Single-Molecule Visualization of Conformational Changes in the SecA ATPase

by

Jacob D. Sargent

B.S. Physics
Rice University, 2007

SUBMITTED TO THE HARVARD-MIT DIVISION OF HEALTH SCIENCES AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN HEALTH SCIENCES AND TECHNOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2015

Signature redacted

Signature of author: _____
Harvard-MIT Division of Health Sciences and Technology
May 18, 2015

Signature redacted

Certified by: _____
Joseph J. Loparo, PhD
Assistant Professor, Biological Chemistry & Molecular Pharmacology, Harvard Medical School
Thesis Supervisor

Signature redacted

Accepted by: _____
Emery N. Brown, MD, PhD
Director, Harvard-MIT Health Sciences and Technology
Associate Director, Institute for Medical Engineering and Science
Professor, Computational Neuroscience and Health Sciences and Technology
Warren M. Zapol Professor of Anesthesia, Harvard Medical School and Mass. General Hospital

Single-Molecule Visualization of Conformational Changes in the SecA ATPase

by

Jacob D. Sargent

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on May 18, 2015 in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Health Sciences and Technology

ABSTRACT

The need for new antibiotics is great as bacterial strains with single and multiple drug resistance have continued to grow more prevalent since the 1980's[1,2]. At the same time, the rate of approval of new antibiotics has dropped precipitously[1]. Existing antibiotics commonly target the bacterial ribosome[3,4] or cell wall synthetic pathways[5]: two targets that are essential for bacterial survival. However, another option is to target a pathway which is more intimately connected to bacterial pathogenesis: protein secretion[6].

In bacteria, most secreted polypeptides are pushed accross the membrane, via the SecYEG channel, by the SecA ATPase[7]. Relatively little is understood of how SecA couples ATP hydrolysis to polypeptide translocation. X-ray crystallography and many biochemical studies support a model in which the two-helix finger (2HF) of SecA pushes the polypeptide through the SecYEG channel[8-12], however some evidence is contradictory[13]. We aim to directly measure conformational changes of the 2HF by utilizing single-molecule Förster resonance energy transfer (smFRET). Directly measuring conformational changes in an ATPase will also provide further insight into the guiding principles of ATPase function.

First, we will build a smFRET microscope and assemble a software package to analyze the data it collects. We will then validate these tools by reproducing results currently in the literature from Holden et al.[14] and McKinney et al.[15]. Next, we will assess the potential limitations of current tools for smFRET data analysis, especially as applied to ATPases. We will propose a new approach that may be useful in these systems. Finally, we will use the smFRET microscope to measure ATP-dependent conformational dynamics of the 2HF. This evidence will help differentiate between three proposed models: the 2HF (1) is not directly involved in polypeptide translocation, (2) moves unidirectionally, directly driving translocation, or (3) moves back and forth but in a way that is coordinated by ATP hydrolysis with progress capture elsewhere in SecA.

Thesis Supervisor: Joseph J. Loparo
Title: Assistant Professor of Biological Chemistry and Molecular Pharmacology
Harvard Medical School

Acknowledgements

I would like to express my sincere gratitude to those whose support and guidance has allowed me to shape my graduate school experience into everything I wanted it to be. My supervisor, Dr. Joseph Loparo, has been continually supportive and encouraging. He has created a laboratory atmosphere which fosters curiosity, friendship, and the highest level of scientific inquiry. I am grateful to have had the opportunity to grow in this environment and learn from the extremely talented individuals around me: HyeongJun Kim, Seungwoo Chang, Lizz Thrall, James Kath, Thomas Graham, and Linda Song.

I would also like to thank my collaborator, Benedikt Bauer, whose tireless efforts and biochemical prowess allowed us to conduct single-molecule experiments on a system that most would have considered too complex to reconstitute *in vitro*. I would also like to thank Benedikt's supervisor, Tom Rapoport, for his support of these cutting-edge and somewhat risky experiments.

I am also grateful to have had the immeasurable support of the HST community. My academic advisor, Brett Bouma, has been a constant and level-headed source of guidance and support. Of course I am also grateful for the personal support of my friends and colleagues in HST who are too numerous to mention by name.

Lastly, I would like to express gratitude to those who enriched my life outside of graduate school. I thank the MIT Outing Club for providing a constant source of rejuvenating experiences and friendships. I thank Zach Epstein and other friends for their support and help taking my mind off of stressful thoughts. I especially thank Sarah Robison for brining me the peace and clarity to see who I am and what I want. Finally, I thank my parents for whom my gratitude is too great for words.

Table of Contents

List of Figures

List of Abbreviations

ADP -- adenosine diphosphate
ADP BeFx -- adenosine diphosphate beryllium fluoride
ATPase -- adenosine triphosphatase
ALEX -- alternating laser excitation
DNA -- deoxyribonucleic acid
EMCCD -- electron multiplying charge-coupled device
FRET -- Förster resonance energy transfer
GUI -- graphical user interface
HMM -- hidden Markov model(ing)
PEG -- polyethylene glycol
P_i -- inorganic phosphate (PO_4^{2-})
smFRET -- single-molecule Förster resonance energy transfer
TDP -- transition density plot
TRE -- target registration error

1. Motivation and Background

The need for new antibiotics is great as bacterial strains with single and multiple drug resistance have continued to grow more prevalent since the 1980's[1,2]. At the same time that antibiotic resistance has been growing, the rate of approval of new antibiotics has dropped precipitously along with the number of companies working to develop antibiotics for approval[1]. This sets the stage for a significant global health problem with rampant untreatable bacterial infections. An exciting approach to solve this problem is to develop new antibiotics with different targets than existing antibiotics in an attempt to circumvent resistance. Existing antibiotics, to which some bacteria have developed resistance, commonly target the bacterial ribosome[3,4] or cell wall synthetic pathways[5]. These two targets are essential for bacterial survival, but another option is to target a pathway which is more intimately connected to bacterial pathogenesis: protein secretion[6].

1.1 Polypeptide translocation in bacteria: the SecA/SecYEG system

In bacteria, most secreted polypeptides cross the membrane via the SecYEG channel complex[7]. SecYEG is a passive pore and requires either the ribosome or the SecA adenosine triphosphatase (SecA ATPase) to push the polypeptide through the channel[7]. Co-translational translocation, in which the ribosome feeds the nascent polypeptide directly into the SecYEG channel, is typical for polypeptides destined to be inserted into the plasma membrane after escaping out the side of the SecYEG channel[7,16]. Most secreted proteins, however, are translocated by the combination of SecYEG and the SecA ATPase[7].

1.1.1 The passive SecYEG pore complex

The SecYEG pore complex is a heterotrimer made up of SecY, SecE, and SecG (these are also referred to respectively as the α, γ, and β subunits)[17]. SecY makes up the bulk of the channel and contains most of the transmembrane segments[17,18]. SecE is smaller but still essential for polypeptide translocation, while SecG is not essential for channel function in vitro[19,20] or in vivo[21]. The opening in the middle of the SecY complex is an hourglass shape with cytoplasmic and periplasmic funnels, reducing interaction with the polypeptide chain to a ring of six hydrophobic residues at the narrowest point: the pore ring (Fig. 1A)[17]. The pore ring is about 5-8 Å in diameter[17], which, if left open, would easily allow ions and small molecules to flow freely from one side of the membrane to the other. To preserve the integrity of the membrane, there is a plug formed by a helical tilted transmembrane segment on the periplasmic side [18,22,23]. The plug swings out into the periplasm and then tucks in on the periphery of the complex to allow polypeptides to pass through the pore (Fig. 1A&B)[17].

Polypeptides which are to be transported by the SecA/SecYEG system start as preproteins with a signal sequence which targets them to both SecA and SecYEG[24]. When this signal sequence and/or SecA interacts with SecYEG, the helices that form the channel shift open and the plug moves (Fig. 1B), priming the channel for polypeptide translocation[8,24]. The signal sequence is removed after translocation by signal peptidases[7]. There is some disagreement in the

field as to the oligomeric state of SecYEG in the active translocon[7,12,25,26], but a recent study indicates that a single copy of SecYEG is sufficient for polypeptide translocation[27].

The human homolog of SecY is Sec61 and it functions similarly in the endoplasmic reticulum membrane: inserting proteins into the membrane, transporting proteins across the membrane, playing a role in calcium signaling, and more[28]. Mutations in Sec61 have been associated with diabetes mellitus[29] and glioblastoma multiforme[30]. Like SecY, Sec61 associates with an ATPase (BiP) during post-translational translocation of polypeptides. A Brownian ratchet model has been validated for this process[31], but it is still unknown if SecA/SecYEG operates via a similar mechanism.

1.1.2 The SecA ATPase

SecA is an ATPase which interacts with phospholipids, keeping it associated with the membrane and increasing its local concentration around SecY[11,32]. SecA then binds to both SecY[8] and the substrate polypeptide[33,34] and coordinates ATP hydrolysis with pushing the polypeptide through the SecYEG channel. Its ATPase activity is stimulated by SecY and the preprotein substrate[35,36]. Several experiments have indicated that SecA binds tightly to the polypeptide substrate in its ATP-bound state but allows the polypeptide to slide more freely in its ADP bound state[11,37]. Upon ADP release, there is a large conformational change[38,39] which may recapture the polypeptide. From these results, a "push and slide" mechanism has been proposed in which SecA alternately pushes the polypeptide through the channel while hydrolizing ATP, and then allows the polypeptide to diffuse while bound to ADP[11].

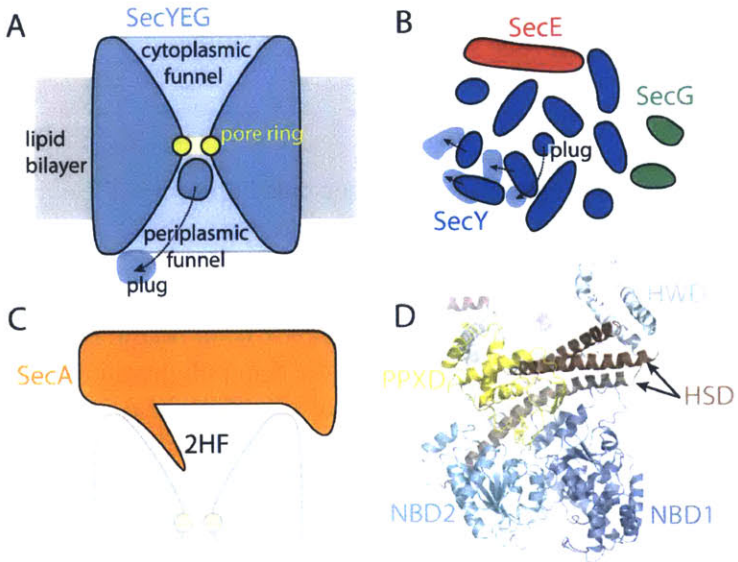


Fig. 1: Basic structures of SecYEG and SecA. (**A**) Cartoon of a vertical cross-section of the SecYEG channel with the motion of the plug upon channel opening indicated. (**B**) Cartoon of a horizontal cross-section of SecYEG with opening motion of transmembrane helices indicated. (**C**) Cartoon of SecA docked on SecYEG with the 2HF indicated. (**D**) Top (cytosolic) view of SecA with domains labeled: helical wing domain (HWD), nucleotide binding domains (NBD1 & NBD2), polypeptide cross-linking domain (PPXD), and helical scaffold domain (HSD). The 2HF is formed by the shorter 2 helices of the HSD. Structure in (D) and inspiration for others from Zimmer et al.[39].

SecA consists of two nucleotide binding domains, a helical wing domain, a helical scaffold domain (HSD), and a polypeptide-cross-linking domain (PPXD)[8,40]. The PPXD acts as a clamp that traps the polypeptide[41], holding it in an extended conformation during translocation[10]. The two shorter helices of the HSD are called the two-helix finger (2HF) and they sit within the cytoplasmic funnel of SecY with the loop connecting the two helices directly above the SecY pore ring (Fig. 1 C&D)[8]. Protein cross-linking studies also confirm the close proximity of the 2HF with the substrate polypeptide during active translocation[9]. The HSD is linked to a nucleotide binding domain and may transmit conformational changes to the 2HF[8]. Mutation of residues in the loop of the 2HF indicate that a tyrosine (or a similarly bulky and hydrophobic residue) is required for translocation activity[9]. From all this data, it is presumed that the 2HF is responsible for actively pushing the polypeptide through the channel, though one study indicates that cross-linking the 2HF of SecA to SecY does not abolish translocation [13].

1.2 Our approach: single-molecule Förster resonance energy transfer

Single-molecule Förster resonance energy transfer (smFRET) is a powerful tool to measure the distance between two fluorescent dyes (a donor and an acceptor) by measuring the efficiency of energy transfer between them (the FRET efficiency)[42]. The donor is excited by laser light, and then some fraction of that energy is transferred to the acceptor while the rest is emitted by the donor. Thus the relative intensities can be used to calculate the FRET efficiency. The FRET efficiency is strongly dependent on the distance between the two dyes,

$$\text{FRET efficiency} = \frac{\text{acceptor intensity}}{\text{donor intensity} + \text{acceptor intensity}} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}$$

where R is the distance between the dyes and R_0 is the distance at which the FRET efficiency is one half (the Förster radius) and is dependent on the properties of the two dyes. The strong dependence of FRET efficiency on inter-dye distance allows for precision measurements of distances between about 2nm and 8nm[42] which is ideal for studying many conformational changes in both DNA and proteins. smFRET has been used extensively to observe the conformational dynamics of the ribosome[43-45]. However, it has not been used to characterize conformational changes in the other driver of SecYEG-mediated translocation: the SecA ATPase.

Previous biochemical work and even x-ray crystallographic studies suffer from the limitation that they average over an ensemble of molecules. In order to get out vital mechanistic information, the system must be locked in different conformations or perturbed rather severely. In smFRET, we measure the distance between the donor and acceptor dyes in real time, for individual molecules, in a minimally perturbed system. This is why the proposed study is well-positioned to resolve the role of the 2HF in the coupling of ATP hydrolysis to polypeptide translocation.

1.3 Specific aims

Aim 1. Demonstrate ability to measure dynamic conformational changes with smFRET

A smFRET microscope was newly built when I joined the lab. I will write software to run the microscope and also analyze the data it acquires. I will then repeat two smFRET experiments in the literature to validate our smFRET microscope as comparable to the current

state of the art: one without dynamic FRET changes[14] and one with dynamic FRET changes[15]. I will then use simulated data to develop analytical tools to prepare to analyze and interpret smFRET data from a complex molecular motor.

Aim 2. Use smFRET to characterize ATP-dependent conformational changes in SecA

I will collaborate with Benedikt Bauer from the Rapoport Lab who has biochemical expertise with the SecA/SecYEG system[10,11]. We will purify labeled SecA/SecYEG complexes and embed them in nanodiscs with biotinylated lipids. These can then be immobilized on a glass slide and subjected to different buffer conditions as we take data with our smFRET microscope. We will track the motion of the 2HF in real time when subjected to different nucleotide states. This data can be used to analyze transitions and differentiate between existing models for the 2HF.

1.4 Significance

Recently, our understanding of ATPases has begun to evolve from a machine that cycles through relatively static conformations based on nucleotide state, to one in which the ATPase continuously explores all of its conformation space and the nucleotide state merely alters the bias of this exploration[46]. Thus far, the majority of the support for this change in understanding has come from molecular dynamics (MD) simulations[46], but the proposed study will provide an empirical visualization of SecA exploring its conformation space. We will be able to directly test the conceptual framework suggested by MD simulations by observing conformational changes in SecA in different nucleotide states.

We plan to use smFRET in an innovative way. While smFRET has been used to observe the conformational changes of the ribosome and tRNAs during translation[47,48], it has not been used in a similar way on relatively small processive ATPases like SecA. We hope that our success encourages others to use similar methodology to decipher the conformational changes of many more enzymes.

In addition, our results add to a body of evidence helping to resolve the dispute over the role of the SecA 2HF in the translocation of polypeptides. Our functional assay can be used to study the SecA/SecYEG system as well as assess the functional impact of small molecules. Our assay and results may provide crucial insight and allow the development of antibiotics which target the SecA/SecYEG system.

2. Single-Molecule Förster Resonance Energy Transfer

2.1 Introduction

Single-molecule Förster resonance energy transfer (smFRET) is the powerful combination of a single-molecule approach with FRET microscopy allowing the detection of nanometer-scale distance changes on individual substrates with high time resolution (~100 ms). We intend to use this approach to measure conformational changes in the SecA 2HF. First, we must create and validate a microscope and software package capable of performing such experiments and analyzing the collected data.

We will begin with a brief discussion of the microscope components used and the software which drives the system and records the movies. We will then review the flow cell platform which allows us to sparsely immobilize constructs for direct observation with the microscope. We will then discuss the two sets of control experiments which demonstrate the ability of our smFRET setup to generate publication-quality data. The first is a static FRET construct which demonstrates the ability of our microscope to resolve distance differences on the order of 0.34nm. The next experiment utilizes a dynamic FRET construct and demonstrates our ability to detect multiple states and transitions between them.

2.2 Materials and Methods

For a more complete review of materials, methods, and physics involved in a typical smFRET experimental setup the reader is referred to Roy *et al.*[42]. Here, we will only present a brief overview of the setup with more detail on the most salient points and those elements which were developed or optimized by the author.

2.2.1 Optics, flow cell, imaging conditions

2.2.1.1 Optics

Experiments were conducted on an Olympus™ inverted microscope with an Olympus™ UPlanSApo 100x objective with a numerical aperture of 1.40. Two Coherent™ lasers at 532nm (Sapphire™) and 641nm (Cube™) were used to directly excite the Cy3 and Cy5 dyes, respectively. The optics utilized along the beam path are outlined in Figure 2-1A. Each beam passes through a beam expander lens pair and an aperture to remove the edges of the beam where the intensity is non-uniform.

When the excitation beam enters the objective, the beam is focused and refracts such that it undergoes total internal reflection at the boundary between the glass slide and the sample (see Figure 2-1B). This creates an evanescent excitation wave which falls off exponentially, greatly reducing any background fluorescence by only exciting a small volume of sample closest to the objective. This technique is called total internal reflection fluorescence (TIRF) microscopy[42].

Reflected or scattered excitation light, along with fluorescence emissions, emanate from the sample and pass through the objective on their way to our detector. A dual view set-up (see Figure 2-1C) uses dichroic and plane mirrors to separate the two emission wavelengths and then

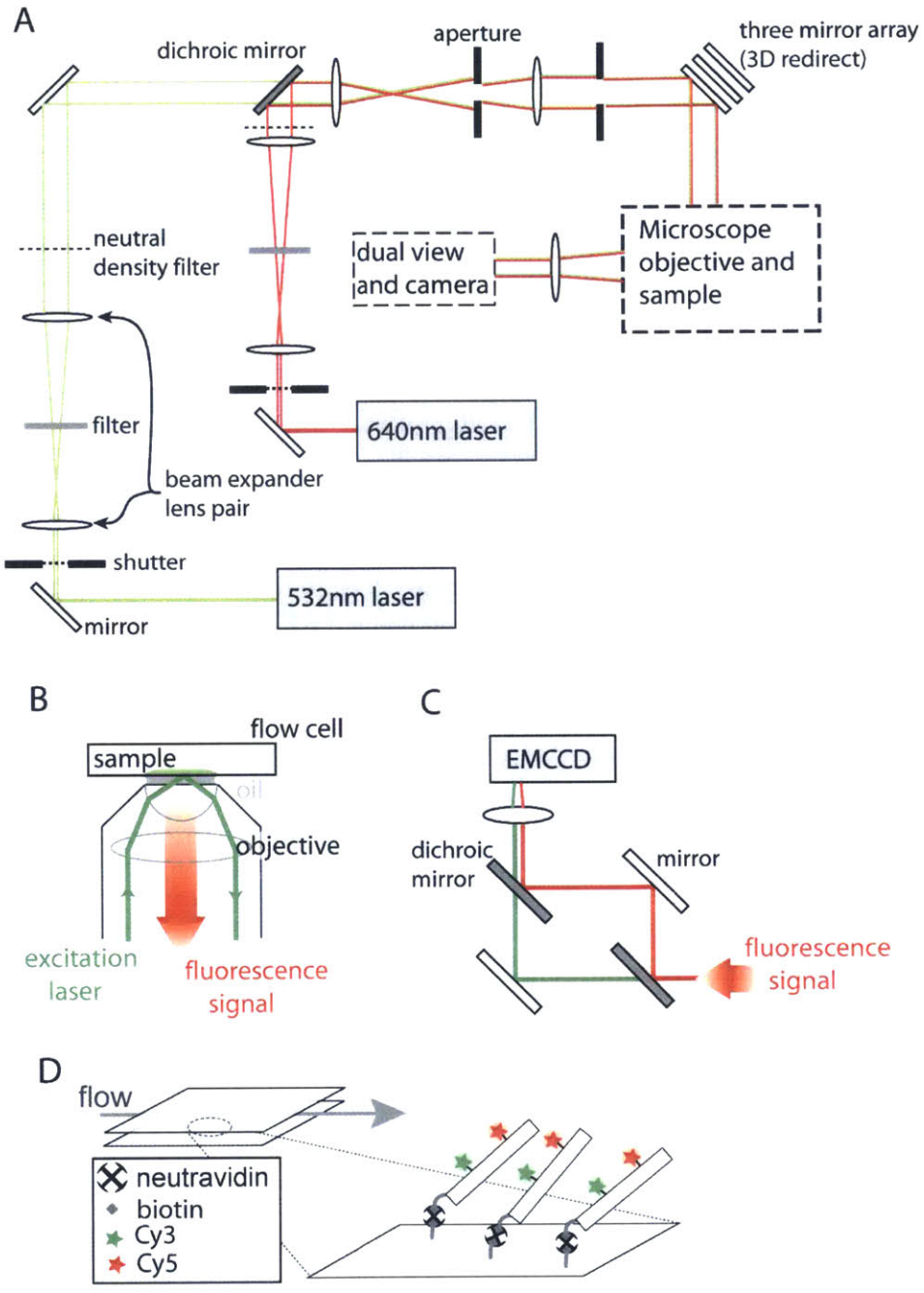


Fig. 2-1: Microscope optics and flow cell platform. (**A**) A diagram showing the optics along the beam path of each of the two lasers. Drawn as though both shutters are open, but one shutter is open at a time during experiments. (**B**) Illustration of how the TIRF evanescent field is achieved. (**C**) Diagram of the dual view and camera setup which separates Cy3 emission wavelengths and Cy5 emission wavelengths and focuses the two resulting images side-by-side on the detector. (**D**) Cartoon illustrating how the flow cell is used to immobilize dually labeled constructs for imaging.

focus them next to each other on the detector. This creates an image where the left side is the signal from Cy3 emission wavelengths and the right side is the signal from Cy5 emission wavelengths. The detector is a Hamamatsu EMCCD camera (model C9100-13).

2.2.1.2 Flow cell for sample imaging

We use a simple flow cell with a functionalized glass coverslip on the bottom to facilitate sparse immobilization of substrates, simplify buffer exchange, and allow all this to be done while imaging on the inverted microscope (see Figure 2-1D). The flow cell is made by cutting a channel out of double-sided tape (a 2 mm by 17 mm rectangle), sandwiching this tape between a quartztop and a functionalized coverslip, and sealing the edges with epoxy. Fluid can be injected or drawn through the channel through holes in the quartztop. The glass coverslip at the bottom of the flow cell is functionalized with a mixture of polyethylene glycol (PEG) and <5% biotin-PEG. The surface of the flow cell is incubated with neutravidin, which then allows us to immobilize biotinylated constructs on the glass while the PEG helps prevent non-specific binding. Our lab has used this approach successfully in many previous single-molecule experiments[49-51]. A detailed protocol for the functionalized cover slips and the flow cell assembly can be found elsewhere[52].

2.2.1.3 Alternating laser excitation

Each beam passes through a Uniblitz™ shutter (model VS14S2TO) which allows us to rapidly switch excitation lasers during an experiment, a technique called alternating laser excitation (ALEX). Code was written in LabView™ to interface with the shutters and the camera through a National Instruments USB-6009 data acquisition card, linking the laser shutters to the camera trigger for each frame. This way, the user can set the number of frames for exposure to each laser and the software will switch between the lasers automatically when the camera begins to fire. The timing of the shutters must be very precise to avoid bleed-through. The timing was optimized using an oscilloscope connected to both a photodiode in the beam path and the output from the camera trigger. Thus the timing could be set in the software to ensure that each laser's on or off occurred within 2ms of the desired camera trigger (see Figure 2-2). This is sufficient precision as the exposure times used are either 50ms or 100ms.

2.2.1.4 Imaging conditions

Photobleaching (when a fluorophore loses its fluorescence after absorbing too many photons) is of great concern in single-molecule fluorescence experiments. We utilize a protocatechuic acid/protocatechuate-3,4-dioxygenase (PCA/PCD) oxygen scavenging system to reduce photobleaching[53] and add a triplet state quencher (Trolox) to reduce blinking[54]. PCA, PCD, and Trolox are added directly to the buffer we intend to use during imaging to a final concentration of 10mM, 50nM, and 200μM respectively. They are added between a few minutes and a few hours before imaging to allow the PCA/PCD to reduce the oxygen concentration, but not allow them to lose their activity[53]. However, photobleaching still occurs in our experiments. It will be quite obvious if the donor photobleaches because the spot will disappear, but if only the acceptor photobleaches it will appear as a decrease in FRET efficiency. We overcome this ambiguity by utilizing our ALEX setup to occasionally probe the acceptor dye directly.

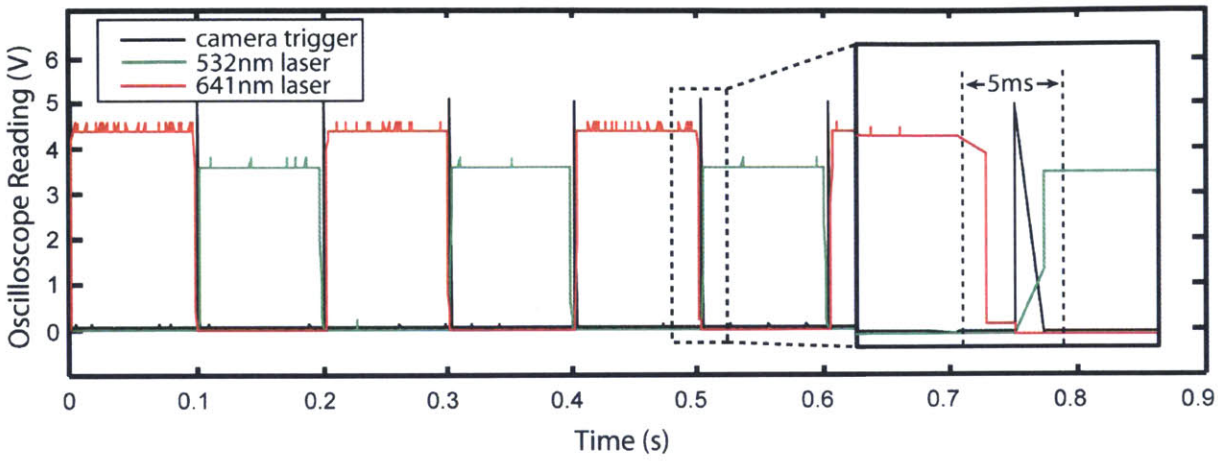


Fig. 2-2: Alternating laser excitation timing optimization. Note that the camera triggers fire at the same time that the lasers switch every 100ms. There is almost no gap time between laser on/ off and the camera trigger (it is less than 2ms).

2.2.2 Image registration of two channels

The images that we collect have Cy3 emission on the left half and Cy5 emission on the right half. However, the two halves of the image represent the same physical space in the sample. In order to make sense of these images, we must be able to transform coordinates from the Cy3 image into the same physical location represented in the Cy5 image. This process is called image registration.

In order to calculate a transformation function, we must have a series of points for which we know the coordinates in both images, called control points. It is best if these points are distributed evenly over the image; perhaps ideally one would have an array of many evenly spaced control points. We accomplished this by shining white light (which is visible in both channels of the dual view) through an array of holes in a metal film. We call this array our nanoGrid. It was nano-fabricated for us by Dr. Daniel Floyd. The holes are so small (~100 nm in diameter) that the bright spots we see are diffraction-limited. The centers of each spot in one channel must be matched with the center of the same spot in the other channel. This is done by taking a separate image of the corner of the array of holes. Pairing the corner hole in each channel generates a linear transformation "guess" that is close enough for us to pair all the spots across the field of view in the original image filled with spots. Next, we can fit the spots to Gaussians and calculate the exact center of each with sub-pixel accuracy. We do this for each one-second exposure in a ten frame movie and average the center coordinates for each spot across the ten frames. Using these precise coordinates of control points from both image channels, we can create a locally-weighted mean transformation function to map coordinate changes between the images[55]. With this method, we typically achieve transformation functions with a target registration error of less than 10 nm, allowing us to confidently colocalize spots appearing in both channels.

2.2.3 Data analysis software

We created a graphical user interface (GUI) in MATLAB™ to assist in and automate the data analysis. The GUI guides the user through the process of creating the image registration function, selecting the desired movie files, identifying local maxima in the movies (bright spots), filtering these spots, fitting the spots to determine their intensity in each frame, and finally selecting intensity traces which correspond to FRET-positive spots with appropriate stoichiometry of dyes.

2.2.3.1 Finding the transformation function

After opening the parent GUI (fretAnalyzeGUI), the user enters the nanoGridGUI where they are guided through selecting the nanoGrid corner image (described in 2.2.2) and evaluating the approximate X and Y linear translation from one channel to the other. Once these values are set, the user will load the movie file of the nanoGrid array (usually 10 frames with 1 s exposure times). It is essential that this movie be acquired under conditions that do not saturate the detector; otherwise the identification and fitting of spots will be quite poor due to the lack of strong local maxima. Next, the user will set the boundaries for the two image channels, removing any overlap. Next, the program will find and pair all spots above a threshold which are within both channel boundaries. Then the user can proceed to fitting the spots, or adjust parameters as needed. Finally the nanoGridGUI will use the fitted centers of the spots to determine the locally weighted mean registration function. In addition, the target registration error (TRE) will be calculated as a measure of the accuracy of the transformation function. This is calculated by removing one pair of control points, calculating a new locally weighted mean transformation function, and using this to predict the location of the missing control point. The error in this prediction is averaged over all control points and this is reported as the TRE. The program reports the TRE in units of pixels; one pixel is about 117nm in the sample. The transformation function and the TRE will be automatically saved (along with all the other parameters) inside the fretAnalyzeGUI figure handle in a data structure called nanoGrid.

2.2.3.2 Analysis settings and movie file selection

Next, the user will return to fretAnalyzeGUI and set the settings in the settings GUI. Here, we set the ALEX scheme (number of frames of each excitation laser), the fitting method, and the spot finding algorithm type, along with several less important and more self-explanatory settings. The fitting method can be set to calculate spot intensities in four different ways: (1, fastRawSum) find the background from a gaussian fit to an average of several frames and estimate intensity using the raw sum of pixel values minus the background, (2, integrated) find the intensity from a gaussian fit to every spot in every frame, (3, rawSum) use a gaussian fit to find the background in every frame but use the raw sum of pixels minus background for intensity, or (4, fastFit) find the intensity from a gaussian fit but fix the center position to the center found at the start of the movie and fix the sigma to the predicted sigma of the point spread function for our microscope geometry. We have had the best results with fastFit as it is less computationally intensive and is able to tolerate noise in the data without being overly sensitive. The spot finding algorithm type can be set to find spots by starting in different channels. The algorithm finds spots in one channel and then checks to see if these spots also exist at that

location in the other channels. We have the best luck with starting in the direct excitation of the acceptor channel (AexAem) as this usually has the lowest background and the best signal to noise ratio.

2.2.3.3 Finding the regions of interest

Then the user will enter the spotFindGUI which guides the user through the process of finding and filtering potential FRET spots. This works by averaging the first several frames of each laser excitation separately in each emission channel. This creates 3 meaningful images: donor excitation donor emission (DexDem), donor excitation acceptor emission (DexAem), and acceptor excitation acceptor emission (AexAem). The DexAem signal is from FRET. There is no AexDem because FRET cannot happen from the acceptor to the donor. The user sets pixel value thresholds for each of these three images. We are typically very permissive with these thresholds as raw pixel value is very susceptible to noise. We set the minimum thresholds just above background (about 2 standard deviations above the mean of the background signal, in practice this is about 1000 above the mean background raw pixel value). We often leave the maximum thresholds set at infinity, and only use them when there are large aggregates in the field of view that must eliminated. The program finds colocalized spots that satisfy the thresholds in all channels by thresholding the images, finding local maxima, and pairing across channels. The result should be a very large number of candidate spots (several hundred per movie). Next, the user sets fitting thresholds (such as spot width, background, and intensity) and the program fits all the spots on the list and only keeps them if they satisfy these fitting thresholds. The most important threshold here is the sigma value as this is the best marker of a true point source. We have had success with using thresholds on the sigma values of 0.5 to 2.0. To set the background thresholds for the fitting, check the approximate value of the background in the raw movie and go a bit above and below this (we usually use a plus or minus 1000 as described above). The output after the application of the fitting thresholds is a much shorter list of spots. All of this data and the parameters for the filtering are saved in the fretAnalyzeGUI figure handle in a data structure called spotFind.

2.2.3.4 Analyzing and filtering the data

Next, the user returns to fretAnalyzeGUI and analyzes the data. This takes quite some time as the program fits every spot, in every channel, for every frame. Next, the user can apply some automated filtering if desired. This allows the user to input various thresholds to be applied to smoothed or raw data which will automatically crop and/or reject traces. This does not alter the raw data, only the filtering indices that produce the final data. We only recommend applying these filters when the output is well-characterized and the user knows exactly what to look for. Additionally, the user can go through each trace and crop or filter the data manually as desired. This allows spots with multiple acceptor dyes (too bright a signal or two step photobleaching) to be filtered out. It also allows each trace to be cropped before photobleaching occurs. All of this data is stored in the fretAnalyzeGUI figure handle; the raw data is stored in the data structure called fretData while the filtered data is stored in a data structure called final. All of the data structures can be saved permanently as a *.mat file from the fretAnalyzeGUI.

We later developed a wrapper script (runme.m) to run this full GUI on a directory full of movie files so long as all the parameters are saved in a file called Tform also within the directory.

This makes analyzing larger data sets much easier. Further details of the software and its use can be found in the program documentation.

2.2.4 Static FRET control experiments

For these experiments, we utilized the 16, 17, and 18 base-pair separation DNA constructs described by Holden *et al.* except that we used Cy3/Cy5 as the donor/acceptor pair instead of Cy3B and ATTO647N[14]. These constructs have a biotin at one end of the DNA duplex to facilitate immobilization and TIRF imaging. They also contain a FRET pair with varying base-pair separation between the two dyes (see Figure 2-3A). The constructs were assembled by annealing and ligating synthetic oligonucleotides ordered from Integrated DNA Technologies, Inc (IDT). All labeling was performed by IDT. These constructs were imaged in a TE buffer (20mM Tris, 2mM EDTA, 50mM NaCl, pH 7.5) with PCA, PCD, and Trolox as described in section 2.2.1.4. Movies were taken with continuous 100ms exposures and the EM gain set to maximum (255). The ALEX scheme used was 5 frames of 532nm excitation followed by 1 frame of 641nm excitation. All data were analyzed by our in-house software package described in 2.2.3.

2.2.5 Dynamic FRET control experiments

For these experiments, we utilized the Holliday junction FRET construct from McKinney *et al.*, ordering the oligonucleotides they report[15] from IDT and annealing them to recreate the Holliday junction construct depicted in Figure 2-4A. We utilized the FRET imaging buffer that they report (10mM Tris, 50mM NaCl, pH 8.0) except that we used the PCA/PCD and Trolox system above instead of the glucose/glucose oxidase system that they report[15]. Movies were taken with continuous 100ms exposures and the EM gain set to maximum (255). The ALEX scheme used was 5 frames of 532nm excitation followed by 1 frame of 641nm excitation. All data were analyzed by our in-house software package described in 2.2.3.

2.3 Results

2.3.1 Static FRET control results

Utilizing the dually labelled DNA duplexes described in 2.2.4 and depicted in Figure 2-3A, the data analysis software package easily picked out many FRET-positive spots in each field of view (see Figure 2-3B). The program then followed those spots through each frame of the movie to create donor and acceptor intensity traces as well as FRET efficiency traces (see Figure 2-3C). These traces maintain a constant FRET value with small oscillations around a mean. The FRET values obtained by following each spot in one field of view for one second are binned into histograms shown in Figure 2-3C. Note that the mean FRET efficiency value of each histogram shifts lower as the distance between the donor and acceptor fluorophores increases by 0.34nm with each additional base-pair.

2.3.2 Dynamic FRET control results

We tracked dynamic FRET changes due to stacking conformer exchange in the Holliday junction construct depicted in Figure 2-4A. Traces demonstrated transitions between two FRET states as shown in Figure 2-4B. The donor and acceptor intensities are well anti-correlated in these transitions. The histogram of FRET efficiencies shows two peaks (see Figure 2-4C), one for each of the states seen in the single-molecule trajectories. While the static FRET constructs generated histograms well fit by a single gaussian (see Figure 2-3C), the dynamic Holliday junction construct generates a histogram which is fit very poorly by a singe gaussian (see Figure 2-4C). However, it is fit well by the sum of two gaussian distributions, one centered at 0.19 and the other centered at 0.73.

2.4 Discussion

The static FRET control experiments in Figure 2-3 demonstrate our ability to visualize individual dually-labeled molecules and measure their FRET efficiency over time. We are able to measure this FRET efficiency with similar noise to that reported in the literature[14]. Our

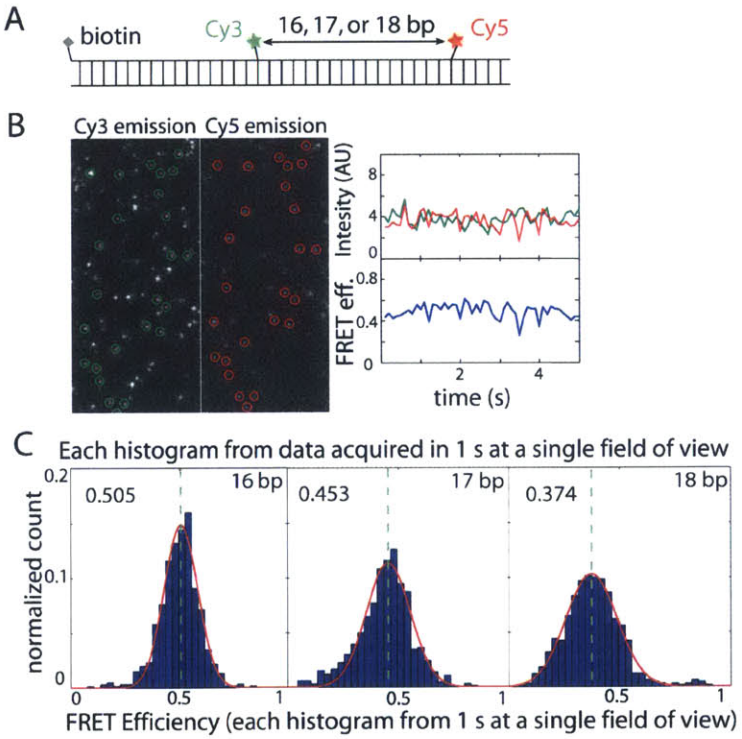


Fig. 2-3: Static smFRET control experiments. (**A**) A cartoon of the DNA constructs used. A biotin allows for immobilization within the flow cell. The donor (Cy3) and acceptor (Cy5) fluorophores are a set distance apart as indicated. (**B**) Sample of our raw imaging data with spots identified in both channels (1/4 field of view shown) and sample single-molecule intensity and FRET efficiency traces. (**C**) FRET efficiency histograms for each of three inter-dye distances collected in one field of view for one second. Even with this small amount of data, we can see the difference in mean FRET due to the different inter-dye distances.

normalized FRET efficiency histograms for the data collected from short DNA duplexes demonstrate our ability to resolve inter-dye distance changes that are the length of a single base pair or 3.4Å (see Figure 2-3C). Our ability to capture this difference with one second of data demonstrates the quality of our time resolution.

The dynamic FRET control experiments in Figure 2-4 demonstrate our ability to measure multiple FRET states explored by a single molecule. We are able to show that two distinct FRET states exist in this population of molecules, and that both states are explored by each molecule. This gives us confidence to proceed to working with more complex experimental systems.

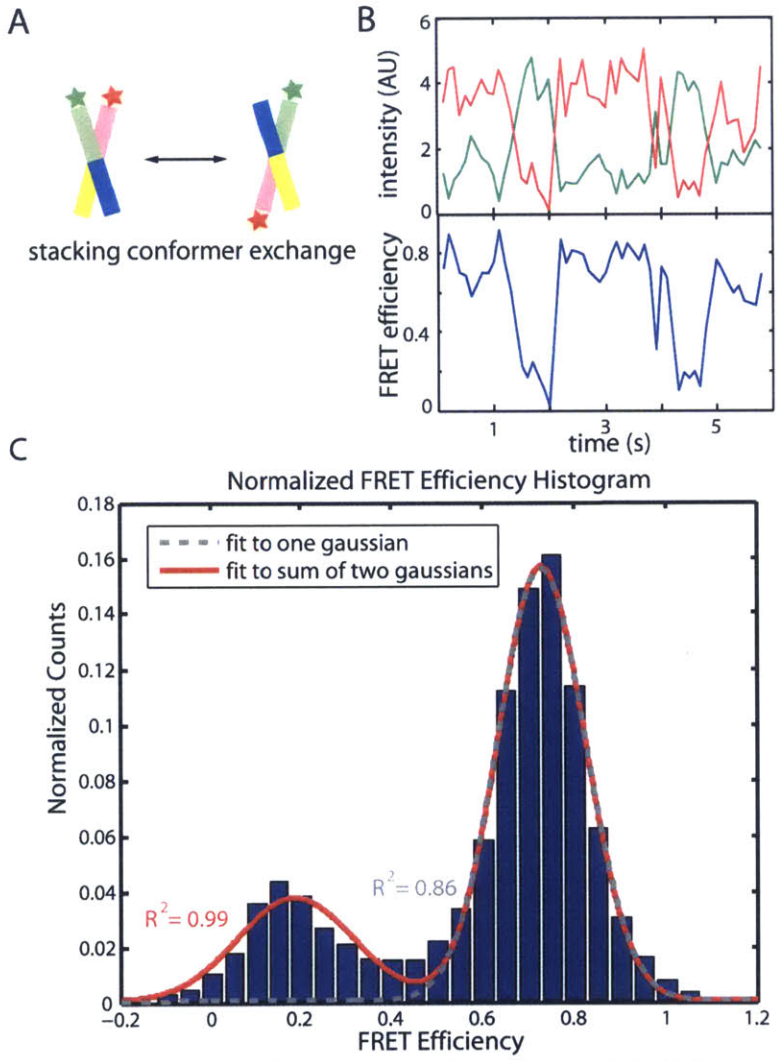


Fig. 2-4: Dynamic smFRET control experiments. (**A**) A cartoon of the DNA constructs used. The donor (Cy3) and acceptor (Cy5) fluorophores are indicated by the colored stars. Cartoon adapted from{McKinney:2005ia} (**B**) Sample single-molecule trajectory showing well anti-correlated Cy3 and Cy5 intensities and transitions between two FRET states. (**C**) Normalized FRET efficiency histogram fin with a one state and a two state model. The two state model has a much higher R^2 value.

Taken together, these results demonstrate our ability to generate publication-quality smFRET data for both static and dynamic FRET systems. In the next chapter we will discuss better ways to analyze dynamic FRET traces so that multiple FRET states and transitions between them can be more reliably detected.

3. Hidden Markov Modeling and Analysis of smFRET Trajectories

3.1 Introduction

While the technique of fitting the raw FRET histogram with a sum of gaussians shown in Figure 2-4C allowed us to detect the presence of two states in the system, the approach becomes less viable as the number of FRET states increases. We would like to analyze the individual traces in such a way that each FRET state can be detected as well as each transition between the states. This way we can reliably extract information about the types, frequency, and kinetics of transitions as well the precise values of each FRET state. To do this, we create a model which we think may describe the system (states and transitions) and then optimize the model parameters to best fit the data. We do this for several models containing different numbers of states and then use statistical methods to pick the one which best describes the data with a minimal number of states.

3.1.1 Hidden Markov models

In order to create a potential model for the system, we must make certain assumptions about how the system behaves. In the literature on analysis of dynamic smFRET traces, it is always assumed that the system is Markovian[42]; that is, its behavior at the next time step is solely determined by its current state. Because we always assume that the system is Markovian, we model the system using a hidden Markov model (HMM). A cartoon of a HMM is shown in Figure 3-1 with panel A showing the "hidden" model and panel B showing the data that might come out. It is said to be "hidden" because all we measure is the apparent FRET efficiency, but behind that measurement is a system transitioning between states that we cannot directly observe. Two other assumptions in addition to the Markovian assumption are typical in this HMM: (1) the observed FRET values for each state are gaussian distributions around a constant mean value, and (2) transitions between states are governed by a single transition probability matrix containing the probabilities for each state of remaining in that state or transitioning to any other state. It follows directly from assumption (2) that the transition kinetics are exponential, consistent with a single rate constant dominating the kinetics of each transition. These assumptions are limitations on the power of HMM, but many biological systems behave in accordance with these assumptions. Later in this chapter, we will see what happens with the assumptions start to break down.

One example of a system which we expect to be described well by a HMM is the stacking conformer exchange in the Holliday junction construct described in chapter 2. We expect there to be two states, and transitions between these states likely occur in a single kinetic step (we expect exponential kinetics). The transition probabilities and therefore kinetics are expected to be invariant over time and should not depend on any memory in the system of its prior state path. We will show that a HMM describes this system well in section 3.3.1.

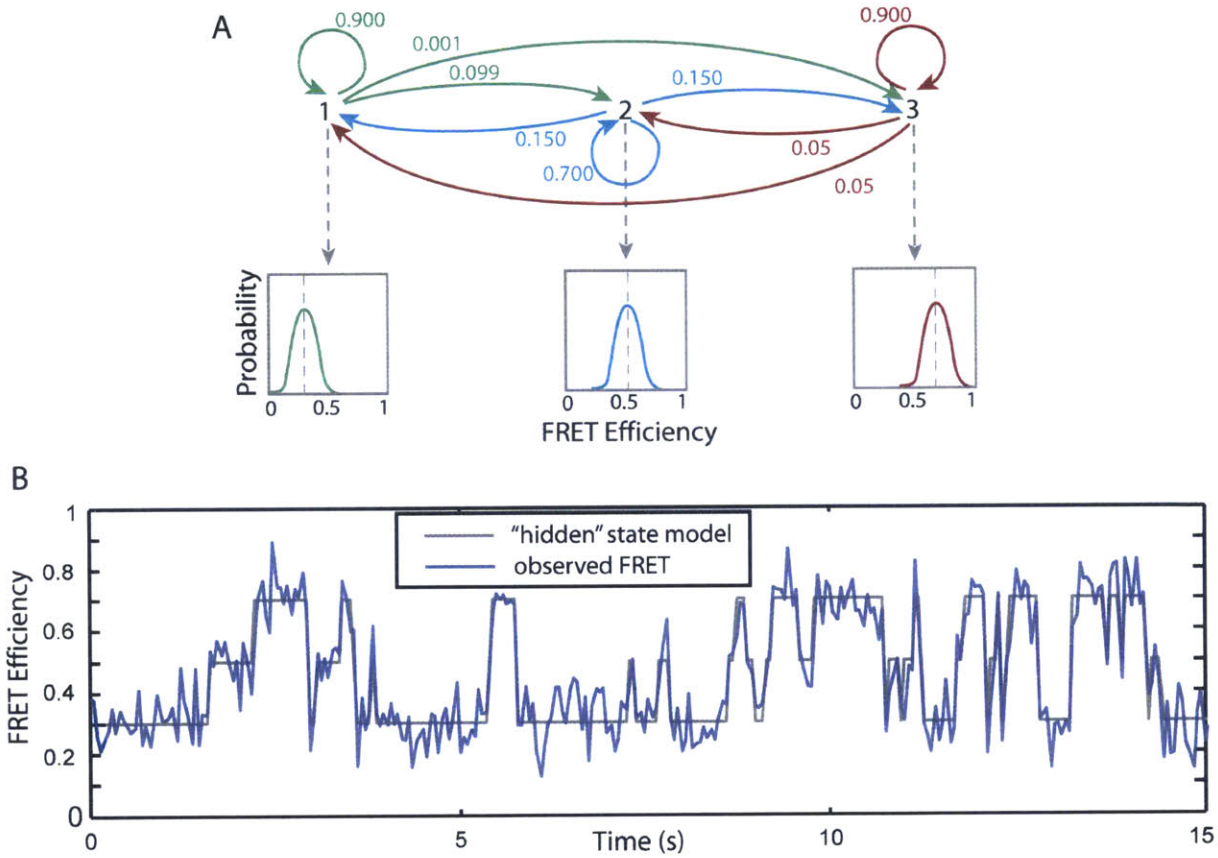


Fig. 3-1: Hidden Markov models (HMMs). (**A**) Diagram of a sample HMM. Three states, and the transition probabilities for each transition are shown. Each time the system is measured, it returns a FRET value selected from the gaussian probability distribution for its current state. (**B**) Sample simulated trace for the HMM in **A**. The state model is in grey and the output is in blue.

3.1.2 Toy Models for smFRET studies of ATPases and other enzymes

We are interested in using smFRET to study a small ATPase: SecA. Due to the defined nucleotide hydrolysis cycle of an ATPase, it is entirely possible that an ATPase has effective memory of its previous state path. Analysis of the resulting data by HMM may fail due to the incongruence with the assumption of a Markovian system. We will explore this by creating a series of toy models, each with four FRET states (we chose four thinking of the nucleotide states: ATP, ADP*P_i, ADP bound, and no nucleotide). Some of these models will be obligate cycles, always moving forward through the defined states, and others will be reversible to varying degrees. Some models will be Markovian, and others will have behavior dependent upon their prior state path. We will see how HMM fits the data simulated from each toy model and try do determine when the HMM breaks down and what can be done in these situations.

3.1.3 Determining the number of states

After fitting data with a series of HMMs with different numbers of states, one must determine which model best describes the data with a minimal number of states. There are three major methods of accomplishing this utilized in the current literature: (1) fitting the raw FRET

histogram to a sum of gaussian distributions and identifying the best fit with the fewest terms[56]as in Figure 2-4C, (2) compiling a histogram of FRET transitions detected by the HMM, fitting to a theoretical model and minimizing the Bayesian information criterion[57], and (3) finding the model which maximizes the evidence (another Bayesian statistic)[58]. We will compare the ability of these methods to accurately determine the number of states in our toy models and discuss the limitations of each.

3.2 Methods

3.2.1 Generation of simulated data

Data was simulated from toy models using code written in MATLAB™ utilizing user-defined transition probabilities and reaction rates. First, the model was simulated, generating a trace with no noise and infinite time resolution. Next, we create simulated experimental data to reflect this model by averaging over the exposure time of each frame, and adding gaussian noise. The amount of noise matches the noise observed in our initial control experiments.

3.2.2 Software packages utilized

Throughout this work, we have tested several different freely available HMM software packages: HaMMy[57], vbFRET[59], and ebFRET[58]. The first uses maximum likelihood to optimize each model and relies on the user to then apply the Bayesian information criterion to find the optimal number of states[57]. Both vbFRET[59] and ebFRET use maximum evidence to optimize the model to the data[58,59]. vbFRET finds the HMM that maximizes the evidence for each trace individually, while ebFRET finds the HMM which maximizes the evidence across all traces for several numbers of states[58,59]. The developers of ebFRET also claim that maximum evidence can be used to determine the appropriate number of states[58]. We have found that vbFRET and ebFRET seem to be the most useful; we use ebFRET for all HMM of data presented in the figures herein.

3.3 Results

3.3.1 Hidden Markov models applied to dynamic FRET control results

When we use ebFRET to fit the Holliday junction data from 2.3.2, we see that the model easily detects the two states and transitions between them (see Figure 3-2A). If we make a 2-dimensional histogram counting each transition such that its initial FRET value is on the horizontal axis and its FRET value after the transition is on the vertical axis (this histogram is called the transition density plot, or TDP), we can see two clear peaks (see Figure 3-2B). This indicates that there are two states and the molecules are transitioning between these states. Additionally, we can look across several HMMs with different numbers of states and see how they perform in maximizing the evidence (see Figure 3-2C). The two-state model maximizes the evidence, confirming what we suspected given the FRET histogram (see Figure 2-4C) and the TDP (see Figure 3-2B). This is how one justifies the selection of one model over another.

This analysis works very well for the simple Holliday junction system. Below, we examine how this approach performs in the context of our toy models and simulated data.

3.3.2 Cyclic models with and without memory

The first toy models we consider are those that move through FRET states on a defined path, always moving in the same direction through a cycle. This is a simple model chosen based on our desire to study an ATPase. Since the ATPase cycle is well-defined and usually thought of as unidirectional, we expect the FRET state cycle to also be well-defined and unidirectional.

Two such irreversible cycles are depicted in the leftmost panels of Figures 3-3A and 3-3B. The first of these moves through four distinct FRET states in a prescribed order. For this toy model, the HMM with four states maximizes the evidence and fits the data well except for occasionally missing transitions which occur too quickly to resolve (see Figure 3-3A). We also note that the TDP is asymmetric because only forward transitions and no reverse transitions are

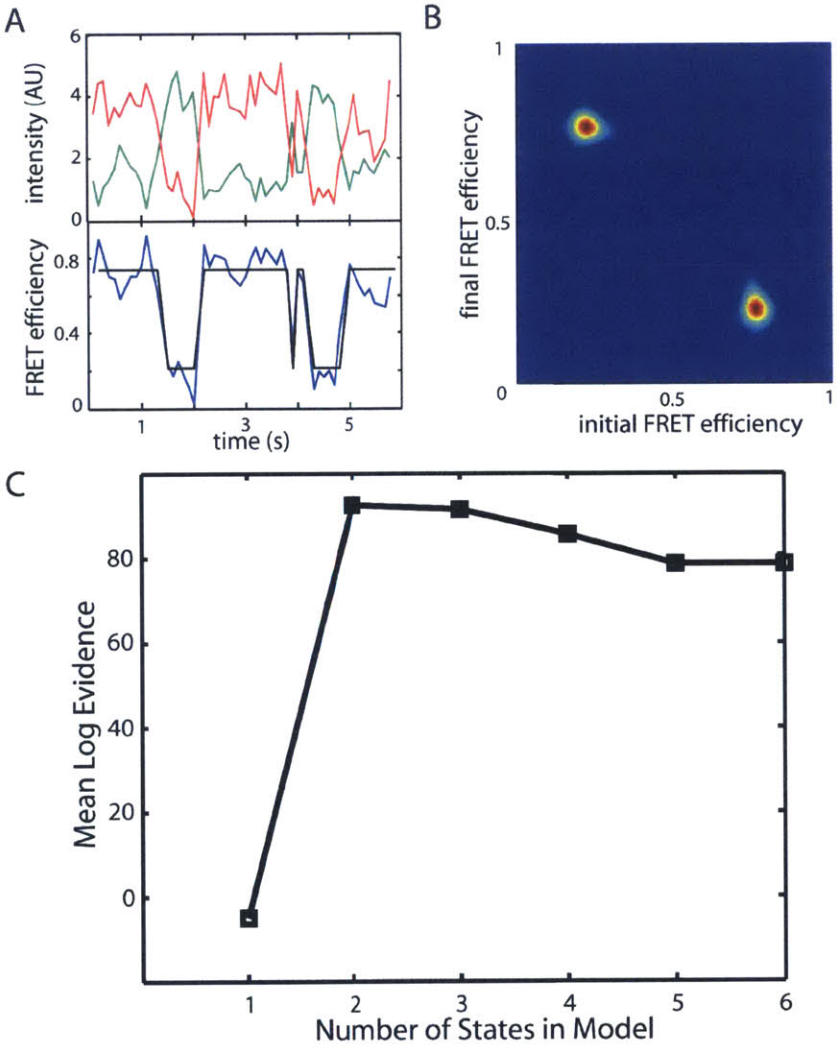


Fig. 3-2: HMM of the dynamic smFRET control data. (**A**) The same sample trace shown in Figure 2-3B but now fit with the 2-state HMM. (**B**) Transition density plot (TDP) tabulating all transitions detected by the 2-state HMM. (**C**) The mean log evidence for HMMs with different numbers of states. The 2-state model maximizes the evidence.

allowed (see Figure 3-3A). This is what we would expect if the SecA 2HF moves in a unidirectional cycle coordinated with ATP hydrolysis.

It is possible, however, that the 2HF could move in such a way that two distinct states on the reaction coordinate have the same FRET value. This case is depicted in the leftmost panel of Figure 3-3B. If we consider the two degenerate FRET states to be identical, then this system has memory, that is, it is non-Markovian. This is because it remembers which direction it is moving (increasing or decreasing FRET). Surprisingly, the HMM with four states still fits the data quite well and matches with the sequential movement through states (the memory of the system is captured), but it maximizes the evidence by a much less convincing margin. The TDP is now symmetric because the second half of the cycle looks like the reverse of the first half. The question now is how to appropriately determine the number of states for a case like this. We will address this in section 3.3.4.

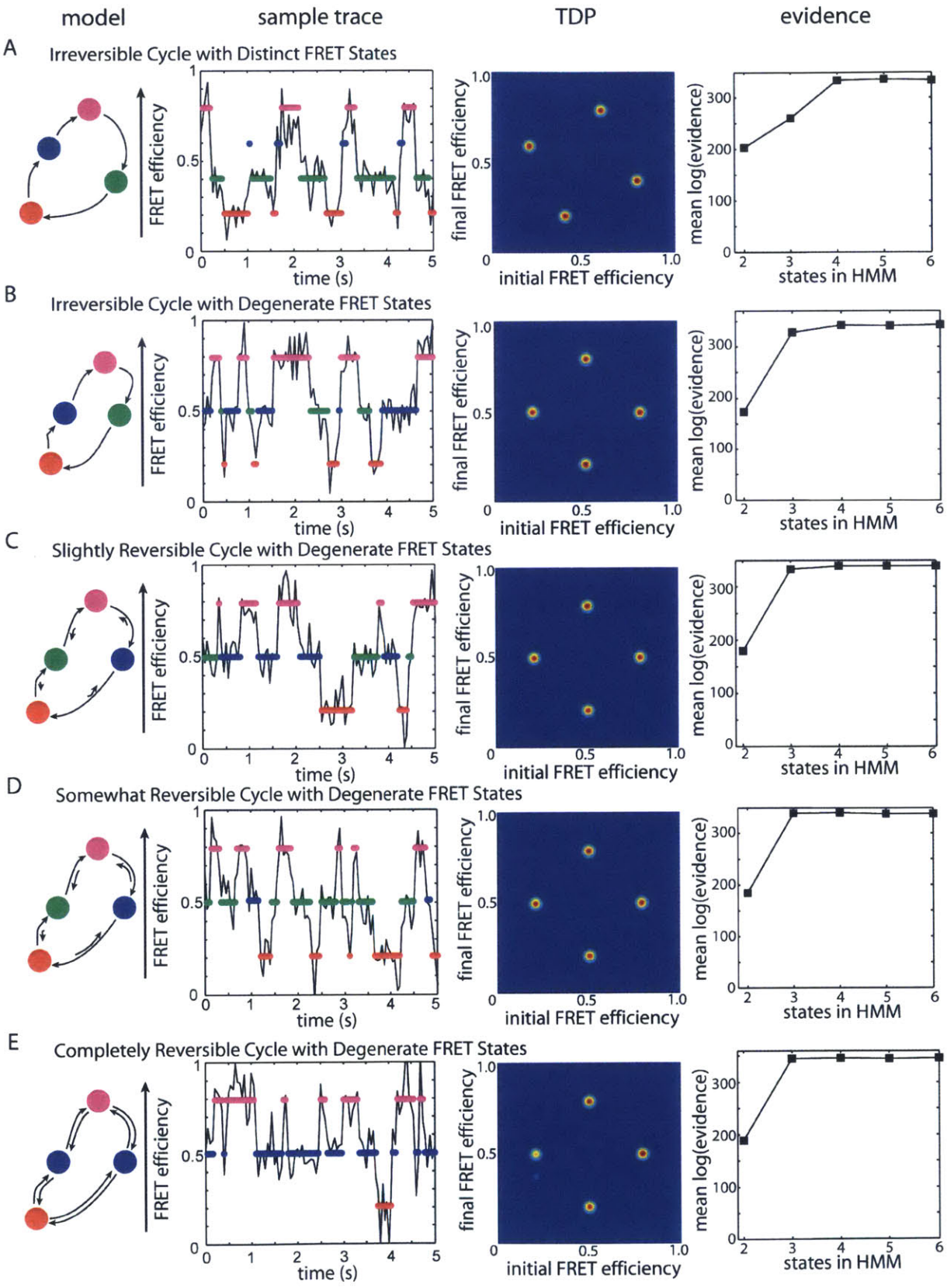
3.3.3 Reversible models with and without memory

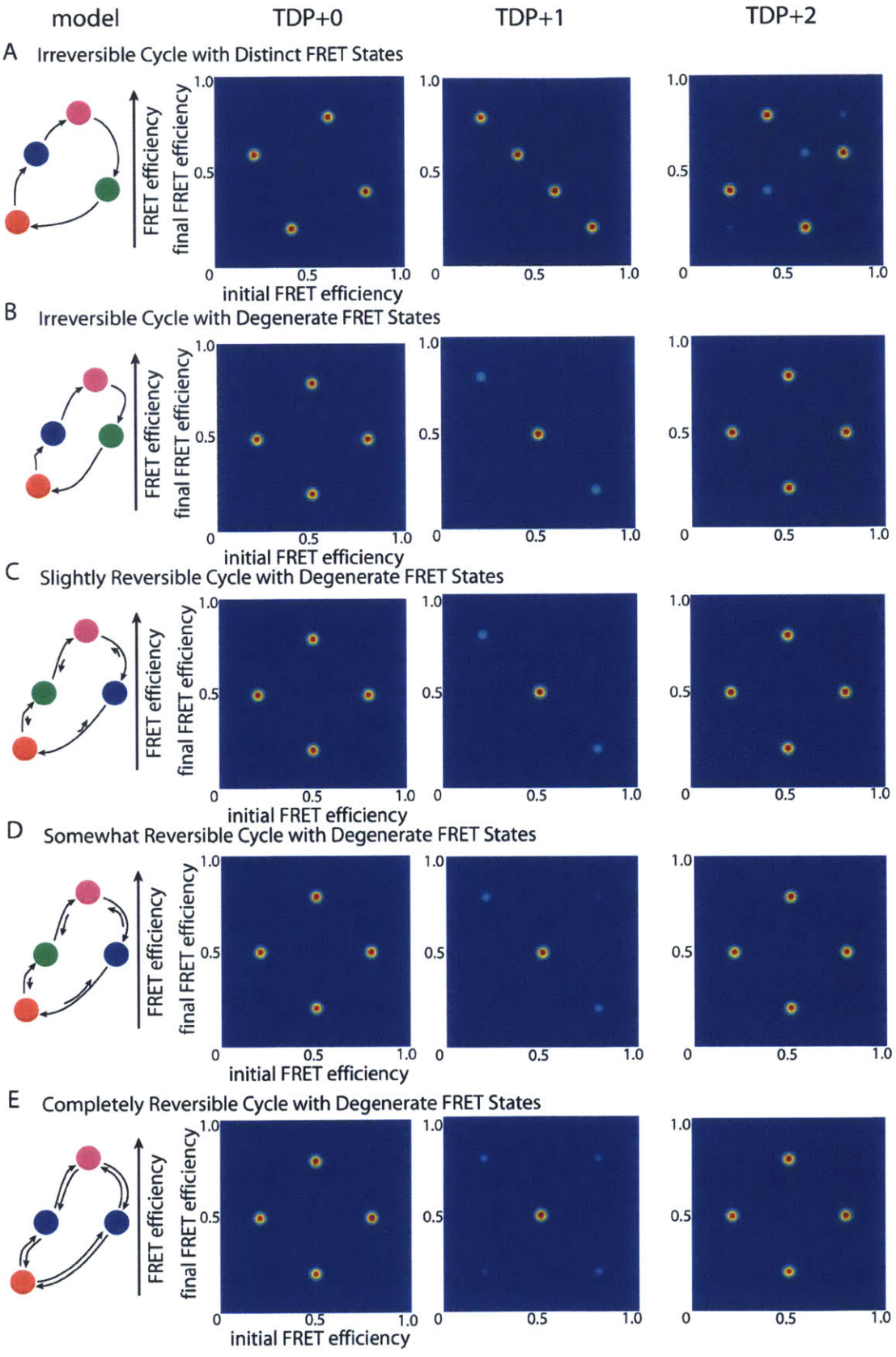
To further probe the limits of HMM in our toy model, we next investigated the effect of adding some reversibility to the cycle. In the "slightly reversible" case (see Figure 3-3C), the system has a 90% chance of moving forward and a 10% chance of moving backward at each transition. For consistency, the traces and TDPs shown are still for the HMM with four states, but we now see that the maximum evidence favors the three-state model. The evidence plots and TDPs look nearly identical for the "somewhat reversible" (75% forward, 25% backward), and the "completely reversible" (50% forward, 50% backward) cases shown in Figures 3-3D and 3-3E respectively. It is worth noting that the "completely reversible" case is truly a three-state model as the two degenerate FRET states are now completely indistinguishable. Again, we ask if we can do anything else to appropriately determine the number of states in these not completely reversible cases.

3.3.4 A new approach in determining the number of states

To rephrase the problem outlined above, current approaches to HMM model selection and number of states determination have difficulty with systems which have memory. Three methods to determine the number of states were presented in section 3.1.3: (1) fitting of the FRET histogram with multiple gaussians, (2) the TDP and Bayesian information criterion approach, and (3) the maximum evidence approach. The first runs into difficulty with larger numbers of states and cannot hope to distinguish between degenerate FRET states. The second relies solely on the TDP so it misses the issue of memory as can be seen by comparing the TDPs in Figure 3-3B through 3-3E. These TDPs are essentially identical yet they represent systems with varying degrees of memory. The third approach uses maximum evidence, but this also fails

Fig. 3-3 (next page): Simulated data. For each model there is a color-coded model cartoon, a sample single-molecule trace, a TDP, and a plot of the mean log(evidence). (**A**) A Markovian irreversible cycle with distinct FRET states. (**B**) A non-Markovian irreversible cycle with degenerate FRET states. (**C**) Same as B but with a 10% probability of going backwards at each transition. (**D**) Same as B but with a 25% probability of going backwards at each transition. (**E**) Same as B but with equal probabilities (50%) of going forward and backward at each transition. This is truly a three state system as the two degenerate states are indistinguishable.


model
sample trace
TDP
evidence
A
Irreversible Cycle with Distinct FRET States
B
Irreversible Cycle with Degenerate FRET States
C
Slightly Reversible Cycle with Degenerate FRET States
D
Somewhat Reversible Cycle with Degenerate FRET States
E
Completely Reversible Cycle with Degenerate FRET States
FRET efficiency
time (s)
final FRET efficiency
initial FRET efficiency
mean log(evidence)
states in HMM


model
TDP+0
TDP+1
TDP+2
A Irreversible Cycle with Distinct FRET States
B Irreversible Cycle with Degenerate FRET States
C Slightly Reversible Cycle with Degenerate FRET States
D Somewhat Reversible Cycle with Degenerate FRET States
E Completely Reversible Cycle with Degenerate FRET States
FRET efficiency
final FRET efficiency
initial FRET efficiency
0
0.5
1.0

to convincingly capture additional states when memory is present (see Figure 3-3B through E).

The memory of the system is evident only in the single-molecule traces where one can observe the ordering to the transitions. To aggregate and present the data in an intuitive way, but retain the information captured by the individual traces, we developed a new TDP that counts later transitions. The normal TDP will be called TDP+0 because it counts each transition and plots it using the FRET value before and after that transition. We will also look at the TDP+1 which looks at each transition and plots the FRET value before the transition and the FRET value not after this transition but after the next transition. We will also look at the TDP+2 which similarly plots each transition with the FRET value before and the FRET value after 3 transitions have occurred. This allows us to see if there is any pattern or bias to the way the molecules are moving between states. In Figure 3-4, we show these three TDPs for each of the toy models presented in Figure 3-3. One difference is that we used the HMM with the minimum number of states suggested by the evidence: four for the first model and three for each of the rest. For the models we tested, the TDP+1 provides a clear metric of the degree of memory. If we consider the 0.2 and 0.8 FRET states and look at the intensity of peaks on the bottom left to the top right diagonal (hereafter called "the diagonal") compared to the intensity of peaks off this diagonal, we see that the on-diagonal peaks are extremely small in systems with memory and the off-diagonal peaks are much larger (see Figure 3-4A and 3-4B). As the system loses memory, intensity is taken away from the peaks off of the diagonal and moved to peaks on the diagonal (see Figure 3-4C through E). This is quantified in Figure 3-5 where we look at slices from the TDP+1 for the 0.2 and 0.8 FRET states. We can now take the ratio of the intensities of the peaks off the diagonal to those on the diagonal. This ratio is close to one for reversible systems (see Figure 3-5 D) but much larger than one for ordered systems (see Figure 3-5A). We believe that this approach can be used in conjunction with maximum evidence to determine the true number of states in the system and if the system demonstrates memory.

3.4 Discussion

3.4.1 Limitations of the transition density plot

The traditional TDP (i.e. TDP+0) is often shown in publications as a way to aggregate and summarize a large number of dynamic single-molecule traces. However, it suffers from limitations that have been exposed in the above simulations. The TDP+0 does not capture any kinetic information, nor does it necessarily capture the order in which a system moves through its FRET states. As shown in Figure 3-3, systems with different patterns of moving through states can all have the same TDP+0. This is especially problematic for those who intend to use the TDP+0 to determine the number of states in the system as has been suggested in the literature[57]. We propose the TDP+n series to visualize the presence of ordering in the movement between states and differentiate between many models that produce the observed TDP+0.

Fig. 3-4 (previous page): A new TDP to visualize transition ordering. For each model there is a model cartoon, a TDP similar to those in Figure 3-3 now called TDP+0, a TDP+1, and a TDP+2. Models (**A**) - (**E**) are the same as in Figure 3-3 and are described in that figure caption.

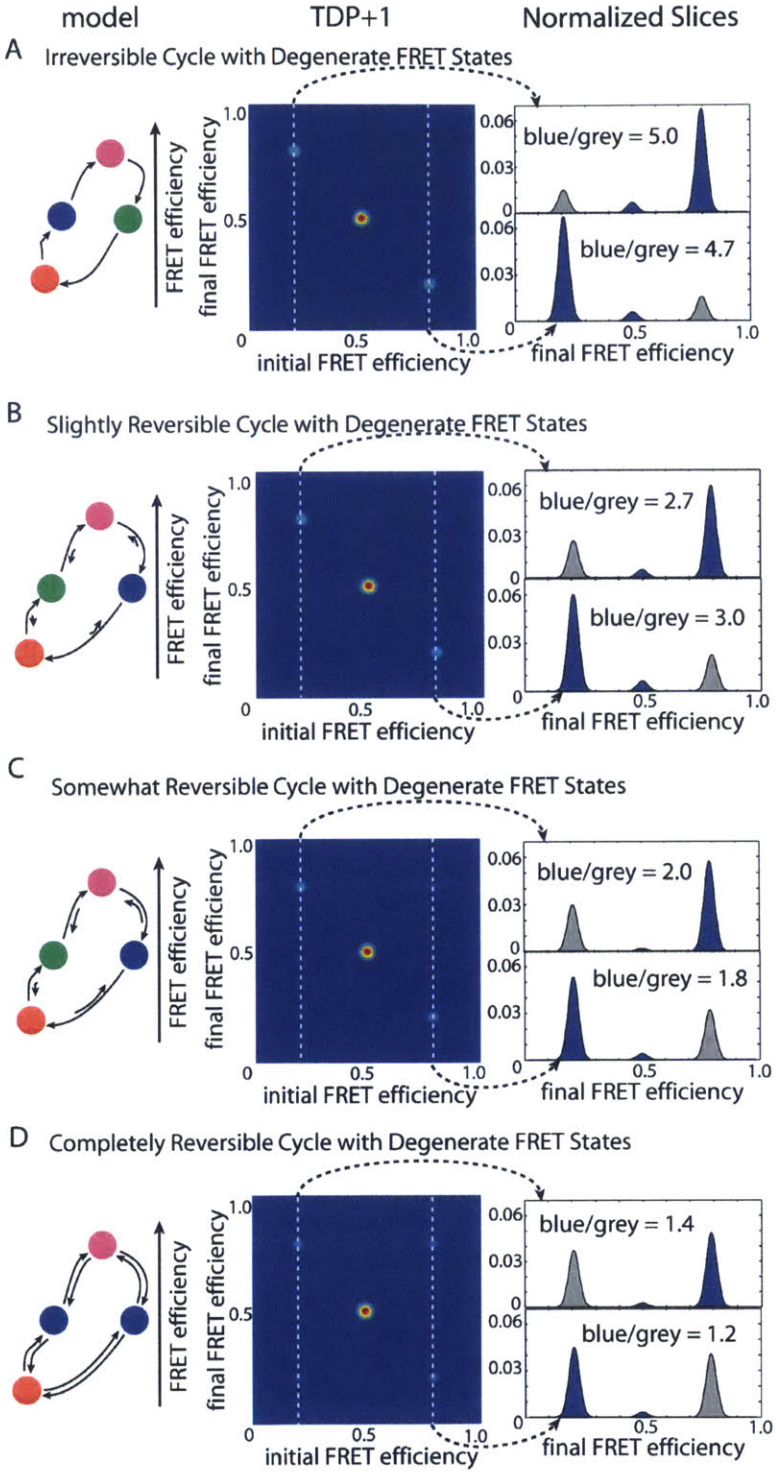


Fig. 3-5: Quantification of the TDP+1 for detection of ordering. For each model there is a model cartoon, the TDP+1 from Figure 3-4 with the location of vertical slices indicated, and a normalized histogram of the data contained in each slice. The ratio of the off diagonal (blue) to the on diagonal (grey) intensity is shown for each slice. Models (**A**) - (**D**) are the same as models **B** - **E** in Figure 3-3, respectively.

3.4.2 Limitations of Markov models

Before performing these simulations, we expected that the HMM software might not be able to fit data from processes with memory because these are by definition non-Markovian. We were pleasantly surprised to find that ebFRET does a good job detecting ordering of states when memory is perfect (see the trace in Figure 3-3B). While ebFRET fits the traces well, the maximum evidence does not strongly suggest the correct number of states (four in our case). However, using the HMM to find the three non-degenerate FRET states, we can then use the TDP+n to identify the presence of ordering or memory in the transitions. This enables us to determine the correct number of states and the ordering of occupancy. All this information was present in the HMM fit but traditional approaches to finding the number of states are not able to capture it.

3.4.3. Determining the number of states

We propose a new method to supplement existing HMM techniques by looking at the data in a new way to detect ordering in the system. The goal is to detect ordering in a state path that would otherwise be hidden by traditional methods of aggregating data (see Figure 3-3). Our approach works under two assumptions that we feel are reasonable for the investigation of ATPase motor proteins: (1) FRET states are stepped through in order with almost no FRET transitions displaying large changes in FRET by skipping over intermediate states, and (2) there are only three FRET states distinguishable by their FRET efficiency. Our approach is likely applicable to systems with more FRET states with some minor alterations, but we will continue to treat only the three state case for simplicity. As for the first assumption, we expect this to hold because an ATPase must be moving through a defined cycle of conformations. If each nucleotide state had a different FRET value, the ordering would be apparent from the traditional TDP (see Figure 3-3A). If each nucleotide state does not have a distinct FRET value, but there are still FRET transitions for each nucleotide state transition, this system is in agreement with our assumption.

Our approach to determine the number of states and the presence of ordering is to (1) analyze the data with ebFRET and choose the HMM which maximizes the evidence, (2) ensure that this HMM satisfies the assumptions above, (3) compile the TDP+1 and take vertical slices at the lowest and highest FRET states, and (4) calculate the ratio of the intensity of the peaks off the diagonal versus those on the diagonal (as in Figure 3-5). There are three cases: (1) both ratios are much greater than one indicating that the system is well-ordered, (2) the *average* of the ratios is close to one indicating that the system is likely completely reversible, and (3) both ratios are much less than one. In case three nearly all of the TDP+1 intensity is on the diagonal indicating that the dominant behavior is to return to the previous state. This describes a system which is reversible, but remembers where it was before and is therefore non-Markovian.

4. smFRET Experiments with the SecA/SecYEG System

4.1 Introduction and working models

With the validated smFRET microscope and data analysis tools described above, we look to directly observe the role of the two-helix finger (2HF) in SecA's ATP-driven translocation of polypeptides through the SecYEG channel. We will dually label the SecA ATPase and directly observe large ATP-dependent conformational changes in the 2HF utilizing our smFRET microscope. Before looking at any data, however, let us explore what we might expect to see and propose two models that we may be able to help distinguish based our data.

SecA is a motor protein, coupling ATP hydrolysis to progressive pushing motion to translocate a polypeptide substrate unidirectionally. The most intuitive model of how this might work is a "unidirectional pushing model" (see Figure 4-1A). The 2HF moves in a defined unidirectional cycle such that it only pushes the substrate in one direction. We expect that the resulting FRET model would be similar to one of of the cyclic models presented in section 3.3.2.

Less intuitive, but equally plausible, is a "separate progress capture model" (see Figure 4-1B). In this model, the 2HF moves back and forth. If that were all that happened, then no translocation would take place. However, in this model we propose that ATP coordinates the timing of the 2HF pushing with a clamp that captures progress such that the clamp is released during a forward push, but engaged during retrograde motion of the 2HF. We expect that the resulting FRET model for this type of motion would be similar to one of the reversible models presented in section 3.3.3.

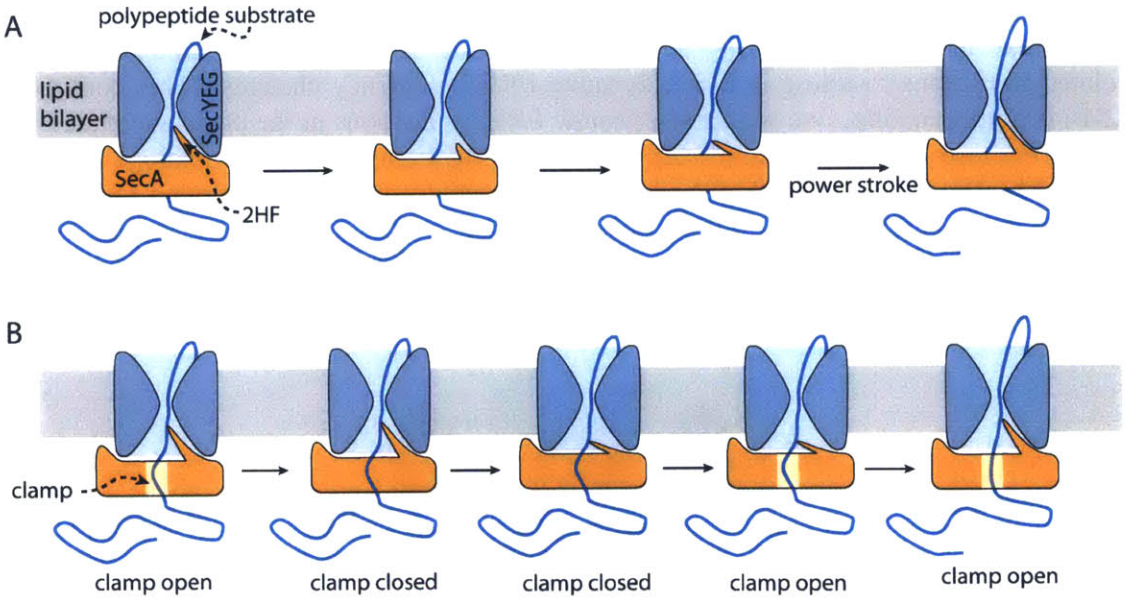


Fig. 4-1: SecA/SecYEG working models. The cartoons match Figure 1, but now the polypeptide substrate is also drawn in. (**A**) A unidirectional pushing model. (**B**) A separate progress capture model where the 2HF moves back and forth in concert with an opening and closing of a substrate clamp.

4.2 Materials and methods

4.2.1 Protein labeling, purification, and experimental conditions

This portion of the work was carried out by our collaborator, Dr. Benedikt Bauer, in the lab of Dr. Tom Rapoport (professor of cell biology at Harvard Medical School and an HHMI investigator). A full description of sample preparation and imaging conditions can be found in Dr. Bauer's PhD thesis.

Briefly, SecA was dually labeled at the 2HF and the PPXD. Labeled SecA was purified in complex with SecYEG and polypeptide substrate. This complex was then embedded in nanodiscs containing a lipid bilayer with a small fraction of biotinylated lipids (see Figure 4-2). In preparation for imaging, the complexes were immobilized in a flow cell (described in section 2.2.1.2) via the biotinylated lipids. Next, imaging was conducted in an imaging buffer with PCA, PCD, and Trolox, as well as either ATP or ADP Beryllium Fluoride (ADP BeFx).

4.2.2 Data analysis

Data analysis was performed using the software described above: the package developed by the author for identifying spots and fitting them to create traces, and ebFRET to perform the HMM. To be sure we only accepted traces from actively translocating complexes, only those traces with more than two transitions were accepted in the final analysis.

4.3 Results

Donor and acceptor intensity traces for individual SecA/SecYEG complexes show anti-correlated transitions resulting in large, dynamic FRET efficiency changes. This confirms that the 2HF is highly mobile. We see these dynamic FRET transitions in the presence of ATP, as we expected, but also in the presence of ADP BeFx (a tightly bound nucleotide analog) as shown in Figure 4-3A. We expected SecA to be locked in one conformation when bound to ADP BeFx, but instead find that it still transitions between FRET states.

Next, we compiled TDPs and raw FRET histograms (see Figure 4-3B). The most striking feature of the TDPs is that they are both very symmetric, showing that forward and reverse

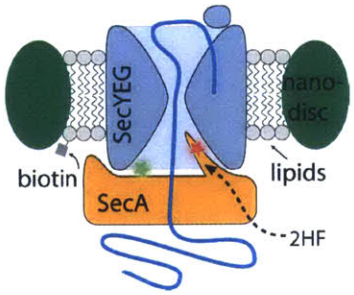


Fig. 4-2: SecA/SecYEG smFRET construct. A cartoon the dually-labeled SecA/SecYEG/ substrate complex embedded in a lipid bilayer supported by a nanodisc (toroidal shaped protein scaffold). The approximate location of the Cy3 and Cy5 labels are indicated.

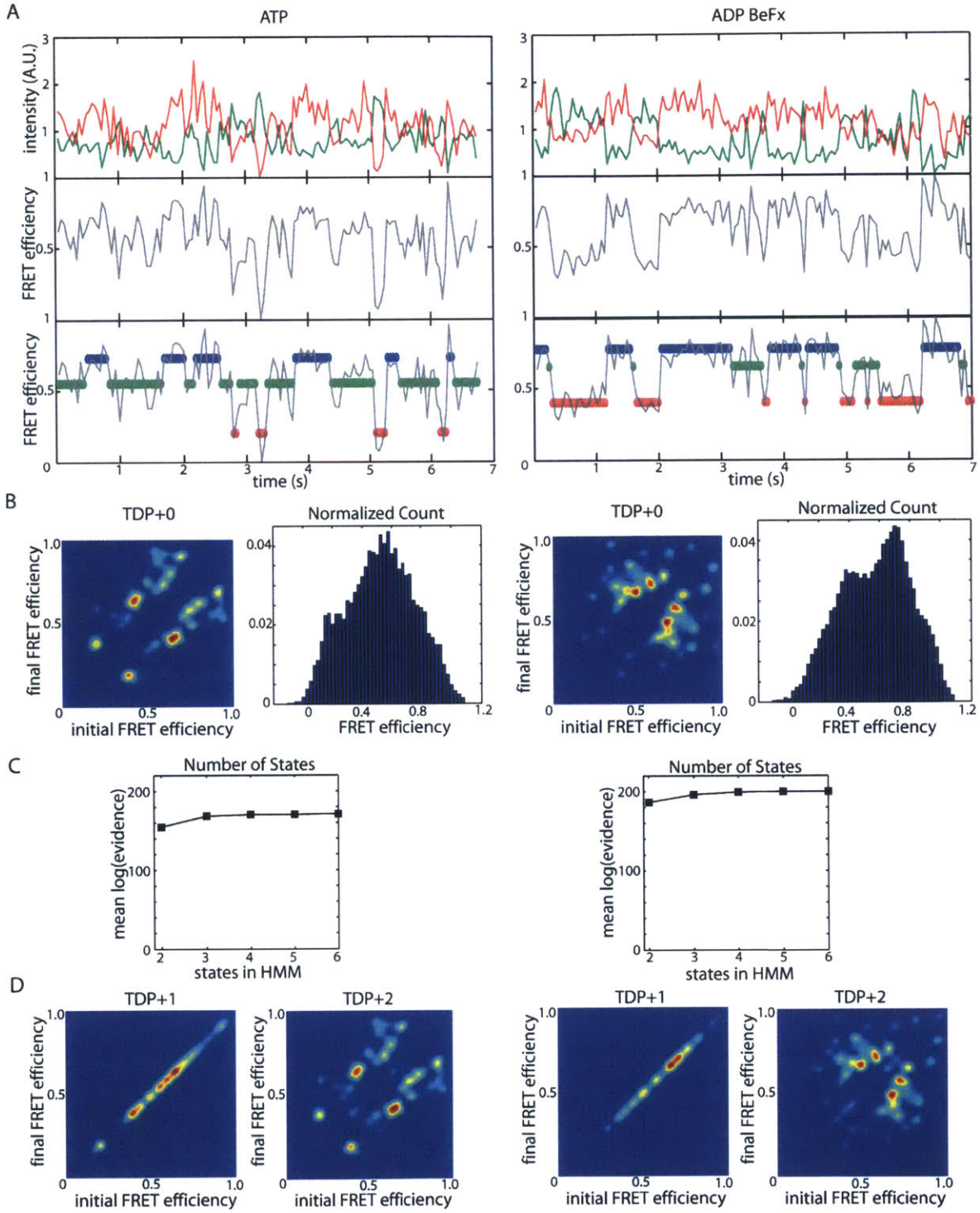


Fig. 4-3: SecA/SecYEG preliminary results. (**A**) Sample smFRET traces for the 2HF of SecA and their fit by the HMM in the presence of ATP or ADP BeFx as indicated. (**B**) Traditional TDP and normalized raw FRET histogram for the ATP (50 traces, 1100 transitions) and ADP BeFx (60 traces, 1600 transitions) data sets. (**C**) Plot of the mean log(evidence) for each data set. (D) TDP+1 and TDP+2 for each data set utilizing the 3-state HMM output.

transitions happen with equal frequency. We also notice that the TDP for ATP is spread out along either side of the diagonal. This is consistent with the 2HF actively exploring a large conformational space by taking smaller but discrete steps along the way. The large peaks in the TDP for ADP BeFx are restricted to a much smaller range of FRET values. This is consistent with the 2HF being much more restricted in its exploration. It can still access all the FRET states as we can see from the width of the raw FRET histogram, but it ventures out much less frequently.

The maximum evidence method of selecting the number of states indicates a three state model (see Figure 4-3C), but we showed in chapter 3 that this may be misleading. The TDP+n series is shown in Figure 4-3D indicates no detectable ordering to the transitions. We can see from the TDP+1 in particular that the dominant activity is always to return to the state from which you just came. This is consistent with the 2HF diffusing mostly randomly through FRET states.

4.4 Discussion

Preliminary results from smFRET studies of SecA/SecYEG indicate that the 2HF undergoes large conformational changes in the presence of ATP. In the presence of ADP BeFx, this same conformational space is explored, but much less actively. The 2HF is restricted more to intermediate FRET states rather than readily exploring the two extremes. These results provide direct evidence of the relatively new model for ATPases presented in section 1.4: that ATPases continuously explore all of their conformational space and the nucleotide state merely alters the bias of this exploration[46]. Our results also show that the 2HF of SecA undergoes large ATP-dependent conformational changes consistent with it playing a central role in polypeptide translocation.

We also show that the 2HF seems to diffuse through these FRET states in a mostly random manner, with the FRET change of each transition being small compared to the range of FRET values explored. This suggests that the 2HF diffuses back and forth between two extreme conformations, but with intermediate FRET states along the way making multiple stopping points on this random walk. This is inconsistent with the unidirectional pushing model proposed in Figure 4-1A, and suggests that the separate progress capture model (see Figure 4-1B) is more likely.

5. General Conclusions

5.1 Overview of current progress

We have successfully developed and validated a smFRET microscope as well as the software tools to run the microscope and analyze the data collected. We have also begun to test the limits of HMM software using simulated data and developed a new way to aggregate and visualize data to assist in the interpretation of the number of states and the elucidation of the "hidden" model even when it is non-Markovian.

We were able to apply these tools to experiments conducted on the SecA/SecYEG system to investigate the role of the 2HF in polypeptide translocation. While our results are preliminary, they support several key insights: (1) the 2HF undergoes large ATP-dependent conformational changes, (2) the 2HF continuously explores all of its conformational space regardless of nucleotide state but the relative populations of these conformations are altered by nucleotide state, (3) the motion between states is mostly random which is consistent with the separate progress capture model (see Figure 4-1B).

5.2 Future directions

The TDP+n is currently a valuable tool for the researcher to quickly assess and share qualitative results, but this approach could be made more rigorous. More work is needed on the TDP+n concept to develop tests of statistical significance related to the ratios calculated from the slices in Figure 3-5.

Further experiments are needed on the SecA/SecYEG system to establish reproducibility, improve signal-to-noise, and provide additional statistical power for fitting and conclusions. It is also necessary to compare our results with a control in which SecA is dually labeled but not on the 2HF. This should confirm that ATP-dependent conformational changes occur much more actively at the 2HF than they do elsewhere in SecA. This control experiment will provide vital support for our conclusions about the role of the 2HF.

References and Literature Cited

1. Cooper, M. A. & Shlaes, D. Fix the antibiotics pipeline. *Nature* **472,** 32–32 (2011).
2. White, A. R. Effective antibacterials: at what cost? The economics of antibacterial resistance and its control. *Journal of Antimicrobial Chemotherapy* **66,** 1948–1953 (2011).
3. Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat Rev Micro* **12,** 35–48 (2014).
4. Poehlsgaard, J. & Douthwaite, S. The bacterial ribosome as a target for antibiotics. *Nat Rev Micro* **3,** 870–881 (2005).
5. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nat Rev Micro* **8,** 423–435 (2010).
6. Lee, V. T. & Schneewind, O. Protein secretion and the pathogenesis of bacterial infections. *Genes & Development* **15,** 1725–1752 (2001).
7. Park, E. & Rapoport, T. A. Mechanisms of Sec61/SecY-Mediated Protein Translocation Across Membranes. *Annual Review of Biophysics* **41,** 21–40 (2012).
8. Zimmer, J., Nam, Y. & Rapoport, T. A. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455,** 936–943 (2008).
9. Erlandson, K. J. *et al.* A role for the two-helix finger of the SecA ATPase in protein translocation. *Nature* **455,** 984–987 (2008).
10. Bauer, B. W. & Rapoport, T. A. Mapping polypeptide interactions of the SecA ATPase during translocation. *Proceedings of the National Academy of Sciences* **106,** 20800–20805 (2009).
11. Bauer, B. W., Shemesh, T., Chen, Y. & Rapoport, T. A. A 'Push and Slide' Mechanism Allows Sequence-Insensitive Translocation of Secretory Proteins by the SecA ATPase. *Cell* **157,** 1416–1429 (2014).
12. Osborne, A. R. & Rapoport, T. A. Protein translocation is mediated by oligomers of the SecY complex with one SecY copy forming the channel. *Cell* **129,** 97–110 (2007).
13. Whitehouse, S. *et al.* Mobility of the SecA 2-helix-finger is not essential for polypeptide translocation via the SecYEG complex. *The Journal of Cell Biology* **199,** 919–929 (2012).
14. Holden, S. J. *et al.* Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy. *Biophysical Journal* **99,** 3102–3111 (2010).
15. McKinney, S. A., Freeman, A. D. J., Lilley, D. M. J. & Ha, T. Observing spontaneous branch migration of Holliday junctions one step at a time. *Proceedings of the National Academy of Sciences* **102,** 5715–5720 (2005).
16. Egea, P. F. & Stroud, R. M. Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proceedings of the National Academy of Sciences* **107,** 17182–17187 (2010).
17. van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427,** 36–44 (2003).
18. Breyton, C., Haase, W., Rapoport, T. A., Kühlbrandt, W. & Collinson, I. Three-dimensional structure of the bacterial protein-translocation complex SecYEG. *Nature* **418,** 662–665 (2002).

19. Duong, F. Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. *The EMBO Journal* **16,** 2756–2768 (1997).
20. Brundage, L., Hendrick, J. P., Schiebel, E., Driessen, A. J. & Wickner, W. The purified E. coli integral membrane protein SecYE is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* **62,** 649–657 (1990).
21. K Nishiyama, M. H. H. T. Disruption of the gene encoding p12 (SecG) reveals the direct involvement and important function of SecG in the protein translocation of Escherichia coli at low temperature. *The EMBO Journal* **13,** 3272 (1994).
22. Park, E. & Rapoport, T. A. Preserving the membrane barrier for small molecules during bacterial protein translocation. *Nature* **473,** 239–242 (2011).
23. van den Burg, B. & Eijsink, V. Selection of mutations for increased protein stability. *Current opinion in biotechnology* (2002).
24. Hizlan, D. *et al.* Structure of the SecY complex unlocked by a preprotein mimic. *Cell reports* **1,** 21–28 (2012).
25. Bessonneau, P., Besson, V., Collinson, I. & Duong, F. The SecYEG preprotein translocation channel is a conformationally dynamic and dimeric structure. *The EMBO Journal* **21,** 995–1003 (2002).
26. Deville, K. *et al.* The Oligomeric State and Arrangement of the Active Bacterial Translocon. *Journal of Biological Chemistry* **286,** 4659–4669 (2011).
27. Kedrov, A., Kusters, I., Krasnikov, V. V. & Driessen, A. J. M. A single copy of SecYEG is sufficient for preprotein translocation. *The EMBO Journal* **30,** 4387–4397 (2011).
28. Haßdenteufel, S., Klein, M. C., Melnyk, A. & Zimmerman, R. Protein transport into the human ER and related diseases, Sec61-channelopathies 1. *Biochem. Cell Biol.* **92,** 499–509 (2014).
29. Lloyd, D. J., Wheeler, M. C. & Gekakis, N. A Point Mutation in Sec61 1 Leads to Diabetes and Hepatosteatosis in Mice. *Diabetes* **59,** 460–470 (2010).
30. Lu, Z. *et al.* Glioblastoma Proto-oncogene SEC61 Is Required for Tumor Cell Survival and Response to Endoplasmic Reticulum Stress. *Cancer Research* **69,** 9105–9111 (2009).
31. Liebermeister, W., Rapoport, T. A. & Heinrich, R. Ratcheting in post-translational protein translocation: a mathematical model. *Journal of Molecular Biology* **305,** 643–656 (2001).
32. Ding, H., Mukerji, I. & Oliver, D. Nucleotide and phospholipid-dependent control of PPXD and C-domain association for SecA ATPase. *Biochemistry* **42,** 13468–13475 (2003).
33. Hunt, J. F. *et al.* Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. *Science* **297,** 2018–2026 (2002).
34. Gelis, I. *et al.* Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. *Cell* **131,** 756–769 (2007).
35. Gouridis, G., Karamanou, S., Gelis, I., Kalodimos, C. G. & Economou, A. Signal peptides are allosteric activators of the protein translocase. *Nature* **462,** 363–367 (2009).
36. Lill, R., Dowhan, W. & Wickner, W. The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. *Cell* **60,** 271–280 (1990).
37. Erlandson, K. J., Or, E., Osborne, A. R. & Rapoport, T. A. Analysis of Polypeptide Movement in the SecY Channel during SecA-mediated Protein Translocation. *Journal of Biological Chemistry* **283,** 15709–15715 (2008).

38. Robson, A., Gold, V. A. M., Hodson, S., Clarke, A. R. & Collinson, I. Energy transduction in protein transport and the ATP hydrolytic cycle of SecA. *Proceedings of the National Academy of Sciences* **106,** 5111–5116 (2009).
39. Zimmer, J. & Rapoport, T. A. Conformational flexibility and peptide interaction of the translocation ATPase SecA. *Journal of Molecular Biology* **394,** 606–612 (2009).
40. Osborne, A. R., Clemons, W. M. & Rapoport, T. A. A large conformational change of the translocation ATPase SecA. *Proceedings of the National Academy of Sciences* **101,** 10937–10942 (2004).
41. Gold, V. A. M., Whitehouse, S., Robson, A. & Collinson, I. The dynamic action of SecA during the initiation of protein translocation. *Biochem. J.* **449,** 695–705 (2013).
42. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nature Methods* **5,** 507–516 (2008).
43. Chen, J., Petrov, A., Tsai, A., O'Leary, S. E. & Puglisi, J. D. Coordinated conformational and compositional dynamics drive ribosome translocation. *Nat Struct Mol Biol* **20,** 718–727 (2013).
44. Blanchard, S. C. Single-molecule observations of ribosome function. *Current Opinion in Structural Biology* **19,** 103–109 (2009).
45. Wang, L. *et al.* Allosteric control of the ribosome by small-molecule antibiotics. *Nat Struct Mol Biol* **19,** 957–963 (2012).
46. Grant, B. J., Gorfe, A. A. & McCammon, J. A. Large conformational changes in proteins: signaling and other functions. *Current Opinion in Structural Biology* **20,** 142–147 (2010).
47. Fei, J., Kosuri, P., MacDougall, D. D. & Gonzalez, R. L. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. *Molecular Cell* **30,** 348–359 (2008).
48. Munro, J. B., Altman, R. B., O'Connor, N. & Blanchard, S. C. Identification of two distinct hybrid state intermediates on the ribosome. *Molecular Cell* **25,** 505–517 (2007).
49. Loparo, J. J., Kulczyk, A. W., Richardson, C. C. & van Oijen, A. M. Simultaneous single-molecule measurements of phage T7 replisome composition and function reveal the mechanism of polymerase exchange. *Proceedings of the National Academy of Sciences* **108,** 3584–3589 (2011).
50. Graham, T. *et al.* ParB spreading requires DNA bridging. *Genes & Development* **28,** 1228–1238 (2014).
51. Kath, J. E. *et al.* Polymerase exchange on single DNA molecules reveals processivity clamp control of translesion synthesis. *Proceedings of the National Academy of Sciences* **111,** 7647–7652 (2014).
52. Tanner, N. A. & van Oijen, A. M. Chapter Eleven - Visualizing DNA Replication at the Single-Molecule Level. *Methods in enzymology* (2010).
53. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments. *Biophysical Journal* **94,** 1826–1835 (2008).
54. Zheng, Q. *et al.* On the Mechanisms of Cyanine Fluorophore Photostabilization. *J. Phys. Chem. Lett.* **3,** 2200–2203 (2012).
55. Goshtasby, A. Image registration by local approximation methods. *Image and Vision Computing* **6,** 255–261 (1988).
56. Blanco, M. & Walter, N. G. Analysis of complex single-molecule FRET time trajectories. *Methods in enzymology* **472,** 153–178 (2010).

57. McKinney, S. A., Joo, C. & Ha, T. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophysical Journal* **91,** 1941–1951 (2006).
58. van de Meent, J.-W., Bronson, J. E., Wiggins, C. H. & Gonzalez, R. L., Jr. Empirical Bayes Methods Enable Advanced Population-Level Analyses of Single-Molecule FRET Experiments. *Biophysical Journal* **106,** 1327–1337 (2014).
59. Bronson, J. E., Fei, J., Hofman, J. M., Ruben L Gonzalez, J. & Wiggins, C. H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophysical Journal* **97,** 3196–3205 (2009).