Hierarchical dynamics of individual RNA helix base pair formation and disruption

Jason Hon

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2017

ABSTRACT

Hierarchical dynamics of individual RNA helix base pair formation and disruption

Jason Hon

This thesis explores the RNA folding problem using single-molecule field effect transistors (smFETs) to measure the lifetimes of individual RNA base-pairing rearrangements. In the course of this research, considerable computational, chemical, and engineering contributions were developed so that the single-molecule measurements could be conducted and quantified. These advancements have allowed, on the basis of the smFET data collected herein, the quantification of a kinetic model for RNA stem-loop structures which has been generalized to quantitatively explore the phenomenological observation that an RNA found in the *bacillus subtilis* strain acts as a metabolite-sensing switch, allowing RNA polymerase to transcribe the messenger RNA when the metabolite is present and preventing transcription when the metabolite is absent. Together, the data presented quantify a simple model for the base pairing rearrangements that underlie RNA folding.

Table of Contents

List of Figures……..……..v

List of Tables……...…..vi

List of Appendices...……...…..vii

Acknowledgements……..…viii

Chapter 1 Introduction..... 1

Part 1: Development of single-molecule field effect transistors and computational methods for analysis of single-molecule trajectories..... 5

Chapter 2 Operation, Mechanism, Fabrication, and Chemical Reactions of Single-Molecule Field Effect Transistors..... 9

2.1 Introduction..... 9

2.1.1 Mesoscopic Conduction..... 10

2.1.2 Fabrication of field effect transistors made from carbon nanotubes..... 14

2.1.3 Fabrication strategy..... 16

2.2 Electrochemically regulated Diazonium Functionalization of Isolated Single Walled Carbon Nanotube Field Effect Transistors 19

2.2.1 Results..... 22

2.2.2 Discussion..... 23

2.3 Pyrene Adsorption to CNTs..... 27

2.4 Fabrication of single-molecule field effect transistors using sequential reactions within nanowells..... 28

2.3.1 Results..... 30

2.3.2 Discussion 31

2.5 Reaction of CNTFETs with diazonium salts creates a point of heightened sensitivity 34

Chapter 3 Single-Molecule Computational Methods 42

3.1 Introduction 42

3.1.1 Dynamics 45

3.1.1.1 Continuous-time Markov Processes 47

3.1.1.2 Discrete-time Markov Chains 54

3.1.1.3 Dynamics of finite state tridiagonal matrices 59

3.1.2 Variational Inference 62

3.2 Practical Applications 68

3.2.1 Unified, Bayesian Inference-based Framework for Analyzing Single-molecule Fluorescence Microscopy Experiments 69

3.2.1.1 Introduction 70

3.2.1.2 Methods and Results 73

3.2.1.2.1 Identification 76

3.2.1.2.2 Registration 81

3.2.1.2.3 Absolute Registration 83

3.2.1.2.4 Intensity Estimation 86

3.2.1.3 Analysis 91

3.2.1.3.1 Movie Statistics 91

3.2.1.3.2 Trajectory Analysis 92

3.2.1.4 Conclusion 92

3.2.2 A Bayesian Approach to Hierarchical Hidden Markov Modeling Allows Direct Measurement of Conditional Kinetic Rates 93
3.2.2.1 Introduction 94
3.2.2.2 Theory 97
3.2.2.3 Results 101
3.2.2.4 Discussion 111
3.2.2.5 Materials and Methods 114
3.2.3 A Bayesian Approach to Single-Molecule Trajectories with Diffusing Observables 116
3.2.3.1 Introduction 117
3.2.3.2 Methods 118
3.2.3.3 Results and Discussion 118
3.2.3.4 Conclusion 119
Part 2: Dynamics of nucleic acids on the microsecond timescale 121
Chapter 4 Direct measurement of base pair-by-base pair zipping and unzipping of individual RNA stem-loops 124
4.1 Introduction 124
4.2 Results 130
4.3 Discussion 138
Chapter 5 Single-molecule observation of riboswitch zipping dynamics on the microsecond timescale 144
References 164

Appendices **190**

List of Figures

Figure 1.1: Examples of RNA structures. 3

Figure 2.1 Electrochemical modulation affects both the rate and extent of diazonium-dependent conductance quenching. 26

Figure 2.2 CNTFET bandgap affects kinetic rates of FBDP conductance quenching. 27

Figure 2.3 BHT, a radical scavenger, does not prevent FBDP-mediated reactions on CNT sidewalls. 28

Figure 2.4 Noncovalent smFET fabrication cycle 29

Figure 2.5 Effect of nanowell-confined chemistry on carbon nanotube devices. 32

Figure 2.6 Real-time sensitivity to secondary reactions on the primary single-molecule probe. . 33

Figure 2.7 Collection and analysis of dc- and tmSGM data. 35

Figure 2.8 Comparison of dcSGM signal with a tmSGM signal on the same CNTFET. 37

Figure 2.9 tmSGM signals on pristine CNTs. 39

Figure 2.10 Precise geometric introduction of sp3 defects using diazonium salts. 41

Figure 3.1 Identification of light-emitting chromophores. 75

Figure 3.2 Chromophore intensity *versus* time trajectory estimation. 89

Figure 3.3 Hierarchically organized dynamic heterogeneity. 95

Figure 3.4 Selection between distinct hierarchical models using variational inference. 103

Figure 3.5 Dynamic heterogeneity of PRE complexes. 106

Figure 3.6 Dynamic model for fluctuations between GS1 and GS2 110

Figure 3.7 Design and validation of the emission drift Hidden Markov Model. 119

Figure 4.1 Schematic of smFET measurement geometry, stem-loop constructs, and experimental conditions. 130
Figure 4.2 Transition pattern and population evolution of stem-loop constructs. 132
Figure 4.3 Thermodynamic and kinetic measures calculated from trajectories in the absence of any competitor DNA. 136
Figure 5.1 smFET experimental setup, RNA sequence design, and wild-type aptamer smFET trajectory overview. 149
Figure 5.2 Dynamic heterogeneity of the P1 stem. 153
Figure 5.3 Kinetic model for basepair-level fluctuations of the wild-type P1 stem in the presence of adenine. 157
Figure 5.4 Dynamics of the G3C aptamer. 158

Figure A. 1 Plots of vbscope intensity analysis for TIRF microscopy movie of 10 nM Cy5-mutRF1 binding to surface-immobilized Cy3-RC. 212
Figure A. 2 Plots of vbscope colocalization analysis for TIRF microscopy movie of 10 nM Cy5-mutRF1 binding to surface-immobilized Cy3-RC. 213
Figure A. 3 Plots of microscope parameters as analyzed by vbscope determined from a TIRF microscopy movie of 10 nM Cy5-mutRF1 binding to surface-immobilized Cy3-RC. 215

Figure C. 1: Characterization of CNT transistors. 243
Figure C. 2: Optimization of smFET functionalization. 244
Figure C. 3: Real-time invasion of the P1 stem. 244
Figure C. 4: Ligand-free fluctuations of the wild-type aptamer. 245

Figure C. 5: Bulk fluorescence binding data of 2-aminopurine to the wild-type aptamer. 246
Figure C. 6: ΔG separation between conductance classes of P1A and P1B for the wild-type and G21C aptamers..... 247
Figure C. 7: Models used for analysis of smFET data in Chapter 5..... 248
Figure C. 8: First passage distributions from the *pbuE* riboswitch trajectories described in Chapter 5..... 249
Figure C. 9: Segmented alignment of *pbuE* riboswitch sequences described in Chapter 5..... 250
Figure C. 10: Dynamics of the stable aptamer..... 251

List of Tables

Table C 1: Average rate constants for the wild-type *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals..... 252
Table C 2. Average rate constants for the G21C *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals. 252
Table C 3. Rate constants for the wild-type *pbuE* adenine-sensing aptamer contingent on occupancy in $P1^A$ under conditions of increasing adenine; error bars are 95% confidence intervals..... 253
Table C 4. Rate constants for the wild-type *pbuE* adenine-sensing aptamer contingent on occupancy in $P1^B$ under conditions of increasing adenine; error bars are 95% confidence intervals..... 253
Table C 5. Fractional occupancy of $P1^A$ or $P1^B$ for the wild-type *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals. 253

Table C 6. Rate constants for the G21C *pbuE* adenine-sensing aptamer contingent on occupancy in $P1^A$ under conditions of increasing adenine; error bars are 95% confidence intervals. 254

Table C 7. Rate constants for the G21C *pbuE* adenine-sensing aptamer contingent on occupancy in $P1^B$ under conditions of increasing adenine; error bars are 95% confidence intervals. 254

Table C 8. Fractional occupancy of $P1^A$ or $P1^B$ for the G21C *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals. 254

Table C 9. Average rate constants for the G3C *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals. 255

Table D. 1: Average rate constants for the stem-loop constructs tested in the absence of competitor DNA; error bars are 95% confidence intervals. 256

List of Appendices

Appendix A Probability Basics, Assorted Proofs, Update Equations..... 190

Appendix B Examples in Mesoscopic Conductance..... 233

Appendix C Additional Controls for Chapter 5 236

Appendix D Additional information for Chapter 4..... 256

Acknowledgements

While I certainly took a large role in organizing the data, theory, and writing composing this thesis, so many people – family, friends, collaborators – voluntarily took a role in shaping this work that it is a deep stretch of the imagination to call this work "my thesis" (a stretch which, to be clear, I absolutely assert.) For example, the experiments described in Chapter 5 were carried out on chips co-produced with three collaborators – Dr. Nathan Daly, Dr. Sefi Vernick, and Scott Trocchia – and were designed with Nathan's help, measured on an instrument designed and built by Scott. The theory presented came out of a few months of discussion with Nathan. The original project was suggested by my mentor Dr. Ruben Gonzalez, who also made critical remarks without which I could not have been led to my current opinion of the physical interpretation of the recordings. So in one chapter, many peoples' thoughts and work are tangled up. Every chapter has a story like this; I fervently believe that such projects, while rarely completed by single student-mentor teams, are much more likely completed through a process of rigorous debate amongst groups of skeptics who know enough to understand the stakes and care enough to seriously evaluate the evidence. For me, this group was mainly Dr. Nathan Daly, Dr. Ruben Gonzalez, Dr. Delphine Bouilly, Scott Trocchia, Dr. Sefi Vernick, Dr. Kenneth Shepard, Dr. Colin Nuckolls, Dr. Bridget Huang, Dr. Colin Kinz-Thompson, Dr. Jan Willem van de Meent, Yan Feng, Sarah Dubnik, Dr. Kelvin Caban, and Dr. Somdeb Mitra.

The work described in this thesis has been, however, just a small part of life. It would not have been worth the effort without the support of my wife, Weiling Liu, who, besides directly writing significant core functions of my code, has consistently dragged me around on various adventures, the fruits of which have been consistently rewarding.

Chapter 1 Introduction

In some sense, life is only possible because the rate of stacking two bases of a connected nucleic acid is three orders of magnitude slower than the rate of stacking two disconnected bases. As a result, information transfer, folding of complex structures, catalysis, *etc* can occur on the same timescale and even within the same molecule[1]. This thesis describes single-molecule recordings of fluctuations of DNA and RNA on the microsecond to second timescale to quantify a model for how these nucleic acids dynamically fold and how these fluctuations assist their many functions.

I have organized the presentation so that in-depth introductions to specific material lie in their respective chapters. The first two of these chapters describe methods-development. Chapter 2 outlines the carbon nanotube single-molecule field-effect transistor – how is it made, and how does it work? Chapter 3 introduce computational algorithms that identify, classify, and quantify trajectories of single molecules. Next follow two chapters that describe biophysical measurements of nucleic acid dynamics, using the measurement technique described in Chapter 2 to record trajectories of single molecules and the computational techniques described in Chapter 3 to analyze them. Chapter 4 describes rearrangements of conserved stem-loop structures with four-base loops known as tetraloops, which are vastly overrepresented in large RNAs such as the ribosome, are thought to hierarchically organize the RNA folding landscape and, furthermore, are known to mediate tertiary structure interactions. Chapter 5 describes metabolite-mediated rearrangements of an RNA switch, as well as a model describing how the

messenger RNA biases base-pair-level kinetic reshuffling to regulate its own transcription is developed based on single-molecule trajectories.

The main contributions of this thesis can be conceptually gathered into a bigger picture – the study of RNA folding. The RNA folding problem can be, by analogy with the protein folding problem, expounded through a hierarchical tree-like organization: as a sequence (primary) which arranges quickly into organized substructures (secondary), which eventually interact with other substructures (tertiary) and other folding sequences (quaternary) to complete the complex. RNA secondary structure is typically a form of helix. Which helices form between which parts of the sequence, as well as their stabilities, affect the likely permutations of tertiary structures that form in configurational space. These permutations can be thought of as existing on tiered timescales[2] which stochastically arrange and rearrange on the second-timescale because the elementary rearrangements that compose them depend on the base-pairing lifetimes of the individual RNA subunits. Remodeling secondary or tertiary structure in RNA may depend on many such individual lifetimes, and such transition pathways underlie essential processes from the operation of RNA switches to the formation of intersubunit bridges in the ribosome. Some examples of RNA structures are shown in Figure 1.1.

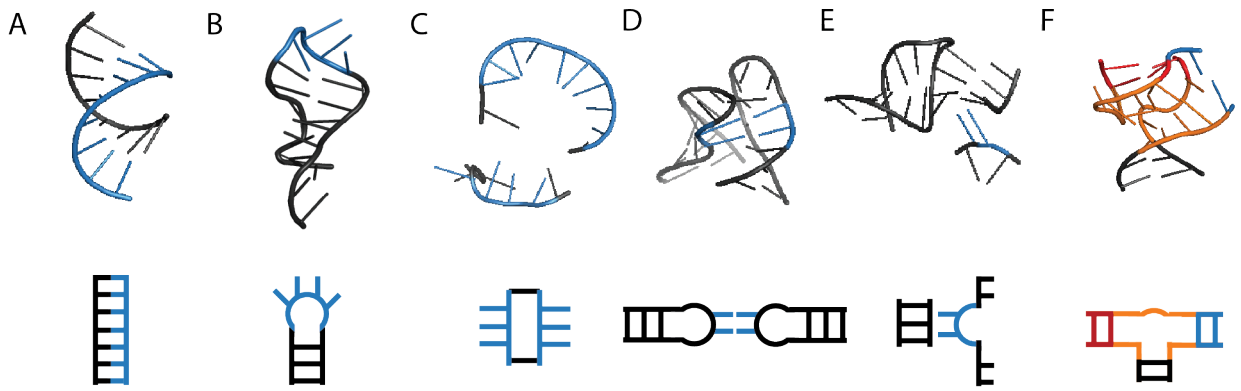


Figure 1.1: Examples of RNA structures.
(A) A-form RNA helix. The example shown is the first helix (helix P1) of the *pbuE* riboswitch aptamer discussed in Chapter 5 (pdb: 3IVN)[3]. (B) RNA stem-loop, which consists of a strand of RNA that loops on itself to form a paired region. The example shown is the sarcin-ricin loop found in ribosomal large subunit 23S rRNA (pdb: 5IBB)[4]. (C) A bulge in an RNA helix. The example shown is intersubunit bridge B7b between the ribosomal small subunit 16S rRNA and the large subunit L2 protein (not shown; pdb: 5IBB). (D) Kissing-loop interaction between two RNA loop structures. The example shown is from a different portion of the molecule in (A). (D) An A-minor motif, wherein an adenine residue interacts with the minor groove of another RNA helix. The example shown is the A-minor interaction between tRNA anticodon paired with a cognate codon to form a small helix, and two adenine residues in helix 44 of the ribosomal small subunit 16S rRNA (pdb: 1IBL)[5]. (E) An RNA junction, in this case a 3-way junction. The example shown is from a different portion of the molecule in (A).

The insight into the RNA folding problem provided by this thesis is the realization that single-molecule field effect transistors (smFETs) may be used to measure the lifetimes of individual base-pairing rearrangements. Considering the hierarchical tree-like picture of RNA

folding, smFETs are here used to quantify the rate constants associated with secondary structure rearrangements. This subject is described in Chapters 4 and 5: in the former, I investigate whether loop stability was conferred directly to the stem it caps and whether the subsequent data support a model describing one facet of how tetraloops serve as folding nuclei for secondary structure formation of larger RNA structures; in the latter I hypothesized that a messenger RNA switch, because of its proposed dynamic mechanism, likely regulates the rate at which its secondary structure remodels. The data support a model wherein the junction between two helices of the RNA switch stabilizes a third only if it binds a target molecule, and thereby prevents invasion by another the rest of its messenger RNA, all while being concurrently transcribed.

In conclusion, I argue that static structures, while occasionally capable of describing essential features of biological molecules, cannot easily communicate the myriad of possible motions leading to and regulating their function[6]. These motions are best described on a wide timescale that ranges from small-scale conformational motions to the large-scale motions they regulate, an investigation requiring methods that independently resolve their likely diverse conformational ensembles. With these remarks, I begin part 1, which describes how investigators typically formalize the language of these motions as well as contributes a unified quantitative stochastic framework for their description, and finally a description of the development of the technique this thesis predominantly utilizes, the single-molecule field effect transistor applied to fluctuations of nucleic acids.

Part 1: Development of single-molecule field effect transistors and computational methods for analysis of single-molecule trajectories.

In the majority of biophysical systems, experimenters typically deal with ensembles of molecules that exist in a complex equilibrium involving many interconverting states rather than as a single species. Because of the enormous number of possible random jostlings, interactions, and reactions, if one were to tag every atom in a molecule and monitor their configurational motions, after a short period, it would be impractical to predict their relative positions. If multiple atomic configurations are similarly stable, collective measurement of the properties of this equilibrium averages out the differences between subpopulations, leaving both the number of- and relationships between- subpopulations of the molecules comingled in the subsequent average. Single-molecule measurements seek to bypass this statistical mixing by considering one molecule at a time, assigning each to its appropriate configurational subpopulation, and subsequently reconstructing the ensemble. This approach has proven fruitful over the past few decades. This is because single-molecule measurements allow investigators to directly resolve the number, lifetime, and occupancy of individual states of a labeled molecule in real time.

Yet, single-molecule resolution comes at a cost. Some costs are simple to describe – for instance, labeling a single molecule with two fluorophores to generate a donor-acceptor pair, a configuration that allows one to monitor interfluorophore distance *via* the fluorescence resonance energy transfer (FRET) effect[7], can lead to ambiguities in interpretation due to perturbations arising from addition of the dyes. While these perturbations are diverse, for instance in the study of nucleic acid dynamics, care must be taken lest the highly aromatic dyes introduced to the sequence stack into a nucleic acid helix because the subsequent change in fluorescence could

possibly be incorrectly interpreted as a drastic change in interfluorophore separation[8]. On the other hand, the costs may manifest as complex effects – for instance, tethering a single molecule between two optical traps, a configuration known as optical tweezers, and systematically pulling in a given direction biases all molecular motion in that direction[9,10]. As a consequence, the force exerted by the optical trap on the molecule changes not only the enthalpic landscape but also the entropic landscape of the molecule, leading to the generation of potential artifacts that must be interpreted with care[10,11]. However, many such costs are, for the most part, obviated by careful experimental design, creating potential problems with interpretation mitigated by clever controls[12].

Problems associated with limited time resolution, however, are not so easy to sidestep with current technology. When observation of a state is limited by time resolution, the problem is that the information is simply not recorded in the trajectory[13]. Because interconversion between species may be fast, the time resolution of a given single-molecule technique must compete with the shortest of the lifetimes of these states lest they are erroneously missed[14,15]. A great number of single-molecule experiments reported in the past decade rely on photon emission from a site-directed fluorophore to, as a proxy, report the state of the molecule[6,16]. Photons are emitted randomly in every direction with some overall rate and, with some intermediate efficiency, converted into an electrical current for subsequent detection. The combination of both steps subsequently determine the signal amplitude per frequency and therefore the experimental time resolution. Typically, photon counts are infrequent, and because at least picoamps of current are required for detection by standard equipment, this creates a tension between the requirement of sufficient photon emission to support a given time-resolution, typically compensated by increasing laser power to compensate for increasing time-resolution, and the technical limitations

of modern dyes, which can undergo chemical reactions which halt photon emission. Therefore, many single-molecule optical techniques cannot simultaneously operate with high time-resolution and maintain a stable signal for a long period of time[13,17]. In other words, a transient atomic configuration may be probed, at the cost of not knowing how its dynamics change over time; or the long-time behavior may be captured, at the cost of not observing transient states.

In Part 1 of this thesis, Chapter 2, I describe my role in the development of single-molecule field effect transistors (smFETs) using carbon nanotubes (CNTs)[18–20]. This technology allows for label-free, tens of microsecond time-resolution of molecular trajectories, with stable signals persisting for hours or days which address many of the problems described above. To achieve this, the technique relies on the persistence of intrinsic molecular properties such as the average charge distribution, which do not cease so long as the molecule is intact. In this vein, the principal concerns of Part 1, Chapter 2 of this thesis are: how does one generate and validate the fabrication of an smFET? And what physical principles underlie the operation of an smFET?

However, even if one possessed a perfect account of the position and momentum of every atom of every molecule in a system as a function of time, significant difficulty would yet arise in organizing and therefore describing system behavior. Typically, one addresses this difficulty by breaking the system into sets of configurations that are in some sense equivalent and on the same basis distinct from any other such set of configurations. Then, one can describe how those equivalence classes interact or interconvert – a computationally tractable, if intricate, task[21]. As it stands, single-molecule trajectories typically record a handful of variables – for example, a single donor-intensity and acceptor-intensity *versus* time trajectory speak only to changes in interfluorophore distance as time goes by. Therefore, the computational task is both easier and harder – easier because the number of variables is smaller, and harder because many variables

contribute to a one dimensional signal. In Part 1 of this thesis, Chapter 3, I describe computational methods designed to break such a compound signal into its constituent parts, enabling a richer and more exact description of the single-molecule trajectory.

Overall, the methods described in Part 1 are honed to the task of generating and analyzing long, high time-resolution single-molecule trajectories. Part 2 of this thesis utilizes the methods described in Part 1 to quantify RNA folding on the single-molecule level.

Chapter 2 Operation, Mechanism, Fabrication, and Chemical Reactions of Single-Molecule Field Effect Transistors

2.1 Introduction

A major research focus of modern materials chemistry and semiconductor fabrication has been placed upon a targeted body of thought designed to take the abstract conception of a transistor and shrink it. As such, many mesoscopic materials, such as carbon nanotubes, have been deemphasized as potential components in next-generation electronics, because, as I will discuss below, they are too diverse to be ideal. In the wake of this shift away, however, has been a renewed interest in using these mesoscopic materials as small, label-free, electronic sensors – tools to directly answer fundamental questions in chemical and biological fields[20,22]. Here, after outlining the basic theory governing devices constructed using carbon nanotubes as conducting substrates, as well as a straightforward rationale behind the utility of carbon nanotubes as single-molecule probes of nucleic acid dynamics, I will present and discuss my efforts, in collaboration with many other excellent researchers in the laboratories of Profs. Kenneth Shepard and Colin Nuckolls, to use micro- and nanofabrication methods to design single-molecule field effect transistors (smFETs). Finally, I will characterize fundamental chemical transformations on field effect transistors whose channel consists of a carbon nanotube (CNTFETs) and outline the collaborative work behind the statistical validation of these transformations, as well as preliminary measurements demonstrating the robustness of the platform to detect fluctuations in quantal defects in the nanotube lattice at room temperature in solution. I will close with a study

of the consequence of controlled introduction of chemical defects on conductance through the 1D channel.

2.1.1 Mesoscopic Conduction

Because carbon nanotubes in particular and mesoscopic conductors in general are small, quantum effects must be taken into account when calculating their conductance. The central idea, using the non-equilibrium Green's function (NEGF) method[23], is that the transmission spectrum of carriers through a small channel is given by:

$$T(E) = Tr\left(\Gamma_S G^r \Gamma_D G^{r+}\right)$$

where $\Gamma_S = \frac{i}{2}(\Sigma_S - \Sigma_S^+)$ and $\Gamma_D = \frac{i}{2}(\Sigma_D - \Sigma_D^+)$ are the source and drain broadening matrices, respectively, Σ_S and Σ_D describe the interaction potential between the channel of the device and the source and drain, respectively; G^r is the retarded Green's function, given by:

$$G^r = [(E + i0^+)\, I - H - \Sigma_S - \Sigma_D]^{-1}$$

The transmission spectrum allows calculation of the current given a temperature:

$$I = \int T(E)\, f_T(E) dE$$

where $f_T(E)$ denotes the thermal distribution function of E, the energy. Because it has R types of repeating units, in the tight-binding approximation, the Hamiltonian H of a carbon nanotube has a block-tridiagonal structure:

$$\left[H_{i,i+1}\right] = \beta_k, k = i \ \ mod(R)$$

$$\left[H_{i,i-1}\right] = \beta_{k'}, k' = (i-1)\, mod(R)$$

$$[H_{ii}] = \alpha_i$$

where mod refers to the remainder function, the α_i are self-interactions within a repeating subunit and vary as a result of applied potentials, and the β_k are interactions between subunits[24]. In this model, carriers are created along the length of the nanotube and propagate from repeating unit to unit, according to the block-tridiagonal Hamiltonian, until they reach one of the termini and absorb. The structure of the repeating subunit, which governs the efficiency of the propagation, is determined by the chirality of the carbon nanotube. Loosely, there are many types of propagation depending on which site in which repeating unit the carrier has to jump from as well as where it has to jump to. Because a carbon nanotube consists of a contiguous lattice of hexagonal rings made of carbon atoms at the vertices, which have been collectively rolled up onto each other to form a cylinder from a 2D sheet, the transmission efficiency of these propagation types is determined by the symmetry of how the cylinder was wound. This symmetry is given by the chirality indices (m,n) which are, given an atom in a hexagonal unit, how many hexagons in the (m) longitudinal or (n) latitudinal direction must be traversed before arriving at the atom that coincides with the given atom in the next unit cell[25]. As anticipated by the symmetries of the above Hamiltonian[25], the nanotube chirality has a large effect on how the nanotube propagates charge carriers: if $n = m$, the nanotube is metallic, because there are only two types of propagating sites which are efficiently coupled into a helix oriented transverse to the nanotube axis; if $(n - m)\ mod(3) = 0$, the nanotube is semiconducting, because none of the propagating types are compatible with a transverse mode; else, the nanotube is "semimetallic," because some but not all of the propagating types are compatible with a transverse mode. These designations are made on the basis of measurements that gate the nanotube channel with an external electric field to make the nanotube a field effect transistor (FET) – by modulating the external field at the surface of a semiconducting or semimetallic nanotube, the conductance can

be decreased in many cases arbitrarily close to zero – intuitively, the field rotates the carriers so that they can either flow or not flow through the channel, but only if they are in non-metallic conductance modes.

Early attempts to reconcile conductance theory and experimental results focused on the local density of states (LDOS) of the carbon nanotube, loosely the measure of how many carriers fit in a volume at a defined spatial location as a function of their energy. The pattern in the Hamiltonian, in theory, translates into regularities in the LDOS, and on this basis the LDOS was calculated using the tight-binding model and compared to direct measurements using Scanning-Tunneling microscopy (STM)[26,27]. The predicted LDOS was expected to have peaks – known as van Hove singularities – at energies defined by nanotube chirality. Because of the high spatial resolution of the technique, the nanotube chirality was imaged concurrently with the LDOS[25–27]. When directly probed, the measured LDOS matched the LDOS calculated from the tight-binding model. In the cited studies, there were hints of the complexities associated with nanotube defects as well as surface defects which could interact with the nanotube, both of which strongly modulated the density of states. In Appendix B, I use the NEGF formalism to discuss the effect of perturbations of the Hamiltonian on conduction in a simple 1D lattice, which may be helpful to sort out some of the details. The principal idea is that any local perturbation has a significant effect on the conductance because all the carriers have to go through the perturbed site on their path from the source to the sink, and therefore a perturbation changes the carrier density, or local density of states, after itself by attenuating what is allowed to pass it by.

As control over surface features as well as quality of nanotube preparation increased, investigators began to notice that the conductance, sampled regularly with time, had large quantal fluctuations[28–34]. In analogy with experiences with metal-oxide field effect transistors

made from Si substrates (MOSFETs), these quantal fluctuations were originally supposed to arise mainly from surface effects, because they mainly affected non-metallic nanotubes and because the nanotubes were hypothesized to be covered in surface defects, as evidenced by the fact that they are usually not straight even though they should be if they were perfect crystals. The fluctuations could be quite dramatic – experimentally, nearly the entirety of the conductance through a nanotube could be spontaneously quenched[28]. Eventually, these defects were observed in metallic nanotubes as well[35]. Concurrently with these observations, investigators noticed that even at room temperature, Coulomb blockade was directly observed[36], confirming proposals based on STM observations[26].

Recently, theorists have attempted to tease out the nature of these quantal fluctuations. Through calculations it was found that a single charge could extinguish nanotube conductance, given a certain oxide dielectric, if it was just located in the right point in the channel[37]. By charging an atomic force microscopy (AFM) tip and scanning the length of the nanotube, investigators were able to find evidence of precise geometric positioning of these defects and, in some cases, were able to produce mechanical damage to the nanotube surface near the source or drain electrodes, and subsequently probe changes in the conductance response map[38,39]. It was found, in some cases, that the presence of an electrochemically introduced defect[40] matched or exceeded the expectations of theory[41], leading to significant modulation or loss of conductance through the channel, the drop itself shown to be localized in the channel towards the source or drain electrodes[18,42]. Immediately, these sensitive and chemically modifiable defects were used to construct the first single-molecule field effect transistors[18,43] (smFETs), capable of monitoring changes in nanotube conductance caused by rearrangements in individual molecules attached to the defect site. Subsequent experimental investigations found that even in the absence of an

introduced covalent defect, these smFETs could still be fabricated and be used as probes of molecular conformational changes[20,44,45]. In this chapter, I will discuss my efforts to fabricate carbon nanotube field effect transistors (CNTFETs), functionalize them with molecules with specific chemical handles, and either use those handles to directly attach individual molecules to the nanotube surface or directly probe them to verify that the defects have an impact on the channel conductance. This work was highly collaborative; as a footnote of each section, I will annotate the specific contributions of individual researchers.

2.1.2 Fabrication of field effect transistors made from carbon nanotubes

Field effect transistors (FETs) are fundamental objects in the modern world mainly because of their use as switches. The general idea is that a channel with a given bias between two leads attached to the channel, known as a source drain bias, can be tuned between a conductive and a nonconductive state based on tunable capacitive charging between the channel and a third electrode[46]. In this thesis, field effect transistors fabricated using carbon nanotubes will be used to measure the conformational dynamics of nucleic acids, and in this chapter I will characterize the properties and study the mechanism of conductance through those channels, with the ultimate goal of contributing a scheme by which operation of smFETs may be understood as well as methods by which this operation may be coupled to chemical transformations upon the nanotube surface.

FETs are typically named according to the design of the capacitor that applies the modulating field and thus operates the switch. In this thesis, this capacitor, often termed the "gate," will either be the Si substrate separated from all the electrodes by a thick SiO_2 oxide, an AFM tip, or a solution-coupled electrode. The first two are simple capacitors through a static

dielectric. The latter is diffusive. Therefore I will begin by noting the elementary theory for how an electrolyte solution can control conductance through a nanotube channel. Measurements of intermolecular interactions or intramolecular conformational changes through measurement of their effect on carbon nanotube conductance manifesting as- or mediated through- a defect in the nanotube lattice must always be complicated by the fact that, for almost all cases of interest, conformational changes only take place in electrolytic solutions. In contrast to a static dielectric, an electrolytic solution contains both positively and negatively charged ions, which are mobile, and which therefore negate any putative permanent charge. In order to detect a change in molecular conformation using the nanotube, therefore, the constantly shifting ion cloud must fluctuate slow enough so that, when by chance intermolecular collisions alter the ion density in the cloud to reveal somewhat the electronic configuration of the molecule, the change lasts long enough to both affect the conductance through the nanotube channel and be detected by a measurement apparatus. Because the rate of diffusion due to thermal fluctuations in an electrostatic field provides the quantity that determines whether measurements are possible, the crucial quantity is the distance correlation function $h_{ij}^{D}(r)$ between two molecules with opposite charges. This is given by, where z_i and z_j are the charge per ion molecule and e_c is the charge of an electron, $\epsilon_0\epsilon$ is the permittivity of the material, k_BT is Boltzmann's constant times the absolute temperature, r is the distance between the two ions:

$$h_{ij}^{D}(r) = -\frac{z_i e_c z_j e_c}{4\pi\epsilon_0\epsilon k_B T}\frac{e^{-\frac{r}{r_D}}}{r}$$

And finally, $r_D = \sqrt{\frac{\epsilon_0\epsilon k_B T}{e_c^2 \sum_l \rho_l z_l^2}}$ is the Debye length and ρ_l is the concentration of charged species l, valid for low concentrations and assuming ions which are points. Intuitively, this correlation

function quantifies the ion density in a shell around which measurements are subsequently possible; this physical theory has been explicitly tested for measuring rearrangements of proteins and nucleic acids on Si nanowire FETs and carbon nanotube smFETs[18,47]. The characteristic length of the shell is given by r_D, and this is roughly equal to the shell around which the carbon nanotube, as a sensor, can effectively probe the state of an attached molecule. Because it is immobile, the nanotube also assembles a charged layer known as the Helmholtz layer which is typically no more than a few molecules in thickness.

To sum up, the general strategy will be to fabricate a single crystal field effect transistor using a carbon nanotube as a substrate. In the following experiments, the field provided to the nanotube will be from the either a single molecule attached to the nanotube (an smFET), the solution itself (a solution gated CNTFET), the silicon substrate upon which the nanotube was grown, or an atomic force microscopy (AFM) tip.

2.1.3 Fabrication strategy[1]

Devices for CNTFET and smFET measurements were fabricated using standard photolithographic and electron-beam lithography methods (see [18,48] for examples most relevant to this application.) CNTs were grown by chemical vapor deposition using a ferritin catalyst and ethanol as a carbon source[49]. Patterning electrodes using photolithography proceeded through use of either a positive or a negative photoresist, which is a chemical spin-coated onto a chip and crosslinked after exposure to light to provide differential solubility. Two layers of two chemically distinct photoresists were applied to 1x1 cm chips, first Shipley S1813 and

[1] Strategy developed with Drs. Nathan Daly, Sefi Vernick, Delphine Bouilly, as well as Scott Trocchia, Jaeeun Yu, and Yan Feng.

subsequently LOR3A (Microchem, both), in series, by dripping the resist onto the surface and then spinning the chip at a defined frequency and heat-curing, resulting in a more or less even layer of light-sensitive material on the chip surface. A micron scale mask was separately prepared by taking a soda-lime glass plate coated with Cr and a separate photoresist (ip3500, Microchem) and crosslinking the photoresist on the surface by programmed patterning with a laser (Heidelberg DWL66, Singh Center for Nanotechnology, University of Pennsylvania). Crosslinked ip3500 was removed with CD-26 (Microchem) and Cr was removed using a strong acid (Chrome Etchant, Microchem). Chips covered in bilayer Shipley 1813/LOR3A were placed underneath a Hg lamp and exposed to light through the mask to crosslink both photoresists; subsequently the crosslinked photoresist was dissolved in AZ400 MIF, leaving behind trenches in the photoresist layer down to the SiO_2 surface located at the positions defined by the mask. Because crosslinked Shipley 1813 dissolves faster than crosslinked LOR3A in AZ400 MIF, these trenches have an "undercut" structure, so that some LOR3A masks a photoresist-free portion of the chip surface without actually contacting that surface. Subsequently, metal was deposited onto the surface by electron-beam evaporation (Angstrom) to some defined height. Because of the undercut structure, following exposure to a chemical in which un-crosslinked resist is soluble (Remover PG), metal that contacts the SiO_2 surface remained and the rest was removed from the surface, resulting in micron-scale defined electrode features. This method was used in series to deposit 75nm Ti as a source-drain electrode material, then 100nm Pt as a solution gate electrode as well as a pad buffer for wiring the chip. Ti was selected for the source-drain electrodes because its surface oxide passivates it against most chemical reactions as well as forming efficient electrical contact with the CNT[50]. Pt was selected for the gate electrode because it efficiently couples electrochemically to both organic and aqueous solutions. The

channel length of 4 microns was chosen because CNTs are expected to be diffusive, as opposed to ballistic, conductors when the channels have such a length[23]; furthermore, longer channels are undesirable for smFET measurements as gate-associated noise increases as the square root of the channel length. At the completion of any photolithographic step, chips were vacuum annealed at 350C and ~10^{-6} torr for 30 minutes to remove small fragments and chemical residues.

After deposition of at least the Ti source-drain electrodes, the chip effectively possessed a detailed map printed onto its surface. This was used to locate all the CNTs using a scanning electron microscope (SEM, Hitachi 4700). CNTs were selected for isolation as CNTFETs on the basis of their length, apparent optical quality, Raman spectra, and transport characteristics. CNTs were isolated by photolithographically defining a region of remaining resist to cover them and protect them from subsequent exposure to an O_2 plasma. For experiments in which they were used, nanowells were defined using electron beam lithography following spin-coating the chip with polymethylmethacrylate (PMMA[48]). Two types of device were used for the experiments in this chapter – one, developed by Steven Warren and Scott Trocchia is a wide field of densely packed electrodes with Ti contacts; a second, developed by Jaeeun Yu and Delphine Bouilly is a dense packed array of Ti/Pd/Au electrodes used for the collection of conductance statistics associated with nanowells (discussed in detail below).

2.2 Electrochemically regulated Diazonium Functionalization of Isolated Single Walled Carbon Nanotube Field Effect Transistors [2]

As outlined in the introduction to this chapter, defects in CNTs are common and these defects occasionally cause stochastic, quantal fluctuations in conductance through the channel. This thesis is concerned with programming these defects with biomolecules whose stochastic fluctuations are both of biological interest and at the timescale accessible to a CNTFET. Therefore, there are two significant challenges: first, CNTs are for the most part chemically inert[51]; second, CNTFETs obey electrochemical rules as, for the most part, semiconducting working electrodes[52], instead of a more canonical role as bulk reaction species – in particular, charge is not conserved because it can exit *via* the drain.

In bulk, many methods for modifying nanotube surfaces have been developed, and extensively reviewed, including cycloadditions[53], azide modification[54,55], and oxidation[40,56], both UV-assisted and acid-catalyzed. In particular, acid catalyzed methods have been used on CNTFETs as a method for fabricating smFETs[18,22,40,42]; however, the method is frenetic and uncontrolled, and a full account of its mechanistic details rendered difficult by challenges inherent in characterizing an event that just happens once.

Two more promising approaches to smFET fabrication have been recently described – first, noncovalent adsorption of pyrene moieties bearing useful chemical handles to nanotube surfaces[20], which has the advantage of a straightforward and reversible protocol that I will discuss in the next section; second, covalent functionalization of CNT surfaces using diazonium reagents, which I will describe in this section.

---

[2] Experiments designed and data analyzed with the assistance of Dr. Sefi Vernick.

Diazonium functionalization of nanotubes typically proceeds *via* a radical chain reaction that proceeds in three coupled steps[57]. First, in solution, diazonium ions interact with water to form aryl radicals; second, these adsorb to nanotubes; finally, they react pairwise with the nanotube. The reaction proceeds in accordance with the phenomenological Hammett parameter[57] and typically has a greater affinity for metallic nanotubes than for semiconducting nanotubes[57–59]. Covalent addition of diazonium is both predicted and observed to have distinct effects on the electronic structure of the CNT, in particular, the addition of so-called mid-gap states resulting from depletion of specific reactive modes[60,61].

Taking this last fact as a starting point, first, note that a nanotube has a countable number of reactive modes. I will here distinguish between a reactive mode, which is the particular wavefunction that interacts with the reactant, and a reactive site, which is where a reactant happens to add itself. Enumerating the "reactive state" of the nanotube by an integer, the probability that $j \leq N$ of these modes will have reacted by time t is governed by the following master equation[62]:

$$\frac{dp_j(t)}{dt} = -jkp_j(t) + (j+1)kp_{j+1}(t)$$

This is a pure death process with rate parameter k. The states are enumerated according to the number of remaining reactive sites which are no longer reactive following functionalization. To solve this equation, consider instead an infinite chain – it makes no difference, as the initial condition will make the chain finite since the chain is pure death and therefore it never accesses the virtual states. The generating function:

$$g(t,\xi) = \sum_{j=0}^{\infty} p_j(t)\xi^j$$

obeys the equation:

$$\frac{\partial g}{\partial t} = k(1-\xi)\frac{\partial g}{\partial \xi}$$

which, on being solved with the initial condition $g(0,\xi) = \xi^N$, gives:

$$g(t,\xi) = \left(1 - e^{-k}\ (1-\xi)\right)^N$$

This is the generating function for a binomial distribution with $p = e^{-kt}$. The probabilities themselves are given by differentiation of $g(t,\xi)$:

$$p_j(t) = e^{-jkt}(1-e^{-kt})^{N-j}\frac{\Gamma(N)}{\Gamma(N-j+1)\Gamma(j+1)}, j \leq N$$

2-1

As noted above, the probability of any number of reactive sites above the initial value vanishes because the gamma function diverges on negative integers. Finally, the expected number of reactive modes remaining at a given time is given by:

$$E[modes] = \frac{\partial g}{\partial \xi}|_{(t,1)} = Ne^{-kt}$$

Successful fabrication of an smFET occurs when the nanotube leaves the state N, enters the state $N-1$, and remains there when the assay terminates. As a function of time, the probability that this happens is given by:

$$p_{N-1}(t) = e^{-(N-1)kt}(1-e^{-kt})$$

This expression, which is positive and starts and ends at 0 and therefore must have a local maximum, implies that there is only a very short time after initiating a reaction for which only a single reactive event may be expected. However, reasoning that the rate parameter k depends entirely on the band-gap properties of the CNT as well as the electronic configuration of the smFET as well as diazonium concentration, I investigated the validity of this theory as these

parameters were modified by simultaneously monitoring many nanotubes as their reactive modes were depleted by diazonium. First, arguing that a radical chain mechanism would be difficult to harness kinetically, I will describe a method whereby diazonium functionalizes CNTs *via* electrophilic attack. Taking this as a starting point, I will develop a simple kinetic assay for monitoring the extent to which a nanotube has reacted, by monitoring conductance as a function of solution gate voltage, known as the IV characteristic, and further verifying *via* confocal Raman scattering spectroscopy[63,64]. The assay naturally lends itself to electrochemical modulation using the source-drain electrodes, and I will describe its effect on the rates and extent of conductance quenching caused by the reaction, as well as the consequences of these observations on the proposed mechanism.

2.2.1 Results

An electrophilic attack mechanism for diazonium addition to electrodes has been proposed for functionalization of metallic substrates. Arguing that the reaction mechanism either involved a radical chain reaction in solution coupled to the CNT or direct electrophilic attack, the validity of the former was investigated by exposing CNTs to the reaction of the reagent formylbenzenediazonium[65] (FBDP) at 10mM concentration in acetonitrile, supported by 100mM $(N\text{-}(n\text{-}bu)_4)^+$ PF_6^-, as opposed to aqueous solvent supported by buffer, within a polydimethylsiloxane (PDMS) flowcell for specified periods of time then thoroughly flushing out the reagent and measuring the IV characteristic. 74 CNTs, 19 large bandgap (semiconducting) and 55 small bandgap (semimetallic), were thus treated. Because the reaction was insensitive to 10x addition of dibutylhydroxytoluene, BHT, (N=10), a radical scavenger, and further because the reaction in general followed an apparent exponential rate law, I argue that the

rate parameter *k* was constant, whereas it would be steadily increasing were a radical chain coupled to the CNT reaction, and would be quenched by addition of BHT. In contrast to reports[61], discrete events were not observed; however, substituting the Ti electrodes for Au, discrete events appeared, suggesting that these result from reaction of diazonium with gold.

While the rate parameter was constant, both the rate and extent of conductance quenching were observed to be dependent on the band-gap properties of the CNT and the electronic configuration of the CNTFET. In particular, setting the CNT at a negative (-1V) bias relative to solution (N=6) greatly increased the rate relative to a neutral bias; a positive (+1V) bias and neutral bias had nearly equivalent rates, but a positive bias (N=5) caused near extinction of conductance and much more intense Raman disorder bands (Figure 2.1). Finally, it was observed that small bandgap CNTs reacted much faster than large bandgap CNTs (Figure 2.2). These results are discussed below, assuming an electrophilic attack mechanism.

2.2.2 Discussion

By comparison of the apparent reaction kinetics to an exponential rate law with and without radical scavengers, I conclude first that under conditions of the assay described above, the diazonium functionalization of CNTs proceeds *via* an electrophilic mechanism with a fixed rate parameter. Because the rate parameter is fixed, subsequent reactions do not affect the overall rate, so that reactions are not correlated, *i.e.*, the remaining reactive modes are unaffected by a reaction at any of the other modes. Here I highlight the distinction between reactive modes and reactive sites – reactive modes are the electronic structure elements that interact directly with diazonium and are delocalized; reactive sites are real space monuments to depletion of reactive modes, commemorated with the diazonium adduct.

The accelerated rate of small *versus* large bandgap is consistent with the proposal that conductive modes are the reactive modes, since there are more conductive modes in small bandgap CNTs than in large bandgap CNTs in the energy region high enough to react efficiently with the ion. However, the question remains why the extent of conductance quenching varies between the three bias conditions tested. The solution can perhaps be found by considering what else is correlated with high current – as a semiconducting electrode in solution[52], the more carriers that flow through the channel, the more ions will be associated with the CNT. Some of these ions are the reactive FBDP ions; since the local concentration is higher around small bandgap than large bandgap CNTs, this contribution will also explain why the reaction apparently proceeds faster, and this conceptual framework presents a more useful mnemonic, *vide infra*.

This simple theory can be extended to explain why the extent of conductance quenching varies between the three bias conditions tested. First, note that just because the conductance ceases to be quenched does not necessarily mean that the reactive modes have all reacted, while this is very likely the case. The most general conclusion is that all reactions that affect conductance have stopped. In other words, if a reaction that has already occurred quenches a conductive mode that is the only mode affected by a subsequent reaction, the subsequent reaction will not be detected by this assay. Therefore, it is reasonable to suppose that the cause of the variegated asymptotic conductance quenching across the three bias conditions is a geometrically distinct reaction mechanism. Solving the Poisson-Boltzmann equation for various gate biases of a CNT in a cylindrically symmetrical dielectric[24], one notes immediately that the expected charge distribution varies across the length of the channel (for example, see Figure B.1). The diazonium ion is therefore expected to be more concentrated in certain geometric positions than

others. Therefore, I propose that at positive biases, the reaction proceeds in a very concentrated position in the CNT, leading to more efficient conductance quenching due to 1D effects, whereas at the neutral and negative biases the reaction tends to be more diffuse.

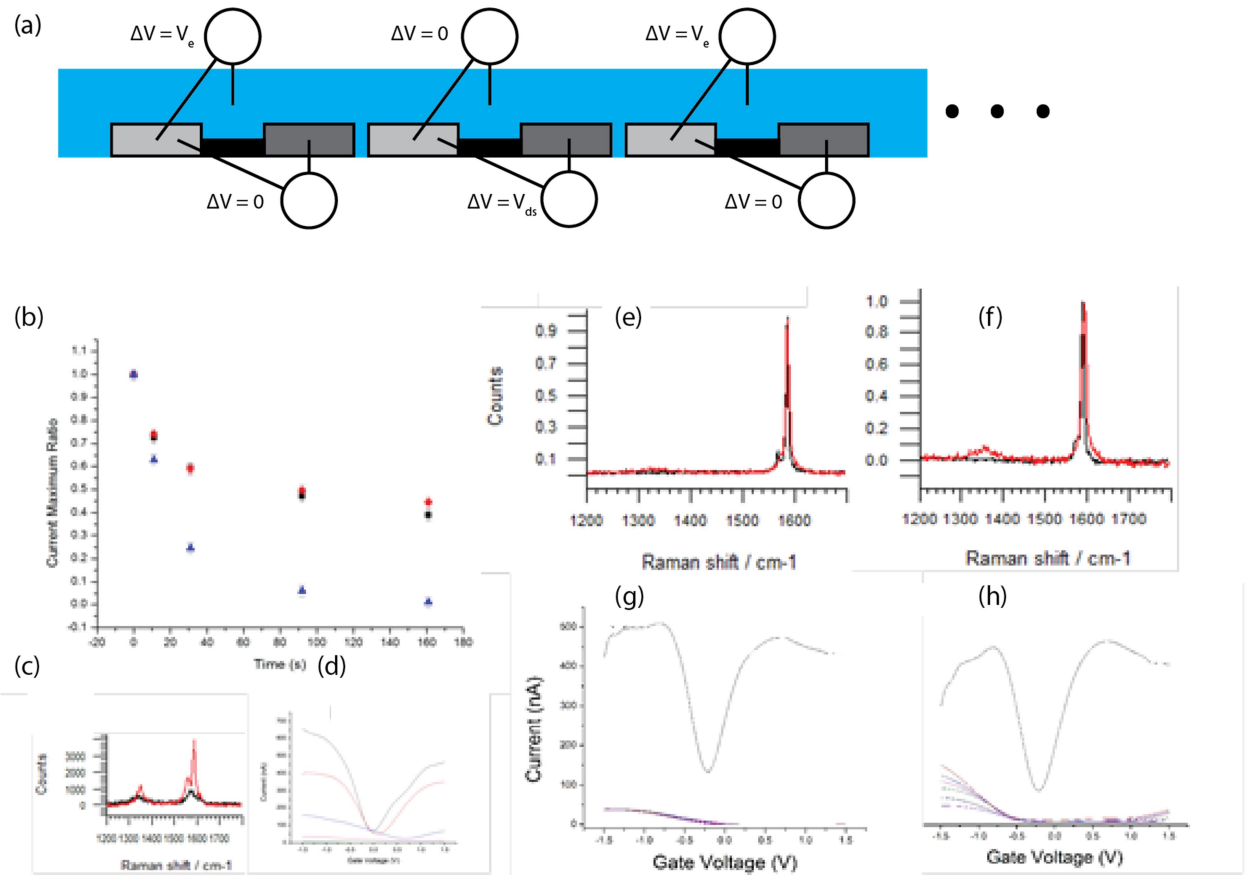


Figure 2.1 Electrochemical modulation affects both the rate and extent of diazonium-dependent conductance quenching.
(a) CNTFETs alternated between unbiased channels with a potential difference relative to solution and CNTFETs with a small source-drain bias and no potential difference relative to solution. (b) Kinetics of conductance quenching monitored by measuring the IV characteristics as function of total exposure time to FBDP. Blue has V_e=+1V, black and red are spatially adjacent CNTs with V_e=0. (c) Raman spectra of V_e=1V (black) and V_e=0V (red). (d) IV characteristics taken at the time-points shown in (b) for V_e=1V. (e) Raman spectra for V_e=-1V, before and after functionalization. (f) Analogous Raman spectra for V_e=0V. (g) IV characteristics for the time-points in (b), for V_e=-1V. Kinetics are too fast to measure. (h) Analogous IV characteristics for V_e=0V. Error bars in kinetic plots are from repeated measurement (#scans = 5) of IV characteristics.

While electrochemical control of geometric reaction distribution appears to be a promising method the fabricate smFETs, a more straightforward synthetic route is to simply cover the nanotube so that only a small number of reactive sites are exposed to the reagent. This method, using nanowells, is described below.

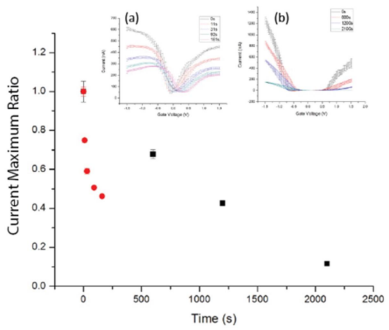


Figure 2.2 CNTFET bandgap affects kinetic rates of FBDP conductance quenching. (a) Small bandgap CNTs have slower kinetic rates than (b) large bandgap CNTs, as monitored by the IV characteristics measured after exposure to FBDP.

2.3 Pyrene Adsorption to CNTs

Since 2001, it has been well-understood that pyrene moieties, in particular pyrene-N-hydrosuccinimide ester (pyrene-NHS) can be used to deliver any sort of molecule, in particular protein or nucleic acid, to the surface of a CNT[66]. Since noncovalent attachments based on pyrene adsorption to CNT sidewalls have been used by others as the basis of smFET design[20], I reasoned that by tuning the concentration of primary amine moiety, it would be trivial to find an assay condition where the inter-molecule distance on CNTFET surfaces would roughly match the dimensions of the channel. To find this assay condition, I held all other parameters constant and varied the concentration of 10nm gold nanoparticles (Nanocs) bearing polyethyleneglycol (PEG) units terminated with primary amine handles, referred to as AuNPs. To quantify, the distribution of heights associated with the AuNPs was identified using the vbscope model (see Chapter 2) and the number of molecules within 3 pixels of a nanotube was counted by hand. The optimal protocol was a 15 minute exposure to 100mM pyrene-NHS dissolved in DMSO followed by a thorough DMSO rinse and 90 minutes exposure to 400nM AuNPs in 10mM phosphate buffered to pH 8.4; since each nanoparticle had more than one amine group, the protocol was adjusted accordingly to our labeled RNA which only has one labile primary amine (~10 μM). For example, in Figure 2.4a, 9 particles are seen on a 16 micron nanotube; correcting for background coincidence in the image (5 particles) as well as nonspecific

adsorption (1 particle, Figure 2.4b), we come to 3 particles/12 microns in this image. The total image area was 400 μm^2, the total imaged CNT length was 48 μm, the total number of particles imaged was 2231, the number of particles on CNTs was 38, leading to an estimate of 4.3±0.8 μm/particle. An equivalent area was measured (no pyrene-NHS) on a separate chip.

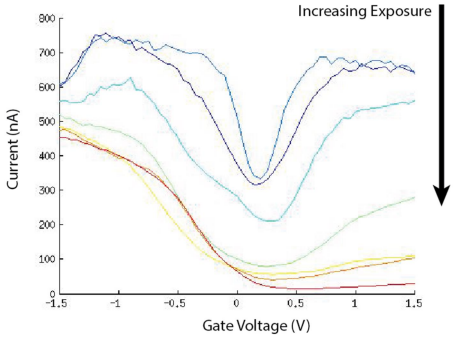


Figure 2.3 BHT, a radical scavenger, does not prevent FBDP-mediated reactions on CNT sidewalls. Reactions in the presence of 10mM BHT, as described in the text, and monitored by the IV characteristic.

2.4 Fabrication of single-molecule field effect transistors using sequential reactions within nanowells[3]

A special case of the solutions discussed above is that if $N = 1$ then there is not a local maximum; instead the probability of a single functionalization event rises rapidly to certainty. This is equivalent to the absorbing-state master equation solved in Chapter 2. Therefore, since the diazonium adduct must add to a specific position on the nanotube surface, it is reasonable to suppose that using nanolithographic methods, the exposed area of the nanotube surface could be confined to a handful of reactive sites and, with optimization, perhaps just one. Even if this specific goal is unattainable, lowering the effective length of the nanotube will still have the effect of raising the smFET fabrication probability, as discussed above.

[3] With Dr. Delphine Bouilly and Dr. Nathan Daly. Section is adapted from work published in *Nano Letters* in 2016.

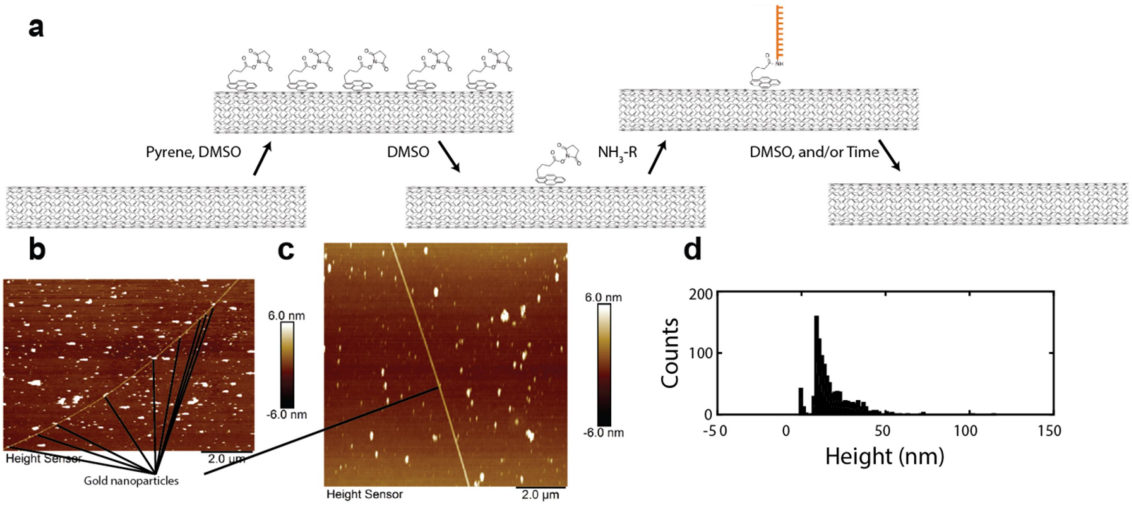


Figure 2.4 Noncovalent smFET fabrication cycle
(a) Pyrene moieties, in this example bearing an NHS handle, are nonspecifically adsorbed onto the CNT surface (as well as everywhere else). DMSO is used to wash away the residual pyrene. Subsequently, a molecule attached to a primary amine is introduced to the solution causing it to conjugate to the pyrene. Subsequently, after time has passed or addition of DMSO, the last remaining pyrene will eventually desorb as well. (b) Example CNT that has gone through the first four steps with ~10nm gold nanoparticles. (c) Same, except without addition of pyrene-NHS, as a control. (d) Height profile of all the replicate images from (b), generated with the help of the vbscope model.

A second problem with the above approaches is that there is no true negative control – all the CNTs are exposed to all the reagents. The nanolithographic process that allows for design and production of nanowells allows for positive controls because CNTs can be entirely covered instead of exposed within a nanowell, allowing for verification that functionalization with certain reagents is required for signals to manifest.

A third problem is that the charge distribution becomes increasingly inhomogeneous as the gate bias is more extreme relative to the source-drain bias. Assembly of the smFET using a nanowell allows this variable to be controlled because the molecule will eventually attach to the same place in the nanotube for every recording even on different chips.

Finally, as compared with noncovalent functionalization methods, nanowells allow for covalent attachment of biomolecules to the CNT, increasing the time available for experiment 10-fold. This is especially important for measurements of G-quadruplexes, which have kinetic rates that must be measured for several hours in order to gain precision and characterize dynamic heterogeneity.

2.3.1 Results

A mask was designed following the schematic shown in Figure 2.5a. The size of the nanowell was varied between 0 (positive control) and 4 microns, with an example 20nm nanowell shown in Figure 2.5b. The CNTs were then exposed to paracarboxydiazonium (CBDT) overnight in 100mM PO_4 pH 8, likely resulting in a free-radical mediated grafting of the reagent to all available reactive sites[57]. Following, approximately 40 CNTs from each nanowell size were probed for conductivity using the Si as a back gate, results shown in Figure 2.5c. To analyze this data, the number of CNTs with equal or greater conductance from their initial probed values were counted, using the expectation value from a lognormal distribution to calculate the G/G_0 ratio threshold (Figure 2.5 c, d). Assuming that the number of reacted sites is Poisson-limited, $\lambda^n e^{-\lambda}/n!$, this count is used to estimate the Poisson parameter for each nanowell size. Shown in Figure 2.5e is the optimal size, 20nm ($\lambda = 1.39$, 20% conductance drop; theoretical optimum $\lambda = 1$) , compared with the next size up, 40nm.

As an additional test, it was found in previous studies[43] that addition of 50μM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) to CNTs functionalized with carboxy groups led to fluctuations between two conductance classes. Following exposure of CNTs below 20nm nanowells (Figure 2.6a) or entirely covered CNTs with CBDT, ~30% of CNTs (2 out of 7) with 20nm nanowells showed fluctuations between two conductance classes (Figure 2.6c), and none of the positive controls showed such fluctuations (predicated on not showing quantal fluctuations in buffer without EDC.) These fluctuations were observed to have dynamic heterogeneity – a mixture of two periods, one in which fluctuations between the two conductance classes were common and one in which fluctuations were rare. These fluctuations were interpreted as follows: the periods with rare fluctuations were likely CNTs without EDC, whereas periods with fluctuations were interpreted as internal rearrangements of EDC bound to the carboxylic acid functional group. The average lifetime of the fluctuations is shown in Figure 2.6d. Taking a running 1-sec window, these fluctuations were characterized using a Hidden Markov Model (HMM) in each window, and the average lifetimes and frequency of each equilibrium constant are shown in Figure 2.6 e and f.

2.3.2 Discussion

Chemistry in nanowells allows unprecedented geometric control over the reactive sites as well as considerable choice of chemical handles. The yield of the method was high, approaching 35% both predicted from the conductance distributions as well as measured empirically by checking for two-level fluctuations using EDC. Nanowells can be fabricated on the wafer scale. Furthermore, additional chemical reactions can be continued on the same chip allowing fabrication of smFETs.

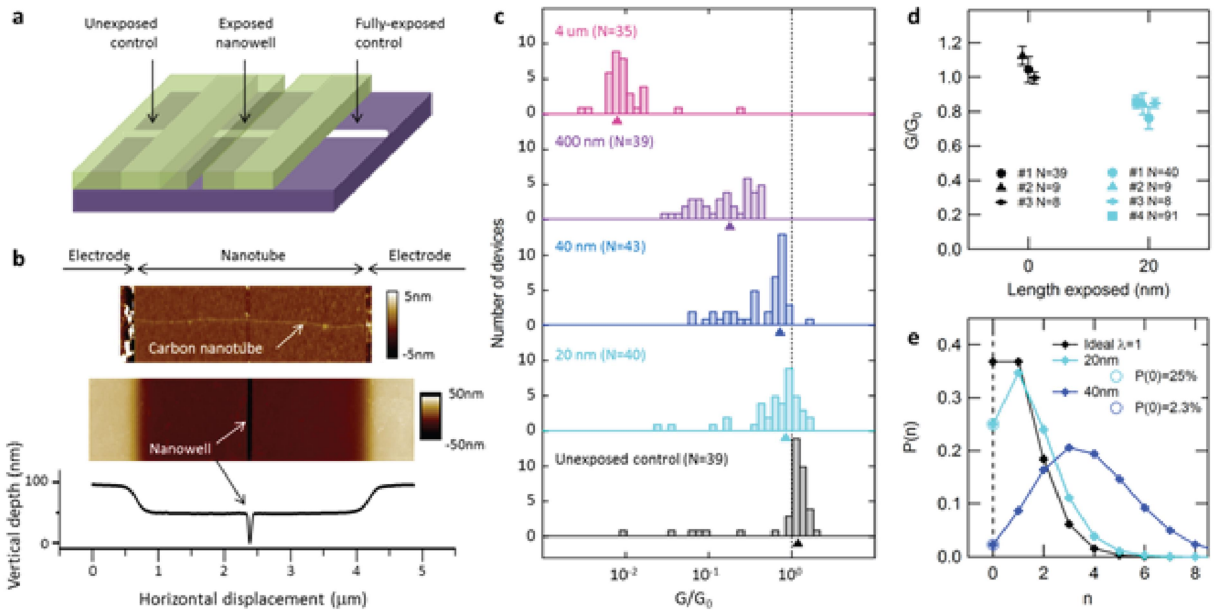


Figure 2.5 Effect of nanowell-confined chemistry on carbon nanotube devices.

a. Mask design. A small window is opened in the channel. A positive (entirely exposed) and negative (entirely covered) control is included as part of the experimental design for Raman spectroscopy. **b**. AFM characterization of nanowells. **c**. Distribution of conductance change G/G_0 after CBDT exposure within nanowells of different sizes, compiled on N individual devices from the same nanotube ($N_{total} = 196$). Arrows indicate the expected conductance ratio from a log-normal fit. **d.** Conductance ratio using 20 nm nanowell masks compared to control devices following CBDT exposure, as in **c.** averaged on N (indicated) devices from distinct CNTs. **e.** Probability of n functionalizations given a nanowell width. Circles represent measured points based on data in panel c; others extrapolated from a Poisson model. This figure is adapted without modification from Bouilly, D., et al (2016). Single-Molecule Reaction Chemistry in Patterned Nanowells. *Nano Letters*, *16*(16), 6–12. Link: <http://pubs.acs.org/doi/abs/10.1021/acs.nanolett.6b02149>. Further permission requests to

reproduce this figure should be directed to the ACS.

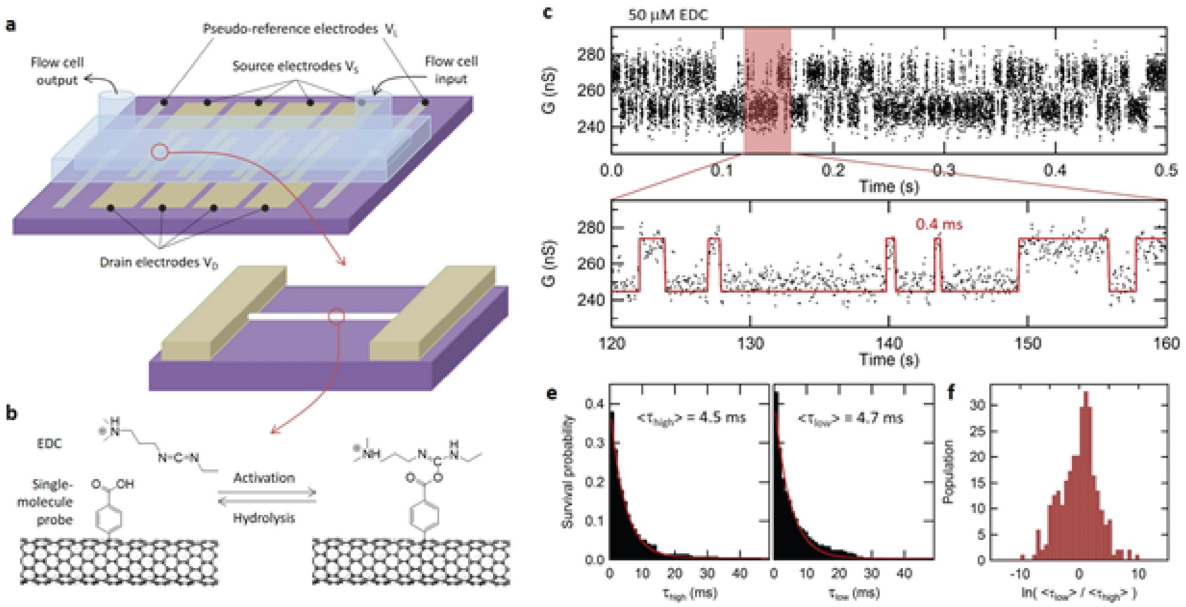


Figure 2.6 Real-time sensitivity to secondary reactions on the primary single-molecule probe.
a. Design of smFET devices used in this study. **b.** CBDT-nanotube interacting with the carbodiimide group of EDC. **c.** Real-time response of a device in the presence of 50 µM EDC, after baseline correction, showing an active phase with two-state activity characteristic of rapid fluctuations in a single carboxy-EDC adduct. **d.** Zoomed trajectory and Viterbi trajectory obtained using parameters from a Hidden Markov model, revealing sub-millisecond fluctuations between two conductance classes. **e**. CDF of dwell-times for the high and low conductance classes, fitted using a single-exponential model to obtain average lifetimes $<\tau_{high}>$ and $<\tau_{low}>$. **f**. Dynamic heterogeneity of EDC adduct, roughly described by measuring the average rate constant. This figure is adapted without modification from Bouilly, D., et al (2016). Single-Molecule Reaction Chemistry in Patterned Nanowells. *Nano Letters*, *16*(16), 6–12. Link: <http://pubs.acs.org/doi/abs/10.1021/acs.nanolett.6b02149>. Further permission requests to

reproduce this figure should be directed to the ACS.

2.5 Reaction of CNTFETs with diazonium salts creates a point of heightened sensitivity[4]

The basic scheme whereby diazonium salts react with CNTs is described above. However, in all the measurements described, it is difficult to tease apart the effect of a covalent modification on the spatial density of carriers across the nanotube, as well as to argue directly from the data that the point at which a sp^3 defect in the generally sp^2 CNT lattice will be of subsequent use as an attachment site for smFET fabrication. To make these arguments directly, I reasoned that the conductance response of a CNT could be mapped by passing an AFM tip (SCM-PIT, Bruker) over the surface as a mobile capacitor, a technique known as scanned gate microscopy (SGM)[38,67], and recording the current map, I(x,y), translating that into the CNT coordinates as I(z). Finally, I designed an experiment using nanowells to place a reactive site at a predefined location, so that it could be subsequently mapped by SGM to directly visualize the effect of a point defect on the conductance response. In a limited number of cases, this response was centered at the putative defect site, directly explaining how in such cases an smFET interacts with the carrier distribution in the CNT.

The general method for measuring the response of CNTs to a local scanned gate is schematized in Figure 2.7. The data was analyzed with a diffusing Lorentzian, fitting their current map I(x,y) so that it could be deconvoluted into CNT coordinates, I(z), see Appendix B.

[4] Electronics designed by Scott Trocchia including the printed circuit board used to record the measurements.

Two methods have been discussed previously – dcSGM, which scans a biased tip at a predefined height, and tmSGM, which scans a biased tip at a sine wave about a predefined height with frequency equal to the tip resonant frequency[68]. Comparing the response of dcSGM and tmSGM directly (Figure 2.8) and determined that for CNTs with high baseline current, enough to populate frequencies equal to the resonant frequency of the tip, tmSGM has better sensitivity and spatial resolution than dcSGM. Unfortunately not every CNT has significant power at the 60-120kHz range to give a significant response even prior to functionalization with diazonium. Therefore the technique here is limited to those CNTs whose source-drain current, before and after reaction, is significant enough to be detected at high frequencies.

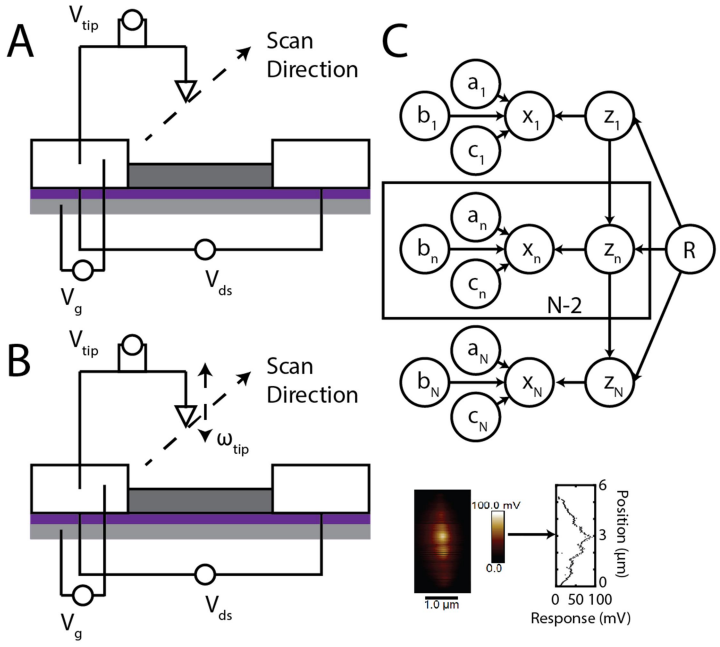


Figure 2.7 Collection and analysis of dc- and tmSGM data.
(A) In conventional SGM, a tip is held at a bias V_{tip} relative to the source; carriers flow as a result of a bias

V_{ds} between source and drain, and a gate bias V_g between the gate electrode and channel is applied as well. The tip is scanned to create a conductance map I(x,y). (B) In tmSGM, this is the same except the tip is oscillated at its resonant frequency and the map I(x,y) is filtered with a lock-in to that frequency. (C) Data analysis of tmSGM or SGM data consists of fitting an image line-by-line to lorentzian functions ; R denotes diffusion either from growth of the nanotube, it is not necessarily a line in the image, or continuous variation in I(x,y). Generally, what will be shown in what follows is the max current variation, *i.e.* the peak of the individual lorentzians as a function of the CNT coordinate.

Furthermore, tmSGM signals are proportional to the derivative of the capacitance rather than, as dcSGM signals, directly proportional to the capacitance[68]. Therefore it is necessary, when compiling maps, to carefully measure a wide array of tip-bias conditions in order to measure an accurate and representative map. Examples on pristine CNTs are shown in Figure 2.9. From a complete map, one can directly observe the Schottky barrier[69] which forms at the entry point of current carriers into the channel. The Schottky barrier results from different Fermi levels between the channel and metal at the junction, and therefore maps the direction of current. In Figure 2.9a, the current flows from top to bottom whereas in Figure 2.9b, the current flows from bottom to top. The CNT shown here is ambipolar and semi-metallic, and Figure 2.9a/Figure 2.9b are two separate fragments of the same CNT. A large amount of peaked variation localized distal to the Schottky barriers is in most cases observed and, as the tip bias is brought to extremes, the landscape of the

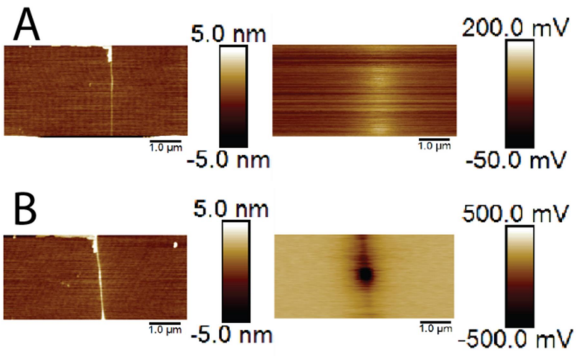


Figure 2.8 Comparison of dcSGM signal with a tmSGM signal on the same CNTFET. (A) dcSGM signal. (B) tmSGM signal. The tmSGM signal has better spatial resolution and 20-fold higher response.

CNT changes, with the addition of peaks in locations that did not previously possess them. A possible explanation for this is that at the extreme potentials, $|V_{tip} - V_{bg}| < 2V$ for instance, secondary effects such as accessibility of previously inaccessible current modes or direct contact between the tip and CNT can occur. Defining, therefore, the biases between $|V_{tip} - V_{bg}|$<2V as the region of linear response, one can see that even within this radius, the positive and negative biases are not strictly mirror images of each other, though they are close. This slight variation is likely a result of the long acquisition time required to generate these images – close to 8 hours for each full dataset in Figure 2.9 – though direct localization is further complicated by the fact that at high tip biases the surface itself will deflect the tip at an angle as it is scanned. Direct transfer of current carriers into the CNT through the tip is unlikely as this is not observed when the metal contacts are scanned.

Aside from the Schottky barrier, none of the variations observed in the I(z) map of these CNTs are *a priori* predicted. Two separate hypotheses can be put forward, neither of which is easy to directly test. First, the CNT could possess, already, numerous defects in its surface[70]. Indeed, CNTs that were not straight and had kinks were observed to possess SGM sensitivity peaks at those kinks[68]. Second, the CNT response could be a function of the wavefunction of the particular mode that dominates at room temperature, which could have peculiar spatial variation, as predicted from the theory of 1D conductance[23,71]. However, because the chirality of the CNT

in many of these cases is unknown and the CNT-metal junction difficult to model even were it known, and further, because defects in the CNT are likely and perhaps even generated by the tip itself[39], comparing theory to experiment at this stage possesses its own challenges.

Therefore, an experiment was designed to directly observe the resulting carrier response to introduction of diazonium covalent adducts to the CNT sidewall at geometrically defined positions, so that the potential response could be mapped when defects of a known type were intentionally introduced. First, the CNT was scanned at one potential using tmSGM (to preclude tip-mediated defect introduction). Second, a nanowell, described in the previous section, was introduced into the center of the channel and the chip was immersed in 10mM carboxydiazonium (CBDT) at aqueous pH 8 overnight, including positive and negative control regions for Raman spectroscopy. The acetonitrile chemistry described above could not be used because the acetonitrile solvent was unfortunately found to dissolve the PMMA window, and therefore as it took place in aqueous conditions the reaction likely proceeded *via* a free-radical mech-

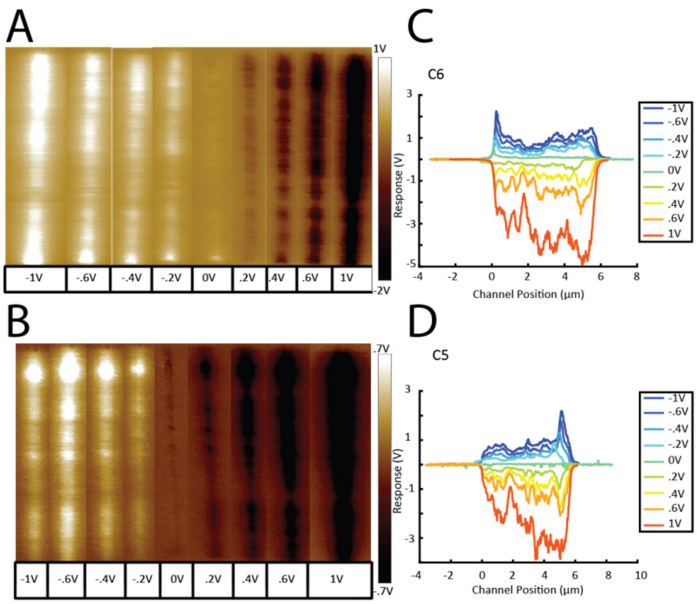


Figure 2.9 tmSGM signals on pristine CNTs. (A) and (B) show the raw data for each CNT at the various potentials, whereas (C) and (D) show the I(z) in CNT coordinates as a function of channel position, following analysis with the model described in Figure 2.7.

anism described elsewhere[57] leading to the introduction of covalent diazonium grafts to the CNT sidewall within the confines of the nanowell. The chip was subsequently washed, wired, and mapped again using tmSGM, this time at numerous bias potentials, shown in Figure 2.10. A highly significant response was observed in the region where the nanowell was placed, which did not previously exist in the channel (Figure 2.10b, c). This modulation was likely associated with covalent reactions with the CNT sidewall because comparison of the Raman spectra of positive and negative control regions revealed a D-band peak[72] in the unexposed region (Figure 2.10e). Furthermore, while thermal annealing at 350C in vacuum removed the D-band peak in this control window, subsequent tmSGM images compiled after the anneal step continued to reveal localized sensitivity at the point where the nanowell was introduced, indicating that the defect was introduced in such a way that the CNT could not "self-repair" [73] (Figure 2.10d).

Careful examination of the tmSGM response reveals that the peak shown in Figure 2.10c is not symmetric – like the Schottky barriers themselves, the response peak presents an initial

barrier and subsequently tapers off in the direction of current. This result is anticipated from the theory of 1D conductance: the covalent perturbation presents a barrier which has the effect of purifying the carriers and enriching modes that possess a higher energy and therefore traverse the barrier. This perspective is supported by the fact that much of the natural variation in this CNT has vanished subsequent to introduction of the peak. The mode structure in the channel has changed, and the "ground" set of states observed in the pristine CNT is no longer the dominant current carrying family. Either these states are at a high enough energy level that the kinks are invisible or their spatial variation is distinct from the "ground" set of states.

These results in general give a framework for how smFETs prepared using covalent adducts to the CNT sidewall subsequently sense the conformational change of single molecules. smFETs prepared using the conditions describe here likely have a point of dominant sensitivity which is also the point at which the CNT is most reactive, if chemical handles are introduced concurrently to the diazonium functionalization. Therefore when a molecule is attached to the CNT, it does so at a point that, because of the covalent attachment, is the most sensitive to changes in charge composition.

How then, do smFETs designed around noncovalent attachment function? The answer is shown in Figure 2.9 – my argument is that the CNT possesses enough natural variation, either due to its peculiar electronic structure of pre-existing points of sensitivity owing to defects introduced during growth or fabrication, that noncovalent attachment at one of these points by chance may give enough signal for measurement of intra- or inter-molecular interactions through its use as a point site of a noncovalent graft.

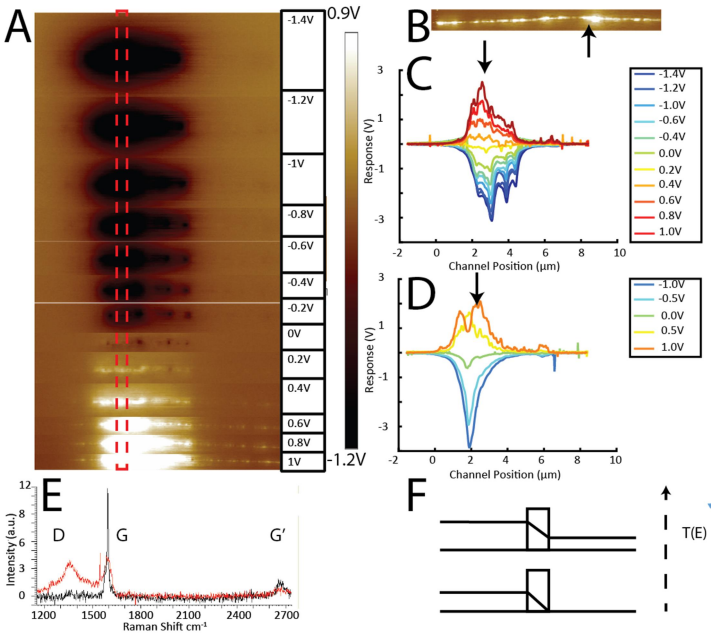


Figure 2.10 Precise geometric introduction of sp3 defects using diazonium salts.

(A) I(x,y) maps at various tip biases, indicates to the right, following exposure to carboxybenzenediazonium, as described in the text. The red box indicates the approximate location of the nanowell. (B) I(x,y) map of the same device before exposure to diazonium. The arrow indicates the eventual location of the nanowell. (C) I(z) map after exposure to diazonium. The arrow indicates the location of the nanowell. (D) I(z) map after annealing the to remove diazonium adducts. The arrow indicates the location of the nanowell. (E) Raman spectra of the positive (red, fully exposed) and negative (black, fully covered) control locations. (F) Argument for peak shape – higher energy conductance states have a larger probability of passing through a barrier created by an sp3 defect than lower energy conductance states.

Chapter 3 Single-Molecule Computational Methods

3.1 Introduction

Single-molecule observable *versus* time trajectories, from recordings of quantum transitions between the polarity of the magnetic moment of an isolated electron[74] to recordings of ribosomal fluctuations in solution[75], bear strong mathematical analogies. Because the timescales of different types of fluctuations are widely separated, to the experimentalist recording the observable *versus* time trajectory of the isolated molecule, the molecule appears to spontaneously and discontinuously jump from one state into another; whereas in reality, the molecule continuously varies between its competing atomic configurations in what may be conceptualized as a series of mechanistic steps. Examples of such trajectories may be found in Chapters 2, 4, and 5.

The principal problem facing analysis of single-molecule data is reverse engineering: using a trajectory or group of trajectories collected in an experiment, one must simultaneously discover the set of dynamic equations that could possibly have led to the data and estimate the parameters that tune the frequency of transitions between discrete states. In many cases, this amounts to taking the observation of discrete states of a label, assigning those states to configurations of the molecule, and measuring the rates of interconversion between the various states. The trajectory or collection of trajectories is understood when the number of states has been counted and the rates of interconversion between those states have been measured. In this thesis, I identify and computationally address three specific problems associated with the analysis of single-molecule data.

First, to accurately reconstruct the dynamic equations and therefore the behavior of the ensemble, a first step is rigorous identification of molecules. The prototypical case described here is the identification of isolated molecules from an optical microscopy image. In this case, the challenge is finding all subpopulations of chromophores in an image without systematically missing any given types. The problem of, in an unbiased way, computationally locating all the molecules in a series of observations of single-molecules in an optical microscopy image is referred to as the molecule identification problem.

Next, on the level of the observable *versus* time trajectories, as a result of timescale separation, the system is assumed to possess no history at a coarse enough timescale, so that the time at which measurement began is in many experimental designs arbitrary. Therefore it is typically assumed that the equations governing dynamic evolution of the trajectory come from the family of Markov jump stochastic differential equations. Because they are discretely sampled by some apparatus, trajectories generated by such continuous-time equations are typically modeled using discrete-time Markov chains. However, because of technical limitations inherent in current technology, the observable in single-molecule experiments does not typically report the position and momentum of every atom in the system. Therefore, while the system in general may be adequately modeled with a Markov chain with respect to all its components resulting for example from numerous and rapid thermal fluctuations mediated by collisions with the environment, with respect to the limited observable, the trajectory may appear to have dynamic organization or patterns that result from rearrangements in parts of the molecule that are hidden from the measurement. Such patterns can manifest as sudden and random changes in the frequency of transition between states, and the problem this presents for analysis of single-molecule data is referred to as the dynamic heterogeneity problem.

Lastly, in conductance *versus* time trajectories collected using single-molecule field effect transistors (smFETs), the amplitude and fluctuations of the observable arise from many physical processes distinct from the fluctuations of the molecule of interest. These separate fluctuations, while unrelated to the molecular motions under study, nevertheless diffuse as well. This leads to a fluctuation in the amplitude of the observable associated with a given state, as both the conformational changes of the molecule and the environmental fluctuations are, in this example, detected using the same electric current, and the problem this presents for the analysis of single-molecule trajectories is referred to as the emission drift problem.

In this chapter, three computational algorithms are presented which offer solutions to the molecule identification, dynamic heterogeneity, and emission drift problems. They present significant improvements to current techniques tackling the same issues. All the algorithms described here utilize Bayesian inference using the variational approximation. Algorithms utilizing Bayesian inference make use of the insight that all measurements have some form of uncertainty, and the technique uses specific assumptions about the error inherent in each individual observation in order to measure, instead of just the parameters, a probability distribution over all possible parameters. This is known as the posterior distribution[76]. This stands in contrast to a maximum likelihood approach which endeavors to measure point estimates of parameters only[77–80]. When applied to the molecule identification problem, the unified approach presented in this chapter allows complete identification of isolated molecules in optical microscopy images, registration of disparate imaging channels, and superior intensity estimation. This ultimately leads to a two-fold increase in the signal-to-noise of the subsequent chromophore intensity *versus* time trajectory[17,81,82]. When applied to the dynamic heterogeneity problem, the approach presented in this chapter leads to rigorous quantification of the rates of

change of dynamic behavior, which have previously been unquantifiable[77–80], even those utilizing a Bayesian inference framework[83–85]. Finally, when applied to the emission drift problem, by integrating previous approaches[86] with Markov models, the approach presented in this chapter leads to a four-fold increase in the accuracy of the quantification of rates with respect to the amount of diffusing noise corrupting the observable *versus* time trajectory of the molecule over current approaches that do not account for emission drift. Together the solutions to these problems allow for, in particular, rigorous quantification of the conductance *versus* time trajectories of molecules attached to field effect transistors, as described in Chapters 2, 4, and 5.

In summary, this chapter describes the development of single-molecule computational methods that find, pick apart, and quantify observable *versus* time trajectories. I begin with a theoretical description of the information these trajectories contain--in particular, I outline continuous, Markovian dynamics. Next, I describe the theoretical computational framework, based on the variational approximation to Bayesian inference, used for all practical applications in this chapter. Finally, I describe each of the methods representing contributions from this thesis work – for the molecule identification problem, for the dynamic heterogeneity problem, and for the emission drift problem – utilizing the variational approximation to Bayesian inference.

3.1.1 Dynamics

This thesis is concerned with “rates” between “states”. A “state” may be defined as an equivalence class of atomic configurations that a system may take, and a transition may be defined as a jump between equivalence classes. Physically, a molecule jumps between states when there exist mixtures of timescales so that the configurational subspaces associated with the wait times between transitions can be decomposed and considered independently from one

another. An account in terms of the precise size of the minima of a rugged multidimensional energy landscape is given in Langer *et al*[87].

In general, investigators typically assume that, as a result of numerous and rapid thermal fluctuations coupling the environment to a molecule, the configuration of the molecule at a given instant only depends on its most recent previous configuration. A caveat, however, is that an experimental apparatus does not measure any property in continuous time, but instead regularly samples a continuous process with some time-resolution. This is the difference between a continuous time Markov process, which characterizes the actual molecular motion, and a discrete time Markov chain, which characterizes what is actually recorded by an apparatus. From a practical point of view, single-molecule observable *versus* time trajectories are constrained to be the latter, while generated by the former; the goal of analysis is to estimate the parameters of the continuous time process using its recorded representation as a discrete chain. In this subsection, which forms preliminary material for the practical material subsequently presented, the relationship between the discrete and continuous models is examined in detail. Following this largely mathematical discussion, the theoretical technique underlying the computational algorithms, which has been previously developed by others (see [76] for an alternate exposition), is introduced in generality. The overall goal to this section is to provide the mathematical underpinnings to the computational models developed below, which utilize Hidden Markov Models (HMMs) in a Bayesian inference setting using the variational approximation, as well as to develop the underpinnings necessary to describe the computational model developed for fluorescence imaging analysis.

3.1.1.1 Continuous-time Markov Processes

In this section I will describe a conventional, empirical, non-equilibrium, dynamical theory that will be henceforth be used to discuss the experiments described in subsequent chapters. Specifically, I will discuss the Markov jump master equation, synchronicity, equilibrium and approach to equilibrium, fractional occupancies and distribution measures over ΔG values in dynamical systems, discrete likelihoods for dynamical state spaces which jump between states and arbitrary conditions, and conclude with analytical results on tridiagonal matrices. These properties all manifest as the result of the dynamics of a continuous-time Markov process.

Equilibrium and non-equilibrium dynamics can be abstractly considered in the context of a transition function which governs the probability of exchange of the system between a finite or countable number of discrete states. I will loosely follow the exposition presented by Todorovic[62]. This level of theory is sufficient to describe dynamics that are well-separated, such as binding events, microsecond-level spontaneous or factor-dependent conformational changes, and in the small barrier limit, transport. Suppose there is some state space $\{\Omega_i \subset \Omega \mid \Omega_i \cap \Omega_j = \emptyset, i \neq j, (i,j) \in \mathbb{N}\}$ which is here either finite or countably infinite. The transition function is defined by

$$A(t,s) \equiv \left\{p\left(t,s,\Omega_i,\Omega_j\right) = p\left(\Omega(t) = \Omega_i \mid \Omega(s) = \Omega_j\right), t > s \geq 0\right\}$$

and is called homogeneous if

$$A(t-s,0) = A(t,s), t > s \geq 0$$

A process is Markovian if the transition function defined above equals the conditional describing its entire history. I will assume that there is an $\epsilon > 0$ such that $p\left(t+\epsilon,t,\Omega_i,\Omega_j\right) > 0$ and that

the transition functions are a non-Abelian (i.e. time ordered and therefore non-commutative) semigroup according to the Kolmogorov-Chapman equation:

$$p\left(t,s,\Omega_i,\Omega_j\right)=\sum_k p(t,0,\Omega_i,\Omega_k)p\left(s,0,\Omega_k,\Omega_j\right),t>s\geq 0$$

3-1

which allows the semigroup condition to be satisfied for the transition function, considered as a matrix:

$$A(t+s,0)=A(t,0)A(s,0),t>s\geq 0$$

It can be shown[62] with these conditions the associated semigroup admits a generator known as the rate matrix:

$$Q\equiv\left\{q_{ij}=\lim_{t\to h+}\frac{p\left(t,h,\Omega_i,\Omega_j\right)-\delta_{ij}}{t+h},(i,j)\in\mathbb{N}\times\mathbb{N}\right\}$$

3-2

For example, a homogeneous rate matrix may be constructed by setting $\sum_{j\neq i}q_{ij}=-q_{ii}$, setting $0\leq q_{ij}<\infty$, and setting $\frac{dQ}{dt}=0$, which collectively give rise to a homogeneous transition function. If there are a finite number of states, which is true in many practical cases, and no state is absorbing, that is, $-q_{ii}<\infty$ for all i, and the rate matrix is homogeneous, then:

$$\frac{dA_{ij}}{dt}\equiv\lim_{h\to 0}\frac{A_{ij}(t+h,0)-A_{ij}(t,0)}{h}=\sum_k p(t,0,\Omega_i,\Omega_k)\left(\lim_{h\to 0}\frac{p\left(h,0,\Omega_k,\Omega_j\right)-\delta_{kj}}{h}\right)$$

$$=\sum_k p(t,0,\Omega_i,\Omega_k)\,q_{kj}=\sum_k A_{ik}\,q_{kj}$$

which is known as the forward equation. However, this equation could also be defined in the dual space:

$$\frac{dA_{ij}}{dt}=\sum_k\left(\lim_{h\to 0}\frac{p(h,0,\Omega_i,\Omega_k)-\delta_{ik}}{h}\right)p\left(t,0,\Omega_k,\Omega_j\right)=\sum_k q_{ik}p\left(t,0,\Omega_k,\Omega_j\right)=\sum_k q_{ik}A_{kj}$$

which is known as the backward equation. These equations immediately give rise to the master equation for the evolution of some occupancies associated with the state space $\rho(t) = \{p(t, \Omega_i) \text{ for all } i = \{1 \dots n\}\}$:

$$\rho(0)\frac{dA}{dt} = \frac{d\rho(t)}{dt} = \rho(0)A(t)Q(t) = \rho(t)Q$$

3-3

When the rate matrix is homogeneous, the master equation has the formal solution associated with a measurement at $t = 0$:

$$\rho(t) = \rho(0)e^{tQ}$$

3-4

There are three extensions worth mentioning which find no use in this thesis and are therefore not developed here. A simple extension of these arguments, i.e., requiring them to be self-referential, allows derivation of the chemical master equation, for when Q depends on the size of the overall population. Secondly, the master equation on a finite or countably infinite state space can be easily extended to the Fokker-Planck, or sometimes Smoluchowski, equation[88] to describe diffusion over a continuous state space – in this case a rate matrix is not sufficient and the equation takes the flavor of an inhomogeneous wave equation. Finally, I note that these equations are all results on classical probabilities of which an underlying density matrix formulation may be constructed instead by allowing for complex amplitudes in a quantum mechanical framework, for example, using the Lindblad equation[89] or Generalized Master equation[90] which, again, finds no use here.

As a simple example, consider a two state Markov process with a rate matrix given by:

$$Q = \begin{bmatrix} -k_{12} & k_{12} \\ k_{21} & -k_{21} \end{bmatrix}$$

In this case, solving the master equation is trivial, since:

$$Q^k = \left(-(k_{12} + k_{21})\right)^{k-1} Q$$

$$\rightarrow e^{tQ} = \sum_{k=0}^{\infty} \frac{t^k}{k!} Q^k = I - \frac{1}{k_{12} + k_{21}} \sum_{k=1}^{\infty} \frac{(-1)^k (k_{12} + k_{21})^k t^k}{k!} Q$$

$$= I + \frac{1}{k_{12} + k_{21}} Q - \frac{e^{-t(k_{12}+k_{21})}}{k_{12} + k_{21}} Q$$

where I is the identity matrix. When at least one of these rates does not vanish, synchronizing the process into the first state gives:

$$\rho(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left[I + \frac{1}{k_{12} + k_{21}} Q - \frac{e^{-t(k_{12}+k_{21})}}{k_{12} + k_{21}} Q \right] = \begin{bmatrix} \frac{k_{12} e^{-t(k_{12}+k_{21})} + k_{21}}{k_{12} + k_{21}} \\ \frac{k_{12}\left(1 - e^{-t(k_{12}+k_{21})}\right)}{k_{12} + k_{21}} \end{bmatrix}$$

Note that this process has a well-defined equilibrium distribution, which is reached regardless of the initial conditions (which can be shown with an arbitrarily normalized initial distribution):

$$\lim_{t \to \infty} \rho(t) = \begin{bmatrix} \frac{k_{21}}{k_{12} + k_{21}} \\ \frac{k_{12}}{k_{12} + k_{21}} \end{bmatrix}$$

On the other hand, in the limit where one of the two states is absorbing – ie, after entering that state, there are no transitions out – we can write:

$$\lim_{k_{21} \to 0} e^{tQ} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} e^{-tk_{12}} & -e^{-tk_{12}} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} e^{-tk_{12}} & 1 - e^{-tk_{12}} \\ 0 & 1 \end{bmatrix}$$

which gives a simple model for the population evolution near an absorbing state, such as an irreversible chemical reaction:

$$\rho(t) = \begin{bmatrix} e^{-tk_{12}} \\ 1 - e^{-tk_{12}} \end{bmatrix}$$

The preceding example illustrates three important concepts that are deeply entwined. The first is equilibrium – in the two state system, an equilibrium distribution was reached out of the balance of the flux in and flux out of the respective states. The second is first-passage near an absorbing boundary – when the second state had no way of transitioning out, it simply charged up at the expense of the first with a characteristic time occasionally called the "first passage time." The final concept, which I will use mainly as a graphical device, is synchronicity – when measuring ensembles of time-series, synchronizing them to a shared initial condition, as opposed to a random as-measured initial condition, can yield descriptive plots of time evolution. I will discuss each in turn.

For homogeneous rate matrices, equilibrium can be defined quite simply as the condition for which the following holds:

$$0 = \pi Q \qquad \textbf{3-5}$$

where π is called the equilibrium distribution. This is condition is derived by setting the derivatives to vanish in the master equation. There are two natural questions – how many equilibria are consistent with a matrix Q, and what conditions are necessary and sufficient for the system to eventually reach its equilibrium? If a continuous time markov process with a countable state space has the following properties – (1) every state can be reached from every other ("irreducible"), (2) the process continues to transition between states forever ("regular," or "non-explosive;" a more technical description is, the sequence defined by the times at which transitions occur, which is monotonically increasing, has no finite limit), (3) the process always eventually returns to every state ("positive recurrent") – then these are necessary and sufficient for the existence of a unique equilibrium state, i.e., the solution to $0 = \pi Q$, which, given:

$$\mu = \frac{1}{diag(Q)R}$$

where $R = \left\{E_i\left[\int_0^{R_i} ds\right], i \in \{1, \dots, n\}\right\}$ is the expected time it takes to return to a state after leaving it, is equal to:

$$\pi = \frac{\mu}{\sum \mu_i}$$

The collected conditions (1)-(3) are referred to as the conditions for the process to be ergodic (for a full account see [91]). Therefore, there is a strong relationship between the equilibrium distribution and the first-passage times associated with the rate matrix.

Clearly then, there is a quantitative connection between the concept of chemical equilibrium and the straightforward representation afforded by the equation above. This connection may be found by considering the long-run time in a state. For example, suppose, for simplicity, that N separate draws, i.e. independent molecules, from an equivalent synchronized point with an equivalent homogeneous rate matrix Q are given, i.e., a Gibbs-type state preparation, where each molecule follows the same overall laws yet remains random. At any given time, the fraction of draws from a state matches the formal solution of the master equation, and therefore the fraction itself follows:

$$\rho(t) = \rho(0)e^{tQ}$$

and more importantly, the number of molecules in any given configuration, n_i, at any given time follows:

$$p\big(n_i|\rho_i(t)\big) = \frac{\Gamma\big(\sum_j n_j + 1\big)}{\prod_j \Gamma\big(n_j + 1\big)} \prod_j \rho_j(t)^{n_j}$$

where Γ denotes the gamma function. While for simplicity, I have assumed that all the molecules begin with an equivalent initial condition, one could easily conceive of solutions where the initial condition was random or arbitrarily deterministic across the set of molecules. This latter equation is a multinomial distribution, and is interesting because it may be used to define the distribution about the expected free energy at equilibrium; first noting that we must invert:

$$p(\rho_i(t)|n_i) = \frac{p\big(n_i|\rho_i(t)\big)p\big(\rho_i(t)\big)}{p(n_i)}$$

Following observation of a given realization:

$$p(\{\rho_i(t)\}|\{n_i\}) = \frac{p\big(n_i|\rho_i(t)\big)p\big(\rho_i(t)\big)}{p(n_i)} = \frac{\prod_j \Gamma\big(n_j(t)+1\big)}{\Gamma(N+1)}\prod_j \rho_j(t)^{n_j}$$

This is a Dirichlet distribution. It can be used to derive the free-energy landscape distribution at a given temperature ($\beta \equiv \frac{1}{k_B T}$ where here and only here does T refer to temperature):

$$p(\{\Delta G_i(t)\beta\}|\{n_i\}) \equiv p\big(\{\log\big(\rho_i(t)\big)\}|\{n_i\}\big) = \frac{\prod_j \Gamma\big(n_j(t)+1\big)}{\Gamma(N+1)}\prod_j \frac{e^{\rho_j(t)\left(n_j+1\right)}}{\rho_j(t)}$$

3-6

Ergodicity is not required to derive this distribution and it therefore admits absorbing states, as does the master equation above, as shown in the example. This equation is nothing more mysterious than the law of mass-action.

Finally, first passage times and synchronicity, which are two concepts which are very closely related, will be briefly discussed. The first passage distribution of a given state is, given a master equation and initial state distribution, the distribution of the first time that state gets occupied. Calculating this distribution occurs by setting the state of interest as an absorbing state and calculating the rate of change of its occupancy given an initial distribution. Most questions

of interest regarding first passage times regard how a state is occupied following synchronization into a given state or after a given condition. For example, the recurrence condition involved in the calculation of the equilibrium distribution would consider first passage times following exit from a state $i \in \{1, \dots, n\}$ and terminating once more in the same state.

Having described the dynamic theory, I now begin working towards a description required for the practical contributions of this thesis in this section. These require, (1) inference on the conditional discrete-time master equation and (2) analytical results for occupancies of finite-state tridiagonal matrices.

3.1.1.2 Discrete-time Markov Chains

A countable set of sequential draws from a homogeneous, regular, continuous time markov chain has a representation as a discrete-time markov chain. This representation is closer to the practical world, in which regularly sampled measurements of a given chain are observed so that the process appears to evolve in discrete time although the underlying physical process evolves in continuous time. In this case, the time evolution for the master equation, can be transformed:

$$\rho(t) = \rho(0)e^{tQ} \rightarrow \rho(t + \Delta t) = \rho(t)e^{\Delta tQ} \equiv \rho(t)A$$

where A is known as the transition matrix associated with the sampling time resolution Δt. The discrete case is important to consider because it describes what many measurements, such as single-molecule measurements described in this thesis, directly observe. This requires, in general, derivation of the distribution governing the likelihood of the various realizations of a physical process.

I will first derive the likelihood function in the simplest case – where there is a single rate matrix and therefore a single transition matrix. Suppose that a family of independent sequences

of states $\{z_{tn}, t \in 1 \dots T_n, n \in 1 \dots N\}$ has been given and each observation in each sequence has been assigned without uncertainty to one of K states. I will derive the likelihood of observing this family of sequences as well as the distributions over the transition matrix and initial state probabilities. To begin with, following standard references[76], a few definitions – since the sequence is perfectly observed, it may be summarized by the running counts of the state-to-state transitions $\{c_{ijn}, (i,j) \in 1 \dots K, n \in 1 \dots N\}$ and the running counts of the initial state occupancies $\{m_i, i \in 1 \dots K\}$. The likelihood function for one of the sequences is:

$$p(z_{T_n n} \dots z_{1n}) = p(z_{T_n n} | z_{T_n - 1, n} \dots z_{1n}) p(z_{T_n - 1, n} \dots z_{1n}) = p(z_{1n}) \prod_{t=2}^{T_n} p(z_{tn} | z_{t-1,n} \dots z_{1n})$$

$$= p(z_{1n}) \prod_{t=2}^{T_n} p(z_{tn} | z_{t-1,n}) = \pi_l \prod_{i=1}^{K} \prod_{j=1}^{K} A_{ij}^{c_{ijn}}$$

where the second-to-last line follows from the markov property, π_l is the initial state probability (supposing that the sequence starts in some state l), and the last line is simple accounting. Overall, since they are independent, the sequences have the likelihood, defining $\sum_{n=1}^{N} c_{ijn} = n_{ij}$:

$$L = p(\{z_{T_n n} \dots z_{1n}, t \in 1 \dots T_n, n \in 1 \dots N\}) = \prod_{l=1}^{K} \pi_l^{m_l} \prod_{i=1}^{K} \prod_{j=1}^{K} A_{ij}^{n_{ij}}$$

3-7

Applying a constrained optimization to the log-likelihood function, the following may be observed:

$$\lambda_i + \frac{\partial \log(L)}{\partial A_{ij}} = \lambda_i + \frac{n_{ij}}{A_{ij}} = 0 \rightarrow \lambda_i \sum_{j=1}^{K} A_{ij} = \lambda_i = -\sum_{j=1}^{K} n_{ij} \rightarrow A_{ij} = \frac{n_{ij}}{\sum_{j=1}^{K} n_{ij}}$$

Similarly, the most likely initial state probabilities can be given:

$$\pi_i = \frac{m_i}{\sum_{j=1}^{K} m_j}$$

However, this is not the most informative set of equations, because while the underlying premise of using the likelihood function is that some sequences are more likely than others, which will affect the estimate of the transition matrix, there is not yet any information on how the matrix elements and initial probabilities themselves are distributed. In other words, we have no measure as yet as to the precision to which the parameters are estimated. This can be rectified by noticing that the likelihood may be expressed as separable factors, q:

$$L = \left[\prod_{l=1}^{K} \pi_l^{m_l}\right] \prod_{i=1}^{K} \left[\prod_{j=1}^{K} A_{ij}^{n_{ij}}\right] = q(\{m_i\}|\pi) \prod_{j=1}^{K} q\left(\{n_{ij}\}\middle|A, i\right) \qquad \textbf{3-8}$$

Normalizing each of the factors gives $K + 1$ separate multinomial distributions:

$$p\left(\{n_{ij}\}|A, i\right) = \frac{\Gamma\left(\sum_j n_{ij} + 1\right)}{\prod_j \Gamma\left(n_{ij} + 1\right)} \prod_j A_{ij}^{n_{ij}}$$

$$p(\{m_i\}|\pi) = \frac{\Gamma\left(\sum_j m_j + 1\right)}{\prod_j \Gamma\left(m_j + 1\right)} \prod_j \pi_j^{m_j}$$

It is instructive to notice that these are the probability of drawing a set number of times from K categories with replacement. These probabilities supply $p\left(\{n_{ij}\}|A, i\right)$ and $p(\{m_i\}|\pi)$; we can now use Bayes' theorem to supply $p\left(A, i|\{n_{ij}\}\right)$ and $p(\pi|\{m_i\})$. The conjugate prior for the K distributions over the second index of A_{ij} as well as the distribution over the initial state probabilities π (and subsequently, their posterior predictive distributions), are each independent Dirichlet distributions[92]. With this parameterization, the average transition matrix elements and

initial state probabilities are now equal to an expectation over the posterior predictive distribution:

$$A_{ij} = E\left[Dir\left(A, i|n_{ij}, \alpha_{0,ij}^{A}\right)\right] = \frac{n_{ij} + \alpha_{0,ij}^{A}}{\sum_{j=1}^{K}\left(n_{ij} + \alpha_{0,ij}^{A}\right)}$$

$$\pi_i = E\left[Dir\left(\pi|m_i, \alpha_{0,i}^{\pi}\right)\right] = \frac{m_i + \alpha_{0,i}^{\pi}}{\sum_{j=1}^{K}\left(m_i + \alpha_{0,i}^{\pi}\right)}$$

where α_0 are the prior expectations for how many counts will be observed. In Appendix A, I will show the solution to the case where which state each sequence occupies in a given time is unknown, (i.e, a Hidden Markov Model.)

I next derive the likelihood function for the case where there are arbitrarily complex layers of conditions required for transitions between certain types of classes. First, I define a multilevel state variable $\{z_{nt}^{d}, n \in 1 \dots N, t \in 1 \dots T_n, d \in 1 \dots D\}$ for N independent sequences, each of which have a separate length T_n and each of which has at most D distinct conditions that are simultaneously met whenever the state is occupied. This state arrangement can be visualized as a tree of conditions, with stratified levels. Suppose that this family of independent sequences has been observed without uncertainty. At each time step, each multilevel state variable transitions entirely to the next one in a manner analogous to the markov chain above, which leads to the following likelihood function:

$$L = p\left(\left\{z_{T_n n}^{d}\right\} \dots \left\{z_{1n}^{d}\right\}\right)$$

$$= \prod_{n=1}^{N}\left[\prod_{d=1}^{D}\pi_{d,z_{n1}^{d}}\right]\left[\prod_{t=2}^{T_n}\prod_{d=1}^{D}A_{d,z_{nt}^{d},exit}^{\delta_{nt}^{d}}A_{d,z_{nt}^{d},z_{n,t-1}^{d}}^{\delta_{nt}^{d+1}(\delta_{nt}^{d}-1)}\pi_{d,z_{n,t+1}^{d}}\right]\left[\prod_{d=1}^{D}A_{d,z_{nT_n}^{d},exit}\right]$$

$$\delta_{nt}^{d} \equiv \left\{1 \; if \; z_{nt}^{d-1} = z_{n,t+1}^{d-1} \; or \; d > D, else \; 0\right\}$$

3-9

As for the simpler case above, suppose that, having gone through the sequence carefully, a count has been provided for each time a transition occurs. As above, there are many distinct types of counts: n_{dij} transitions between the subset of the Ω_d states at level d that they communicate with: that is, $i, j \in 1, \dots, \frac{\Omega_{d+1}}{\Omega_d} \equiv \widetilde{\Omega}_d$, e_{di} transitions from a state $i \in 1, \dots, \widetilde{\Omega}_d$ in the d level to its parent condition in $d-1$ level, and b_{di} transitions from a state at level d to one of the ones below it, in $i \in 1, \dots, \widetilde{\Omega}_{d+1}$ which also include the initial state counts. As before the likelihood can be rewritten in terms of these counts as follows:

$$L = \prod_{d=1}^{D} \prod_{i=1}^{\widetilde{\Omega}_d} \pi_{di}^{b_{di}} A_{d,i,exit}^{e_{di}} \prod_{j=1}^{\widetilde{\Omega}_d} A_{dij}^{n_{dij}}$$

3-10

Just as above, because the probability space at each of the levels is closed since conditional probabilities are closed within their restricted sample space, this more complex likelihood may be analogously normalized into a factors consisting of multinomials with conjugate Dirichlet priors which are then used, in conjunction with the counts, to calculate posterior distributions that describe the expected parameters as well as their density in probability space. In the case where the state designation of the sequence is uncertain, this likelihood is used to construct a Hierarchical Hidden Markov Model. While both of these models are approximations to the true physical process in continuous time, in fact, both can be "glued back together" using a more complex likelihood that uses pseudo-counts to re-estimate the continuous-time generator[93]. The central point here, however, is that for discrete-time likelihoods, the sufficient goal is to accurately calculate counts, and the implicit distributions take care of updates, credible intervals, *etc*, from there. This is a specific case of a more general theory discussed later.

3.1.1.3 Dynamics of finite state tridiagonal matrices

Tridiagonal rate matrices describe population evolution in sequential systems. In this thesis I will be concerned with these systems in terms of zipping – transition from a helical form to a dissociated form one monomer at a time – and unzipping – the reverse process. However, they appear in more quixotic ways in this thesis as well – for example, the space of transport modes in a carbon nanotube possesses a tridiagonal Hamiltonian in the tight-binding approximation. The model could, obviously, be extended to describe diffusion of any polymer through a nanoscale pore. I will consider two cases: first, the case where the monomers are indistinguishable and the chain is infinitely long; second, the case when the monomers are distinguishable and the chain is finite. In both cases, I will derive the exponential for a finite n-state tridiagonal rate matrix which, as described above, immediately gives the time dependence of all the states; I will also show the solution for the inverse of a finite n-state tridiagonal matrix.

The simplest possible model for the zipping or unzipping of a nucleic acid is a model wherein a stacking nucleus forms by pairing of two adjacent bases, and the rest of the bases assemble by pairing contiguously – and only contiguously – to the nucleus. In other words, the probability of a state with n bases paired, designated $\rho_n(t)$, has the rate matrix Q which is designated according to some assumptions of how the underlying physics are organized. I will first assume that there is a single pairing rate, λ, and a single unpairing rate, μ. It can be shown, using for example the method of generating functions, that this model is ill-posed in the infinite n limit, where $n$must be positive, because nothing prevents, in the long run, the occupancy of $\lim_{n\to\infty,t\to\infty} \rho_n(t) = 1$ subsequently depleting the occupancies of the rest; however, in the finite n limit, one may proceed by locating the eigenvectors $\{\Xi_i\}$ and eigenvalues $\{\xi_i\}$ and using them to calculate the matrix exponential of Q according to $e^{Qt} = \Xi D \Xi^{-1}, D = Diag\left(e^{\xi_1 t}, \dots, e^{\xi_n t}\right)$. I

will calculate $\{\Xi_i\}$ and $\{\xi_i\}$ by adapting the method of Yueh[94]. First noting the following definitions:

$$\cos\theta = \frac{\xi_i + \lambda + \mu}{2\sqrt{\lambda\mu}}$$

$$\nu = \sqrt{\frac{\lambda}{\mu}}$$

Adapting equation (5) in Yueh for n states, i.e., possible base pairs:

$$\sin\big((n+1)\theta\big) - \frac{\lambda+\mu}{\sqrt{\lambda\mu}}\sin(n\theta) + \sin\big((n-1)\theta\big) = 0$$

3-11

This immediately gives, by symmetry of the sine function and choice of what makes the middle term vanish:

$$\theta = \frac{\pm k\pi}{n}, k \in \{1, \dots, n-1\}$$

Which allows, using equation (7) in Yueh:

$$\xi_k = -(\lambda+\mu) + 2\sqrt{\lambda\mu}\cos\frac{k\pi}{n}, k \in \{1, \dots, n-1\}$$

3-12

Noting that the cosine is bounded above by 1, we use the following to show that all the eigenvalues are negative:

$$-(\lambda+\mu) + 2\sqrt{\lambda\mu} = -\big(\sqrt{\lambda} - \sqrt{\mu}\big)^2 < 0$$

These eigenvalues allow us to use equation (8) in Yueh to give the eigenvectors:

$$\Xi_j^{(k)} = \nu^{j-1}\sin\frac{jk\pi}{n} - \nu^j \sin\frac{(j-1)k\pi}{n}, k \in \{1, \dots, n-1\}, j \in \{1, \dots, n\}$$

3-13

The last eigenvector is found by noting that $\Xi_n = \mathbf{1}$ is an eigenvector with eigenvalue $\xi_n = 0$. Together these give the entire solution space for any sized nucleic acid. Combining this result with the expression the preceding section gives the time evolution of the uncertainty/entropy as the system evolves in time. More complex models for nucleic acid dynamics can be constructed by restricting the state space to stacked states and allowing for a vocabulary, i.e. A-T, A-U, G-C, *etc*, to describe possible interactions[95,96]. This is similar to models developed for helix-coil transitions in the study of secondary-structure formation of proteins[97]. In general, investigations of this type have helped to define a most likely base pairing event rate, on the order of nanoseconds, as well an explanation for non-Arrhenius behavior – i.e., when more than one eigenvalue of the rate matrix dominates the dynamics, there is a glass transition. The experiments described below go beyond the observations that could possibly be made in these models, as discussed in Chapter 4 and Chapter 5, because I investigate the effect of tertiary structure and loop interactions on the dynamics, and the master equation formalism has no spatial structure.

To optimize the estimates of a diffusing noise variable below, I will require the inverse of an arbitrary finite-state tridiagonal matrix. This is given in Usmani *et al*[98] to be, if

$$M_{ij} = 0 \;\; |i-j| > 1$$

$$M_{ii} = b_i, M_{i,i+1} = c_i, M_{i,i-1} = a_i, i \in \{1, \dots, n\}$$

then

$$M_{ij}^{-1} = \frac{(-1)^{i+j} \prod_i^{j-1} c_i \, \theta_{i-1} \phi_{j+1}}{\theta_n}, i > j$$

$$M_{ij}^{-1} = \frac{\theta_{i-1} \phi_{i+1}}{\theta_n}, i = j$$

$$M_{ij}^{-1} = \frac{(-1)^{i+j} \prod_{j+1}^{i} a_i \, \theta_{j-1} \phi_{i+1}}{\theta_n}, i > j$$

where

$$\theta_i = b_i\theta_{i-1} - a_i c_{i-1}\theta_{i-2},$$

$$\theta_{-1} = 0,$$

$$\theta_0 = 1$$

and

$$\phi_i = b_i\phi_{i+1} - c_i a_{i+1}\phi_{i+2}$$

$$\phi_{n+1} = 1$$

$$\phi_{n+2} = 0, i \in \{1, \dots, n\}$$

3-14

Collectively this result immediately gives rise to an efficient linear-time algorithm for solving any linear equation involving a tridiagonal coefficient matrix.

3.1.2 Variational Inference

Bayes' theorem plays a central role in the design and execution of probabilistic machine learning, to the point where that process is commonly called Bayesian inference. Bayesian inference is the process of combining observations with previous knowledge or assumptions, quantified as a prior probability measure, in order to increase knowledge, challenge the prior assumptions, and measure the uncertainty in the new probability measure, quantified as a posterior. Denoting the observations collectively by d and the parameters describing the probability measure by θ, Bayes' theorem can be restated:

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)} = \frac{p(d|\theta)p(\theta)}{\int p(d|\theta')p(\theta')d\theta'}$$

3-15

In many cases the posterior $p(\theta|d)$ contains the information of primary interest, describing how likely a set of parameters are given the present data. Deriving an exact posterior is often

computationally intractable even in simple cases. Approximations must be made. In this presentation, I will first derive the variational approximation in terms of local computations on a mean-field graph. This will form the computational toolbox used to solve all the models below.

The following develops precisely what variational inference in the form presented by Winn *et. al*[99] aims to achieve, and may be skipped by those most interested in Chapter 4 and 5. Inference on a mean-field graph using this form refers to a class of problem-solving using repetitive distribution-inference and information-passing equations collectively referred to, in analogy to belief propagation, as variational message-passing (occasionally called coordinate ascent). To show this explicitly, following Winn *et. al*[99], I begin by writing the definition of joint probability in terms of data D and hidden states S which are typically described by some parameter space denoted θ, a state space of all the possible hidden states Ω_{S}, a state space of all the possible observations Ω_{D}, and the entire state space denoted by the Cartesian product of those, namely $\Omega = \Omega_S \times \Omega_D$. Writing,

$$p(D \subset \Omega_D) = \frac{p(\{\omega = (\omega_S, \omega_D) \subseteq \Omega \; \forall \; S \subset \omega_S, D \subset \omega_D\})}{p(\{\omega = (\omega_S, D) \subseteq \Omega \; \forall \; S \subset \omega_S\})} \equiv p(D) = \frac{p(S \cap D)}{p(S|D)}$$

3-16

where the latter statement, while an abuse of notation, is much easier to read, and will therefore be used from now on. We typically consider the log probability:

$$\log\big(p(D)\big) = \log\big(p(S \cap D)\big) - \log(p(S|D))$$

For some probability measure $q(S)$, where all integrals are understood to be over the entirety of the sample space associated with the hidden states possessing some appropriate measure,

$$\log\big(p(D)\big) = \int_S q(s)\log\left(\frac{p(s\cap D)}{q(s)}\right)ds + \int_S q(s)\log\left(\frac{q(s)}{p(s|D)}\right)ds$$
$$\equiv \mathcal{L}(q) + D_{KL}(q(S)||p(S|D))$$

3-17

The quantity $\mathcal{L}(q)$ is a lower bound on $\log\big(p(D)\big)$ because $D_{KL}(q(S)||p(S|D))$, which is sometimes called the Kullback-Leibler divergence,

$$D_{KL}(q(A)||p(B) = \int q(A)\log\frac{q(A)}{p(B)}dA$$

is always non-negative (alternatively, we could note Jensen's inequality for some measurable function $q(S)$

$$\log\big(p(D)\big) = \log\int p(S\cap D)\frac{q(S)}{q(S)}dS \geq \int q(S)\log p(S\cap D)\,dS - \int q(S)\log q(S)\,dS$$

noting that the logarithm is concave, which holds for generalized functions[100].) The standard procedure is to find a probability measure $q(S)$ which maximizes $\mathcal{L}(q)$. This is sometimes referred to as maximizing the lower bound of the evidence, $\log\big(p(D)\big)$. Using A. 3 (see Appendix A), there must be some family of disjoint subsets $S_i \subseteq S, S_i \cap S_j = \emptyset, i \neq j$, so that we can define $q(S) \equiv \prod_i q_i(S_i)$ where each factor $q_i(S_i)$ is separately normalized and has no measure outside its defined support set S_i. Then,

$$\mathcal{L}(q) = \int_S \prod_i q_i(s_i) \log\left(\frac{p(s \cap D)}{\prod_j q_j(s_j)}\right) ds$$

$$\mathcal{L}(q) = \int_S \prod_i q_i(s_i) \log\big(p(s \cap D)\big)\, ds - \int_S \prod_i q_i(s_i) \sum_j \log\Big(q_j(s_j)\Big)\, ds$$

3-18

This equation simplifies because the log factors vanish outside their disjoint subsets of definition, so this lets us write:

$$\mathcal{L}(q) = \int_S \prod_i q_i(s_i) \log\big(p(s \cap D)\big)\, ds - \sum_j \int_{S_j} q_j(s_j) \log\Big(q_j(s_j)\Big)\, ds_j$$

This equation can be decomposed for each S_j:

$$\mathcal{L}(q) = \int_{S_j} q_j(s_j) \int_{S \setminus S_j} \prod_{i \neq j} q_i(s_i) \log\left(\frac{p(s' \cap D)}{q_j(s_j)}\right) ds_i\, ds_j - \sum_{i \neq j} \int_{S_i} q_i(s_i) \log\big(q_i(s_i)\big)\, ds_i$$

This may be made compact by recalling the form of the Kullback-Leibler divergence as well as the definition of information (negative entropy) as $H\big(q(x)\big) = \int q(x) \log\big(q(x)\big)\, dx$ over some probability measure $q(x)$. Defining

$$\log \varphi_j(s_j \cap D) = \int_{S \setminus S_j} \prod_{i \neq j} q_i(s_i) \log\big(p(s' \cap D)\big)\, ds_i - \log(Z_{\varphi_j})$$

where Z_{φ_j} normalizes φ_j gives:

$$\mathcal{L}(q) = -D_{KL}\big(q_j(s_j) || \varphi_j(s_j \cap D)\big) - \sum_{i \neq j} H\big(q_i(s_i)\big)$$

3-19

Which, noting the range of D_{KL} or, alternatively, taking the functional derivative, yields an update equation for maximizing $\mathcal{L}(q)$, that is, by setting $q_j\big(s_j \cap S_j\big) = \varphi(s_j \cap S)$ in turn, until

$\mathcal{L}(q)$ converges to a value. It is interesting at this point to note that the lower bound is simply a differential information term combined with a series of information terms, and that we are, indeed, maximizing entropy.

The only remaining task is to simplify calculation of $\varphi(s_j \cap S)$ in terms of local operations. First, I define what I mean by locality: I assume that there exists a graphical representation $G = \left(\{c_{ij}\}, \{\Omega_i\}, i, j \in \mathbb{N}\right)$ on $\Omega = \Omega_S \times \Omega_D$ where $\{c_{ij}\}$ is an antisymmetric matrix, known as a connectivity, adjacency, or Kirchhoff, matrix, which contains a value of 1 in the i,jth position if Ω_i points to Ω_j (a -1 if the other way around), and 0 if those two nodes are not directly connected. If $c_{ij} = 1$ then Ω_i is a parent to Ω_j (sometimes denoted $\Omega_i \subseteq pa(\Omega_j)$) and if $c_{ij} = -1$ then Ω_i is a child of Ω_j (sometimes denoted $\Omega_i \subseteq ch(\Omega_j)$). The connectivity matrix is assumed free of cycles – ie, no parent node is a child of a child of one of its children, *etc*. With that assumption, and the assumption that $\Omega_i \cap \Omega_j = \emptyset, i \neq j$ the probability measure can be decomposed:

$$\mathrm{p}(\Omega) = \prod_i p\left(\Omega_i | pa(\Omega_i)\right)$$

3-20

Colloquially G is referred to as the graph of conditional relationships, which is directed (because $\{c_{ij}\}$ is antisymmetric) and additionally, by assumption on $\{c_{ij}\}$, acyclic. Clearly not every probability measure may be decomposed like this; I will only treat with those that can, a set of probability measures commonly called Dynamic Bayesian Networks (DBNs[76].) A calculation on Ω_i will be called local if it only requires computations that make use of its parents, children, and its children's other parents (sometimes denoted , $cp(\Omega_j) \equiv pa\left(ch(\Omega_j)\right) \backslash \Omega_j$) a set collectively referred to as the Markov blanket. Typically in this case the S_i above are chosen to coincide with

the parts of Ω_i involving its hidden states. In the sense that I restrict inference to DBNs and variational distributions that coincide with G, I am concerned solely with a mean-field graph though, strictly, I will show that computations depend on the entire Markov blanket. With this restriction and change of notation (for example, $\Omega_{\mathrm{S,j}} \equiv \mathrm{S_j}$ and some disjoint family of subsets $\omega_j \subseteq \Omega_{\mathrm{S,j}}$) to reflect G, I continue my treatment:

$$\log \varphi_j(\omega_j \cap D) = \int_{\Omega_{\mathrm{S}} \setminus \Omega_{\mathrm{S,j}}} \prod_{i \neq j} q_i(\omega_i) \sum_{m \subseteq pa(\Omega_i)} \log\big(p(\Omega_i|m)\big)\, d\omega - const$$

Given an arbitrarily complex Markov blanket of $\Omega_{\mathrm{S,j}}$,

$$\begin{aligned} &\log \varphi_j(\omega_j \cap D) \\ &\quad = \int_{\Omega_{\mathrm{S}} \setminus \Omega_{\mathrm{S,j}}} \prod_{i \neq j} q_i(\omega_i) \Bigg[\sum_{m \subseteq pa(\Omega_j)} \log\Big(p\big(\omega_j \cap D|m\big)\Big) \\ &\quad + \sum_{m \subseteq pa\left(ch(\Omega_j)\right)} \sum_{n \subseteq ch(\Omega_j)} \log\big(p(n|m)\big) \Bigg] d\omega - \log(Z_{\varphi_j}) \end{aligned}$$

3-21

All the terms for the factored variational distribution $\varphi_j(\omega_j \cap D)$ except those arising from the Markov blanket of $\Omega_{\mathrm{S,j}}$ are absorbed directly into the normalizing function $\log(Z_{\varphi_j})$. Therefore calculation of the lower bound and updating the variational distribution requires only calculations within the Markov blanket and are thus local.

Next, I will derive optimal factorized distributions over the various types of observations that will be encountered, as well as their conjugate priors, which will, using Bayes' theorem, allow calculation of the posterior distribution necessary for computing the updates to $\varphi_j(\omega_j \cap D)$. First, I will give a general prescription for calculating optimal, in the sense that they minimize

the gained information from the data, priors given a distribution and show that in specific cases, these optimal priors are the conjugate priors: following the analysis of Gutierrez-Pena *et. al*[101], consider the distributions $p\big(\Omega_i|pa(\Omega_i)\big)$ to be constrained to the exponential family of distributions, that is, a distribution of the form

$$\log f_X(X|\theta) = \log\big(h(x)\big) + \eta(\theta)\cdot T(x) - A(\theta)$$

where, in this context, $h(x)$, $A(\theta)$, $\eta(\theta)$, and $T(x)$ are, respectively, called the base measure (scalar), partition function (scalar), natural parameter (vector), and sufficient statistic (vector). The principal result for this family, known as the Pitman-Koopman-Darmois theorem[102,103], is that inclusion in the exponential family of distributions is necessary for $T(x)$ to possess a set number of dimensions upon collection and incorporation of additional data[102,103]. Conjugacy has significant advantages for inference. Importantly, the form of the posterior distribution does not change with additional observations. In fact, it can be shown that with the choice of conjugacy, information gain upon observation of data is minimized[101]. When conjugate priors from the exponential family are used, it has been shown that the update scheme described above amounts to a message-passing algorithm known as variational message passing[99], and is equivalent to the variational approximation for Bayesian inference. This algorithm will be used extensively in the practical applications below.

3.2 Practical Applications

I will discuss the development of three practical applications of the variational approximation which are important for the analysis of single-molecule data. First, I will describe an algorithm that uses variational inference to learn the positions of isolated molecules in a movie, estimate the intensities of their light emission, and identify equivalent molecules detected

at disparate wavelengths. Second, I will describe an algorithm that allows the quantification of static and dynamic heterogeneity. Third, I will describe an adaptation of existing baseline-correction Gaussian mixture models to a Hidden Markov Model with baseline-correction, which allows analysis of trajectories with non-gaussian noise distributions.

3.2.1 Unified, Bayesian Inference-based Framework for Analyzing Single-molecule Fluorescence Microscopy Experiments[5]

Comprehensive quantification of the underlying biomolecular processes observed in single-molecule fluorescence microscopy experiments requires the implementation of multiple, complex methodologies in order to transform fluorescence intensity images into informative quantities such as rate constants and free energy landscapes. In part due to the computational and scientific complexity required to complete such transformations, no comprehensive standard to do this exists within the field. Thus the analysis of such experiments is often performed with methods that are disjointed, subjective, and even arbitrary. To address this shortcoming, we have developed a software package, which we call vbscope, that uses a Bayesian inference-based framework and modern machine learning algorithms in order to, in a statistically rigorous manner, unify the various tasks required to analyze an ensemble of single-molecules while remaining conscientious of the underlying physical processes involved in such experiments. As a result of the increased consideration of such processes, we show that the use of vbscope enables identification of nearly all the light-emitting chromophores in an image as well as an at least two-

[5] Co-written with Dr. Colin Kinz-Thompson, who additionally provided significant scientific insight as well as the labeled RF1 and performed the smFET experiments described in this work. This work is a manuscript in preparation as of 2017.

fold increase in the signal-to-noise ratio of individual single-molecule intensity *versus* time trajectories. Furthermore, the use of Bayesian inference allows separate experiments analyzed by vbscope to be compared on the same statistical footing thus enabling a novel hierarchical approach to the analysis of single-molecule fluorescence microscopy experiments, which allows quantitative comparison between disparate datasets.

3.2.1.1 Introduction

The analysis of single-molecule fluorescence microscopy experiments requires a demanding mixture of physicochemical knowledge and complex statistical methodology.[6] The process of extracting information from these experiments begins by accurately estimating the fluorescence intensity originating from diffraction-limited spots and ends by using this data to rigorously characterize the behavior of the ensemble of single-molecules. Because these steps are non-trivial, systematizing the procedure by which such single-molecule data is collected and analyzed is necessary in order to minimize bias in the quantification of these intricate datasets. Towards this end, we have developed a computational tool that is robust enough to deal with the diversity of experimental data, and yet is flexible enough to assist in testing the validity of hypotheses a researcher may make about a single-molecule fluorescence microscopy dataset. In this work, we describe these methods and the software package we have developed to facilitate their implementation, as well as the application of these methods to single-molecule fluorescence microscopy experiments.

Our principal approach is to leverage modern machine learning tools in order to enable computers to adaptively identify light-emitting chromophores, and then to use those identifications to rigorously analyze the experiment. In the physical sciences, such machine

learning approaches have found great success in fields such as astrophysics,[104] but, despite their potential, have not been widely adopted by the single-molecule fluorescence microscopy community.[81,105,106] We begin by defining what it means to probabilistically infer the presence of a light-emitting chromophore in a diffraction-limited, fluorescence microscopy image or time-ordered series of images (*i.e.*, a movie) collected using a wide-field microscope. Essentially, this is simply the process of locating a bright spot (*i.e.*, a local maximum) in an image. However, due to the many sources of noise inherent to single-molecule experiments, there are often many local maxima in an image, not all of which correspond to a chromophore. Therefore, in order to infer the presence of a light-emitting chromophore in such an image, the significance of the local maximum in question must be evaluated relative to the sources of noise present in the experiment.

To do this, we have adapted a Bayesian inference-based framework,[76,107] which 'learns' how to characterize intensity maxima of interest by identifying features in a movie and using those identifications to find additional features. Bayesian inference is a statistical method that mirrors the scientific process by allowing initial hypotheses (*i.e.*, prior probability distributions) to be mathematically updated in response to the acquisition of new data from an experiment. In the context of single-molecule fluorescence experiments, Bayesian inference provides an increasingly popular conceptual approach and computational toolkit, since use of Bayesian techniques naturally quantifies the experiment- and ensemble-derived uncertainty that a particular amount of observed data brings to any calculated parameters.[83–85,108–114] Additionally, Bayesian inference approaches are able to optimally select between different mechanistic models in a way that prevents over-interpretation and encourages parsimony.[76,108] Both of these aspects are very important for single-molecule fluorescence experiments, because the fluctuations in the

fluorophore intensity, which often serve as proxies for the dynamics of underlying biomolecular complexes, are complicated to interpret due to the photophysics of the chromophores.[115] Such complications can drastically limit the amount and the usefulness of the data collected in these experiments. In our implementation, which uses the variational Bayes method in the context of the variational message passing algorithm,[92,99] Bayesian inference functions in a manner that not only standardizes the analysis of single-molecule fluorescence microscopy movies, but that also enables a statistically rigorous comparison of different movies in an unbiased manner.

Following this, we discuss methods that automatically register all of the images in a movie into a common coordinate system, and thus create a universal map of chromophore locations in the single-molecule fluorescence movie. This is necessary in order to recognize distinct chromophores across the successive images in a movie, and also across the multiple color channels imaged in multi-wavelength microscopy experiments. For instance, in single-molecule fluorescence resonance energy transfer (smFRET) experiments,[6,17] after a common coordinate system is determined, a map of the chromophore locations in different color channels can be used to associate the fluorescence intensity emitted from the donor and acceptor fluorophores with one another in order to calculate the time-averaged efficiency of resonance energy transfer (E_{FRET}) in the acquired multi-wavelength image. In addition to their use in smFRET experiments, these maps of chromophore locations *versus* time form a foundational step for many analysis routines involving single molecules, such as single-particle tracking,[116] colocalization,[117] or multicolor super-resolution experiments.[118]

Next, we discuss an optimal approach for estimating the emission intensity of a chromophore given its location in an image by considering the utility of several different methods. Using an algorithm that optimally estimates the amplitude of any point-spread function

(PSF), we adaptively correct for background contamination and, both in theory and in practice, increase the signal-to-noise of chromophore intensity *versus* time trajectories by nearly a factor of two when compared to methods currently used in the field.[17] Finally, we connect these successive methods of analysis with a comprehensive set of statistics to describe the underlying processes occurring in the analyzed single-molecule fluorescence microscopy movie, as well as with a flexible tool for the visualization and analysis of intensity *versus* time trajectories, which interfaces with various probabilistic analysis models, such as various Bayesian hidden Markov models (HMMs)[83–85,108] in order to facilitate the diversity of tasks required for the analysis of data from single-molecule kinetics experiments.[15] In short, this collection of single-molecule fluorescence microscopy analysis methods, which we have assembled into an open-source and freely available software package called vbscope, enables a significant improvement in the accuracy and statistical rigor of the analysis of single-molecule fluorescence microscopy experiments with an approach that unifies the entire hierarchy of single-molecule data analysis from the level of processing raw fluorescence microscopy movies to the level of analyzing individual single-molecule trajectories.

3.2.1.2 Methods and Results

The computational task of analyzing multi-wavelength, single-molecule fluorescence microscopy images can be subdivided into three parts (Figure 3.1A). First, a molecule must be identified and its position in the image must be determined; second, the locations of the individual molecules, or the sets of associated molecules that appear in different wavelength images must be mapped into a common coordinate system, also known as image registration; and finally, the position and intensity *versus* time trajectories must be estimated from the raw

data using these registration maps. As such, we have divided our exposition into sections dealing individually with each step.

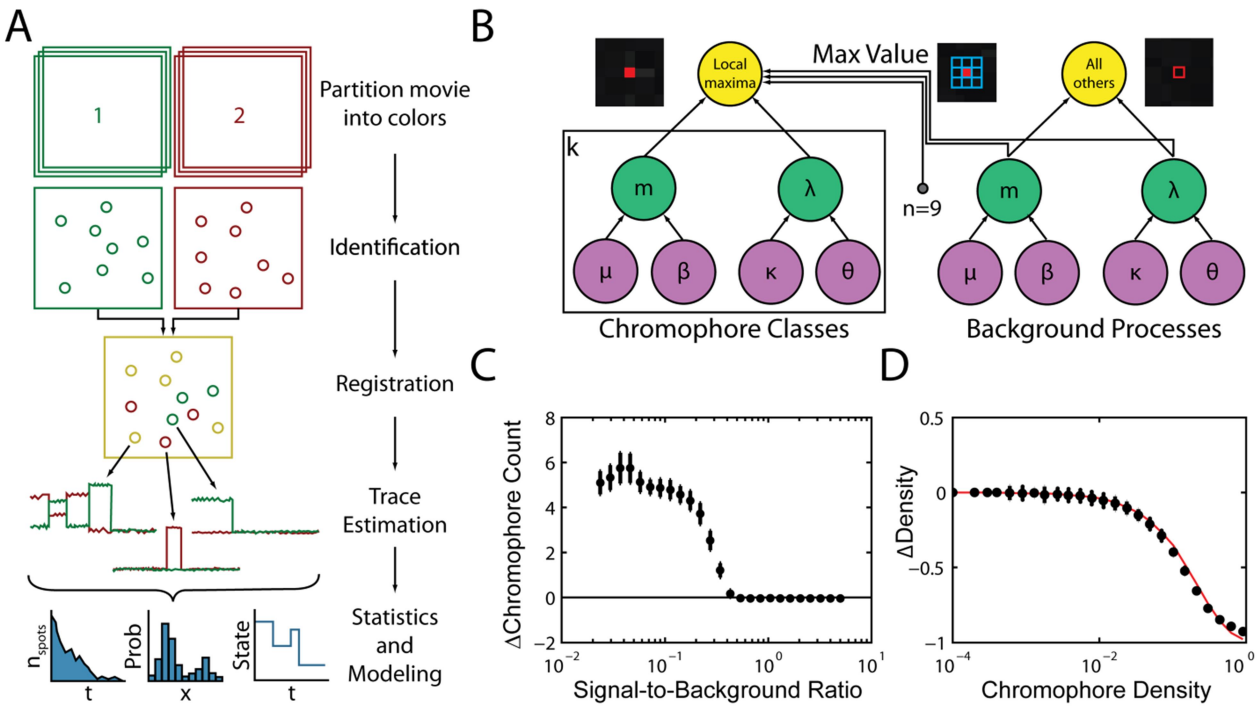


Figure 3.1 Identification of light-emitting chromophores.

(A) Schematic of analysis of single-molecule fluorescence microscopy movie. (B) Graphical model used to identify light-emitting chromophores. First, an image of a single-molecule fluorescence microscopy movie is split into two parts: pixels with intensities that are local maxima and those which are not. The pixels that are not local maxima are used to estimate the background distribution, as they do not contain light-emitting chromophores. Pixels that are local maxima are a mixture of noise-associated coincidences and light-emitting chromophores. The maximum value distribution associated with the background distribution describes the coincidences, and a mixture model is used to describe the chromophores. Data nodes are shown in yellow, parameter nodes are shown in green, and prior nodes are shown in violet. (C) Plot of chromophore identification error as a function of signal-to-background ratio (SBR). The absolute error between the number of simulated and the number of identified chromophores increases

when the SBR of is less than 0.5. (D) Plot of normalized chromophore identification error as a function of density of chromophores per pixel. The normalized error (density) between the number of simulated and the number of observed spots (black) follows the number of non-coincidentally co-localized, uniformly randomly-distributed chromophores (red).

3.2.1.2.1 Identification

The first, and sometimes principal, task of analyzing wide-field, single-molecule fluorescence microscopy images is to identify all of the light-emitting chromophores that were imaged. In such images, each pixel corresponds to a distinct region in the microscope's field of view, and this region can contain zero, one, or multiple light-emitting chromophores. However, in order to maintain single-molecule resolution, the density of light-emitting chromophores in the entire field of view must be low enough such that the light emitted from each chromophore is recorded in a distinct set of pixels with minimal contribution from neighboring chromophores. Under such conditions, the location of each light-emitting chromophore in a field of view can be mapped to a corresponding pixel, and this pixel will be a local intensity maximum in the image. Unfortunately, because background noise sources such as scattering and/or detection noise necessarily create local maxima in sets of pixels that correspond to regions lacking light-emitting chromophores, not all of the local maxima in an image will correspond to light emitted from a chromophore. As such, we propose to identify all of the individual light-emitting chromophores in such a single-molecule fluorescence microscopy image by determining whether each particular local maximum in the image resulted from the presence of a light-emitting chromophore or whether it corresponds to a local maximum generated by background noise.

With this distinction in mind, we have developed a probabilistic model to classify all of the local maxima located in such an image as either light-emitting chromophores or as coincidences associated with background noise (Figure 3.1B). We treat the probability of the intensity value of each local maximum pixel with a mixture model having a conditional probability of belonging to one of k different classes of intensity (*i.e.*, 'types' of chromophores), or being the local maximum of eight neighboring background intensities (*i.e.*, the search radius involves only nearest neighbors). By then weighing the likelihood of the local-maximum pixel belonging to any of the chromophore classes, in other words by marginalizing out the distinction between the different classes, the probability that the intensity value, I, of that pixel is best explained by any of the set of k chromophore classes, $\{1 \dots k\}$, is given by

$$p(pixel \in \{1 \dots k\}) = \frac{\sum_{i=1}^{k} p(I|\theta_i)}{\sum_{i=1}^{k} p(I|\theta_i) + p(I|max_n(\theta_b))}. \quad \textbf{3-22}$$

In this notation, θ_i denotes the distribution parameters associated with chromophore class i, and $max_n(\theta_b)$ denotes the maximum value distribution[119] associated with taking the local maximum of n pixels whose intensities are independent random variables identically distributed according to the background distribution governed by parameters θ_b. For a background that is distributed according to the normal distribution (*e.g.*, a background dominated by a combination of high background electron counts and instrumental noise), this maximum value distribution is given by

$$p(I|\mu,\lambda) = n\left(\frac{1+\mathrm{erf}\left(\frac{(I-\mu)\sqrt{\lambda}}{\sqrt{2}}\right)}{2}\right)^{n-1} \cdot \mathcal{N}(I|\mu,\lambda), \quad (2) \quad \textbf{3-23}$$

where $\mathcal{N}$ denotes a normal distribution, I is the intensity value of the local maximum, μ is the mean of the background distribution, λ is the precision (i.e., inverse of the variance) of the

background distribution, erf denotes the error function of the normal distribution, and n is nine – the local maximum and surrounding eight pixels. With the marginalized model given by Equation 3-22, users can set a probability cutoff (*e.g.*, greater than 0.5) that determines whether or not the local maximum pixel under consideration corresponds to a light-emitting chromophore or is best explained as a coincidence associated with the background noise distribution. This approach is conceptually similar to simultaneously evaluating the significance of a local maximum while concurrently defining the null hypothesis. In other words, a local maximum pixel with an intensity value that corresponds to the presence of a light-emitting chromophore is a pixel whose intensity value is inconsistent with the maximum value probability distribution associated with the random variable governing the background intensity values.

In our algorithm, parameter learning to determine the θ_i and θ_b takes place as part of a variational-Bayes expectation maximization (VBEM) routine[76] that is implemented using variational message passing.[92,99] However, in order to use the variational message passing algorithm to execute VBEM, we must deal with the fact that the maximum value distribution in Equation 3-22 associated with the background is not of the exponential family of probability distributions, and therefore the message-passing equations do not directly apply. To handle this complication, we note that the maximum value probability distribution is parametrized directly by the number of pixels over which the local maximum is determined, and by the background noise distribution, which also describes the pixels that do not contain light-emitting chromophores and are not local maxima. Therefore, we first use the pixels that are not local maxima to estimate the background distribution parameters, θ_b, using variational message passing, and then we use these parameters to directly compute the probability of the local maximum intensity value being a light-emitting chromophore (*c.f.*, Equation 3-22) by using the

analytical form of the maximum value distribution (*c.f.*, Equation 3-23). The VBEM algorithm alternates between two steps: (i) estimating the probability that each pixel belongs in either the background or light-emitting chromophore classes, and then (ii) using these occupation probabilities to re-estimate the values of the parameters associated with the chromophore classes, θ_i, and the background, θ_b. These steps are iterated until an estimate of the probability of observing the data (*i.e.*, the evidence lower-bound) converges,[76,99] and at this point the current values of θ_i and θ_b are taken as the inferred parameters.

With regard to the prior probability distributions used in this Bayesian inference procedure, we typically use prior probabilities that are estimated from a representative image of a control experiment, though we note that in absence of such a control, a representative image of the movie can be used instead. This is conceptually similar to an empirical Bayes (EB) approach,[76] though we do not use EB updates, because the time dependence of the molecular events occurring during the experiment (e.g., fluorophore photobleaching) yields time-dependent expectation values. Therefore, the successive images of a given movie will contribute time- and experiment-dependent contributions in the EB update procedure, which contributes to our opinion that each image in a movie is more readily modeled as a distinct experiment. Though, notably, modeling this time dependence is quite powerful.[111] Regardless, overall, this approach of using a control image to determine the prior probability distributions for the Bayesian inference procedure allows for rigorous comparison between different single-molecule fluorescence microscopy movies with a common statistical foundation (*i.e.*, with a common initial hypothesis).

In order to test the accuracy of this algorithm, we simulated images according to the generative model given in Figure 3.1B, and used the inference routine described above to

analyze these simulated images and therefore validate our implementation. First, we simulated a given number of light-emitting chromophores, with locations in image space drawn from a uniform distribution over the limits of the image, from the model in Fig. 1B, and counted the number of simulated chromophores identified *via* Equation 3-22. By varying the signal-to-background ratio (SBR) in this simulation, we identified the threshold at which the distinction between emitted light and background light breaks down (Figure 3.1C). This threshold is at an SBR of 0.5, roughly when the probability distributions for light emitted by a chromophore and that of background noise begin to significantly overlap. Furthermore, in order to evaluate whether our model can locate all of the light-emitting chromophores in an image, we performed another simulation in which we varied the density of light-emitting chromophores, with locations drawn as above, with a set SBR of 1 (Figure 3.1D). As shown, our model essentially finds all of the chromophores that do not, by coincidence, localize into the same pixel.

Finally, it is important to note that experimentally the illumination profiles in wide-field microscopy can be non-uniform, and that the model shown in Figure 3.1B does not capture this irregularity. Therefore, only when identifying chromophores and not when estimating chromophore intensities, we adaptively remove the local background inhomogeneity by subtracting the local minimum from the intensity of each pixel. Given a suitably low density of chromophores, these local minima will reasonably represent the variable amounts of scattering and background-fluorescence created by an inhomogeneous illumination profile. In our implementation, this calculation completes concurrently with the routine that locates the local maxima in the single-molecule fluorescence microscopy images in order to minimize computational cost. On a 3.6 GHz Intel Core i7 processor with four parallel threads, the entire chromophore identification process takes approximately 2 minutes of computer time on a

512x512 pixel movie containing 1200 frames; this means that, given a 10 Hz exposure time that is typical of many wide-field, single-molecule fluorescence microscopes, our chromophore identification algorithm can be performed in real-time.

3.2.1.2.2 Registration

In multi-wavelength microscopy images, the Cartesian coordinate systems describing the different color channels will not be the same due to imperfections in the alignments of different cameras or multi-wavelength imaging devices. Therefore, any maps describing the locations of light-emitting chromophores identified as described in Section 2.1 are only valid for the particular color channel used during identification. In order to reconcile these different color channel coordinate systems, we utilize two different methods to register the spatial differences in the alignments of different color channels. One method, which we call "deterministic," is guaranteed to produce a reliable registration of different color channels, but requires additional experimental information, while the second method, which we call "stochastic," does not require this additional effort, but instead relies on the accurate identification of some set of the same chromophores in all of the different color channels.

The deterministic registration procedure uses a pre-obtained image of fiducial markers, which must appear in all of the different color channels, as control points;[120,121] practically, we suggest using arrays of sub-diffraction limit, nanofabricated structures, such as zero-mode waveguides,[122] because such regularly repeating structures provide excellent coverage over an entire field of view. By fitting the locations of the fiducial markers to PSFs in order to determine their locations in each color channel,[123] we can then find an interpolating polynomial function that, for instance, transforms the apparent Cartesian coordinates x and y of color "1", x_1 and y_1, to those in color "2"

$$x_2 = \sum_i \sum_j w_{ij} \cdot x_1^i y_1^j,$$

3-24

where i and j are indices that run over the degree of the interpolating polynomial, and the weighting parameters, w_{ij}, can be estimated, for instance, using a non-linear least-squares fitting algorithm. An equivalent equation can be written, and determined for y_2 from these control points. The benefits of this approach are: (i) the particular interpolating functions can be used until the microscope alignment changes, which practically, we find, can be several months, (ii) once the coordinates of the fiducial markers have been obtained in each color channel's coordinate system, deterministic registration is effectively instantaneous, (iii) this method does not require chromophores to be identified in all of the different color channels, which might be important for experiments where one color channel is associated with a low-affinity interaction that is effectively too transient to localize, and (iv) in practice, we find that this method can accurately correct for arbitrarily complex optical distortions created by different optical components in the various optical paths for the different color images.

On the other hand, the stochastic registration procedure uses the locations of the chromophores in each color channel, which were learned with the identification algorithm described in Section 2.1. These locations are used to find registration maps between the different color channels by successive affine transformations of the identification map of one color channel to the identification map of another color channel until the affine transformation that maximizes the overlap between the chromophore locations is found. The shortcomings of this method are that several chromophores must be identified in all of the different color images in order to have a quantity to maximize, and also that it is impractical to accurately correct for

optical distortions with stochastically dispersed control points whose associations between the different color channels must be inferred.

Additionally, in general, the microscope field of view may undergo a global random walk *via,* for example, thermal fluctuations affecting the microscope stage. In this case, a drift correction may be necessary for optimal analysis. For this purpose, we have also provided a tool whereby each image of a particular color channel in a movie is globally registered to the first image in that color channel by registering the $i+1^{th}$ to the i^{th} image. This is done *via* successive affine transformations determined using the iterative closest point algorithm[124] on the map of the chromophore identifications discussed in Section 2.1. Such a correction can also be performed locally with single-particle tracking algorithms, of which variational Bayesian implementations have been recently developed.[110]

3.2.1.2.3 Absolute Registration

Importantly, we note that most multi-wavelength single-molecule localization microscopy experiments utilize registration functions (*e.g.*, Eqn. (3)) that transform the coordinate systems of the different color channels into the coordinate system of an arbitrarily-chosen reference color channel.[120,121] It is worth noting that the coordinate system of such a reference color channel is located in the image-plane, and is therefore distorted because of imperfections in the optical components of the imaging system. Thus, calculations of the relative distances between features in different color channels will not be accurate, since they are not of the distances in the object-plane (*i.e.*, real-space). In order to accurately quantify such relative distances, the coordinate systems of each color channel must be transformed into the rectilinear coordinate system of the object.

Experimentally, this absolute registration into the coordinate system of the object may be performed by finding a transformation between nanofabricated structures localized in a color channel of an optical microscopy image, and the exact same nanofabricated structures in the coordinate system defined experimentally, for instance, by an atomic-force microscopy (AFM) image. Practically, we find that, at minimum, fourth-order polynomials in Equation 3-24 are necessary to yield transformations that adequately remove optical distortion. Once such absolute registrations are obtained, one is typically interested in the magnitude of the distance vector, $|\boldsymbol{D}|$, between two chromophores in object-space. To obtain an expression for the probability distribution of $|\boldsymbol{D}|$ given chromophores localized in image-space, we first consider that generally, any transformation from a set of source coordinates, $\mathbf{x}_s = (x_s^1, \ldots, x_s^m)$, to the i[th] coordinate of set of target coordinates, $\mathbf{x}_t = (x_t^1, \ldots, x_t^i, \ldots, x_t^n)$, can be written as

$$x_t^i = \sum_{k=1}^{m} \frac{\partial x_s^i}{\partial x_t^k} x_s^k \equiv \sum_k B_k^i x_s^k,$$

3-25

where defining B_l^i allows the entire transformation to be succinctly written as $\mathbf{x_t} = \mathbf{B}\mathbf{x_s}$. Similarly, we can also transform the metric tensor associated with the source coordinates, $g_{kl,s}$, which *a priori* may not be known, to the target coordinates by writing

$$g_{ij,t} = \sum_{k=1}^{m} \sum_{l=1}^{m} \frac{\partial x_s^k}{\partial x_t^i} \frac{\partial x_s^l}{\partial x_t^j} g_{kl,s} \equiv \sum_k \sum_l A_{ij}^{kl} g_{kl,s},$$

3-26

where commonly the target coordinates are rectilinear coordinates in absolute space, and the metric is therefore the Kronecker delta δ_{ij}. If we assume that the undistorted chromophores have a symmetric Gaussian PSF form in object-space, then the distance vector in object-space between chromophores at positions $\mathbf{x}_1$ and $\mathbf{x}_2$ in the image-spaces of color channels 1 and 2,

respectively, can be considered a random variate from the multivariate normal distribution specified by

$$\mathcal{N}(\boldsymbol{D}|\mathbf{B}_1\boldsymbol{\mu}_1 - \mathbf{B}_2\boldsymbol{\mu}_2, \sum_k \sum_l A_{ij}^{kl}(\boldsymbol{\Sigma}_{kl,1} + \boldsymbol{\Sigma}_{kl,2})) \equiv \mathcal{N}(\boldsymbol{D}|\mathbf{B}\boldsymbol{\mu}_D, \boldsymbol{\Sigma}_D),$$

3-27

where $\boldsymbol{\mu}$ denotes the position of a chromophore in image-space, $\boldsymbol{\Sigma}_{kl}$ denotes the covariance of coordinates k and l in image-space, subscripts 1 and 2 denote the respective chromophore and color channel, and the subscript D denotes parameters related to the relative distance. Equation 3-27 follows from the convolution properties of normal distributions, and noting that the covariance transforms the same way as the metric tensor (Equation 3-26). Notably, if one is interested in the magnitude of the relative distance between these two points, the moment generating function for $|\boldsymbol{D}|$ can be written as

$$M_{|\boldsymbol{D}|}(t) = \frac{1}{\sqrt{\det(\boldsymbol{I}-2t\boldsymbol{\Sigma}_D)}} Exp(-\frac{1}{2}(\boldsymbol{B}\boldsymbol{\mu}_D)^T(\boldsymbol{I} - (\boldsymbol{I} - 2t\boldsymbol{\Sigma}_D)^{-1})(\boldsymbol{\Sigma}_D)^{-1}(\boldsymbol{B}\boldsymbol{\mu}_D)),$$

3-28

where $\boldsymbol{I}$ is the identity matrix. This formula follows from Theorem 3.2a.1 in Mathai *et al* after noting that the metric tensor in rectilinear coordinates (*i.e.*, δ_{ij}) is symmetric and positive semi-definite.[125] Typically, the matrix $\boldsymbol{\Sigma}^D$ will be diagonal because the variation in the position of each rectilinear coordinate is independent of the other coordinates. Finally, we note that while it is possible to write the moment generating function for $|\boldsymbol{D}|$ in an arbitrary coordinate system such as those of a reference color channel, the form given above with diagonal covariance matrices is exceedingly simple in comparison, and doing so would still require one to find the transformations into and out of the rectilinear object-space coordinates in order to compute the magnitude of the distance between the chromophores. To the best of our knowledge, this is the only demonstration of such an expression.

3.2.1.2.4 Intensity Estimation

After identifying the location of a light-emitting chromophore in all of the color channels in an image, the next task in our algorithm is to quantify the intensity of that chromophore in each color channel (Figure 3.1A). Generally, the intensity of light emitted by a chromophore is spatially distributed in an image according to some PSF, $\psi(\theta_{PSF}, \lambda)$, which is a function of parameters θ_{PSF} that describe the imaging system (*i.e.*, microscope) and is wavelength-dependent, λ. Assuming that the microscope is a linear imaging system, then each photon of a particular color that is emitted by a chromophore contributes additively to the image in a manner specified by the PSF. Thus, the monochromatic image, I, of a point source or other object with a linear spatial distribution of photon emission that is created by the emission of N_{λ_0} individual photons of wavelength λ_0 is a repeated convolution, which can be written as

$$I = \left(N_{\lambda_0} \cdot \rho\left(\theta_{Obj.}\right) \cdot \delta(\lambda = \lambda_0)\right) * \psi(\theta_{PSF}, \lambda) = N_{\lambda_0} \cdot \psi\left(\left\{\theta_{PSF}, \theta_{Obj.}\right\}, \lambda = \lambda_0\right), \quad \textbf{3-29}$$

where $\rho\left(\theta_{Obj.}\right)$ describes the spatial distribution of the object, δ is a delta-function to specify the wavelength of the photons, and $\psi\left(\left\{\theta_{PSF}, \theta_{Obj.}\right\}, \lambda = \lambda_0\right)$ denotes the PSF at wavelength λ_0 and accounting for the object described by $\rho\left(\theta_{Obj.}\right)$. Interestingly, Eqn. (8) amounts to the product of a prefactor described by the rate of photon emission from the chromophore, and a density function describing where those photons intersect the image plane; this is a general result for any linear imaging system, regardless of choice of PSF. Additionally, we note that, while chromophores emit multiple different wavelengths of light, given a sufficiently narrow band of wavelengths as defined by a band-pass filter, $\psi(\theta_{PSF}, \lambda)$ may be fairly independent of λ, and thus Equation 3-29 is also applicable to experimentally collected microscopy images. Finally, in

a wide-field microscopy experiment, we must also account for the discretization caused by recording the image on camera, as well as the presence of the n different chromophores in the field of view. As such, the expected intensity value, d_{xy}, of a particular pixel at position (x, y) with area c_{xy}, can be written as

$$d_{xy} = \sum_{i=1}^{n}\left(\int_{c_{xy}} I_i\right) + b_{xy} = \sum_{i=1}^{n}\left(N_{\mathrm{i}} \cdot \Psi_{i,xy}\right) + b_{xy}\ ,$$

3-30

where b_{xy} represents a convolution of the background photon counts, electron dark counts, and instrumental noise which contribute to the measured intensity for the pixel at position (x, y), I_i is the image of the i[th] chromophore given in Equation 3-29, N_{i} is an amplitude that corresponds to the intensity or the number of photons emitted by the i[th] chromophore, and $\Psi_{i,xy}$ denotes the discretized density function of the i[th] molecule that was integrated over the pixel at position (x, y). Considering that distinct equations like Equation 3-30 can be written for each of the pixels in an image, and that all of these equations are linear in all of the N_{i}, we can obtain an analytical formula for the estimate of each N_{i} by using the maximum-likelihood (ML) framework, after certain assumptions about the noise in these measurements are made (see Appendix A). This procedure yields

$$N_i = \frac{\sum_{x,y}\left(d_{xy} - b_i - \sum_{j\neq i} N_j \cdot \Psi_{j,xy}\right)\cdot \Psi_{i,xy}}{\sum_{x,y}\left(\Psi_{i,xy}^2\right)},$$

3-31

$$b_i = \frac{\sum_{xy}\left(d_{xy} - \sum_{j} N_j \cdot \Psi_{j,xy}\right)}{\sum_{xy} 1},$$

3-32

which applies regardless of the choice of PSF model (*e.g.*, Airy disk or Gaussian). A full discussion of the assumptions leading to these formulae and the initial conditions that we use in our EM implementation of these equations is given in Appendix A.

The advantage of this approach is that observed intensity values from multiple pixels can be used to estimate the N_i, and this added information increases the precision with which the values of the N_i can be inferred. As a result, this approach yields an almost two-fold improvement to the SNR of an estimated intensity versus time trajectory as compared to commonly used methods in the field (*i.e.*, summing neighboring pixels, or using just the central pixel), which exactly matches the theoretical improvement from including multiple measurements in the inference procedure (Figure 3.2A) (see Appendix A). Additionally, we investigated the effect that chromophore motion has on these calculations for a single-chromophore intensity *versus* time trajectory (Figure 3.2B). While directly estimating all of the PSF parameters provides an estimate of the amplitude that is invariant of chromophore motion, it is very computationally expensive, taking almost 400-fold longer to run than the EM implementation (Figure 3.2C). We find that our approach of estimating only the amplitude of the PSF is still robust against such motion relative to other methods, and therefore provides the best balance between effectiveness and computational expense. Finally, we note that because we simultaneously use a PSF for every chromophore in the image, our methods eliminates cross-contamination of intensity estimates from neighboring, identified molecules.

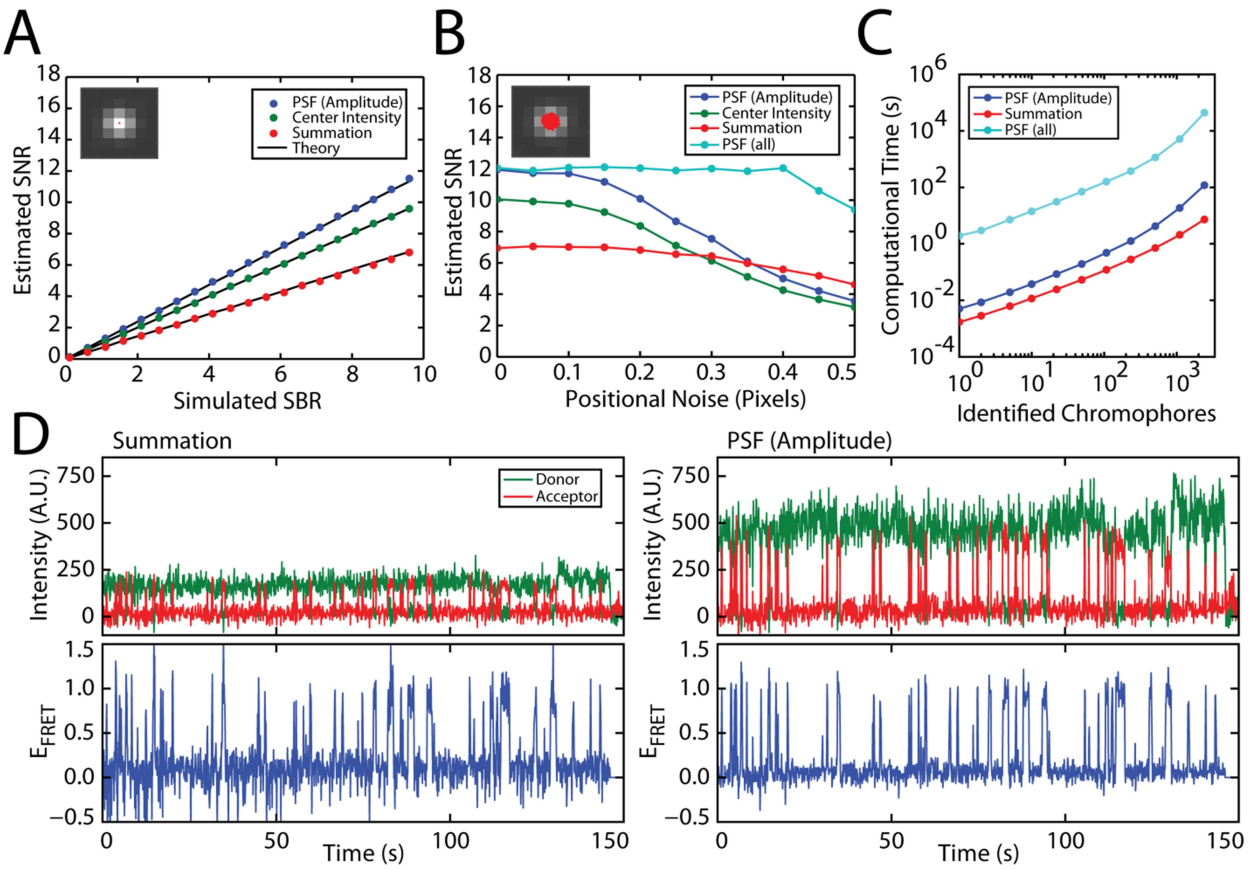


Figure 3.2 Chromophore intensity *versus* time trajectory estimation.
(**A**) Plot of estimated SNR versus simulated signal-to-background ratio (SBR) using the maximum-likelihood formula for the amplitude of an arbitrary PSF given by Eqn. 3 (blue), the local maximum pixel (green), and the sum of the local-maximum and nine-neighboring pixels (red). The PSF method is best, as it uses many pixels to obtain a better estimate of the amplitude and background. Black lines denote theoretical curves for each method (see Appendix A). (**B**) Plot of estimate SNR versus noise in emitter position. As the noise in the position of the molecule increases, the summation of neighboring pixels method remains more robust relative to the maximum-likelihood amplitude formula or central pixel methods, which maintain singular chromophore locations; however, the SNR estimated by fitting all the parameters of a PSF does

not degrade with chromophore motion. (**C**) Plot of the computational cost of the various methods of intensity estimation. We simulated ten, 100 x 100 pixel movies with 1000 frames each and benchmarked the amount of time required to calculate the intensity amplitudes in each case, as a function of the number of chromophores. Estimating the intensity using Eqns. (10) and (11) is roughly as fast as directly summing the neighboring pixel intensities, both of which perform approximately 10^6 estimates per second, whereas also fitting the PSF to determine the location of each chromophore results in a roughly 400-fold increase in computational time. (**D**) Comparison of summation (left) and PSF amplitude estimation (right) methods on experimental smFRET data of a translation factor binding to and dissociating from a ribosome. Here, using the PSF estimation method results in a two-fold increase in SNR, in agreement with theory (panel A).

With this optimal method of intensity estimation, we can take full advantage of the optimally identified spots found using the methods in Sec. 2.1. Importantly, because Equations 3-31 and 3-32 are analytic expressions, calculation using this direct PSF method takes approximately the same amount of time as directly summing the neighboring pixels in the region of interest. Thus, by careful consideration of the image, one obtains a two-fold increase to the SNR at no expense. We have demonstrated this on experimental data from a smFRET experiment of an acceptor-fluorophore labeled prokaryotic release factor 1 (RF1) variant binding to and dissociating from the aminoacyl-tRNA binding site of a surface-tethered prokaryotic ribosome containing a donor-fluorophore labeled tRNA in the peptidyl-tRNA binding site (Figure 3.2D) (see Appendix A for experimental details). In this case, not only do the individual donor and acceptor fluorophore intensity *versus* time trajectories originating from the same ribosome exhibit this SNR improvement, but the E_{FRET} *versus* time trajectory also is better resolved. This improvement means that, practically, higher concentrations of fluorophore-labeled

biomolecules (e.g., the acceptor-fluorophore labeled RF1) can be present in solution when imaging molecules of interest (*e.g.*, the donor-fluorophore labeled, surface- tethered ribosomes) before the ability to maintain single-molecule resolution is compromised. Effectively, our approach is a computational method to push past this 'concentration barrier', which limits the ability to observe transient and/or rare molecular inter-molecular interactions using single-molecule fluorescence approaches.[14,126] Furthermore, combined with the chromophore identification algorithm presented in Sec. 2.1, we find that, in practice, vbscope alleviates the concentration barrier by at least one order of magnitude, and in a complementary manner, significantly improves the acquisition time-resolution for smFRET experiments.

3.2.1.3 Analysis

3.2.1.3.1 Movie Statistics

Throughout the process of analyzing single-molecule fluorescence microscopy movies with vbscope in order to identify chromophores and calculate intensity *versus* time trajectories, a large number of informative statistics are simultaneously calculated. These statistics effectively describe the movie and its molecular underpinnings, and as such, can be utilized in the 'big-picture' evaluation of experimental datasets composed of multiple movies. Thus, vbscope can facilitate the task of optimizing experimental parameters, such as the chromophore loading density and optimal length of recording, in addition to the analysis of experiments that require less intricate approaches, such as calculation of co-localization probability distributions. To this end, vbscope provides the following statistics: the number of light-emitting chromophores *versus* time, the average SBR *versus* time, the total autocorrelation function of the chromophore intensity *versus* time trajectories, the cross-correlation functions between distinct color channels of the chromophore intensity *versus* time trajectories, the posterior probability density of the co-

localization probability for each color channel, the probability of co-localization by coincidence, the illumination profile of the microscope, and the multi-wavelength registration profile. Technical descriptions, as well as an example plot for each of these statistics from the analysis of an experimental obtained smFRET movie are provided in Appendix A.

3.2.1.3.2 Trajectory Analysis

While vbscope is designed to identify chromophores and estimate intensity *versus* time trajectories, we have also developed and included a tool to facilitate the downstream analysis of chromophore intensity *versus* time trajectories. This tool enables the classification of chromophore intensity *versus* time trajectories, creation of traditional smFRET-related plots, and further analysis *via* hidden Markov modeling using, for example, the Bayesian-inference based HMMs deployed in the vbFRET[83,108] and ebFRET[84,85] HMM packages. With this tool, for example, the analyst can sort multi-wavelength intensity *versus* time trajectories by amount of anticorrelation, in order to expedite cursory analysis of experiments. Additionally, this tool automatically classifies photobleaching events with a Bayesian-inference based, variational, Gaussian mixture model, which learns the instantaneous changes in intensity *versus* time trajectories; these points can also be manually corrected for each trajectory. The algorithms used for anticorrelation sorting and photobleach detection are described in the Appendix A. Finally, this tool connects the analyst to additional tools for creating transition density plots and 1- and 2-dimensional histograms, as well as to analysis suites such as vbFRET[83,108] and ebFRET[84,85] for rigorous kinetic analysis.

3.2.1.4 Conclusion

The vbscope approach presented here provides a unified, Bayesian-inference based framework that takes an analyst with little-to-no programming experience from data collection

all the way to the end stages of data analysis. By employing novel machine-learning algorithms, vbscope finds essentially all identifiable chromophores in a movie, a task crucial to production of unbiased data. Furthermore, vbscope uses analytical solutions in conjunction with any PSF to extract intensities in a manner that significantly increases the SNR of intensity *versus* time trajectories when compared to traditional methods, while striking an optimal computational cost balance. The post-analysis tools provided and interfaced with comprise a comprehensive and powerful analysis suite for single-molecule fluorescence experiments that minimizes the inherent bias present in automated and non-Bayesian approaches. Finally, as an open source and freely distributed software package, we hope that its adoption will increase the quality and efficiency of data collection and analysis of wide-field, single-molecule fluorescence experiments.

3.2.2 A Bayesian Approach to Hierarchical Hidden Markov Modeling Allows Direct Measurement of Conditional Kinetic Rates[6]

Time-resolved, single-molecule biophysical experiments and analyses can allow direct determination of the minimal number of states needed to describe a biological process of interest as well as direct quantification of the kinetics of the mechanistic steps connecting those states, thus characterizing the minimal kinetic scheme describing the biological process. In most cases, the signal *versus* time trajectories that are recorded in time-resolved, single-molecule biophysical experiments directly report on only a single coordinate (*e.g.*, a single distance change associated with a particular conformational change), while information from other coordinates is only indirectly reflected in the direct coordinate (*e.g.*, additional conformational changes that modulate the single distance change that is recorded). It is often of mechanistic interest to

[6] I would like to thank Dr. Kelvin Caban for his help writing this section, especially the discussion of translation.

characterize whether and how transitions along indirect coordinates influence the frequency of transitions along the direct coordinate – such an approach is necessary, for example, in order to quantitatively characterize the allosteric influence of ligand-binding to a biomolecular receptor on conformational dynamics of a structural element on the receptor distal to the ligand binding site. Here, we report the development of a computational algorithm capable of inferring the number of states and the rates of transitions between those states along both direct and indirect coordinates from data recorded as single-molecule signal *versus* time trajectories.

3.2.2.1 Introduction

Single-molecule biophysical experiments aim to facilitate definition and detailed quantification of the most parsimonious kinetic model describing a biological process of interest[6,15]. Kinetic models describe transitions between local configurational minima upon a multi-coordinate free energy surface[87]. While time-resolved experiments are undertaken to describe the relative depths of these free energy minima *via* the rates at which the biological process spontaneously fluctuates between them, single-molecule biophysical experimental designs typically define only a small handful of coordinates to observe directly, leaving variation along other coordinates indirectly observed: for example, in many cases, the presence of indirect coordinates may be inferred by sudden alteration in the rates at which the amplitudes associated with direct coordinates of free energy minima change – termed dynamic heterogeneity, schematized in Figure 3.3 – or by comparing separate molecules which apparently possess distinct kinetic models – termed static heterogeneity. As a result of the statistical inference involved, quantifying the number of, complexity of, and kinetic rates between indirect coordinates remains a significant computational challenge.

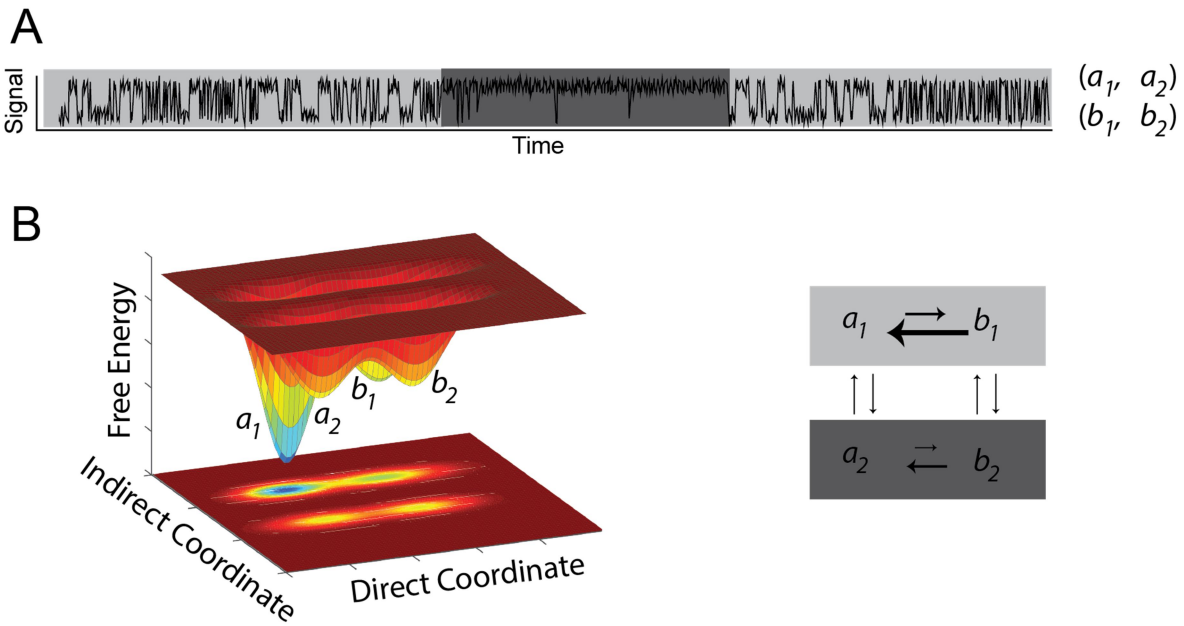


Figure 3.3 Hierarchically organized dynamic heterogeneity. (A) A simulated signal versus time trajectory composed of contiguous periods exhibiting distinct kinetic properties, indicated by the variable grayscale backgrounds, between two observable signal classes, denoted as (a_1 and a_2) and (b_1 and b_2.) (B) A sample energy landscape (left-hand side) corresponding to a generative kinetic scheme (right-hand side) for the trajectory shown in (A). A molecule on this landscape can undergo transitions both along a directly observed coordinate, or direct coordinate, describing $a_1 \overset{\rightarrow}{\leftarrow} b_1$ and $a_2 \overset{\rightarrow}{\leftarrow} b_2$ transitions, as well as along an indirectly observed coordinate, or indirect coordinate, describing $a_1 \overset{\rightarrow}{\leftarrow} a_2$ and $b_1 \overset{\rightarrow}{\leftarrow} b_2$ transitions. Information in the trajectory describing the indirect coordinate is obtained by inferring the frequency of transition between the two observable signal classes.

The information contained in time-resolved single-molecule trajectories abstractly consists of a continuous time Markov process discretely sampled by an apparatus into an apparent discrete-time Markov chain[127]. Because the experimental machinery record an imperfect reconstruction of the configuration of the molecule, either as a result of measurement noise or information loss during discrete sampling of a quickly varying continuous process[128,13], noisy reconstructions of single-molecule trajectories are commonly analyzed using Hidden Markov Models (HMMs)[129,77,80,83–85]: these probability distributions combine three elements – a kinetic model for transitions between free energy minima along direct coordinates; an observational, or emission, model to numerically define distributions of amplitudes associated with the direct coordinates of free energy minima; and a computational algorithm to quantitatively define both from a dataset[76]. HMMs have been fruitfully applied to single-molecule data when the probability of transition between free energy minima with distinct direct coordinates is slow compared to the experimental time resolution yet the speed at which a transition occurs is rapid in the same comparison. Such examples abound in, for instance, single-molecule recordings of ion channel currents[130], fluorescence resonance energy transfer (smFRET) recordings of RNA and molecular machines[6,7,17], force spectroscopy recordings[6,131], and field effect transistors[19,20,48,132]. Many of these recordings yield evidence of dynamic heterogeneity[19,133–135,45,136–139]: indirectly observed coordinates remodel the free energy landscape along which direct coordinate transitions occur in a manner identifiable by a burst, or a change in the rate of fluctuations between discrete observation classes of the direct coordinate amplitudes (Figure 3.3). In many cases this indirect coordinate corresponds to states of the complex with an external factor present or absent, the bound state biasing dynamics along the direct coordinate into a preferred configuration[136]. Such allosteric behavior is anticipated and commonly observed

in the single-molecule recordings of highly orchestrated and rectified motions of biomolecular machines[140,75]. However, the kinetic model of HMMs is highly uncorrelated – the molecular configuration at each time point is presumed to depend only on a single coordinate[76]. This assumption presents significant limitations to quantifying allosteric dynamic networks within biomolecular machines. While many methods have been developed to quantify such transitions[85,141–144], including methods capable of quantifying the limiting case of one indirect and one direct coordinate[145], general solutions remain elusive.

To rigorously address this problem, we have here adapted a class of inference tools based on a Markov chain subclass known as a hierarchical Markov chain, whose corresponding probability distribution is known as a Hierarchical Hidden Markov Model (HHMM)[146–148]. Description of the experimentally recorded trajectory with a hierarchical Markov chain allows each free energy minimum to be expressed by a full set of coordinates, both direct and indirect, and therefore allows definition of a hierarchical kinetic model describing transitions between free energy minima in a fully specified coordinate system. Using the variational approximation to Bayesian inference[76,99,100], fruitfully utilized for HMMs[83–85], we demonstrate a method to fully quantify a parsimonious hierarchical kinetic model for a population of single-molecule trajectories, and, as an example, apply these methods to experimental smFRET data demonstrating dynamic heterogeneity. These methods allow the experimentalist to rigorously measure the amplitudes of free energy minima along direct coordinates while simultaneously quantifying cooperative conformational changes within intricate biological complexes.

3.2.2.2 Theory

We begin by describing a hierarchical Markov chain – as discussed above, hierarchical Markov chains are subsets of Markov chains whose states are parameterized in terms of multiple

coordinates as opposed to a single coordinate[146]. The hierarchical Markov chain with obeys a Kolmogorov-Chapman equation[62] propagating one full set of D coordinates $\{z_{nt}^{d}\}$ for the nth trajectory at time t into the next in the subsequent time point $\{z_{n,t+1}^{d}\}$, giving rise to the following likelihood function, L, for a given population of N mutually independent trajectories each of length T_n:

$$L = \prod_{n=1}^{N} p\left(\{z_{nT_n}^{d}\} \ldots \{z_{n1}^{d}\}\right) = \prod_{n=1}^{N} p\left(\{z_{n1}^{d}\}\right) \prod_{t=2}^{T_n} p\left(\{z_{nt}^{d}\} | \{z_{n,t-1}^{d}\}\right)$$

3-33

We separate these coordinates into those which specify the emission distribution – direct coordinates denoted z_{nt}^{1}, setting up the production level of the state space, indirect coordinates z_{nt}^{2} that specify differences in interconversion rates between production states, and arbitrarily higher order indirect coordinates z_{nt}^{d} which specify differences in interconversion rates between the indirect coordinates at the level below z_{nt}^{d-1}. These coordinates are given natural number values that abstractly distinguish coordinates of free-energy minima. The nested, conditional dependencies of this state-space coordinate system may be visualized as a tree[146–148], which may be thought of as specifying the order in which the coordinates of free energy minima are always found on a chart. Using the conditional dependencies of this state-space coordinate system, the likelihood of the hierarchical Markov chain L can be decomposed[147] beginning with the direct coordinates at the production level with the direct coordinates and iteratively specify indirect coordinates until none remain:

$$L = \prod_{n=1}^{N} \left[\prod_{d=1}^{D} \pi_{d,z_{n1}^{d}} \right] \left[\prod_{t=2}^{T_n - 1} \prod_{d=1}^{D-1} A_{d,z_{nt}^{d},exit}^{\delta_{z_{nt}^{d+1},z_{n,t+1}^{d+1}}} A_{d,z_{nt}^{d},z_{n,t-1}^{d}}^{\delta_{z_{nt}^{d},z_{n,t+1}^{d}} \left(1 - \delta_{z_{nt}^{d+1},z_{n,t+1}^{d+1}}\right)} \pi_{d,z_{n,t+1}^{d}} \right] \left[\prod_{d=1}^{D} A_{d,z_{nT_n}^{d},exit} \right]$$

3-34

Where we have introduced the standard notation:

$$\delta_{ij} = \begin{Bmatrix} 1, \text{if } i = j \\ 0, \text{if } i \neq j \end{Bmatrix}$$

$$p\left(z_{n1}^{d} = i\right) = \pi_{di}$$

$$p\left(z_{nt}^{d} = i | z_{n,t+1}^{d} = j\right) = A_{dij}$$

$$p\left(z_{nt}^{d+1} \neq z_{n,t+1}^{d+1}, z_{nt}^{d} = i\right) = A_{di,exit}$$

with the final statement indicating the probability that indirect coordinates at level d specifying interconversion rates at level $d-1$ have changed. We note that a kinetic model for static heterogeneity may be derived from Equation 3-34 by simply requiring a single indirect coordinate that can take specify an arbitrary number of minima between which transitions are forbidden.

Equation 3-34 specifies the kinetic model. We use the variational approximation to Bayesian inference to specify both the emission distributions and the machine learning algorithm for the HHMM, which we briefly sketch here leaving the details of update equations to the *SI*. We seek to maximize the lower bound of the log probability, denoted the "evidence," or F, of a parameter distribution, denoted θ, and a set of observations, denoted $\{x_{nt}\}$, given prior information, denoted ψ_0:

$$F = \ln p\left(\{x_{nt}\}, \{z_{nt}^{d}\}, \theta | \psi_0\right) \geq \int d\theta \sum_{n} \sum_{z_{nt}^{d}} p\left(\{z_{nt}^{d}\}, \theta \middle| \{x_{nt}\}, \psi_0\right) \ln \frac{p(\{x_{nt}\}, \{z_{nt}^{d}\}, \theta | \psi_0)}{p\left(\{z_{nt}^{d}\}, \theta \middle| \{x_{nt}\}, \psi_0\right)}$$

3-35

The variational approximation assumes that there is no dependence between the coordinates and parameter distributions so that the joint probability may be written

$$p\left(\{z_{nt}^{d}\}, \theta \middle| \{x_{nt}\}, \psi_0\right) = q\left(\{z_{nt}^{d}\} | \psi_0\right) q(\theta | \psi_0)$$

Though we will here assume that the emission distributions are normal distributions with distinct parameters for each production state, this assumption can be generalized as necessary. Inference of the parameters of an HHMM proceeds by iteratively locating parameters that optimize a lower bound for the evidence. Iterations proceed by optimizing $q(\{z_{nt}^d\}|\psi_0)$ then by optimizing $q(\theta|\psi_0)$ and finally calculating the evidence lower bound; convergence is achieved with the evidence lower bound remains virtually unchanged between iterations.

By factorizing the joint probability in the variational approximation, we may also decompose the distribution of the kinetic model as follows. First, we simplify the hierarchical Markov chain likelihood in terms of its transition counts, that is, to:

$$L = \prod_{d=1}^{D} \prod_{i=1}^{\Omega_d} \pi_{di}^{b_{di}} A_{di,exit}^{e_{di}} \prod_{j=1}^{\Omega_d} A_{dij}^{n_{dij}}$$

where b_{di} denote the number of trajectories that begin with coordinates $z_{nt}^d = i$ together with the number of times $z_{nt}^d = i$ following a transition at the level above, e_{di} denotes the number of transitions from $z_{nt}^d = i$ and $z_{nt}^{d+1} \neq z_{n,t+1}^{d+1}$, and n_{dij} denotes the number of times that $z_{nt}^d = i$ given that $z_{n,t+1}^d = j$, and Ω_d denotes the number of distinct values of indirect coordinates at level d. It is trivial to observe that normalizing L implies that the kinetic factors decompose into multinomial distributions:

$$q(\{b_{di}\}, \{e_{di}\}, \{n_{dij}\}|\{\pi\}, \{A\}, \psi_0)$$

$$= \prod_{d=1}^{D} Mult(\{b_{di}\}|\pi_d, d, \psi_0) \prod_{i=1}^{\Omega_d} Mult(\{e_{di}\}, \{n_{dij}\}|A_{dij}, i, \psi_0)$$

$$= \prod_{d=1}^{D} q(\{b_{di}\}|\pi_d, d, \psi_0) \prod_{i=1}^{\Omega_d} q(\{e_{di}\}, \{n_{dij}\}|A_{dij}, i, \psi_0)$$

Therefore, considering the state space as a tree of connected points, each point can be considered as an independently operating Markov chain, and to infer the parameters and parameter distributions of the hierarchical kinetic model it is sufficient to calculate the transition counts specified above.

3.2.2.3 Results

Analysis of simulated data

We first sought to calibrate our model by investigating whether the algorithm presented could be used to accurately select among kinetic models. To do this, we simulated data from a known kinetic model, inferred optimal parameters for models of increasing complexity, calculated the lower bound of the evidence (see *Theory*), comparing the latter as a means for model selection. This task is important in the analysis of experimental data with indirect coordinates as in this case, the statistical techniques the only means for counting and distinguishing amongst alternative kinetic models.

We simulated data from a model with two indirect coordinates that specify eight distinct free energy minima (Figure 3.4, right; for simulation parameters see *Methods*). We optimized parameters using the HHMM with the state space on the left and on the right of Figure 3.4 – a simple model with two values of one indirect coordinate, and the correct model with two values for each of the two indirect coordinates respectively. Both models possess the correct number (two) of direct coordinates. We used the methods in the SI to calculate a lower bound for the evidence in both cases and find, as expected, a significantly higher value for the correct model than the simplified model demonstrating that the evidence lower bound may be used to specify the most parsimonious kinetic model.

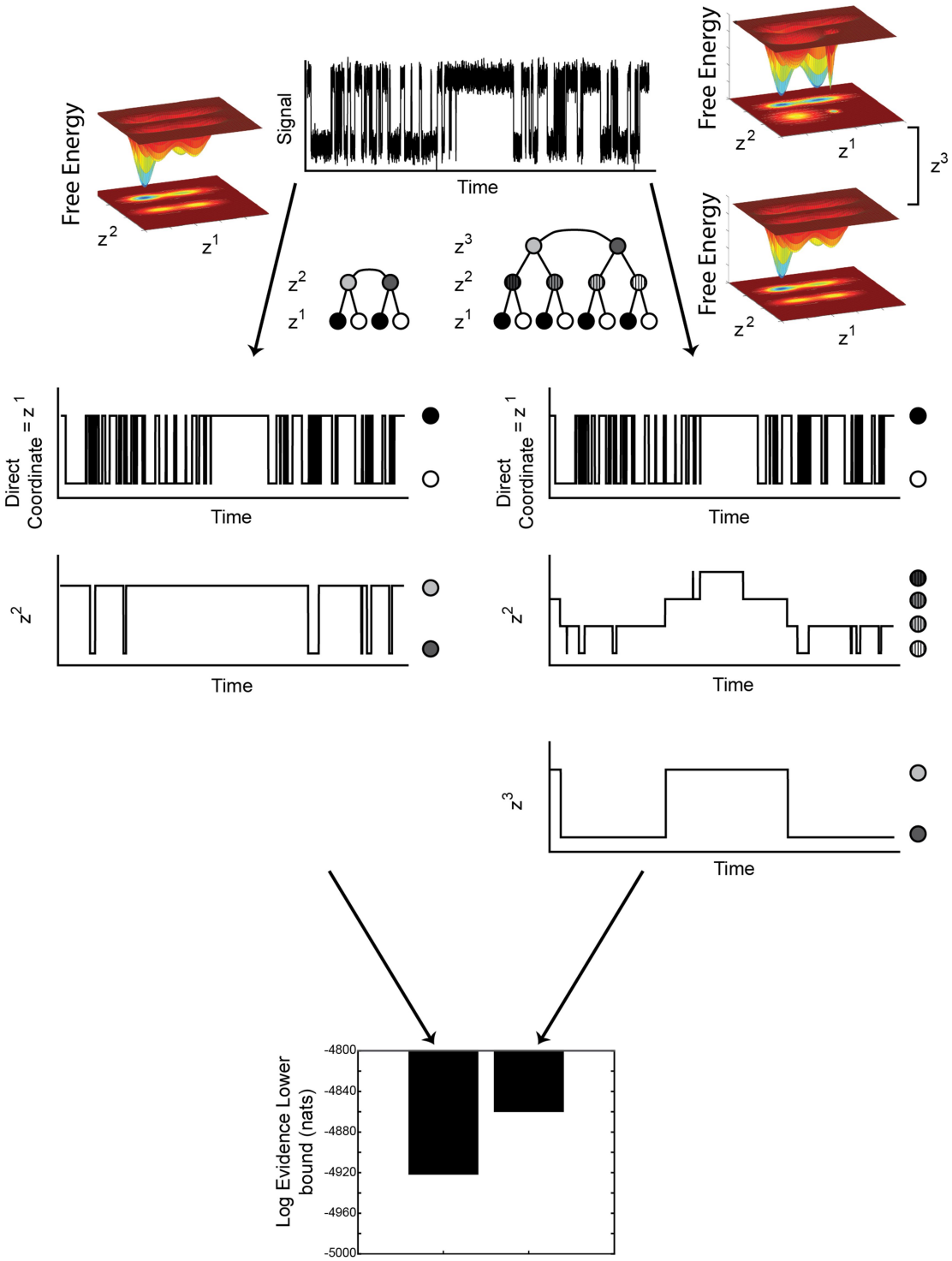

Free Energy
z2
z1
Signal
Time
Free Energy
z2
z1
z3
Free Energy
z2
z1
z2
z1
z3
z2
z1
Direct Coordinate = z1
Time
Direct Coordinate = z1
Time
z2
Time
z2
Time
z3
Time
Log Evidence Lower bound (nats)
-4800
-4840
-4880
-4920
-4960
-5000

Figure 3.4 Selection between distinct hierarchical models using variational inference. The performance of two distinct hierarchical models on the same simulated trajectory, which was simulated using one direct coordinate and two indirect coordinates and part of which is shown at the center of the top of the figure, were compared. The first model, describing the free energy surface on the left-hand side, is comprised of a direct coordinate, z1, and one indirect coordinate, z2. The second model, describing the free energy surface on the right-hand side, is comprised of a direct coordinate, z1, and two indirect coordinates, z2 and z3. We infer all kinetic parameters of the simulated data using the two distinct models and compare the inferred evidence lower bounds to find that the hierarchical model with two indirect coordinates, as expected, best describes the simulated dataset. The grayscale circles schematically denote distinct values of the direct or indirect coordinates, a full set of coordinates being required to specify the address of a particular free energy minimum on the surface.

Analysis of experimental data

In all domains of life, messenger RNA (mRNA) templates are decoded by transfer RNA (tRNA) substrates into proteins by a RNA-protein complex known as the ribosome a process known as translation. During the elongation stage of translation, tRNA substrates cycle sequentially through three ribosomal binding sites – the A, P, and E sites. Translational elongation proceeds by transfer of a nascent polypeptide chain from the P-site tRNA to the amino acid attached to the A-site tRNA; this step precedes translocation, defining a pretranslocation, or PRE complex. Next, translocation, wherein binding of EF-G and hydrolysis of a GTP to a GDP catalyzes movement of the newly deacylated P-site tRNA into to the E site,

the polypeptide-bound A-site tRNA into the P-site, and the concurrent movement of the mRNA template, sets up the next cycle of elongation. Ribosomal complexes at this step define a posttranslocation, or POST complex, which binds and decodes a new cognate aminoacyl-tRNA in the A-site in order to set up the next round of translational elongation.

PRE complexes fluctuate between two global states involving numerous configurational rearrangements of the ribosome and its bound tRNA substrates, termed GS1 and GS2[149]. In particular, in the GS2 state, tRNA substrates take configurations that allow them to interact with multiple tRNA binding sites simultaneously: a hybrid P/E state wherein the P-site tRNA interacts with both the P and E site, and a hybrid A/P state wherein the A-site peptidyl-tRNA interacts with both the A and P site. In this notation, the classical states of each tRNA are denoted P/P for the P-site tRNA and A/A for the A-site peptidyl tRNA.

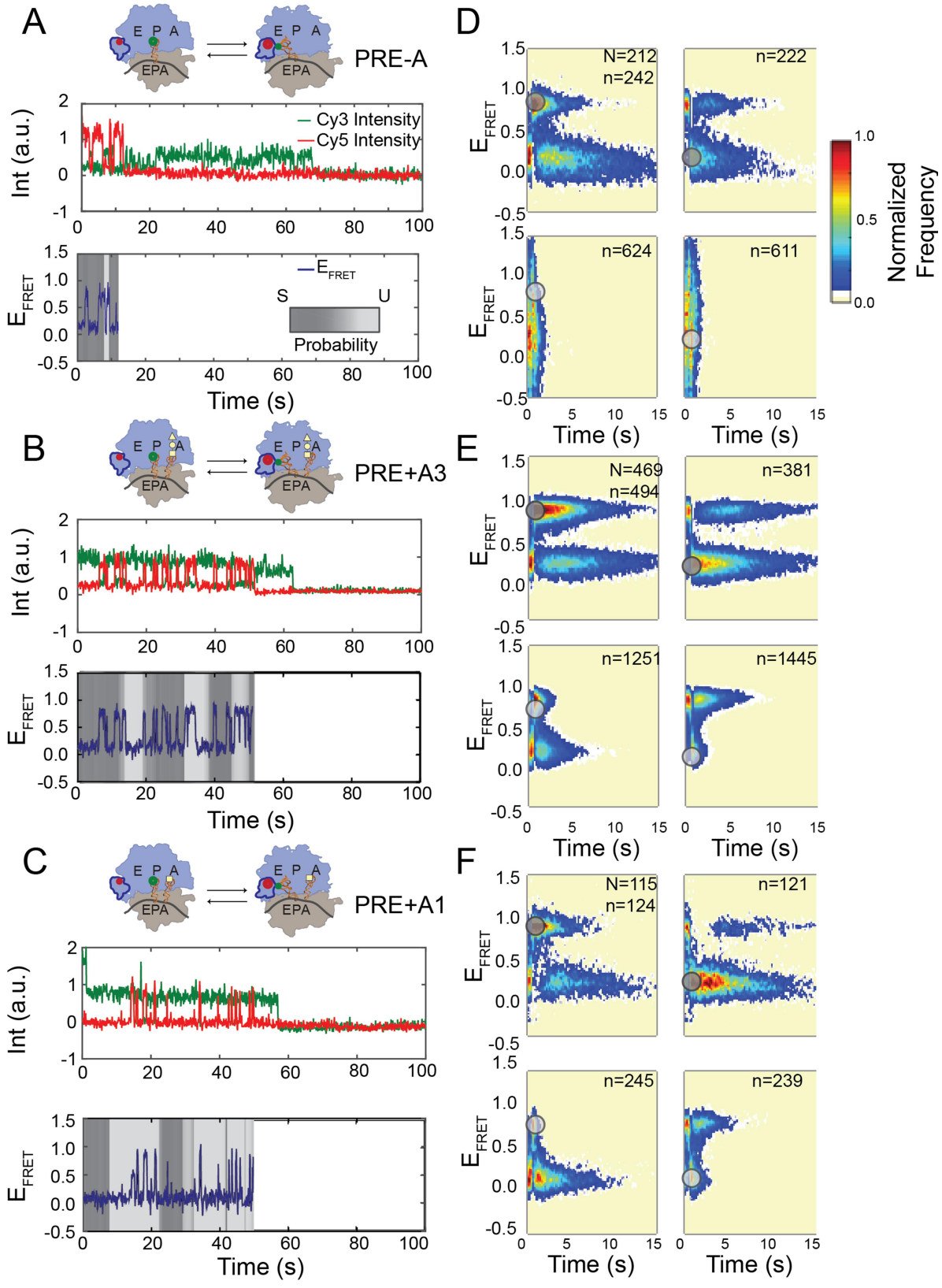

A
E P A
EPA
PRE-A
Int (a.u.)
Cy3 Intensity
Cy5 Intensity
E_FRET
S
U
Probability
Time (s)
B
PRE+A3
C
PRE+A1
D
N=212
n=242
n=222
n=624
n=611
E
N=469
n=494
n=381
n=1251
n=1445
F
N=115
n=124
n=121
n=245
n=239
Normalized Frequency

Figure 3.5 Dynamic heterogeneity of PRE complexes. Cartoons of the PRE complexes (top), sample fluorophore intensity *versus* time trajectories (middle), and the corresponding sample E_{FRET} *versus* time trajectories (bottom) are shown for (A) PRE-A (B) PRE+A3, and (C) PRE+A1. In the cartoons, the large ribosomal subunit is shown in blue with the L1 stalk structural element outlined in dark blue; the small ribosomal subunit is shown in tan; the mRNA is shown in black; the tRNAs are shown in orange; the FRET donor and acceptor fluorophores are shown in green and red, respectively; the amino acids are shown in white; and the A, P, and E tRNA binding sites are denoted on both the small and large subunits. The grayscale regions on the E_{FRET} *versus* time trajectory are linearly grayscale-weighted by the probability that a region of the trajectory belongs to either type S, dark gray, or type U, light gray. Post-synchronized 2D histograms (see main text) for (D) PRE-A, (E) PRE+A3, and (F) PRE+A1. In each panel, the initial and final observations of the sub-trajectories begin and end with the same overall set of direct and indirect coordinates, giving rise to four types of post-syncronized 2D histograms showing E_{FRET} recurrence distributions – clockwise, post-syncronized to $GS2^S$, $GS1^S$, $GS1^U$, or $GS2^U$. Initial conditions beginning with an indirect coordinate of Type S are shaded dark gray and of Type U are shaded light gray. N specifies the number of distinct trajectories, shared for all four 2D histograms, and n specifies the number of sub-trajectories for each type of post-synchronization.

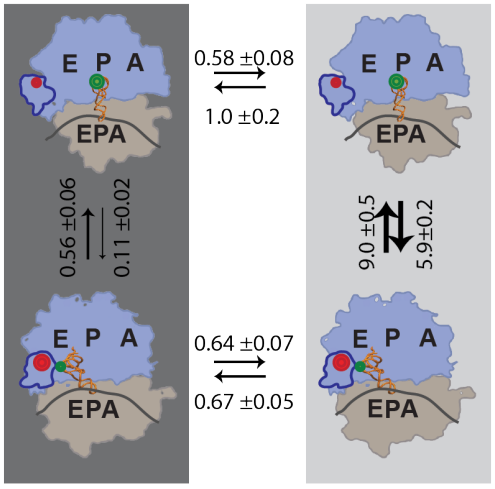
In work described by Fei *et al*[137], fluorophore-labeled ribosomal PRE complexes were prepared in order to determine the rate at which tRNA substrates in the P site fluctuate between classical and hybrid configurations and to determine the influence of the A-site tRNA on the

classical-hybrid fluctuations of the P-site tRNA. To do this, three pretranslocation complexes were prepared: PRE-A, containing a OH-(Cy3)tRNA^{Phe} but with a vacant A site; PRE+A1, containing OH-(Cy3)tRNA^{Phe} in the P site and Lys-tRNA^{Lys} in the A site; and PRE+A3, containing OH-(Cy3)tRNA^{Phe} in the P site and fMet-Phe-Lys-tRNA^{Lys} in the A site[137]. With the donor Cy3 fluorophore labeling the P-site tRNA, an smFRET signal reporting on classical-hybrid fluctuations was generated by reconstituting ribosomal complexes with acceptor Cy5 fluorophore-labeled L1 stalk, a mobile protein near the E site of the ribosome – in the hybrid P/E state, the E_{FRET} was predicted to be approximately 0.85 and in the classical P/P state the E_{FRET} was predicted to be approximately 0.20.

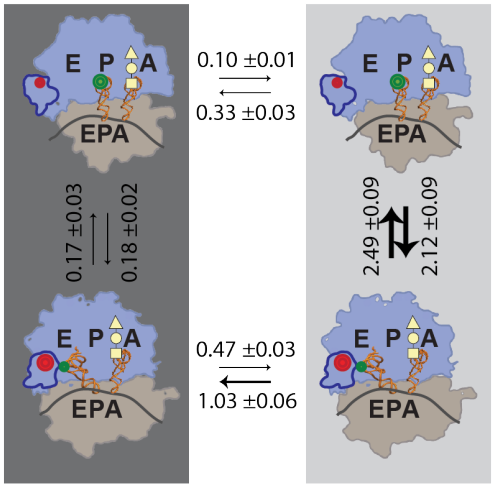
PRE-A, PRE+A1, and PRE+A3 complexes were assembled and visualized as discussed in *Methods*, and Cy3 (I_{don}), Cy5 (I_{acc}), and E_{FRET} *versus* time trajectories recording fluctuations of the P-site tRNA were extracted (representative trajectories shown in Figure 3.5 A, B, and C). In all three PRE complexes, E_{FRET} *versus* time trajectories fluctuated between a low E_{FRET} and a high E_{FRET} state, consistent with the classical (denoted GS1) and hybrid (denoted GS2) states of the P-site tRNA, respectively. The population of each of these trajectories were analyzed according to the method discussed in the previous section and the most parsimonious model in each was found to be one in which the direct coordinate fluctuates between two states, GS1 and GS2, with one indirect coordinate that can also take two values, termed S and U for stable and unstable, respectively, as the S has slower rates of conversion from GS1 to GS2 and from GS2 to GS1 than the U. This model possesses four distinct free energy minima corresponding to GS1^{S}, GS2^{S}, GS1^{U}, and GS2^{U}. Eight rate constants may be defined in each model – four along the direct coordinates GS1 and GS2 holding the indirect S or U coordinates constant, k^{S}_{GS1-G} ,

k^{S}_{GS2-G} , k^{U}_{GS1-G} , k^{U}_{GS2-G} , and four along the indirect coordinates S and U holding the direct coordinates GS1 or GS2 constant, k^{S-U}_{GS1}, k^{S-U}_{GS2}, k^{U-S}_{GS1}, and k^{U-S}_{GS2}.

A


E P A
EPA
0.58 ±0.08
1.0 ±0.2
0.56 ±0.06
0.11 ±0.02
9.0 ±0.5
5.9±0.2
0.64 ±0.07
0.67 ±0.05

B


E P A
EPA
0.10 ±0.01
0.33 ±0.03
0.17 ±0.03
0.18 ±0.02
2.49 ±0.09
2.12 ±0.09
0.47 ±0.03
1.03 ±0.06

C

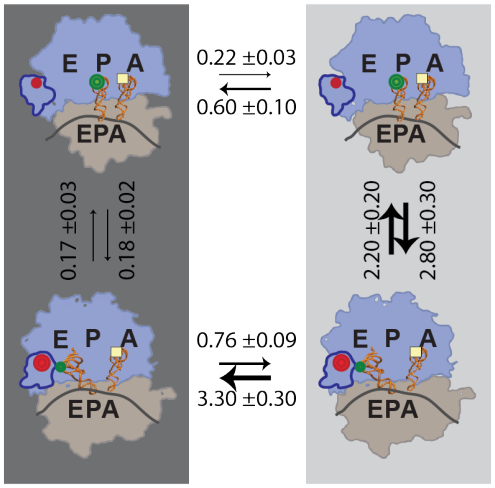

E P A
EPA
0.22 ±0.03
0.60 ±0.10
0.17 ±0.03
0.18 ±0.02
2.20 ±0.20
2.80 ±0.30
0.76 ±0.09
3.30 ±0.30

Figure 3.6 Dynamic model for fluctuations between GS1 and GS2. Fully quantified kinetic model describing the rates of transitions between GS1S, GS1U, GS2S, and GS2U for (A) PRE-A, (B) PRE+A3, and (C) PRE+A1. Note that all rates are in units of s–1. Transitions between GS1 and GS2 holding the indirect S coordinate constant, GS1S $\rightleftarrows$ GS2S, are enclosed within a dark gray box and the corresponding transition holding the U coordinate constant, GS1U $\rightleftarrows$ GS2U, are enclosed within a light gray box. On comparison between PRE-A, PRE+A3, and PRE+A1, we argue that PRE+A3^{S} and PRE+A3^{U} are equivalent with PRE+A1^{S} and PRE+A1^{U}, respectively; on the other hand, PRE-A^{S} and PRE-A^{U} are distinct from the corresponding inferred from PRE+A3 and PRE+A1.

We reasoned that the S-U indirect coordinate could correspond to changes in P-site tRNA dynamics arising from steric clashes with the A-site tRNA or conformational changes caused by the A-site tRNA. To graphically compare the three pretranslocation complexes, we first constructed a population of subtrajectories from trajectories collected from each complex where each subtrajectory begins when a trajectory enters one of the four free energy minima and terminates once the trajectory re-enters it again. These subtrajectories contain information on direct coordinate amplitudes sampled during first-order recurrence events, which we term E_{FRET} recurrence distributions. We visualize E_{FRET} recurrence distributions graphically by preparing 2D histograms of these subtrajectories recording frequency of E_{FRET} values *versus* time (Figure 3.5 D, E, and F). While the two complexes with A-site tRNA substrates, PRE+A1 and PRE+A3, possess very similar E_{FRET} recurrence distributions associated with all four free energy minima, but while PRE-A^{S} possesses similar E_{FRET} recurrence distributions to PRE+A1^{S} and PRE+A3^{S},

PRE-A^U appears distinct. Indeed, because of its high noise and short lifetimes, we surmise that PRE-A^U is an E_{FRET} statistical artifact obscured in PRE+A1 and PRE+A3 trajectories by the high evidence values associated with the true S-U indirect coordinate instead of a pair of distinct conformations of the ribosome.

Because of its strong dependence on the presence of a bound A-site tRNA, we argue that the S-U coordinate has minima with two distinct configurations of the A-site tRNA – a classical A/A state and a hybrid A/P state: difference in P-site tRNA classical hybrid configurations are affected by those of the A-site tRNA by steric hindrance between the two tRNA substrates. Having identified a clear distinction between pretranslocation complexes with and without bound A-site tRNA, we undertook a more quantitative comparison between the PRE complex with a bound A-site aminoacyl tRNA (PRE+A1, Figure 3.6B) and the PRE complex with a bound A-site peptidyl tRNA (PRE+A3, Figure 3.6C). While the rates associated with $\mathsf{GS1}^{\mathsf{U}} \rightleftarrows \mathsf{GS2}^{\mathsf{U}}$ and with $\mathsf{GS1}^{\mathsf{S}} \rightleftarrows \mathsf{GS2}^{\mathsf{S}}$ appear similar between the two PRE complexes, transitions between $\mathsf{GS1}^{\mathsf{U}} \rightleftarrows \mathsf{GS1}^{\mathsf{S}}$ and $\mathsf{GS2}^{\mathsf{U}} \rightleftarrows \mathsf{GS2}^{\mathsf{S}}$ are distinct between the two, a subtle effect distinguishing the classical-hybrid equilibrium between the two complexes, leading us to surmise that the PRE+$A1^S$ and PRE+$A3^S$ complexes correspond to those in which the A-site tRNA is in the classical conformation and the PRE+$A1^U$ and PRE+$A3^U$ complexes correspond to those in which the A-site tRNA is in the hybrid conformation. This interpretation of the analysis is discussed below.

3.2.2.4 Discussion

This work demonstrates a rigorous approach to quantifying single-molecule data whose kinetic model obeys a hierarchical Markov chain: experimental configurations wherein

fluctuations along directly observed coordinates of a biomolecular complex are affected by changes in the complex along indirectly observed coordinates.

Our model, which uses the variation approximation to optimize the evidence lower bound, provides both the machinery for estimating the interdependent parameters of a hierarchical Markov chain and for selecting the simplest kinetic schemes required to describe populations of single-molecule trajectories.

In contrast to existing methods for quantifying kinetic rates in the presence of direct and indirect coordinates[85,141–144], including a very similar model released as this manuscript was in the final stages of preparation[145], our algorithm enables experimentalists to directly quantify and select between kinetic schemes generated from hierarchical markov chains of arbitrary complexity.

We have subsequently applied this method to an smFRET signal reporting fluctuations between classical and hybrid configurations of a tRNA bound to the P site of a ribosome stalled prior to translocation. These results have revealed a kinetic model with four free energy minima, highly dependent on the A-site tRNA. In fact, we observed that the presence of an A-site tRNA and the length of the peptide attached to the A-site tRNA affects the rate of interconversion along the indirect coordinate (S/U). What is the origin of this indirect coordinate?

The indirect coordinate cannot be explained by, as proposed before[137], independent trajectories recorded from different types of molecules (static heterogeneity): we find evidence of numerous, reciprocal fluctuations along the S-U coordinate within a single trajectory (Figure 3.5.) It is possible that combination of our algorithm for dynamic heterogeneity with our algorithm for static heterogeneity, a more complex model than that presented here, is warranted; here, we simply present the most parsimonious kinetic scheme identified with our method.

As argued in *Results*, we conclude that the kinetic differences between the complex with an empty A site, PRE-A, and the two complexes with bound A-site tRNAs warrant the conclusion that the A-site tRNA modulates the classical-hybrid equilibrium of the P-site tRNA. The A-site tRNA fluctuates between a classical (A/A) and hybrid (A/P) state as well, as determined by structural and smFRET studies[150–152]. In the classical state, the terminal end of the A-site tRNA, bearing the CCA motif and covalently bound to the amino acid, contacts the 23S rRNA at a stem-loop in the A site known as the A-loop[153–156]; in the hybrid state, the same portion of the A-site tRNA contacts the 23S at a stem-loop in the P site known as the P-loop (reviewed in [150]). The lifetimes of the classical and hybrid states of the A-site tRNA are on the same timescale as those of the P-site tRNA.

We therefore argue that steric hindrance between the A-site tRNA in the hybrid state and mutually exclusive interaction with the P-loop or A-loop residues give rise to a coordinate that indirectly affects fluctuations of the P-site tRNA. With the A-site tRNA transiently occluding the P site, the rate at which the P-site tRNA returns form the hybrid to the classical configuration will depend also on the length of time required for the A-site tRNA to re-enter its classical configuration. This compounded rate is slower than the corresponding rate wherein the A-site tRNA does not transiently occlude the P site. Therefore, we will assign S to a set of configurations wherein the A-site tRNA can transiently sample the hybrid state and U to a set of configurations wherein the A-site tRNA remains in the classical configuration. We propose that the A-loop of the 23S rRNA stabilizes the A-site tRNA and leads to the faster P-site dynamics when holding U constant, but that lack of the extra configurational constrains provided by this contact lead to interactions between the two tRNAs and the slower P-site dynamics observed

when holding S constant. Therefore we propose that the observed dynamic heterogeneity arises from interaction between the two flexible tRNA substrates.

To conclude, we note possible generalizations of this work. (1) The probability distribution presented here could be modified with the goal of connecting observations from separate fluorophore labeling positions. At present, it is difficult to demonstrate the link between such experiments because they cannot be rigorously represented as a group; however, taking smFRET studies of the ribosome as an example (reviewed in [140] for example), there are a wealth of studies that are heuristically interpreted as a group, a situation calling for a unified, quantitative model. Combining information between two signals can be done in the context of this model. If static heterogeneity requires that the trees do not possess a common root, trees can be mixed just as disconnected nodes can, to construct a mixture of copies. (2) A multi-production-level tree could connect theoretical Markov-state models from molecular dynamics simulations to those directly observed in a single-molecule experiment. This model currently suffers from two ends – first, a full Bayesian approach does not exist for MD simulations, and second, the time resolution of current single-molecule experiments is too low to be directly relevant to all but extremely expensive simulations. We note, however, that recent work[20,22] has all but removed this experimental restriction, and we expect that in the near future such bridging models will become highly relevant.

3.2.2.5 Materials and Methods

Simulated data was prepared by adding white noise, to a signal-to-noise ratio of 5, to a sequence of index values of a direct coordinate generated from a hierarchical Markov chain whose inter-coordinate transition rates depended additionally upon two distinct indirect coordinates, given in tree-representation by the diagram on the right hand side of Figure 3.3. The

rate constants at each level were separated sufficiently in value so that no two sets of indirect coordinate gave rise to near-equivalent transition rates between direct coordinates to prevent ambiguity during inference by either of the models in Figure 3.3.

Single-molecule fluorescence resonance energy transfer (smFRET) data discussed in this work consists of a dataset previously reported by Fei *et al*[137]. Ribosomal protein L1 was site-specifically labeled with an smFRET acceptor fluorophore, Cy5, and reconstituted into purified 50S ribosomal subunits; a $tRNA^{Phe}$ labeled at the dihydrouridine at position 47 with an smFRET donor fluorophore, Cy3, was incorporated into the P-site of all ribosomal complexes. Three pretranslocation complexes were prepared: PRE-A, generated by deacylating the P-site tRNA with puromycin prior to peptide transfer, containing a OH-(Cy3)$tRNA^{Phe}$ but with a vacant A site; PRE+A1, containing OH-(Cy3)$tRNA^{Phe}$ in the P site and Lys-$tRNA^{Lys}$ in the A site generated by delivering an EF-Tu:Lys-$tRNA^{Lys}$ to PRE-A complexes; and PRE+A3, containing OH-(Cy3)$tRNA^{Phe}$ in the P site and fMet-Phe-Lys-$tRNA^{Lys}$ in the A site generated by undergoing three rounds of consecutive translation[137]. Ribosomal complexes were assembled onto biotin-labeled mRNA templates which were bound to streptavidin tetramers surface-immobilized by pre-conjugation to a biotin-labeled polyethylene glycol layer coating a quartz slide[137,157]. Fluorescence intensity *versus* time trajectories were collected at 50 ms time-resolution from successive images of wide-field, prism-based, total internal reflection fluorescence (TIRF) movies simultaneously recording donor Cy3 intensity, $I_{don}(t)$, and acceptor Cy5 intensity, $I_{acc}(t)$, following excitation of the donor Cy3 fluorophore with a 532 nm laser, allowing calculation of $E_{FRET}(t) = \frac{I_{acc}(t)}{I_{acc}(t)+I_{don}(t)}$. The E_{FRET} values are proportional to the donor-acceptor distance, given by $E_{FRET}(R) = \frac{1}{1+\left(\frac{R}{R_o}\right)^6}$, and the Forster radius R_o, assuming

constant quantum efficiency and free isotropic rotation of both flourophores, is 54 Å[7]. Donor and acceptor intensities from each ribosomal complex were collected in identical buffer conditions excluding excess components, such as labeled tRNA molecules or puromycin, required to generate them[137].

$I_{don}(t)$ and $I_{acc}(t)$ trajectories corresponding to single ribosomal complexes labeled with Cy3 and Cy5 were identified and extracted from TIRF movies as described elsewhere[137]. Data were truncated according to the single-step photobleaching event and the $I_{don}(t)$ and $I_{acc}(t)$ trajectories were baseline corrected by subtracting the average intensity of the last ten time points following the photobleaching event of either Cy3 or Cy5. Truncated and baseline corrected trajectories were used to calculated E_{FRET} *versus* time trajectories. These were analyzed using an HHMM as discussed in *Results*, and 2D histograms were prepared as discussed in *Results*. Software for generation of simulated trajectories and analysis of simulated and experimental trajectories was written using MATLAB R2015a.

3.2.3 A Bayesian Approach to Single-Molecule Trajectories with Diffusing Observables

The parameters describing how a measurement reports the occupancy of a state are known as the emission distribution. Typically the emission distribution is assumed to be unchanging through time. While this assumption avoids severe difficulties when the emission distribution is static, when the emission distribution varies continuously with time, the trajectory becomes inscrutable to the model. Here, a Bayesian inference framework based on the variational approximation allows for the emission distributions to vary continuously and thus enables analysis of trajectories that not only jump between discrete states, but also continuously alter the observable definition of those states.

3.2.3.1 Introduction

Single-molecule observable *versus* time trajectories, or simply trajectories, are typically modeled with emission distributions that possess constant parameters over time[80,83,85]. This is appropriate when the observable has strict constraints – for example, in single molecule fluorescence resonance energy transfer (smFRET) measurements, the FRET efficiency reports an interdye distance and reflects the fact that the overlap integral, which presents the main contribution to the energy transfer from the donor fluorophore to the accepter fluorophore, is typically constant with time at any given separation[7]. Not all observables possess such physical constraints. For instance, passive clamp optical tweezer experiments are calibrated to a given force, but there is no *a priori* reason to suspect that a given extension must be assigned to a given rip force, and indeed, those changes vary somewhat over time as a result of low frequency noise[158]. As another example, ion channel measurements occasionally possess an incomplete seal about the membrane, causing a small leakage current that continuously varies conductance class assignment[86]. More practically, measurements of conductance *versus* time trajectories using carbon nanotube single-molecule field effect transistors described in Chapters 4 and 5 possess significant low-frequency noise, arising from fast, correlated charge fluctuations in the solution gate electrode, which can be essentially considered a manifestation of a random walk. Collectively, this phenomenon may be referred to as "emission drift."

Solutions to the emission drift problem have been proffered by the ion channel community. In particular, two solutions exist: first, a "metastate" maximum likelihood hidden markov model (HMM) utilizing time lags and direct fits to the Yule-Walker equation was implemented to estimate rate constants with small emission drift[79]; second, a maximum

likelihood Gaussian mixture model was designed to deconvolute multiscale emission drift using a direct diffusion model[8 6].

To improve these models, my approach has been to employ a Bayesian inference based framework to the graphical model implied by the diffusion model for emission drift[76]. Bayesian inference mimics the scientific process by allowing assumptions (prior distributions) to be updated by observations, leading to better estimates (posterior distributions.) In doing so, here I improve on existing tools by naturally quantifying the experimental- and ensemble-derived uncertainty. This is particularly crucial for datasets exhibiting emission drift, as the noise associated with a measurement can approach a level wherein the human eye is an insufficient, whereas variational inference, which encourages parsimony, can provide quantitative measures to aid proper model selection.

3.2.3.2 Methods

Implementation of a HMM that corrects for emission drift using a diffusion process utilizes the same Bayesian network, and therefore identical equations, as the usual HMM except for the presence of an additional variable which tracks the diffusive state of the trajectory. The values of this variable, called the "baseline," are assigned by inverting the finite-difference Laplacian uniquely associated with the particular trajectory. Parameter estimation is schematized in Appendix A. Simulated data to test the model was prepared by generating a trajectory and adding 1/f noise of varying amplitude directly to the state space variable z_{nt}.

3.2.3.3 Results and Discussion

The primary use of the emission drift HMM in this thesis will be to deconvolute conductance *versus* time trajectories recorded from single-molecule field effect transistors. Therefore simulations have been prepared to demonstrate that the algorithm can actually perform

this task, schematized in Figure 3.7. For a wide range of noise amplitudes, the algorithm accurately reconstructs the rate matrix, and comparatively increases the SNR available for analysis by a factor of 4. This is especially important for long trajectories containing dynamic heterogeneity, such as those in Chapter 5.

3.2.3.4 Conclusion

Combining the emission drift model presented here with all the other computational methods presented in Chapter 3 presents a toolkit that can deconvolute almost any jump-markov problem. This code has been prepared, tested, and will be used extensively throughout subsequent chapters to quantify single-molecule field effect transistor data.

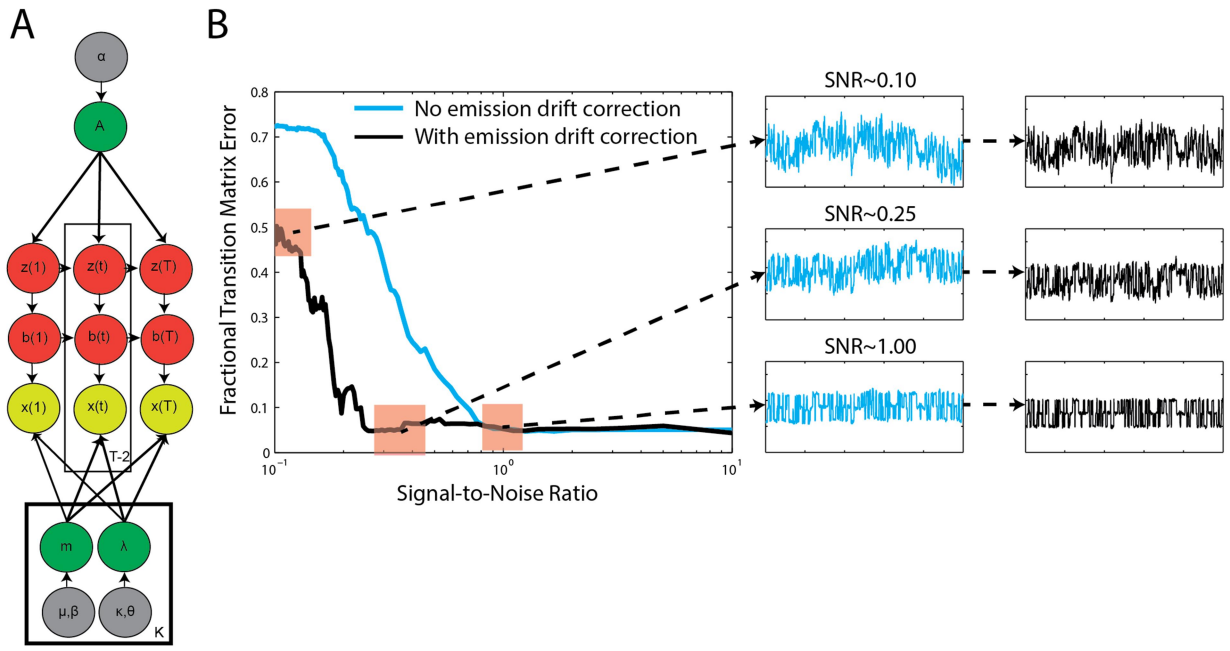


Figure 3.7 Design and validation of the emission drift Hidden Markov Model.
(A) Graphical model of the Dynamic Bayesian Network. Gray, prior parameters; red, model realizations; green; expected parameters; yellow, observed data. (B) Noise limitations of the model, and comparison to conventional HMM, evaluated using a 3-state markov chain.

Fractional transition matrix error is defined by $Error = \sum_{i,j} \frac{\left|A_{ij}^{sim} - A_{ij}^{est}\right|}{A_{ij}^{sim}}$ where A_{ij}^{sim} is the simulated transition matrix and A_{ij}^{est} is the transition matrix measured from the simulated data with a mixture of $1/f^{1.7}$ noise, a common value measured in trajectories recorded using the methods in Chapters 2, 4, and 5. The signal-to-noise ratio is defined by dividing the separation between two states ΔS by the coefficient B of the noise amplitude (i.e., the power spectrum is $B/f^{1.7}$.) The same signal at several SNR values is shown on the right.

Part 2: Dynamics of nucleic acids on the microsecond timescale

The past half-century has witnessed an explosion in research demonstrating that DNA is not simply a carrier of information, and that RNA plays substantially more roles than as a transcribed intermediate between DNA and protein[1]. Like proteins, to execute their roles, these molecules fold up into complex three-dimensional shapes to assemble conserved recognition, catalytic, and regulatory sites. The question has become, not whether nucleic acids can play these roles, but how; and this question deeply entwines with how the molecule gains its shape and interconverts between competing conformations.

Compared with proteins, nucleic acids at first glance have very few monomer components: in contrast to the twenty-one canonical amino-acids making up most proteins, there four distinct bases in DNA, four distinct bases in RNA, and both molecules have consistent negative charge resulting from the phosphate backbone, which is always respectively segmented by the same deoxyribose or ribose sugar. This comparative lack of diversity is illusory. First, in modern organisms, RNA and DNA rarely carry out regulatory functions alone, often balanced or augmented synergistically with protein partners: for example, Cas9, a key component of the bacterial immune system, must be loaded with a proper guide RNA which it then uses to search for potential foreign nucleic acid targets to cleave[159]. Second, although the monomer composition is limited, the interaction space is enormous, because all the monomers have potential interactions with each other: for example, other than a Watson-Crick base pair, nucleotides may form Hoogsteen base pairs on the opposite edge; bulges or loops may insert into helices to form A-minor motifs; disparate stem-loops may coalesce into a “kissing” interaction;

bulges may protrude from a helix to provide an interaction surface for other RNA or protein molecules; the nucleic acid may bend over onto itself to form d-loops, t-loops, or pseudoknots; four-way junctions may diffuse up and down an enormous helix; and of course, all at once (see Figure 1.1 for a few selected examples)[160]. Third, the bases themselves possess many post-transcriptional modifications – for example, human genomic DNA is extensively marked by methylation, and functional RNAs such as tRNAs bear extensively regulated post-transcriptional modifications[1]. Finally, like proteins, RNA especially and to some extent DNA, possesses programmed modules which are thought to rapidly pre-form by way of competition with alternate structures[161].

Conformational heterogeneity and subsequent interconversion between individual atomic configurations, therefore, has been proposed as a major mechanism by which biological outcomes are regulated[2]. These interconversions do not take place instantaneously in a single step. In many cases, the elementary steps that compose entry and exit from the transition state involve breaking or re-forming individual base pairs. While the timescale of this process has been measured, events like these are difficult to resolve on the single-molecule level because they are exceedingly rapid and distance changes within the nucleic acid are negligible compared to the resolution of contemporary techniques. Because the single-molecule field effect transistors described in Chapter 2 sense very small changes in charge density in the carbon nanotube vicinity with microsecond time resolution, these devices are uniquely suited to the study of DNA and RNA dynamics.

In Chapters 4 and 5, I present two distinct studies of RNA dynamics. Chapter 4 is primarily concerned with the formation of specific folding modules, in particular, how the formation of conserved loop modules guides the kinetics of base pair rearrangements within the

context of an RNA stem-loop. Finally, culminating all these viewpoints, Chapter 5 investigates how the dynamics of an RNA switch are guided by shifting structural modules above, and shows how this dynamic heterogeneity offers a detailed theory potentially accounting for the operation of the switch. Collectively, these studies provide some foundational research into the fast-paced and conserved motions underlying the RNA world.

Chapter 4 Direct measurement of base pair-by-base pair zipping and unzipping of individual RNA stem-loops

4.1 Introduction

RNA secondary and tertiary structure formation has been commonly represented as a non-directed graph bearing varying thermodynamic weights of occupancy and kinetic weights of formation and disruption of the edges. The polymer navigates the graph *via* transient population of intermediates, eventually populating each according to its equilibrium distribution[95,96,162]. In conditions typical of biology, these dynamics resemble a random walk[88]. Steps are considered stochastic because the environment constantly fluctuates and typically relative extrema of environmental fluctuations are required to qualitatively perturb the graph[87]; a brief discussion of some aspects of stochastic dynamics are described in Chapter 3. The purpose of this chapter is to describe the use of the single-molecule field-effect transistor (smFET) technique to measure the rate constants of the pairing and unpairing of individual RNA base pairs in the context of highly conserved folding motifs.

Two questions immediately jump out given the view of RNA folding as a random walk with discrete steps: first, what is the diversity and what are the timescales of potential steps? And second, given that state space complexity increases exponentially with the number of subunits as the number of ways that the polymer can interact with itself increases combinatorically with the

number of compositional monomers, how does RNA both utilize and manage that complexity? To answer the first question, investigators have turned to NMR and temperature jump hypochromicity studies, the former of which indicate that, once begun, the lifetime of a paired base in a helix lies between 1 and 100ms[163–166]; and the latter of which indicate that the rate of transfer from the paired to the unpaired state and *vice versa* lies in the 100ns range[95,96,167] (a rather complete reference can be found in [160]). So-called toeholding experiments which monitor the invasion of one strand of DNA by another typically predict a similar lifetime[168,169]. To answer the second question, investigators have taken to comparisons of sequences of RNA that are highly conserved across domains of life, under the banner of predicting structure from sequence. In the early 1990s, such questions and comparison of sequences of ribosomal RNA led to the astounding observation that more than half of all ribosomal stem-loops are capped by loops consisting of four nucleotides and that these four nucleotides were most usually either GNRA (guanine, anything, purine, adenine) or UNCG (uracil, anything, cytosine, guanine)[170]. Additionally, UNCG stem-loops were also identified in some viral transcripts (T4 bacteriophage)[171]. Unsurprisingly, biophysical characterization using UV melting at high salt concentrations revealed that there was something special about this sequence – stems capped with GNRA or UNCG tetraloops melt at temperatures 10-20 ˚C higher than their more random counterparts[172,173].

This observation, as well as their phylogenetic prevalence, has led to two separate but nonexclusive proposals for tetraloop function. First, by comparing sequences of known RNA interactions containing tetraloops[174], it was predicted[175] and later confirmed[176–180] that tetraloops act as the donor of a donor-acceptor motif in order to mediate RNA tertiary interactions as well as to act as binding sites for other RNAs . A prototypical example is the role of GNRA tetraloops

in spliceosome assembly, a topic reviewed in [181]. Alternatively, tetraloops can serve as adaptor sites for binding of proteins. A prototypical example is ribosome assembly, reviewed in many places but [182,183] containing a fascinating perspective. A more specific example is the GTPase associated center of the ribosome, which contains a conserved GAGA loop, type GNRA, called the sarcin-ricin loop, or SRL, which serves as part of an essential binding and regulatory site for GTPases during the elongation and termination phases of translation. Second, it was proposed, on the basis of their unusual stability, that RNA structures bearing loops of this type serve as co-transcriptional folding nuclei – for example, ribosomal RNA, bearing hundreds of bases, has on its face a very complicated folding ensemble, which is conceptually simplified[184] when one considers that stems capped by hyperstable tetraloops likely fold first and are unlikely to fully unwind before subsequently transcribed modules can begin to interact[185–188]. In this matter, the simplicity of the loop is key – to create the proto-structure of a tetraloop the only interactions required are that three adjacent bases stack and two bases pair[189]. Indeed, this folding process begins the moment the RNA polymerase slides out of the way, just after the RNA is transcribed[190].

These proposals bring the search for a conservation pattern beyond the RNA sequence itself and into the realm of biophysics. To evaluate proposed function, generation of structural data was a logical first step. Early comparisons of GNRA, UNCG, and other loops based on their crystal and NMR structures led to a phylogenetic analysis of sorts revealing that patterns in structural morphology, which have much to do with donor-acceptor motifs and even more relevance for stability, can be clustered independently of sequence, i.e. there are numerous "letter" violations, for example GNRA-sequence tetraloops that actually look like UNCG tetraloops. However, there are not so many configurational violations – most loops can be

classified[191,192]. Further comparison of these structural clusters, with the addition of even higher resolution ribosomal and RNA structures, has allowed the presentation of what may be considered the core set of structural strategies[193]: numbering the tetraloop 1-4 from the 5' end, a donor-acceptor interaction between two RNAs will typically require an interaction between the ribose sugar of the base in position 1 and the nucleobase in position 4; furthermore, structures can be directly classified into two types of loop fold regardless of sequence – a U-turn, favored in general by GNRA-type, and a Z-turn, favored in general by UNCG type . The latter is, in general, more stable. Enhanced stability is proffered if the closing base pair (0-5) are a C and a G, in that order[173,194]. A U-turn consists of a reversal in phosphate backbone direction stabilized by a base-phosphate hydrogen bond between the first and fourth base respectively as well as a stacking interaction between the first and third base; a Z-turn is distinguished from a U-turn, because there is a base pairing interaction between the first and fourth base and a stacking arrangement between the third and fourth bases. It is important to observe that the two types may interconvert within a single sequence, and furthermore, that with this definition it is not required that the loop consists of four bases. These rules have crystallized what may be considered the core family of conserved stem-loops folding strategies.

What characterizes the dynamics of tetraloop folding motifs? As mentioned above, two key questions have been thus far investigated. First, how does the native structure assemble from an unfolded ensemble? And second, how do tertiary docking interactions involving tetraloops assemble and what are, and what principles govern, their timescales?

Assessing the assembly of tetraloop structures has been mainly viewed *via* experiments focusing on their disassembly. Experimentally, this can be achieved by replacement of an adenine in a loop with its fluorescent near analogue 2-aminopurine (2AP) and measuring bulk

fluorescence following a sudden increase in temperature, or using the single-molecule optical tweezers technique to pull on either end of a stem-loop within a laser force trap. Theoretically, this can be achieved by all-atomistic molecular dynamics (MD) simulations. These assays have, collectively, suggested the following model for folding of stable tetraloops: the nucleic acid in its single-stranded form is not a random coil; rather it is a collapsed form which is already near a state where a loop forms. Following stochastic nucleation of a proto-loop, or globular intermediate state, which is predicted to possess an approximately 10-nanosecond-scale lifetime[186], subtle rearrangements in the loop that force more optimal stacking between the 3rd and 4th position have been proposed on the basis of 2AP relaxation occur within the next microsecond[195]. After these two events have occurred, the hairpin explores a sequence-dependent dynamic transition state involving rapid rearrangement of the stem base pairs, which under conditions of force is centered at least 3 base pairs distal to the loop itself[196–198]; the rate constant of these individual transitions, under force, is faster than 100 microseconds[197], *via* NMR, estimated at 1-100 milliseconds[163–166], and *via* temperature jump experiments, estimated as fast as 100 nanoseconds[95,96,199–201]. At equilibrium, lower than the melting temperature, only the latter, base pair rearrangements, are expected to be populated to a large degree. It should be noted that, unsurprisingly for a polyanion such as RNA, these rate constants are highly salt dependent – folding is accelerated in the presence of high concentrations of salt (a salient discussion may be found in [160]).

Dynamics of docking interactions, on the other hand, have been mainly studied *via* single-molecule fluorescence resonance energy transfer (smFRET), fluorescence correlation spectroscopy (FCS), and NMR, and are thus not typically modeled beyond two-state kinetics involving transfers between an “unfolded” and a “folded” state, and are dynamic on the 100ms-

10s timescale. NMR structures have revealed a mainly U-turn motif following docking of GNRA tetraloops into the minor groove of the acceptor RNA (a so-called A-minor motif[202]), indicating that rearrangement of the loop architecture may not be necessary beyond distortion of base stacking within the loop. Interestingly, smFRET studies have revealed a magnesium cation dependence that relies on two states of the loop as well as two states of the docking site, both of which are modulated into a more favorable configuration by its addition[203] (reviewed in [204]). It is not known, however, what the consequences of docking are on stem stability.

This mixed approach of phylogenetic, structural comparative, kinetic, and theoretical analysis has been extraordinarily fruitful with respect to the study of tetraloop folding. However, while this approach has been successful in some cases in explaining function as well as conservation pattern, questions remain. The question investigated here is: how does the loop and loop-type affect dynamics of the bases in the stem? Such questions require high time-resolution methods to answer and furthermore, because the kinetics of nucleobase rearrangements are highly correlated to each other, single-molecule techniques and single-molecule field effect transistors (smFETs) in particular are highly suited to this study. The implications of answering this question are clear – beyond enabling a kinetics-based classification of stems, as described above, smFETs are currently the only equilibrium single-molecule tool able to probe stem dynamics, one of the primary tools used by nature to fold RNA tertiary structures and gauge interactions between stems and other molecules, and.

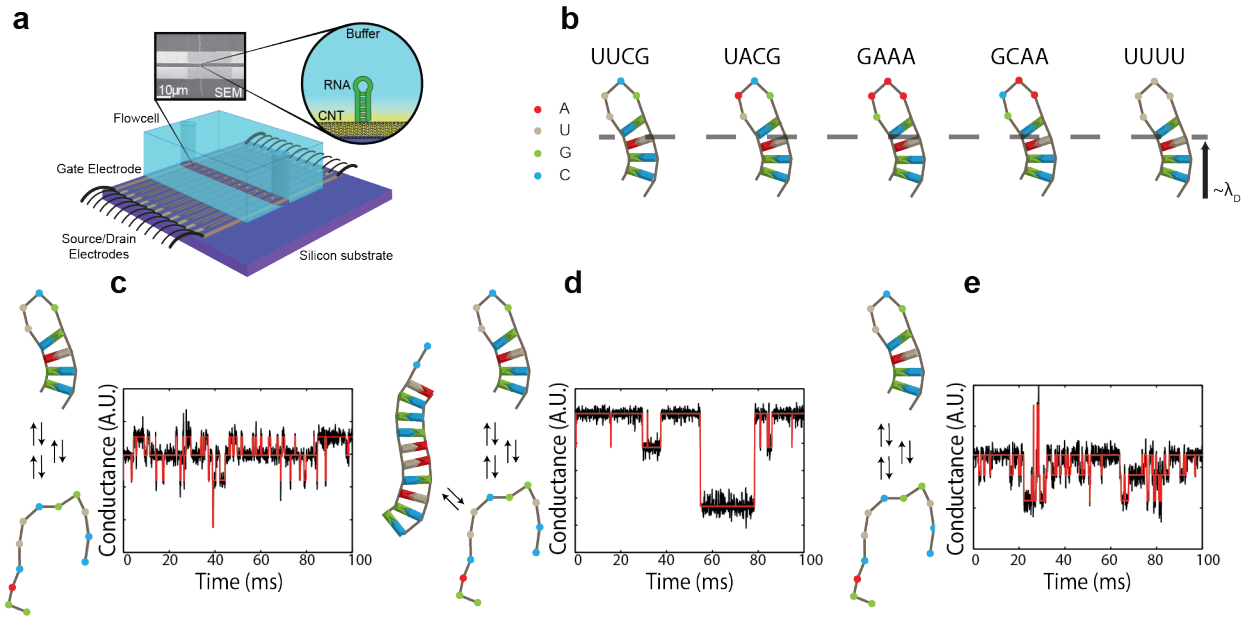


Figure 4.1 Schematic of smFET measurement geometry, stem-loop constructs, and experimental conditions.
(a) Flowcell setup for an smFET, wherein 60 devices are simultaneously measured within a PDMS flowcell. (b) Predicted secondary structure cartoons of each of the 5 stem-loop constructs tested. (c) A subset of a sample trajectory recorded in the absence of complementary or non-complementary DNA for the UUCG stem-loop; (d) in the presence of 1 μM complementary DNA; (e) in the presence of 1 μM non-complementary DNA.

4.2 Results

RNA constructs, 5'-$NH_3(CH_2)_6$-GGAC$L_1L_2L_3L_4$GUCC-3' were purchased from IDT, with $L_1L_2L_3L_4$ consisting of either the GAAA, GCAA, UACG, UUCG, or the relatively unstable[172] UUUU stem-loop motifs and whose 5' bases were modified with a primary amine were attached to carbon nanotubes (CNTs) using the pyrene-NHS anchor as schematized in Figure 4.1 (see Figure 3.3 in Chapter 2 for CNT functionalization methods within

polydimethylsiloxane (PDMS) flow cells (identical stems have been studied in [172,173]); RNA was purified through an anion exchange hitrap QHP column across a 0.1-1M NaCl gradient at pH 7. This procedure yields two major peaks for each construct. Isolating either peak and pushing it once more through the column yields again the same two peaks. Subsequent purity was assessed using a 20% D-PAGE with 20% formamide and staining by toluidine blue.

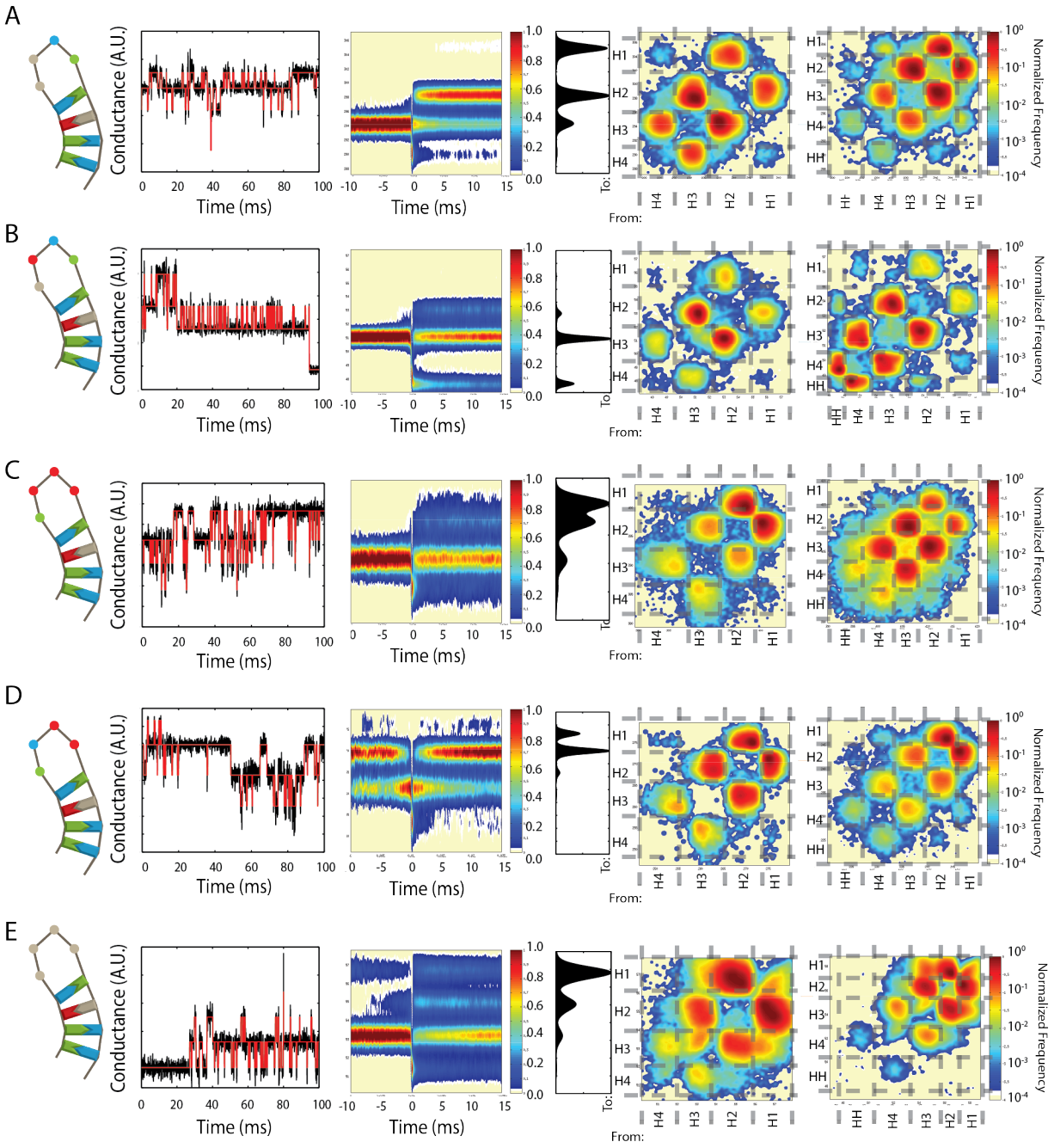


Figure 4.2 Transition pattern and population evolution of stem-loop constructs.
From left to right, cartoon representation, sample trace, 2D histogram (see text), transition density plot in the absence of complementary or non-complementary competitor DNA, and transition density plot in the presence of 1μM competitor DNA for the (A) UUCG (B) UACG

(C) GAAA (D) GCAA (E) UUUU stem-loops.

Following extensive washing with buffer (10mM phosphate, buffered to pH 7 by mixing mono- and disodium phosphate, 100mM NaCl), three separate 10 minute recordings of conductance *versus* time trajectories, or just trajectories, were collected (gate voltage V_g=-300 mV, see Chapter 2 for a FET overview) on each of the five constructs at 50 μs time-resolution: (1) a recording in the presence of 1 μM of a DNA sequence complementary to the RNA stem loop; (2) a recording in the presence of 1 μM of a DNA sequence with no complementarity to the stem loop; (3) a recording in just the buffer. A sample trajectory from each of these conditions is shown in Figure 4.1 c-e. If trajectories showed quantal, competitor sequence-specific fluctuations, then these fluctuations were supposed to arise from fluctuations of the stem loop and were analyzed further using a Hidden Markov Model with baseline correction (see Chapter 3).

Comparison of the trajectories revealed that in buffer with non-complementary DNA in solution or in buffer without competitor DNA of any sort, every smFET with stem loop-dependent fluctuations possessed four conductance classes, numbered CC1-4 from highest to lowest conductance. A fifth conductance class, CC5, appeared in trajectories when DNA complementary to the stem loop sequence was added to the flow cell. I argue that these five conductance classes arise from decreased or increased flexibility of the phosphate backbone arising from zipping and unzipping base pairs of the stem loops because the calculated Debye length under the assay conditions, λ_D defined in Chapter 2, predicts that the phosphate backbone of at least three base pairs are expected to contribute to the signal and because addition of DNA

complementary to the stem loop causes the appearance of an additional conductance class with a very specific transition pattern, discussed below.

To interpret the conductance classes, as in previous studies[18,205], the more compact form of the nucleic acid is associated with the lower conductance class. This interpretation is supported by the transition pattern of CC5 in the presence of complementary DNA, discussed below. It is important to note that every smFET has a different baseline conductance owing to variability in the structure, defect density, or electrode contact of the CNT fragment used to form each device. For these reasons, absolute smFET conductance values cannot be assigned to specific molecular conformations. Instead, relative changes in device conductance are correlated to predicted structural rearrangements of the attached molecule. Apprehended of these desiderata, I argue that these conductance classes correspond to distinct states of the stem-loop that have a different terminally paired base in the stem-loop: for example, the class with the 5' base paired is proposed to correspond to CC4, and CC1, the class with the highest conductance, is proposed to correspond to a mixture of the state with the only intact base pair as loop-closing pair and the unfolded state. These base pairing states of the stem-loops will be referred to as H1-4 (CC1-4). On the other hand, CC5, which is unique to trajectories recorded in the presence of complementary DNA, will be referred to as HH as it likely only arises when the RNA and DNA strands are in a paired state.

To quantify the rates of transition between the base pairing states, trajectories were analyzed using a four- or five- state HMM and the rate constants were extracted from the transition matrix of the fit (see Appendix A and Chapter 3). The plots described below are given in Figure 4.2. To show the unperturbed base pair lifetimes, post-synchronized 2D histograms were prepared for each stem-loop construct, in the presence of buffer without any DNA in

solution, by splitting the trajectory into subtrajectories that each begin with occupation in H4 and terminate after occupation and subsequent dwell in H1. These subtrajectories correspond to events that unpair the portion of the stem-loop helix that the smFET can detect. Transition density plots, which depict the originating and terminating conductance class after a jump-transition, were prepared for the trajectories recorded with buffer and, separately, in the presence of complementary DNA. In agreement with a zipping and unzipping model for the helix, while H1-H4 primarily transition between adjacent base pairing states. HH may transition to and from H1 and H2 directly, bypassing H3 and H4, in agreement with a model wherein the invading DNA only productively interrogates the RNA when the RNA bases are unpaired. There are additionally reciprocal transitions between HH and H4, perhaps indicating that the DNA-RNA hybrid possesses similar dynamics to the stems; however, the fact that HH may transition to H1 or H2 directly indicates that unfolding of the DNA-RNA duplex is in some cases initiated from the U-A/A-T rich region which formed the erstwhile loop.

To compare the thermodynamic weights of each of the base pairs of the stem-loop, two thermodynamic measures were prepared for the measurements in the absence of complementary or non-complementary competitor DNA. First, the fractional occupancy of each state H1-H4 was measured directly from the trajectory (Figure 4.3a). From this, one immediately concludes that the UACG stem-loop is the most thermodynamically stable of the five stem-loops tested in the sense that it spends the least amount of time in H1, and that the UUUU stem-loop is the least thermodynamically stable of the five stem-loops tested according to the same measure. Because most transitions were between adjacent conductance classes, in accord with a zipping and unzipping model for adjacent base pairing states, a second thermodynamic measure Δg_i, shown

in Figure 4.3b, can be calculated from the kinetic rates $k_{i,i+1}$ and $k_{i+1,i}$ between adjacent base pairing states as $\Delta g_i \equiv \frac{\Delta G}{k_B T} = \log\left(\frac{k_{i+1,i}}{k_{i,i+1}}\right)$.

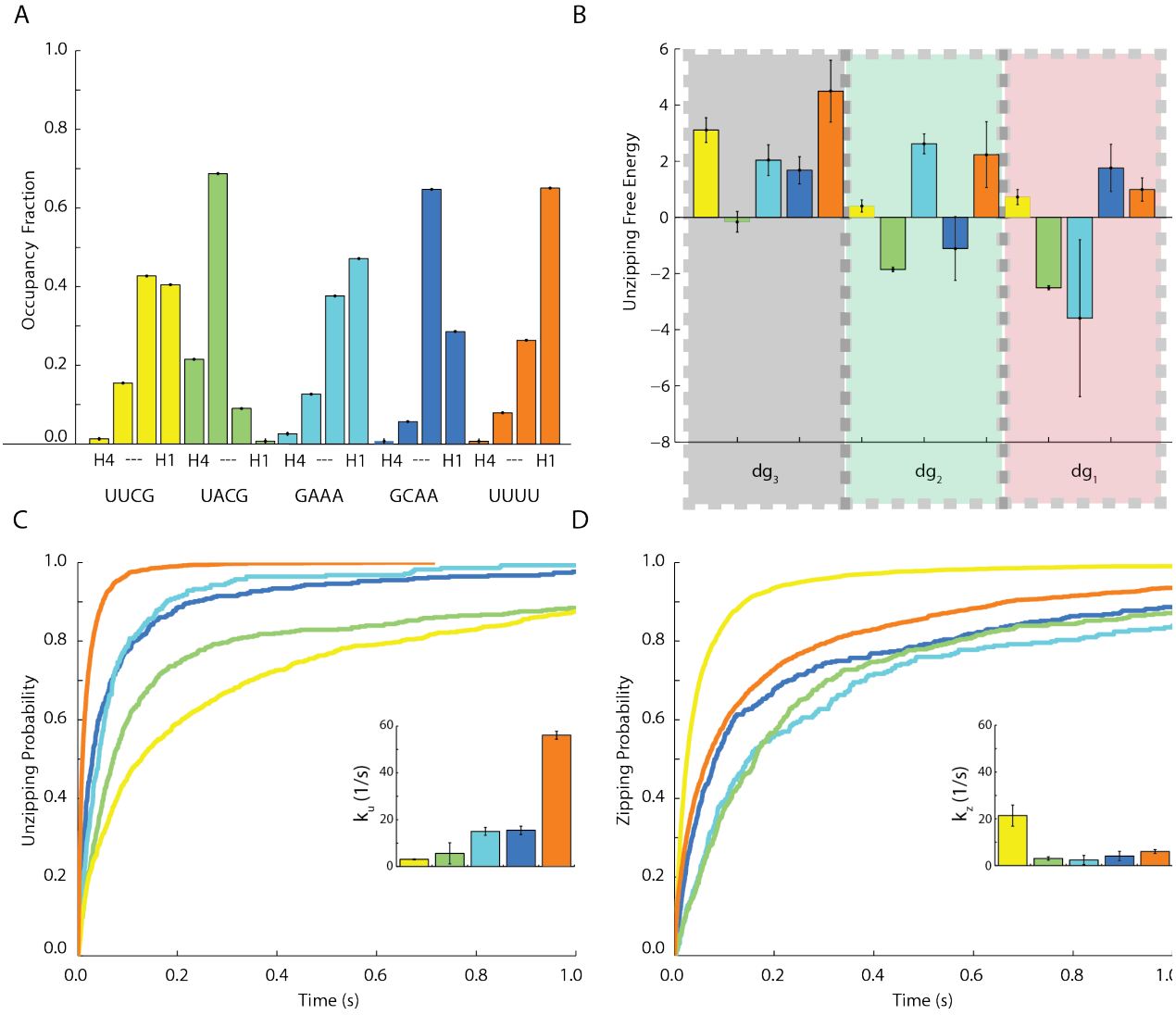


Figure 4.3 Thermodynamic and kinetic measures calculated from trajectories in the absence of any competitor DNA.
(a) Occupancy fractions of each of the four base pairing states H1, H2, H3, and H4 for the five stem-loops. (b) Δg_i for each base pair of each of the five stem-loops, described in the text. Error bars are 70% confidence intervals obtained from the transition matrix distribution. (c) Unzipping

cumulative probabilities for the first-passage times beginning in H4 and terminating in H1; inset is k_u. (d) Zipping cumulative probabilities for the first-passage times beginning in H1 and terminating in H4; inset is k_z. Both apparent rate constants were obtained by directly fitting cumulative distribution to the equation $p(t > T) = e^{-kt}$. Error bars are 70% confidence intervals obtained from the fit.

This measure, which I call the "unzipping free energy difference," estimates the relative free energy difference between subsequent terminally paired states. Comparison of the five stem-loops using this measure reveals that, while the terminal base pairs of each construct have near-equivalent Δg_i, the closer the terminal base pair to the loop the more diversity between the five constructs. The exception to this pattern is the UACG stem-loop, which has a consistently negative Δg_i, which is also reflected by the fact that this stem-loop spends most of its time in H3.

To compare the stem dynamics of the five loop constructs, two dynamic measures were prepared. First, by sorting through the trajectory and identifying when the stem-loop entered H4, we count how much time passes before it enters H1; by fitting the cumulative distribution of these first-passage times to an exponential, these give the unzipping mean first passage rate for the stem loop, k_u. Likewise, by counting how much time passes before entering H4 starting at entry into H1 and fitting the cumulative distribution of the subsequent first-passage times to an exponential, one can measure the zipping mean first passage rate of the stem loop, k_z. The cumulative distributions and apparent rate constants are shown in Figure 4.3 c and d. k_u increases in order UUCG<UACG<GAAA≅GCAA<UUUU, indicating the relative kinetic stability of the respective stem-loops, in agreement with the previously reported melting temperatures. On the other hand, the k_z of each stem loop is essentially equivalent, with the notable exception of the

UUCG loop, which is approximately 5-fold faster than the others. The fact that the UUCG stem-loop has both a faster k_z and slower k_u than the UACG stem-loop would appear to be at odds with the fact that the occupancy of H1 of the UACG stem-loop is 10-fold lower than the occupancy of H1 of the UUCG stem-loop. This apparent discrepancy between the thermodynamic and kinetic measures is explained by the consistently negative Δg_i values of the UACG loop – in particular Δg_1, which is positive for the UUCG and negative for the UACG stem-loop.

4.3 Discussion

This chapter has described fluctuations in smFET conductance which have been interpreted as arising from alterations of base pairing states in RNA helices. In the introduction to the chapter, I outlined three ways that the results and their interpretation could be compared to the copious body of previous work: first, the measured melting temperatures[172,173,188,194,206] can be compared to the thermodynamic properties inferred from the smFET trajectories; second, the thermodynamic barriers between individual base pairs may be compared to expectations founded on predictions from the nearest-neighbor model[160]; third, opening and closing rate constants of individual base pairs may be compared to the expectations of a zipping model on the basis of lifetimes estimated *via* NMR[163–166,207] and force spectroscopy[197,198]. I will discuss each of these as a separate topic. Finally, I will introduce two more comparisons, which I see as opportunities for future study suggested by these results: first, between the two GNRA stem-loops and the two UNCG stem-loops; and second, looking at the trajectory-derived role of the so-called closing base pair, *i.e.* the last base before the loop.

On its face, the melting temperature trend from previous work, which suggests the following order of stability, UUCG>UACG>GAAA>GCAA>UUUU, for the four stem-loops studied, is at odds the order of stability obtained from comparison of the observed population of the non-H1 states in the smFET trajectories, UACG>GCAA>UUCG>GAAA>UUUU. However, the ratio of the kinetic measures k_u and k_z, corresponding to a pseudo-two state equilibrium constant agree with the melting temperature trend from previous work. Additionally, each of these aggregate rate constants are quantitatively similar to rate constants obtained by fitting the results of single-molecule fluorescence correlation spectroscopy experiments to a two-state model, an approach that yields correlation times between 0.1 and 10 ms.[208,209] Likewise, the aggregate rate constants obtained here are quantitatively similar to rate constants obtained from the analysis of single-molecule FRET melting experiments recorded on stem-loops with extended 30 nucleotide loops, experiments that typically yield two interconverting states with rates of transitions between 1 s^{-1} and 60 s^{-1}.[210] Why then the thermodynamic discrepancy? I argue that the melting temperatures, evaluated on the basis of bulk measures and on the assumptions of two-state models, are evaluated on a footing more commensurate with the summary kinetic measures rather than the occupancy measures. However, the smFET trajectories provide a wealth of new information which is inaccessible to hypochromicity experiments. For example, analysis of this data has provided the individual forward and reverse rate constants for each individual base in the stem-loop, with the possible exception of that of the closing base. From this study, one may make a strong argument that two-state models for stem-loop unfolding are overly simplistic because the stem-loop does actually spend a significant amount of its folding and unfolding time paused at individual intermediate paired states. This is, in fact, no different than

recent proposals for stem-loop folding and unfolding mechanisms, which have recently been applied to hypochromicity data[96,162].

While this work did not describe a comprehensive set of pairwise base switching experiments, the unzipping free energy differences measured in this work are in qualitative agreement with the nearest neighbor model. This can be seen in Figure 4.3b, wherein the coordinate with the most bearing on kinetic differences between base pairing lifetimes is simply proximity to the loop sequence, which varies between the five constructs; a notable exception being the UACG stem-loop. This is what would be predicted from the nearest neighbor model, because the only difference contributing to kinetic difference between base pairs of the different stem-loops is their proximities to different stems.

The individual terminal base pair lifetimes measured from the smFET trajectories all lie within the 1-100ms range, in agreement with measurements from NMR studies[163–166,207], but are significantly longer-lived than those estimated *via* force spectroscopy[197,198]. This is in accord with an entropic argument stating that, under force, configurational diffusion is restricted and therefore the folding rates are in general more rapid, whereas extrapolation to no force is therefore difficult to ascribe[9]. The kinetics of individual base pairs can be resolved *via* the smFET technique in such a way that heterogeneity in the rate constants can be resolved as well, though such an analysis on the present trajectories and constructs was deemed unwarranted by the observations. An interesting future study would explore the magnesium dependence, reviewed in [211], of the dynamic heterogeneity of the type described in Chapter 3 of the various stem-loops. In particular, it is possible that under high-magnesium or in the presence of appropriate co-factors, individual stem-loops could undergo U-Z transitions described in the introduction above, contributing to distinct base-pairing kinetics within the stem-loop.

By way of comparison, the two GNRA tetraloops tested had near-equivalent kinetic properties whereas the two UNCG tetraloops were distinct from one another. The subtle differences in kinetic measure between the two GNRA tetraloops can possibly be ascribed to differential stacking within the loops[195]. The differences between the two UNCG tetraloops, however, is a somewhat surprising result, especially given the fact that the two stem-loops have near-equivalent melting temperatures[173]. These differences are mitigated somewhat on comparison of k_u and k_z, which suggests that the two are, by these measures, near-equivalent. The UUUU stem-loop, however, displayed decreased stability of nearly every base pair, though no significant departure from the zipping model was detected for this construct, indicating that unzipping and rezipping still began distal to the loop.

The most intriguing result meriting future study implied from these trajectories, however, is the role of the closing base pair in stem-loop stability. For example, the UUCG stem-loop has a melting temperature of 60C, under high salt conditions, for a construct bearing only one base-pair in the stem: the C-G closing pair[188]. The results described here are consistent with a mechanism whereby the closing base pair contributes to a long-lived +1 terminal base pair, *i.e.* the base pair right below the closing base pair is relatively long lived on pathway to entry into H1 for 2 out of the 5 constructs (UUCG and GCAA, Figure 4.2). It is possible that the specific identity of the closing base is required for stabilization of this particular base pairing intermediate in these constructs. Indeed, for the UUCG and GCAA stem-loops, it has been shown that switching the closing base-pair from C-G to G-C lowers the apparent melting temperature by 10-15 °C[173,194,191,192].

The data presented here, in particular the sequential zipping and unzipping transition pathways describing helix unwinding, leave room for at minimum three types of paired states in

an RNA helix, depending on the number of adjacent paired bases – 2, 1, or 0, denoted Pr_2, Pr_1, and Pr_0. Because the zipping pathways we observe are sequential, our measurements are consistent with the identity of a limiting step to unpairing of a base pair as the unpairing of an adjacent base pair – i.e., that a base pair unpairs following entry into Pr_1. In the stem-loop system, this begins in one of two ways – first, from the terminal end, furthest from the loop sequence, and second, from the closing base pair of the loop. We find that the latter pathway, accounting for transitions from H4 to H1, is comparatively rare, accounting for two orders of magnitude fewer transitions than the former (Figure 4.2). These two pathways are consistent with recent high-resolution NMR measurements highlighting the role of transient intermediate base pairing configurations in RNA structures. Specifically, the rate limiting step of interconversion between the most stable and less stable configurations of these RNA structures, which involves secondary-structure remodeling, was found to occur on the 100 μs to 100 ms timescale, the same timescale I have observed here[212]. smFET measurements at higher time-resolution could potentially be used to define the rate of transitions between *syn* to *anti* conformations of the glycosidic bond of the nucleoside and therefore transitions between canonical and Hoogsteen forms of base pair of the RNA helix, estimated from the bulk NMR studies to possess lifetimes approximately an order of magnitude shorter than those probed here[213–215].

To conclude, the work presented here has implications beyond secondary-structure rearrangements. These fall into three categories – consequences for secondary-structure remodeling following tertiary rearrangements; consequences for remodeling following binding of exogenous factors; and consequences for RNA structure prediction. In the first category, this data supports a model wherein the paired state of a stem sequence is remodeled by tertiary

interactions, as has been proposed before, for example in folding of the group I intron [178]; this is explored in a different RNA system in Chapter 5, wherein I argue from measurements of the kinetics of a paired region of a riboswitch that the stability of specific base pairs in the paired region under study depend on two distinct but correlated tertiary interactions. In the second category, it has been found that binding of protein factors to conserved stem-loop motifs directly modulates the paired state of the stem; an intriguing future study using the smFET platform would therefore probe these changes in real time using the experimental design discussed in this chapter[216]. Finally, these studies imply that RNA structure prediction requires a stronger set of predictive models than those focused on secondary-structure prediction, because secondary structure can be actively kinetically altered by structural modules as simple as a four-nucleotide loop. These studies should be dynamically focused and will likely take the form of coarse grained simulations, the first step of which are beginning to be developed[169,217–219].

This chapter has described a base pair-by-base pair model for the kinetics of five stem-loops. Because of the timescales and small distances involved, such single-molecule studies are near-inaccessible by other modern means. The next chapter will describe how more complex RNA structures automatically regulate stem responses in concert with binding of small-molecules in order to the auto-assemble as switch-like structures under kinetic control.

Chapter 5 Single-molecule observation of riboswitch zipping dynamics on the microsecond timescale[7]

Long considered a mere information-carrying intermediate, RNA has now been established to possess a startling diversity of functional roles. The modern prevalence of functional RNA is attributed to (1) inevitability: RNAs must be produced to make protein anyway, so selection for functional roles, which occurs *in vitro* resulting in high affinity aptamers[220,221] and functional RNAs[222], will inevitably arise and some subset enhance fitness; (2) speed: RNAs that regulate translation, transcription, splicing, and mRNA stability[223,224] often have an advantage over protein factors as they do not have to diffusively discover their targets; (3) selectivity: RNAs assist as DNA or RNA sequence specific templates that act either in conjunction with protein factors, for example in the context of bacterial immunity, telomerase-mediated telomere extension, or RNA interference, or alone, for example in micro-RNA-mediated messenger RNA degradation. This chapter will focus on the folding of a transcription-regulating RNA switch.

Riboswitches are genetic control elements located within the 5' untranslated region (UTR) of messenger RNAs (mRNAs) that undergo metabolite-dependent structural rearrangements so as to regulate mRNA transcription, splicing, translation, or stability[223,224].

[7] With Dr. Nathan Daly. As of 2017 this chapter composes a manuscript in preparation.

While they occur in all domains of life, specific phylogenetic types of riboswitches are only found in eubacteria, marking them as targets for antibiotic drug design[225]. Additionally, as a result of their relative simplicity and selectivity, riboswitch motifs, especially aptamer motifs, have also been utilized in the field of genetic engineering[226,227]. The successful design and implementation of riboswitches and riboswitch-targeted antibiotics rely on an understanding of how these RNA molecules undergo structural rearrangements, how they recognize and selectively bind their target ligand, and how these two actions collectively operate the switch.

One of the most well-studied riboswitches is the adenine-sensing *pbuE* riboswitch found in *Bacillus subtilis*[196,228–232], a member of the broader class of purine-binding riboswitches. The *pbuE* riboswitch sequence consists of an aptamer domain that is responsible for recognition and binding of the target metabolite adenine, an expression platform domain that is responsible for forming the terminator hairpin that interacts with transcription machinery, and a switching sequence that belongs to both domains[228]. In the absence of adenine the aptamer structure gives way to formation of the more stable terminator hairpin, which arrests production of mRNA. Adenine binding, however, provides stability to the aptamer, inhibiting the formation of the terminator hairpin and allowing mRNA production and expression of the *pbuE* gene to continue. Because the *pbuE* gene encodes an adenine efflux pump, the overall cycle forms a negative feedback loop. The rate of adenine association and uptake into the aptamer *versus* the rate of transcription forms the central competition affecting the outcome of riboswitch regulation[233,234]. This kinetic balance affects the generation of the expression platform and the relative position of RNA polymerase (RNAP) to the riboswitch, both of which are thought to play a more significant role in the regulatory ability of the *pbuE* riboswitch than ligand binding thermodynamics alone[230,233,234].

Following adenine binding to the *pbuE* riboswitch, further regulation is required to communicate the decision to RNAP. This is accomplished by adenine-induced stabilization of the first five and last five nucleotides of the aptamer domain into a five base pair helix, known as the P1 stem[235]. Formation of the P1 stem sequesters base pairs required by the expression platform to form the terminator hairpin, making it a critical component of the *pbuE* riboswitch regulatory mechanism[229]. Criticality of balanced P1 stem stability has been suggestively demonstrated by *in vivo* experiments monitoring efficiency of *pbuE* riboswitch regulation of a reporter gene with increasing mismatch density on the P1 stem[236]. However, dynamics of the P1 stem, consisting entirely of base pair rearrangements, are difficult to characterize because both the lifetimes as well as the distance scales of base pairing are short[163–166]. Structures of purine family riboswitches[3,229,237], as well as NMR characterizations of solution conformations[238–240] and single-molecule fluorescence resonance energy transfer (smFRET) studies[231,241], have established that the tertiary structure of the *pbuE* riboswitch transitions to its native state following binding of adenine. Furthermore, in the bound state, more than 90% of the solvent exposed surface of adenine is surrounded by RNA, leading to the suggestion that significant structural changes in the aptamer domain are required for adenine to enter the binding pocket[229,237–239]. It is unclear, however, how stabilization of the binding pocket and formation of the tertiary structure of the aptamer domain of the riboswitch influence the stability of the P1 stem and therefore the transcription of the *pbuE* gene.

Because single-molecule techniques allow for direct observation of discrete, rare events as well as direct characterization of structural motions in a diverse ensemble, single-molecule biophysical techniques, most notably smFRET and single-molecule force spectroscopy, have been employed to investigate riboswitch dynamics. However, these techniques are hampered by

difficulties in measuring millisecond-scale dynamics, such as base pair rearrangements, as well as the inability to observe unperturbed individual molecules for long times. The recent development of a single-molecule technique using carbon nanotube-based field-effect transistors (smFETs)[18,20,44,45,48,205,242] allows for observation of the dynamics of charged biomolecules, such as charged amino acids in proteins or the phosphate groups of nucleic acid backbones, with 50 microsecond (μs) temporal resolution. Structural rearrangements of a single charged molecule on the surface of a single-walled carbon nanotube (CNT) transistor can lead to measureable fluctuations in conductance through the CNT channel. As the technique is label-free, particularly chromophore-free, smFETs exhibit stability for extended single-molecule observation so long as the molecule itself is not labile. The combined wide-bandwidth of this technique allows us to simultaneously characterize fast events and to quantify how the frequency of these events change over long time-scales.

Here we identify adenine-dependent dynamics of the P1 stem of the *pbuE* riboswitch aptamer on the microsecond timescale by correlating smFET conductance fluctuations with predicted RNA structural rearrangements and by mutagenesis of the aptamer sequence. We demonstrate how long-lived, ligand-dependent intermediates form at a base pair level and describe their consequence for riboswitch-regulation by adenine uptake into the aptamer. Using smFET technology we have achieved the first observation of RNA zipping and unzipping at the single-molecule level, as well as label-free observations of the effects of a three-way junction motif on helix zipping and unzipping.

Devices, or transistors for smFETs consisting of isolated, individual CNTs, were prepared as described (also see Chapter 2)[18,22,48]. We prepared a custom-printed circuit board capable of measuring up to 60 smFETs simultaneously in a microfluidic channel (Figure 5.1a, see

Chapter 2). We attached single riboswitch molecules by conjugating the RNA to CNT-adsorbed 1-pyrenebutyric acid *N*-hydroxysuccinimide ester, an adaptation of existing protocols[66] (see Chapter 2). Approximately one out of ten smFETs at this stage show conductance fluctuations corresponding to a signal of interest, as determined by repeating previously reported signals[18,22,48], below the yield predicted from a Poisson distribution of 37%. We speculate that this reduced yield either from reduced yield due to using RNA instead of gold nanoparticles or from the need for pyrene adsorption at CNT sites rendered sensitive to charge fluctuations by geometric or defect-driven effects, whereas most locations are insensitive, see the discussion in Chapter 3.

To observe structural rearrangements in the RNA, we measured the conductance through the smFET as a function of time. The noise spectrum of the conductance *versus* time trajectory of these devices has significant drift noise (typically the power spectral density has a power law frequency dependence $\propto \frac{1}{f^{\alpha}}, \alpha \sim 1.3$), and thus the baseline current undergoes a restricted random walk over the course of the measurement on top of random noise characteristic of typical normally distributed fluctuations. This was accounted for in our analysis using an adaptation of the algorithm presented in Bruno *et. al*[86]. We also note that every smFET has a different baseline conductance owing to variability in the structure, defect density, or electrode contact of the CNT fragment used to form each device. For these reasons, absolute smFET conductance values cannot be assigned to specific molecular conformations. Instead, relative changes in device conductance are correlated to predicted structural rearrangements of the attached molecule.

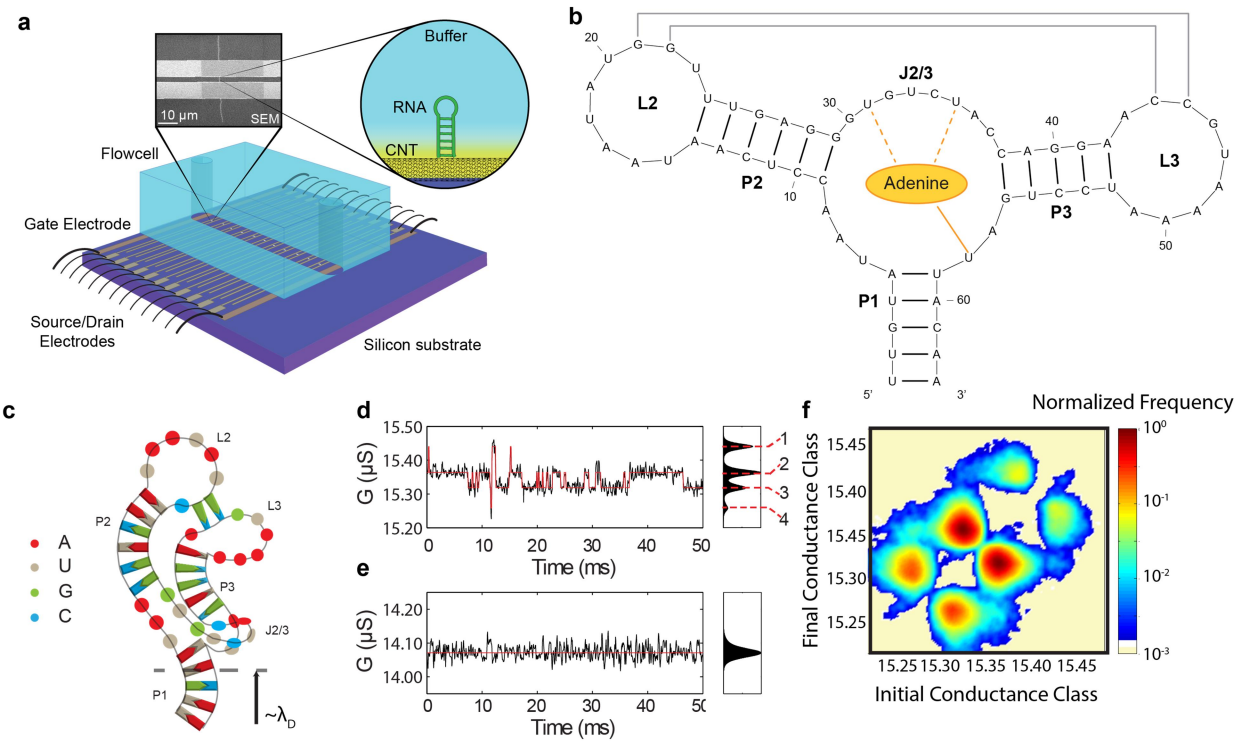


Figure 5.1 smFET experimental setup, RNA sequence design, and wild-type aptamer smFET trajectory overview.

a, smFET chip design allows for up to 60 devices to be generated from the same CNT. The inset shows a SEM image of one device and a cartoon of an RNA molecule tethered to the CNT surface. b, Secondary structure of the pbuE riboswitch aptamer. Adenine interactions in the aptamer binding pocket shown in orange. Tertiary contacts between loops L2 and L3 are shown in gray. The nucleotide at position 1 denotes the site for amine modification and subsequent tethering to smFET devices. c. Cartoon representation of the pbuE riboswitch aptamer. Dashed line shows the calculated Debye length for the experiment relative to the CNT tethering site. d. Sample trace and total population histogram of an smFET device following aptamer tethering and adenine (3 μM) exposure. The trajectory has been baseline corrected, in an adaptation of Bruno et. al[86]. e, The same device and conditions as in d, following extensive DMSO washing to remove the aptamer. f. Transition density plot of the dataset shown in d. The device current

fluctuates consecutively between four discrete states.

We began our investigation by attaching *pbuE* riboswitch aptamers to CNTFETs and performing an adenine titration. Upon addition of adenine to riboswitch-functionalized smFETs following attachment of the *pbuE* riboswitch aptamer (see Figure 5.1b and Figure 5.1c) to the surface of the nanotube *via* pyrene linker, the smFET conductance began to fluctuate between four discrete conductance classes. We argue that these conductance classes originate from rearrangements of the P1 stem of the *pbuE* riboswitch aptamer for four reasons: (1) extensive washing with dimethylsulfoxide (DMSO) caused the signal to vanish, consistent with its dependence on the pyrene linker (Figure 5.1 d, e); (2) addition of DNA complementary to the P1 stem to a distinct but similar device caused the appearance of a distinct mixture of 4- and 2-conductance class behavior (Figure C. 3); (3) though adenine is not a charged molecule at the experimental pH, the signal was adenine dependent in a way that is consistent with MFOLD calculations (Figure C. 4); (4) the calculated Debye length λ_D under our conditions predicts that four base-pairs of the P1 stem can reasonably contribute to the signal, consistent with the observation of four distinct conductance classes. With DNA smFET signals under similar conditions, the lower conductance classes were shown to be the more compact form of the nucleic acid whereby more net negative charge is localized near the nanotube surface[18,22,48]. Therefore we assigned the conductance classes sequentially from highest to lowest, 1-4, with 1 representing the most unfolded state we could detect and 4 representing the most folded state we could detect. Our argument is that each conductance class is a state with a different terminally paired base in the P1 stem, and that the changes in signal arise from increased or decreased

flexibility of the phosphate backbone of the P1 stem following unpairing or pairing of the terminal pair.

Structures of purine family riboswitches[3,229,237], as well as NMR characterizations of solution conformations[238–240] have established that the conserved base pairs between the tandem guanine nucleotides (nts) in L2 (G21 and G22) and the tandem cytosine nts in L3 (C44 and C45) form a tertiary "kissing-loop" interaction to complete the folded conformation of the riboswitch. smFRET studies of both adenine and guanine sensing riboswitches have revealed that this interaction possesses a lifetime on the order of seconds[231,241], and the folded conformation is stabilized by the presence of ligand[235,238,241]. Formation of this tertiary contact by necessity greatly restricts the available conformations of J2/3, the three-way junction element between P2 and P3, which caps the P1 stem (see Figure 5.1b and Figure 5.1c)[237,238]. Therefore, to characterize the fluctuations of the terminal base pairs of the P1 stem and assess the role of the tertiary contact on P1 stem rearrangements, we performed two distinct experiments – first, to smFETs functionalized with wild-type aptamer, we varied the concentration of adenine between 0.030 and 3.0 μM and measured conductance fluctuations arising from conformation changes in single aptamers for ~30 minutes at each concentration with 50 μs time resolution; second, to smFETs functionalized with an aptamer with a disrupted tertiary interaction, the G21C mutant aptamer, we varied the concentration of adenine between 0.050 and 50 μM and measured conductance fluctuations arising from conformation changes in single aptamers for ~30 minutes at each concentration with 50 μs time resolution. The concentration ranges of each were chosen on the basis of the experimental K_d under our conditions (See Figure C. 5.)

Intriguingly, both constructs required at least two distinct transition rate matrices to describe the fluctuations between the four conductance classes. These conductance classes

primarily transition between adjacent classes (schematized for the wild-type aptamer in Figure 5.1f). The dynamics of these classes were analyzed with the aid of a hierarchical hidden Markov model by comparing three models – one with a single rate matrix, one with two, and one with three – and noting that while the model with two transition rate matrices was much more likely and greatly changed the interpretation of the data, populating with three had no such effect because the third state was unpopulated, reducing it to the two-rate state model. We refer to the two rate matrices as arising from two distinct types of P1 stem, which we call $P1^A$ and $P1^B$ (Figure 5.2a and Figure 5.2e). To compare the wild-type and G21C aptamer $P1^A$ and $P1^B$, we prepared post-synchronized 2D histograms by cutting each conductance *vs* time trajectory into fragments that begin with occupation of the third conductance class of $P1^A$ or $P1^B$ (3-$P1^A$ and 3-$P1^B$) and terminate with occupation of- and subsequent dwell time in- 1-$P1^A$, and compared these for the wild-type and G21C aptamers (see Figure 5.2b, c, f, and g; these 2D histograms are normalized for every time point, and scaled to the most populated bin). This synchronization is chosen because in our model conductance class 3 corresponds to the second base-pair from the terminal end of the P1 stem, *i.e.,* the startpoint of a hypothetical invasion event, while conductance class 1 corresponds to the most unpaired state of the P1 stem that we can detect. We also compared the individual ΔG barriers between each conductance class in $P1^A$ and in $P1^B$ (see Supplementary information). On the basis of these two analyses, we reason that the two classes observed for each aptamer are comparable and possibly arise from equivalent rearrangements of the RNA. Although we cannot see such rearrangements with our signal, we propose that $P1^A$ and $P1^B$ are distinct because of rearrangements in J2/3 which, in crystal structures of similar purine-sensing aptamers, has been shown to interact with the conserved A-U and U-A base pairs of the P1 stem, proximal to the binding pocket (nts 5:59 and 4:60 in our numbering.)

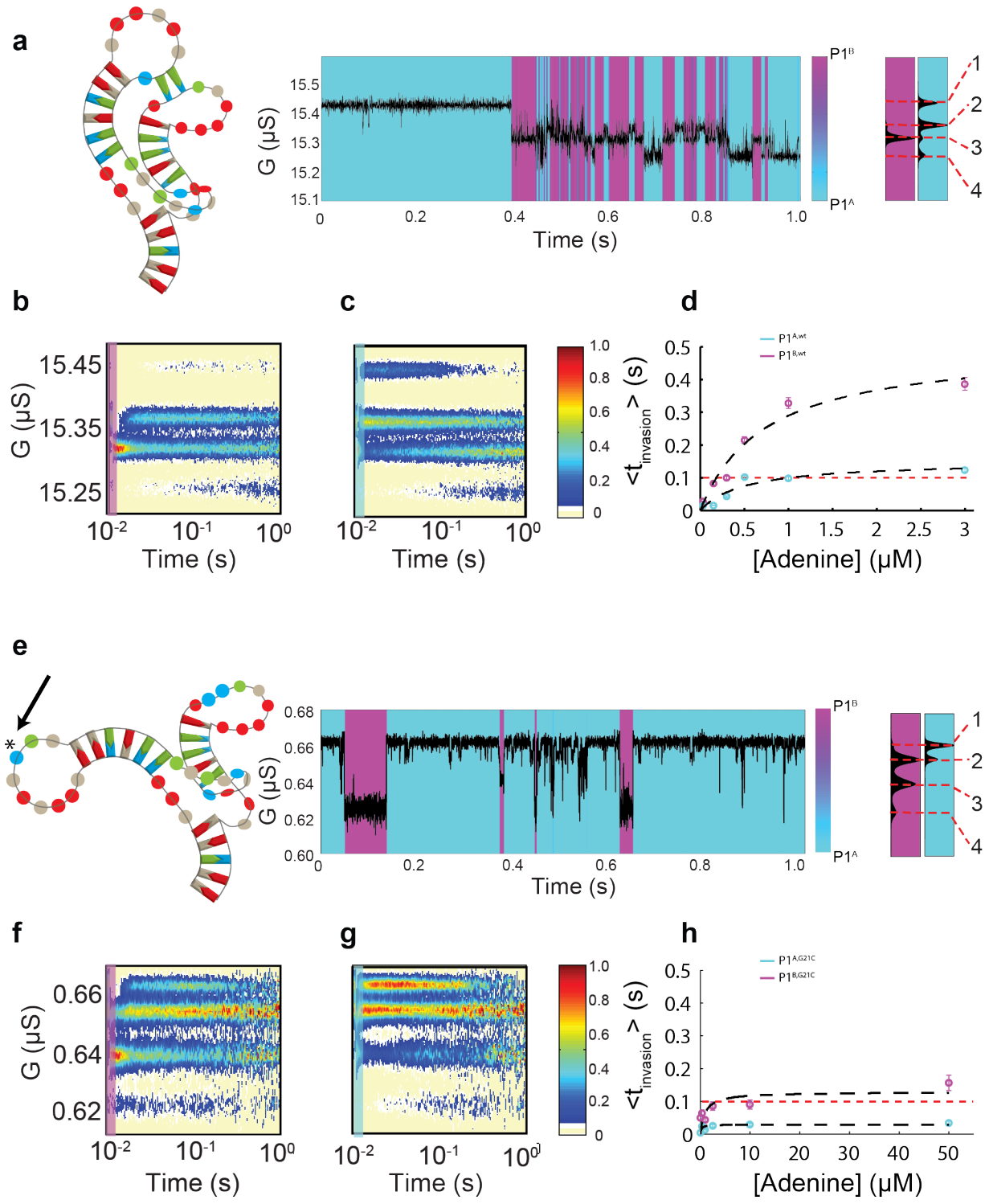


Figure 5.2 Dynamic heterogeneity of the P1 stem.

Dynamics of the **a-d,** Wild-type and **e-h**, G21C adenine-sensing *pbuE* riboswitch aptamers. **a, e,** The fluctuations possess dynamic heterogeneity consistent with two types of P1 stem, shown in cyan ($P1^A$) and magenta ($P1^B$) for each. Population evolution following **b, f,** post-synchronization into the third conductance class of $P1^B$, 3-$P1^B$, and **c, g** following post-synchronization into 3-$P1^A$ are shown as 2D histograms. Finally, for both constructs, **d, h,** the mean lifetime of events that begin in 3-$P1^B$ or 3-$P1^A$ and terminate at 1-$P1^A$, indicated as $< t_{invasion} >$, is shown as an increasing function of adenine concentration. Red dashed lines indicate the mean time for RNAP to transcribe the terminator hairpin and black dashed lines indicate a least-squares fit to a hyperbolic equation: $y = \frac{t_{max}[A]}{K_{d,app}+[A]}$.

As the mRNA is transcribed, invasion of the nascent aptamer by the expression platform is under kinetic control, and the time it takes to unwind the P1 stem, $< t_{invasion} >$, presents a first-passage chance for operation of the switch, the result of which is an intrinsically non-equilibrium process[230,233,234,243,244]. Using the trajectory fragments contributing to the post-synchronized 2D histograms in Figure 5.2b and Figure 5.2f, we recorded the lengths of each, which correspond to the time it takes to unzip the P1 stem, and used this population to calculate $< t_{invasion} >$, shown in Figure 5.2d and Figure 5.2h. As adenine was increased beyond the K_D of either construct, though G21C required roughly 100x more adenine than the wild-type for this effect to manifest, the resistance to unfolding increased, indicating that binding of adenine stabilizes the P1 stem

To uncover the kinetic details behind this stabilization, we compared the interconversions between conductance classes for $P1^{A,wt}$, $P1^{B,wt}$, $P1^{A,G21C}$, and $P1^{B,G21C}$. Compared to the 2-$P1^{B,wt}$, 2-$P1^{B,G21C}$ is relatively less stable, as with this mutation the aptamer more easily transitions

directly into conductance class 1 with a 9-fold reduced reverse rate, from 6500 ±300 s^{-1} to 700 ±200 s^{-1}. Next, we examined transitions between $P1^A$ and $P1^B$. We first note that, while the overall partition of the G21C aptamer conductance *versus* time trajectory between $P1^A$ and $P1^B$ does not change as the concentration of adenine is increased, for the wild-type aptamer the occupancy of $P1^B$ increases as the concentration of adenine is raised, from 10% at 30nM adenine to ~40% at 3 μM adenine. This is primarily driven by a 3-fold decrease in the rate of transition from $P1^B$ into $P1^A$ *via* conductance class 2, from 370 ±10 s^{-1} to 144 ±6 s^{-1} indicating that addition of adenine raises the free energy barrier between $P1^B$ and $P1^A$. As discussed below, this latter rate is now commensurate with the rate of transcription[245–247]. Together, these kinetic results imply that the L2-L3 tertiary interaction has two major effects: first, it prevents $P1^B$ from entering the predominantly unpaired state (conductance class 1), and second, it prevents $P1^B$ from entering $P1^A$ any faster than the transcription rate. These observations suggest that the role of the tertiary contact is to populate $P1^B$, with the help of adenine, by preventing transitions to $P1^A$ (schematized in Figure 5.3). Because this effect is synergistic with the presence of adenine, we propose that the tertiary contact mainly acts to organize the binding pocket elements of J2/3 that interact with the P1 stem.

Therefore, we find that the P1 stem of the *pbuE* riboswitch aptamer is highly dynamic even in the presence of adenine, a result that is loosely consistent with Nozinivic *et al*[235]. However, we suspect, as others have, that this highly dynamic kinetic pattern characterizes all nucleic acid structures[95,96,199]. According to NMR results, the lifetime of an paired base, while highly sequence dependent, is on the order of 1-100 ms[163,165,166]. So-called toeholding experiments which monitor the invasion of one strand of DNA by another typically predict a similar lifetime[168,169]. Furthermore, from studies of the zipping of hairpins, the rate at which

bases transition from the paired to the unpaired once that motion has begun, and *vice versa*, is on the order of hundreds of nanoseconds[95,96,167,197], in loose agreement with our observation of instantaneous transitions even at 50 μs time resolution. These two features of pairing and unpairing dynamics lead to a highly complex ensemble of possible and constantly interconverting configurations – a hairpin consisting of just 10 base pairs has over 1500 unique configurations[96]. It is notable, therefore, that under our conditions, the P1 stem for the most part only accesses two pathways for discrete and sequential zipping and unzipping, $P1^{A}$ and $P1^{B}$.

In the case of the wild-type aptamer, our data suggest two ways in which the RNA modulates invasion by the expression platform. First, enhanced occupancy of $P1^{B}$ in its stabilized intermediate, 3-$P1^{B}$, indirectly increases the occupancy of the fully folded P1 stem, 4-$P1^{A}$ or 4-$P1^{B}$ (Figure 5.3). We assert that in this configuration the expression platform cannot invade the P1 stem because it cannot interact with any bases in the helix. Second, assuming an invasion has begun at the first base pair of the P1 stem, while the aptamer will unfold almost immediately if it is in 3-$P1^{A}$ (Figure 5.2 c, g), our model suggests that if the aptamer is in 3-$P1^{B}$ (Figure 5.2 b, f) it will unfold up until the third base pair from the 5' end, *i.e.* the G-C base pair. If adenine concentrations are low, transitions between $P1^{B}$ and $P1^{A}$ are significantly faster, ~370 s^{-1}, than the rate of transcription of the antiterminator, ~10-90 nt/s[245–247] for *E. coli* and perhaps faster for *B. subtilis*[248], though ~10 nt, or an RNAP footprint[190,249] must be transcribed for RNAP to escape, an argument introduced in this context by Wickiser *et al*[230]. This lifetime is indicated as a dotted line in Fig. 2d and 2h. As the reverse rate back into $P1^{B}$ is slow, ~25 s^{-1}, the aptamer enters $P1^{A}$ and continues to unfold. However, if adenine concentrations are high, transitions between $P1^{B}$ and $P1^{A}$, ~140 s^{-1}, are competitive with the rate of transcription, and the refolding rate is very high – the aptamer resists invasion.

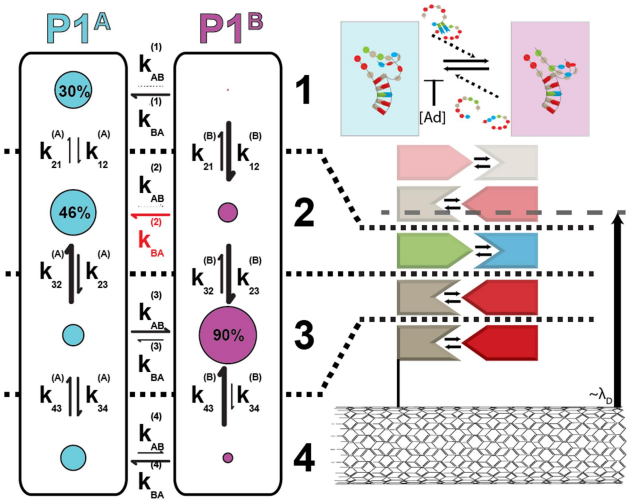


Figure 5.3 Kinetic model for base pair-level fluctuations of the wild-type P1 stem in the presence of adenine.
Specific rate constants can be found in Supplementary table S2-4. In red is shown the adenine dependent rate of transition from $P1^B$ into $P1^A$. These fluctuations are affected by formation of tertiary contacts as well as by adenine concentration.

Intriguingly, the second mechanism implies a central role for the third G-C base pair. Previous studies of the wild-type *pbuE* riboswitch sequence have suggested a crucial role for the central G-C base pair of the P1 stem[196,230,232]. In force pulling studies by Greenleaf *et al.* and Frieda *et al.*, an observed transition state for P1 stem unfolding was located at a pulling distance corresponding to the G-C base pair, and obliterating that pair results in a loss of riboswitch activity[196,232,236]. To assess the role of the central G-C base pair in P1 stem stability, we performed smFET studies on *pbuE* aptamer mutants with modifications in the P1 stem, beginning with a P1-destabilized mutant referred to as the G3C aptamer. Between 0.030 and 3.0 μM adenine, of the G3C aptamer revealed fluctuations between two conductance classes whose rates of transition were strongly adenine dependent (Figure 5.4a, b, c). The rates of transition were heterogeneous analogous to the wild-type or G21C dynamics; addition of adenine primarily reduced the heterogeneity in favor of fast dynamics between the two conductance class, as summarized in Fig. 4c. Overall these effects are consistent with a dynamic model (Figure 5.4d), similar to that proposed for the wild-type P1 stem (Figure 5.3).

For the G3C aptamer, these results are consistent with an assignment of conductance class 2 being a P1 stem conformer with the A1 and U63 paired, which is relatively stabilized by adenine binding, and conductance class 1 being a state with a predominantly unpaired P1 stem. This is because binding pocket rearrangements are in close proximity to the predominantly unpaired state of the helix and are likely to have a much more pronounced effect, rearranging the phosphate backbone to a state close to the paired state when adenine is in the bound state. On the other hand the fact that the fully paired state is weakly destabilized as adenine is added is consistent with the interpretation that the C-C mismatch tends to destabilize the helical conformation, in addition to abolishing the intermediate stacked states observed in the wild-type aptamer. Importantly, comparison of the adenine dependence of the G3C aptamer in our smFET assay with its adenine-analog (2AP) dependence as measured in a bulk fluorescence ligand binding assay (Fig. S5) suggest that the binding pocket plays a direct role in stabilizing the paired conformation of the P1 stem, but under our conditions, disruption of P1 stem rearrangements has little to no effect on the binding pocket.

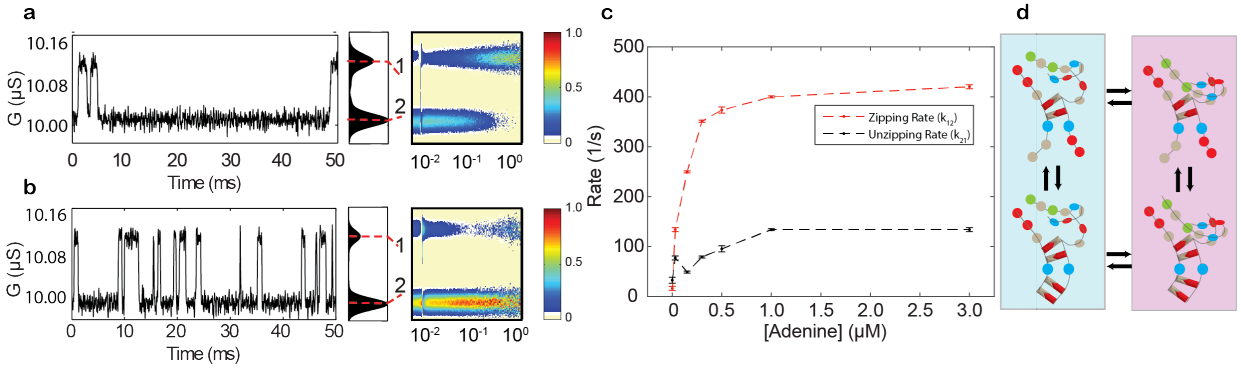


Figure 5.4 Dynamics of the G3C aptamer.
a, Dynamics in the absence of ligand possesses two conductance classes whose transitions rates are heterogeneous with a wide mixture of timescales – the aptamer occasionally spends entire

seconds in either the high or low conductance class. 2D histograms are post-synchronized to transitions from conductance class 1 to class 2 and are normalized identically to those in Figure 2. **b,** In the presence of 3 μM adenine, two conductance classes remain but there is drastically reduced heterogeneity. **c,** As the adenine concentration is varied, this effect leads to a steady increase in the overall aggregated rate constant. **d.** The data are consistent with a two rate-state model, analogous to that observed in the wild-type and G21C aptamers.

A second aptamer with a P1 stem stabilized without altering the relative distance between the CNT surface and the *pbuE* aptamer structure by exchanging the two terminal A-U base pairs for G-C base pairs, referred to as the stable aptamer, was prepared. We compared the bulk binding of the analogue 2-aminopurine (2AP) of the four aptamers – wild-type, G21C, G3C, and stable. 2AP had similar binding affinity for the wild-type, G3C, and stable aptamers, and poor binding affinity for the G21C aptamer, as described previously (Figure C. 5)[230,231]. During the course of our measurements of the stable aptamer, we were unable to find a signal possessing a strong adenine dependence. However, a signal with two conductance classes whose fluctuations were on the millisecond timescale was repetitively observed, leading us to tentatively speculate that this signal corresponded to fluctuations of the P1 stem of the stable aptamer (see Supplementary Figure C.10). Although the signal was readily washed from the smFET, it is difficult to speculate further as to its significance.

By way of conclusion, our data suggest the existence of a dynamic correlation between L2-L3 kissing loop formation[231,241] and the base pair level dynamics of the P1 stem, over a distance of 32 Angstroms, which orient and dominate the fluctuations of the G-C base pair in the third position. This picture is supported by the observation that, in crystal structures of purine riboswitches[3,229,237], the junction element J2/3 caps the P1 stem, and further, that NMR studies

have revealed that this junction element is disordered in the absence of adenine[238–240]. Together with these observations, our data suggest that rearrangements of the ligand-enclosing flap, which are strongly influenced by the tertiary interaction between L2 and L3, reduce the probability of transition from $P1^B$ to $P1^A$, and $P1^B$ is inefficiently unzipped, holding the P1 stem stable long enough for an RNAP to escape the vicinity of the riboswitch before the terminator hairpin forms. This conclusion on the basis of our conductance *versus* time trajectories is fully consistent with recent theoretical work on the closely-related guanine-sensing *gsw* riboswitch suggesting an allosteric collaboration between the P1 stem, L2-L3 tertiary interaction, and J2/3 whose overall function is to stabilize the P1 stem in the presence of cognate ligand or leave the P1 stem invasible by the expression platform in the absence of cognate ligand[250].

To augment our understanding of this dynamic correlation between tertiary structure formation and base-pair level zipping and unzipping of the P1 stem, we compare our results to the pattern of conservation of purine riboswitches. In general, the P2 and P3 stems are conserved in the sense that they remain paired regions of a certain length, but their precise sequences are variable (RF00167)[251]. L2 and L3 have a conserved length as well as conserved G-C kissing-loop pair. The binding pocket element J2/3 is almost universally conserved, as are the two A-U base pairs of the P1 stem closest to the pocket. The third base pair of the P1 stem, which is a G that forms the central G-C base pair in the case of the *pbuE* aptamer, (Figure 5.1b) is not strictly conserved: it varies between an adenine (51%) and a guanine (49%). Examining these results in more detail (see Figure C. 9), we notice that if the P1 stem contains an A instead of a G in the third (G-C) position, it also contains an AUG on the 3' end of the P1 stem sequence, the U of which base pairs with an A in the third (Figure 5.1b) position on the 5' end of the P1 stem sequence. Furthermore in such cases the P1 stem always contains an additional four bases (two

A-U pairs and two G-C pairs). On the basis of our results we propose that the purine riboswitch family has at least two classes of P1 stems – those that begin zipping and unzipping at the AU-UA pair in position 5 (Fig. 5.1b), and those that begin zipping and unzipping at the AUG three base pairs down but contain four extra base pairs in the P1 stem. While the *pbuE* riboswitch falls into the former class, the guanine-sensing *xpt* riboswitch is an example of the latter class. smFRET and single-molecule force pulling studies have revealed that the *pbuE* riboswitch does not fully fold in the absence of adenine[196,231], whereas NMR studies of the *xpt* riboswitch have revealed that the riboswitch is close to its bound state even in the absence of its ligand[239,240]. We propose that these two families have differing designs of P1 stem in order to transfer metastability from the tertiary kissing-loop formation of L2-L3 to the secondary structure of the P1 stem. Furthermore, from comparison of the heterogeneity in wild-type sequences with sequences that cannot form stable L2-L3 interactions, we conjecture that the L2-L3 interaction exists on one side of a dynamic network whose consequence is zipping of the P1 stem.

Our methods utilizing single-molecule field effect transistors contrast from those employed in previous studies, as we are able to provide a label-free high time-resolution single-molecule measurement of the *pbuE* riboswitch aptamer using its natural sequence without continuous perturbation. Previous studies have provided crucial information as to the organization of the relative structural elements of the switch, as well as the order in which they fold and the likely slow-timescale rearrangements that accompany ligand binding. In this work we present the first wide-bandwidth single-molecule measurements of base-pair level stability of the P1 stem. We find that zipping and unzipping of the stem is sequential and dependent on the presence of specific base pairs in the organization of the stem. Finally, we find that these rearrangements are heterogeneous on the second timescale yet occur on the microsecond to

millisecond timescale, making measurements from smFETs essential in order to describe the potential ensemble of unzipping pathways.

Riboswitches other than the purine riboswitch family discussed above also utilize kinetic competition between distinct RNA structures. For example, the $PreQ_1$ riboswitch exists as a stem-loop with a highly structured tail[252] which possesses a ribosome binding site that becomes occluded in the ligand-dependent fully folded form of the RNA[253]. In this RNA, the rate limiting steps of RNA folding do not involve formation and disruption of base pairs, but instead rely on the formation and disruption of numerous A-minor interactions with the ligand-bound stem-loop; the platform described here could be modified to determine the rate of each step of pseudoknot formation of each A-minor interaction with the stem, because the mechanism of this switch is kinetically homologous to operation of the switch discussed in this chapter.

Next, note that the mechanism described in Figure 5.3 possesses many rates that are on the millisecond to second scale. These rates depend on tertiary structure interactions. Therefore, we argue that computational simulations of RNA switch behavior will need to encompass this timescale, and thus the development of highly efficient and spatially descriptive coarse grained models of RNA dynamics will be required for their *in silico* description[169,217–219].

Finally, this study highlights the role of a three-way junction element in stability of its connected paired regions, but there are many examples of such elements[254]. In particular, these fall into three types, depending on the order in which the strand pairs with itself, and whether additional strands are required. Such junctions are ubiquitous in nature: for example, the hinge region of the L1 stalk of the large subunit of the ribosome contains a three-way junction whose topology is similar to the riboswitch described in this chapter[149,255]. Future smFET studies of three-way junction elements could include the a detailed study of the allosteric influence of a

three way junction on the kinetics of distant base pairs, indicating how this influence on distant base pairing states allow these three-way junctions to operate as hinges, as well as in general the kinetics of base pairing states of each of the three-way junction families.

In further future studies, smFETs could be used to probe the magnesium (Mg^{2+}) dependence of folding, as well as to study rearrangements of each portion of the riboswitch relative to every other. For instance, *xpt* riboswitches include a rearrangement between the P1 and P2 stems that cannot be detected using the 5' terminal tethering presented here. However, the rate constants of this process have a very strong dependence on the concentration of Mg^{2+} ions, which quickly exit the observable time-resolution of most single-molecule methods. We expect that the smFET experimental platform can be applied to such situations, as well as, in general, systems with local charge fluctuations that have a wide mixture of time-scales.

References

1. Gesteland, R. F., Cech, T. R. & Atkins, J. F. *The RNA World*. (Cold Spring Harbor Laboratory Press, 2005).
2. Mustoe, A. M., Brooks, C. L. & Al-Hashimi, H. M. Hierarchy of RNA Functional Dynamics. *Annu. Rev. Biochem.* **83,** 441–466 (2014).
3. Delfosse, V. *et al.* Riboswitch structure: An internal residue mimicking the purine ligand. *Nucleic Acids Res.* **38,** 2057–2068 (2009).
4. Rozov, A., Westhof, E., Yusupov, M. & Yusupova, G. The ribosome prohibits the G???U wobble geometry at the first position of the codon-anticodon helix. *Nucleic Acids Res.* **44,** 6434–6441 (2016).
5. Ogle, J. M. *et al.* Recognition of Cognate Transfer RNA by the 30 S Ribosomal Subunit. **897,** 897–903 (2013).
6. Tinoco, I. & Gonzalez, R. L. Biological mechanisms, one molecule at a time. *Genes Dev.* **25,** 1205–1231 (2011).
7. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. (Springer, 2006).
8. Iqbal, A. *et al.* Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 11176–81 (2008).
9. Berkovich, R., Garcia-Manyes, S., Klafter, J., Urbakh, M. & Fern??ndez, J. M. Hopping around an entropic barrier created by force. *Biochem. Biophys. Res. Commun.* **403,** 133–137 (2010).
10. Cossio, P., Hummer, G. & Szabo, A. On artifacts in single-molecule force spectroscopy. *Proc. Natl. Acad. Sci.* **112,** 201519633 (2015).

11. Dudko, O. K., Graham, T. G. W. & Best, R. B. Locating the barrier for folding of single molecules under an external force. *Phys. Rev. Lett.* **107,** 3–6 (2011).

12. Wen, J.-D. *et al.* Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophys. J.* **92,** 2996–3009 (2007).

13. Rosenstein, J. K., Lemay, S. G. & Shepard, K. L. Single-molecule bioelectronics. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **7,** 475–493 (2015).

14. Kinz-Thompson, C. D. & Gonzalez, R. L. smFRET studies of the 'encounter' complexes and subsequent intermediate states that regulate the selectivity of ligand binding. *FEBS Lett.* **588,** 3526–3538 (2014).

15. Kinz-Thompson, C. D., Bailey, N. A. & Gonzalez, Jr., R. L. Precisely and Accurately Inferring Single-Molecule Rate Constants. *Methods Enzymol.* **581,** In Press (2016).

16. Kapanidis, A. N. & Strick, T. Biology, one molecule at a time. *Trends Biochem. Sci.* **34,** 234–243 (2009).

17. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5,** 507–516 (2008).

18. Sorgenfrei, S., Chiu, C., Johnston, M., Nuckolls, C. & Shepard, K. L. Debye Screening in Single-Molecule Carbon Nanotube Field-Effect Sensors. 3739–3743 (2011).

19. Sorgenfrei, S. *et al.* Label-free single-molecule detection of DNA-hybridization kinetics with a carbon nanotube field-effect transistor. *Nat. Nanotechnol.* **6,** 126–132 (2011).

20. Choi, Y. *et al.* Single-Molecule Lysozyme Dynamics Monitored by an Electronic Circuit. *Science (80-.).* **335,** 319–324 (2012).

21. Bowman, G. R., Pande, V. S. & Noé, F. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. **797,** (2014).

22. Sorgenfrei, S. *et al.* Label-free single-molecule detection of DNA-hybridization kinetics with a carbon nanotube field-effect transistor. *Nat. Nanotechnol.* **6,** 126–132 (2011).

23. Datta, S. *Electronic Transport in Mesoscopic Systems*. *Electronic Transport in Mesoscopic Systems* **3,** (1997).

24. Guo, J., Datta, S., Lundstrom, M. & Anantam, M. P. Toward Multiscale Modeling of Carbon Nanotube Transistors. *Int. J. Multiscale Comput. Eng.* **2,** 257 (2004).

25. Dresselhaus, M. S., Dresselhaus, G. & Saito, R. Physics of carbon nanotubes. *Carbon N. Y.* **33,** 883–891 (1995).

26. Kim, P., Odom, T. W., Huang, J.-L. & Lieber, C. M. Electronic Density of States of Atomically Resolved Single-Walled Carbon Nanotubes: Van Hove Singularities and End States. 4 (1998). doi:10.1103/PhysRevLett.82.1225

27. Wilder, J. W. G., Venema, L. C., Rinzler, A. G., Smalley, R. E. & Dekker, C. Electronicstructureof atomically resolved carbonnanotubes. *Nature* **391,** 59–62 (1998).

28. Liu, F. *et al.* Giant random telegraph signals in the carbon nanotubes as a single defect probe. *Appl. Phys. Lett.* **86,** 1–3 (2005).

29. An, Y., Rao, H., Bosman, G. & Ural, A. Random telegraph signal and 1/f noise in forward-biased single-walled carbon nanotube film-silicon Schottky junctions. *Appl. Phys. Lett.* **100,** 1–5 (2012).

30. Liu, F. & Wang, K. L. Correlated Random Telegraph Signal and Low-Frequency Noise in Carbon Nanotube Transistors. *Nano Lett.* (2007).

31. Liu, F., Bao, M., Wang, K. L., Zhang, D. & Zhou, C. Coulomb attractive random telegraph signal in a single-walled carbon nanotube. *Phys. Rev. B - Condens. Matter Mater. Phys.* **74,** 6–10 (2006).

32. Wang, N. P., Heinze, S. & Tersoff, J. Random-telegraph-signal noise and device variability in ballistic nanotube transistors. *Nano Lett.* **7,** 910–913 (2007).

33. Warren, S. B., Vernick, S., Romano, E. & Shepard, K. L. Complementary Metal-Oxide-Semiconductor Integrated Carbon Nanotube Arrays: Toward Wide-Bandwidth Single-Molecule Sensing Systems. *Nano Lett.* **16,** 2674–2678 (2016).

34. Peng, H. B., Hughes, M. E. & Golovchenko, J. a. Room-temperature single charge sensitivity in carbon nanotube field-effect transistors. *Appl. Phys. Lett.* **89,** 1–3 (2006).

35. Son, Y. W., Ihm, J., Cohen, M. L., Louie, S. G. & Choi, H. J. Electrical switching in metallic carbon nanotubes. *Phys. Rev. Lett.* **95,** 1–4 (2005).

36. Postma, H. W. C. Carbon Nanotube Single-Electron Transistors at Room Temperature. *Science (80-.).* **293,** 76–79 (2001).

37. Wang, N.-P. & Xu, X.-J. Effects of defects near source or drain contacts of carbon nanotube transistors. *EPL (Europhysics Lett.* **100,** 47009 (2012).

38. Tans, S. J. S. & Dekker, C. Potential modulations along carbon nanotubes. *Nature* **404,** 834–5 (2000).

39. Prisbrey, L., Roundy, D., Blank, K., Fifield, L. S. & Minot, E. D. Electrical characteristics of carbon nanotube devices prepared with single oxidative point defects. *J. Phys. Chem. C* **116,** 1961–1965 (2012).

40. Mannik, J., Goldsmith, B. R., Kane, A. & Collins, P. G. Chemically induced conductance switching in carbon nanotube circuits. *Phys. Rev. Lett.* **97,** 1–4 (2006).

41. Ashraf, M. K., Bruque, N. A., Pandey, R. R., Collins, P. G. & Lake, R. K. Effect of localized oxygen functionalization on the conductance of metallic carbon nanotubes. *Phys. Rev. B - Condens. Matter Mater. Phys.* **79,** 1–11 (2009).

42. Hunt, S. R., Wan, D., Khalap, V. R., Corso, B. L. & Collins, P. G. Scanning Gate Spectroscopy and Its Application to Carbon Nanotube Defects. 1055–1060 (2011).

43. Goldsmith, B. R., Coroneus, J. G., Kane, a a, Weiss, G. a & Collins, P. G. Monitoring single molecule reactivity on a carbon nanotube. *Nano Lett.* **8,** 189–194 (2008).

44. Choi, Y. *et al.* Dissecting single-molecule signal transduction in carbon nanotube circuits with protein engineering. *Nano Lett.* **13,** 625–631 (2013).

45. Olsen, T. J. *et al.* Electronic measurements of single-molecule processing by DNA polymerase i (Klenow fragment). *J. Am. Chem. Soc.* **135,** 7855–7860 (2013).

46. Lundstrom, M. S. & Guo, J. *Nanoscale transistors: Device physics, modeling and simulation*. *Nanoscale Transistors: Device Physics, Modeling and Simulation* (2006). doi:10.1007/0-387-28003-0

47. Stern, E. *et al.* Importance of the debye screening length on nanowire field effect transistor sensors. *Nano Lett.* **7,** 3405–3409 (2007).

48. Bouilly, D. *et al.* Single-Molecule Reaction Chemistry in Patterned Nanowells. *Nano Lett.* **16,** 6–12 (2016).

49. Li, Y. *et al.* Growth of single-walled carbon nanotubes from discrete catalytic nanoparticles of various sizes. *J. Phys. Chem. B* **105,** 11424–11431 (2001).

50. Chen, Z., Appenzeller, J., Knoch, J., Lin, Y. & Avouris, P. The Role of Metal / Nanotube Contact in the Performance of Carbon Nanotube Field-Effect Transistors. *Nano Lett.* **5,** 1497–1502 (2005).

51. Niyogi, S. *et al.* Chemistry of single-walled carbon nanotubes. *Acc. Chem. Res.* **35,** 1105–1113 (2002).

52. Bard, A. & Faulkner, L. R. *Electrochemical Methods: Fundamentals and Applications*.

(Wiley, 2001).

53. Tagmatarchis, N. & Prato, M. Functionalization of carbon nanotubes via 1 , 3-dipolar cycloadditions. *J. Mater. Chem.* **14,** 437–439 (2004).

54. Devadoss, A. & Chidsey, C. E. D. Azide-modified graphitic surfaces for covalent attachment of alkyne-terminated molecules by 'click' chemistry. *J. Am. Chem. Soc.* **129,** 5370–5371 (2007).

55. Barrière, F. & Downard, A. J. Covalent modification of graphitic carbon substrates by non-electrochemical methods. *J. Solid State Electrochem.* **12,** 1231–1244 (2008).

56. Zhang, J. *et al.* Effect of Chemical Oxidation on the Structure of Single-Walled Carbon Nanotubes. *J. Phys. Chem. B* **107,** 3712–3718 (2003).

57. Schmidt, G., Gallon, S., Esnouf, S., Bourgoin, J. P. & Chenevier, P. Mechanism of the coupling of diazonium to single-walled carbon nanotubes and its consequences. *Chem. - A Eur. J.* **15,** 2101–2110 (2009).

58. Nair, N., Kim, W. J., Usrey, M. L. & Strano, M. S. A structure-reactivity relationship for single walled carbon nanotubes reacting with 4-hydroxybenzene diazonium salt. *J. Am. Chem. Soc.* **129,** 3946–3954 (2007).

59. Bahr, J. L. *et al.* Functionalization of carbon nanotubes by electrochemical reduction of aryl diazonium salts: A bucky paper electrode. *J. Am. Chem. Soc.* **123,** 6536–6542 (2001).

60. Bouilly, D., Janssen, J. L., Cabana, J., Côté, M. & Martel, R. Graft-induced midgap states in functionalized carbon nanotubes. *ACS Nano* **9,** 2626–2634 (2015).

61. Wilson, H. *et al.* Electrical monitoring of sp3 defect formation in individual carbon nanotubes. *J. Phys. Chem. C* **120,** 1971–1976 (2016).

62. Todorovic, P. *An Introduction to Stochastic Processes and Their Applications*. *Learning*

26, (1992).

63. Brown, S., Jorio, a., Dresselhaus, M. & Dresselhaus, G. Observations of the D-band feature in the Raman spectra of carbon nanotubes. *Phys. Rev. B* **64,** 3–6 (2001).

64. Dresselhaus, M. S., Dresselhaus, G., Saito, R. & Jorio, A. Raman spectroscopy of carbon nanotubes. *Phys. Rep.* **409,** 47 (2005).

65. Gavrilyuk, J., Ban, H., Nagano, M., Hakamata, W. & Barbas, C. F. Formylbenzene Diazonium Hexa fl uorophosphate Reagent for Tyrosine-Selective Modi fi cation of Proteins and the Introduction of a Bioorthogonal Aldehyde. *Bioconjug. Chem.* **23,** 2321–2328 (2012).

66. Chen, R. J., Zhang, Y., Wang, D. & Dai, H. Noncovalent Sidewall Functionalization of Carbon Nanotubes for Protein Immobilization. *J. Am. Chem. Soc.* **123,** 3838–3839 (2001).

67. Mceuen, P. L., Fuhrer, M. & Park, H. Single-Walled Carbon Nanotube Electronics. *Nanotechnology, IEEE Trans.* **1,** 78–85 (2002).

68. Wilson, N. R. & Cobden, D. H. Tip-modulation scanned gate microscopy. *Nano Lett.* **8,** 2161–2165 (2008).

69. Sah, C. T. *Fundamentals of Solid-State Electronics*. (World Scientific Publishing, 1991).

70. Kim, Y., Oh, Y. M., Park, J.-Y. & Kahng, S.-J. Mapping potential landscapes of semiconducting carbon nanotubes with scanning gate microscopy. *Nanotechnology* **18,** 475712 (2007).

71. Odom, T. W., Huang, J. L., Kim, P. & Lieber, C. M. Structure and electronic properties of carbon nanotubes. *J. Phys. Chem. B* **104,** 2794–2809 (2000).

72. Dresselhaus, M. S., Jorio, A., Souza Filho, A. G., Saito, R. & Saito, A. R. Defect characterization in graphene and carbon nanotubes using Raman spectroscopy. *Philos.*

Trans. R. Soc. A Math. Phys. Eng. Sci. **368,** 5355–5377 (2010).

73. Börrnert, F. *et al. In situ* observations of self-repairing single-walled carbon nanotubes. *Phys. Rev. B* **81,** 201401 (2010).

74. Dehmelt, H. A Single Atomic Particle Forever Floating at Rest in Free Space: New Value for Electron Radius. *Phys. Scr.* **T22,** 102–110 (1987).

75. Frank, J. & Gonzalez, R. L. Structure and Dynamics of a Processive Brownian Motor: The Translating Ribosome. *Annu. Rev. Biochem.* **79,** 381–412 (2010).

76. Bishop, C. M. *Pattern Recognition and Machine Learning*. *Springer* **1,** (2006).

77. McKinney, S. A., Joo, C. & Ha, T. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophys. J.* **91,** 1941–1951 (2006).

78. Qin, F. & Li, L. Model-Based Fitting of Single-Channel Dwell-Time Distributions. *Biophys. J.* **87,** 1657–1671 (2004).

79. Qin, F., Auerbach, a & Sachs, F. Hidden Markov modeling for single channel kinetics with filtering and correlated noise. *Biophys. J.* **79,** 1928–1944 (2000).

80. Qin, F., Auerbach, a & Sachs, F. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys. J.* **79,** 1915–1927 (2000).

81. Juette, M. F. *et al.* Single-Molecule Imaging of Non-Equilibrium Molecular Ensembles on the Millisecond Timescale. *Nat. Methods* **13,** 1–7 (2016).

82. Preus, S., Noer, S. L., Hildebrandt, L. L., Gudnason, D. & Birkedal, V. iSMS: single-molecule FRET microscopy software. *Nat. Methods* **12,** 593–594 (2015).

83. Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L. & Wiggins, C. H. Learning rates and states from biophysical time series: A Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* **97,** 3196–3205 (2009).

84. van de Meent, J.-W., Bronson, J. E., Wood, F., Gonzalez, R. L. & Wiggins, C. H. Hierarchically-coupled hidden Markov models for learning kinetic rates from single-molecule data. *J. Mach. Learn. Res. Work. Conf. Proc.* **28,** 361–369 (2013).

85. 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,** 1327–1337 (2014).

86. Bruno, W. J., Ullah, G., Daniel Mak, D. O. & Pearson, J. E. Automated maximum likelihood separation of signal from baseline in noisy quantal data. *Biophys. J.* **105,** 68–79 (2013).

87. Langer, J. S. Statistical theory of the decay of metastable states. *Ann. Phys. (N. Y).* **54,** 258–275 (1969).

88. Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*. *North-Holland Personal Library* **2,** (1981).

89. Lindblad, G. On the generators of quantum dynamical semigroups. *Commun. Math. Phys.* **48,** 119–130 (1976).

90. Kenkre, V. M., Montroll, E. W. & Shlesinger, M. F. Generalized master equations for continuous-time random walks. *J. Stat. Phys.* **9,** 45–50 (1973).

91. Brémaud, P. *Markov Chains*. (1999). doi:10.1007/978-1-4757-3124-8

92. Bishop, C. M. & Winn, J. Structured Variational Distributions in VIBES.

93. Liu, Y.-Y., Li, S., Li, F., Song, L. & Rehg, J. M. Efficient Learning of Continuous-Time Hidden Markov Models for Disease Progression. *Adv. Neural Inf. Process. Syst.* 3599–3607 (2015).

94. Yueh, W. Eigenvalues of Several Tridiagonal Matrices. **5,** 66–74 (2005).

95. Kuznetsov, S. V. & Ansari, A. A kinetic zipper model with intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophys. J.* **102,** 101–111 (2012).

96. Zhang, W. & Chen, S.-J. RNA hairpin-folding kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1931–6 (2002).

97. Rezácová, P., Borek, D., Moy, S. F., Joachimiak, A. & Otwinowski, Z. The helix-coil transition revisited. *Proteins* **70,** 311–319 (2008).

98. Usmani, R. A. Inversion of a Tridiagonal Jacobi Matrix. *Linear Algebra Appl.* **10010,** 413–414 (1994).

99. Winn, J. M. & Bishop, C. M. Variational message passing. **6,** 661–694 (2005).

100. Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. Variational Inference: A Review for Statisticians. *arXiv* 1–33 (2016). doi:10.1080/01621459.2017.1285773

101. Gutiérrez-Peña, E. & Muliere, P. Conjugate Priors Represent Strong Pre-Experimental Assumptions. *Scand. J. Stat.* (2004). doi:10.1111/j.1467-9469.2004.02-019.x

102. Pitman, E. J. G. Sufficient statistics and intrinsic accuracy. *Math. Proc. Cambridge Philos. Soc.* **32,** 567–579 (1936).

103. Koopman, B. O. Urn ,. **222,** 399–409 (1935).

104. Ball, N. M. & Brunner, R. J. Data Mining and Machine Learning in Astronomy. *Int. J. Mod. Phys. D* **19,** 1049–1106 (2010).

105. Sage, D. *et al.* Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat. Methods* **12,** 1–12 (2015).

106. Rolfe, D. J. *et al.* Automated multidimensional single molecule fluorescence microscopy feature detection and tracking. *Eur. Biophys. J.* **40,** 1167–1186 (2011).

107. Sivia, D. S. & Skilling, J. *Data Analysis: A Bayesian Tutorial*. (Oxford University Press,

2006).

108. Bronson, J. E., Hofman, J. M., Fei, J., Gonzalez, R. L. & Wiggins, C. H. Graphical models for inferring single molecule dynamics. *BMC Bioinformatics* **11,** S2 (2010).

109. Taylor, J. N., Makarov, D. E. & Landes, C. F. Denoising single-molecule FRET trajectories with wavelets and Bayesian inference. *Biophys. J.* **98,** 164–173 (2010).

110. Persson, F., Lindén, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat. Methods* **10,** 265–9 (2013).

111. Cox, S. *et al.* Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nat. Methods* **9,** 195–200 (2012).

112. Okamoto, K. & Sako, Y. Variational Bayes analysis of a photon-based hidden Markov model for single-molecule FRET trajectories. *Biophys. J.* **103,** 1315–24 (2012).

113. Yoon, J. W., Bruckbauer, A., Fitzgerald, W. J. & Klenerman, D. Bayesian inference for improved single molecule fluorescence tracking. *Biophys. J.* **94,** 4932–47 (2008).

114. Rubin-Delanchy, P. *et al.* Bayesian cluster identification in single-molecule localization microscopy data. *Nat. Methods* **12,** 1072–1076 (2015).

115. Ha, T. & Tinnefeld, P. Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.* **63,** 595–617 (2012).

116. Yildiz, A. *et al.* Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization. *Science (80-.).* **300,** 2061–2065 (2003).

117. Larson, J. D., Rodgers, M. L. & Hoskins, A. a. Visualizing cellular machines with colocalization single molecule microscopy. *Chem. Soc. Rev.* **43,** 1189–200 (2014).

118. Bates, M., Huang, B., Dempsey, G. T. & Zhuang, X. Multicolor Super-Resolution Imaging with Photo-Switchable Fluorescent Probes. *Science (80-.).* **317,** 1749–1753

(2007).

119. David, H. A. & Nagaraja, H. N. *Order Statistics*. (John Wiley & Sons, Inc, 2003).

120. Churchman, L. S., Okten, Z., Rock, R. S., Dawson, J. F. & Spudich, J. A. Single molecule high-resolution colocalization of Cy3 and Cy5 attached to macromolecules measures intramolecular distances through time. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 1419–23 (2005).

121. Pertsinidis, A., Zhang, Y. & Chu, S. Subnanometre single-molecule localization, registration and distance measurements. *Nature* **466,** 647–651 (2010).

122. Kinz-Thompson, C. D. *et al.* Robustly passivated, gold nanoaperture arrays for single-molecule fluorescence microscopy. *ACS Nano* **7,** 8158–8166 (2013).

123. Smith, C. S., Joseph, N., Rieger, B. & Lidke, K. A. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat. Methods* **7,** 373–375 (2010).

124. Besl, P. J. *et al.* A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **14,** 239–256 (1992).

125. Mathai, A. M. & Provost, S. B. *Quadratic Forms in Random Variables*. (Marcel Dekker, Inc, 1992).

126. Holzmeister, P., Acuna, G. P., Grohmann, D. & Tinnefeld, P. Breaking the concentration limit of optical single-molecule detection. *Chem. Soc. Rev.* **43,** 1014–1028 (2014).

127. Colquhoun, D. & Hawkes, A. G. On the stochastic properties of single ion channels. *Proc R. Soc. Lond. B.* **211,** 205–235 (1981).

128. Schuler, B. & Eaton, W. a. Protein folding studied by single molecule FRET. *Curr. Opin. Struct. Biol.* **18,** 16–26 (2008).

129. Andrec, M., Levy, R. M. & Talaga, D. S. Direct determination of kinetic rates from

single-molecule photon arrival trajectories using hidden Markov models. *J. Phys. Chem. A* **107,** 7454–7464 (2003).

130. Hille, B. *Ion Channel Excitable Membranes*. (Sinauer Associates, Inc, 2001). doi:10.1007/3-540-29623-9_5640

131. *Handbook of Molecular Force Spectroscopy*. (Springer Science, 2008). doi:10.1007/978-0-387-49989-5

132. Vernick, S. *et al.* Electrostatic melting in a single-molecule field-effect transistor with applications in genomic identification. *Nat. Commun.* **8,** 15450 (2017).

133. Colquhoun, D. & Hawkes, A. On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Proc. R. Soc. London Ser. B* **300,** 1–59 (1982).

134. Tan, E. *et al.* A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc. Nat. Acad. Sci. USA* **100,** 9308–9313 (2003).

135. Solomatin, S. V, Greenfeld, M., Chu, S. & Herschlag, D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463,** 681–684 (2010).

136. Fei, J. *et al.* Allosteric collaboration between elongation factor G and the ribosomal L1 stalk directs tRNA movements during translation. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 15702–15707 (2009).

137. Fei, J., Kosuri, P., MacDougall, D. D. & Gonzalez, R. L. Coupling of Ribosomal L1 Stalk and tRNA Dynamics during Translation Elongation. *Mol. Cell* **30,** 348–359 (2008).

138. Lee, J. Y., Okumus, B., Kim, D. S. & Ha, T. Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 18938–18943 (2005).

139. English, B. P. *et al.* Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat. Chem. Biol.* **2,** 87–94 (2006).

140. Caban, K. & Gonzalez, R. L. The emerging role of rectified thermal fluctuations in initiator aa-tRNA- and start codon selection during translation initiation. *Biochimie* **114,** 30–38 (2015).

141. Rosales, R. a. MCMC for hidden Markov models incorporating aggregation of states and filtering. *Bull. Math. Biol.* **66,** 1173–1199 (2004).

142. Li, C.-B. & Komatsuzaki, T. Aggregated Markov Model Using Time Series of Single Molecule Dwell Times with Minimum Excessive Information. *Phys. Rev. Lett.* **111,** 58301 (2013).

143. Uni, S., Qin, F., Auerbach, a & Sachs, F. Maximum likelihood estimation of aggregated Markov processes. *Proc. Biol. Sci.* **264,** 375–83 (1997).

144. Blanco, M. R. *et al.* Single Molecule Cluster Analysis dissects splicing pathway conformational dynamics. *Nat. Methods* **12,** 1077–1084 (2015).

145. Hwang, W., Lee, I. B., Hong, S. C. & Hyeon, C. Decoding Single Molecule Time Traces with Dynamic Disorder. *PLoS Comput. Biol.* **12,** 1–29 (2016).

146. Fine, S., Singer, Y. & Tishby, N. The Hierarchical Hidden Markov Model : Analysis and Applications. *Mach. Learn.* **32,** 41–62 (1998).

147. Wakabayashi, K. & Miura, T. Forward-Backward Activation Algorithm for Hierarchical Hidden Markov Models. *Adv. Neural Inf. Process. ...* 1–9 (2012).

148. Murphy, K. Linear time inference in hierarchial HMMs. *Neural Inf. Process. Syst.* (2001).

149. Frank, J. & Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406,** 318–322 (2000).

150. Jin, H., Kelley, A. C. & Ramakrishnan, V. Crystal structure of the hybrid state of ribosome in complex with the guanosine triphosphatase release factor 3. *Proc. Natl. Acad.*

Sci. U. S. A. **108,** 15798–803 (2011).

151. Munro, J. B., Altman, R. B., O'Connor, N. & Blanchard, S. C. Identification of Two Distinct Hybrid State Intermediates on the Ribosome. *Mol. Cell* (2007). doi:10.1016/j.molcel.2007.01.022

152. Blanchard, S. C., Kim, H. D., Gonzalez, R. L., Puglisi, J. D. & Chu, S. tRNA dynamics on the ribosome during translation. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 12893–12898 (2004).

153. Brunelle, J. L., Youngman, E. M., Sharma, D. & Green, R. The interaction between C75 of tRNA and the A loop of the ribosome stimulates peptidyl transferase activity. *Rna* **12,** 33–39 (2006).

154. Blanchard, S. C. & Puglisi, J. D. Solution structure of the A loop of 23S ribosomal RNA. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 3720–3725 (2001).

155. Widerak, M., Kern, R., Malki, A. & Richarme, G. U2552 methylation at the ribosomal A-site is a negative modulator of translational accuracy. *Gene* **347,** 109–114 (2005).

156. Voorhees, R. M. & Ramakrishnan, V. Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* **82 VN-r,** 203–236 (2013).

157. MacDougall, D. D., Fei, J. & Gonzalez Jr., R. L. Single-molecule fluorescence resonance energy transfer investigations of ribosome-catalyzed protein synthesis. *Mol. Mach. Biol.* 93–116 (2011).

158. Greenleaf, W. J., Woodside, M. T., Abbondanzieri, E. A. & Block, S. M. Passive all-optical force clamp for high-resolution laser trapping. *Phys. Rev. Lett.* **95,** 1–4 (2005).

159. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347–55 (2014).

160. Tinoco, I., Turner, D., Crothers, D. in (University Science Press, 2000).

161. Gracia, B. *et al.* RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly. *J. Mol. Biol.* **428,** 3972–3985 (2016).

162. Ansari, a, Kuznetsov, S. V & Shen, Y. Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 7771–7776 (2001).

163. Guéron, M. & Leroy, J.-L. Base-Pair Opening in Double-Stranded Nucleic Acids. *Nucleic Acids Mol. Biol.* **6,** 1–22 (1992).

164. Kochoyan, M., Lancelot, G. & Leroy, J. L. Study of structure, base-pair opening kinetics and proton exchange mechanism of the d-(AATTGCAATT) self-complementary oligodeoxynucleotide in solution. *Nucleeic Acids Res.* **16,** 7685–7702 (1988).

165. Leroy, J. L., Kochoyan, M., Huynh-Dinh, T. & Gu??ron, M. Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J. Mol. Biol.* **200,** 223–238 (1988).

166. Leroy, J., Broseta, D. & Gueron, M. Proton Exchange and Base-pair Kinetics of Poly (rA) poly (rU) and Poly (r1) poly (rC). **184,** 165–178 (1985).

167. Porschke, D. A direct measurement of the unzippering rate of a nucleic acid double helix. *Biophys. Chem.* **2,** 97–101 (1974).

168. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131,** 17303–17314 (2009).

169. Srinivas, N. *et al.* On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **41,** 10641–10658 (2013).

170. Woese, C. R., Winker, S. & Gutell, R. R. Architecture of ribosomal RNA: constraints on the sequence of 'tetra-loops'. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8467–8471 (1990).

171. Tuerk, C. *et al.* CUUCGG Hairpins: Extraordinarily Stable RNA Secondary Structures Associated with Various Biochemical Processes. *PNAS* **85,** 1364–1368 (1988).

172. Antao, V. P. & Tinoco, I. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.* **20,** 819–824 (1992).

173. Antao, V. P., Lai, S. Y. & Tinoco, I. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res.* **19,** 5901–5905 (1991).

174. Hedenstierna, K. O. F., Siefert, J. L., Fox, G. E. & Murgola, E. J. Co-conservation of rRNA tetraloop sequences and helix length suggests involvement of the tetraloops in higher-order interactions. *Biochimie* **82,** 221–227 (2000).

175. Michel, F. & Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216,** 585–610 (1990).

176. Wu, L., Chai, D., Fraser, M. E. & Zimmerly, S. Structural Variation and Uniformity among Tetraloop-Receptor Interactions and Other Loop-Helix Interactions in RNA Crystal Structures. *PLoS One* **7,** (2012).

177. Zheng, M., Wu, M., Jr, I. T. & Tinoco, I. Formation of a GNRA tetraloop in P5abc can disrupt an interdomain interaction in the Tetrahymena group I ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 3695–3700 (2001).

178. Wu, M., Tinoco, I. & Jr., I. T. RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 11555–11560 (1998).

179. Qin, P. Z., Feigon, J. & Hubbell, W. L. Site-directed spin labeling studies reveal solution conformational changes in a GAAA tetraloop receptor upon Mg2+-dependent docking of a GAAA tetraloop. *J. Mol. Biol.* **351,** 1–8 (2005).

180. Jaeger, L., Michel, F. & Westhof, E. Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* **236,** 1271–1276 (1994).

181. Will, C. L. & Lührmann, R. Spliceosome structure and function. TL - 3. *Cold Spring Harb. Perspect. Biol.* **3 VN-re,** 1–23 (2011).

182. Harish, A. & Caetano-Anolles, G. Ribosomal History Reveals Origins of Modern Protein Synthesis. *PLoS Genet.* **7,** (2012).

183. Caetano-Anollés, G. & Caetano-Anollés, D. Computing the origin and evolution of the ribosome from its structure - Uncovering processes of macromolecular accretion benefiting synthetic biology. *Comput. Struct. Biotechnol. J.* **13,** 427–447 (2015).

184. Chen, K. *et al.* Assembly of the five-way junction in the ribosomal small subunit using hybrid MD-Gō simulations. *J. Phys. Chem. B* **116,** 6819–6831 (2012).

185. Sorin, E. J., Rhee, Y. M., Nakatani, B. J. & Pande, V. S. Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophys. J.* **85,** 790–803 (2003).

186. Sorin, E. J., Engelhardt, M. a, Herschlag, D. & Pande, V. S. RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.* **317,** 493–506 (2002).

187. Sorin, E. J., Rhee, Y. M. & Pande, V. S. Does water play a structural role in the folding of small nucleic acids? *Biophys. J.* **88,** 2516–2524 (2005).

188. Molinaro, M. & Tinoco, I. Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: Thermodynamic and spectroscopic applications. *Nucleic Acids Res.* **23,** 3056–3063 (1995).

189. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M. & Pardi, A. A Network of

Heterogeneous Hydrogen Bonds in GNRA Tetraloops. *J. Mol. Biol* **264,** 968–980 (1996).

190. Monforte, J. a, Kahn, J. D. & Hearst, J. E. RNA folding during transcription by Escherichia coli RNA polymerase analyzed by RNA self-cleavage. *Biochemistry* **29,** 7882–90 (1990).

191. Correll, C. C. & Swinger, K. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. *RNA* **9,** 355–363 (2003).

192. Huang, H.-C., Nagaswamy, U. & Fox, G. E. The application of cluster analysis in the intercomparison of loop structures in RNA. *RNA* **11,** 412–23 (2005).

193. D'Ascenzo, L., Leonarski, F., Vicens, Q. & Auffinger, P. Revisiting GNRA and UNCG folds: U-turns versus Z-turns in RNA hairpin loops. *RNA* (2016). doi:10.1261/rna.059097.116

194. Blose, J. M., Proctor, D. J., Veeraraghavan, N., Misra, V. K. & Bevilacqua, P. C. Contribution of the closing base pair to exceptional stability in RNA tetraloops: Roles for molecular mimicry and electrostatic factors. *J. Am. Chem. Soc.* **131,** 8474–8484 (2009).

195. Menger, M., Eckstein, F. & Porschke, D. Dynamics of the RNA hairpin GNRA tetraloop. *Biochemistry* (2000). doi:10.1021/bi992297n

196. Greenleaf, W. J., Frieda, K. L., Foster, D. a N., Woodside, M. T. & Block, S. M. Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **319,** 630–633 (2008).

197. Woodside, M. T. *et al.* Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* **314,** 1001–4 (2006).

198. Woodside, M. T. *et al.* Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 6190–6195

(2006).

199. Cocco, S., Marko, J. F. & Monasson, R. Slow nucleic acid unzipping kinetics from sequence-defined barriers. *Eur. Phys. J. E* **10,** 153–161 (2003).

200. Stancik, A. L. & Brauns, E. B. Rearrangement of partially ordered stacked conformations contributes to the rugged energy landscape of a small RNA hairpin. *Biochemistry* **47,** 10834–10840 (2008).

201. Sarkar, K., Meister, K., Sethi, A. & Gruebele, M. Fast folding of an RNA tetraloop on a rugged energy landscape detected by a stacking-sensitive probe. *Biophys. J.* **97,** 1418–1427 (2009).

202. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 4899–903 (2001).

203. Hodak, J., Downey, C. D., Fiore, J. L., Pardi, A. & Nesbitt, D. J. Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. **102,** (2005).

204. Fiore, J. L. & Nesbitt, D. J. An RNA folding motif: GNRA tetraloop-receptor interactions. *Q. Rev. Biophys.* **46,** 223–64 (2013).

205. Sorgenfrei, S. *et al.* Label-free single-molecule detection of DNA-hybridization kinetics with a carbon nanotube field-effect transistor. *Nat Nanotechnol* **6,** 126–132 (2011).

206. Proctor, D. J., Schaak, J. E., Bevilacqua, J. M., Falzone, C. J. & Bevilacqua, P. C. Isolation and characterization of stable tetraloops with the motif YNMG that participates in tertiary interactions. *Biochemistry* **41,** 12062–12075 (2002).

207. Snoussi, K. & Leroy, J. L. Imino proton exchange and base-pair kinetics in RNA

duplexes. *Biochemistry* **40,** 8898–8904 (2001).

208. Wallace, M. I., Ying, L., Balasubramanian, S. & Klenerman, D. FRET Fluctuation Spectroscopy: Exploring the Conformational Dynamics of a DNA Hairpin Loop. *J. Phys. Chem. B* **104,** 11551–11555 (2000).

209. Yin, Y. *et al.* Panorama of DNA hairpin folding observed via diffusion-decelerated fluorescence correlation spectroscopy. *Chem. Commun.* **48,** 7413 (2012).

210. Tsukanov, R. *et al.* Detailed study of DNA hairpin dynamics using single-molecule fluorescence assisted by DNA origami. *J. Phys. Chem. B* **117,** 11932–11942 (2013).

211. Lipfert, J., Doniach, S., Das, R. & Herschlag, D. *Understanding nucleic Acid-ion interactions. Annual review of biochemistry* **83,** (2014).

212. Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negroni, A. & Al-hashimi, H. M. Visualizing transient low-populated structures of RNA. *Nature* **491,** 724–8 (2012).

213. Nikolova, E. N. *et al.* Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470,** (2011).

214. Sathyamoorthy, B. *et al.* Insights into Watson–Crick/Hoogsteen breathing dynamics and damage repair from the solution structure and dynamic ensemble of DNA duplexes containing m 1 A. *Nucleic Acids Res.* **45,** 1–16 (2017).

215. Zhou, H. *et al.* New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. *Nucleic Acids Res.* **43,** 3420–33 (2015).

216. Yang, X., Gérczei, T., Glover, L. T. & Correll, C. C. Crystal structures of restrictocin-inhibitor complexes with implications for RNA recognition and base flipping. *Nat. Struct. Biol.* **8,** 968–73 (2001).

217. Ouldridge, T. E., Šulc, P., Romano, F., Doye, J. P. K. & Louis, A. A. DNA hybridization

kinetics: Zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* **41,** 8886–8895 (2013).

218. Mosayebi, M., Romano, F., Ouldridge, T. E., Louis, A. A. & Doye, J. P. K. The role of loop stacking in the dynamics of DNA hairpin formation. *J. Phys. Chem. B* **118,** 14326–14335 (2014).

219. Ouldridge, T. *Coarse-Grained Modeling of DNA and DNA Self-Assembly*. *Journal of Chemical Information and Modeling* **53,** (2012).

220. Ellington, a D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–22 (1990).

221. Ellington, A. & Szostak, J. Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature* **355,** 850–852 (1992).

222. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–510 (1990).

223. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–6 (2002).

224. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* **113,** 577–586 (2003).

225. Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nat Biotechnol* **24,** 1558–1564 (2006).

226. Penchovsky, R. & Breaker, R. R. Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. *Nat. Biotechnol.* **23,** 1424–1433 (2005).

227. Desai, S. K. & Gallivan, J. P. Genetic screens and selections for small molecules based on

a synthetic riboswitch that activates protein translation. *J. Am. Chem. Soc.* **126,** 13247–13254 (2004).

228. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* **11,** 29–35 (2004).

229. Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **11,** 1729–1741 (2004).

230. Wickiser, J. K., Cheah, M. T., Breaker, R. R. & Crothers, D. M. The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **44,** 13404–13414 (2005).

231. Lemay, J. F., Penedo, J. C., Tremblay, R., Lilley, D. M. J. & Lafontaine, D. A. Folding of the Adenine Riboswitch. *Chem. Biol.* **13,** 857–868 (2006).

232. Frieda, K. L. & Block, S. M. Direct Observation of Cotranscriptional Folding in an Adenine Riboswitch. *Science (80-.).* **338,** 397–400 (2012).

233. Lemay, J. F. *et al.* Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet.* **7,** (2011).

234. Frieda, K. L. & Block, S. M. Direct Observation of Cotranscriptional Folding in an Adenine Riboswitch. *Science (80-.).* **338,** 397–400 (2012).

235. Nozinovic, S. *et al.* The importance of helix P1 stability for structural pre-organization and ligand binding affinity of the adenine riboswitch aptamer domain. *RNA Biol.* **11,** 83–82 (2014).

236. Marcano-Velázquez, J. G. & Batey, R. T. Structure-guided mutational analysis of gene regulation by the Bacillus subtilis pbuE adenine-responsive riboswitch in a cellular context. *J. Biol. Chem.* **290,** 4464–4475 (2015).

237. Gilbert, S. D., Stoddard, C. D., Wise, S. J. & Batey, R. T. Thermodynamic and Kinetic Characterization of Ligand Binding to the Purine Riboswitch Aptamer Domain. *J. Mol. Biol.* **359,** 754–768 (2006).

238. Buck, J., Fürtig, B., Noeske, J., Wöhnert, J. & Schwalbe, H. Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 15699–15704 (2007).

239. Noeske, J. *et al.* Interplay of 'induced fit' and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Res.* **35,** 572–583 (2007).

240. Ottink, O. M. *et al.* Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined induced fit mechanism. *Rna* **13,** 2202–2212 (2007).

241. Brenner, M. D., Scanlan, M. S., Nahas, M. K., Ha, T. & Silverman, S. K. Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine. *Biochemistry* **49,** 1596–1605 (2010).

242. Choi, Y. *et al.* Single-molecule dynamics of lysozyme processing distinguishes linear and cross-linked peptidoglycan substrates. *J. Am. Chem. Soc.* **134,** 2032–2035 (2012).

243. Wickiser, J. K., Winkler, W. C., Breaker, R. R. & Crothers, D. M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell* **18,** 49–60 (2005).

244. Lutz, B., Faber, M., Verma, A., Klumpp, S. & Schug, A. Differences between cotranscriptional and free riboswitch folding. *Nucleic Acids Res.* **42,** 2687–2696 (2014).

245. Tolic-Norrelykke, S. F., Engh, A. M., Landick, R. & Gelles, J. Diversity in the Rates of

Transcript Elongation by Single RNA Polymerase Molecules. *J. Biol. Chem.* **279,** 3292–3299 (2004).

246. Adelman, K. *et al.* Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 13538–13543 (2002).

247. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311,** 1600–1603 (2006).

248. Artsimovitch, I. *et al.* RNA Polymerases from Bacillus subtilis and Escherichia coli Differ in Recognition of Regulatory Signals In Vitro RNA Polymerases from Bacillus subtilis and Escherichia coli Differ in Recognition of Regulatory Signals In Vitro. (2000). doi:10.1128/JB.182.21.6027-6035.2000.Updated

249. Komissarova, N. & Kashlev, M. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 14699–14704 (1998).

250. Hanke, C. A. & Gohlke, H. Ligand-mediated and tertiary interactions cooperatively stabilize the P1 region in the guanine-sensing riboswitch. *PLoS One* **12,** 1–29 (2017).

251. Gardner, P. P. *et al.* Rfam : updates to the RNA families database. **37,** 136–140 (2009).

252. Eichhorn, C. D. *et al.* Unraveling the structural complexity in a single-stranded RNA tail: Implications for efficient ligand binding in the prequeuosine riboswitch. *Nucleic Acids Res.* **40,** 1345–1355 (2012).

253. Rinaldi, A. J., Lund, P. E., Blanco, M. R. & Walter, N. G. The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs show ligand-dependent accessibility bursts. *Nat. Commun.* **7,** 1–10 (2016).

254. Lescoute, A. & Westhof, E. Topology of three-way junctions in folded RNAs. *RNA* **12,**

83–93 (2006).

255. Selmer, M. *et al.* Structure of the 70S Ribosome Complexed with mRNA and tRNA. *Science (80-.).* **313,** 1935–1942 (2006).

256. Ober, R. J., Ram, S. & Ward, E. S. Localization accuracy in single-molecule microscopy. *Biophys. J.* **86,** 1185–1200 (2004).

257. Sternberg, S. H., Fei, J., Prywes, N., McGrath, K. a & Gonzalez, R. L. Translation factors direct intrinsic ribosome dynamics during translation termination and ribosome recycling. *Nat. Struct. Mol. Biol.* **16,** 861–868 (2009).

258. Fei, J. *et al. Chapter 12 - A Highly Purified, Fluorescently Labeled In Vitro Translation System for Single-Molecule Studies of Protein Synthesis*. *Methods in Enzymology* **Volume 472,** (Elsevier Inc., 2010).

259. GE Dharmacon. Deprotection 2 â€™ - ACE Protected RNA. 9880 (2014).

260. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31,** 3406–3415 (2003).

261. Ward, D. C., Reich, E. & Stryer, L. Fluorescence studies of nucleotides and polynucleotides. *J. Bioloi Chem.* **244,** 1228–1237 (1969).

Appendix A Probability Basics, Assorted Proofs, Update Equations

Axioms of Probability

The mathematician Andrey Kolmogorov is generally credited with the three axioms of probability which, as they seek to give structure to common sense, have intuitive underpinnings. Before discussing the axioms themselves, I want to give the reader a few helpful definitions – first, the idea of a "sample space" denoted S – an abstract, nonempty set that contains elements called "events," which are things that can admissibly happen. Finally, the concept of a "measure" is useful because we will define a specific kind associated with probabilities: a "measure" is a function that takes a set into the positive real line that is "linear" in the sense that the measure of a collection of disjoint subsets is equal to each measured independently then summed together. Loosely then, the axioms of probability are the rules by which we can provide a measure, in the rigorous sense, of how likely an event is to occur, given the body of things that could possibly happen, i.e., given the sample space.

A. 1: The probability measure is given by an operator called p that acts on the sample space, a nonempty set S. $1 \geq p(s) \geq 0 \; \forall \; s \in S$, that is, p maps the elements of the set S, whatever they are, onto the positive real line between 0 and 1.

Explanation: being positive gives the number "zero" relevance, as a number that describes events that can never happen; indeed this axiom is necessary to show that the probability measure of the empty set is zero.

Appendix A

A. 2: $\boldsymbol{p(S) = 1}$
Explanation: all possible events are described by the sample space S. This axiom gives the number 1 relevance – it denotes certainty.

A. 3: $\boldsymbol{p(\cup_i E_i) = \sum_i p(E_i)}$ given a collection $\boldsymbol{E_i}$ such that $\boldsymbol{E_i \cap E_j = \emptyset}$
Explanation: probability can be measured in units and does not depend on how mutually entwined the events are (in fact, those parts cancel out, as I will show below).

These axioms trivially give rise to the sum rule for events, the product rule for independent events, and an important result known as Bayes' theorem. We now derive each of these. To begin with, we define a conditional probability by the following formula: $p(A|B) = \frac{p(A \cap B)}{p(B)}$ in terms of the events A and B, assuming that B can actually occur.

The sum rule can be seen as a direct consequence of the third axiom (the assumption that disjoint sets are additive):

$$\boldsymbol{p(A \cup B) = p(A) + p(B \backslash (A \cap B))}$$
$$\boldsymbol{p(B) = p(B \backslash (A \cap B)) + p(A \cap B)}$$

Rearranging gives:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

Next, independence is defined as:

$$p(A \cap B) = p(A)p(B)$$

This definition, with the definition of conditional probability, shows what independence means in terms of conditionals:

$$p(A|B) = p(A)$$

which has a simple intuition – any two events are independent if the probability of one event occurring is not dependent on the occurrence of the other. Finally, it is simple to derive Bayes'

theorem, which is a consequence of commutativity of the intersection property and for which we require that A and B can actually occur:

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{p(B \cap A)}{p(B)} = \frac{p(B|A)p(A)}{p(B)}$$

A. 4

It is important to note that the sum rule and Bayes' theorem both manifest an important property of the measure definition of probability – it preserves the algebraic set structure. For this reason, the sample space is often restricted, in its definition, to a σ-algebra.

Random Variable Algebra

I begin with the notions, familiar from deterministic functions on the real line, of sums and products. I will point out that these are simply manifestations of different types of convolutions, one using the Fourier kernel and one using the Mellin kernel. When measuring physical objects it is often necessary to form a sum of two random variables which are independent of each other, for instance when X and Y are independent, we can form the sum:

$$Z = X + Y$$

The probability measure of Z, p_Z, can be found directly from the probability measures of X and Y, p_X and p_Y. To do so, we construct p_Z by setting a number on the real line and enumerating all the possibilities in the sample space (and understanding that integration is summation over limits of simple functions in the sample space, which we achieve for free using Kolmogorov's axioms):

$$p_Z(t) = \int p(X = t - \tau \cap Y = \tau) d\tau$$

which, using independence, gives:

Appendix A

$$p_Z(t) = \int p_X(t-\tau)p_Y(\tau)d\tau$$

The right hand side is a Fourier convolution (or occasionally, simply a convolution), and is related to the Fourier transform, denoted by $F[f](\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t)dt$ and its inverse $F^{-1}[F[f]](t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t} F[f](\omega)d\omega$. The function $F[p](\omega)$ is occasionally called the characteristic function because it has a fundamental connection to some important descriptors of the probability measure known as moments. This operation is conceptually similar to solving problems in momentum space rather than position space in quantum mechanics. The convolution theorem allows:

$$F[p_Z](\omega) = F[p_X](\omega)F[p_Y](\omega)$$

and therefore:

$$p_Z(t) = F^{-1}[F[p_X]F[p_Y]](t)$$

A. 5

Similarly, for two independent random variables which only map numbers to the positive real line, X and Y (the negative parts of random variables that map to the whole line can always be found by a coordinate transform and splitting the product distribution into four parts) the product can be defined:

$$Z = XY$$

The probability measure of Z, p_Z, can be found directly from the probability measures of X and Y, p_X and p_Y. To do so, construct p_Z by setting a number on the real line and enumerating all the possibilities in the sample space (and understanding that integration is summation over limits of simple functions in the sample space, which we achieve using Kolmogorov's axioms):

Appendix A

$$p_Z(w < t) = \int p\left(X < \frac{t}{u} \cap Y = u\right) du$$

$$p_Z(w < t) = \int p_X\left(w < \frac{t}{u}\right) p_Y(u) du$$

$$p_Z(t) = \int \frac{1}{u} p_X\left(\frac{t}{u}\right) p_Y(u) du$$

In the same way that the probabilistic sum was related to the Fourier convolution, this random variable, the probabilistic product, is related to the Mellin convolution. Typically the Mellin transform is defined by $M[f](s) = \int_0^\infty x^{s-1} f(x) dx$, where s is a complex number, and its inverse by $M^{-1}[M(f)](x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s} M[f] ds$, where c is any number within the strip in the complex plane upon which the transform is defined. Using the Mellin convolution theorem (also known as the exchange formula), we can write:

$$M[p_Z](s) = M[p_X](s) M[p_Y](s)$$

and therefore:

$$p_Z(t) = M^{-1}[M[p_X] M[p_Y]](t)$$

A. 6

I will require one final formula, which is non-algebraic but is simple to derive. Given some collection of independent random variables $\{X_i\}$, the probability measure over their extreme values may be found:

$$p_{max}(\max(X_i) < t) = \prod_i p(X_i < t) = \prod_i \int p_{X_i}(t) dt$$

A. 7

The distribution over the minimum can be found by considering the complement of the cumulative distribution.

Appendix A

The following may be observed: first, independent normal random variables, which are not necessarily identically distributed, form an abelian semigroup with no inverse, also known as a monoid, under their probabilistic sum or difference. To show this, we note that the characteristic function of a normally distributed random variable parameterized by μ_x and σ_x is given by:

$$F[p_X](\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(t-\mu_x)^2}{2\sigma_x^2}} dt = e^{i\mu_x\omega - \frac{(\omega\sigma_x)^2}{2}}$$

Therefore the probabilistic sum (or difference) of two independent random variables X and Y has the characteristic function:

$$F[p_Z](\omega) = e^{i\mu_x\omega - \frac{(\omega\sigma_X)^2}{2}} e^{\pm i\mu_Y\omega - \frac{(\omega\sigma_Y)^2}{2}} = e^{i(\mu_x \pm \mu_Y)\omega - \frac{(\sigma_x^2+\sigma_x^2)\omega^2}{2}}$$

which is still the characteristic function of a normally distributed random variable with suitably modified parameters. Note that the limit:

$$\lim_{\sigma_X \to 0} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{t^2}{2\sigma_x^2}} = \delta(t)$$

which defines the identity element. It is simply something deterministic. There is no inverse because it is not possible to arrive at the identity element by adding two normal random variables together, because one cannot add two positive numbers and get zero. Put another way, any normal randomness precludes certainty: it is not possible to arrive at certainty using uncertain methods.

Second, we note that independent Poisson random variables form an abelian semigroup without an inverse under the probabilistic sum. The characteristic function is given by:

$$F[p](\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} p(t) dt = \sum_{t=0}^{\infty} \frac{e^{-i\omega t} e^{-\lambda} \lambda^t}{t!} = e^{-\lambda} \sum_{t=0}^{\infty} \frac{(e^{-i\omega}\lambda)^t}{t!} = e^{(e^{-i\omega}-1)\lambda}$$

Therefore the probabilistic sum (but not the difference!) of two independent random variables X and Y has the characteristic function:

$$F[p_Z](\omega) = e^{(e^{-i\omega}-1)\lambda_X} e^{(e^{-i\omega}-1)\lambda_Y} = e^{(e^{-i\omega}-1)(\lambda_Y+\lambda_X)}$$

which is still the characteristic function of a Poisson distributed random variable with a modified rate parameter. The difference, however, is not a Poisson distributed random variable (i.e., background subtraction with random variables is a complicated matter.) This is easy to see, because the characteristic function of the difference is:

$$F[p_Z](\omega) = e^{(e^{-i\omega}-1)\lambda_X} e^{(e^{i\omega}-1)\lambda_Y} = e^{(e^{-i\omega}-1)\lambda_X+(e^{i\omega}-1)\lambda_Y}$$

which may be inverted to give the Skellam distribution.

Supporting Information for Unified, Bayesian Inference-based Framework for Analyzing Single-molecule Fluorescence Microscopy Experiments

Note: code for all projects described in this dissertation may be found on the Gonzalez lab server on the path < /home/jtemp2/Desktop/code>, and is also available upon request directed to the author at < jhon0882@gmail.com >.

Theoretical Analysis of Trace Estimation Methods

1. Derivation of the ML formula for the Amplitude

The random variable d_{xy} which corresponds to the value of the intensity measured in a pixel at position (x,y) in an image is given by

$$d_{xy} = C + B_{xy} + N_{xy},$$

A. 8

where $C \sim \mathcal{N}(\mu_c, \Sigma_c)$ and $B_{xy} \sim Poisson(k_{B,xy})$ respectively denote the instrumental noise, normally distributed and assumed independent of position, and electron and photon counts from

dark current or background fluorescence, Poisson distributed, and possibly inhomogeneous due to the illumination profile. $N_{xy}{\sim}Poisson(k_{N,xy})$ denotes the random number of photon incidences from all the chromophores at the given pixel at (x,y). Note that here we use the notation for the normal distribution of $\mathcal{N}(\text{mean}, \text{variance})$ and for the Poisson distribution of $Poisson(\text{rate})$.

In the following derivation, we will make use of several assumptions:

(1) Distinct molecules are independent of each other.

(2) The number of background electron or photon counts $k_{B,xy}$ collected in any pixel at any given time interval, corresponding to the Poisson rate, is high (>100).

(3) The variance of the variable d_{xy} is homogenous in a larger neighborhood than $\Psi_{i,xy}$ falls off to zero.

These assumptions allow us to arrive at the formulas provided in the main text, as well as to clarify technical details. However, we begin our derivation with more general equations where these assumptions have not yet been made, which will support implementation of other optimization techniques in future work. First, we note that because the sum of two Poisson random variables is itself Poisson distributed with a modified mean, such that the sum $B_{xy} + N_{xy}{\sim}Poisson(k_{B,xy} + \sum_i N_i \Psi_{i,xy})$, a result that makes use of assumption (1). Next, given assumption (2), we note that the Poisson probability distribution governing this sum may also be approximated by a normal distribution as

$$B + N_{xy}{\sim}\mathcal{N}\left(k_{B,xy} + \sum_i N_i \Psi_{i,xy}\,, k_{B,xy} + \sum_i N_i \Psi_{i,xy}\right).$$

At this point we can write the probability distribution of the intensity random variable, which is

Appendix A

$$d_{xy} \sim \mathcal{N}\left(\mu_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy}\,, \Sigma_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy}\right) \equiv \mathcal{N}\left(\mu_{I,xy}, \Sigma_{I,xy}\right).$$

A. 9

Noting that each pixel is independent given the form above, we give the likelihood of an image

$$p(image) = \prod_{x,y} \mathcal{N}\left(d_{xy} | \mu_{I,xy}, \Sigma_{I,xy}\right)$$

$$\mathcal{L} \equiv ln\left(p(image)\right) = -\sum_{x,y} ln(2\pi) + \sum_{x,y}\left(-\frac{1}{2} ln\left(\Sigma_{I,xy}\right) - \frac{1}{2\Sigma_{I,xy}}\left(d_{xy} - \mu_{I,xy}\right)^2\right)$$

$$\mathcal{L} = -\frac{1}{2}\sum_{x,y} ln(2\pi)$$

$$-\frac{1}{2}\sum_{x,y}\left(ln\left(\Sigma_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy}\right)\right.$$

$$\left. - \frac{\left(d_{xy} - \left(\mu_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy}\right)\right)^2}{\left(\Sigma_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy}\right)}\right),$$

A. 10

where sums and products over *x,y* are over all pixels in the image. A version of this likelihood without using assumption (2) and therefore not utilizing a Normal approximation to the Poisson distribution may be found in Ober *et al*.[256] Finding the most likely parameters is equivalent to maximizing $\mathcal{L}$, which is equivalent to maximizing $p(image)$ because logarithms are monotonic.

We seek the most likely set of incident photon rates N_i, and find them by setting

Appendix A

$$\frac{\partial \mathcal{L}}{\partial N_i} = 0 = \sum_{x,y} \left(\frac{\Psi_{i,xy}}{2\left(\Sigma_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy}\right)} \right.$$

$$- \frac{\Psi_{i,xy}}{\Sigma_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy}} \left(\left(d_{xy} - \left(\mu_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy} \right) \right) \right.$$

$$\left. \left. - \frac{(d_{xy} - (\mu_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy}))^2}{2(\Sigma_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy})} \right) \right).$$

A. 11

This formula has no analytic solution, is self-referential, and has a singularity when assumption (2) breaks down. While it is possible to stop here and use gradient descent to determine the various N_i, continuing to simplify this expression has two advantages: first, it lowers the number of parameters, which is necessary because there are more parameters than observations due to the Poisson background rates; second, it allows us to use a highly stable technique with theoretically guaranteed convergence known as expectation maximization (EM).[76] Making use of assumption (3), we get a simpler formula, which we expound upon by defining an operator $L(i)$, which gives the (x, y) pixel position of the light-emitting chromophore indexed by i, and another operator $O(L(i))$, which gives the neighborhood about $L(i)$ on which $\Psi_{i,xy}$ is, for the most part, nonvanishing. This allows us to write

$$\frac{\partial \mathcal{L}}{\partial N_i} = 0 = \sum_{(x,y) \in O(L(i))} \left(\frac{1}{\Sigma_{d,L(i)}} \left(d_{xy} - \left(\mu_c + k_{B,L(i)} + \sum_{j \neq i} N_j \Psi_{j,xy} \right) - N_i \Psi_{i,xy} \right) \Psi_{i,xy} \right),$$

where $\Sigma_{d,L(i)} = \Sigma_c + k_{B,xy} + \sum_i N_i \Psi_{i,xy}$ which is assumed constant in $O\big(L(i)\big)$. . Rearrangement of this equation and defining $b_{L(i)} = \mu_c + k_{B,L(i)}$, immediately gives

$$N_i = \frac{\sum_{(x,y)\in O(L(i))} \Big(\big(d_{xy} - b_{L(i)} - \sum_{j\neq i} N_j \Psi_{j,xy} \big) \Psi_{i,xy} \Big)}{\sum_{(x,y)\in O(L(i))} \Psi_{i,xy}^2}.$$

A. 12

An analogous calculation, for a small enough neighborhood and utilizing assumption (2), yields an expression for $b_{L(i)}$ of

$$b_{L(i)} = \frac{\sum_{(x,y)\in O(L(i))} \big(d_{xy} - \sum_i N_i \Psi_{i,xy} \big)}{\sum_{(x,y)\in O(L(i))} (1)}$$

A. 13

Optimal values of these parameters are found using an expectation-maximization routine (EM) iterating successive updates of these parameters using the formulas above. To this end, we typically initialize the algorithm (iteration index shown in parenthesis in the superscript) by providing the following guesses for the photon rates and background. First, we define the following

$$b_{L(i)}^{(0)} = \frac{\sum_{(x,y)\in O(L(i))} d_{xy}}{\sum_{(x,y)\in O(L(i))} (1)}$$

which constitutes an initial guess for the background of a local area, and

$$N_i^{(0)} = d_{L(i)} - b_{L(i)}^{(0)}$$

A more robust guess could be given by taking into account the contributions of neighboring light-emitting chromophores. Observing that:

$$d_{L(i)} = \sum_j \Psi_{j,L(i)} N_j + b_{L(i)}$$

and defining the square matrix $\boldsymbol{A} = \{\Psi_{j,L(i)}\}$, the vector of the amplitudes as $N = \{N_i\}$, and the vector of the observed maxima as $M = \{d_{L(i)} - b^{(0)}_{L(i)}\}$ we can write a formula for an alternative initial guess of the N_j:

$$N = \boldsymbol{A}^{-1}M$$

which may be solved, for instance, utilizing the *linsolve* function in MATLAB. With either of these guesses, the algorithm typically converges within five iterations. When fitting the locations of chromophores alongside the amplitude and per-frame background, we assume that while the amplitudes and background vary per-frame, the position is fixed; with this restriction, we numerically maximize the likelihood function, making use of all of our assumptions above, and the constraint that the chromophore is actually located within a pixel of the one in which we identified it.

2. Variance analysis of the ML formula for the Amplitude

With an expression for the estimate of N_i and the variance in that estimate, $Var(N_i)$, we calculate the theoretical signal-to-noise ratio (SNR) of a chromophore intensity estimated with a particular method as $SNR = N_i/\sqrt{Var(N_i)}$. To assist in the derivations of the SNR for estimation methods discussed in the main text, we note that

$$Var\left(\sum_i b_i X_i\right) = \sum_i \left(b_i^2 Var(X_i)\right), \text{ and}$$

$$Var(a + bX) = b \cdot Var(X),$$

when X_i are normally distributed, and a, b, and b_i are constants.

Appendix A

2.1 Direct PSF

In the case of only one light-emitting chromophore

$$N_i = \frac{\sum_{(x,y)\in O(L(i))}\left(\left(d_{xy} - b_{L(i)}\right)\Psi_{i,xy}\right)}{\sum_{(x,y)\in O(L(i))}\Psi_{i,xy}^2}.$$

Defining

$$\gamma \equiv \sum_{(x,y)\in O\left(L(i)\right)} \Psi_{i,xy}^2 \text{, and}$$

$$a \equiv \frac{-\sum_{(x,y)\in O(L(i))} b_{L(i)}\Psi_{i,xy}}{\gamma},$$

allows us to rewrite the formula for the amplitude as

$$N_i = \frac{1}{\gamma}\sum_{(x,y)\in O(L(i))} \Psi_{i,xy} d_{xy} + a.$$

Noting that the d_{xy} are normally distributed, and utilizing the assumptions leading to the derivation of the amplitude formula (particularly assumption (3)), yields

$$Var(N_i) = \frac{Var\left(d_{xy}\right)}{\gamma} \cong \frac{\Sigma_{d,L(i)}}{\gamma}.$$

This result holds for an arbitrary form of $\Psi_{i,xy}$. Applying this result to the case of a symmetric 2D Gaussian PSF parameterized as

$$\psi(\theta_{PSF}) = \frac{1}{\sqrt{2\pi\Sigma_{i,xy}}} Exp\left(-\frac{1}{2\Sigma_{i,xy}}((x - x_i)^2 + (y - y_i)^2)\right),$$

and exchanging the double sum for a double integral gives

$$Var(N_i) = \frac{\Sigma_{d,L(i)}\Sigma_{i,xy}}{\pi}.$$

Appendix A

2.2 Direct Summation

Estimating N_i by summing the pixels in the region $O(L(i))$ surrounding the i[th] chromophore, and then removing the background contribution estimated from the area surrounding $O(L(i))$, defined as $O'\big(L(i)\big)$, yields

$$N_i \equiv \frac{1}{Card\left(O\big(L(i)\big)\right)} \sum_{(x,y)\in O(L(i))} d_{xy} - \frac{1}{Card\left(O'\big(L(i)\big)\right)} \sum_{(x,y)\in O'(L(i))} d_{xy},$$

where $Card(region)$ is the cardinality of the specified region (i.e., number of pixels). Given this formula, and the assumption made in Sec. 2.1 about the variance of the d_{xy} yields

$$Var(N_i) = \Sigma_{d,L(i)} \left(\frac{1}{Card(O(L(i)))} + \frac{1}{Card(O'(L(i)))}\right).$$

We note that $O(L(i))$ is typically defined as the one-pixel neighborhood about the pixel identified as containing the light-emitting chromophore, and $O'(L(i))$ is typically defined as one pixel further around $O(L(i))$, but not containing any pixels from $O(L(i))$.

3. Photobleaching Correction

Since photobleaching of a chromophore involves an effectively irreversible chemical reaction that results in an effectively instantaneous drop in the photon-emission rate of a chromophore, we choose to locate photobleaching points by learning the types of instantaneous changes in intensity that occur in a collection of intensity versus time trajectories. As such, we learn where these instantaneous changes occur by using a Gaussian mixture model (GMM) and utilizing VBEM.[76]

First, we define $\Delta I_k^c(t)$ as the derivative of the intensity in color channel c at time t in chromophore k. We are primarily concerned with whether a particular $\Delta I_k^c(t)$ belongs to a class

of $\Delta I_k^c(t)$ describing an increase, lack of change, or decrease in intensity. Thus, we can write a factorized, joint probability distribution as

$$p(z, \theta) \cong q(\theta)q(z),$$

where z is a one-of-K binary vector describing whether a data point belongs to a particular class (denoted $+, 0,$ or $-$ for increase, same, or decrease, respectively), and θ describes the parameters of the Gaussians. This particular factorization is tractable in a GMM, and enables us to utilize a variational approximation. By defining

$$r_+(t,k) = \frac{\mathcal{N}(\Delta I_k^c(t)\ |\mu_+, \lambda_+)}{R},$$

$$r_0(t,k) = \frac{\mathcal{N}(\Delta I_k^c(t)\ |\mu_0, \lambda_0)}{R},$$

$$r_-(t,k) = \frac{\mathcal{N}(\Delta I_k^c(t)\ |\mu_-, \lambda_-)}{R}, \text{ and}$$

$$R = \mathcal{N}(\Delta I_k^c(t)\ |\mu_+, \lambda_+) + \mathcal{N}(\Delta I_k^c(t)\ |\mu_+, \lambda_+) + \mathcal{N}(\Delta I_k^c(t)\ |\mu_+, \lambda_+),$$

we arrive at the following expressions for the occupancy and parameter posteriors

$$q(z) = \prod_k \prod_t r_+(t,k)^{z(+,t,k)} r_0(t,k)^{z(0,t,k)} r_-(t,k)^{z(-,t,k)} \text{ , and}$$

$$q(\theta)$$

$$= Dir(\boldsymbol{\pi}|\boldsymbol{\alpha})\mathcal{N}(\mu_+|m_+, B_+)\Gamma(\lambda_+|a_+, b_+)\mathcal{N}(0|0, B_0)\Gamma(\lambda_0|a_0, b_0)\mathcal{N}(\mu_-|m_-, B_-)\Gamma(\lambda_-|a_-, b_-).$$

A. 14

$Dir(\boldsymbol{\pi}|\boldsymbol{\alpha})$ denotes the Dirichlet distribution over the mixing coefficients, $\mathcal{N}(\Delta I_k^c(t)\ |\mu, \lambda)$ denotes a normal distribution over the intensity derivatives, which is parameterized in terms of precision (inverse variance), $\mathcal{N}(\mu|m, B)$ denotes the posterior distribution over the mean, and $\Gamma(\lambda|a, b)$ denotes the gamma posterior distribution over the precision. A full discussion of these

distributions, and the variational analysis to solve the model can be found in standard references.[76]

The photobleaching event in a particular intensity versus time trajectory is defined as the last occupancy of the distribution governed by $\mathcal{N}(\Delta I_k^c(t)|\mu_-, \lambda_-)$ before the termination of recording. With high quality data when the assumptions of the model are met, this model works reasonably well for >90% of the trajectories. We suggest the use of a point-and-click method to fix those that do not conform; such functionality is included in vbscope.

4. Anti-correlation Sorting

Defining the un-normalized cross-correlation function as

$$CCF(t, c, c') = \langle I_k^c(t) I_k^{c'}(t - t') \rangle,$$

we simply sort the trajectories by $CCF(t = 0, c, c')$.

5. Release Factor 1 and smFRET Methods

These methods primarily expand upon the work of Sternberg et al.[257] Further details can be found elsewhere.[258]

5.1 Recombinant Gene Construction, Expression, Purification, and Labeling of Cy5-mutRF1

Beginning with the plasmid containing the gene encoding the N-terminal hexahistidine tagged, single-cysteine variant of release factor 1 (RF1) described by Sternberg et al,[257] a variant of this construct was created using site-directed mutagenesis by introducing two mutations – G896C and G902C in the DNA sequence of the prfA gene numbering of *Escherichia coli (E.*

coli) K12. As described previously, this plasmid was co-transformed into *E. coli* BL21-(DE3) with a plasmid containing the gene for the N^5-glutamine methyltransferase encoded by prmC, which is responsible for methylating the GGQ motif of class 1 release factors.[257] Over-expression of both the resulting RF1 mutant (mutRF1) and the methyltransferase was induced with isopropyl β-D-1-thiogalactopyranoside. MutRF1 was then purified using nickel-affinity chromatography. The hexahistidine tag was then removed using TEV protease to cleave the tag at the down-stream TEV protease site, and purified again using nickel-affinity chromatography. This single-cysteine mutRF1 was labeled using Cy5-maleimide (GE Healthcare Life Sciences), and purified using size exclusion chromatography (Superdex 75 pg; GE Healthcare Life Sciences) to remove unreacted Cy5 and then hydrophobic interaction chromatography (Phenyl 5PW; Tosoh Biosciences) to remove unlabeled mutRF1, as previously described.[257] This procedure yields a 100% labeling efficiency of mutRF1 by Cy5 (Cy5-mutRF1).

5.2 Ribosomal Release Complex Formation

Cy3-labeled, prokaryotic, ribosomal release complex (Cy3-RC) programmed with a stop-codon in the aminoacyl-tRNA site was formed as previously described.[257,258] Briefly, Cy3-RC was prepared enzymatically in tris-polymix buffer (50 mM tris-acetate (pH=7.5), 100 mM KCl, 5 mM ammonium acetate, 0.5 mM calcium acetate, 0.1 mM EDTA, 10 mM 2-mercaptoethanol, 5 mM putrescine, 1 mM spermidine, and 5 mM magnesium acetate) using fMet-$tRNA^{fMet}$ and Phe-$tRNA^{Phe}$, which was labeled with a Cy3-succinimidyl ester at the primary amine-containing 3-(3-amino-3-carboxypropyl)-uridine at position 47 of $tRNA^{Phe}$. The mRNA message was *in vitro* transcribed from a construct derived from the gene encoding gene product 32 from the T4 bacteriophage, such that the gene to be translated was AUG-UUU-UAA. Assembled complexes

were then purified using sucrose density gradient ultracentrifugation; the mRNA message was hybridized to a biotinylated DNA oligo in order to tether it to the surface of a microscope slide.

5.3 TIRF Microscopy of Cy5-mutRF1 Binding to Cy3-RC

As described previously, Cy3-RC was immobilized on biotinylated-quartz slides using a biotin-streptavidin-biotin bridge.[257] Briefly, prior to imaging, immobilized Cy3-RC samples were incubated with 10 nM Cy5-mutRF1 in tris-polymix buffer supplemented with 15 mM magnesium acetate, 1% (w/v) β-D-glucose, 300 mg/mL glucose oxidase (Sigma-Aldrich), 40 mg/mL catalase (Sigma-Aldrich), 1mM 1,3,5,7-cyclooctatetraene (Sigma-Aldrich), and 1 mM *p*-nitrobenzyl alcohol (Fluka). When illuminating samples with a 532 nm laser (Gem532; Laser Quantum), fluorescence was collected from through a 60x objective (PlanApo; Nikon) with a prism-based total-internal reflection fluorescence (TIRF) microscope (Ti-U; Nikon). Fluorescence intensity was imaged through a wavelength splitter (DV2; Photometrics) onto an electron-multiplying charge-coupled-device camera (iXon3 897; Andor) at 10 Hz.

Plot	**Example**	**Description**
Molecule Count	Fig. S1A	For each color channel (here a two-color movie), we define $$Counts(t) = \sum_{x,y} \delta_{xy}(t)$$ $$\equiv \sum_{x,y} H(p((x,y) \in \{molecules\}, t) - \gamma)$$ where H is the Heaviside step function, γ is the

		probability threshold of significance set by the user during the molecule search, and t is the frame number. This metric tracks the number of molecules and should, in most experiments involving light-emitting chromophores, decay to zero with a characteristic lifetime that roughly determines the optimal length of recording.
Signal-to-Background Ratio	Fig. S1B	For each color channel, we calculate $$SBR(t) = \frac{Mean\left(\delta_{xy}(t)d_{xy}(t)\right)}{\sqrt{Var((1-\delta_{xy}(t))d_{xy}(t))}}$$ where $d_{xy}(t)$ denotes the intensity of a pixel in frame t. This metric is useful for tracking global changes in the movie, such as the bleaching of a background chromophore, as well as for evaluating the general reliability of the identified chromophores.
Intensity Autocorrelation	Fig. S1C	For each color channel, we calculate the intensity auto-correlation function as $$ACF(t,c) = \frac{1}{K}\sum_k \frac{\langle I_k^c(t) I_k^c(t-t')\rangle}{\langle I_k^c(0)^2\rangle}$$ $$= \frac{\sum_k \int e^{i\omega t}\left(\int e^{-i\omega t'} I_k^c(t)dt'\left(\int e^{-i\omega t'} I_k^c(t)dt'\right)^*\right)d\omega}{\sum_k \int\int e^{-i\omega t'} I_k^c(t')dt'\left(\int e^{-i\omega t'} I_k^c(t')dt'\right)^* d\omega}$$ where $I_k^c(t)$ is the de-meaned intensity vs time trajectory

		of molecule k in color channel c, and t' is the lag time. We calculate this using a fast Fourier transform. This quantity is, for a stationary family of time series, equivalent to the ensemble autocorrelation function, and thus reports on the rate constants involved in the equilibrium governing the system.
Intensity Cross-Correlation	Fig. S1D	For each pair of color channels, we calculate the intensity cross-correlation function as, $$CCF(t,c,c') = \frac{1}{K}\sum_k \frac{\langle I_k^c(t) I_k^{c'}(t-t')\rangle}{\langle I_k^c(0) I_k^{c'}(0)\rangle}$$ $$= \frac{\sum_k \int e^{i\omega t}\left(\int e^{-i\omega t'}\Delta I_k^c(t)dt'\left(\int e^{-i\omega t'}\Delta I_k^{c'}(t)dt'\right)^*\right)d\omega}{\sum_k \int\int e^{-i\omega t'}\Delta I_k^c(t')dt'\left(\int e^{-i\omega t'}\Delta I_k^{c'}(t')dt'\right)^* d\omega}$$ where $\Delta I_k^c(t)$ is the derivative of the intensity vs time trajectory of molecule k as measured in color channel c. This quantity is the ensemble cross-correlation function between the intensity in color c and the intensity in color c'. This metric reports whether the ensemble of intensity vs. time trajectories possesses correlation between the pair of color channels. In smFRET experiments, this quantity should be negative.
Spatial Coincidence	Fig. S2A	We define R_{xy}^c as a function that is 1 if there is a molecule in the registration map (i.e. all chromophore

Appendix A

Counts		locations in color channel c transformed to the reference color channel coordinates) at point (x,y), and 0 otherwise. For each color channel, these plots show $\sum_{x,y} R_{xy}^{c}$. Colocalization counts between channels c and c' are then defined as $$Counts(t) = \sum_{x,y} H\left(\sum_{(x,y)\in O(L(i))} R_{xy}^{c} R_{(x,y)\in O(L(i))}^{c'} \right)$$ where $O(L(i))$ is the pixel neighborhood of a chromophore located at $L(i)$.
Spatial Coincidence Probability	Fig. S2B	Defining $$z^{c} = \frac{\sum_{x,y} R_{xy}^{c} \left(Card\left(O\left(L(i) \right) \right) \right)}{Card(frame)}$$ we calculate the posterior probability of co-localization as $$p(s) = \beta(s|a,b)$$ $$a \equiv z^{c} z^{c'}$$ $$b \equiv \sum_{x,y} H\left(\sum_{(x,y)\in O(L(i))} R_{xy}^{c} R_{(x,y)\in O(L(i))}^{c'} \right) - z^{c} z^{c'}$$ where $\beta(s|a,b)$ is the beta distribution with support s.
Co-localization Probability	Fig. S2C	We calculate the posterior probability of co-localization as $$p(s) = \beta(s|a,b)$$

Appendix A

		$a \equiv \sum_{x,y} H\left(\sum_{(x,y)\in O(L(i))} R^{c}_{xy} R^{c'}_{(x,y)\in O(L(i))}\right)$ $b \equiv \sum_{x,y} H\left(\sum_{(x,y)\in O(L(i))} R^{c}_{xy}\left(1 - R^{c'}_{(x,y)\in O(L(i))}\right)\right)$
Illumination Profile	Fig. S3A	The illumination profile shown is defined by $M_{xy} = \min_{O((x,y))} d_{xy}$
Registration	Fig. S3B	The registration profile is a series of X's drawn in the size of the channel color and then transformed according to the registration function into the principal color.

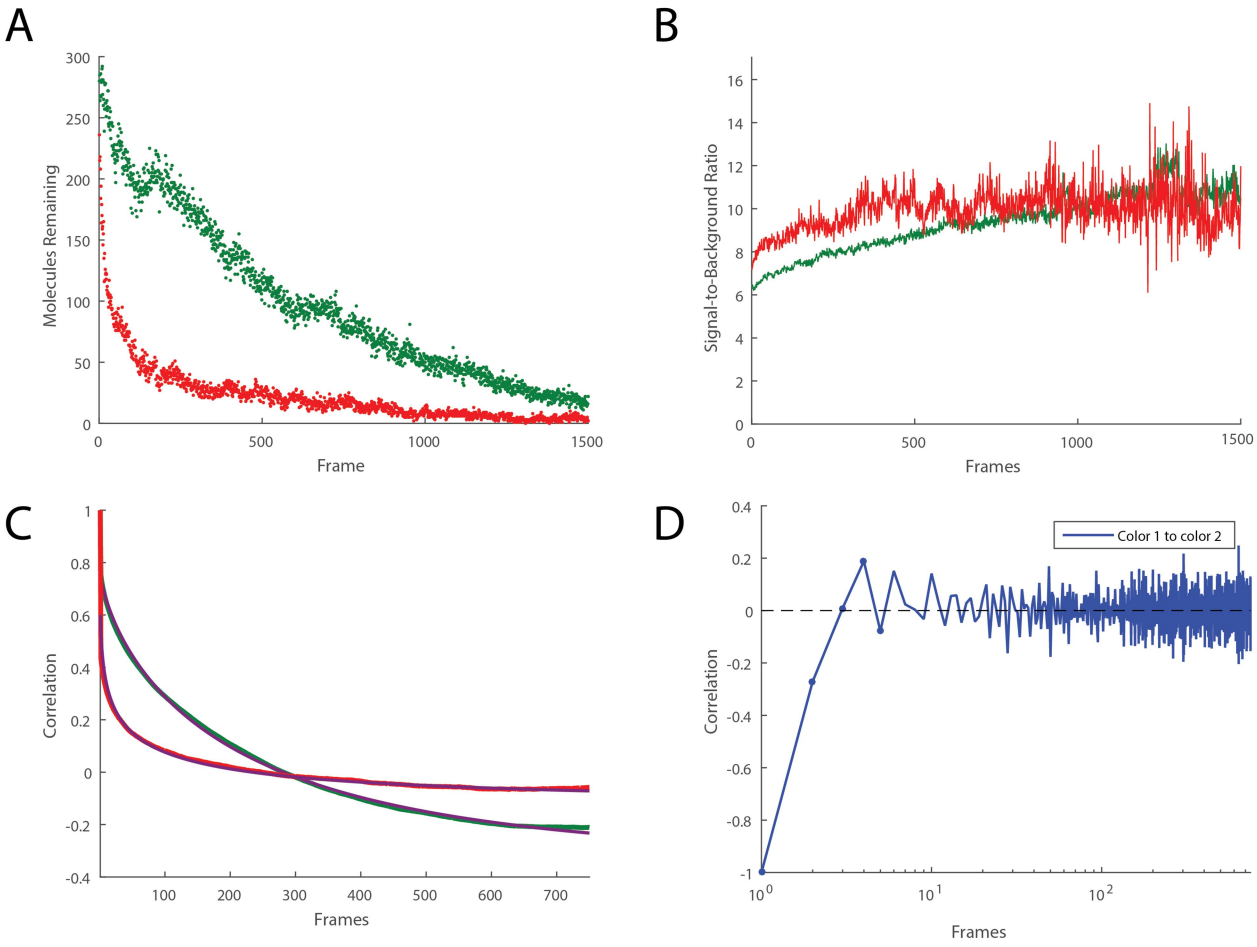


Figure A. 1 Plots of vbscope intensity analysis for TIRF microscopy movie of 10 nM Cy5-mutRF1 binding to surface-immobilized Cy3-RC.
(A) Plot of number of molecules identified in each image of the movie for the Cy3 (green), and Cy5 (red) color channels. (B) Plot of average signal-to-background ratio for molecules identified in each image of the movie for the two color channels. (C) Plot of the intensity autocorrelation function for the molecules identified in the movie for the two color channels. (D) Plot of the intensity cross-correlation between color channel 1 (Cy3) and color channel 2 (Cy5) for all identified molecules. The value at the first frame is caused by the extreme anticorrelation of two-color FRET.

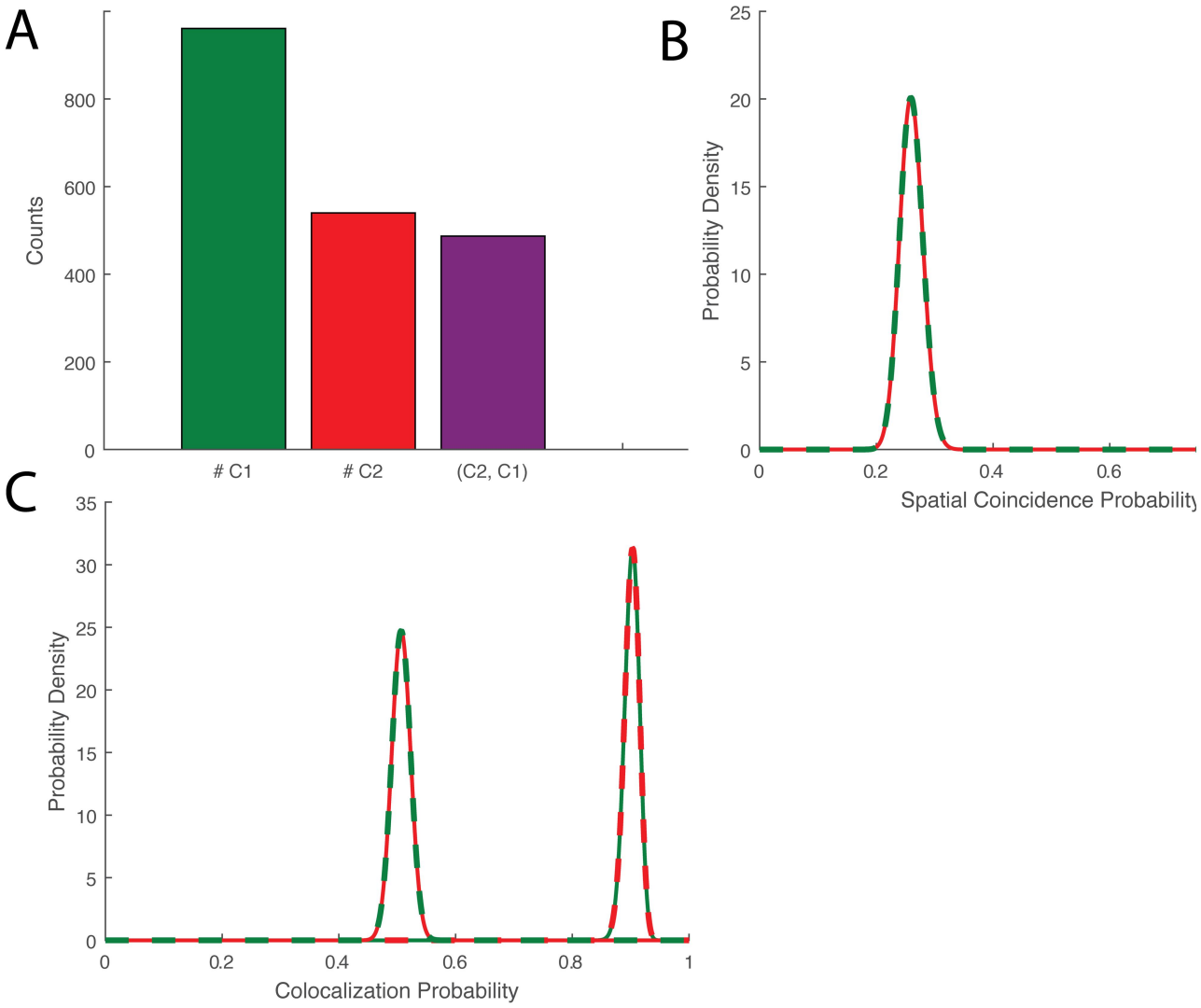


Figure A. 2 Plots of vbscope colocalization analysis for TIRF microscopy movie of 10 nM Cy5-mutRF1 binding to surface-immobilized Cy3-RC.
(A) Plot of number of distinct, identified molecules in color channel 1 (C1; Cy3) and color channel 2 (C2; Cy5), and the number of these molecules with are colocalized to the same pixel ((C2,C1)). (B) Plot of the posterior probability distribution of the probability that an independent chromophore identified in one color channel coincidently overlaps with a separate, non-associated, independent chromophore identified in the other color channel. (C) Plot of the

posterior probability distribution of the probability that a molecule identified in C1 colocalizes with a molecule identified in C2 (thick green, thin red), and that a molecule identified in C2 colocalizes with a molecule identified in C1 (thick red, thin green). Comparing these curves to the plot in panel (B) shows that the observed colocalization is not random.

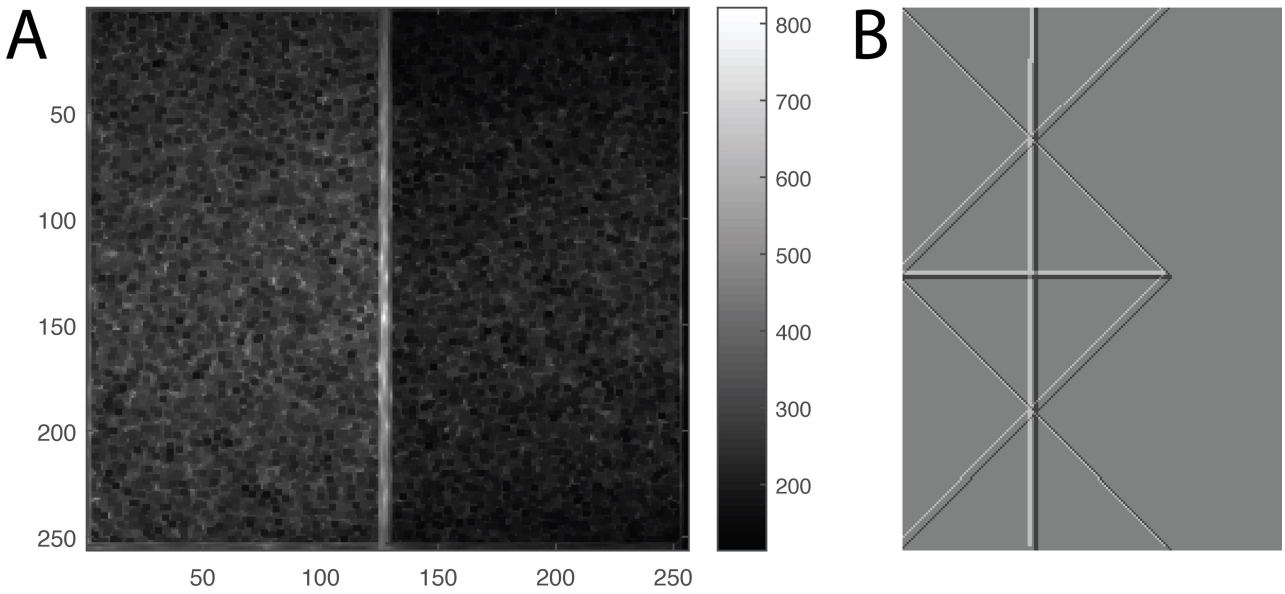


Figure A. 3 Plots of microscope parameters as analyzed by vbscope determined from a TIRF microscopy movie of 10 nM Cy5-mutRF1 binding to surface-immobilized Cy3-RC. (A) Image of the estimated illumination profile of the movie. (B) Difference image of the registration function used to align the two color channels. The color channel on the right shows no difference, because it is the reference channel. The specific line patterns were chosen to show the registration function across the entire color channel.

Appendix A

Supporting information for A Bayesian Approach to Hierarchical Hidden Markov Modeling Allows Direct Measurement of Conditional Kinetic Rates and for A Bayesian Approach to Single-Molecule Trajectories with Diffusing Observables

S1 Generative Model for Hierarchical Hidden Markov Models (HHMMs)

S1.1 Overview

In this section we will first define all the variables used to describe the algorithms for static and dynamic heterogeneity. Next, we show how these variables are organized to optimize the evidence – the probability that the parameters, state occupancies, and observations are all found and inferred in the same dataset. To explain this quantity, we will begin with a formal definition of the evidence, then follow by defining the emissions model used herein, describe the prior distributions, then close with general outlines of the two algorithms.

S1.2 Variable Definitions

Observations of a trajectory $n \in \{1, \dots, N\}$ at time $t \in \{1, \dots, T_n\}$ x_{nt}

State of the molecule in trajectory $n \in \{1, \dots, N\}$ at time $t \in \{1, \dots, T_n\}$. The model for dynamic heterogeneity has $d \in \{1, \dots, D\}$ and the model for static heterogeneity has $d \in \{1,2\}$. z_{nt}^{d}

Ω_d

Appendix A

Size of the state space at level $d \in \{1, \dots, D\}$	
Accessible state space at level $d \in \{1, \dots, D\}$	$\widetilde{\Omega}_d \equiv \frac{\Omega_d}{\Omega_{d-1}}$
Collectively, parameters for a population of trajectories	θ
Collectively, the parameters for the emission distribution	ϕ
Emission distribution for a given production state $i \in \{1, \dots, \widetilde{\Omega}_D\}$	ϕ_i
Emission Normal distribution mean for a given production state $i \in \{1, \dots, \widetilde{\Omega}_D\}$	μ_i
Emission Normal distribution precision for a given production state $i \in \{1, \dots, \widetilde{\Omega}_D\}$	λ_i
Variational estimate for the mean μ_i of the Emission Normal distribution, $i \in \{1, \dots, \widetilde{\Omega}_D\}$	m_i
Variational estimate for the precision of the mean μ_i of the Emission Normal distribution, $i \in \{1, \dots, \widetilde{\Omega}_D\}$	β_i
Variational estimate for the scale of the precision λ_i of the Emission Normal distribution, $i \in \{1, \dots, \widetilde{\Omega}_D\}$	a_i

Appendix A

Variational estimate for the rate of the precision λ_i of the Emission Normal distribution, $i \in \{1, \dots, \widetilde{\Omega}_D\}$	b_i
Initial-state probabilities $d \in \{1, \dots, D\}, i \in \{1, \dots, \widetilde{\Omega}_d\}, k \in \{1, \dots, \Omega_{d-1}\}$	$\pi_i^d(k)$
Transition matrices $d \in \{1, \dots, D\}, i \in \{1, \dots, \widetilde{\Omega}_d\}, j \in \{1, \dots, \widetilde{\Omega}_d + 1\}, k \in \{1, \dots, \Omega_{d-1}\}$	$A_{ij}^d(k)$
Probability of transitioning between branches of the tree at level d	$A_{i,\widetilde{\Omega}_d+1}^d(k)$
Variational estimate for the number of time a trajectory is first observed in state $d \in \{1, \dots, D\}, i \in \{1, \dots, \widetilde{\Omega}_d\}, k \in \{1, \dots, \Omega_{d-1}\}$	$\rho_i^d(k)$
Variational estimate for the number of times a trajectory makes a transition between $i \in \{1, \dots, \widetilde{\Omega}_d\}$ and $j \in \{1, \dots, \widetilde{\Omega}_d + 1\}$ at level $d \in \{1, \dots, D\}$ positioned at the path $k \in \{1, \dots, \Omega_{d-1}\}$	$\alpha_{ij}^d(k)$
Collectively, hyperparameters for the prior distribution	ψ_0
Prior estimate for the mean μ_i of the Emission Normal distribution, $i \in \{1, \dots, \widetilde{\Omega}_D\}$	$m_{0,i}$
Prior estimate for the precision of the mean μ_i of the Emission Normal distribution, $i \in \{1, \dots, \widetilde{\Omega}_D\}$	$\beta_{0,i}$
Prior estimate for the scale of the precision λ_i of the Emission Normal distribution, $i \in \{1, \dots, \widetilde{\Omega}_D\}$	$a_{0,i}$

Appendix A

Prior estimate for the rate of the precision λ_i of the Emission Normal distribution, $i \in \{1, \dots, \widetilde{\Omega}_D\}$	$b_{0,i}$
Prior estimate for the number of time a trajectory is first observed in state $d \in \{1, \dots, D\}, i \in \{1, \dots, \widetilde{\Omega}_d\}, k \in \{1, \dots, \Omega_{d-1}\}$	$\rho_{0,i}^{d}(k)$
Prior estimate for the number of times a trajectory makes a transition between $i \in \{1, \dots, \widetilde{\Omega}_d\}$ and $j \in \{1, \dots, \widetilde{\Omega}_d + 1\}$ at level $d \in \{1, \dots, D\}$ positioned at the path $k \in \{1, \dots, \Omega_{d-1}\}$	$\alpha_{0,ij}^{d}(k)$
Evidence	$L\left(q(\{z_{nt}^{d}\}), q(\theta)\right)$
Expected occupancy of the production state $i \in \{1, \dots, \widetilde{\Omega}_D\}$ of a molecule	γ_{nt}^{i}
Expected counts of the number of transitions between $i \in \{1, \dots, \widetilde{\Omega}_d\}$ and $j \in \{1, \dots, \widetilde{\Omega}_d + 1\}$ at level $d \in \{1, \dots, D\}$ positioned at the path $k \in \{1, \dots, \Omega_{d-1}\}$ in trajectory $n \in \{1, \dots, N\}$ at time $t \in \{2, \dots, T_n\}$	$\xi_{ntij}^{d}(k)$
Expected counts of the number of time a trajectory begins in state $i \in \{1, \dots, \widetilde{\Omega}_d\}$ at level $d \in \{1, \dots, D\}$	g_{ni}^{d}
Forward-backward scale variable	$c_{nt}(k)$
Forward variable	$\hat{\alpha}_{nti}^{d}(k)$
Backward variable	$\hat{\beta}_{nti}^{d}(k)$
Forward-upward variable	$\hat{\alpha}_{b_{n,t}^{d}}^{i}(k)$
Forward-downward variable	$\hat{\alpha}_{e_{n,t}^{d}}^{i}(k)$

Appendix A

Backward-upward variable	$\hat{\beta}^{i}_{b^{d}_{n,t}}(k)$
Backward-downward variable	$\hat{\beta}^{i}_{e^{d}_{n,t}}(k)$
The set of nodes in the state space graph that point at x	$par(x)$
The k^{th} super-parent of x	$par_k(x)$
The set of nodes in the state space graph that x points at	$ch(x)$
The set of nodes in the state space graph that share nodes that point to x	$sib(x)$
Gamma function	$\Gamma(z) = \int_0^{\infty} x^{z-1}e^{-x}dx$
Digamma function	$\psi(z) = \frac{d\left[\ln\left(\Gamma(z)\right)\right]}{dz}$

S1.3 Evidence

The evidence is the probability the current set of observations was obtained from an experiment given any possible set of parameters, given as well some prior data. Specifically, using Bayesian inference, we seek parameter distributions that optimize the evidence, given by:

$$p(\{x_{nt}\}, \theta|\psi_0) = \int p(\{x_{nt}\}|\theta)p(\theta|\psi_0)\, d\theta$$

This value is, in the case of the present model, analytically intractable, so we instead seek parameter distributions that maximize a lower bound for the evidence, given by[76]:

$$L\left(q\left(\{z^{d}_{nt}\}\right), q(\theta)\right) = \int d\theta \sum_{n}\sum_{z^{d}_{nt}} q(\theta)q(z^{d}_{nt}) \ln\frac{p(x_{nt}, z^{d}_{nt}, \theta|\psi_0)}{q\left(z^{d}_{nt}\right)q(\theta)}$$

This sum runs over all possible values of the possible states of the molecule trajectory. This expression assumes that the joint probability may be factorized:

$$p\left(z_{nt}^{d}, \theta \middle| x_{nt}, \psi_0\right) = q\left(z_{nt}^{d}\right) q(\theta)$$

and this assumption is the basis of the variational approximation.

S1.4 Emissions Model

The emissions model is the probability that an observation was obtained from an experiment given a particular set of parameters as well as a particular production state of the molecule. This is given by:

$$p(\{x_{nt}\}|z_{nt}^{D}, \theta) = \prod_{n=1}^{N} \prod_{t=1}^{T_n} p\left(x_{nt} \middle| \phi_{z_{nt}^{D}}\right)$$

Furthermore we assume that $p\left(x_{nt} \middle| \phi_{z_{nt}^{D}}\right)$ follows a Normal distribution:

$$p\left(x_{nt} \middle| \phi_{z_{nt}^{D}}\right) = \left(\frac{\lambda_{z_{nt}^{D}}}{\pi}\right)^{\frac{1}{2}} e^{-\frac{\lambda_{z_{nt}^{D}}}{2}\left(x_{nt} - \mu_{z_{nt}^{D}}\right)^2}$$

This may in general be modified to any appropriate distribution.

S1.5 Prior Distributions

Prior information using the variational approximation and conjugate exponential distributions allows us to write down the form of the prior distributions that maximizes their informative character and minimizes the information provided by observations[101]. These are given by:

$\mu_i, i \in \{1, \ldots, \widetilde{\Omega}_D\}$

$$p(\mu_i|\psi_0) = p\left(\mu_i | m_{0,i}, \beta_{0,i}\right) = \left(\frac{\beta_{0,i}}{2}\right)^{\frac{1}{2}} e^{-\frac{\beta_{0,i}}{2}\left(\mu_i - m_{0,i}\right)^2}$$

Appendix A

$\lambda_i, i \in \{1, \dots, \widetilde{\Omega}_D\}$

$$p(\lambda_i|\psi_0) = p(\lambda_i|a_{0,i}, b_{0,i}) = \frac{{b_{0,i}}^{a_{0,i}}}{\Gamma(a_{0,i})} {\lambda_i}^{a_{0,i}-1} e^{-b_{0,i}\lambda_i}$$

$\{\pi_i^d(k)\}, i \in \{1, \dots, \widetilde{\Omega}_d\}$

$k \in \{1, \dots, \Omega_{d-1}\} d$

$\in \{1, \dots, D\}$

$$p(\{\pi_i^d\}(k)|\psi_0) = p\left(\{\pi_i^d\}(k)\middle|\{\rho_{0,i}^d\}(k)\right)$$

$$= \frac{\Gamma\left(\sum_{j=1}^{\widetilde{\Omega}_d} \rho_{0,j}^d(k)\right)}{\prod_{j=1}^{\widetilde{\Omega}_d} \Gamma\left(\rho_{0,j}^d(k)\right)} \prod_{j=1}^{\widetilde{\Omega}_d} \left(\pi_j^d(k)\right)^{\rho_{0,j}^d(k)-1}$$

$A_{ij}^d(k), i \in \{1, \dots, \widetilde{\Omega}_d\}$

$j \in \{1, \dots, \widetilde{\Omega}_d + 1\}$

$k \in \{1, \dots, \Omega_{d-1}\}$

$d \in \{1, \dots, D\}$

$$p(\{A_{ij}^d\}(k)|\psi_0, i) = p(\{A_{ij}^d\}(k)|\{\alpha_{0,ij}^d\}(k), i)$$

$$= \frac{\Gamma\left(\sum_{j=1}^{\widetilde{\Omega}_d+1} \alpha_{0,ij}^d(k)\right)}{\prod_{j=1}^{\widetilde{\Omega}_d+1} \Gamma\left(\alpha_{0,ij}^d(k)\right)} \prod_{j=1}^{\widetilde{\Omega}_d+1} \left(A_{ji}^d(k)\right)^{\alpha_{0,ij}^d(k)-1}$$

S1.6 Algorithm

Iterate an Expectation – E – and a Maximization – M – step until the change in the value of the evidence changes negligibly between consecutive iterations.

The E-step determines the expected state of the molecule in each trajectory at each time point and also calculates the value of the likelihood function. The M-step takes the expected state occupancies of the molecule calculated in the E-step and uses these occupancies to re-estimate all of the parameters. We do not re-estimate the prior distributions and thus it is important that they not overwhelm the contributions from observations in the M-step.

Appendix A

S2 Variational Bayes Expectation Maximization (VBEM)

S2.1 E-Step

The E-Step estimates the likelihood function while concurrently using the current estimates for parameters, θ, to decode the state of each molecule n at time t, z_{nt}^{d}. This is done in conceptually distinct ways for static and for dynamic heterogeneity. In the static case, we present an algorithm which may be called a mixture of HMMs, and determine mixture coefficients that describe the degree to which each trajectory belongs to each subpopulation. In the dynamic case, we present an algorithm, first described by Wakabayashi *et al*[147], which estimates the contribution of every potential conditional transition by determining the degree to which it has been activated. As this degree at level d depends on whether $d+1$ has activated a transition, this algorithm takes on a vertical character at each time point. The goal of each algorithm will be to return occupancies γ_{nt}^{i} and counts $\xi_{ntij}^{d}(k)$ with which we will re-estimate all parameters in the M-Step, below. The E-Step is, in both cases, the most time-consuming step of the algorithm.

S2.1.1 Forward-Backward Mixture Algorithm – Static Heterogeneity

The Forward-Backward Mixture (FBM) algorithm, which is adapted from standard references[76], sequentially decodes the optimal expected state occupancies without considering the entirety of the exponentially-scaled state space by iteratively decoding an optimal occupancy at time t in trajectory n and using that information to decode an optimal occupancy at time $t+1$, finding the corresponding conditional through a similar process in reverse, and combining them using Bayes' theorem to acquire the desired parameters. Therefore the complexity of the

algorithm is simply $O(NTK^2D)$ where D is the number of static subpopulations and K is the number of production states.

The forward and backward variables reduce in the d dimension because the populations that do not interconvert do not have subsequent transitions at any higher level:

$$\hat{\alpha}^d_{nti}(k) = \hat{\alpha}_{nti}(k) \equiv p(x_{nt}|\phi_i)\sum_j \hat{\alpha}_{n,t-1,j}(k)A_{ji}(k)$$

$$\hat{\beta}^d_{nti}(k) = \hat{\beta}_{nti}(k) \equiv \sum_j \hat{\beta}_{n,t+1,j}(k)p\big(x_{n,t+1}|\phi_j\big)A_{ij}(k)$$

The boundary conditions are as follows:

$$\hat{\alpha}_{n1i}(k) = p(x_{nt}|\phi_i)\pi_i(k)$$

$$\hat{\beta}_{nT_ni}(k) = p\big(x_{nT_n}\big|\phi_i\big)$$

where one may notice that we have removed an unnecessary index from $\pi^d_i(k)$. These variables are normalized to supply the likelihood function as well as a convenient scale and obvious recursion, all to guarantee computational precision:

$$c_{nt}(k) \equiv \sum_j \hat{\alpha}_{ntj}(k)A_{ji}(k)$$

$$\widehat{\boldsymbol{\alpha}}_{nti}(k) \equiv \hat{\alpha}_{nti}(k)\prod_{t'=1}^{t} c^{-1}_{nt'}(k)$$

$$\widehat{\boldsymbol{\beta}}_{nti}(k) \equiv \hat{\beta}_{nti}(k)\prod_{t'=1}^{t} c^{-1}_{nt'}(k)$$

To complete the algorithm we use the above variables to calculate the variables of primary interest:

$$p(\{x_{nt}\}) = q\big(z^d_{nt}\big) = \prod_{n,k,t} c_{nt}(k)$$

$$\gamma_{nt}^{i} = \sum_{k} \widehat{\boldsymbol{\alpha}}_{nti}(k)\widehat{\boldsymbol{\beta}}_{nti}(k)\ \frac{\prod_t c_{nt}(k)}{\sum_k \prod_t c_{nt}(k)}$$

$$g_{ni} = \widehat{\boldsymbol{\alpha}}_{n1i}(k)\widehat{\boldsymbol{\beta}}_{n1i}(k)\frac{\prod_t c_{nt}(k)}{\sum_k \prod_t c_{nt}(k)}$$

$$\xi_{ntij}(k) = \frac{c_{nt}(k)p(x_{nt}|\phi_i)\widehat{\boldsymbol{\alpha}}_{n,t-1,j}(k)\widehat{\boldsymbol{\beta}}_{nti}(k)A_{ji}(k)}{\prod_t c_{nt}(k)}$$

At this point all variables required for the M-Step have been prepared.

S2.1.2 Forward-Backward Activation Algorithm – Dynamic Heterogeneity

The Forward-Backward Activation (FBA) algorithm, first described by Wakabayashi *et al*[147], sequentially decodes the optimal expected state occupancies at every level of the state space while concurrently counting the number of transitions between each state space branch. While the skeleton of the algorithm bears similarities to the FBM algorithm, the requirement of determining when indirectly observed underlying conditions are in force adds significant complexity – for example, if each level of the tree has K children, then the algorithm is $O(NTK^D)$.

The forward-upward and forward-downward variables are calculated according to the following recursion, beginning at the top of the tree:

$$\hat{\alpha}_{b_{nt}^1}^{i}(1) = \sum_{j} \hat{\alpha}_{e_{nt-1}^1}^{j}(1)A_{ji}^{1}(1)$$

As with the FBM algorithm, we define a scale factor which will eventually be used to calculate the likelihood as well as to keep the entirety within computational precision.

Appendix A

$$\hat{\alpha}^{i}_{e^{D}_{nt}}(k) = \hat{\alpha}^{i}_{b^{D}_{nt}} p(x_{nt}|\phi_i)$$

$$c_{nt} = \sum_{k\in\Omega_{D-1}} \sum_{i\in ch(k)} \hat{\alpha}^{i}_{e^{D}_{nt}}(k)$$

Using these, we continue the recursion:

$$\hat{\boldsymbol{\alpha}}^{i}_{e^{D}_{nt}}(k)$$

$$= \hat{\alpha}^{i}_{e^{D}_{nt}}(k) \prod_{t'=1}^{t} c^{-1}_{nt'}$$

$$\hat{\alpha}^{i}_{b^{d}_{nt}}(k) = \hat{\alpha}^{k}_{b^{d-1}_{nt}}\big(par(k)\big)\pi^{d}_{i}(k) + \sum_{j\in ch(k)} \hat{\alpha}^{j}_{e^{d}_{nt-1}}(k) A^{d}_{ji}(k)\,, d \in \{2,\dots,D\} \hat{\alpha}^{i}_{b^{D}_{nt}}(k)$$

$$= \hat{\alpha}^{k}_{b^{D-1}_{nt}}\big(par(k)\big)\pi^{D}_{i}(k) + \sum_{j\in ch(k)} \hat{\boldsymbol{\alpha}}^{j}_{e^{D}_{nt-1}}(k) A^{D}_{ji}(k)$$

$$\hat{\alpha}^{i}_{e^{d}_{nt}}(k)$$

$$= \sum_{j\in ch(i)} \hat{\alpha}^{j}_{e^{d+1}_{nt}}(i) A^{d+1}_{j,\tilde{\Omega}_{d}+1}\,, d \in \{1,\dots,D-1\}$$

The forward-upward variables have time-boundary conditions:

$$\hat{\alpha}^{i}_{b^{1}_{n1}}(k) = \pi^{1}_{i}$$

$$\hat{\alpha}^{i}_{b^{d}_{n1}}(k) = \hat{\alpha}^{k}_{b^{d-1}_{n1}}\big(par(k)\big)\pi^{d}_{i}, d \in \{2,\dots,D\}$$

Similarly, the backward-upward and backward-downward variables are calculated according to the following recursion:

Appendix A

$$\hat{\beta}^i_{e^1_{nt}}(1) = \sum_j \hat{\beta}^j_{b^1_{nt+1}}(1) A^1_{ij}(1)$$

$$\hat{\beta}^i_{b^D_{nt}}(k) = \hat{\beta}^i_{e^D_{nt}} p(x_{nt}|\phi_i) \prod_{t'=1}^{t} c_{nt}^{-1}$$

$$\hat{\beta}^i_{e^d_{nt}}(k) = \hat{\beta}^k_{e^{d-1}_{nt}}\big(par(k)\big) A^d_{i,\widetilde{\Omega}_d+1}(k) + \sum_{j \in ch(k)} \hat{\beta}^j_{b^d_{nt+1}}(k) A^d_{ij}(k)\,,\, d \in \{2, \dots, D\}$$

$$\hat{\beta}^i_{b^d_{nt}}(k) = \sum_{j \in ch(i)} \hat{\beta}^j_{b^{d+1}_{nt}}(i) \pi_j^{d+1}\,, d \in \{1, \dots, D-1\}$$

One will note that we have introduced the scale alongside the backward-downward variables. The backward-downward variables have time-boundary conditions:

$$\hat{\beta}^i_{e^1_{nT_n}}(k) = A^1_{i,\widetilde{\Omega}_1+1}$$

$$\hat{\beta}^i_{e^d_{n1}}(k) = \hat{\beta}^k_{b^{d-1}_{n1}}\big(par(k)\big) A^d_{i,\widetilde{\Omega}_d+1}, d \in \{2, \dots, D\}$$

Finally, we need to prepare the variables needed for the M-Step, as well as calculate the likelihood function. This is done by setting:

$$g^d_{ni} = \hat{\alpha}^i_{b^d_{n1}}(k) \hat{\beta}^i_{b^d_{n1}}(k) + \sum_{t=1}^{T_n-1} \hat{\alpha}^k_{b^{d-1}_{n,t+1}}\big(par(k)\big) \pi_i^d \hat{\beta}^i_{b^d_{n,t+1}}(k)$$

$$\gamma^i_{nt} = \sum_{k \in \Omega_{D-1}} \hat{\alpha}^i_{e^D_{nt}}(k) \hat{\beta}^i_{e^D_{nt}}(k)$$

$$\xi^d_{ntij}(k) = \sum_{t=1}^{T_n-1} \hat{\alpha}^i_{e^d_{n,t}}(k) A^d_{ij} \hat{\beta}^j_{b^d_{n,t+1}}(k)$$

$$\xi^d_{nti,\widetilde{\Omega}_d+1} = \hat{\alpha}^i_{e^d_{nT_n}}(k) \hat{\beta}^i_{b^d_{nT_n}}(k) + \sum_{t=1}^{T_n-1} \hat{\alpha}^i_{e^d_{n,t}}(k) A^d_{i,\widetilde{\Omega}_d+1} \hat{\beta}^k_{e^d_{n,t+1}}\big(par(k)\big)$$

$$p(\{x_{nt}\}) = \prod_{n,t} c_{nt}$$

Appendix A

S2.2 M-Step

Parameters are updated in the M-Step. This is done simultaneously and as such, all parameters on the right hand side of the equations belong to the previous iteration and those on the left hand side belong to the current iteration. Priors are not updated in this model as we do not utilize the Empirical Bayes' framework. This step is iterated as necessary with the E-Step above.

$$\beta_i = \beta_{0,i} + \sum_{n,t} \gamma_{nt}^i \, , i \in \{1, \dots, \widetilde{\Omega}_D\}$$

$$a_i = a_{0,i} + \frac{1}{2} \sum_{n,t} \gamma_{nt}^i \, , i \in \{1, \dots, \widetilde{\Omega}_D\}$$

$$b_i = b_{0,i} + \frac{1}{2} \left(\beta_{0,i} m_{0,i}^2 + \sum_{n,t} x_{nt}^2 \gamma_{nt}^i - \frac{\left(\beta_{o,i} m_{0,i} + \sum_{n,t} x_{nt} \gamma_{nt}^i\right)^2}{\beta_i} \right), i \in \{1, \dots, \widetilde{\Omega}_D\}$$

$$\lambda_i = \frac{a_i}{b_i}, i \in \{1, \dots, \widetilde{\Omega}_D\}$$

$$m_i = \frac{\lambda_i}{\beta_i} \left(\sum_{n,t} x_{nt} \gamma_{nt}^i + m_{0,i} \beta_{0,i} \right) , i \in \{1, \dots, \widetilde{\Omega}_D\}$$

$$\mu_i = m_i, i \in \{1, \dots, \widetilde{\Omega}_D\}$$

$$\rho_i^d(k) = \rho_{0,i}^d(k) + \sum_n g_{ni}^d \, , \; d \in \{1, \dots, D\}, i \in \{1, \dots, \widetilde{\Omega}_d\}, k \in \{1, \dots, \Omega_{d-1}\}$$

Appendix A

$$\pi_i^d(k) = e^{\psi\left(\rho_i^d\right) - \sum_i \psi\left(\rho_i^d\right)} \;,\; d \in \{1, \dots, D\}, i \in \left\{1, \dots, \widetilde{\Omega}_d\right\}, k \in \{1, \dots, \Omega_{d-1}\}$$

$$\alpha_{ij}^d(k) = \alpha_{0,ij}^d(k) + \sum_{n,t} \xi_{ntij}^d(k)\,, i \in \left\{1, \dots, \widetilde{\Omega}_d\right\}, j \in \left\{1, \dots, \widetilde{\Omega}_d + 1\right\}, k \in \{1, \dots, \Omega_{d-1}\},$$

$d \in \{1, \dots, D\}$

$$A_{ij}^d(k) = e^{\psi\left(\rho_i^d\right) - \sum_i \psi\left(A_{ij}^d(k)\right)}, i \in \left\{1, \dots, \widetilde{\Omega}_d\right\}, j \in \left\{1, \dots, \widetilde{\Omega}_d + 1\right\}, k \in \{1, \dots, \Omega_{d-1}\},$$

$d \in \{1, \dots, D\}$

S2.3 Calculation of Evidence Lower Bound

The evidence lower bound is given by:

$$L\left(q\left(\{z_{nt}^d\}\right), q(\theta)\right) = p(\{x_{nt}\}) - D_{KL}(\phi||\psi_0) - D_{KL}\left(\{\rho_i^d\}||\psi_0\right) - D_{KL}\left(\{\alpha_{ij}^d(k)\}||\psi_0\right),$$

$$D_{KL}(\phi||\psi_0) = \left(\left[(a_i - 1)\psi(a_i) + \log\left(\frac{a_i}{b_i}\right) - a_i + \log\left(\frac{\Gamma\left(a_{0,i}\right)}{\Gamma(a_i)}\right) + a_{0,i}\log\left(b_{0,i}\right) - \left(a_{0,i} - 1\right)(\psi(a_i) + \log(b_i)) + \frac{a_i b_i}{b_{0,i}}\right] + \left[\log\left(\frac{\beta_{0,i}}{\beta_i}\right) + \frac{\beta_i + \left(m_i - m_{0,i}\right)^2}{2\beta_{0,i}} - \frac{1}{2}\right]\right)$$

Appendix A

$$D_{KL}\left(\{\rho_i^d\}||\psi_0\right)$$

$$= \log\sum_{i=1}^{\widetilde{\Omega}_d}\Gamma\left(\rho_i^d\right) - \log\sum_{i=1}^{\widetilde{\Omega}_d}\Gamma\left(\rho_{0,i}^d\right) + \sum_{i=1}^{\widetilde{\Omega}_d}\log\Gamma\left(\rho_i^d\right) - \sum_{i=1}^{\widetilde{\Omega}_d}\log\Gamma\left(\rho_{0,i}^d\right)$$

$$+ \sum_{i=1}^{\widetilde{\Omega}_d}\left(\rho_i^d - \rho_{0,i}^d\right)\left(\psi\left(\rho_i^d\right) - \psi\left(\sum_{i=1}^{\widetilde{\Omega}_d}\rho_i^d\right)\right)$$

$$D_{KL}\left(\{\alpha_{ij}^d(k)\}||\psi_0\right)$$

$$= \sum_{j=1}^{\widetilde{\Omega}_d+1}\left(\log\sum_{i=1}^{\widetilde{\Omega}_d}\Gamma\left(\alpha_{ij}^d(k)\right) - \log\sum_{i=1}^{\widetilde{\Omega}_d}\Gamma\left(\alpha_{0,ij}^d(k)\right) + \sum_{i=1}^{\widetilde{\Omega}_d}\log\Gamma\left(\alpha_{ij}^d(k)\right)\right.$$

$$- \sum_{i=1}^{\widetilde{\Omega}_d}\log\Gamma\left(\alpha_{0,ij}^d(k)\right)$$

$$\left.+ \sum_{i=1}^{\widetilde{\Omega}_d}\left(\alpha_{ij}^d(k) - \alpha_{0,ij}^d(k)\right)\left(\psi\left(\alpha_{ij}^d(k)\right) - \psi\left(\sum_{i=1}^{\widetilde{\Omega}_d}\alpha_{ij}^d(k)\right)\right)\right)$$

S3 Calculation of Kinetic Rates

S3.1 Static Heterogeneity

Calculation of the kinetic rates follows:

$$k_{ij}^d \approx A_{ij}^d, i \neq j$$

where the rate constants are in units of time-steps.

S3.2 Dynamic Heterogeneity

Calculation of the kinetic rates follows:

Appendix A

$$d^* \equiv \min\left(d | par_d(i) = par_d(j)\right)$$

$$k_{ij}^d \approx \left[\prod_{m=1}^{d-1} \sum_j A_{j,\tilde{\Omega}_m+1}^m \left(ch_m(i)\right)\right]$$

$$\prod_{m=d}^{d^*} A_{par_{m-d}(i),\tilde{\Omega}_m+1}^m \left(par_{m-d+1}(i)\right) \pi_{par_m(j)}^m \left[A_{par_{d^*}\left(par_{d^*-1}(i)\right),par_{d^*}\left(par_{d^*-1}(j)\right)}^{d^*}\left(par_{d^*+1}(i)\right)\right]$$

$$i \neq j$$

where the rate constants are in units of time-steps.

Emission Drift

Emission drift is assumed to be distributed so that the difference between the values of any two sequential samples from a trajectory is itself normally distributed, the variance of which is ratiometrically related to the expected variance of the trajectory on the basis of, for example, the Gaussian mixture updates in the M-step described above. This ratio is defined as $R = \sigma_b / E[\sigma_t]$ where σ_b is the standard deviation of the random walk and $E[\sigma_t]$ is the expected standard deviation of the trajectory. This value is used in a E-step to find the expected values of the baseline, according to, for an individual trajectory x_t,

$$b_t = (-\Delta + R^2)^{-1}\left(x_t - \sum_k \gamma_t^k \mu_k\right) \quad \textbf{A. 15}$$

where γ_t^k are the occupancy posteriors from the HMM (see above), μ_k are the centers of the Gaussian mixture model, from the M-step, and Δ is the finite difference operator defined by:

$$\Delta n_t = n_{t+1} + n_{t-1} - 2n_t, t = 2, \ldots, T-1$$

$$\Delta n_1 = n_2 - n_1$$

Appendix A

$$\Delta n_T = n_{T-1} - n_T$$

The required inverse is calculated according to the method of Usmani[98]. To complete the model, all that is required is an estimate for R^2 as an addition to the M-step. This is given by:

$$\sigma_t = \sum_k \gamma_t^k \sigma_k$$

$$s_1 = 0, s_t = b_t - b_{t-1}$$

This latter estimate is understood to have been either initialized or to have originated in a previous EM iteration. Next, defining:

$$\lambda = \sum_t s_t \sigma_t$$

$$A = 128 - 96\lambda + 24\lambda^2 - 2\lambda^3 + 288\lambda T^2 - 36\lambda^2 T^2 + 18\lambda^3 T^2 + 108\lambda^2 T^4$$

We arrive at:

```
R^2 = ((2*(2+d))/(3*(-1+T^2))-(2^(1/3)*(-4*(2+d)^2-3*d*(8+d)*...
    (-1+T^2)))/(3*(-1+T^2)*(128-96*d+24*d^2-2*d^3+288*d*T^2-36*d^2*T^2+...
    18*d^3*T^2+108*d^2*T^4+sqrt((128-96*d+24*d^2-2*d^3+288*d*T^2-36*d^2*...
    T^2+18*d^3*T^2+108*d^2*T^4)^2+4*(-4*(2+d)^2-3*d*(8+d)*(-
1+T^2))^3))^(1/3))+...
    (1/(3*2^(1/3)*(-1+T^2)))*((128-96*d+24*d^2-2*d^3+288*d*T^2-
36*d^2*T^2+18*...
    d^3*T^2+108*d^2*T^4+sqrt((128-96*d+24*d^2-2*d^3+288*d*T^2-
36*d^2*T^2+18*...
    d^3*T^2+108*d^2*T^4)^2+4*(-4*(2+d)^2-3*d*(8+d)*(-1+T^2))^3))^(1/3))
```

Appendix B Examples in Mesoscopic Conductance

As an instructive and much simpler case, consider instead the matrix implied by one-type 1D channel:

$$H_{11} = \alpha + \Sigma_{S,11}$$

$$H_{i,i-1} = H_{i,i+1} = \beta, H_{ii} = \alpha, i \in \{2, \dots, n-1\}$$

$$H_{nn} = \alpha + \Sigma_{D,nn}$$

where α is the self-energy, assumed constant, β is the exchange integral, assumed constant, Σ_S vanishes except the (1,1) element, and Σ_D vanishes except the (n,n) element. The retarded Green's function satisfies:

$$G^r = [E\ I - H]^{-1}$$

Using the methods in Chapter 2 which originate in Yueh, I apply the following eigendecomposition to the matrix H, noting that the $E\ I - H$ has the same decomposition offset by E:

$$\cos\theta = \frac{\xi - \alpha}{2\beta}$$

$$\beta^2 \sin\big((n+1)\theta\big) - (\Sigma_S + \Sigma_D)\sin(n\theta) + (\Sigma_S\Sigma_D)\sin\big((n-1)\theta\big) = 0$$

This equation as-written has a very complex closed-form solution that is not very useful. As a base case, when the lattice is "connected" to a source (drain) that is exactly itself, that is, $\Sigma_S = \Sigma_D = \beta$, the equation:

$$\sin\big((n+1)\theta\big) - \frac{2}{\beta}\sin(n\theta) + \sin\big((n-1)\theta\big) = 0$$

Appendix B

Has the same solutions as in Chapter 2, namely,

$$\theta = \frac{k\pi}{n}, k \in \{1 \dots n-1\}$$

This leads to the following eigenvalues (ξ_k) and eigenvectors ($\Xi_j^{(k)}$) for $E\ I - H$:

$$\xi_k = E - \alpha - 2\beta \cos\frac{k\pi}{n}, k \in \{1 \dots n-1\}$$

$$\Xi_j^{(k)} = \sin\frac{jk\pi}{n} + \sin\frac{(j-1)k\pi}{n}, j \in \{1 \dots n\}$$

As before there is one missing eigenvalue and eigenvector which are $\xi_n = E - \alpha - 2\beta$ and $\Xi^{(k)} = \mathbf{1}$. Because $E\ I - H$ is Hermitian, the inverse eigenvector matrix can be trivially written as the transpose:

$$\left(\Xi_j^{(k)}\right)^{-1} = \Xi_k^{(j)}$$

which immediately gives the retarded Green's function as:

$$G^r = \Xi[Diag(\xi_1^{-1}, \dots, \xi_n^{-1})]\Xi^{\mathrm{T}}$$

These solutions are standing waves. The standing waves are "tilted" by the potential so that they become transverse modes that empty into the drain. Examining the transmission spectrum in some detail, it can be seen that the transmission spectrum consists of peaks which are approximately lorentzian and centered about the eigenvalues, and that there are roughly as many significant eigenvalues as there are subunits, at low bias, whereas at high bias these wash out. The transmission spectrum for a prototypical case is shown in Figure B.1. Perturbations on the channel also have a relationship to the bias. At a low bias, weak perturbations close the channel, whereas at high bias, only perturbations near the source can close the channel.

Appendix B

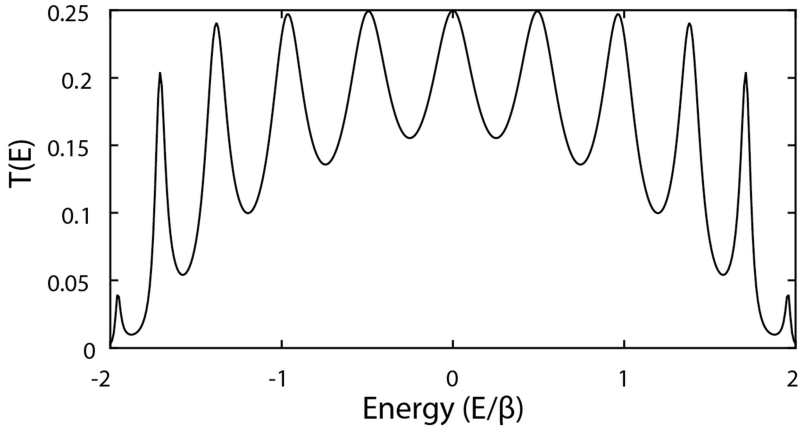


Figure B.1: Transmission spectrum of an 11 atom chain at low bias. There are 11 approximately lorentzian peaks centered about the 11 eigenvalues.

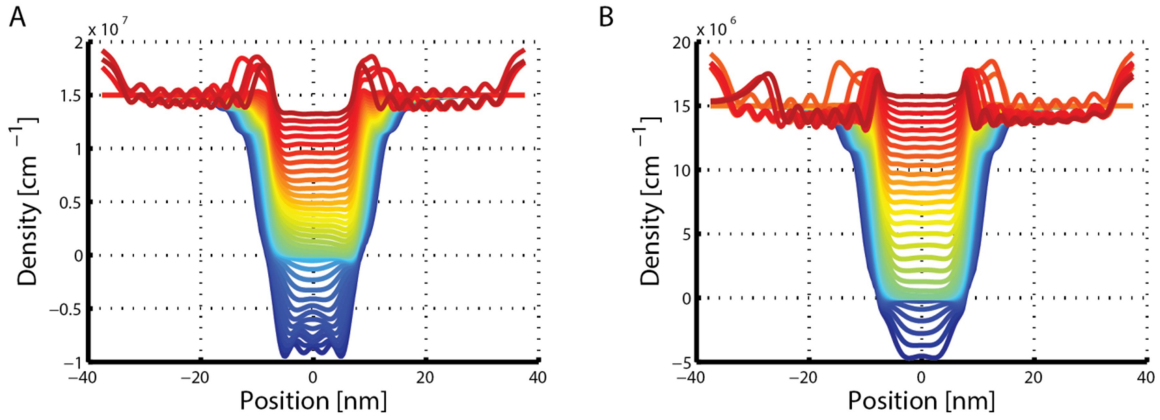


Figure B.2: Predicted electron density of a metallic (10,10) CNT from a cylindrically symmetric dielectric with V_g = -1 to V_g = 1V, with (A) V_{ds} = 0.40V and (B) V_{ds} = 0.04V simulated using the moscnt model provided by Guo *et al*[24], demonstrating spatial variations in charge density across the nanotube at severe gate biases.

Appendix C Additional Controls for Chapter 5

Optimization of smFET fabrication and molecule attachment:

To generate devices carbon nanotubes were grown on 1x1 cm^2 silicon substrates (525 μm degenerately-doped silicon covered with 285 nm thermally-grown oxide) via chemical vapor deposition[49]. Briefly, a 1:200 dilution of 10 mg/mL ferritin cationized from horse spleen in 0.15 M NaCl (Sigma Aldrich) was deposited onto the edge of the silicon substrate, followed by annealing under flow of argon/hydrogen gas (80 sccm total in a 5:1 ratio) for 20 minutes at 750°C. Nanotubes were grown by bubbling an argon/hydrogen gas mixture (50 sccm total in a 9:2 ratio) through ice-cold ethanol and over the annealed iron catalyst for 1 hour at 890°C.

Evaporation of 75 nm of titanium was used to create alignment marks and 32 parallel electrode pairs with 4 μm source-drain separation. A second evaporation of 100 nm of platinum was used to create two on-chip pseudo-reference electrodes. Substrates were then annealed in vacuum at 350°C. Scanning electron microscopy (Hitachi 4700) and confocal Raman microscopy (Renishaw inVia using a 532 nm laser) were used to visually characterize CNTs and to determine their diameter and chirality[64], respectively (Figure C. 1). Once an individual single-walled CNT was selected, oxygen plasma reactive-ion etching was used to isolate each device (by removal of CNT sections not located between source and drain electrodes) as well as to remove all other CNTs on the substrate. A microfluidic cell (7 mm long x 800 μm wide x 400 μm tall) was made out of polydimethylsiloxane (PDMS, cured at 80C) and stamped onto the

substrate. The mold for PDMS microfluidic cells was fashioned from crosslinked SU-8 2150 negative photoresist spun onto a silicon wafer at 500 rpm and subsequently exposed to light for 12 minutes.

smFET assay experimental conditions:

All single-molecule experiments were performed in 0.1X PBS (1mM Na/Na_2PO_4, 13.7mM NaCl, 0.3mM KCl, pH 7.4) supplemented with 10mM $MgCl_2$. Using the formula for the Debye length:

$$\lambda_D = \sqrt{\frac{\epsilon k_B T}{\sum n_i q_i^2}}$$

we calculate a Debye length of 1.30 nm.

2'-ACE protected RNA sequences for single-molecule experiments were purchased from GE Dharmacon with primary amine functionalities (5'-Amino modifier C6). To deprotect the bases, RNA samples were incubated at 60°C in 100 mM acetic acid, pH 3.8 for 30 minutes followed by ethanol precipitation, lyophilization, and resuspension in 0.1X PBS[259]. Before experiments of any type, RNA samples were heated to 95°C for 2 minutes followed by slow cooling to 25°C before addition of 10 mM $MgCl_2$. smFET measurements were conducted at a constant bias of -0.3V relative to the gate, as this was found to be the hold condition for the Pt metal.

Real-time invasion of the P1 stem:

Appendix C

Following attachment of the riboswitch aptamer to a CNT, to confirm that we were observing rearrangements of the P1 stem, we incubated the wild-type aptamer with 3 μM adenine and 1 μM DNA with a sequence complementary to the P1 stem, which we refer to as EPDNA (5'-TCCTGATTACAA-3'). In contrast to the signal in the absence EPDNA (Figure C. 3a), a two-state signal occasionally extinguishes the more complicated four-state signal (Figure C. 3b).

Ligand-free and alternate ligand experiments:

Following attachment of the riboswitch aptamer to a CNT, we observed that structural rearrangements of the riboswitch resulted in three discrete smFET conductance classes in the absence of adenine (Figure C. 4a). We analyzed the data using an adaptation of baseline correction algorithms to merge with those normally used in analysis of smFRET intensity *versus* time trajectories, and use the transition matrix to infer the rate constants[84,86]. The lifetimes of all observed classes fell within a range of 100 μs to 10 ms (Table S1). Subsequent addition of adenine led to transitions between four discrete conductance classes (Figure C. 4b), including an apparently new class characterized by a very short lifetime and significantly-lowered conductance (Figure C. 4b). Titration of adenine from 0.30 nM to 3 μM stabilized the lowest conductance class, extending its average lifetime by 13-fold between 30 nM and 3 μM adenine (Figure C. 4d). This result implies that the lowest conductance class is actively stabilized by addition of adenine. Together with the general observation that more negative charge near the CNT surface tends to lead to lower conductance under our experimental conditions[18,48,205], these results led us to hypothesize that the lowest conductance class represents a fully base-paired P1 stem, an observation consistent with the results of NMR spectra[235]. With this interpretation, addition of adenine leads to stabilization of the fully paired conformation of the P1 stem. It is

likely that this conductance class was too short lived to be observed in the absence of adenine, consistent with RNA secondary structure calculations using MFOLD we initiated suggesting that the terminal base is unpaired, and that its total contribution to the structure is approximately -0.8 kcal/mol, consistent with highly transient dynamics[260].

Addition of 3 μM 2AP followed the same trend as addition of adenine; however, the rates were subtly different as a result of different proportions in $P1^A$ and $P1^B$, as well as a lowered rate of transition between $P1^B$ to $P1^A$ while in conductance class 3, leading to much more transient population of the fully paired state (Figure C. 4c and Figure C. 4e).

Bulk Fluorescence assays:

It has been shown that the *pbuE* riboswitch binds adenine and 2-aminopurine (2AP) with similar affinity[228], though the measured rate of association of these metabolites varies by a factor of three[230]. As has been reported elsewhere, 2AP can be selectively excited and its fluorescence, which is quenched by base stacking[261], can be monitored to measure quenching caused by binding to the *pbuE* riboswitch[231]. We performed fluorescence-quenching assays using a Perkin Elmer LS55 luminescence spectrophotometer and collected spectra over the wavelength range 330-450 nm with 300 nm excitation. RNA samples (prepared as described above) were heated to 95°C for 2 minutes in either 0.01X, 0.1X, or 1X phosphate-buffered saline (PBS) followed by slow cooling to 25°C before addition of 10 mM $MgCl_2$. Data was collected at 25°C with a fixed 2AP concentration of 100 nM and a range of RNA concentrations in excess of 2AP to simplify the binding equation to the following:

$$\Delta F/F = (1-\alpha)[RNA]/(K_D + [RNA])$$

where ΔF/F is the percent fluorescence intensity lost upon the addition of a known concentration of riboswitch, [RNA]. The parameter α is proportional to the quantum yield of 2AP fluorescence and K_D represents the equilibrium dissociation constant of 2AP from the riboswitch[231]. These assays show that, in the presence of 10 mM $MgCl_2$, a change in monovalent salt between 0.01X and 1X PBS does not significantly affect the binding of 2AP to the *pbuE* aptamer (Figure C. 5a). Additional quenching assays were used to measure the K_D of 2AP for mutated aptamer sequences (Figure C. 5b).

Conductance *versus* time trajectories statistics; rate constants; hierarchical model selection:

Rate constants were estimated using two separate models. In the first, shown in the text (Figure C. 4b) and below (Figure C. 7a), the conductance *versus* time trajectory was split into parts small enough to contain 10 events each, evaluated by eye, independently fit to a baseline-correcting markov chain, and strung back together (Figure C. 7b) by taking the occupancy posteriors and using them to recalculate a consensus baseline before running a baseline correction Hidden Markov Model (HMM) on the entirety as independent trace fragments with consensus emissions and baseline. A first-order approximation to the learned aggregated transition matrix from all the expected pseudocounts from the last process (Figure C. 7b) was used to estimate the average rate constant at each condition. Error bars were estimated using the Dirichlet distribution implicit in the HMM to generate 95% confidence intervals. In the second, a hierarchical markov model was used (Figure C. 7c). However, two conceptual differences were applied – first, the baseline correction from the first method was directly applied to the conductance *versus* time trajectory before inference, making this method a reflection of the Viterbi, or idealized, trajectory through

the data; second, all the trajectories from a series of experiments from a single device were entered into the model at once, so that the model posterior would be a reflection of the entire body of data. Error bars were, again, 95% confidence intervals calculated using the implicit Dirichlet distribution of the hierarchical markov model. Using the consensus model, rate constants for each condition were obtained from the stored expectation values for each of the relevant distributions.

Model selection for the hierarchical model consisted of fitting with an extra kinetic class and quantifying how populated it was. Our expectation was that we had enough data to do "automatic" model selection by depopulating an unnecessary class. For the wild-type aptamer, the overall fraction of every adenine condition in the third kinetic class was 0.0006 and for the G21C aptamer, the fraction of every adenine condition in the third kinetic class was 0.001. We interpreted these results as implying that the third class is unnecessary to explain the data, and on this basis present only two kinetic classes, $P1^A$ and $P1^B$.

Calculation of ΔG separations for $P1^A$ and $P1^B$ for the wild-type and G21C aptamers:

In order to compare the kinetic states present in the wild-type and G21C time series, we first calculate, for each conductance class in each kinetic state:

$$\Delta G_i^A = ln \frac{k_{i,i+1}^A}{k_{i+1,i}^A}$$

$$\Delta G_i^B = ln \frac{k_{i,i+1}^B}{k_{i+1,i}^B}$$

which is the thermodynamic partition between entering conductance class i from a conductance class i+1 within $P1^A$ *versus* the reverse rate from conductance class i+1 back to conductance class i. This corresponds to the probability of zipping rather than unzipping. These quantities for the wild-type and G21C aptamers are shown below (Figure C. 6).

Dynamics of the stable aptamer:

Amine labeled stable aptamer differs from wild-type by possessing the following modifications: U1C, U2G, A62C, A63G. We attached the stable aptamer to CNTs and recorded conductance *versus* time trajectories as above, in the presence and absence of 3 μM adenine and of 2AP. In all cases (Figure C. 10) we observed fluctuations between two conductance classes, but the dynamics in each condition were very similar, preventing us from characterizing the stable aptamer further.

Appendix C

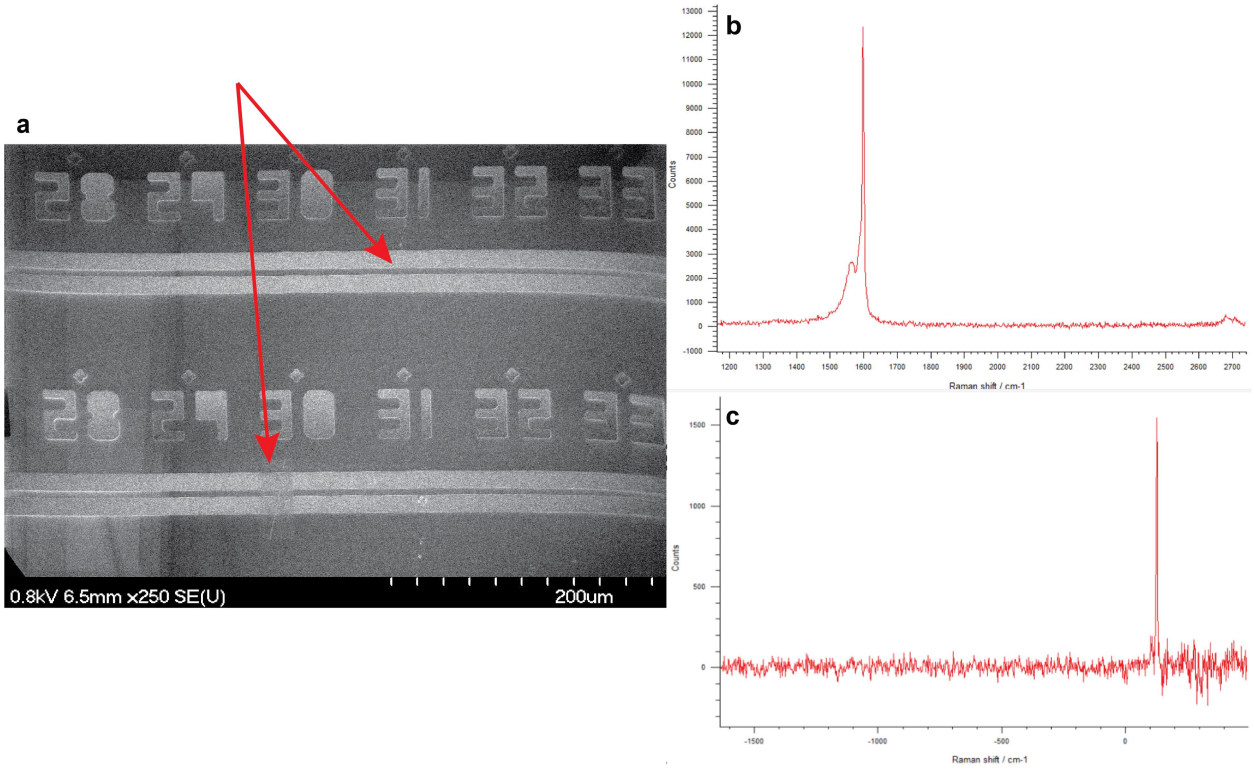


Figure C. 1: Characterization of CNT transistors.
a, Scanning electron micrograph of the mask used in this study. Isolated CNTs are indicated by the red arrows. **b,c**, Raman spectra of the CNT fragment used for collection of the wild-type aptamer data. The nanotube is metallic, as seen by its G band linewidth (**b**) and has a diameter of 1.95nm as seen by its radial breathing mode (**c**).

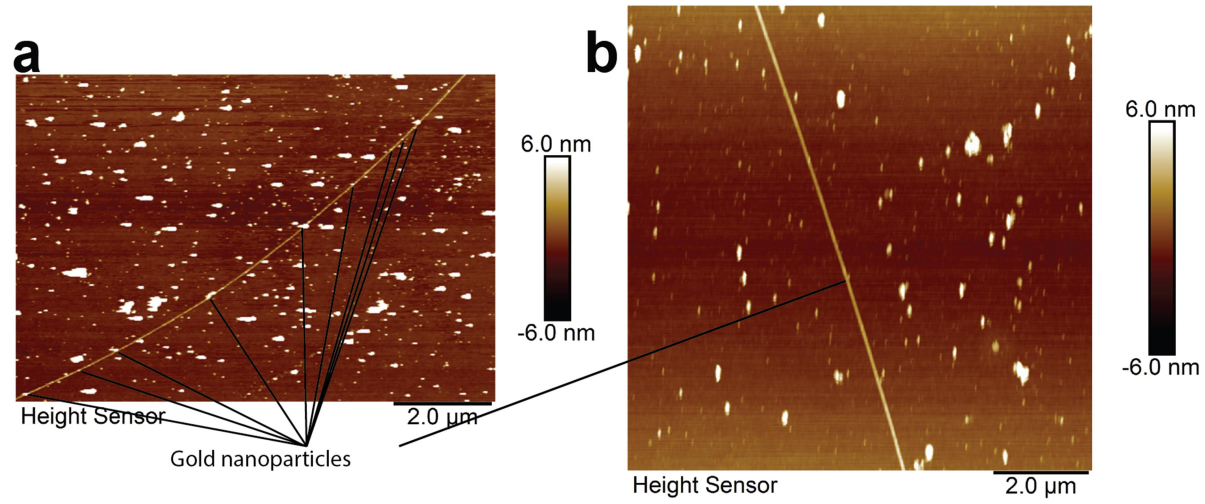


Figure C. 2: Optimization of smFET functionalization.
a, CNT incubated with both pyrene-NHS and gold nanoparticles. **b**, Only incubated with gold nanoparticles.

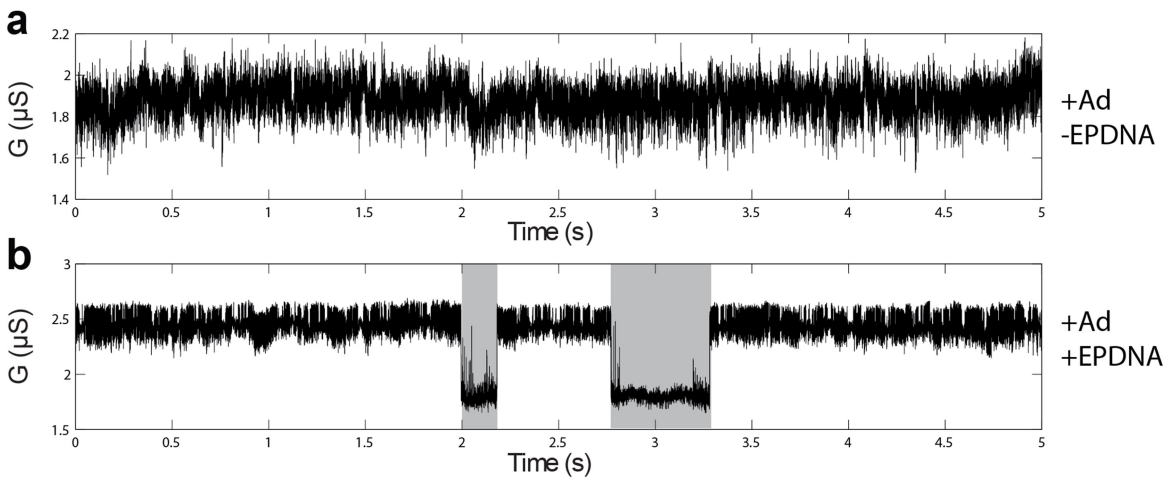


Figure C. 3: Real-time invasion of the P1 stem.
a, wild-type aptamer in the presence of saturating adenine. **b,** wild-type aptamer in the presence of saturating adenine and a DNA sequence matching the expression platform. Shaded regions are novel, EPDNA-dependent events, which we have simply marked by hand.

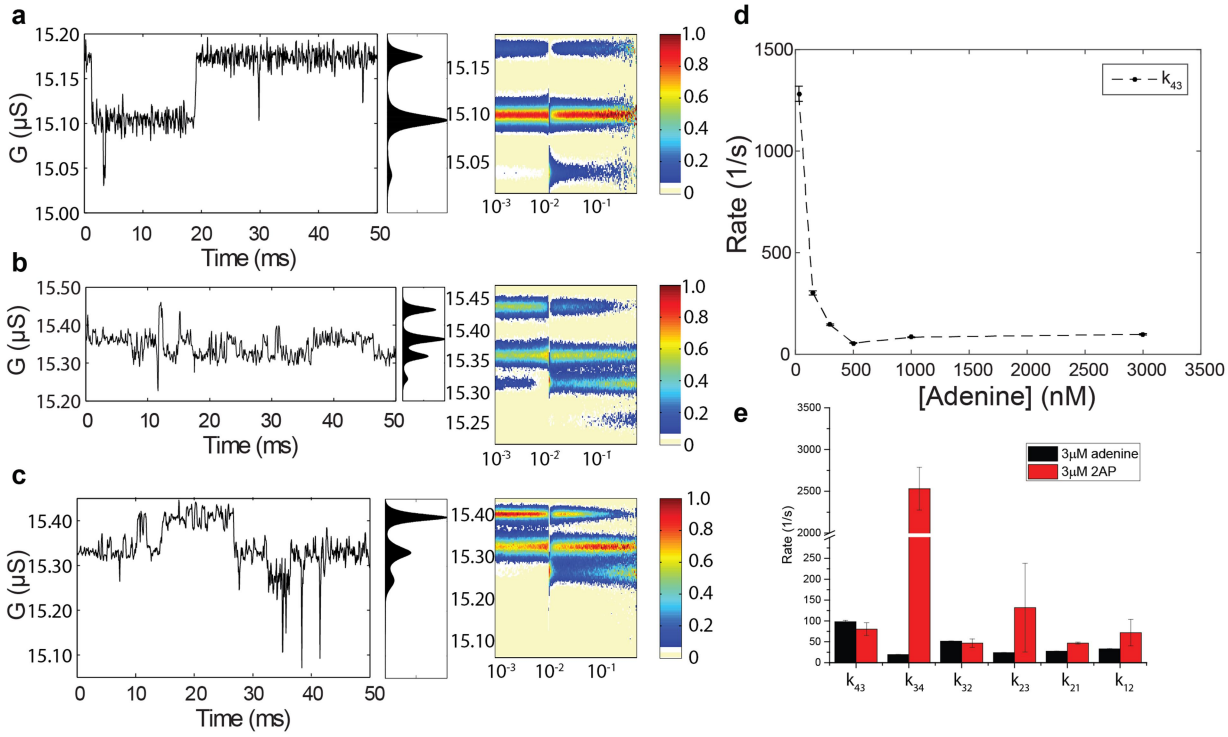


Figure C. 4: Ligand-free fluctuations of the wild-type aptamer.
a, sample trace, histogram, and post-synchronized 2D histogram of the wild-type aptamer before any addition of ligands. The trajectory possesses 3 conductance classes instead of the normal four. 2D histogram is synchronized to entry into conductance class 3. **b**, Reproduced sample trace, histogram, from Figure 1d of the main text (i.e., with 3 μM adenine). 2D histogram of the entire population, starting in conductance class 3 (either $P1^A$ or $P1^B$) and terminating in 1-$P1^A$, identical to those in the main text. **c**, sample trace, histogram, and post-synchronized 2D histogram as in **b**, with 3 μM 2AP instead of 3 μM adenine. **d**, adenine dependence of k_{43}. **e**, overall changes in average rate constants of 3 μM adenine *vs* 3 μM 2AP.

Appendix C

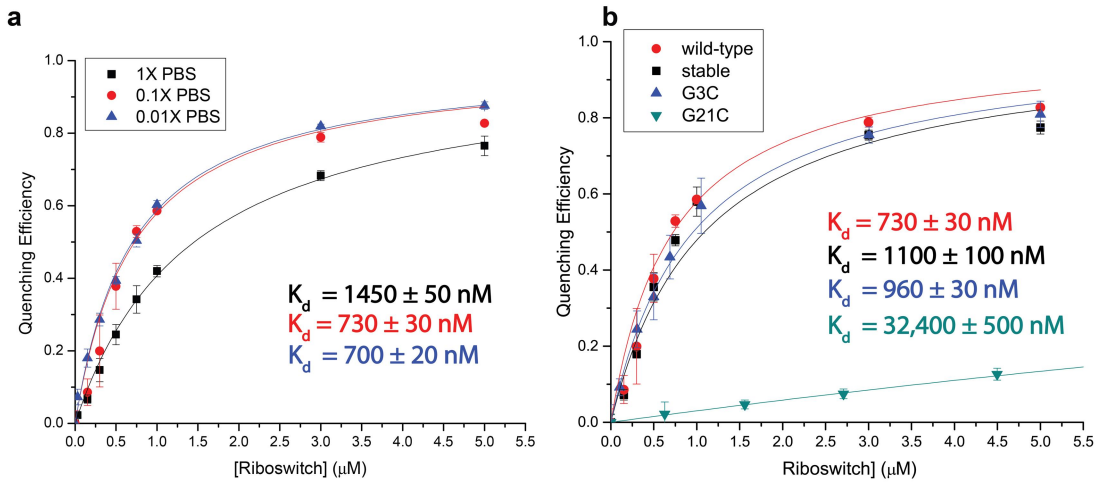


Figure C. 5: Bulk fluorescence binding data of 2-aminopurine to the wild-type aptamer. **a,** As monovalent salt is dropped, the K_d for 2AP, and by assumption adenine, is slightly salt dependent. **b,** Fluorescence binding assay of the mutants characterized in this study, wild-type, stable, G3C, and G21C (see text). Red curve is the same curve as in **a**, for reference.

Appendix C

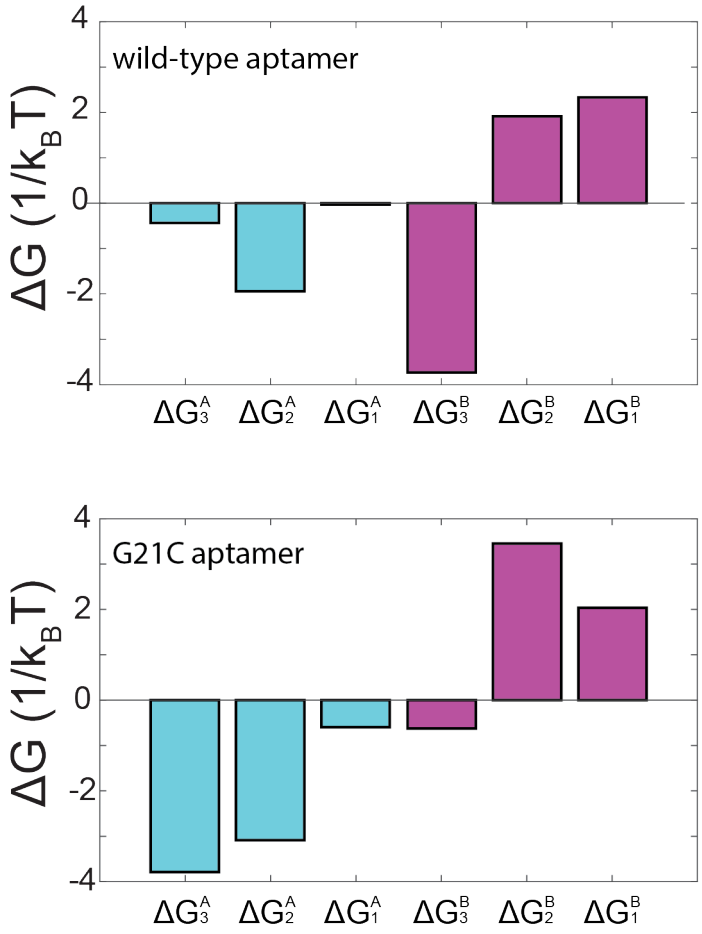


Figure C. 6: ΔG separation between conductance classes of P1A and P1B for the wild-type and G21C aptamers.

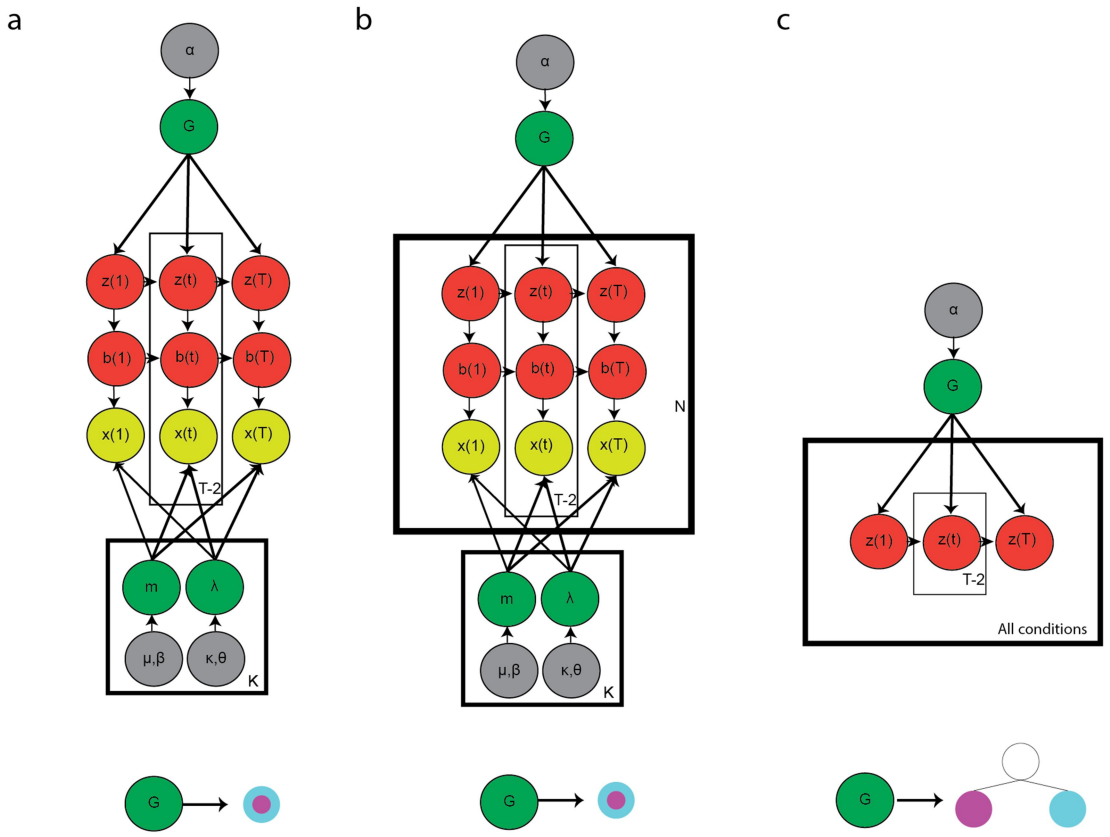


Figure C. 7: Models used for analysis of smFET data in Chapter 5.
a, the data are initially split into groups which are analyzed with independent baseline-correction HMMs. **b**, The individual fragments are strung together into a consensus set of transitions and emissions. Both **a** and **b** assume a single population (G, below). **c**, Setting G with two dynamically interconverting subpopulations, we then combine the idealized trajectories at every condition and analyze them together to get consensus parameters for each signal. In every graph above, graph circles denote prior densities, red denote hidden state variables, green denote expectation values, yellow denote observables. The graphs below denote the kinetic structure of the hierarchical transition matrix used at a given stage of analysis.

Appendix C

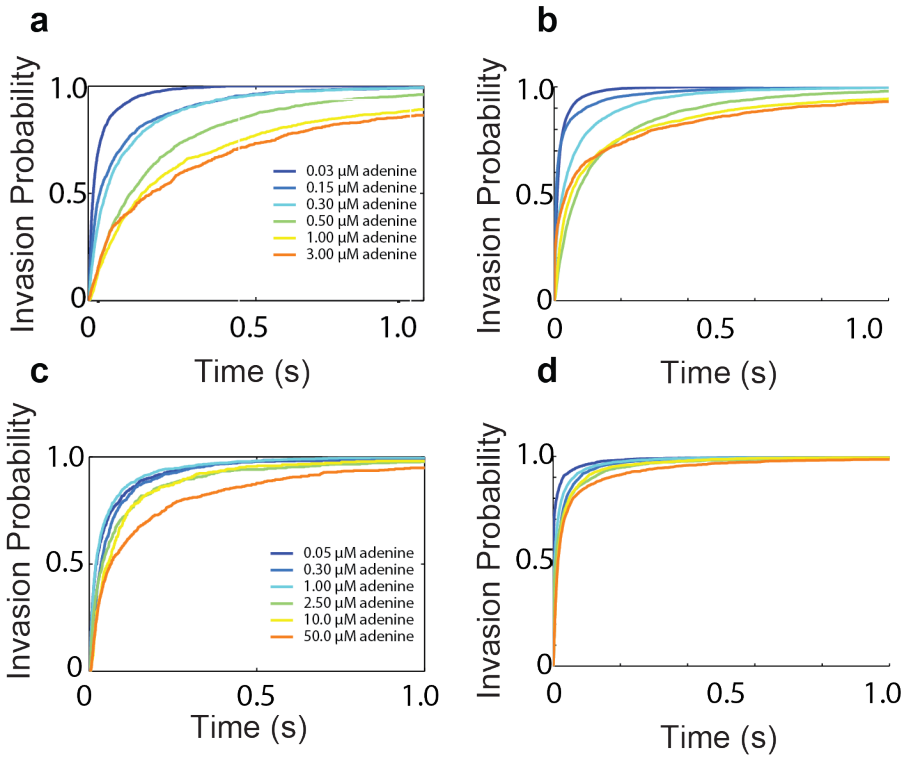


Figure C. 8: First passage distributions from the *pbuE* riboswitch trajectories described in Chapter 5.
a, First-passage time beginning in 3-P1^{B} and terminating in 1-P1^{A} for the wild-type aptamer. **b**, First-passage time beginning in 3-P1^{A} and terminating in 1-P1^{A} for the wild-type aptamer. **c**, First-passage time beginning in 3-P1^{B} and terminating in 1-P1^{A} for the G21C aptamer. **d,** First-passage time beginning in 3-P1^{A} and terminating in 1-P1^{A} for the G21C aptamer.

Appendix C

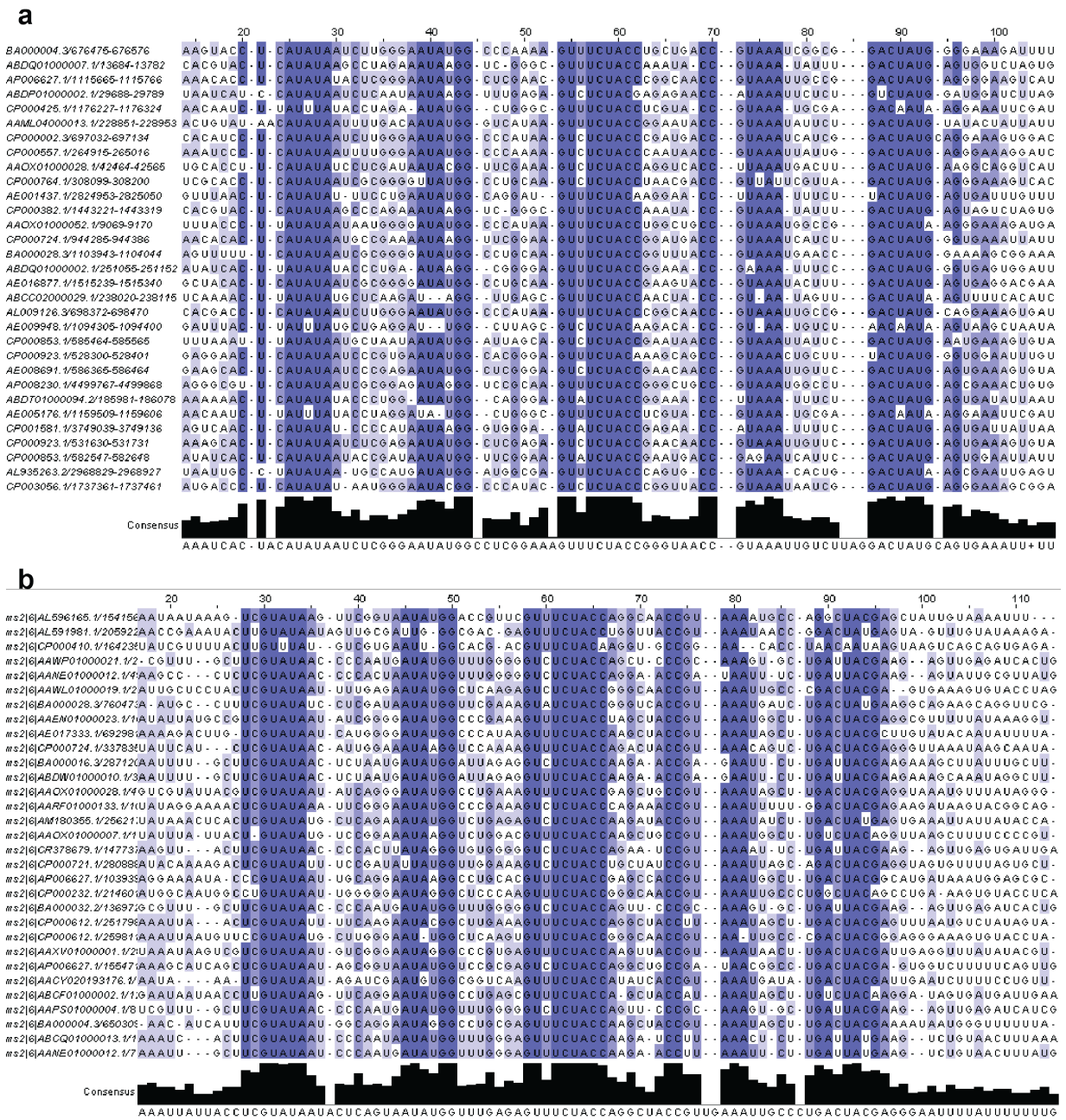


Figure C. 9: Segmented alignment of *pbuE* riboswitch sequences described in Chapter 5. **a**, Alignment of sequences with an A in the third position (here, 25) of the P1 stem. **b**, Alignment of sequences with a G in the third position (here, 30) of the P1 stem, as possessed by the wild-type aptamer under study. Figure generated with jalview using the RF00167 sequence library found on the Rfam database.

Appendix C

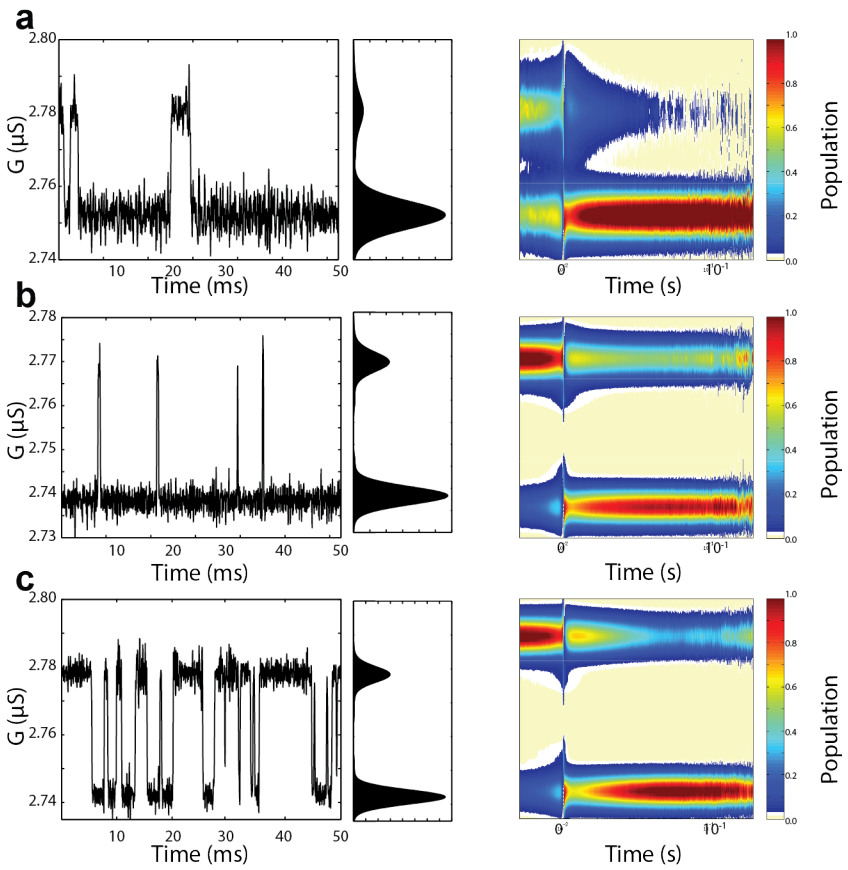


Figure C. 10: Dynamics of the stable aptamer.
a, Recording in the absence of ligand, **b**, in the presence of 3 μM adenine, and **c**, in the presence of 3 μM 2AP.

[Adenine] (μM)	k_{43} (s^{-1})	k_{34} (s^{-1})	k_{32} (s^{-1})	k_{23} (s^{-1})	k_{21} (s^{-1})	k_{12} (s^{-1})
0 (not measured)						
0.03	1280±40	33±1	73±2	22±1	27±1	24±1
0.15	300±10	34±1	109±2	35±1	35±1	34±1

Appendix C

0.30	145±4	27±1	86±2	27±1	37±1	51±1
0.50	55±2	10.5±0.3	36±1	17.4±0.4	38±1	51±1
1.00	84±2	14±0.4	43±1	30±1	42±1	43±1
3.00	98±4	20±1	51±1	24±1	28±1	33±1
3.00 [2AP]	81±15	2500±300	50±10	130±100	47±3	72±30

Table C 1: Average rate constants for the wild-type *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals.

[Adenine] (μM)	k_{43} (s^{-1})	k_{34} (s^{-1})	k_{32} (s^{-1})	k_{23} (s^{-1})	k_{21} (s^{-1})	k_{12} (s^{-1})
0 (not measured)						
0.05	152±12	14±1	150±5	97±4	0.2±0.1	19±0.4
0.30	240±30	25±2	180±8	38±2	7±0.4	23.5±0.5
1.00	100±15	5±0.5	0.2±0.2	3.8±0.4	0.02±0.01	0.9±0.1
2.50	65±16	4.3±0.8	1±1	3.3±0.6	0.03±0.03	1.4±.2
10.0	19±6	0.3±0.2	45±5	20±0.2	0.05±0.02	3.8±0.2
50.0	23±5	0.3±0.2	1.6±0.6	4.9±0.6	0.02±0.02	0.03±0.02

Table C 2. Average rate constants for the G21C *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals.

[Adenine] (μM)	k_{43} (s^{-1})	k_{34} (s^{-1})	k_{32} (s^{-1})	k_{23} (s^{-1})	k_{21} (s^{-1})	k_{12} (s^{-1})	$k_{AB}^{(4)}$ (s^{-1})	$k_{AB}^{(3)}$ (s^{-1})	$k_{AB}^{(2)}$ (s^{-1})	$k_{AB}^{(1)}$ (s^{-1})
0.03	3590±90	330±10	2630±30	229±3	545±4	265±2	61±10	155±5	25±1	2.5±1
0.15	590±15	365±8	1870±20	220±2	183±2	180±2	22±3	143±4	13±0.6	3.8±0.2
0.30	560±10	336±6	1210±20	196±2	62±1	95±2	28±2	135±5	16±0.5	1.7±0.2
0.50	357±6	456±8	1090±15	149±2	71±1	135±2	25±2	178±6	16.6±0.6	0.8±0.2
1.00	363±5	620±10	1100±15	274±3	114±2	116±2	27±2	211±4	26±1	1.0±0.2
3.00	322±5	404±7	960±10	223±2	30±1	48±2	25±2	167±5	20±1	0.5±0.2

Appendix C

Table C 3. Rate constants for the wild-type *pbuE* adenine-sensing aptamer contingent on occupancy in $P1^A$ under conditions of increasing adenine; error bars are 95% confidence intervals.

[Adenine] (μM)	k_{43} (s^{-1})	k_{34} (s^{-1})	k_{32} (s^{-1})	k_{23} (s^{-1})	k_{21} (s^{-1})	k_{12} (s^{-1})	$k_{BA}^{(4)}$ (s^{-1})	$k_{BA}^{(3)}$ (s^{-1})	$k_{BA}^{(2)}$ (s^{-1})	$k_{BA}^{(1)}$ (s^{-1})
0.03	4100±120	175±6	640±10	860±15	1200±20	2830±40	300±35	75±4	370±10	48±5
0.15	3680±80	84±2	303±4	1700±20	540±15	3960±80	100±15	44±2	200±7	65±10
0.30	2920±80	59±2	301±4	1640±20	500±15	5000±100	150±20	83±2	215±8	85±15
0.50	2730±60	54±1	209±3	1830±20	400±10	5500±100	95±10	46±1	170±7	105±20
1.00	2750±50	60±1	212±2	1960±20	334±8	6800±150	90±10	39±1	136±5	90±20
3.00	2620±60	44±1	159±2	1760±20	90±5	6500±300	100±10	50±1	144±6	110±50

Table C 4. Rate constants for the wild-type *pbuE* adenine-sensing aptamer contingent on occupancy in $P1^B$ under conditions of increasing adenine; error bars are 95% confidence intervals.

[Adenine] (μM)	$F(P1^A)$ (%)	$F(P1^B)$ (%)
0.03	7.79±0.01	92.2±0.01
0.15	17.9±0.01	82.1±0.01
0.30	16.7±0.01	83.3±0.01
0.50	28.8±0.02	71.2±0.01
1.00	40.3±0.03	58.7±0.03
3.00	34.2±0.02	65.8±0.02

Table C 5. Fractional occupancy of $P1^A$ or $P1^B$ for the wild-type *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals.

Appendix C

[Adenine] (μM)	k_{43} (s^{-1})	k_{34} (s^{-1})	k_{32} (s^{-1})	k_{23} (s^{-1})	k_{21} (s^{-1})	k_{12} (s^{-1})	$k_{AB}^{(4)}$ (s^{-1})	$k_{AB}^{(3)}$ (s^{-1})	$k_{AB}^{(2)}$ (s^{-1})	$k_{AB}^{(1)}$ (s^{-1})
0.05	3100±200	210±20	2960±60	78±2	250±4	100±2	70±40	200±20	5±0.5	1±0.2
0.30	3400±300	50±6	1790±30	91±2	250±3	153±2	93±60	98±8	6.2±0.5	0.6±0.1
1.00	2400±200	27±3	975±20	77±2	128±2	127±2	61±40	97±6	14.1±0.7	1.5±0.2
2.50	2300±400	15±3	1120±20	65±2	106±2	87±2	70±70	96±7	10.5±0.7	0.5±0.1
10.0	2700±500	25±5	1520±40	83±2	217±4	65±1	93±97	90±10	7.3±0.7	0.38±0.09
50.0	2000±300	33±6	1150±30	96±3	204±5	48±1	83±58	136±10	18±1	0.6±0.1

Table C 6. Rate constants for the G21C *pbuE* adenine-sensing aptamer contingent on occupancy in $P1^A$ under conditions of increasing adenine; error bars are 95% confidence intervals.

[Adenine] (μM)	k_{43} (s^{-1})	k_{34} (s^{-1})	k_{32} (s^{-1})	k_{23} (s^{-1})	k_{21} (s^{-1})	k_{12} (s^{-1})	$k_{BA}^{(4)}$ (s^{-1})	$k_{BA}^{(3)}$ (s^{-1})	$k_{BA}^{(2)}$ (s^{-1})	$k_{BA}^{(1)}$ (s^{-1})
0.05	264±8	175±5	80±4	3100±100	550±50	2200±200	7.5±1	18±2	400±50	420±100
0.30	370±15	153±6	133±6	2700±100	280±35	2700±300	12±3	52±4	280±35	500±150
1.00	290±15	51±3	57±3	1560±60	45±10	300±90	11±3	64±3	150±20	200±80
2.50	205±15	21±2	58±3	1630±70	21±8	400±200	13±4	53±3	150±20	380±200
10.0	98±6	80±5	86±5	2030±100	80±25	1200±400	3±1	56±4	200±40	630±300
50.0	206±6	98±2	42±2	1640±70	40±10	700±200	4±1	28±2	110±20	260±120

Table C 7. Rate constants for the G21C *pbuE* adenine-sensing aptamer contingent on occupancy in $P1^B$ under conditions of increasing adenine; error bars are 95% confidence intervals.

[Adenine] (μM)	$F(P1^A)$ (%)	$F(P1^B)$ (%)
0.05	84.9±0.1	15.1±0.1
0.30	91.6±0.1	8.4±0.1
1.00	84.1±0.1	15.9±0.1
2.50	87.9±0.1	12.1±0.1
10.0	92.3±0.1	7.7±0.1
50.0	78.1±0.1	21.9±0.1

Table C 8. Fractional occupancy of $P1^A$ or $P1^B$ for the G21C *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals.

Appendix C

[Adenine] (μM)	k_{21} (s^{-1})	k_{12} (s^{-1})
0	33±6	16±4
0.03	77±4	134±4
0.15	49±2	250±2
0.30	79±2	351±2
0.50	95±6	374±6
1.00	134±2	400±2
3.00	134±4	420±4

Table C 9. Average rate constants for the G3C *pbuE* adenine-sensing aptamer under conditions of increasing adenine; error bars are 95% confidence intervals.

Appendix D Additional information for Chapter 4

Stem-loop	k_{43} (s^{-1})	k_{34} (s^{-1})	k_{32} (s^{-1})	k_{23} (s^{-1})	k_{21} (s^{-1})	k_{12} (s^{-1})
GAAA	40±3	10±1	20±1	48±1	65±1	15±1
GCAA	118±8	15±1	200±4	15±0.3	7.5±0.2	108±2
UACG	25±1	7.5±0.2	110±1	875±7	90±2	1543±20
UUCG	1040±11	45±1	218±2	80±1	40±1	43±1
UUUU	1250±20	20±1	240±4	43±1	325±2	122±1

Table D. 1: Average rate constants for the stem-loop constructs tested in the absence of competitor DNA; error bars are 95% confidence intervals.