[13,37]. Given that voltage can be controlled across these devices, one could be inspired to suspend membranes inside a bespoke ZMW for voltage control. In this paradigm, solution flow could be handled through coupling to a microfluidic device[38]. Finally, it is possible to abandon ZMWs altogether and still achieve high concentration measurements in a more flexible environment. For example, photoactivation schemes (PhADE[39] and sm-PAFRET[40]) enable ligand binding dynamics on the order of 1-30 µM, sufficient for HCN studies (see **1.3.2**). As this can be performed on a cover glass, integration with electrophysiological tools will be easier. This scheme would require the synthesis of new cAMP derivatives featuring photoactivatable moieties which may affect their affinity[41]. Additionally, this paradigm will be limited in temporal resolutions owning to the photoactivation and diffusion steps needed to resolve binding compared to ZMWs. All the pieces for this new and exciting technology are here waiting for a motivated experimentalist to bring the next phase of this project to fruition.

5.3 References

1. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* 35, 539-546 (2010).
2. Puljung, M.C., DeBerg, H.A., Zagotta, W.N. & Stoll, S. Double electron-electron resonance reveals cAMP-induced conformational change in HCN channels. *Proceedings of the National Academy of Sciences of the United States of America* 111, 9816-9821 (2014).
3. Taraska, J.W., Puljung, M.C., Olivier, N.B., Flynn, G.E. & Zagotta, W.N. Mapping the structure and conformational movements of proteins with transition metal ion FRET. *Nature Methods* 6, 532-U94 (2009).
4. DeBerg, H.A., Brzovic, P.S., Flynn, G.E., Zagotta, W.N. & Stoll, S. Structure and Energetics of Allosteric Regulation of HCN2 Ion Channels by Cyclic Nucleotides. *Journal of Biological Chemistry* 291, 371-381 (2016).
5. James, Z.M. & Zagotta, W.N. Structural insights into the mechanisms of CNBD channel function. *Journal of General Physiology* 150, 225-244 (2018).
6. Zagotta, W.N. et al. Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* 425, 200-205 (2003).
7. Lolicato, M. et al. Tetramerization Dynamics of C-terminal Domain Underlies Isoform-specific cAMP Gating in Hyperpolarization-activated Cyclic Nucleotide-gated Channels. *Journal of Biological Chemistry* 286, 44811-44820 (2011).
8. Lee, C.H. & MacKinnon, R. Structures of the Human HCN1 Hyperpolarization-Activated Channel. *Cell* 168, 111-+ (2017).
9. Akimoto, M. et al. A Mechanism for the Auto-inhibition of Hyperpolarization-activated Cyclic Nucleotide-gated (HCN) Channel Opening and Its Relief by cAMP. *Journal of Biological Chemistry* 289, 22205-22220 (2014).
10. Kusch, J. et al. How subunits cooperate in cAMP-induced activation of homotetrameric HCN2 channels. *Nat Chem Biol* 8, 162-9 (2011).
11. Thon, S., Schulz, E., Kusch, J. & Benndorf, K. Conformational Flip of Nonactivated HCN2 Channel Subunits Evoked by Cyclic Nucleotides. *Biophys J* 109, 2268-76 (2015).
12. Zhu, P. & Craighead, H.G. Zero-Mode Waveguides for Single-Molecule Analysis. in *Annual Review of Biophysics, Vol 41*, Vol. 41 (ed. Rees, D.C.) 269-293 (2012).
13. Crouch, G.M., Han, D. & Bohn, P.W. Zero-mode waveguide nanophotonic structures for single molecule characterization. *Journal of Physics D-Applied Physics* 51(2018).
14. Levene, M.J. et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299, 682-686 (2003).
15. Peng, S.J., Wang, W.J. & Chen, C.L. Breaking the Concentration Barrier for Single-Molecule Fluorescence Measurements. *Chemistry-a European Journal* 24, 1002-1009 (2018).

16. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nature Methods* 5, 507-516 (2008).
17. Hellenkamp, B. et al. Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study. *Nature Methods* 15, 669-+ (2018).
18. Lerner, E. et al. Toward dynamic structural biology: Two decades of single-molecule Forster resonance energy transfer. *Science* 359, 288-+ (2018).
19. Hill, F.R., Monachino, E. & van Oijen, A.M. The more the merrier: high-throughput single-molecule techniques. *Biochemical Society Transactions* 45, 759-769 (2017).
20. Juette, M.F. et al. Single-molecule imaging of non-equilibrium molecular ensembles on the millisecond timescale. *Nature Methods* 13, 341-344 (2016).
21. Chen, J. et al. High-throughput platform for real-time monitoring of biological processes by multicolor single-molecule fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* 111, 664-669 (2014).
22. Steinbach, M., Karypis, G. & Kumar, V. A comparison of document clustering techniques. in *KDD workshop on text mining* Vol. 400 525-526 (Boston, 2000).
23. Pelleg, D. & Moore, A.W. X-means: Extending k-means with efficient estimation of the number of clusters. *Icml* 1, 727-734 (2000).
24. Bowie, J.U. Stabilizing membrane proteins. *Current Opinion in Structural Biology* 11, 397-402 (2001).
25. Civjan, N.R., Bayburt, T.H., Schuler, M.A. & Sligar, S.G. Direct solubilization of heterologously expressed membrane proteins by incorporation into nanoscale lipid bilayers. *Biotechniques* 35, 556-+ (2003).
26. Bayburt, T.H. & Sligar, S.G. Membrane protein assembly into Nanodiscs. *Febs Letters* 584, 1721-1727 (2010).
27. Dorr, J.M. et al. Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K+ channel: The power of native nanodiscs. *Proceedings of the National Academy of Sciences of the United States of America* 111, 18607-18612 (2014).
28. Lee, S.C. et al. A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nature Protocols* 11, 1149-1162 (2016).
29. Moonschi, F.H. et al. Cell-Derived Vesicles for Single-Molecule Imaging of Membrane Proteins. *Angewandte Chemie-International Edition* 54, 481-484 (2015).
30. Fu, X. et al. Brain Region Specific Single-Molecule Fluorescence Imaging. *Analytical Chemistry* 91, 10125-10131 (2019).
31. Wang, S.Z., Vafabakhsh, R., Borschell, W.F., Ha, T. & Nichols, C.G. Structural dynamics of potassium-channel gating revealed by single-molecule FRET. *Nature Structural & Molecular Biology* 23, 31-36 (2016).
32. Wang, S.Z. et al. Potassium channel selectivity filter dynamics revealed by single-molecule FRET. *Nature Chemical Biology* 15, 377-+ (2019).
33. Lee, C.H. & MacKinnon, R. Voltage Sensor Movements during Hyperpolarization in the HCN Channel. *Cell* 179, 1582-+ (2019).
34. Sasmal, D.K. & Lu, H.P. Single-Molecule Patch-Clamp FRET Microscopy Studies of NMDA Receptor Ion Channel Dynamics in Living Cells: Revealing the Multiple Conformational States Associated with a Channel at Its Electrical Off State. *Journal of the American Chemical Society* 136, 12998-13005 (2014).
35. Thon, S., Schmauder, R. & Benndorf, K. Elementary Functional Properties of Single HCN2 Channels. *Biophysical Journal* 105, 1581-1589 (2013).
36. Richards, C.I. et al. Live-Cell Imaging of Single Receptor Composition Using Zero-Mode Waveguide Nanostructures. *Nano Letters* 12, 3690-3694 (2012).

37. Han, D., Crouch, G.M., Fu, K., Zaino, L.P. & Bohn, P.W. Single-molecule spectroelectrochemical cross-correlation during redox cycling in recessed dual ring electrode zero-mode waveguides. *Chemical Science* 8, 5345-5355 (2017).
38. Martin, W.E. et al. Real-Time Sensing of Single-Ligand Delivery with Nanoaperture-Integrated Microfluidic Devices. *Acs Omega* 2, 3858-3867 (2017).
39. Loveland, A.B., Habuchi, S., Walter, J.C. & van Oijen, A.M. A general approach to break the concentration barrier in single-molecule imaging. *Nature Methods* 9, 987-+ (2012).
40. Peng, S.J., Sun, R.R., Wang, W.J. & Chen, C.L. Single-Molecule Photoactivation FRET: A General and Easy-To-Implement Approach To Break the Concentration Barrier. *Angewandte Chemie-International Edition* 56, 6882-6885 (2017).
41. Otte, M. et al. Hydrophobic alkyl chains substituted to the 8-position of cyclic nucleotides enhance activation of CNG and HCN channels by an intricate enthalpy - entropy compensation. *Scientific Reports* 8(2018).

6 Appendix-A Nanopositioning single-molecules into zero-mode waveguides with DNA origami

6.1 Project summary

Nanophotonic devices such as zero-mode waveguides (ZMWs) are powerful tools that enable the resolution of single-molecule dynamics at physiological concentrations. Although primed for high-throughput collection, the high probability of empty or multiply-occupied nanoapertures during typical single-molecule immobilization limits the collection efficiency of each array. In addition, random surface deposition inside each ZMW leads to heterogenous signals from interactions with the metallic cladding. Herein, we aim overcome these limitations by adapting a steric hindrance approach to deposit single-molecules directly into the center of ZMWs using DNA origami. Our method aims to circumvent the Poisson limited deposition process to increase the throughput of ZMWs while homogenizing both signal and dynamics.

This project was not completed during my graduate work in part due to the COVID-19 pandemic of 2020.

White, D.S.[1,2], Zhang, R.[1,2, †], Retterer, S.T.[3], Goldsmith, R.H.[2], Chanda, B.[2,4, †]

[1] Department of Neuroscience, University of Wisconsin-Madison, Madison, WI, USA
[2] Department of Chemistry, University of Wisconsin-Madison, Madison, WI, USA
[3] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[4] Department of Biomolecular Chemistry University of Wisconsin-Madison, Madison, WI, USA

[†] Present address:
Center for Investigation of Membrane Excitability Diseases, Department of Anesthesiology, Washington University School of Medicine, St. Louis, MO, USA

D.S.W., R.H.G., and B.C., conceived the project; D.S.W. and R.Z., collected data; D.S.W. fabricated ZMWs under supervision of S.T.R.; D.S.W. analyzed data and wrote this section with input from R.H.G and B.C.

6.2 Introduction

Single-molecule fluorescence microscopy can resolve details of molecular compositions and dynamics that are otherwise obscured in ensemble measures[1-3]. However, the inherent concentration barrier from the diffraction limit of focused light often leads to an experimental compromise of using low concentrations of fluorescence species which often prevents the application of single-molecule fluorescence in physiologically relevant environments[4]. Advances in nanotechnology have helped circumvent this limitation through the development of zero-mode waveguides (ZMWs) which make use of tiny holes in a 100 nm thick metal layer on a cover glass[5,6]. This architecture restricts the transmission of propagating light through the sub-wavelength sized nanoapertures, allowing only an exponentially decaying evanescent field close the metal-glass interface. The resulting zeptoliter-scale excitation volume enables the resolution of single fluorescence species up to micromolar concentrations and when combined with Förster resonance energy transfer (FRET), millimolar concentrations are achievable[5,7]. ZMWs have been used for a variety of applications, spanning DNA sequencing, ligand associations, and electrochemical studies[6,8-11].

Although a powerful tool for single-molecule resolution, the overall efficiency of data collection with ZMW device is limited by random deposition of molecules onto the surface. Typically, biotinylated biomolecules are immobilized onto streptavidin coated surfaces inside ZMWs[12]. Assuming all ZMWs are equal and the biomolecules do not interact in solution, the probability of a single-molecule being in a ZMW can be modeled as Poisson distribution with the probability density function

$$P(k) = \frac{\lambda^k e^{-k}}{k!} \qquad \text{Eq. 6-1}$$

where k is equal to the number of molecules and λ is the average occupancy. This limits the number of singly-occupied ZMWs to 37% with an equal percentage empty and the remaining 26% multiply-occupied. Further, given a random distribution of streptavidin molecule deposition on the surface, the location of single-molecules within the aperture is also heterogenous. As the proximity between dyes and metallic nanostructures can either enhance or quench fluorescence signals[13-16], the number of useable ZMWs is further reduced due variations in signal-to-noise ratios. In addition, one can easily imagine that the dynamics of a molecule may also vary as close proximity to the metallic walls could restrict conformational dynamics. Therefore, there is a pressing need for tools that can promote non-random and super-Poisson single-molecule deposition to the center of ZMWs to improve the collection efficiency and homogenize signal.

Previous work has demonstrated that the Poisson limit of single-molecule deposition can be overcome. For example, atomic force microscopy (AFM) has been used to directly place labeled DNA fragments directly into the center of ZMWs[17]. While this approach led to increased homogeneity of the florescence signal, the slow deposition rate of single-molecules via AFM renders this method incompatible with high-throughput experiments. Additional work has adapted a size-exclusion approach using large DNA molecules that are similar in size to the diameter of a ZMW which raises the single-occupancy rate to nearly 70%[18,19]. In particular, biotinylated DNA-Origami molecules allowed super-Poisson occupation and homogenized the fluorescence lifetime distribution of immobilized dyes[18]. Although a promising avenue, the current scheme is only valid for DNA-Origami molecules and cannot be used to deposit single biomolecules of interest to the centers of ZMWs. We further note recent work using a hybrid nanopore/ZMW device that uses voltage to reversibly position DNA into the center of ZMWs[20,21]. While this technology is powerful and exciting, the reliance on the negative charge of DNA fragments restricts the diversity of biological molecules compatible in this paradigm.

Herein, we advance previous work adapting DNA-Origami for single-molecule delivery into ZMWs. Our goal is to use a steric hinderance approach to deliver single-streptavidin molecules to the center of ZMWs, creating a functionalized surface primed for single-molecule deposition with a high rate of single-occupancy. The general scheme for the deposition of single-molecule into the centers of ZMWs is shown

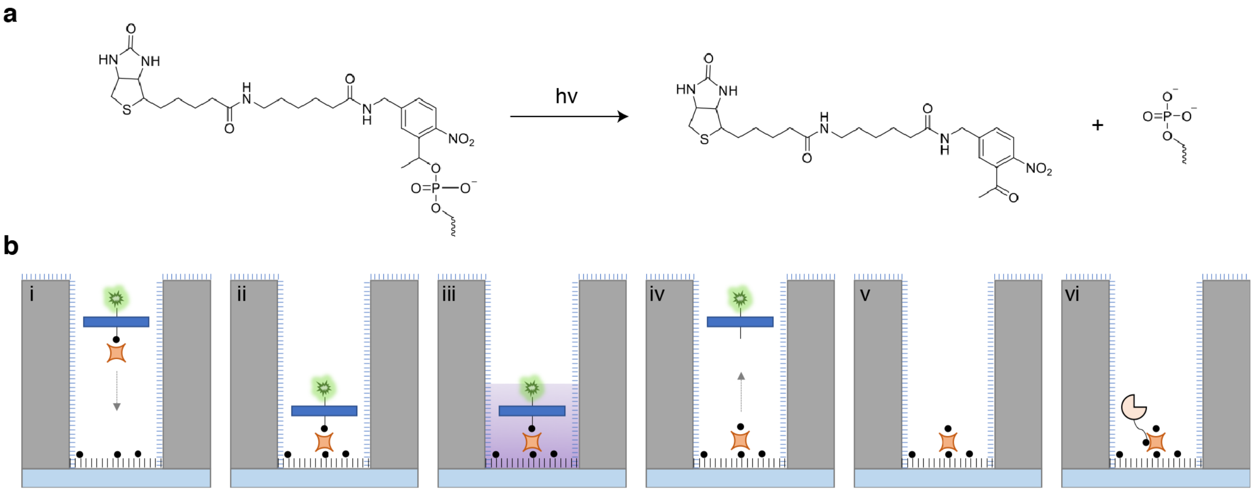

Figure 6-1: Single-molecule delivery to the center of a ZMW using DNA-Origami. *(**a**) Reaction of the photocleavable biotin moiety upon UV irradiation. (**b**) General scheme of DNA-Origami delivery into a ZMW. (I, II) A DNA-Origami bound streptavidin complex is added to a biotinylated ZMW surface. (III, IV) UV excitation cleaves the photocleavable moiety and the DNA-Origami is washed from the ZMW. (V) A single streptavidin is left in the center of the ZMW. (VI) A biotinylated molecule of interest can be immobilized on the ZMW surface by binding to the delivered streptavidin.*

in **Figure 6-1**. Our approach adapts a photocleavable biotin (pc-biotin) phosphoramidite moiety which is commonly used for streptavidin purification of oligonucleotides via irradiation with 300-350 nm light **Figure 6-1a**[22]. To deliver single-molecules into ZMWs, a flat sheet DNA-Origami featuring both a fluorophore label and a pc-biotin moiety in the center of each face will be used. The overall experimental workflow is highlighted in **Figure 6-1b**. First, isolated streptavidin-bound DNA-Origami complexes are immobilized into ZMWs featuring a functionalized PEG-biotin surface. Next, irradiation from a UV source cleaves the biotin from the DNA-Origami complex, leaving a single-streptavidin molecule bound to the center of the ZMW with two free biotin binding sites. Finally, a biotinylated molecule or interest (e.g. biotinylated protein or nanobody) is immobilized by binding streptavidin.

6.3 Results

6.3.1 Photocleavable biotin for single-molecule deposition

A DNA-Origami construct of 2.6 x 68 x 94 nm was designed for single-streptavidin delivery into ZMWs (see **6.5.1**). A Cy3 dye and pc-biotin moiety were introduced to opposite faces of the flat sheet for

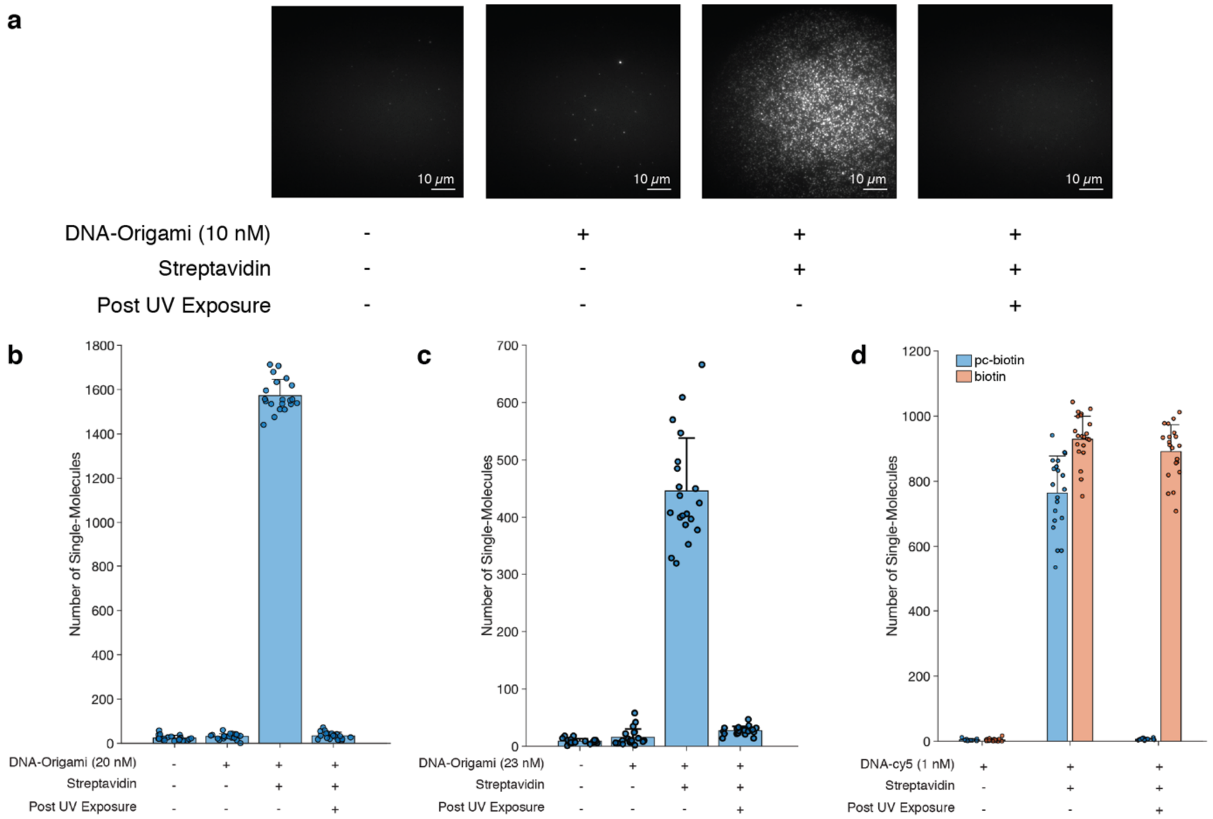

Figure 6-2: Photocleavage of DNA-Origami. *(**a**) Representative images showing the specific binding of the DNA-Origami complex to streptavidin coated surfaces at a high density. The number of single-molecules on the surface is reduced following exposure to UV light. Results of photolysis experiments of pc-biotin-DNA-Origami-Cy3 on a cover glass (**b**) and on a ZMW chip (**c**). (**d**) Photocleavage of pc-biotin-DNA-Cy5 (blue) and biotin-DNA-cy5 (orange) inside of a ZMW. Values are mean ± s.d. across N=20 locations on a chip per condition.*

visualization and immobilization (see **6.5.2**). The DNA-Origami construct was incubated at 10 nM for 10 minutes onto functionalized cover glasses featuring a PEG-biotin surface (**Figure 6-2a**). Notably, the DNA-Origami showed low non-specific absorption and only bound to the surface in the presence of streptavidin. After irradiation with 365 nm light for 30 minutes, the number of single-molecules per field of view returned to background levels suggesting that the DNA-Origami has been cleaved from the surface. By counting the average number of single-molecules per field of view before and after UV exposure across multiple locations (**Figure 6-2b**, see **6.5.3**), we determined these conditions yield a 99.8% photocleavage efficiency (see **6.5.4**).

Next, the experiment was repeated inside aluminum ZMWs featuring a diameter of approximately 160 nm. ZMWs were fabricated via positive-tone electron-beam lithography and functionalized with biotin-PEG and poly(vinylphosphonic acid) as described previously (see **6.5.2**)[5,12]. After incubation of the DNA-Origami molecules, the ZMW chip was exposed to 365 nm for 30 minutes in a bottom-up illumination scheme to ensure that the UV light successfully propagates into the ZMW (**Figure 6-2c**). This resulted in a 97.3% photocleavage efficiency. To ensure that the pc-biotin linker being cleaved is indeed responsible for the loss of Cy3 signal, as opposed to UV photobleaching the dye, short DNA fragments were synthesized with a 5’ Cy5 dye and either a 3’ biotin or pc-biotin (see **6.5.1**). Following immobilization of these DNA fragments into ZMWs, UV light successfully cleaved the fragment with the pc-biotin moiety (99.8% efficiency) but could not significantly reduce the number of observed biotinylated DNA molecules (4.1% efficiency) (**Figure 6-2d**). This result demonstrated that our 365 nm exposure conditions are not photobleaching the dyes and an immobilized DNA-Origami construct with a pc-biotin linker can be successfully removed from a streptavidin surface in ZMWs via UV irradiation (see **6.6.1**).

6.3.2 Purification of streptavidin-bound DNA Origami molecules

A streptavidin-bound DNA-Origami complex was generated and purified to deliver single-streptavidin molecules to the center of ZMWs (**Figure 6-2b**). A 1:100 molar ratio of DNA-Origami to streptavidin was mixed at 4°C for 60 minutes to ensure a high probability of one DNA-Origami molecule per tetrameric streptavidin molecule (see **6.5.5**). The reaction mixture was then dialyzed overnight to remove unbound streptavidin. To test if the streptavidin-bound DNA-Origami complex was functional, the complex was incubated on functionalized cover glasses in the absence of any externally added streptavidin. Importantly, the complex adhered to the surface with a high density and was successfully removed via 365 nm irradiation with a 93% photocleavage efficiency (**Figure 6-3**).

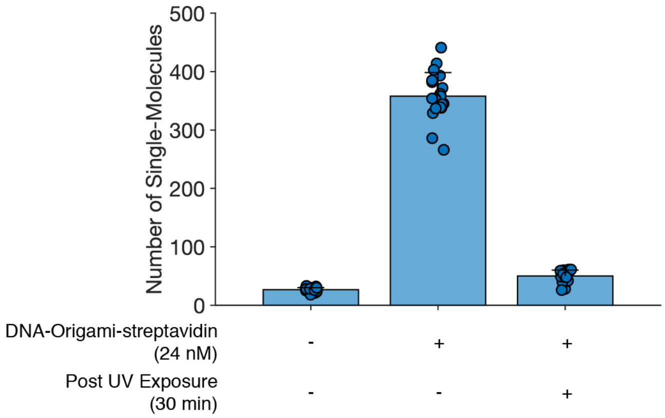

Figure 6-3: Photocleavage of a streptavidin-bound DNA-Origami complex inside a ZMW. Values are mean ± s.d. across N=20 fields of view per condition.

6.3.3 Pausing the project

Although the results have thus far been promising, the experiment shown in **Figure 6-3** was not reproducible after multiple attempts. In fact, both with streptavidin-bound and unbound DNA-Origami constructs, a large amount of non-specific absorption was observed upon repeated attempts. Presumably, this is due to sample degradation of the DNA-Origami complex as a result of storage conditions with EDTA (see **A.6.2**). While this is easy to test, the focus on other projects and the time spent unable to enter the lab due to the Covid-19 quarantine has left this project uncompleted. Next steps and considerations to complete the project are outlined in section **6.6.**

6.4 Discussion

Zero-mode waveguides are powerful tools for advancing single-molecule experiments to physiological concentrations with high-throughput capabilities. Herein, we assert that the Poisson limit of single-molecule deposition can be overcome using a steric hinderance approach. Although uncompleted, this work demonstrates a potential scheme for delivering single-streptavidin molecules to the center of ZMWs for singly-occupied ZMWs. We have demonstrated that a photocleavable biotin linker can be used for DNA-Origami deposition and removal inside of ZMWs, creating a potential avenue for single-molecule delivery. Critical next steps in this work include the validation of streptavidin delivery into ZMWs, as measured by specific binding of biotinylated fluorescent species, and assessments of both occupancy and signal to noise ratio per ZMW. If successful, this work will pave the way for homogenizing the signal of single-molecule dynamics while expanding the high-throughput capabilities of these powerful nanophotonic tools.

6.5 Methods

6.5.1 DNA fragments and origami synthesis

Custom DNA fragments featuring both a biotin linker and Cy5 dye were purchased from Integrated DNA Technologies (IDT) as 5PCBio/AGGACTTGT/3Cy5Sp/ and /5Cy5/AGCACTTGT/3BioTEG/. The 9 base pair sequences were adapted from previous single-molecule study[23]. The DNA-Origami was custom designed by Tilibit Nanosystems. The Cy3 dye and pc-biotin moieties were added to opposite faces of a commercially available 2.6 x 68 x 94 nm template origami (PF-6 flat sheet). Following assembly, the DNA-Origami complex was PEG purified and the architecture was confirmed via electrophoresis and transmission electron microscopy. The DNA-Origami was provided as a 550 µL solution at 240 nM in a buffer of 5 mM Tris, 5 mM NaCl, 1 mM EDTA, 5 mM Mg_2Cl_2. The sample was aliquoted and stored at -20°C. Thawed samples were kept in 4°C and discarded after five days to ensure sample stability.

6.5.2 Substrate preparation and passivation for TIRFM experiments

Cover glass and ZMW preparation for single-molecule experiments were performed as previously described (see **4.5.3** and **7.3**). Briefly, cover glasses were cleaned via sequential sonication in detergent

(Hellma), ethanol, and 1 M KOH for 30 minutes then functionalized overnight 1 mg/mL biotin-PEG-silane (MW = 3400 g mol^{-1}) and 10 mg/mL mPEG-silane (MW = 2000 g mol^{-1}) (Laysan Bio Inc.) dissolved in ethanol with 5% glacial acetic acid. ZMWs were fabricated at Oak Ridge National Labs in the Center for Nanophase Materials Sciences. Cover glasses were SC-1 cleaned, coated with 100 nm of thermally evaporate aluminum, spin-coated with ZEP520A (ZEONREX Electronic Chemicals) at 2000 RPM for 45 seconds, baked at 180°C for 2 minutes, and patterned using electron-beam lithography (JEOL JBX-9300FS)[5]. Substrates were then developed in xylenes for 30 seconds and dry etched with a mixture of 30 standard cubic centimeters (sccm) chlorine (Cl_2) and 10 sscm boron trichloride (BCl_3) gasses at 50°C for 60 seconds (Oxford Plasmalab System 100 Reactive Ion Etcher). Finally, substrates were plasma cleaned and scanning electron microscopy (SEM) was used to characterize the diameter of ZMWs. Prior to experiments, ZMWs were passivated by incubation in 2% poly(vinylphonic acid) (PVPA) (Polysciences) for 3 minutes at 90°C followed by overnight incubation in 1 mg/mL biotin-PEG-silane (MW = 3400 g mol^{-1}) and 10 mg/mL mPEG-silane (MW = 2000 g mol^{-1}) in ethanol with 5% glacial acetic acid[12]. Both cover glasses and ZMWs were further incubated with 10 mg/mL bovine serum albumin (BSA) in PBS (pH = 7.4) for 30 minutes.

6.5.3 Single-molecule imaging and analysis

Single-molecule fluorescence imaging and analysis was performed in a similar manner to previously described methods (see **4.5.2**). Briefly, samples were imaged using total internal fluorescence microscopy (TIRFM) on an inverted microscope (Olympus, IX-71) with a 512x512 EMCCD (Andor iXon Ultra X-9899) under 532 nm or 640 nm laser excitation (Coherant). Fluorescence emission was passed through a through a multiband dichroic and filter cube for imaging Cy3/Cy5 (Semrock Brightline Cy3/Cy5-A-OMF). All imaging was controlled with Metamorph software (Molecular Devices) and images were collected at a 10 frames per second. The locations of single-molecules in image stacks were identified using a threshold approach. For each image stack, a binary mask was created by averaging the first 10 images, removing background with a top-hat filter, and applying a uniform hard-threshold value (see **6.6.4**). Identified spots of at least 2x2-pixel size and 5-pixel separation between neighboring spots were counted as single-molecules. Image processing code can be found at: https://github.com/David-Scott-White/mZAP.

6.5.4 Photolysis of pc-biotin with UV

Cover glasses and ZMW chips were set on-top of a UV flashlight (365 nm, LIGHTFE) and excited for 30 minutes to ensure photocleavage. Importantly, illumination was performed in a bottom-up manner to ensure light enters through bottom of the ZMW and interacts with molecules bound to the surface. Following excitation, the substrates were rinsed to remove any unbound or cleaved molecules. The photocleavage efficiency was computed as the percent difference between the mean number of spots per field of view before and after UV excitation.

$$Photocleavage\ Efficiency = \frac{(\mu_{preUV} - \mu_{bkg}) - (\mu_{postUV} - \mu_{bkg})}{(\mu_{preUV} - \mu_{bkg})} x\ 100\% \qquad \text{Eq. 6-2}$$

6.5.5 Generation of a streptavidin-bound DNA-Origami complex.

A 1:100 molar ratio of DNA-Origami to streptavidin (2.4 nM to 240 nM) was incubated in 1 mL of 5 mM Tris HCl (pH = 8), 5 mM NaCl, 5 mM $MgCl_2$ and 1 mM EDTA for 30 minutes at 4°C. The solution was then subjected to a 1:2000x dialysis overnight in the same buffer at 4°C using a 100 kD cut-off dialysis tube (Millipore Sigma) in a 2 L volume. The dialyzed solution was stored in 4°C and never frozen.

6.6 Future directions and considerations

6.6.1 Improved control experiment

Although we demonstrated that 365 nm exposure was not photobleaching Cy5 dyes, this result does not necessarily translate to Cy3. Therefore, I obtained /5Cy3/AGCACTTGT/3BioTEG/ fragment as a control. Only one experiment was performed using this fragment, which overall showed that Cy3 was not photobleaching as a result of 365 nm exposure on functionalized cover glasses (data not shown). However, due to the high density of the labeling and the lack of a positive control, this control experiment needs to be repeated to quantitatively conclude that photobleaching of Cy3 is not occurring.

6.6.2 Potential problems and possible solutions

The work herein showcases that the pc-biotin moiety can be used to deliver molecules into ZMWs that can then be removed with UV excitation. If this project is continued, I would suggest a few alterations to the experimental design to circumvent challenges that are likely to arise. First, it would be helpful to change the Cy3 dye on the DNA-Origami complex to a more photostable and bright organic dye (e.g. ATTO, Alexa, or DyLight)[ix]. Although the Cy3/Cy5 dyes are fantastic for smFRET[24,25], better dyes can be used to obtain a more reliable signal to identify signal molecules above the background commonly found on cover glasses in non-FRET paradigms. Next, refine the storage conditions for the DNA-Origami complex. While the sample we purchased was provided in EDTA, some reports indicate that long term storage in EDTA can disrupt the origami structure by chelating Mg^{2+} ions[26]. While I do not have direct evidence that this occurred with our substrate, it is my hypothesis that an unstructured DNA origami led to the high prevalence of non-specific binding to the surface following dialysis. Adding an extra experimental step to confirm the integrity of the origami architecture by transmission electron miscopy or similar techniques will mitigate this concern. Finally, consider looking into trivalent streptavidin complexes as an extra precaution for single occupancy of ZMWs[27]. This mutant would allow a single streptavidin to bind a biotin on the surface, the biotin from pc-biotin, and a single biotin from the molecule of interest. This scheme

---

[ix] Admittedly, choosing Cy3 was one of the largest oversights of my thesis work. Tilibit Nanosystems should be able to incorporate any dyes available through IDT to correct my blunder.

alleviates the concern that the fourth and final active binding site on streptavidin can bind a second molecule of interest which would lead to double occupancy.

6.6.3 Achieving even higher collection efficiency

Our typical single-molecule set-up used a 512x512 EMCCD (Andor iXon Ultra X-9899) with a 100x oil immersion objective (Olympus, 1.49 NA) which results in an imaging area of approximately 80 µm x 80 µm. When imaging ZMW arrays with a 2 µm spacing between ZMWs, there are about 1,600 ZMWs per field of view. Although this is already a large number, this number can be dramatically increased by expanding the physical imaging area. For example, using a 60x oil immersion objective (Olympus UPAPO60XOHR, NA = 1.5) with the same 512x512 EMCCD leads to 4,500 ZMWs per field of view. In addition, switching the detector from an EMCCD to an sCMOS would significantly increase the physical imaging space. When coupled with a 2048x2048 sCMOS (e.g. Photometrics), a 100x objective could fit 39,000 ZMWs per field of view, and a 60x fit over 62,000 ZMWs per field of view. This result would surpass the previous reports of high-throughput collection with sCMOS for non-ZMW smFRET experiments and nearly match commercially available ZMW-based sequencing machines[28,29]. Of course, an important limitation of increasing the field of view dramatically is the lower resolution per single-molecule from the reduction in the number of pixels per ZMW. However, the possibility of increasing single-molecule throughput by nearly 3 to 40-fold makes the straightforward work of optimizing the best objective and detector combo a worthwhile endeavor.

6.6.4 Reducing bias in image analysis

The experiments herein rely on counting the number of single-molecules per image to quantify the efficiency of photocleavage and deposition. Regions of interest in this analysis were determined by applying a uniform hard-threshold across all images for a given experiment. Unfortunately, this is an inherently biased analysis when the experimentalist is both collecting data and performing image analysis. Previously, we have automatically counted from image masks using Otsu's method to minimize user bias (see **4.5.5**); however, this method is not perfect and often has trouble distinguishing signal from noise when in the signal is sparse in images (i.e. few molecules). Therefore, to reduce image analysis bias, this step should be blinded. This can be accomplished by having one experimentalist collect data then randomize and rename images for a second experimentalist to perform all image analysis. Once completed, images can be returned to their original file names and statistical analysis can be performed.

6.7 References

1. Ha, T. Single-molecule methods leap ahead. *Nature Methods* **11**, 1015-1018 (2014).
2. Moerner, W.E. & Fromm, D.P. Methods of single-molecule fluorescence spectroscopy and microscopy. *Review of Scientific Instruments* **74**, 3597-3619 (2003).
3. Moerner, W.E., Shechtman, Y. & Wang, Q. Single-molecule spectroscopy and imaging over the decades. *Faraday Discussions* **184**, 9-36 (2015).
4. Holzmeister, P., Acuna, G.P., Grohmann, D. & Tinnefeld, P. Breaking the concentration limit of optical single-molecule detection. *Chemical Society Reviews* **43**, 1014-1028 (2014).
5. Levene, M.J. et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686 (2003).
6. Zhu, P. & Craighead, H.G. Zero-Mode Waveguides for Single-Molecule Analysis. in *Annual Review of Biophysics, Vol 41*, Vol. 41 (ed. Rees, D.C.) 269-293 (2012).
7. Goldschen-Ohm, M.P., White, D.S., Klenchin, V.A., Chanda, B. & Goldsmith, R.H. Observing Single-Molecule Dynamics at Millimolar Concentrations. *Angewandte Chemie-International Edition* **56**, 2399-2402 (2017).
8. Eid, J. et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**, 133-138 (2009).
9. Goldschen-Ohm, M.P. et al. Structure and dynamics underlying elementary ligand binding events in human pacemaking channels. *Elife* **5**(2016).
10. Han, D., Crouch, G.M., Fu, K., Zaino, L.P. & Bohn, P.W. Single-molecule spectroelectrochemical cross-correlation during redox cycling in recessed dual ring electrode zero-mode waveguides. *Chemical Science* **8**, 5345-5355 (2017).
11. Crouch, G.M., Han, D. & Bohn, P.W. Zero-mode waveguide nanophotonic structures for single molecule characterization. *Journal of Physics D-Applied Physics* **51**(2018).
12. Korlach, J. et al. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1176-1181 (2008).
13. Kinkhabwala, A. et al. Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. *Nature Photonics* **3**, 654-657 (2009).

14. Punj, D. et al. A plasmonic 'antenna-in-box' platform for enhanced single-molecule analysis at micromolar concentrations. *Nature Nanotechnology* **8**, 512-516 (2013).
15. Acuna, G.P. et al. Fluorescence Enhancement at Docking Sites of DNA-Directed Self-Assembled Nanoantennas. *Science* **338**, 506-510 (2012).
16. Holzmeister, P. et al. Quantum yield and excitation rate of single molecules close to metallic nanostructures. *Nature Communications* **5**(2014).
17. Heucke, S.F. et al. Placing Individual Molecules in the Center of Nanoapertures. *Nano Letters* **14**, 391-395 (2014).
18. Pibiri, E., Holzmeister, P., Lalkens, B., Acuna, G.P. & Tinnefeld, P. Single-Molecule Positioning in Zeromode Waveguides by DNA Origami Nanoadapters. *Nano Letters* **14**, 3499-3503 (2014).
19. Plenat, T., Yoshizawa, S. & Fourmy, D. DNA-Guided Delivery of Single Molecules into Zero-Mode Waveguides. *Acs Applied Materials & Interfaces* **9**, 30561-30566 (2017).
20. Larkin, J., Foquet, M., Turner, S.W., Korlach, J. & Wanunu, M. Reversible Positioning of Single Molecules inside Zero-Mode Waveguides. *Nano Letters* **14**, 6023-6029 (2014).
21. Larkin, J., Henley, R.Y., Jadhav, V., Korlach, J. & Wanunu, M. Length-independent DNA packing into nanopore zero-mode waveguides for low-input DNA sequencing. *Nature Nanotechnology* **12**, 1169-1175 (2017).
22. Olejnik, J., KrzymanskaOlejnik, E. & Rothschild, K.J. Photocleavable biotin phosphoramidite for 5'-end-labeling, affinity purification and phosphorylation of synthetic oligonucleotides. *Nucleic Acids Research* **24**, 361-366 (1996).
23. Cisse, I.I., Kim, H. & Ha, T. A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nature Structural & Molecular Biology* **19**, 623-627 (2012).
24. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nature Methods* **5**, 507-516 (2008).
25. Di Fiori, N. & Meller, A. The Effect of Dye-Dye Interactions on the Spatial Resolution of Single-Molecule FRET Measurements in Nucleic Acids. *Biophysical Journal* **98**, 2265-2272 (2010).
26. Kielar, C. et al. On the Stability of DNA Origami Nanostructures in Low-Magnesium Buffers. *Angewandte Chemie-International Edition* **57**, 9470-9474 (2018).
27. Dubacheva, G.V. et al. Controlling Multivalent Binding through Surface Chemistry: Model Study on Streptavidin. *Journal of the American Chemical Society* **139**, 4157-4167 (2017).
28. Juette, M.F. et al. Single-molecule imaging of non-equilibrium molecular ensembles on the millisecond timescale. *Nature Methods* **13**, 341-344 (2016).
29. Chen, J. et al. High-throughput platform for real-time monitoring of biological processes by multicolor single-molecule fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 664-669 (2014).

7 Appendix B: Sample Preparation

7.1 Key reagents for single-molecule experiments

Table 7-1: Key single-molecule reagents

Name	Source	Catalog Number	Notes
I. Sample			
DWK Life Sciences Wheaton™ Screw Caps for 900570 Slide Staining Jar	Fisher Scientific	02-912-618	Or similar product
Fisherbrand™ Premium Cover Glasses	Fisher Scientific	12-548C	25x25 mm
Grace Bio-labs SecureSeal™ hybridization chambers	Millipore-Sigma	GBL621505	well diam. × depth 9 mm × 0.8 mm, ports diam. 1.5 mm
Wash-N-Dry™ coverslip rack	Millipore-Sigma	Z688568	
II. Cleaning			
Bath sonicator			e.g. Branson
Ethanol	Millipore-Sigma	64-17-5	HPLC Grade
Hellmanex™ III	Hellma	9-307-011-4-507	
Potassium Hydroxide (KOH)			1 M
III. Passivation			
biotin-PEG-silane	Laysan Bio Inc.	Biotin-PEG-SIL-3400	MW=3400 g mol^{-1}
mPEG-silane	Laysan Bio Inc.	mPEG-Silane-2000	MW=2000 g mol^{-1}
poly(vinylphonic acid) (30% soln)	Polysciences	24297-10	PVPA
Snapware Glass Medium Square Container	Target		Or any similar product
IV. Sample Pull-down			
Bovine Serum Albumin	Millipore-Sigma	9048-46-8	BSA
GFP VHH, biotinylated	Chromotek	gtb-250	GFP-TRAP
Streptavidin	Prospec	PRO-791	
V. Oxygen Scavenging System			
(±)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid	Millipore-Sigma	53188-07-1	Trolox
3,4-Dihydroxybenzoic acid	Millipore-Sigma	99-50-3	PCA
Protocatechuate 3,4-Dioxygenase from Pseudomonas sp.	Millipore-Sigma	9029-47-4	PCD

7.2 General reagent stock preparation

7.2.1 BSA

This protocol is adapted from Cold Spring Harbor[x].

1. Add 1 g BSA powder per 10 mL H_2O (or PBS pH 7.4) in 50 mL skirted Falcon tube.
2. Allow BSA to sit on the water at 4°C for a couple hours/ overnight to gently dissolve.
    a. Note: Do not shake, stir, or vortex the mixture. This will form lots of bubbles which obstruct single-molecule imaging.
3. Aliquot and store in -20°C.

7.2.2 PCD

This protocol is adapted from *Aitken et al., Biophys J, (2008)*[1].

1. Dissolve PCD to a concentration of 1 unit per 25 µL in:
    a. 50% Glycerol, Tris-HCl (pH = 8), 1 mM EDTA, 50 mM KCl.
2. Aliquot into 25 µL volumes (e.g. in strip tubes).
3. Flash freeze using liquid nitrogen. Store in -80°C.
4. When ready to for use, add 25 µL PCD stock to 975 µL imaging buffer containing 2.5 mM PCA (**see 7.2.3**)
    a. Note 1: make sure imaging buffer contains both Trolox and PCA.
    b. Note 2: I typically replaced PCD solutions every 30-60 minutes to ensure enzyme activity.

7.2.3 PCA

Stocks of PCA are not needed since they will degrade. Instead, a solution is prepared for each experiment.

1. Add 15 mg of PCA into a solution of:
    a. 150 µL 1 M KOH to 850 µL H_20 (Milli-Q preferred).
2. Vortex to dissolve.
    a. Note: This will take a minute or two. Solution will be slightly yellow.
3. Add 250 µL PCA solution for ever 10 mL of buffer (2.5 mM solution).

---

[x] http://cshprotocols.cshlp.org/content/2006/1/pdb.rec6.full?text_only=true

7.2.4 Trolox

Trolox is a finicky reagent to work with- it does not like going into solution and can both gain and lose potency based on how long it has been in solution. For a detailed guide to using Trolox, see *Cordes et al., JACS (2009)*[2]. The Supporting Information of that article outlines working with Trolox to maximize its effect. I devised a two-pronged approach merging different literature protocols that seemed to work[xi].

1. Make stock Trolox (100 mM)
    a. Dissolve 2.5 mg per 10 mL in DMSO. Vortex
    b. Aliquot into 100 µL volumes and store at -20°C
2. Trolox in stock imaging buffer (e.g. 500 mL PBS)
    a. To ~1X stock buffer intended for single-molecule imaging, add ~2.5 mg Trolox per 1000 mL volume.
    b. Stir to dissolve with gentle heating (around 40-50°C). Without heat, stir overnight.
    c. Filter with 0.22 µm filter.
    d. Ensure buffer is sealed tightly with a screw cap. Store away from light at 4°C.
        i. Note: Wrap aluminum foil around bottle to prevent light from entering.
        ii. Note: Solution will turn yellow when Trolox has degraded (1-2 weeks)
3. On day of experiment:
    a. Add 100 µL 1 mM stock Trolox (-20°C, DMSO) per 10 mL stock buffer (+ 1 mM Trolox)
    b. Vortex vigorously to mix DMSO into water
    c. Effective Trolox concentration ~ 2 mM.

7.3 TIRFM experiments

7.3.1 Cover glass cleaning

1. Add cover-glasses (~9) to slide staining jar.
2. Cover the samples with 1% Hellmanex™ III. Sonicate in bath sonicator for 30 minutes.
3. Rinse cover glasses (inside staining jar) with water 10 times.
4. Cover the samples with HPLC Grade Ethanol Sonicate in bath sonicator for 30 minutes.
5. Rinse cover glasses (inside staining jar) with water 10 times.
6. Cover the samples with 1 M KOH. Sonicate in bath sonicator for 30 minutes.
7. Rinse cover glasses (inside staining jar) with water 10 times[xii].
8. Dry with Argon[xiii]
9. Plasma clean for 5 minutes

[xi] While this method provides adequate results, I am sure it can be optimized for much better results.

[xii] Most cleaning protocols will show KOH first, then Ethanol. However, no matter how clean Ethanol claims to be (e.g. spectroscopic or HPLC grade, etc...) it is never actually clean, especially in the 488 nm / 532 nm range. KOH slightly etches the glass and works better to ensure Ethanol dirt is removed from the sample prior to plasma cleaning.

[xiii] Nitrogen will also work. The Chanda lab just always had Argon around so I used that.

7.3.2 Cover glass passivation

1. Place a silicone gasket (Grace Bio-well) onto the cleaned cover glass.
    a. Note 1: if working with detergents in your buffers, you may consider using a hole puncher to cut out the center of gasket. This will enable evaporation, but detergents prevent solutions from flowing through the cell and create lots of bubbles inside the chamber.
    b. Note 2: Make sure to press firmly on the Bio-well to ensure it sticks to the cover glass. They will often fall off overnight from poor adhesion.
2. Prepare incubation chamber
    a. We used a "Snapware" container from Target with a rack from a pipette tip box. The bottom of the container was filled with ethanol to reduce evaporation of PEG solution.
3. Make solvent mix: 5% Glacial Acetic Acid + 95% HPLC Grade Ethanol (v/v).
4. Weigh out 1 mg biotin-PEG-silane for every 10 mg mPEG-silane.
5. Dissolve PEGs into 1 mL of solvent mix. Vortex variously to dissolve.
6. Add PEG solution to cover glass with Bio-well.
7. Leave overnight in solution chamber to passivate.
8. On the next morning, prepare a BSA containing solution (e.g. 10% BSA in any buffer)
9. Remove PEG solution from Bio-well
    a. Note: if the solution evaporated overnight, do not bother using the samples.
10. Rinse chamber with HPLC Grade Ethanol (10 volumes worth)
11. Rinse chamber with Milli-Q H_2O (10 volumes worth)
12. Dry the solution with Argon
    a. Note: when adding aqueous solutions into the chamber now, you should see the solution resist spreading out on the surface and hold still as a blob. This indicates good passivation since the water is being repelled by the surface.
13. Incubate with a BSA solution (e.g. 10% BSA) for at least 30 minutes prior to use.
    a. Note: This forms an extra passivation layer and helps to prevent non-specific absorption.

7.3.3 TIRFM control experiment: non-specific absorption

It is a good idea to test that your samples are actually passivated. This can be done by quantifying the amount of non-specific binding of your molecule of interest to the surface.

1. Take blank images of your chip with the excitation wavelength for your molecule.
    a. Note 1: Expect the most noise around 488 nm and less around 640 nm.
    b. Note 2: If there are bright spots everywhere, your sample is dirty and cannot be used.
2. Add < 1 nM of a fluorescently labeled molecule to the surface. Incubate for 10 minutes[xiv]
3. Rinse chamber multiple times to remove excess protein.
4. Image multiple spots on your chip. If passivation was successful, you should only see a few non-specially bound molecules. If there are now spots everywhere, passivation failed.

[xiv] An appropriate concentration of your molecule to use will need to be determined by experimentation. As a general guideline, we used ~10 pM of eGFP tagged proteins for photobleaching experiments.

7.4 ZMW experiments

7.4.1 ZMW fabrication at ORNL CNMS

The fabrication of ZMWs at Oak Ridge National Labs (ORNL) Center for Nanophase Materials Sciences (CNMS) was done using positive-tone electron beam lithography[3]. A general schematic of ZMW fabrication is found in **Figure 4-S3.** Here, I indicate steps specific to ORNL tools in case anyone else needs to make ZMWs at the facility.

1. Cover glass cleaning (SC-1 clean)
    a. Soak cover-glasses in 5 parts deionized water, 1 part 30% hydrogen peroxide, 1 part 35% ammonium hydroxide for 15 minutes at 75°C
    b. Rinse with water, dry with N_2
    c. Plasma clean for 5-10 minutes before starting the fabrication process.
2. Aluminum deposition
    a. Deposit 100 nM of Aluminum via Ebeam evaporation at 2 Å/second with a dual source electron-beam evaporator
        i. ORNL: 28 cover-glasses can be loaded at a time (7 per wafer) and deposited using the "substrate rotation option". If using rotation, increase the Al deposition height to 125 nm to account for the time each chip is not above the source. This will result in ~100 nm height (as measured by profilometry)
3. Spin-coat photoresist
    a. Spin-coat ZEP520A onto cover glasses for 45 seconds at 2000 rpm
        i. ORNL: Program 2
    d. Bake at 180°C for 2 minutes
4. Load samples into the electron-beam lithography tool
        i. ORNL: Can load four cover glasses at a time using cassette 8 of the JEOL JBX-9300FS E-beam lithography system.
5. Expose samples
    a. File: ZEP56.mgn
    b. Should take 20-30 minutes for calibration and exposure of 4 cover-glasses
    c. Note, despite our best efforts, each position on cassette 8 (A-D) receive slightly different exposure and result in different sizes. Adjust dose as needed.
6. Develop
    a. Develop in xylenes for 30 seconds.
    b. Rinse with isopropyl alcohol from squirt bottle. Dry with N_2.
    c. Check for patterns using dark field illumination.
7. Aluminum Etch
    a. Etch 100 nm of Al using Oxford Plasmalab 100 Reactive Ion Etcher
    b. 30 standard cubic centimeters (sccm) chlorine (Cl_2) and 10 sscm boron trichloride (BCl3) gasses at 50°C for one minute.
        i. ORNL: Recipe: OPT-Al Etch-ICP

        ii. ORNL: Run the recipe once on a dummy wafer to make sure the tool is at 50°C.
        iii. ORNL: apply a small drop of krytox to the back of each cover-glass and set on a clean wafer. This will prevent photoresist burning during the etch.
    c. Remove krytox by gently wiping the back of the cover glasses. This is the point where most samples will break so be very careful!
8. Plasma cleaning
    a. Plasma clean samples for 10-15 minutes on high power to remove remaining photoresist
        i. ORNL: Max-strip-O2-Ar-10-min
        ii. You should now be able to use bottom up illumination to see ZMW features.
9. Confirm with SEM
    a. ORNL: Merlin is the best SEM for resolution, but is always booked by users. The Phenom XL has lower resolution but can be used to confirm samples if you did not need accurate measurements of size.
10. Chip transport and storage.
    a. Taping each ZMW chip into wafer carriers works well. Place tape on the corners of each cover glass. Each wafer holder can fit 7 ZMWs. I additionally tape the outside out the holder to prevent air flow and wrap tightly into plastic zip-lock bags while still inside clean room to maintain the dust-free atmosphere.
    b. I have successfully brought back bags of ZMWs in carry-on luggage without any issue from airport security or damage to the samples.

7.4.2 ZMW passivation

1. Plasma clean ZMW chips immediately before passivation.
2. Aluminum Passivation:
    a. Incubate ZMW in 2% PVPA solution at 90°C for 2-3 minutes
    b. Rinses with water, dry with Argon.
3. Glass passivation:
    a. Follow **7.3.2** for procedure to passivate glass of ZMW.

7.4.3 ZMW control experiment: ligand diffusion

To ensure ZMWs work as intended and do not exhibit non-specific absorption of ligands to the surface, it is a good idea to image ~1 µM solution of a diffusing ligand (e.g. fcAMP) in ZMWs. See **Figure 4-S4**.

1. Add 1 µM fluorescent ligand to ZMW chip
    a. Keep ~1% BSA in the solution to reduce non-specific absorption
2. Record ligand binding (a couple minutes is good enough)
    a. You should see that most ZMWs appear empty for the duration of imaging.
    b. Some ZMWs will light up for a brief second then go dark- this is normal. If ZMWs are staying bright, this indicates non-specific binding. If only a few ZMWs (out of the 100-1,000s per field of view) show “sticking” behavior, the chip is fine.
    c. Alternatively, if all ZMWs are illuminated this could indicate that the ZMWs were not fabricated properly (e.g. Al is too thin and the bulk ligand solution is now being excited)

7.4.4 ZMW control experiment: Non-specific absorption

1. Follow **7.3.3** for non-specific protein absorption controls
    a. Note: protein concentrations in ZMWs typically need to be 10-100X higher for surface incubation vs cover glasses.
2. See example in **Figure 4-S4**

7.5 A note on removing oxygen from solutions

Displacing oxygen via inert gas (e.g. nitrogen or argon) is a common avenue to extend the lifetime of organic fluorophores before photobleaching. However, removing oxygen will result in excessive photoblinking. Fortunately, the blinking of organic fluorophores can be quenched via triplet state quenchers like Trolox. Therefore, solution purging followed by addition of oxygen scavengers (PCA/PCD) and triplet state quenchers (Trolox) is a powerful scheme to extend the lifetime of organic fluorophores. However, this avenue is not always beneficial. In particular, the chromophore of fluorescent proteins like eGFP are buried inside a beta-barrel and are inaccessible to solvents and triplet state quenchers. For the case of fluorescent protein photobleaching experiments, a better signal can be achieved by not purging solutions or adding an oxygen scavenging systems. This is a lesson I learned the hard way when starting to measure eGFP-HCN photobleaching steps in **Chapter 4.** I refer readers to the terrific a terrific article by Taekjip Ha and Philip Tinnefeld for a discussion of photophysics[4].

7.6 References

1. Aitken, C.E., Marshall, R.A. & Puglisi, J.D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical Journal* **94**, 1826-1835 (2008).
2. Cordes, T., Vogelsang, J. & Tinnefeld, P. On the Mechanism of Trolox as Antiblinking and Antibleaching Reagent. *Journal of the American Chemical Society* **131**, 5018-+ (2009).
3. Levene, M.J. et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686 (2003).
4. Ha, T. & Tinnefeld, P. Photophysics of Fluorescent Probes for Single-Molecule Biophysics and Super-Resolution Imaging. *Annual Review of Physical Chemistry, Vol 63* **63**, 595-617 (2012).