THE EFFECT OF SINGLE-STRANDED DNA BINDING PROTEIN RPA2 ON XPD HELICASE PROCESSIVITY

BY

BARBARA STEKAS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Aleksei Aksimentiev, Chair
Professor Yann Chemla, Director of Research
Professor Taekjip Ha, Johns Hopkins University
Professor Brian DeMarco

Abstract

Understanding how proteins work together to perform vital cellular functions, such as replicating and repairing DNA, not only extends our understanding of fundamental biology, but can also lead to important medical interventions. As a necessary first step to understanding larger systems, we focus on a two-protein system involved in DNA repair. Xeroderma pigmentosum group D (XPD) is a helicase protein that plays an important role in nucleotide excision repair (NER). Its function is to unwind double-stranded DNA, allowing access to the bases that connect the strands and code genetic information. Previous work has shown that XPD activity is enhanced by the single-stranded DNA binding protein replication protein A (RPA2). However, the mechanism by which unwinding enhancement occurs is unknown. In single-molecule optical trapping experiments, we monitor – with single base-pair precision – the unwinding of a DNA hairpin by XPD in the presence of RPA. We observe the effect of RPA2 on XPD unwinding in real time and distinguish between proposed models of protein cooperation by analyzing changes in unwinding behavior with added RPA. Our data disfavor mechanisms by which RPA2 melts the duplex ahead of XPD as well as RPA2 sequestering ssDNA behind the helicase. We present our own 2-state kinetic model of XPD unwinding that we believe explains our data best. We propose that XPD has two inherent states of unwinding, high and low processivity, and that RPA2 aids unwinding by increasing the likelihood of XPD being in its more processive state.

To my family,
especially my husband, Andrew.

Acknowledgements

First and foremost, I would like to thank my advisor, Prof. Yann Chemla. Under Yann's guidance I have learned a tremendous amount about experimental design, data analysis techniques, and scientific communication. I appreciate his careful and detailed approach to scientific inquiry. Thank you, Yann, for shaping me into the scientist I am today.

Thank you to past and present members of Chemla Lab, especially Zhi Qi, Alice Troitskaia, Kevin Whitley, Sukrit Suksombat, Isaac Li, Matt Comstock, Vishal Kottadiel, Patrick Mears, and Tatyana Perlova. You have all been wonderful colleagues and I learned so much from each of you. I will miss working with you as much as our lunchtime debates and coffee breaks.

I would like to thank Prof. Maria Spies and my collaborators in the Spies Lab. Maria has been a great collaborator, full of ideas and excitement for new experiments. Thank you for sharing your insights, enthusiasm, and protein samples with me.

Thank you to my committee members for your time spent reading and providing helpful feedback on my preliminary exam, dissertation, and defense: Prof. Aleksei Aksimentiev, Prof. Taekjip Ha, and Prof. Brian DeMarco.

I would also like to thank the entire Department of Physics and the Center for Physics of Living Cells for fostering such an engaged and collaborative community. A special thanks to Graduate Director Prof. Lance Cooper for his endless dedication to me and all physics graduate students. From mentoring conversations, to editing thesis chapters, to encouraging countless student-lead initiatives in the department, and more, in many ways you are the heart of this department and it's been my enormous privilege to work with you.

Thank you to all the friends I have made in Champaign. Zack, Shawn, Carolyn, Ben, Brandon, Kristen, Vishal, Farshid and everyone else that made graduate school more than bearable – it's actually been tremendous fun. Thank you for keeping me sane, grounded, and still laughing even when things were tough. And thank you to the Wild Bohrs softball team for the friendships, the exercise, and reminding me that the scoreboard isn't the only way to measure success.

Thank you to everyone in New Jersey who helped me to finish my work remotely. I am especially grateful to Jim Rugolo and everyone at G3 Technologies in Murray Hill, NJ for generously providing me with the office space and endless coffee I needed to finish writing my thesis. Everyone has been so kind and supportive, and the focused atmosphere has been beyond helpful. And thank you to Nathalie Morales for inspiring and motivating me with her incredible work ethic.

Last but not least, I want to thank my family for supporting me every step of the way. To my brother, Nick, thanks for always motivating me, whether with your words of encouragement or through friendly rivalry. To my dad, Jim Stekas, you have definitely earned the award for "Longest Serving Homework Helper." Thank you for teaching me to read and add all the way through to doing qual problems and thesis edits. To my mom, Stephanie Novak, thank you for always encouraging me and inspiring me as both my mom and a fellow scientist. And thank you for your hands-on interest in my work and reading every poster and paper along the way including this thesis. I'm so grateful to have you as my parents. To my husband, Andrew, thank you for your love, support, and patience over the past few years. Words cannot express how happy you make me.

Table of Contents

Glossary

Adenosine triphosphate (ATP): A molecule commonly used as fuel for chemical reactions in cells.

Annealing: The spontaneous formation of base pairing bonds between single strands of DNA.

Deoxyribonucleic acid (DNA): The genetic blueprint of a cell made up of strands of nucleic acid bases. It can be found as a single-strand (ssDNA) or in double-stranded (dsDNA) form in which the bases of one strand bind to those of another.

Helicase: A protein that separates the base pairs of double-stranded DNA, exposing the bases and creating single-stranded DNA.

Melting: The passive breaking of base pairing bonds in double-stranded DNA due to thermal energy fluctuations.

Processivity: The number of bases a helicase is able to unwind in one attempt.

Replication Protein A (RPA): Protein that binds to and protects single-stranded DNA. Involved in many cellular processes including DNA repair.

Translocation: When a protein moves along a substrate without fully detaching.

Unwinding: The processes of applying energy to actively break the base pairing bonds in double-stranded DNA creating single-stranded DNA

Xeroderma Pigmentosum Group D (XPD): A helicase protein involved primarily in DNA repair.

Chapter 1: Introduction

Proteins are the workhorses of living cells, performing vital functions from metabolizing food sources to preserving the genetic blueprint of the cell, DNA. Because cells have a limited number of resources, it is inefficient to create a different protein for each and every task in the cell. Instead, proteins often pull double duty and perform different tasks as part of different cellular processes. But how does a protein, just a big molecule, "know" when to preform task A instead of task B? Often by signals from other proteins! In fact, proteins rarely act alone; they can link together forming complexes and interact in vast networks. Understanding this interplay is vital to understanding how proteins accomplish their cellular duties and not only extends our understanding of fundamental biology but can also lead to important medical interventions. As a necessary first step to understanding larger systems, this thesis will describe a simpler system of two interacting proteins, XPD and RPA2 from the archaeal organism *Ferroplasma acidarmanus*.

1.1 XPD Helicase and Its Role in the Cell

Helicases are proteins whose function is to separate the two strands of double-stranded DNA (dsDNA), allowing access to the bases that connect the strands and code genetic information[1–4]. With few exceptions, helicases accomplish this by translocating unidirectionally along single-stranded DNA (ssDNA) using ATP to fuel their movement. As they move along one strand, the translocating strand, and they encounter regions where there is progress is blocked by base-paired dsDNA and must break those bonds in order to keep moving forward. Helicases are required for a long list of cellular processes including DNA replication, repair, and transcription[1,2].

Xeroderma pigmentosum group D (XPD) is a 5' to 3' superfamily 2 helicase. In humans and other eukaryotes (animal- and plant-like cells), XPD has an important role in DNA repair as part of transcription factor II H (TFIIH)[5]. Transcription factor II H (TFIIH) is a 10-protein complex found in eukaryotes including humans that plays an indispensable role in transcription initiation[5]. Transcription is the process by which a cell "reads" the genetic blueprint in DNA and is the first step in gene expression and the making of proteins. TFIIH separates some dsDNA creating an

ssDNA bubble allowing other proteins to bind and begin transcribing the from the exposed ssDNA template[5–7].

In addition, TFIIH is also vital for nucleotide excision repair (NER), a process that repairs bulky lesions on DNA that are usually the result of UV damage[5]. People with genetic mutations that disrupt NER are subject to high rates of skin cancer and premature aging[8]. XPD derives its name from the condition xeroderma pigmentosum, one of a few genetic disorders directly linked to mutations in human XPD[8–10]. TFIIH opens a bubble in the dsDNA around the damage site allowing other proteins to then bind, cut out the damaged piece of ssDNA, and replace it with undamaged DNA (**Figure 1.2**)[5,11,12].

Opening these bubbles in dsDNA is the job of a helicase and TFIIH has two: XPD and XPB. Evidence shows XPD unwinding is required for NER but is not necessary for transcription initiation, while XPB ATPase activity is not necessary for NER but is required for transcription[12–14]. This division of labor begs the question: what causes this switch and why? The TFIIH performs nearly the same task in both cases, it would make sense for the same proteins to perform the same function in both. Instead, we see a clear change in roles for XPD moving from center stage in NER to a supporting role in transcription initiation. What causes this change in activity? Understanding this stimulation of XPD activity would add to our understanding of the larger TFIIH complex and can shed light on the stimulation of other, similar helicases in different pathways. Unfortunately, human XPD (hXPD) has not been purified outside of the larger TFIIH complex. Instead, to study the single-molecule properties of XPD, we use a homolog that can be isolated found in the archaeal organism *Ferroplasma acidarmanus*, called *Fac*XPD.

Archaea are a distinct branch of single-cellular life that share some common features with eukaryotes. They are often studied, as in this thesis, as simpler models for more complex eukaryotic systems. But a major downside to studying archaeal proteins is that archaea are, in general, much less well studied than other organisms, especially human cells! Thus we do not know the precise function of *Fac*XPD in the cell and our current best guess is based on its resemblance to eukaryotic XPD. *Fac*XPD has a very similar structure to human XPD and we expect that it functions much like its human counterpart. Thus, it is speculated that *Fac*XPD is involved

in an archaeal NER pathway, but the specifics of that pathway or even the *existence* of such a repair pathway in *F. acidarmanus* is not confirmed[15–17].

Previous experiments have shown that a single molecule of *Fac*XPD is an inefficient unwinder, exhibiting low processivity[18,19]. Processivity is defined as the number of base pairs a helicase is able to unwind and some helicases are known to unwind hundreds or even thousands of bases[20,21]. In previous optical trapping experiments from our lab, *Fac*XPD unwound in repetitive bursts of only 12 bp even of the aid of relatively high forces from the traps[19].

1.2 FacRPA2 Regulates FacXPD Activity

Another protein with a known role in eukaryotic NER that also has an archaeal homolog is replication protein A (RPA). RPA is a single-stranded DNA binding protein that, as the name implies, binds to ssDNA to protect it from damage[22–24]. RPA is so ubiquitous that it is present in numerous pathways – any time ssDNA is created, RPA is there to bind to it. In particular, RPA is known to be involved in human NER and is recruited to the damage site by TFIIH[25,26]. As such, helicases like XPD are likely to encounter RPA as they unwind, and indeed human RPA plays a role in the damage recognition that activates NER as well as binds to the ssDNA in the NER bubble[11,12].

1.2.1 Past experiments

Previous bulk studies of *Fac*XPD and *Fac*RPA2 by collaborators in the Spies Lab, mixed XPD, a DNA substrate, and RPA2 in solution and measured the amount of DNA substrate unwound after 15 min. They found that XPD was able to unwind more DNA in the presence of RPA2 and that the effect was concentration dependent[27]. They also determined that RPA2 alone is capable of “melting” dsDNA at low rates. Melting is a passive process in which two strands of dsDNA separate due to an increase in thermal energy. In this case, the presence of RPA2 lowers the energy needed for dsDNA strands to separate making them more likely to melt due to normal thermal fluctuations.

In a related set of experiments from the Spies and Ha Labs, single molecules of XPD and RPA2 were tracked using total internal reflection fluorescence microscopy (TIRF) as they bound to and translocated along ssDNA[28,29]. These experiments showed that RPA2 does not interfere

with XPD binding to ssDNA and that XPD can even translocate past bound RPA2 on the same strand. While they did not measure unwinding, these experiments hint at a cooperative relationship between XPD and RPA2 and offered a method of detecting their interactions using fluorescence. This thesis will follow up on those experiments and attempt to discern the mechanism of XPD and RPA2 cooperativity.

1.3 Proposed models of processivity enhancement

Three possible mechanisms for the enhancement of XPD unwinding by RPA2 have been proposed by the Spies Lab: ssDNA sequestration, duplex melting, and complex formation[27].

1.3.1 ssDNA sequestration model

In the ssDNA sequestration model, RPA2 binds to the ssDNA behind XPD and acts as a physical barrier preventing the helicase from stepping or sliding backward (**Figure 1.3 A**). XPD has been shown to back-step and back-slide often during unwinding[19]. Sequestration would force XPD to more forward or not at all increasing the amount of DNA unwound over time.

1.3.2 Duplex melting model

*Fac*RPA2 and most single-stranded binding proteins in general are able to "melt" a DNA duplex (another word for dsDNA)[22,27,30–32]. We have found that RPA2 can temporarily melt ~6 base pairs of dsDNA when it binds near an ssDNA-dsDNA junction (detailed in Chapter 3:). In the duplex melting model, RPA2 binds to the DNA strand near the ssDNA-dsDNA junction and melts a few base pairs creating ssDNA ahead of XPD (**Figure 1.3 B**). In most cases, helicases move along ssDNA faster than they unwind dsDNA due to the physical and energetic barrier posed by the base pairing hydrogen bonds it must break to move forward during unwinding. Though we do not have measurements of *Fac*XPD's translocation speed on ssDNA to compare to its unwinding speed, past experiments from our lab have shown that XPD unwinds via a "largely passive" process, relying on thermal energy fluctuation to help it break the base pairing bond ahead of it[19]. Thus, RPA2 enhancing the effect of thermal fluctuations may also enhance XPD's unwinding potential.

1.3.3 Complex formation model

In the complex formation model, RPA2 and XPD join together in a complex that is more processive than the sum of its parts. Experiments on human proteins have shown that hRPA forms stable complexes with Werner, Bloom, FANCJ, and RecQ helicases[33–38]. These hRPA-helicase complexes have been shown to be more processive than the helicases themselves and also more likely to unwind past obstacles such as damage sites and other proteins, perhaps because RPA2 binds to XPD in a way that forces XPD into a more processive conformation (**Figure 1.3 C**). However, experiments from the Spies Lab show that *Fac*XPD and *Fac*RPA2 do no form a stable complex in solution. Nevertheless, it is possible that they form a stable but short-lived complex or that they may require DNA to mediate their interaction.

1.4 Why we need our trap assay

Through bulk experiments, the stimulation of XPD unwinding due to the addition of RPA2 was observed, but these were population measurements and could not track unwinding in real time. TIRF microscopy experiments allowed for real-time observations of single molecule motion on ssDNA, but this was translocation along ssDNA and they did could not measure unwinding. Our optical trapping technique offers a unique opportunity to observe the real-time unwinding activity of single molecules of XPD under different RPA2 conditions.

An optical trap, also known as optical tweezers, is a tightly focused laser beam which creates a force field capable of holding small objects in three dimensions. Optical traps have been used extensively and with great success to study a wide range of biological systems, including helicases[19,39–41]. In our experiment, we monitor – with single base pair precision – the unwinding of a DNA hairpin (a self-complementary section of ssDNA that zips together and base-pairs to itself) by XPD in the presence of RPA2. Our aim is to observe the effect of RPA2 on XPD unwinding and to distinguish between the three models of protein cooperation by analyzing changes in unwinding behavior with added RPA2.

First, we will describe the optical tweezers instrument and how we use it to measure the opening of a DNA hairpin by a protein. Then, we will discuss in more detail the results of past experiments on XPD unwinding in the trap (Chapter 2:). Next, we will examine the melting effect

RPA2 alone has on our DNA hairpin substrate (Chapter 3:). After this control has been established, we will observe XPD unwinding in the presence of increasing concentrations of RPA2 and analyze the resulting changes in XPD unwinding behavior. We will test our data against the established models to find evidences either supporting or contradicting them and present our own 2-state model of XPD unwinding that we believe explains our data nicely (Chapter 4:). Finally, we will suggest directions for future research to refine our understanding of this system and present preliminary data from one of those experiments (Chapter 5:).

1.5 Figures

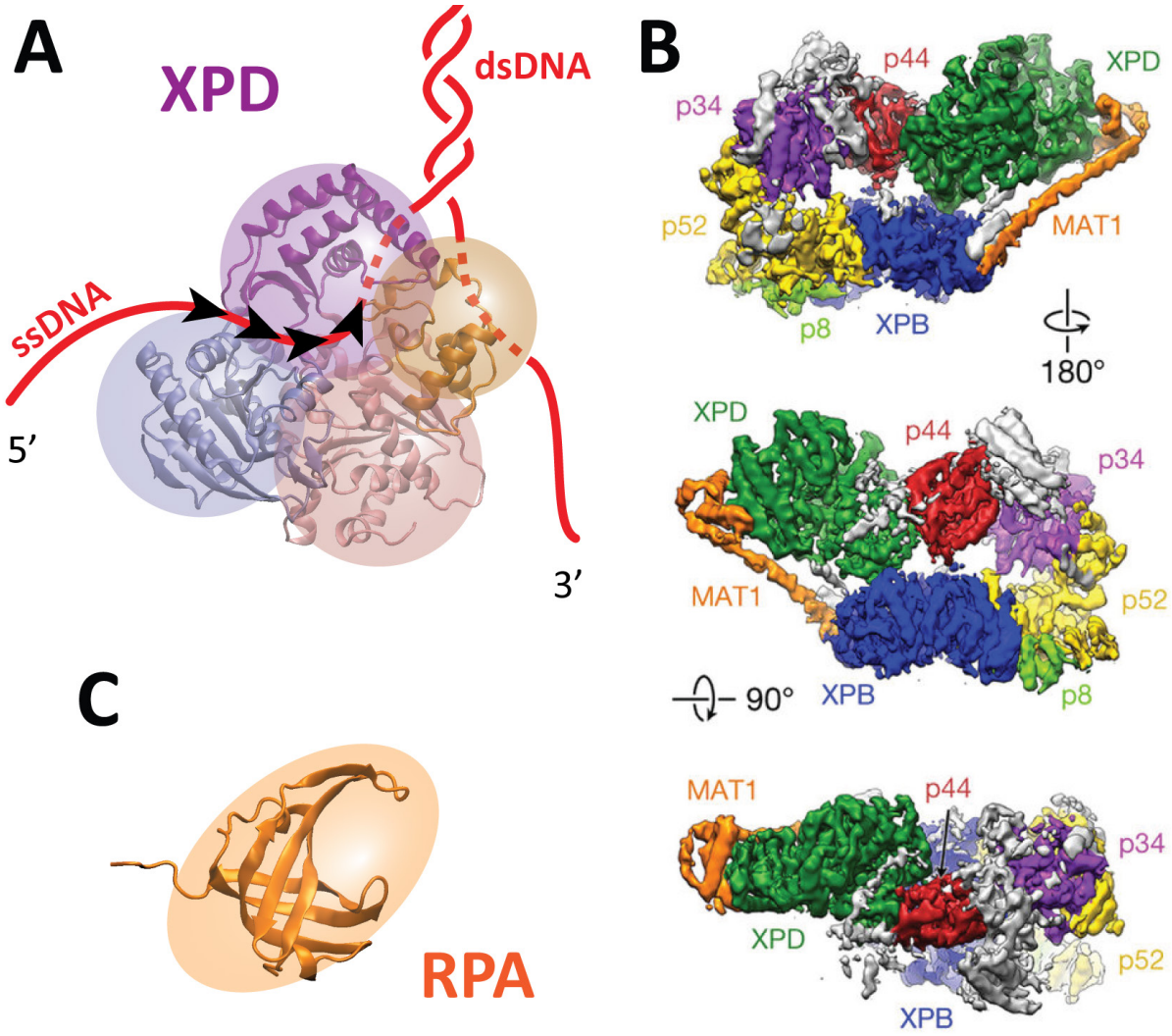

Figure 1.1 Molecular structures of XPD and RPA. **A)** Crystal structure of an archaeal XPD (PDB 3CRV) similar to *Fac*XPD[9]. Arrows indicate direction of XPD translocation along ssDNA (5' to 3') and the binding site for ATP is indicated by a star. XPD is composed of four domains: HD1, HD2, Arch and FeS. HD1 and HD2 are the motor domains that propel XPD along. The Arch domain closes over ssDNA that XPD binds to. The FeS domain helps stabilize the entire structure. A cartoon sphere representation of XPD is overlain on the crystal structure and will appear in other figures throughout this thesis. **B)** Cryo-electron microscope images the human TFIIH core complex including helicases XPD (in green) and XPB (in blue). Views of the 3D structure are along 3 orthogonal axes are shown. Reprinted from Greber *et al.* with permission[42]. **C)** Crystal structure of archaeal RPA (PDB 2K5V) similar to *Fac*RPA2, a single OB fold that binds to ssDNA in the C-shaped cavity. A cartoon ovoid representation of RPA2 is overlain on the crystal structure and will appear in other figures throughout this thesis.

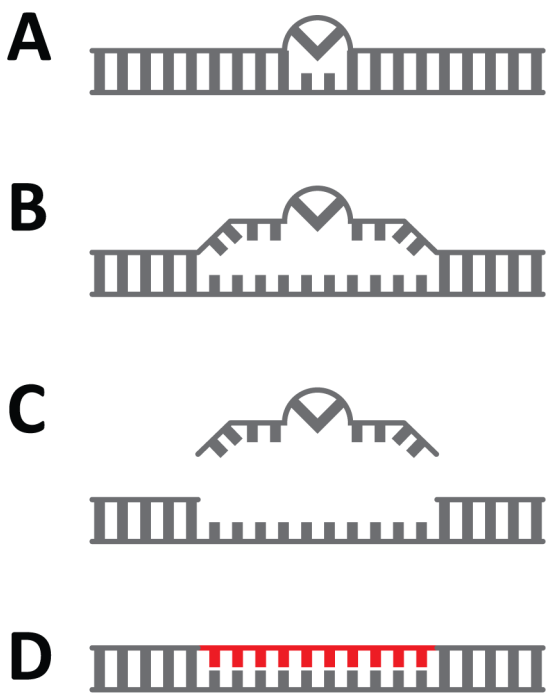

Figure 1.2 The steps of nucleotide excision repair. **A)** Bulky damage caused by UV radiation is detected. **B)** A bubble is opened around the damage site. XPD is believed to help open the bubble by unwinding. RPA2 is thought to stabilize the bubble by binding to ssDNA and preventing the bubble from closing again. **C)** The damaged strand is cut out and removed. **D)** New, undamaged DNA (red) is brought in to replace it.

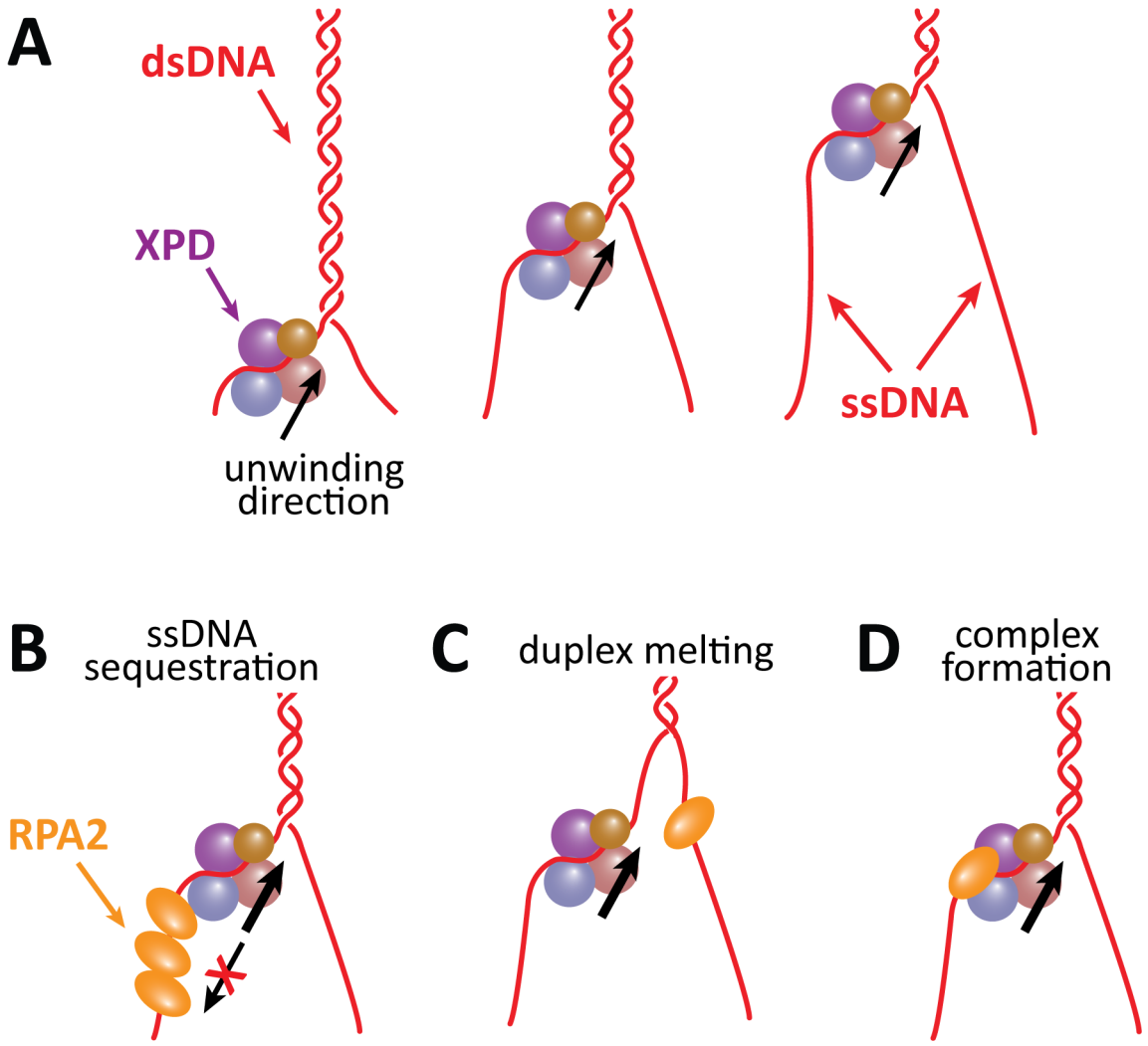

Figure 1.3 Cartoon representations of XPD unwinding in the presence and absence of RPA2. **A)** XPD unwinding in the absence of RPA2. XPD translocates along ssDNA in the 5' to 3' direction (black arrow) with occasional back-steps (not shown). As it progresses, XPD separates the dsDNA ahead of it. Proposed models of RPA2 enhancement of XPD processivity: **B)** In the ssDNA sequestration model, RPA2 binds behind XPD and prevents the helicase from stepping backward**. C)** In the duplex melting model, RPA2 binds near the dsDNA-ssDNA junction and melts a few base pairs ahead of XPD making it easier for the helicase to a step forward. **D)** In the complex formation model, RPA2 and XPD form a complex that is more processive than XPD alone.

Chapter 2: Hairpin Assay Overview

2.1 Dual Trap Instrument Design

Optical traps sound like the stuff of science fiction; a focused beam of light that can grab and trap objects near the focal point. Though current technology has not yet reached the scale of Star Trek tractor beams (towing objects > 200 m in length[43,44]), the ability to trap and control small particles (~10^{-6} m) with laser light has been well established since Arthur Ashkin's experiments in the early 1970s[45,46]. Since then, optical traps (sometimes called optical "tweezers") have been used to trap and cool atoms[47,48], manipulate live cells[41,49,50], and study molecular motor proteins[51–53].

2.1.1 Optical Trapping Theory and Execution

The theoretical underpinning of an optical trap is most readily understood in the case of a trapped sphere whose radius is much larger than the wavelength of the trapping laser. In this case, the forces on the sphere can be described with simple ray optics[39]. (Note: in our experiments, the diameter of the trapped bead (~900 nm) is similar to the wavelength of the trapping laser (1064 nm), therefore this theoretical description is incomplete, but conceptually illustrative.) As shown in **Figure 2.1**, refraction of incident light produces a change in the light's momentum. By Newton's Third Law, this change in momentum results in an equal and opposite force on the illuminated sphere that is proportional to the intensity of the incident light.

For a sphere with an index of refraction greater than that of the surrounding medium, the refraction of light in the sphere can be shown to result in a force on the sphere in the direction of the intensity gradient[39]. Thus, the refraction of light in a spherical particle trapped in a focused, Gaussian beam will cause a lateral force toward the center of the beam. In the axial direction, the particle will be trapped at an equilibrium position slightly downstream of the laser focus where the axial component of the refractive force balances the force caused by scattering[39] (**Figure 2.1 C**). If the particle moves away from equilibrium, the imbalance of forces results in a restorative force back toward the equilibrium position. The trapped sphere experiences a nearly harmonic potential created by these refractive forces; thus, for small displacements from the trap center, the restoring force on the particle obeys Hooke's Law, $F = -k_{trap}\Delta x_{bead}$.

Locating the trapped particles

The location of the particle within the trap is measured using a technique called back-focal-plane interferometry[54]. A particle deflected laterally from the center of the trap will displace the beam of light transmitted through it in the same direction. A quadrant photodiode (QPD) is used to measure the amount the beam is deflected from its equilibrium position, Δx_{bead}. A QPD contains four photodetectors arranged in a square that output a voltage proportional to the total intensity of light shining on each quadrant. For a beam with a Gaussian intensity profile, the difference in intensity between the left and right (and/or top and bottom) quadrants is proportional to the beam's displacement in that direction (**Figure 2.2**). A scaling factor, recalibrated for each bead, converts the output displacement of the bead/beam from volts to nm.

Calibrating the trap

To convert the bead displacement into a distance (nm) and to calculate the force on the bead from the trap we must calibrate the trap. We use a passive calibration method that makes use of the Brownian motion of the bead. The physics of a bead in a harmonic potential is well understood. The power spectrum of the bead's thermal motion in each direction is described by a Lorentzian profile[39]

$$S_{xx}^{(QPD)}(f) = \frac{S_{xx}^{(nm)}(f)}{\alpha^2} = \frac{k_B T}{\alpha^2 \pi^2 \beta (f_0^2 + f^2)} \tag{2.1}$$

where $S^{(QPD)}$ has units of V^2/Hz. Because our displacement data is output as a voltage, a scaling factor, α, is necessary to convert between V and nm, the resulting units of our theoretical model. In that model, β is the hydrodynamic drag coefficient of the trapped object and f_0 is the roll-off frequency, related to trap stiffness, k_{trap}, by $f_0 = k_{trap}(2\pi\beta)^{-1}$. Tracking the motion of a trapped spherical bead of known size – and thus known hydrodynamic drag, $\beta = 6\pi\eta r$ – at a high sampling rate (125kHz) for a sufficiently long time (~10s) provides enough data to fit this spectrum with two parameters (**Figure 2.3**): the trap stiffness, k_{trap} (pN/nm), and the scaling factor, α (V/nm). This calibration is performed for each trapped bead. A full derivation of this equation and more in-depth explanation can be found in Appendix A.1.

2.1.2 Our Dual-Trap Optical Tweezers Instrument

The DNA "dumbbell" assay

We employ a dual trap optical tweezers instrument to monitor protein interactions on DNA[40,55] (**Figure 2.4**). In this type of experiment, sometimes referred to as a "dumbbell" assay, two beads are independently trapped in two optical traps and a DNA "tether" is formed between the two beads. The traps hold the DNA taut and proteins in solution are allowed to bind to the DNA tether. When proteins bind or otherwise interact with the DNA, it affects the end-to-end extension of the molecule. Thoughtful design of varied DNA substrates allows for all manner of different types protein-DNA interactions to be measured using this and related methods. Examples from the literature include proteins wrapping[56,57], stretching[58], unwinding[19,52,59,60], cutting[61], unknotting[62,63], and more (**Figure 2.5**).

Similar optical trapping experiments use only one trap and attach the opposite end of a DNA tether to a glass slide[39,57,60]. While effective, this type of assay has a major limitation in the form of noise introduced by both motion of the attachment surface itself and motion of the trap relative to the surface[64,65]. Even on a levitated optical table, sample stages can move and drift by 5-10 nm/s[64,66]. The length scales of biological processes tend to be on the order of nanometers, and can even be as small as angstroms, so even very small intrinsic instrument noise can drown out relevant signal.

By using two optical traps, our instrument decouples measurements from stage motion, greatly reducing intrinsic noise. By creating two traps from the same laser, we also remove noise and drift that may result along the laser path. The light of two traps follow nearly identical paths, thus any slight deviations due to interference along the beam path will be present in both traps. As our final measurements are only sensitive to the relative position of the two traps and not the absolute location, this source of noise is effectively removed from the instrument.

Creating and controlling the two optical traps through an AOM

To create two traps from the same laser, an acousto-optic modulator (AOM) is used. In an AOM, an acoustic wave is transmitted through a transparent quartz crystal. The wave produces a regular pattern of density variation – and thus refractive index variation – in the crystal that acts as a diffraction grating for the incoming beam of light. Adjusting the wavelength

of the acoustic wave is akin to changing the spacing of a diffraction grating, and thus increasing the frequency of the acoustic wave increases the deflection of the first order maxima[67]. Using a method known as "time-sharing," by rapidly changing the applied acoustic frequency in time, our instrument creates two traps, separated spatially by a distance proportional to the difference in acoustic frequency applied to the AOM, $\Delta X_{traps} = a\Delta f_{AOM}$. In this method, if one trap is "on" at a particular time, the other must be "off," resulting in two traps that are essentially blinking "on" and "off." However, the change in deflection that creates the two traps, and thus the blinking, occurs much faster (267 kHz) than trapped objects are able to respond or diffuse away, and the effective trapping force on the object is constant[40]. The proportionality constant, a, is found using CCD images of two trapped beads and a reticle. The image of the reticle is used to determine the pixel size of the CCD image in nanometers. Then the distance between the two bead centers in the bead image is measured in pixels for given trap separations, Δf_{AOM} in MHz, and converted to nanometers.

Maintaining constant force through a feedback loop

During our experiments, one trap remains in a fixed location while the other is moved relative to the fixed trap. Two polystyrene beads are trapped, one in each trap, and a molecule of DNA is "tethered" between them (**Figure 2.6**). The extension and tension of the DNA tether is controlled by manipulating the relative distance between the traps, X_{traps}. A proportional-integral (PI) feedback control monitors the position of the bead in the trap (Δx_{bead}) and adjusts the location of the moveable trap to maintain the bead position and keep the force on the bead constant. This control allows experiments to be performed under constant tension.

Putting all these pieces together, which is done in full in Appendix A.2, we can calculate the final tension and end-to-end extension of the DNA tether. By monitoring changes in the extension over time at constant force, we can observe protein interactions with our DNA construct.

2.2 Experimental Design

2.2.1 The Hairpin Construct

What is it? Why use it?

We use a hairpin DNA construct to observe helicase unwinding[19,68]. The construct consists of two long dsDNA "handles" that attach to the trapped beads. In the center, there are 10 bases of ssDNA that form a loading site for proteins of interest followed by the DNA hairpin. The hairpin is composed of self-complementary ssDNA that zips and base-pairs with itself to produce dsDNA. When the hairpin opens, either from exerting a force or due to helicase unwinding, the length of the tether increases (**Figure 2.6 A**). A useful feature of the hairpin construct is that for every one basepair unwound, two nucleotides of ssDNA are released, adding to the total tether extension (**Figure 2.6 B**). This "two-for-one" bargain provides a signal boost that, in conjunction with the low noise level of our instrument, allows us to attain a spatial resolution as small as a single base pair in our experiments[19].

Attachment chemistry

The DNA tether is attached to the beads using attachment proteins common in biochemical studies. One end of the construct is labeled with biotin, which binds tightly and specifically to streptavidin-coated polystyrene beads. The other end of the construct is labeled with digoxigenin (DIG) that binds specifically to anti-digoxigenin-antibody (anti-DIG) coated beads. During experiments, a trapped, DNA-coated streptavidin bead is brought close to an anti-DIG bead until the DIG end of the DNA construct binds and forms a tether[69].

Characteristic force response of the DNA hairpin

The hairpin construct has a characteristic response to applied force. At low forces, the hairpin is "closed" and the length change of the full construct by an applied force is solely due to the stretching of the dsDNA handles and can be described by the extensible worm-like chain (XWLC) model for polymers[70,71]. At a particular unfolding force, F_u, bonds between base pairs break and the hairpin unzips. This unfolding force depends on the length and DNA sequence of the hairpin and F_u ~ 16 pN for our construct. At $F > F_u$, the force response of the fully "open" construct can again be described by the XWLC model (**Figure 2.6 C**). All experiments were

performed below F_u at a constant force of 12 pN to avoid mechanically unfolding the hairpin, ensuring all unfolding observed is due to protein activity.

Converting total DNA extension to number of base pairs open

The success of the XWLC model in describing the closed hairpin construct means deviations from the expected length of the closed conformation must be due to the hairpin opening. By comparing the measured tether extension to the predicted length of the closed construct, we can calculate the number of base pairs that have opened, n_{unwound}. In fact, $n_{unwound}$ is directly proportional to the difference in these values. This conversion calculation is explained in full in Appendix A.3.

2.2.2 Laminar flow chamber

All experiments are performed within a specialized sample chamber. The chamber sits between the two objectives that form the optical trap and is made by cutting parafilm to form channels and melting it between two cover slips. Holes in one coverslip at either end of the channels provide access for inlet and outlet tubing connected to syringes to fill the chamber with sample solution (**Figure 2.7 A**). A specially designed mount holds the tubing in place and secures the chamber to a movable stage.

For experiments described in this thesis, bead channels on the top and bottom are connected to the central trapping region by capillaries (~20 µm inner diameter) allowing for a controlled flow of DNA-coated streptavidin beads and anti-DIG beads. The central trapping region is formed by merging 3 separate channels into one (**Figure 2.7**). Each of the three channels contain different solutions that remain well separated due to constant laminar flow provided by syringe pumps, allowing trapped objects to be easily moved between different conditions.

2.2.3 Experimental Procedure

Before experiments begin, the sample chamber is filled with samples with careful attention to rid the channels of any air bubbles. Air bubbles add oxygen molecules to solution which can form reactive species that damage our DNA construct[72]. Anti-DIG beads fill the topmost channel; Streptavidin beads incubated for 30 min at room temperature with our DNA

construct fill the bottommost channel. In the central trapping region, suspended in a buffered Tris solution (Appendix C.2), we add

- Upper channel: 500 µM ATP
- Middle channel: 500 µM ATP-γS
- Lower channel: 60 nM XPD + 500 µM ATP-γS

ATP is the fuel molecule XPD uses to unwind. ATP-γS is a non-hydrolysable form of ATP, meaning XPD can bind to it, but not get energy from it. The binding of ATP, even without hydrolysis, can alter proteins and in some cases allows them to bind their DNA substrates more readily. ATP-γS was added to the XPD and blank channels in an effort to increase the efficiency of XPD binding to the DNA loading site.

Once the chamber is filled, the experiment can begin. First, we move the optical traps near the top capillary and trap an Anti-DIG bead (Step 1 in **Figure 2.8**). Then we move to the bottom capillary and trap a streptavidin+DNA bead (Step 2 in **Figure 2.8**). Then we move upstream of both capillaries and into the center trapping channel (Step 3 in **Figure 2.8**). Here we save 10 sec of Brownian motion data with the syringe pumps off at an ultra-high sampling rate (250 kHz) and calibrate the trap stiffness by fitting the power spectrum of this noise to Equation (2.1).

Next, a DNA "tether" is formed by bringing the two traps, and thus the two beads, close together and waiting for a bond to form between the DIG on the unbound end of a DNA molecule and the Anti-DIG on the surface of the neighboring bead. A tether is detected by moving the beads farther apart and monitoring the force as a function of trap separation. A single molecule of the DNA hairpin construct when stretched will produce a force vs. extension curve like that seen in **Figure 2.6 C**. More than one molecule will produce stiffer curves (force increases more quickly with extension) and will open at higher forces. If these multiple tethers are observed, they are held at a high force until broken and the tether formation process is repeated.

Force feedback is then turned on and set to a constant force below the mechanical unwinding force of the hairpin F_u (**Figure 2.6 C**). Unless otherwise noted, all experiments in this thesis were performed at a constant force of 12 pN. A few seconds of data are saved of the bare hairpin in the blank channel at 12 pN force, establishing a baseline for the molecular extension

at this force and the intrinsic noise of the DNA before unwinding begins (Step 3 in **Figure 2.8** and **Figure 2.9 B**). The force is then lowered (~5 pN) and the tether is moved into the lower XPD stream and incubated there for ~40 sec to allow XPD to bind (Step 4 in **Figure 2.8** and **Figure 2.9 B**). The incubation is performed at low force in another attempt to increase XPD binding efficiency. The addition of ATP-γS and the low incubation force seemed to aid XPD binding, but efficiency was still quite low. However, these methods had a large psychological effect on the experimenter, making her feel as if she had some control over the whims of this capricious helicase.

Following incubation, the force on the tether is increased once again to 12 pN and the tether is moved into the upper ATP channel (Step 5 in **Figure 2.8** and **Figure 2.9 B**). Upon exposure to ATP, an XPD molecule bound to the ssDNA loading site will begin to unwind the hairpin ahead of it. Unwinding data is saved until the tether breaks or XPD dissociates.

2.3 Results of single XPD unwinding in the absense of RPA2

A single XPD helicase unwinds in repetitive unwinding bursts of low processivity, typically unwinding a maximum of ~15 bp per burst (**Figure 2.9 B**). A burst consists of slow, single-base pair steps forward interspersed with occasional back stepping followed by a prolonged back stepping and rapid reannealing caused by back-sliding of the protein back to the beginning of the hairpin (**Figure 2.9 C**). We manually select the beginning and end points of each burst for our analysis and consider each burst to be independent attempts at unwinding the full hairpin.

Previous experiments from our lab analyzed data from single XPD molecules in detail[19]. Qi, *et al.*, with single base pair precision, analyzed the rates of forward and back steps. They found that XPD unwinds in 1 bp steps and its stepping behavior is heavily dependent on the sequence of the DNA it is unwinding[19]. XPD exhibited higher rates of forward stepping at A-T base pairs than G-C base pairs. Based in part on this sequence dependence, they proposed a "mostly passive" model of XPD unwinding where the helicase relies on the base pair ahead of it to break thermally (more likely for A-T than G-C pairs) before it could step forward[19]. They also linked DNA sequence to XPD processivity, showing XPD was unable to unwind through certain G-C rich regions and instead stopped and/or went backward via back-stepping and back-sliding[19].

With this data in mind and work showing RPA2 enhanced XPD activity in bulk experiments, we wondered how RPA2 would affect the unwinding behavior of a single XPD in our trap experiments? With the rich data set this assay provides, we would be able to see the impact of RPA2 on XPD unwinding down to a single forward step offering deeper insights into how the two proteins interact. As a first step toward this goal, in Chapter 3: we look at how RPA2 interacts with the DNA hairpin itself. This provides insight into RPA2 itself and serves as control before combining to the two proteins on DNA.

2.4 Figures

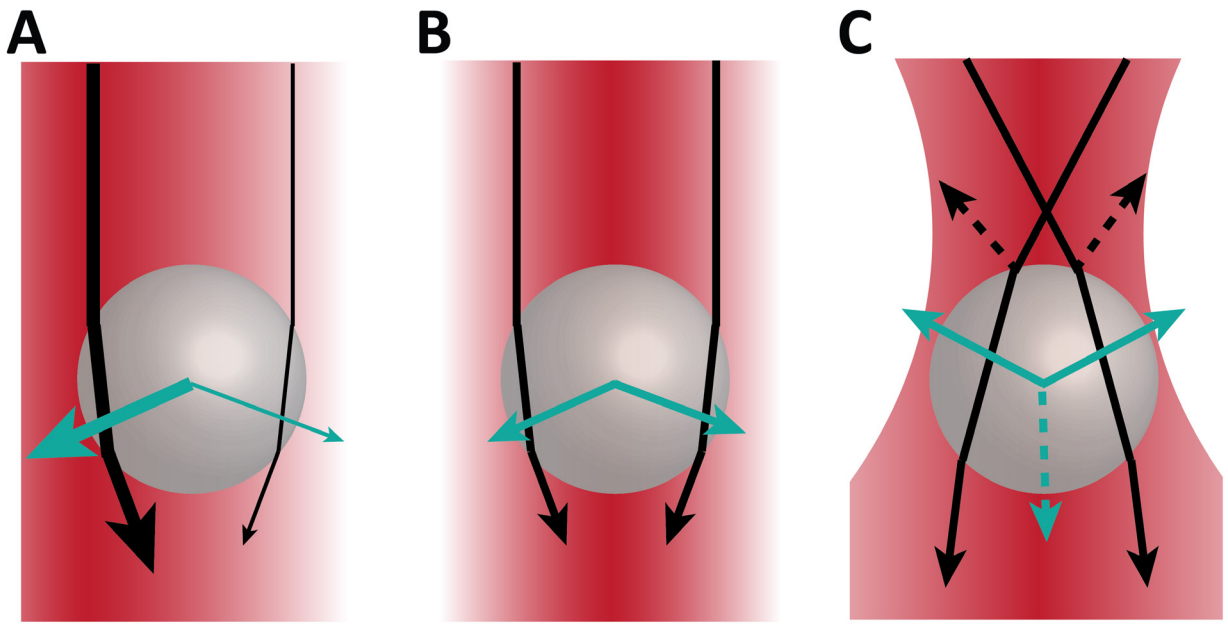

Figure 2.1 Ray optics description of optical trapping forces. A transparent bead is illuminated with rays of light from a trapping laser. When passing through the bead the light is bent twice due to refraction, once when entering the bead and once when exiting back to the surrounding medium. This net change in momentum of the light produces an equal and opposite force on the bead. Rays of light are shown in solid black with thicknesses proportional to their intensity. The forces generated by those rays are shown in solid teal with thicknesses proportional to magnitude of the force. **A)** When illuminated by parallel rays with a horizontal intensity gradient, the net force on the bead is to the left, up the intensity gradient. **B)** In a radial intensity gradient, the bead reaches equilibrium at the center of the beam where radial forces cancel. **C)** In a focused beam with radial intensity gradient, the bead again comes to rest at the center of the beam at point downstream of the focus were the axial force of refraction is balanced by the scattering force. Scattered rays and the scattering force are shown as dashed lines.

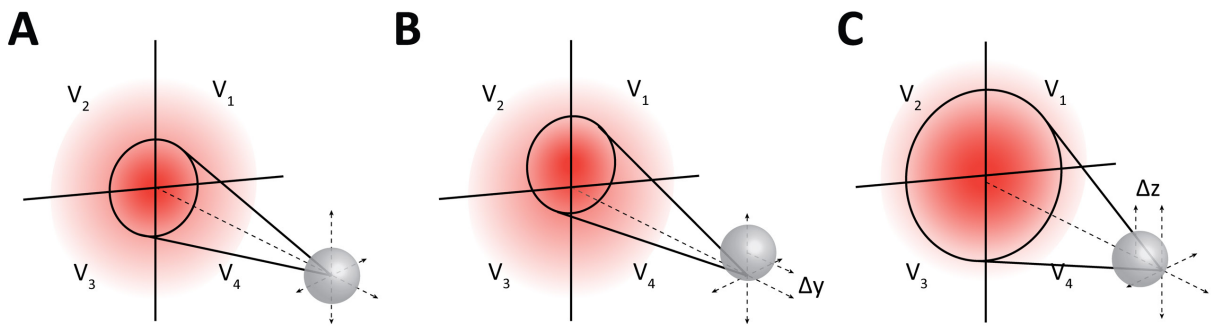

Figure 2.2 Back focal plane interferometry using QPDs. A QPD has four photo detectors that produce a voltage proportional to the intensity of light shining on them. **A)** An un-deflected bead in the trap. **B)** A bead deflected laterally in the *y* direction. The *y* position of the bead (*Δy*) is proportional to $(V_1 + V_2) - (V_3 + V_4) = V_y$. **C)** A bead deflected axially in the *z* direction. The *z* position of the bead is proportional to $V_1 + V_2 + V_3 + V_4 = V_z$. Figure courtesy of Y. Chemla.

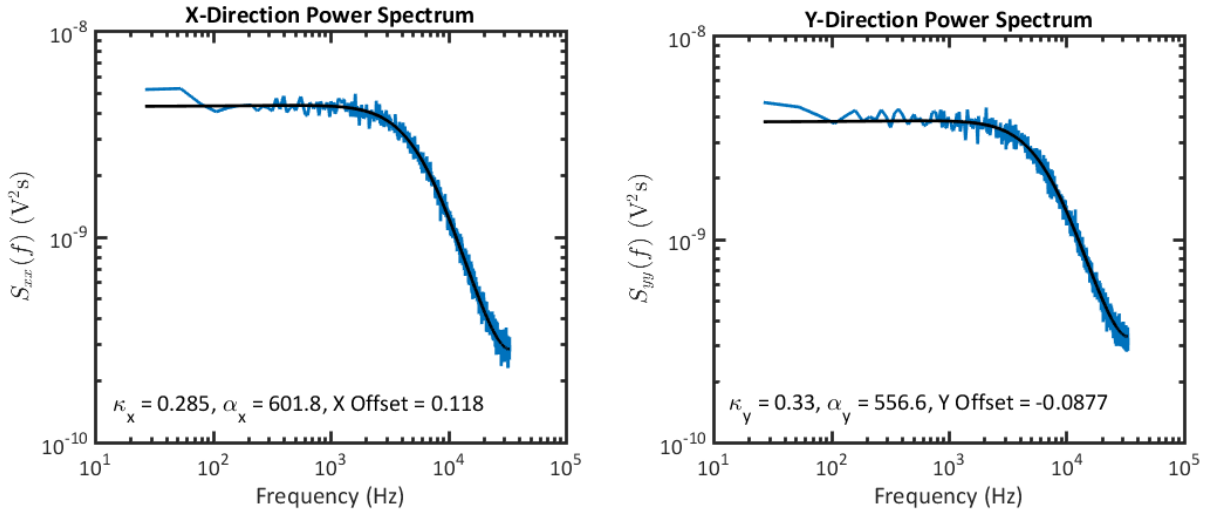

Figure 2.3 Example of high sampling rate QPD voltage data (V_x and V_y) taken of a bead at rest in the trap with no applied force (blue). The fit to the Lorentzian model (black) in the *x* and *y* directions is used to calibrate the trap stiffness. The parameters of the fit measure the trap stiffness (κ) and volts-to-nm conversion for the deflection of the bead (α). The mean value QPD voltage for the bead at rest with no applied force is the offset voltage and subsequent bead displacement is measured in relation to this resting point.

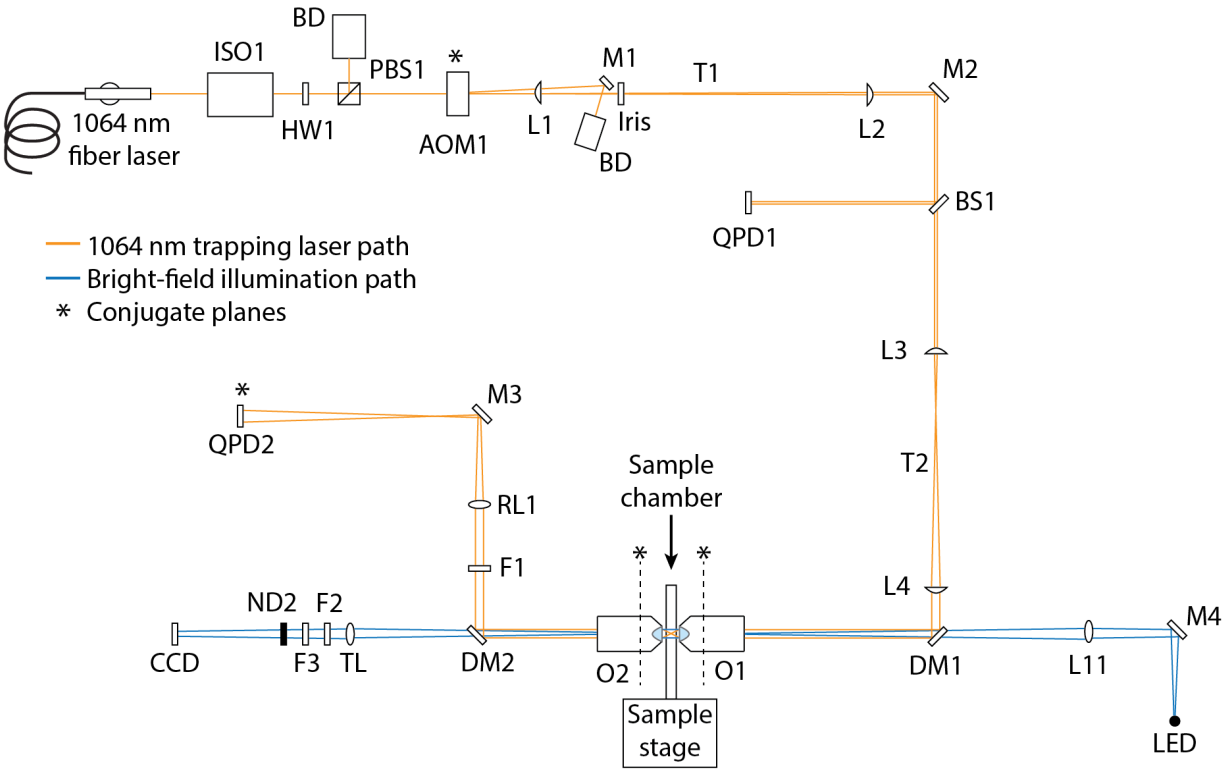

Figure 2.4 Layout of dual trap instrument courtesy of Y. Chemla adapted from a figure in Whitley (2017)[69]. AOM1 controls the interlacing of the traps. Two telescopes (T1 and T2) expand the beams to fill the back of the objective (O1). The trap forms between the two objectives where we place our sample chamber. QPD2 detects motion of the bead within the trap and, because it is placed at the conjugate plane of the AOM and objective focal points, is insensitive to movement of the traps themselves. Because the traps are interlaced, only one QPD is needed and the signal from each bead is separated in time.

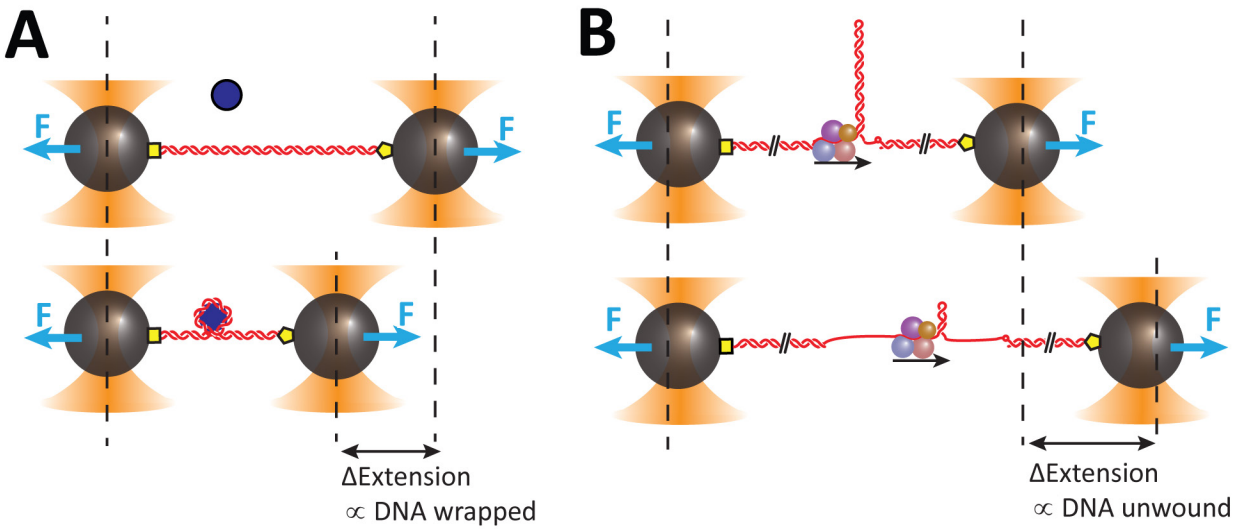

Figure 2.5 Insights of changing DNA extension into protein-DNA interactions. **A)** A protein that wraps DNA will shrink the end-to-end length of DNA when it binds. The resulting change in extension will be related to the amount of DNA wrapped[56]. **B)** In contrast, when a helicase unwinds a DNA hairpin construct, the end-to-end length of the construct increases an amount proportional to the number of base pairs unwound.

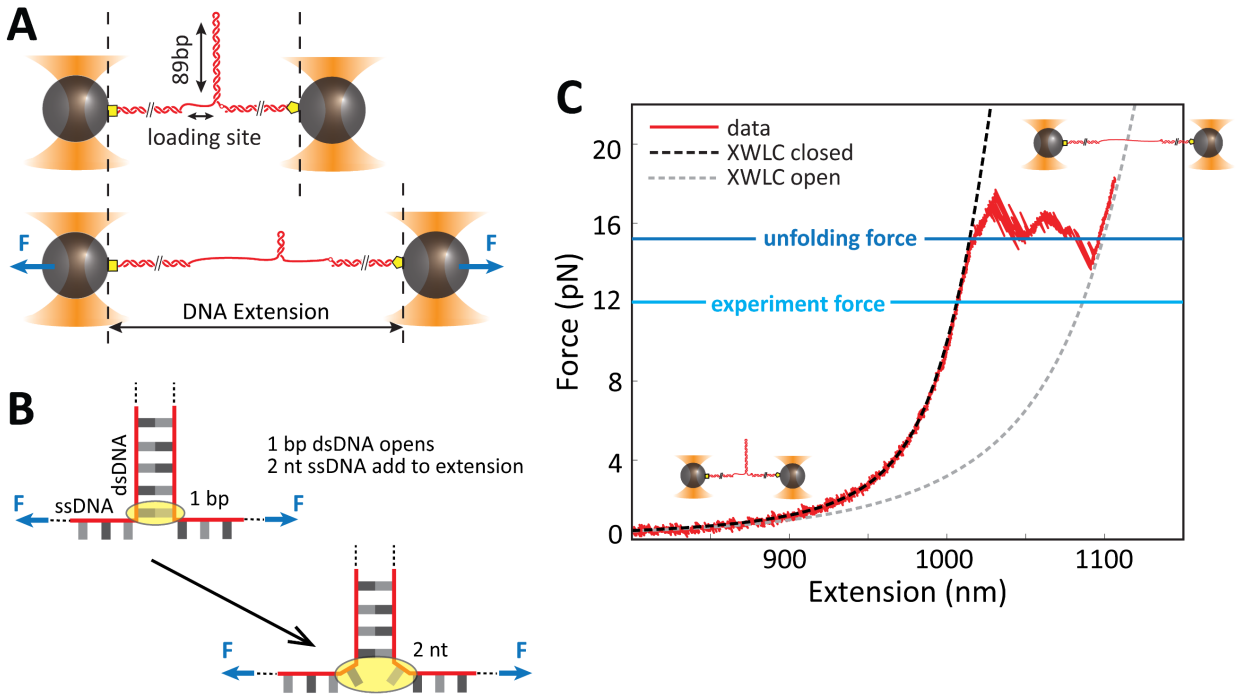

Figure 2.6 Dual trap assay with a DNA hairpin tether. **A)** The DNA construct attaches to the trapped beads via streptavidin-biotin (pentagon) and digoxigenin-anti-digoxigenin (square) linkages. As the hairpin opens, the length of the tether and the distance between the traps increases. **B)** Each single base pair of dsDNA that separates produces two nucleotides of ssDNA. This produces a change in end-to-end length of ~1 nm at a typical force of 12 pN, an amount within resolution of our instrument. **C)** Force-extension curve for a DNA hairpin. The experimental curve (red solid line) is fit to the extensible worm-like chain model (XWLC) [68,70] of the closed (black dashed line) and open (gray dashed line) hairpin.

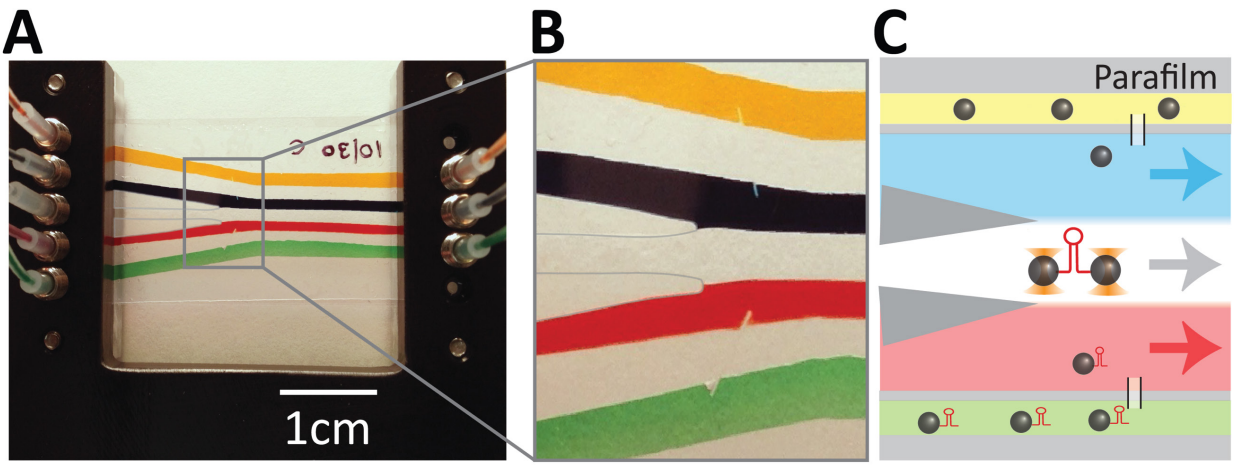

Figure 2.7 Example of a laminar flow chamber. **A)** Photograph of the laminar flow chamber used in this study. Channels are filled with colored food dye. Parafilm in central channel recolored to enhance contrast. **B)** Detailed photograph of the laminar flow channel. The bead channels (yellow and green) are connected via small capillaries to the central trapping region. Three streams (blue, uncolored, and red) meet in the center, but do not mix. **C)** Schematic of the flow chamber. Beads enter the central region through the capillaries. Tethers are formed in the center stream and then moved to neighboring streams as needed to change buffer conditions during experiments.

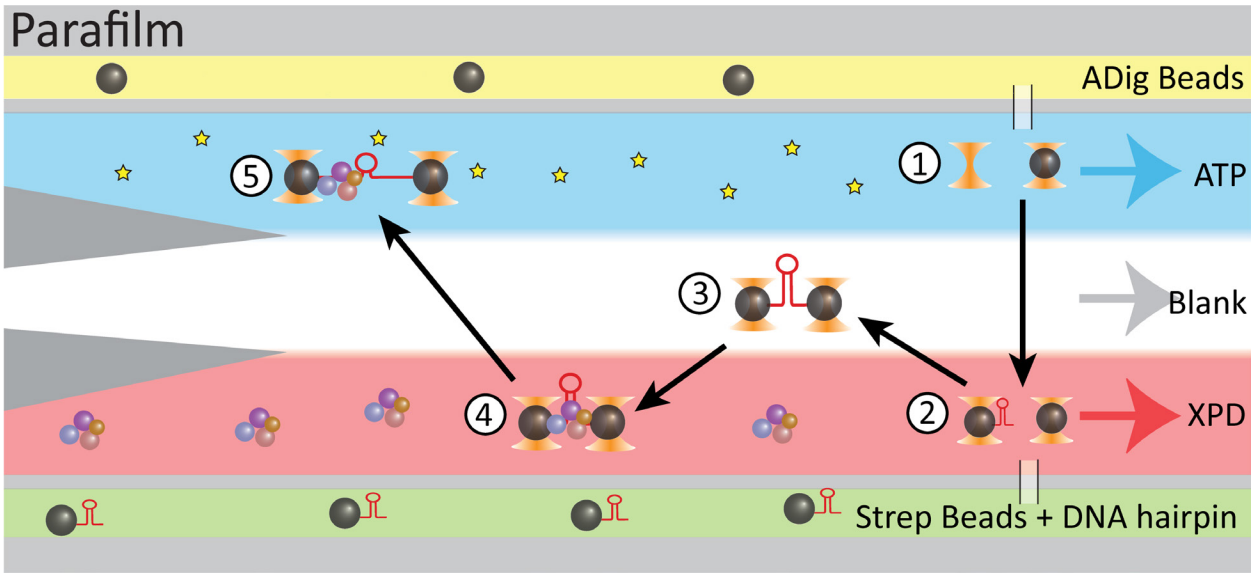

Figure 2.8 Schematic of experimental procedure. The top and bottom channels are filled with Anti-DIG and DNA hairpin-covered streptavidin beads, respectively, and beads flow into the central trapping region through small capillaries. In the central trapping region, the upper channel is filled with 500 µM ATP, the middle with 500 µM ATP-γS, and the bottom with 60 nM XPD and 500 µM ATP-γS. During an experiment, first (1) an Anti-DIG bead is trapped in Trap A near the upper capillary. Second (2), a streptavidin+DNA bead is trapped in Trap B near the lower capillary. Next (3), both beads are moved upstream and into the middle, blank channel. Here high sampling-rate calibration data are taken and then "fishing" begins as the beads are moved close together and we wait until a tether is formed. When a tether is formed, a force-extension curve is taken and is checked against the XWLC model. If the fit is good, the tether is next (4) dipped at low force into the XPD stream and we wait 30-60 seconds for a single XPD to bind. Finally, the force is increased to 12 pN and feedback is turned on to maintain that force. Then, the full construct is moved into the ATP stream and observed until the tether breaks or XPD dissociates.

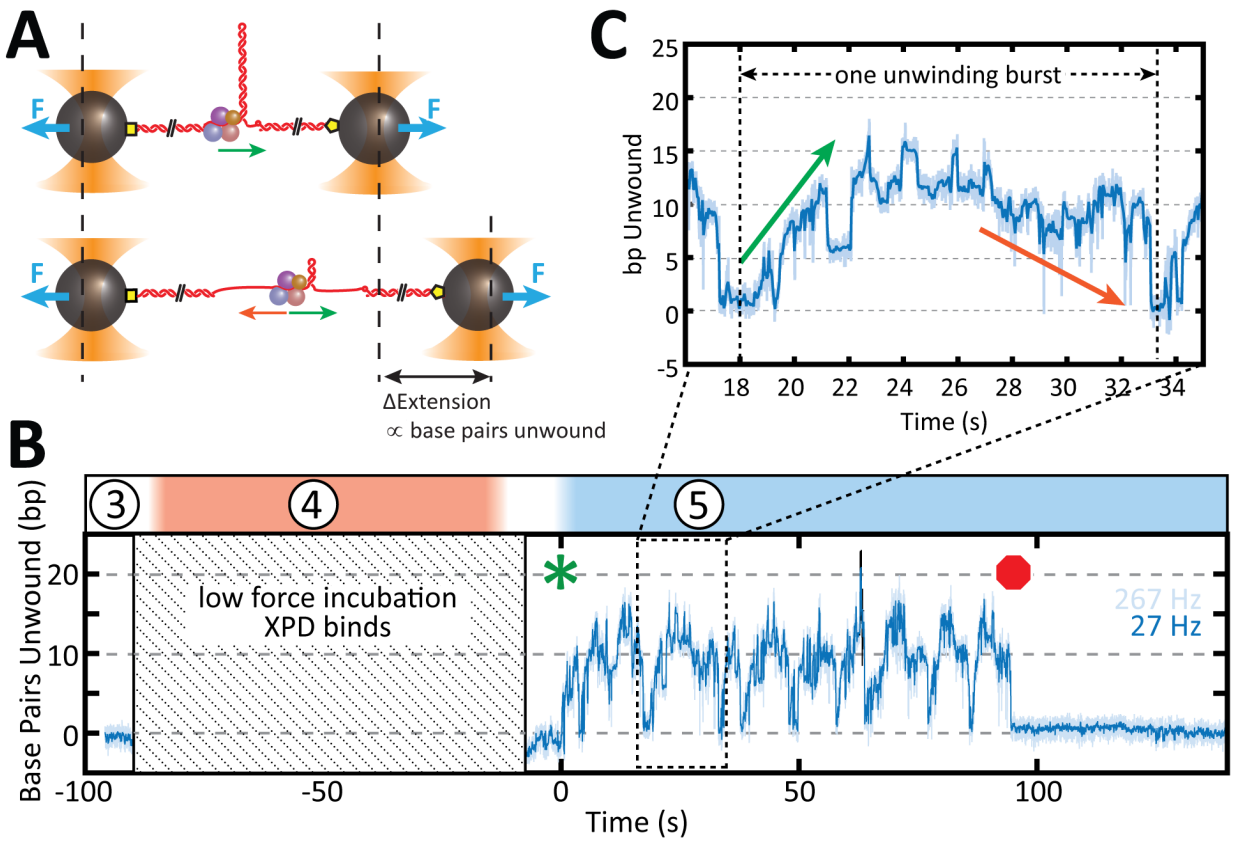

Figure 2.9 Results of a single XPD unwinding a DNA hairpin in the trap. **A)** Schematic of unwinding assay. The DNA tether is held at a constant force of 12 pN. As XPD unwinds the hairpin (green arrow) the total length of the tether increases. This increase in extension is proportional to the number of base pairs XPD has unwound. XPD can also back-step and back-slide (orange arrow) causing the hairpin to rezip and the total length to decrease. **B)** Example unwinding trace of a single XPD. Colored regions and numbers above the trace correspond to chamber locations and experimental procedure outlined in **Figure 2.8**. A tether is formed (3) and moved into 30nM XPD at low force (4) and held there for ~80 sec. Note that at low forces, calculations of base pairs unwound is unreliable hence that segment is blocked out. The force on the tether is increased as it is moved out of the XPD stream and into 500 µM ATP (5). Up to this point XPD cannot unwind because it does not have fuel and the hairpin remains at 0 bp unwound. At t = 0, XPD enters the ATP stream (green star) and we immediately see unwinding activity. XPD unwinds in repetitive "bursts" of activity for ~90 sec until it dissociates (red stop sign) and in its absence the hairpin closes completely once again (0 bp unwound). **C)** We define an unwinding burst as a period when XPD primarily unwinds (green arrow) followed by a period of primarily back-stepping and back-sliding, often back to the base of the hairpin. Bursts are of limited processivity, typically reaching a peak of ~15 bp unwound. This burst has a processivity of 18 bp.

Chapter 3: The effect of RPA2 binding on a DNA hairpin

In Chapter 1:, we noted that RPA2 is able to melt dsDNA on its own[27]. Indeed, this is a feature of many ssDNA binding proteins across many species[22,30–32]. This chapter will discuss RPA2 experiments in which it binds to and melts our DNA construct held under force. We will examine how many basepairs RPA2 can melt at a time, the rates of melting and binding to the substrate, as well as the effect of force and RPA2 concentration on these properties. This serves as both a control before examining XPD and RPA2 together, as well as providing valuable insights into RPA2 itself.

3.1 General Experimental Procedure

Experiments were performed as described in Section 2.2 using the same DNA construct with a binding site of 10 nucleotides. In many experiments, the same 3x1 layout laminar flow cells were used – 3 separate channels converging in the central trapping region (**Figure 2.7**). All channels were filled the same buffered Tris solution (Appendix C.2) to which was added:

- Upper channel: RPA2 (various concentrations)
- Middle channel: (Blank)
- Lower channel: (Blank)

DNA tethers were formed in the middle, blank channel, then force feedback was turned on and set to the desired force. Next, data were saved as the tether was moved into the upper RPA2 channel and melting activity was observed for ~1 min. Because of the redundancy of the middle and lower channels, some data was taken in a modified 2x1 layout laminar flow cell, in which the center trapping region was composed of only two streams meeting rather than three. In that case, tethers were formed in the lower blank channel and then moved at constant force into the upper RPA2 channel.

RPA2 does not hydrolyze ATP, so it is not necessary in these experiments. However, some of the data presented in this chapter were taken during combined experiments with XPD that will be described in Chapter 4:. In these experiments 500 µM ATP was included with RPA2 in the

upper channel. We do not see any evidence of the presence of ATP impacting binding or melting activity.

3.2 Results of Constant Force Experiments

In this set of experiments, the DNA tether was held at a 12 pN force (the same force at which the XPD-only experiments were conducted) and the concentration of RPA2 in solution was varied. In the absence of RPA2, the DNA hairpin remained closed. However, when RPA2 was added, short melting events are observed typically opening 5-10 bp (**Figure 3.1**). This is somewhat larger than the footprint of RPA2, which occludes ~4nt of ssDNA when bound[27]. These melting events become more frequent and appear longer-lived as the concentration of RPA2 is increased.

Figure 3.2 A shows a histogram of the number of basepairs open when exposed to RPA2 in solution at several concentrations. In the absence of RPA2, the hairpin remains closed and the histogram peaks at 0bp open with some spread due to noise. As we increase the concentration of RPA2, the peak corresponding to the closed hairpin persists, but the number of points in the 5-10bp range increases with RPA2 concentration. These points are hairpin melting events and we select a threshold of ≥4bp to pick out these events caused by RPA2 above normal hairpin noise.

Melting traces of the same concentration were concatenated and the time between melting events, t_1, was calculated (see **Figure 3.2 B**). The probability distribution of t_1 is plotted in **Figure 3.2 C** and fitting the distribution gives us the rate at which these melting events occur. The best fit for the probability distribution of t_1 is a bi-exponential with a slower rate ($k_{1,slow}$) and a faster rate ($k_{1,fast}$):

$$PDF = A\exp(k_{1,fast}t_1) + B\exp(k_{1,slow}t_1) \tag{3.1}$$

Due to the vastly different timescales of $k_{1,fast}$ and $k_{1,slow}$ ($k_{1,fast}/k_{1,slow}$ ~10-100), accurate fitting of the distribution is quite difficult. Larger bin sizes capture the longer binding timescale well but miss the fast dynamics entirely. Small bin sizes capture the fast action better but result in many empty bins at longer time scales that skew fitting in favor of faster rates. As a compromise, two different PDFs of t_1 were generated for each RPA2 concentration using different bin sizes, a small bin size of 0.01s and a larger bin size one calculated using the Freedman-Diaconis rule. The Freedman-Diaconis rule calculates a bin size for a histogram to

minimize the difference between the empirical and theoretical probability curves[73]. This was used because it gave consistently nice results across all the different distributions we used. **Figure 3.2 C** shows distributions and fits for the larger bin sizes. Values for $k_{1,slow}$ shown in **Figure 3.2 D** are the result of fitting the PDF with large bins. Values for $k_{1,fast}$ are the result of fitting the PDF with small bins.

We see in **Figure 3.2 D** that $k_{1,slow}$ increases linearly with RPA2 concentration while $k_{1,fast}$ remains roughly constant. Due to this concentration dependence, we believe the slower rate is the rate of RPA2 binding to the hairpin and the faster rate is the intrinsic melting rate of the DNA hairpin when RPA2 is bound. This model would produce fast dynamics when RPA2 is bound to the DNA with longer wait times between binding events.

Next, the duration of each melting event, t_2, was calculated (see **Figure 3.2 B**). The probability distribution of t_2 was fit to a single exponential with the rate, k_2, corresponds to the annealing rate of the hairpin. At low concentrations of RPA2, the annealing rate reaches the limit of our sampling rate with many events only lasting for the length of a single data point. For an accurate measure of this rate, data must to be collected at much higher rates. Despite this limitation, the decrease of k_2 as RPA2 concentration increases is somewhat surprising. This rate should be related to the hairpin reannealing rate as well as the dissociation rate of RPA2, which are generally expected to be concentration independent. However, recent work has demonstrated that a concentration-dependent dissociation rates can occur in the case of multiple binding sites[74]. This is certainly the case for RPA2 in our systems as the loading site can accommodate 2 molecules of RPA2 at once and more when the hairpin is melted. Thus, the next step in understanding RPA2's interaction with DNA is to observe directly the binding and unbinding of individual RPA2 molecules using fluorescence.

3.3 Experiments with Cy3 labelled RPA2

The results of Section 3.2 reveal the binding rates of RPA2, but better still would be to observe directly individual molecules of RPA2 binding to the hairpin. This is possible using fluorescently labeled RPA2 and our "fleezers" instrument that performs confocal microscopy alongside optical trapping. The fluorescent signal indicates when a molecule of RPA2 is bound to

the DNA. The magnitude of the signal is proportional to the number of molecules present and allows us to distinguish between activity due to a single RPA2 vs. multiple bound molecules. Tracking fluorescent RPA2 would also be valuable in experiments including XPD so we know precisely when one or multiple molecules of RPA2 are bound and can correlate the presence of RPA2 with XPD unwinding behaviors.

3.3.1 Fluorescence and Confocal Microscopy

A fluorophore is a molecule that can be excited to a high energy quantum state by a particular wavelength of light. The fluorophore then relaxes, via a different pathway, back to its ground state by emitting a photon of a longer wavelength. This emission is known as fluorescence and when a fluorophore is continuously exposed to light of the right wavelength, the process will happen hundreds of thousands of times, releasing a photon each time[75]. After some time (anywhere from seconds to a minute depending on the type of molecule and conditions) the fluorophore will become permanently damaged and no longer fluoresces[75]. This is known as photobleaching. In biological experiments, fluorophores are chemically bound to proteins of interest at a 1:1 ratio in order to track the protein's motion[76], binding/unbinding kinetics[40,56,77], and even conformational states[52,78].

Our fleezers instrument combines optical trapping with confocal microscopy, in which the excitation laser is focused to a diffraction-limited spot, known as a confocal spot, at the same focal depth (z position) as the tethered DNA[40]. The location of the confocal spot relative to the optical traps (x-y position) is adjustable and for these experiments we locate the spot halfway between the two trapped beads at the loading site of our DNA construct in order to detect molecules. The spot is ~10^7 nm^3 and any fluorophores within that volume will be excited and fluoresce. Our signal is the large increase in fluorescence due to a single molecule of Cy3-RPA2 binding at the loading site and thus the center of our confocal spot. However, there will also be a background signal from unbound molecules of Cy3-RPA2 passing through the confocal volume in solution, but not interacting with the DNA.

Emitted photons are detected using extremely sensitive avalanche photodiodes (APDs) capable of detecting a single photon. The number of photons detected increases linearly with the number of fluorophores present in the excitation volume. If a single fluorophore produces a

signal of 1000 a.u., two will produce 2000 a.u., etc. thus allowing us to count the number of fluorescently labelled molecules. By tracking the fluorescence signal over time, we can correlate data we get from the traps to changes in the fluorescence intensity[40,56].

3.3.2 Experiment Description

We followed the same experimental procedure outlined in Section 3.1, this time labelling RPA2 nonspecifically with a Cy3 fluorophore at the N-terminus using NHS-ester chemistry through a labelling protocol detailed in Honda *et al.* (2009) [28]. A 532-nm confocal excitation laser excited Cy3 directly. The excitation laser was turned on while the hairpin construct was in the blank stream and remained on as it was moved into the RPA2 stream and melting data was collected (**Figure 3.3 A**). Cy3 emission was tracked over time in addition to the usual trap signal.

3.3.3 Results

Despite observing melting events, no increases in Cy3 emission were observed above the background noise (**Figure 3.3 B and C**). This could be due to labelling inefficiency, i.e. RPA2 molecules were binding, but they did not have Cy3 attached. This is possible, but unlikely given concentration measurements that gave a 1.2:1 Cy3:RPA2 ratio in solution. Due to the quite high concentrations of RPA2 necessary to observe binding and melting, there is a large background from unbound molecules in solution passing through the confocal volume. As a general rule of thumb, the fluorescence intensity of one static molecule in the confocal volume is equivalent to the background from 10 nM concentration of molecules in solution. Thus, we expect our background for 50 nM Cy3-RPA2 to be 5x higher than the signal of a single Cy5-RPA2 binding, making small or short-lived signals difficult to detect.

This was surprising because previous experiments from our collaborators[28] and preliminary unpublished data from our lab[79] seemed to show long, clear binding events under similar conditions. We can attribute our collaborators' results to a much larger volume of data – thousands of molecules observed for 1 minute each using TIRF microscopy – making up for the rarity of longer-lived, higher intensity events. However, preliminary data from our own lab taken under similar conditions and using the same instrument seemed to show fluorescence from single RPA2 binding events. Attempting to replicate our lab's past results, the top stream was

filled with 200 nM RPA2. A tether was formed in the blank channel, then moved to the RPA2 stream and incubated for ~30 sec. After incubating, the tether was moved back into the blank stream and then the green excitation laser was turned on to detect Cy3-RPA2 molecules bound to the DNA hairpin (**Figure 3.3 D**). We observed a large, exponentially decaying fluorescence signal that is characteristic of large numbers of fluorophores photo-bleaching over time (**Figure 3.3 E**). This result does replicate the previous results from our lab[79].

However, the same experiment performed *without* a DNA tether produces the same decaying fluorescence signal. With no DNA tether to bind to and no molecules in solution in the central stream, this signal could only come from Cy3-RPA2 molecules bound to the trapped beads either through nonspecific "sticking" to the bead surface or the DNA molecules coating the beads. To test the "sticking" hypothesis, after dipping data were taken, the excitation spot was raster scanned over a 2D plane of both trapped beads and tether. The Cy3 intensity was measured at each point creating a 2D image, and the image was analyzed to find the total intensity of each bead. This total intensity of both beads was weakly correlated with the peak intensity at the loading site with and without a tether (**Figure 3.3 F**) indicating that the signal observed is from Cy3-RPA2 attached to the beads, not bound to the tether, refuting the previous results from our lab [79].

3.4 FRET Experiments with Cy5 labelled RPA2

To decrease the high background fluorescence, we looked to lower the excitation volume. A smaller volume means less unbound molecules in solution are excited, decreasing the background of our fluorescence data. To accomplish this, we used Förster Resonance Energy Transfer (FRET).

3.4.1 What is FRET?

FRET requires the use of two fluorophores, one "donor" and one "acceptor." The donor molecule is directly excited by a laser and fluoresces. The acceptor molecule is chosen to have an absorption spectrum overlapping with the donor's emission spectrum. When an excited donor molecule is sufficiently near to an acceptor molecule, it transfers its energy directly to the acceptor without releasing a photon. The acceptor in turn is excited and fluoresces normally,

emitting a photon of longer wavelength. The rate at which donor energy is transferred to the neighboring acceptor is proportional to R_0^6/r^6 where r is the distance between donor and acceptor and R_0 is the Förster radius of the particular pair of fluorophores. In our experiment, we use Cy3 as the donor and Cy5 as the acceptor, a commonly used FRET pair with Förster radius $R_0 = 5.4$ nm.

The FRET efficiency is the proportion of donor emissions that are absorbed and reemitted by the acceptor. It is typically found by independently tracking donor and acceptor emission intensities and calculated with the equation[80]

$$\text{F.E.} = \frac{I_A}{I_A + I_D} \tag{3.2}$$

where I_A (I_D) is the intensity of light emitted by the acceptor (donor). When the fluorophores are close together, there is high rate of FRET, and the FRET efficiency increases. The opposite holds when the fluorophores are far apart. Due to the $1/r^6$ dependence of the rate of FRET, FRET efficiency can be used to calculate inter-fluorophore distance and is commonly used as a molecular "ruler"[80,81]. In our case, however, we merely take advantage of the small Förster radius to decrease the effective size of our excitation volume. Our excitation volume has a diffraction limited diameter of ~250 nm and thus a volume of ~10^7 nm^3. Any molecules within this volume will be excited and fluoresce. In contrast, FRET will only occur within the Förster radius of a Cy3 donor molecule, a volume of $4\pi R_0^3/3 \approx 900$ nm^3 – a factor of 10^4 smaller! Thus, unbound molecules in solution are largely excluded, decreasing the background of our fluorescence data. In practice, the background is not reduced this much due to other factors described below, but we are able to see a clear FRET binding signal above the noise.

3.4.2 Experiment Design and Procedure

We designed and built a new hairpin construct, with a single donor fluorophore (Cy3) covalently attached between the ssDNA loading site and the dsDNA handle (**Figure 3.4 A**. Details in Appendix C.1). RPA2 is labelled nonspecifically with the acceptor fluorophore (Cy5) at the N-terminus as in Section 3.3. We directly excite the donor fluorophore with our confocal laser and independently detect the intensity of donor and acceptor emissions using two APDs. Because

FRET will only occur within ~5.4 nm of our Cy3 donor, we will only observe a high FRET efficiency when RPA2 is bound at the loading site.

Cy3-DNA tethers were formed in the blank channel, then force feedback was turned on and set to the desired force. Next, trap and fluorescence data were saved as the tether was moved into the Cy5-RPA2 channel (**Figure 3.4 B**). After the tether arrived in the Cy5-RPA2 channel, the 532-nm confocal laser was turned on to excite directly the Cy3 donor molecule. Activity was observed for the lifetime of the Cy3 molecule and saved for about 30 seconds after it photobleached, for ~1 min total.

3.4.3 Calculation of FRET efficiency from data

Now we have trap data showing the melting of our DNA hairpin (**Figure 3.4 C**) *and* fluorescence data showing Cy5-RPA2 binding (**Figure 3.4 D**). In a typical data trace, hairpin melting will be observed immediately upon entering the Cy5-RPA2 channel. Next, when the excitation laser is turned on, a strong Cy3 signal is observed from the donor molecule on our DNA construct. After about 30 s, the Cy3 molecule photobleaches producing an abrupt drop in Cy3 intensity that does not recover. Fluorescence observed after the photobleaching event is the background signal. In the Cy3 channel, this background is due to slight noise in the detector. In the Cy5 channel, the background is due to direct excitation of Cy5-RPA2 in solution by the excitation laser. The Cy5 absorption spectrum is non-zero at 532 nm, the wavelength of our excitation laser, thus there is some direct excitation of Cy5-RPA2 in solution in the confocal volume which produces background fluorescence. However, this background is orders of magnitude lower than the directly excited Cy3-RPA2.

The average intensity after photobleaching is subtracted from the Cy3 and Cy5 signals (**Figure 3.4 E**). FRET efficiency is then calculated from these background-subtracted donor and acceptor intensities (**Figure 3.4 F**). Due to the high background noise in the acceptor Cy5 signal, the background-subtracted acceptor intensity is often negative and produces negative values for the FRET efficiency, both of which are unphysical. However, increases in the FRET efficiency are more sensitive to anti-correlated changes in donor and acceptor intensity than to random noise in these intensities. Thus, high FRET efficiencies corresponding to Cy5-RPA2 binding can be observed rising above the noise (**Figure 3.4 F**).

3.4.4 Results

We again use thresholding to pick from our data high FRET events that correspond to Cy5-RPA2 binding. To select a threshold, we plot a histogram of all FRET efficiencies and find that it is dominated by a large, wide peak around 0 (**Figure 3.5 A**). This makes sense given the relative rarity of high FRET events. On close inspection, there appears to be a high FRET shoulder in the distribution above 0.5, so we select that values as our cut-off for selecting high FRET binding events (**Figure 3.5 B**).

We then use an OR logic gate to pick out all binding and melting events; finding times when FRET efficiency ≥0.5 OR base pairs melted ≥4bp. Because the fluorescence data were taken at a lower sampling rate (100 Hz), the trap data was filtered down from 267 Hz to match the fluorescence sampling rate and make correlating the data easier. For the duration of each event, we found the maximum FRET value and the maximum number of basepairs melted (**Figure 3.5 C**) and made a scatterplot of these values to see how often high FRET is correlated with melting (**Figure 3.5 D**).

Figure 3.5 D shows the scatterplot with each point representing one event. The (x,y) coordinates give the maximum number of base pairs melted and the maximum FRET efficiency during the event, respectively. The northwest quadrant contains events with high FRET efficiency (meaning RPA2 is bound) but during which no melting occurs. The southeast quadrant contains melting events without an increase in FRET (meaning Cy5-RPA2 is not bound). The northeast quadrant contains correlated events where we see simultaneous signals for RPA2 binding *and* hairpin melting.

We had expected to see a majority of correlated events in these data. However, it is immediately obvious that this is not the case and correlated events are quite rare. Only about 13% of all binding events detected by FRET, melt the hairpin. Melting events were linked to FRET-detected RPA2 binding 9% of the time. How could this be?

3.4.5 Discussion

Binding without melting can be explained by the relatively large loading site on our DNA construct. RPA2 has a footprint of 4 nt when bound to ssDNA[27]. Our loading site was designed to be 10 nt to accommodate XPD binding, but this allows RPA2 binding some distance from the

hairpin. If RPA2 must be bound immediately next to the hairpin in order to melt it, then there will be many cases of RPA2 binding too far away to melt the hairpin. 10 nt of ssDNA provide seven distinct 4 nt binding locations. If all of those locations give a high FRET signal (as we would expect), but only one can melt the hairpin, then only 1/7 or ~14% of high FRET events will be associated with hairpin melting, quite close to our result of 13%.

More unsettling is the large number of melting events that are not associated with Cy5-RPA2 binding. After all, how could the hairpin melt if RPA2 is not present? Generally, low labelling efficiency, the fraction of RPA2 molecules that have Cy5 bound, is responsible for this type of mismatch between observed activity and fluorescence detection. If there are many "dark" molecules with no Cy5 attached they may be responsible for the activity but cannot fluoresce. However, fluorescent gel imaging combined with NanoDrop measurements show a very high efficiency near 100% (See Appendix B.5 for details).

The Cy5-RPA2 may also be less capable of binding and melting than unlabeled RPA2 if the attached dye is in a location that interferes. Sometimes the process of adding a fluorescent dye damages the protein. Most labelling protocols require room temperature incubations for several hours and the exchange of the suspending solution. Some proteins can be very sensitive to these changes and lose function entirely. Indeed, even without the fluorescence data, we find the melting rate of Cy5-RPA2 to be lower than that of unlabeled RPA2 (**Figure 3.5 D**).

Alternatively, the particulars of our experiment may be lowering the FRET efficiency between our donor and acceptor. We assume $R_0 = 5.4$ nm for our FRET pair, a value commonly used in the literature, but the exact number depends on many factors including the orientation of the dyes. $R_0^6 \propto \kappa^2$ where κ^2 is an orientation factor for dipole-dipole interactions[82]. In the isotropic case, where both dyes freely rotate in all directions, the average value of $\kappa^2 = 2/3$ and this case is assumed in most experiments. However, by using an internal Cy3 label in our DNA construct and applying force, we are restricting the motion of the donor dye in the direction of the applied force. Thus, the donor can freely rotate around the axis of applied force, but it cannot rotate about any other axis. Even assuming the acceptor Cy5 orientation is isotropic, this could have a significant effect on R_0.

The distance from the Cy3 label on the DNA to the duplex junction is ~4.2 nm, based on length of the 10 nt loading site held at 12 pN tension. If $R_0 = 5.4$ nm as we assumed, then we expect a single molecule of Cy5-RPA2 binding next to the junction to give a FRET efficiency of 0.8. Thus, any Cy5-RPA2 that binds and melts dsDNA would be associated with high FRET. However, if the actual value of R_0 is less than this due the constrained orientation, we may see quite low FRET efficiency from Cy5-RPA2 binding near the junction. Hypothetically, if $R_0 = 4$ nm due to this constraint, this would correspond to a FRET efficiency at the junction of only 0.4, below the "high FRET" threshold we set and lost in the background noise. In this case, Cy5-RPA2 may be simultaneously binding and melting, but we do not register the fluorescence signal as significant and they appear as "melting only" events.

3.4.6 Conclusion

In the end, while our FRET method is exciting, many changes must be investigated to make the method practicable for observing RPA2 binding. If the Cy5 label is interfering with binding, a new labelling technique will need to be used to prevent this interference. If the orientation of the donor dye is causing a decrease in FRET efficiency, changing the labelling method of the DNA hairpin could help this. These and other options are discussed in Section 3.5.

3.5 Future Work

Suggested avenues of future investigation into RPA2.

3.5.1 More force and concentration data

Currently, we have data at 12 pN for a variety of concentrations of RPA2 and data at a variety of forces for 50 nM RPA2. In future work, more data must be taken in the rest of the force-concentration space. This would greatly increase our understanding of the energy landscape of RPA2 interacting with DNA.

3.5.2 Specific labelling of RPA2 mutants

Observing RPA2 binding via fluorescence would add greatly to this work. Given that non-specific labeling appears to interfere with RPA2 binding and melting DNA, specific labeling to an

added cysteine residue could be used. The labelling site could be selected to find an equally active labeled version of RPA2 allowing for more accurate fluorescence observations.

3.5.3 Alternate labelling scheme DNA construct in FRET experiments

Changing the Cy3 dye on our DNA hairpin from an internal label to an external tag hanging off the DNA backbone would allow for isotropic rotation of the dye missing from the current construct. If constrained rotation did indeed decrease R_0, this change may raise FRET efficiency near the junction and increase detection of Cy5-RPA2 molecules simultaneously binding and melting DNA.

Though $R_0 = 5.4$ nm should adequate for detecting binding events in the current geometry, another way to increase the FRET efficiency at the junction is to move the donor dye closer to the junction. This could be accomplished by placing the Cy3 label on the opposite side of the hairpin rather than on the same side as the loading site. There it can be placed closer to the junction with less chance of interfering with XPD unwinding.

3.5.4 Alternative analysis of melting states using Hidden Markov modelling

In this chapter, melting events were selected by simple thresholding – anything above 4 bp was considered "melted." This is enough to approximate of the effect of RPA2 on melting dsDNA, but it has two major drawbacks. First, thresholding treats every melting event above 4 bp as equal and lumps them all into one state. This ignores distinctions between different amounts of melting that may be important. It also erases any hairpin dynamics that occur above 4 bp. We speculate these dynamics are due to multiple RPA2 molecules binding and unbinding and adding them to our analysis would give much more information on RPA2 binding kinetics even without fluorescence measurements. Second, melting events that do rise above normal hairpin noise, but are less than 4 bp are not counted. One such event is visible in **Figure 3.2 B** a bit after t = 30.5 sec. This is necessary trade off and any method will miss some legitimate melting events caused by RPA2, but minimizing false negatives is always preferred.

One option is to still use thresholding, but add more than two states, e.g. creating states around 0, 5, 10, and 15 bp. While this would give more information on melting dynamics, the

threshold boundaries would be somewhat arbitrary. The histogram of base pairs melted in **Figure 3.2 A** does not lend itself to easy divisions into multiple higher order states.

Hidden Markov modelling provides a more objective tool for defining the most likely set of underlying melted states and for fitting data to those states. Melting events have such short lifetimes (often only a single data point) that they are impossible to fit by traditional step-finding methods such as Student's t-test which compare the mean values of neighboring points to locate steps in data. Markov modelling does not rely on these averages, rather finding the best-fit state for each data point independently. Thus, we have found Markov Modelling performed using the HaMMy software[83] reliably selects even single data point events above background noise.

3.6 Figures

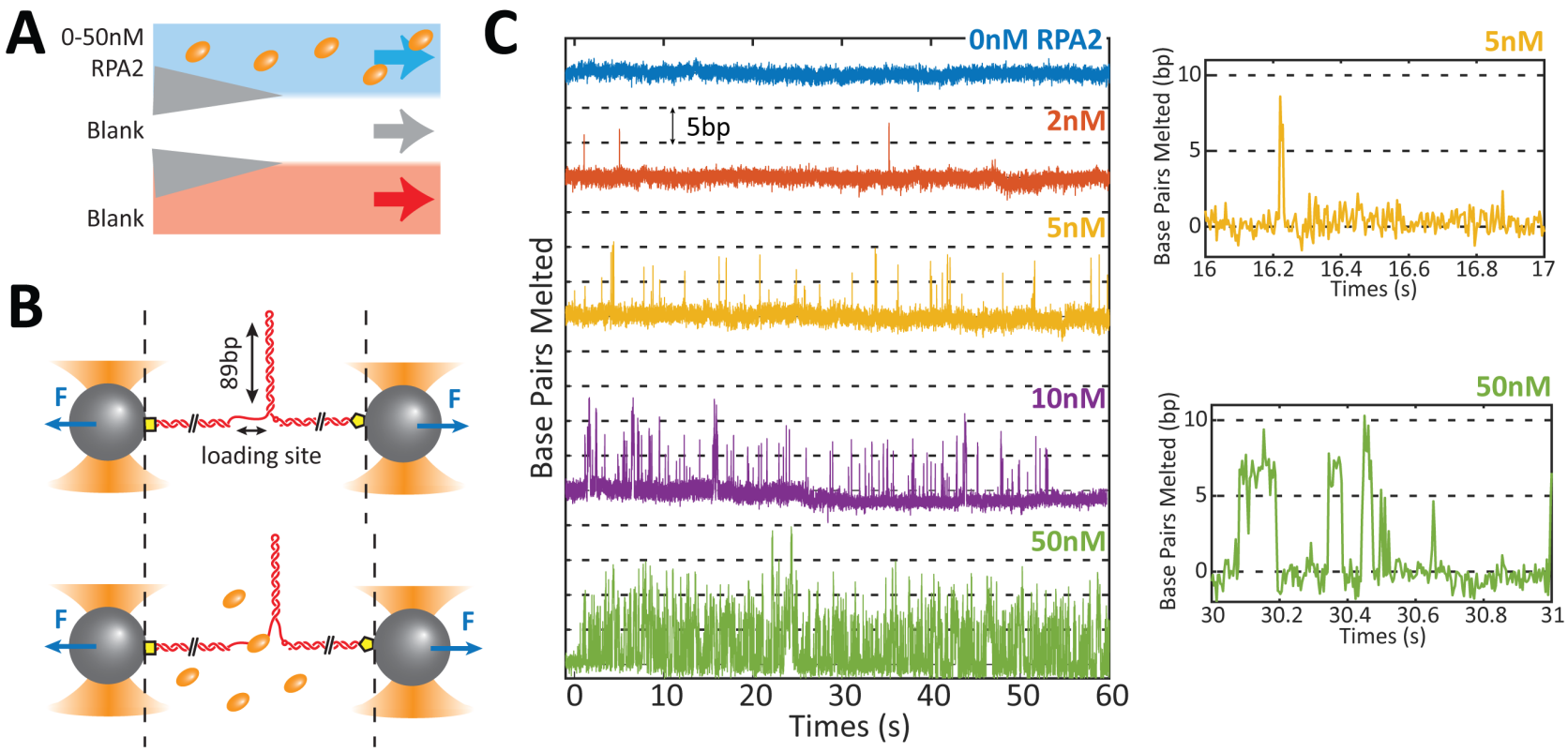

Figure 3.1 Basic RPA2 experimental setup and sample traces. **A)** Schematic layout of the sample chamber, RPA2 in the top stream and blank on the bottom. **B)** The same hairpin construct described in Section 2.2 is held at constant 12 pN tension. When RPA2 binds, dsDNA in its vicinity is destabilized and the hairpin partially melts. **C)** The hairpin enters the upper stream at t=0. When no RPA2 is present, the hairpin remains closed. In the presence of RPA2 however, short-lived 5-10 bp melting events are observed. The frequency of these events increases as RPA2 concentration is increased (top to bottom). Inset: 1 s fragments from the 5nM and 50 nM example traces. At 5 nM RPA2, the melting event is very brief (< 0.1 s). At 50 nM RPA2, melting events are both more frequent and occasionally longer lived.

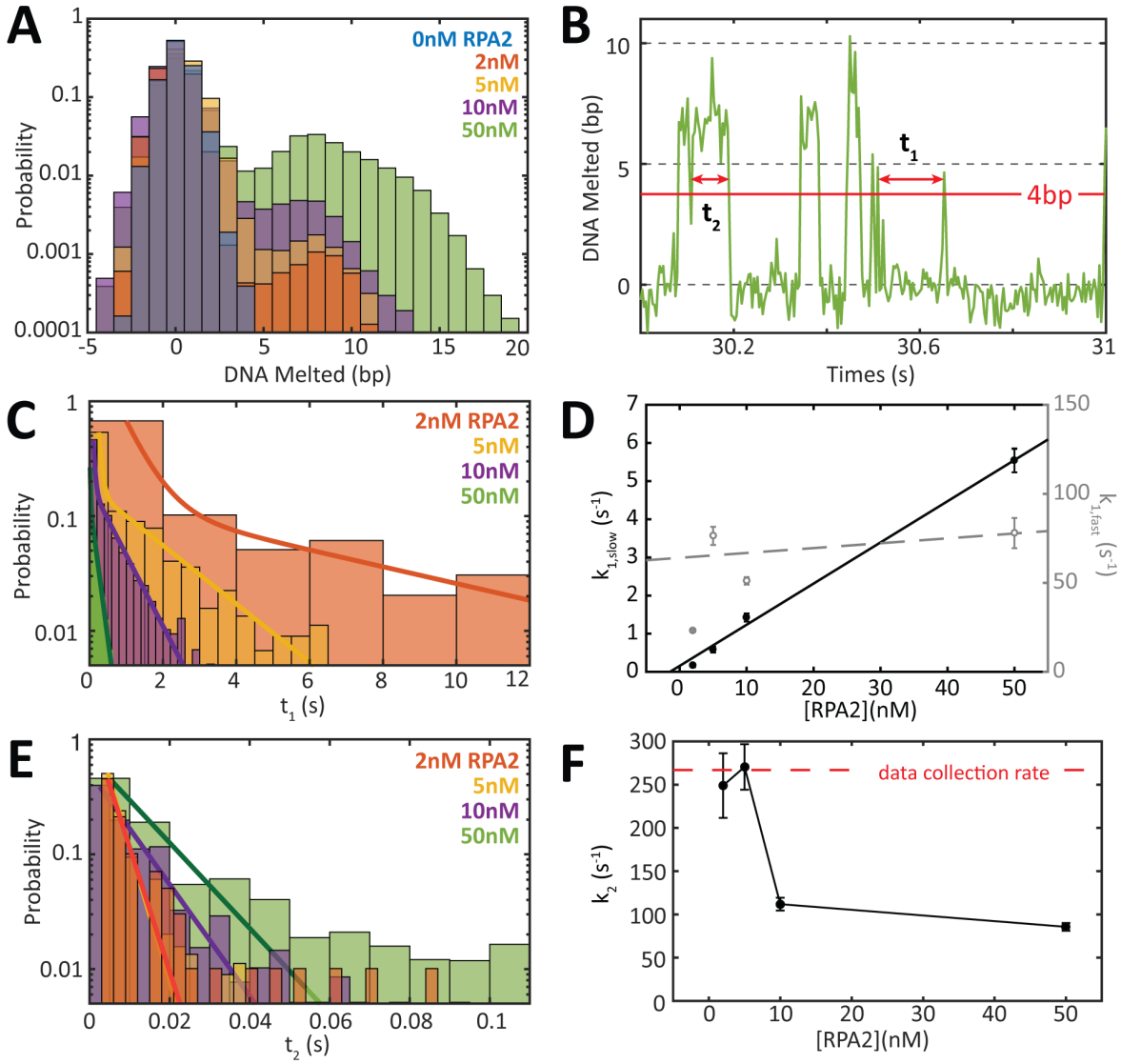

Figure 3.2 Kinetics of hairpin melting by RPA2. **A)** Histogram of DNA melted by RPA2. The large peak at 0 bp represents the completely closed hairpin. A second peak can be seen above 4 bp melted that increases with RPA2 concentration. We use 4 bp as a threshold to pick out melting events in our data. These melting events are relatively rare compared to the closed hairpin, the logarithmic scale makes the peak appear more pronounced. **B)** An example trace in 50 nM RPA2. Melting events above 4 bp are selected. Melting kinetics are determined using the time between events, t_1, and the duration of each event, t_2. **C)** The probability distribution of t_1 is best fit by a bi-exponential with one fast and one slow rate. **D)** Fast and slow melting rates with 95% confidence intervals from t_1 distribution fit. The slow rate, $k_{1,slow}$, increases with RPA2 concentration, while the fast rate, $k_{1,fast}$, remains roughly constant. **E)** The probability distribution of t_2 is exponential with the possible exception at 50 nM RPA2. **F)** The reannealing rate is plotted with 95% confidence intervals from t_2 distribution fit. It reaches the detection rate at low concentrations but decreases as the amount of RPA2 increases.

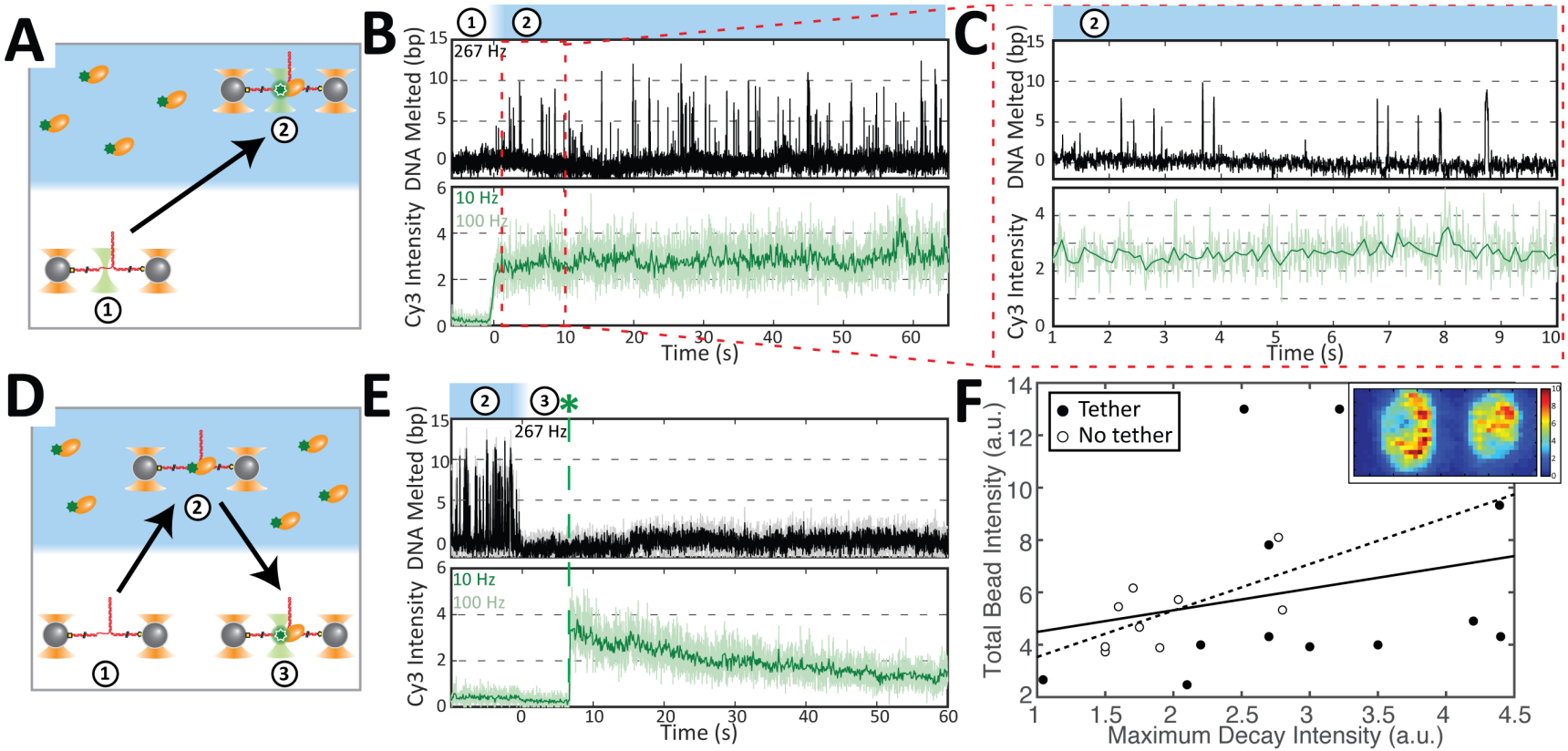

Figure 3.3 Cy3 labeled RPA2 experiment data. **A)** Schematic of experimental procedure for monitoring binding and melting simultaneously. (1) A DNA tether is formed in the blank stream, held at 12 pN tension, and the confocal excitation laser is turned on. (2) The tether is moved into Cy3-RPA2 in the top stream and hairpin melting is measured along with Cy3 emission intensity. **B)** A typical data trace. The tether enters 50 nM Cy3-RPA2 at t = 0 s. An immediate increase in Cy3 emission is observed (green) as well as hairpin melting events. **C)** A zoom in of the previous graph. Note the lack of clear increases in Cy3 emission above the background noise that would correspond to single molecule binding events. **D)** Schematic of experimental procedure recreating previous "dipping" experiments. (1) A DNA tether is formed in the blank stream and held at 12 pN tension. The confocal excitation laser is off. (2) The tether is moved into Cy3-RPA2 in the top stream and hairpin melting is observed. (3) After a 30 s incubation, the tether is moved back to the blank stream and the excitation laser is turned on. **E)** A typical data trace for this experiment. The tether exits 200 nM (**cont.**)

(**Figure 3.3 cont.**) Cy3-RPA2 and enters the blank stream at t = 0 s and melting activity stops immediately. Then the confocal laser is turned on at t ~ 8 s (green star and dashed line) and Cy3 emission is observed. Emission is high initially and decays over time, consistent with many fluorophores present and slowly bleaching. This same emission signal is observed even when no tether is present indicating it is not due to Cy3-RPA2 bound to the DNA. **F)** The Cy3 intensity at the surface of each trapped bead was also measured after dipping for each trace. The total intensity of both beads was plotted against the peak intensity observed between the two beads at the hairpin location and a weak correlation was found with a tether (dotted line r=0.22) and without (solid line r=0.64). Inset: Sample bead image in the Cy3 emission wavelength after incubation in Cy3-RPA2. Anti-digoxigenin bead on the left. DNA-coated streptavidin bead on the right.

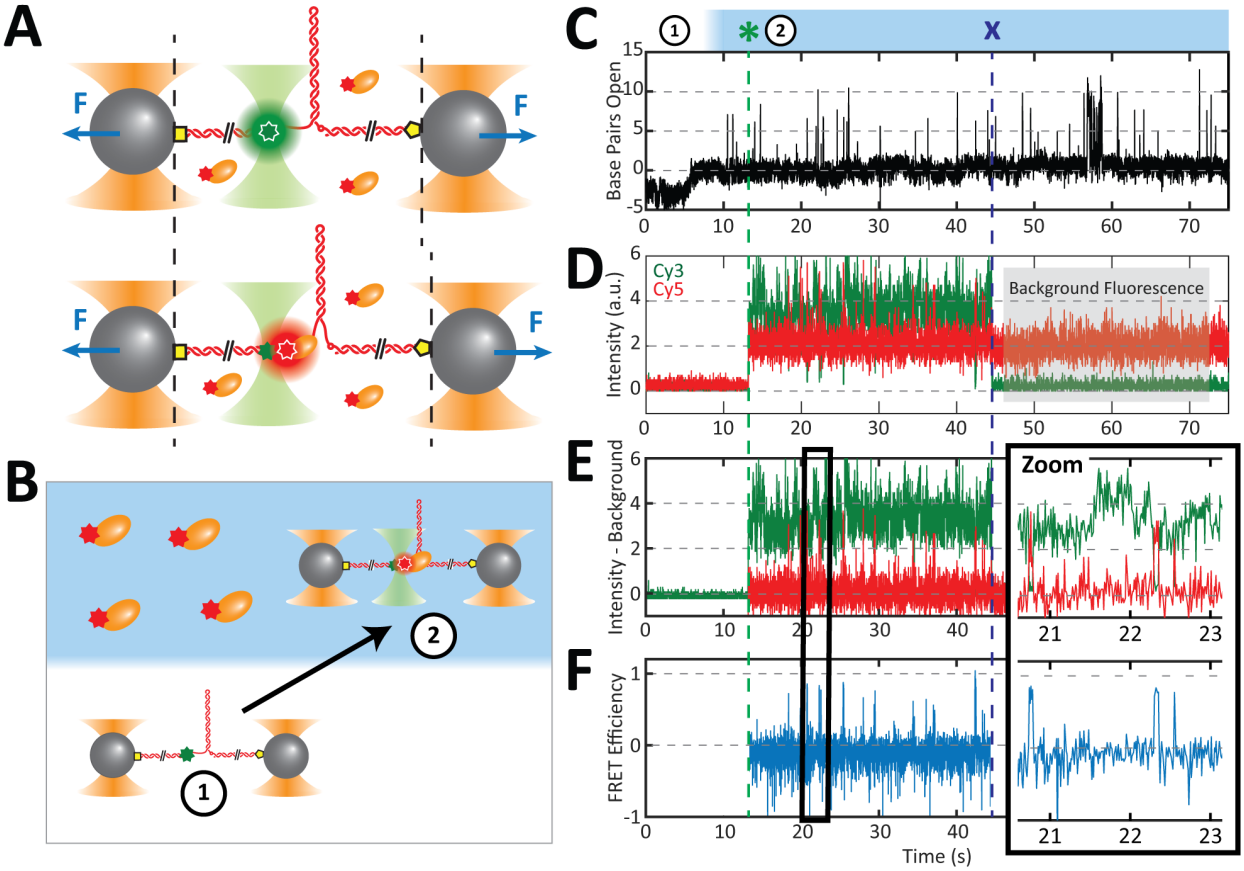

Figure 3.4 FRET Assay overview. **A)** With the Cy3 labeled hairpin construct, we expect to see high Cy3 intensity when there is no RPA2 bound. When RPA2 does bind, we expect the Cy3 intensity to decrease, the Cy5 intensity to increase, and melting to occur simultaneously. **B)** In these experiments, data taking begins at constant force in the blank channel with the excitation laser turned off (1). After arriving in the Cy5-RPA2 channel, the excitation laser is turned on and simultaneous fluorescence and melting data are captured. **C-F)** A sample of melting and fluorescence data. The tether enters the Cy5-RPA2 channel t=~10 s and the excitation laser is turned on just after that (green star and dashed line). **C)** Shows hairpin melting due to Cy5-RPA2 binding. **D)** Shows the raw intensity data collected tracking both Cy3 and Cy5 emission. At t=~45 s the Cy3 dye on the hairpin photobleaches (dark blue X and dashed line). After this point, the Cy3 signal drops to zero and any Cy5 emission is not due to FRET, but direct excitation of the Cy5 dye by the excitation laser. **E)** We subtract the background intensity from our raw signal and use these values to calculate FRET efficiency. **F)** The calculated FRET efficiency shows a majority of points around 0 with occasional high FRET "blips," similar to the melting signal. Inset: Zoom in of fluorescence data. Here the momentary increase in Cy5 intensity along with a decrease in Cy3 intensity is clearer. These points correspond to the high efficiency spikes in the FRET trace. (NOTE: Intensities in a.u. in this figure cannot be directly compared to intensities given in **Figure 3.3**.)

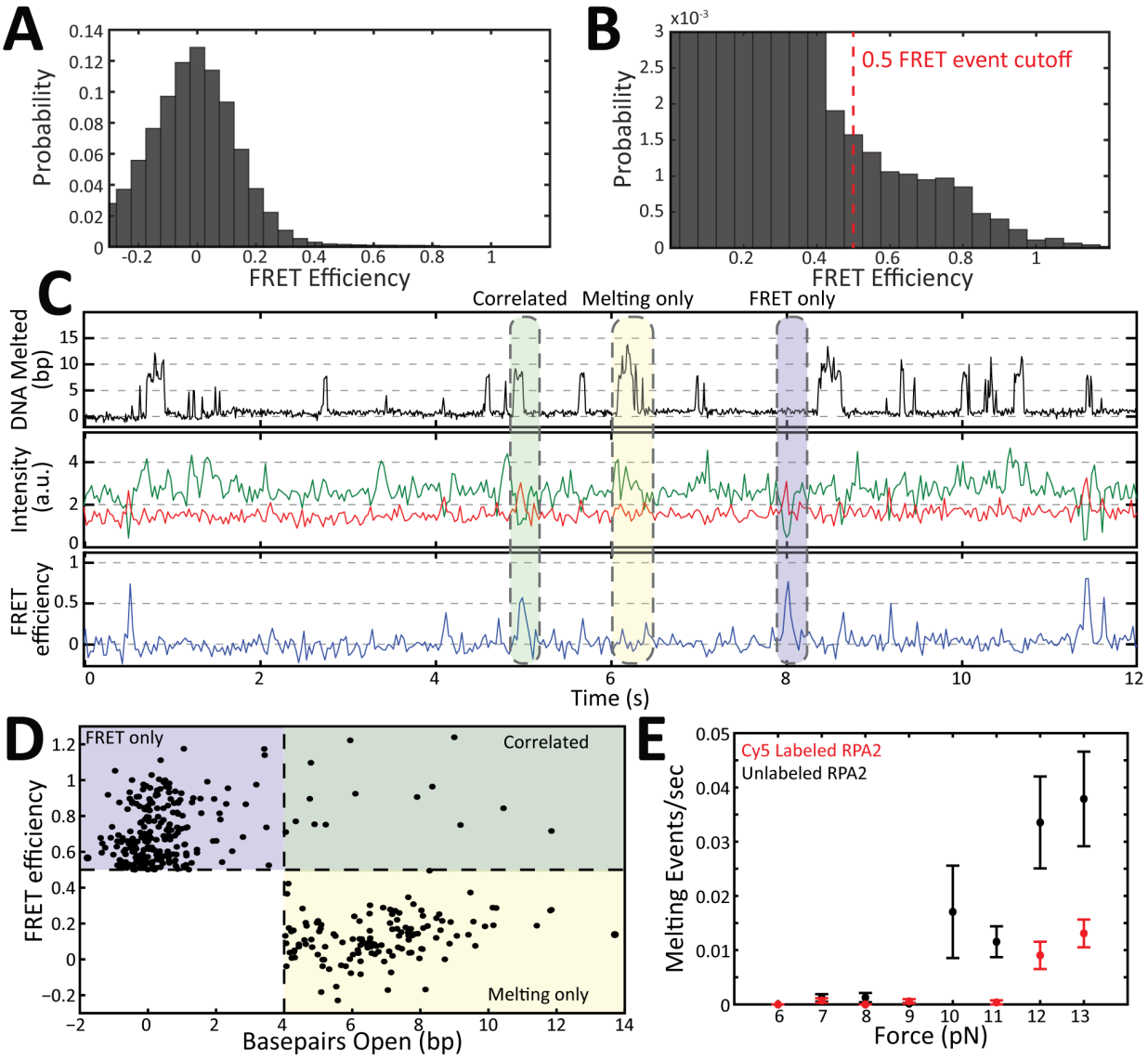

Figure 3.5 Observing binding events with FRET. **A)** A histogram of all measured FRET efficiencies shows a large, wide peak around zero. High FRET events are rare and are only visible if we change the scale. **B)** A zoomed in image of panel **A** showing the distribution of high FRET events. From this we select a cutoff of 0.5 to identify binding events. **C)** A sample trace in 50 nM Cy5-RPA2 highlighting three types of events – correlated events, melting only events, and FRET only events. Ideally, we would expect each melting event to be associated with a high FRET signal indicating binding, i.e. mostly correlated events. **D)** Each point on this graph is one event with the x-y coordinates corresponding to the maximum base pairs melted and FRET efficiency during the event respectively. This clearly shows the vast majority of events are uncorrelated. **E)** A possible explanation for the uncorrelated melting, Cy5-RPA2 melts the DNA hairpin at a slower rate than unlabeled RPA2.

Chapter 4: XPD unwinding with RPA2

In this chapter, we will describe experiments with XPD unwinding in the presence of RPA2 and discuss the effect of RPA2 on XPD unwinding behavior. We will then test the effects of different proposed models of XPD and RPA2 interaction against our data in order to support or refute them as possible mechanisms of unwinding enhancement. Finally, we will propose our own mathematical state model to describe XPD unwinding and fit it to our data.

4.1 Experimental Procedure and Initial Results

4.1.1 Procedure

We performed a variation on the basic single XPD unwinding experiment, detailed in Section 2.2, to investigate XPD unwinding in the presence of RPA2. To do this, we add varying concentrations of RPA2 to the stream with ATP to observe XPD unwinding with RPA2 in solution (**Figure 4.1 A**). In the central trapping region, suspended in a buffered Tris solution (Appendix C.2), we have

- Upper channel: 500 µM ATP + 0-50 nM RPA2
- Middle channel: 500 µM ATP-γS
- Lower channel: 60 nM XPD + 500 µM ATP-γS

As a reminder, ATP is the fuel molecule XPD uses to unwind. ATP-γS is a non-hydrolysable form of ATP, meaning XPD can bind to it, but not get energy from it. The binding of ATP, even without hydrolysis, can alter proteins and in some cases allows them to bind more readily their substrates. ATP-γS was added to the XPD and blank channels in an effort to increase the efficiency of XPD binding to the DNA loading site. Due to the lack of consistent fluorescent detection of labelled RPA2, detailed in Section 3.4, we return to using unlabeled RPA2.

A DNA "tether" is formed in the middle ATP-γS channel and we begin saving constant force data. The tether is held at low force (5 pN), moved into the lower XPD stream and incubated there for ~40 sec to allow XPD to bind. Following incubation, the force on the tether is increased to 12 pN and the tether is moved into the upper ATP+RPA2 channel. Upon exposure to ATP, an XPD molecule bound at the loading site will begin to unwind the hairpin ahead of it. Unwinding

data are saved until the tether breaks or XPD dissociates. We then compare unwinding with several different concentrations of RPA2 in solution to see what, if any, effect RPA2 has on XPD unwinding.

4.1.2 XPD processivity increases with RPA2 concentration

As in previous chapters, our data tracks the number of base pairs open at the base of the DNA hairpin vs. time. XPD unwinds in repetitive "bursts," the start and end points of which are selected manually (**Figure 4.1 B** inset). The processivity of each unwinding burst is the maximum number of base pairs unwound during that burst. Previous results have shown that RPA2 is able to enhance XPD activity, but they could not definitely show an increase in processivity[27,28]. Our results show that RPA2 does increase XPD processivity with the average processivity per burst increasing with RPA2 concentration (**Figure 4.1 B**).

Looking at the distribution of burst processivity (**Figure 4.1 C**), in the absence of RPA2, XPD unwinds primarily in low-processivity bursts, with a large drop off in bursts with > 25 bp processivity. This is consistent with previous work that linked this limited processivity to the sequence being unwound. GC pairs have a higher base pairing energy than AT pairs and XPD tends to stall in GC-rich regions. The sequence we use has one such region around the 15 bp position. As RPA2 concentration increases, more and more bursts fall into the high processivity tail of the distribution with many unwinding the full 89 bp of the hairpin. We use 25 bp as a threshold to distinguish between low- and high-processivity unwinding types and seek to determine any characteristic differences between the two types that could produce the difference in processivity.

4.2 Back-stepping is not affected by presence of RPA2

The next few sections will reintroduce proposed mechanisms of RPA2's enhancement of XPD unwinding from Section 1.3 and, if true, what measurable effects they would have on XPD velocity, stepping rates, etc. We will then compare these expected outcomes to what we observe in our data to confirm or refute the hypotheses.

4.2.1 ssDNA sequestration

In the ssDNA sequestration model, RPA2 binds to the ssDNA behind XPD and acts as a physical barrier preventing the helicase from stepping or sliding backward (**Figure 4.2 A**). XPD has been shown to back-step and back-slide often during unwinding[19]. Sequestration would force XPD to move forward or not at all increasing the amount of DNA unwound over time.

4.2.2 Burst duration is constant vs RPA2 concentration

If RPA2 sequesters ssDNA behind XPD, we would expect longer burst duration in the presence of RPA2 because XPD cannot back-step or backslide effectively. This would be the case even for non-processive bursts since all back-stepping will be inhibited. However, non-processive bursts are of similar duration at all concentrations of RPA2 (**Figure 4.2 B**; gray data). This indicates RPA2 is not inhibiting back-stepping. While we do observe an overall increase in burst duration with RPA2 concentration (**Figure 4.2 B**; black data), this is entirely due to having a higher fraction of high-processivity bursts, which are of longer duration (**Figure 4.2 B**; white data).

4.3 Testing the duplex melting model

4.3.1 Duplex melting model

In the duplex melting model, RPA2 binds to the DNA strand near the ssDNA-dsDNA junction and melts a few base pairs, creating ssDNA ahead of XPD (**Figure 4.3 B** inset). Because XPD unwinds via a "largely passive" process, relying on thermal energy fluctuation to help it break the base pairing bond ahead of it, RPA2 enhancing the effect of thermal fluctuations could also enhance XPD's unwinding potential.

4.3.2 Selecting RPA2 binding events with a step-fitting algorithm

As described in Chapter 3:, RPA2 alone is able to destabilize a DNA hairpin and cause it to melt by 4-8 bp at constant 12 pN force, though the melting is short-lived, ~20 ms. Melting events are also observed while XPD unwinds in the presence of RPA2 (**Figure 4.3 A**). This means RPA2 is binding ahead of or alongside XPD at the fork junction and melting DNA ahead of the helicase. Does this melting aid XPD in unwinding?

We begin by detecting steps in the unwinding data using a chi-squared reduction algorithm by Kerssemakers *et al.*[84]. The unwinding segment of each burst, from the start to the maximum processivity, was fit in its entirety. This algorithm works reliably on short stretches of data with not too many or very regular steps. That is good enough for the short, low processivity traces analyzed in the past[19]. But longer traces and traces with very high processivity are inconsistently fit, over-fit in some places and under-fit in others, making stepping rates calculated by this method unreliable. This algorithm is able to do a decent job of picking out RPA2-like melting events in the data (**Figure 4.3 A**).

RPA2-like melting events are characterized by short-lived melting of ~4 bp of the hairpin. To select RPA2-like events in our fits, we search for step pairs that meet the following criteria: > 2 bp forward step (step n) followed by > 2 bp back-step (step n+1) within 0.06 s of each other (**Figure 4.3 B** outlined in red). While this type of event is observable even in the absence of RPA2, the probability does increase with RPA2 concentration confirming their connection to RPA2 activity (**Figure 4.3 C**).

4.3.3 Alignment of detected RPA2 events shows no effect on forward motion

To test if duplex destabilization aids XPD unwinding, we what to see how these melting events impact the forward progress of the helicase. If duplex destabilization aids in unwinding, XPD may be more likely to advance along the hairpin during melting events. We use the fit steps and selection criteria to locate instances of RPA2-like melting during unwinding (**Figure 4.4 A** schematic). We then normalize each event such that it begins at 0 bp before melting (**Figure 4.4 B** schematic). Finally, we divide each event into melting and reannealing segments that are independently aligned. All melting segments are aligned such that melting begins at time t=0 (**Figure 4.4 C** "before" schematic). And likewise, reannealing segments are aligned such that reannealing occurs at time t=0 (**Figure 4.4 C** "after" schematic).

We use a 2D kernel density estimation to visualize the probability density of these aligned traces. In this method, each point in the aligned data replaced by a kernel function, in this case a 2D Gaussian. The sum of all kernels gives an approximation of the probability density. The darker color in **Figure 4.4 D** and **E** indicate higher probability. **Figure 4.4 D** shows that after a melting event, the hairpin is most likely to reanneal all the way back to where it was before melting

occurred. This means XPD has not stepped forward during melting and indicates the melting event has had no impact on XPD movement.

As a control, we asked how likely XPD was to unwind at all in a similar time period without melting. To accomplish this, we select random points during unwinding and points 0.6 s afterward and aligned them as we did with the melting and reannealing steps. **Figure 4.4 E** shows the results, that XPD is most likely to show no net progress in this amount of time.

The fact that XPD does not move forward during RPA2-llke melting rules out the duplex melting model as an explanation for unwinding enhancement. It may be that the time period of these melting events is too short for XPD to take advantage of.

4.4 Tight binding complex does not form

4.4.1 Complex formation model

In the complex formation model, RPA2 and XPD interact directly and join together in a complex that is more processive than XPD alone, perhaps because RPA2 binds to XPD in a way that forces XPD into a more processive conformation (**Figure 4.5 B**). Experiments from the Spies Lab show that *Fac*XPD and *Fac*RPA2 do no form a stable complex in solution, however they used a co-immunoprecipitation pulldown assay that requires very tight binding to show an effect[27]. It is possible that XPD and RPA2 form an unstable, shorter-lived complex or they may require DNA to mediate their interaction.

4.4.2 Amended experimental design

To test this model, we amended our experiment attempting to load a preformed complex of XPD and RPA2 before entering ATP to unwind (**Figure 4.5 A**). Varying concentrations of RPA2 were added to the stream with XPD. In the central trapping region, suspended in a buffered Tris solution (Appendix C.2), we have

- Upper channel: 60 nM XPD + 500 µM ATP-γS + 0-200 nM RPA2
- Middle channel: 500 µM ATP-γS
- Lower channel: 500 µM ATP

A DNA tether is formed in the middle ATP-γS channel and we begin saving constant force data. The tether is held at low force (5 pN), moved into the upper XPD+RPA2 stream and

incubated there for ~40 sec to allow an XPD+RPA2 complex. Following incubation, the force on the tether is increased to 12 pN and the tether is moved into the lower ATP-only channel. Upon exposure to ATP, the XPD+RPA2 complex bound at the loading site will begin to unwind the hairpin. Unwinding data is saved until the tether breaks or XPD dissociates. We then compare unwinding with several different concentrations of RPA2. If an XPD+RPA2 complex exists, we would expect its formation to be more likely at as we increase the concentration of RPA2 in solution with XPD.

4.4.3 Processivity constant with RPA2 concentration

In complex formation experiments, RPA2 has no effect on processivity (**Figure 4.5 C**). Highly processive unwinding bursts are equally likely with 200 nM RPA2 loaded in solution as without any RPA2. This indicates a complex between RPA2 and XPD is NOT forming. However, while our experiment is more sensitive to short-lived complexes than the pulldown assays of our collaborators, it still takes us about 10 seconds to move from the XPD+RPA2 channel to the ATP channel where we can observe unwinding. Thus, we are only able to observe a complex with a lifetime of at least 10 sec. It is still possible that a very short-lived (<10 sec) direct interaction between RPA2 and XPD occurs when RPA2 is in solution. Mechanistically, direct protein-protein interaction is very different from an indirect, DNA mediated interaction (such as the melting and sequestration models). But from a practical, experimental perspective, such short-lived complexes may be impossible to detect from unwinding data alone. More experiments would need to be done, perhaps with mutated XPD and RPA2.

Previous studies using related human helicases found that human RPA directly binds near the ATP hydrolysis site of WRN and BLM[85]. Mutations in the helicases were able to block RPA enhancement without destroying unwinding ability, confirming a direct interaction with RPA at that location. Mutating similar positions in *Fac*XPD and observing unwinding may deliver more insight into residues that affect unwinding and possible locations for direct RPA2 interaction.

4.5 A Two-State Model of XPD Processivity

The question that spurred the research in this thesis and is the explicit focus of this chapter is: How does RPA2 aid XPD unwinding? So far, we have sought to answer this question

by analyzing our data through the lens of testing existing models. In this next section, we will invert the process; instead of beginning with a model in mind, we will look to our data to see what about XPD unwinding changes when we add RPA2. We have already observed that increasing the concentration of RPA2 in solution increases XPD processivity, which we divide into high and low processivity types. What makes highly processive bursts different from less processive ones? Are those differences RPA2 dependent? Based on these differences and their relationship to RPA2 we seek to build our own model.

4.5.1 Average unwinding bursts show little difference between processivity types

We begin with a qualitative look at unwinding, aligning the beginning of each burst at time $t = 0$. When grouped by RPA2 concentration, some notable features emerge (**Figure 4.6 A**). Long stalling behavior around 10 bp is evident at all concentrations and in both high and low processivity traces. XPD often stalls for several, sometimes tens of seconds at this position in the hairpin. Even after stalling, XPD may continue at high processivity. Again, the increase in the number of bursts that reach high processivity with concentration is clear. Taking this analysis one step further, we average together all unwinding bursts for high and low processivity types at each RPA2 concentration (**Figure 4.6 B**). The average unwinding behavior for both types appears to be the same for all concentrations of RPA2.

4.5.2 Unwinding velocity depends on sequence, processivity; not RPA2 concentration

Taking a quantitative approach to comparing processivity types, we analyze unwinding velocity. One way to travel farther is to travel faster, so a link between speed and processivity is reasonable hypothesis[86,87]. Because XPD unwinding has been shown to be highly sequence dependent, we must consider velocity as a function of the position along the hairpin. **Figure 4.6 D** and **E** plot the average unwinding velocity as a function of hairpin position for low and high processivity types (respectively) at each RPA2 concentration. We observe that the velocity of highly processive bursts is higher than for low processivity bursts particularly near the 20 base-pair position, but within standard error below the 10 base-pair position.

There does not appear to be any link between RPA2 concentration and velocity. Indeed, **Figure 4.6 C** plots the average unwinding velocity at two particular hairpin locations (9 bp and 17

bp) for each RPA2 concentration and processivity type. At 9 bp there does not appear to be any difference regardless of concentration or processivity, a result of the frequent stalling near this location. At 17 bp, average velocity is much larger for high processivity bursts than low, but again does not vary with RPA2 concentration.

4.5.3 XPD has two inherent unwinding states, RPA2 increases the likelihood of one

We have determined that highly processive unwinding in the absence of RPA2 is indistinguishable from highly processive unwinding in the presence of RPA2 and likewise for low processivity unwinding. Instead RPA2 increases the probability of XPD unwinding with high processivity. This seems to indicate that high and low processivity unwinding are inherent states of XPD itself, and RPA2 aids unwinding by increasing the likelihood of XPD being in its more processive state.

4.5.4 Kinetic model of XPD processivity states

We have defined two processivity states for XPD, one high and one low, distinguished by the crossing of a 25 bp unwinding threshold. This system can be described with a simple three-state model of XPD transitioning between low and high processivity states and eventually dissociating from either to the "off" state (**Figure 4.7 A**). In the data we have collected, transitions between the high and low processivity states are ill defined. We can only observe the effect of these states – i.e. that XPD crosses the threshold or it does not. Thus, we adjust our model to describe when XPD first crosses the processivity threshold, a first passage time into the high processivity state.

To calculate first-passage-time kinetics, we disallow transitions out of the state of interest and the dynamics of the remaining system are calculated. Based on the model shown in **Figure 4.7 B** and given that $P_{low}(t)+P_{high}(t)+P_{off}(t)=1$, the probability of XPD being in a given state at time *t*, $P_{state}(t)$, is determined by the mass balance equations below.

$$\frac{dP_{low}(t)}{dt} = -(k_{proc}+k_{off})P_{low}(t)$$

$$\frac{dP_{high}(t)}{dt} = k_{proc}P_{low}(t)$$

$$\frac{dP_{off}(t)}{dt} = k_{off} P_{low}(t)$$

We specify that initially all molecules are in the low processivity state as they have not yet dissociated nor crossed the 25 bp threshold.

$$P_{low}(0) = 1$$

$$P_{high}(0) = 0$$

$$P_{off}(0) = 0$$

The equation for $P_{low}(t)$ is solved by an exponential.

$$\frac{dP_{low}(t)}{dt} = -(k_{off} + k_{proc}) P_{low}(t)$$

$$P_{low}(t) = P_{low}(0) e^{-(k_{off} + k_{proc})\mathrm{t}}$$

$$P_{low}(t) = e^{-(k_{off} + k_{proc})\mathrm{t}} \tag{4.1}$$

And we use the result for $P_{low}(t)$ to calculate $P_{high}(t)$.

$$\frac{dP_{high}(t)}{dt} = k_{proc} P_{low}(t) = k_{proc} e^{-(k_{off} + k_{proc})\mathrm{t}}$$

$$P_{high}(t) = k_{proc} \int_0^t e^{-(k_{off} + k_{proc})\tau} d\tau$$

$$= \frac{k_{proc}}{k_{off} + k_{proc}} \left[-e^{-(k_{off} + k_{proc})\tau} \right]_0^t$$

$$P_{high}(t) = \frac{k_{proc}}{k_{off} + k_{proc}} \left(1 - e^{-(k_{off} + k_{proc})t} \right) \tag{4.2}$$

And by the same derivation, $P_{off}(t)$ is then

$$P_{off}(t) = \frac{k_{off}}{k_{off} + k_{proc}} \left(1 - e^{-(k_{off} + k_{proc})t} \right) \tag{4.3}$$

Based on our model, k_{proc} should increase with RPA2 concentration, but we expect k_{off} to be independent of RPA2. Thus, when we fit our data to the model, we perform a global fit of all concentrations simultaneously with the same value of k_{off} and independent values of k_{proc} for each concentration. **Figure 4.7 D** plots the actual fraction of XPD molecules that have crossed

25 bp by time *t* for each RPA2 concentration. The mathematical model for $P_{high}(t)$, Equation (4.2), gives an excellent fit to the data (**Figure 4.7 D** dashed lines).

k_{proc} does indeed increase linearly with RPA2 concentration (**Figure 4.7 E**), confirming the link between RPA2 and enhanced processivity. k_{proc} has a value of $3.2 \pm 0.4 \times 10^{-3}$ s^{-1}/nM RPA2 – 2 orders of magnitude slower than the melting rate of RPA2 calculated in Chapter 3:, $k_{1,slow} = 0.11 \pm 0.02$ s^{-1}/nM RPA2. This reinforces the conclusion from Section 4.3 that melting of the hairpin is not aiding XPD unwinding. If it did, we would expect a much closer match between the rate of processivity and the rate of RPA2 binding and melting. The success of this mathematical model in fitting our data supports our two-state processivity theory. XPD has two inherent states of processivity, low and high, and it switches between them during unwinding. The presence of RPA2 increases the probability that XPD is in its high processivity state.

4.6 Future Work: Broad Possibilities for Underlying Physical States

The two-state processivity model is agnostic to the underlying physical cause of XPD's two processivity states. This section will explore some possible explanations with suggestions for future work to distinguish them.

4.6.1 Conformational change in XPD – open vs. closed

Two conformational states of XPD have been identified: "open" and "closed" referring to the position of the ARCH domain[78]. Fluorescence experiments by our collaborators in the Spies Lab have demonstrated a link between the closed state and binding to ssDNA[78]. In other helicases, notably UvrD, conformational changes of this type have been linked to forward/backward modes of unwinding[52]. It is a tempting hypothesis that these two conformations produce the high and low processivity states of XPD and that RPA2 directly interacts with XPD causing a conformational change. However, what relation these two states have to unwinding is still unknown.

For their experiments, Ghoneim *et al.* developed a version of XPD with a Cy3-labelled ARCH domain[78]. They make use of the fact that XPD contains and iron-sulfur cluster (Fe-S cluster) that is a natural fluorescence quencher and that the magnitude of quenching is related to the distance between the fluorophore and the Fe-S cluster[29]. In the open state, the Cy3-ARCH domain

is far enough away from the Fe-S cluster to see a high fluorescence signal. In the closed state, the Cy3-ARCH domain and the Fe-S cluster brought close together and fluorescence is quenched. In this way the conformational state of a single XPD can be monitored over time. However, in their experiments they could not observe unwinding only binding to ssDNA.

Our fleezers instrument provides an exciting opportunity to monitor both the conformational changes and unwinding of this labelled XPD molecule and observe any correlations between conformation state and unwinding behavior[40,52]. Performing these experiments with RPA2 as well can show if the presence of RPA2 also changes the probability of being in one conformational state over another.

4.6.2 Displaced strand binding, hindering forward motion

As a helicase unwinds it produces two strands of ssDNA. The first, the translocating strand, is the one the helicase is physically bound to and "walks" along. The other is known as the displaced strand and, in some cases, it can bind to the helicase surface. XPD has been shown to unwind a forked construct with long ssDNA overhangs for binding more efficiently than a one-sided construct where the translocating strand is the only ssDNA overhang[18]. This points to some sort of interaction between XPD and the displaced strand, but the precise effect and strength is unclear. In our experiments because both strands are under tension if XPD binds tightly to the displaced strand the full force from the optical traps will hinder its forward motion. With relatively high 12 pN force necessary to see significant unwinding from this weak helicase, the force would surely stop XPD in its tracks. This pausing effect was observed on RecQ in magnetic tweezers experiments by Harami *et al.*[60]. They were able to circumvent this problem by changing the geometry of their DNA construct. By replacing the hairpin with a gapped construct and placing all the tension on the translocating strand, binding of the displaced strand had no effect unwinding and the pausing disappeared.

If this occurs with XPD, the high processivity state may be unwinding and the low processivity state may be pausing caused by binding of the displaced strand. RPA2, as a ssDNA-binding protein, will readily bind to the displaced strand, preventing it from binding to XPD and causing an increase in the highly processive state. XPD certainly pauses often and for long periods, but these pauses are strong associated with certain positions in the hairpin sequence

with high base-pairing energies. Pausing due to the binding of the displaced strand should be randomly distributed along the sequence. However repeating the gapped construct experiments from Harami *et al.* are a straightforward way to test this theory[60].

4.7 Figures

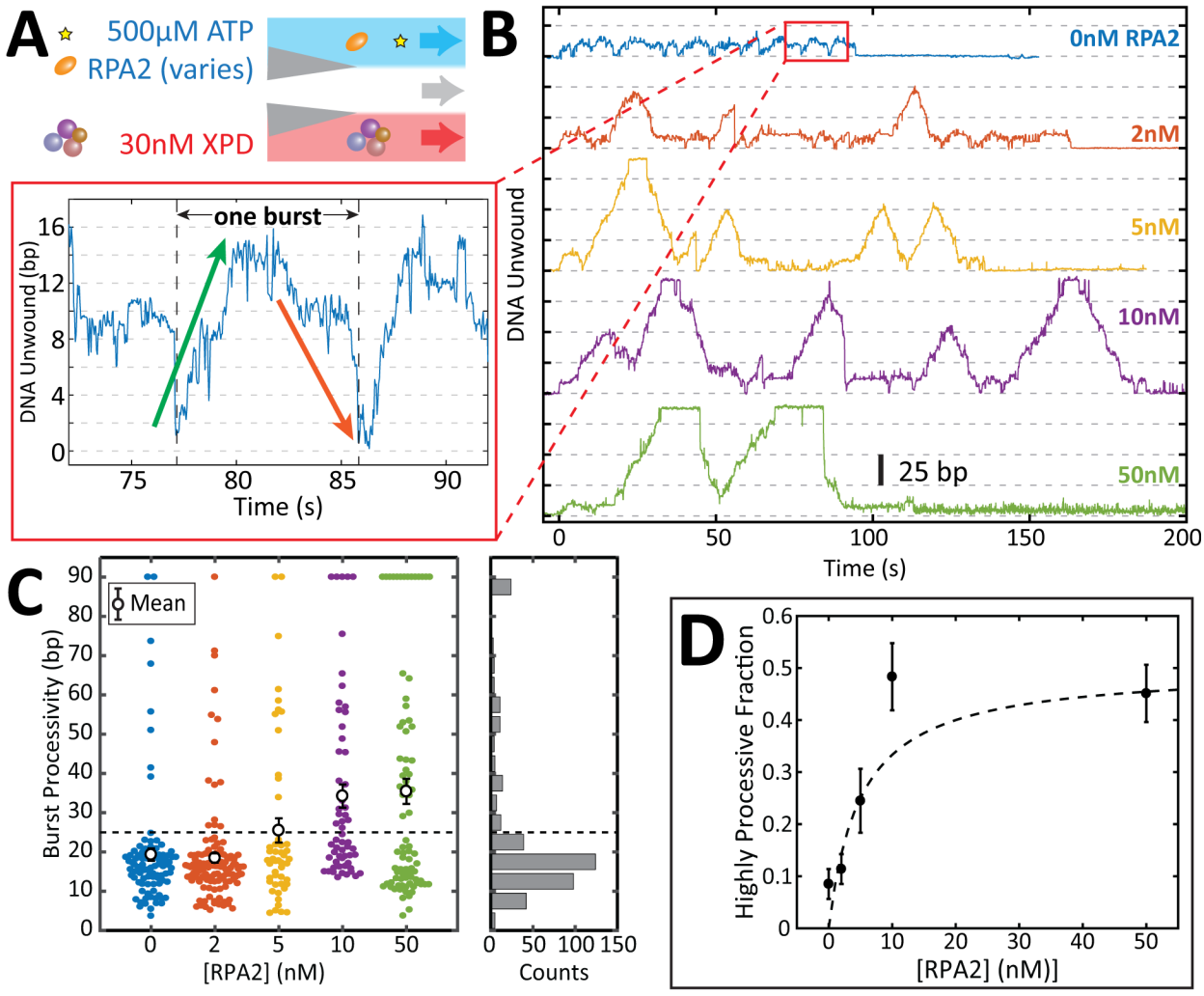

Figure 4.1 XPD processivity increases with RPA2 concentration. **A)** To determine the impact of RPA2 on XPD unwinding, varying concentrations of RPA2 were mixed with ATP in the top stream during experiments. After a single molecule of XPD is loaded on the hairpin construct, it is moved into the ATP+RPA2 stream where it begins to unwind. (For further details on this hairpin assay and interpreting unwinding data, please refer to Section 2.3.) **B)** Examples of a single molecule of XPD unwinding in the presence of varying concentrations of RPA2. XPD enters the ATP+RPA2 stream at t=0 s and begins unwinding in repetitive bursts. Inset: Each burst is composed of forward unwinding (green arrow) followed by backward stepping or sliding (orange arrow). **C)** The processivity of each burst is plotted vs. RPA2 concentration. Overall, there is a large population of bursts with processivity < 25 bp with a long tail of high processivity. This tail gets more dense at higher concentrations of RPA2. The mean processivity (white circles ± s.e.m.) increases with RPA2 concentration. **D)** The fraction of bursts with processivity > 25 bp increases with RPA2 concentration and obeys Michaelis-Menten dynamics with K_D = 5 nM (dashed line).

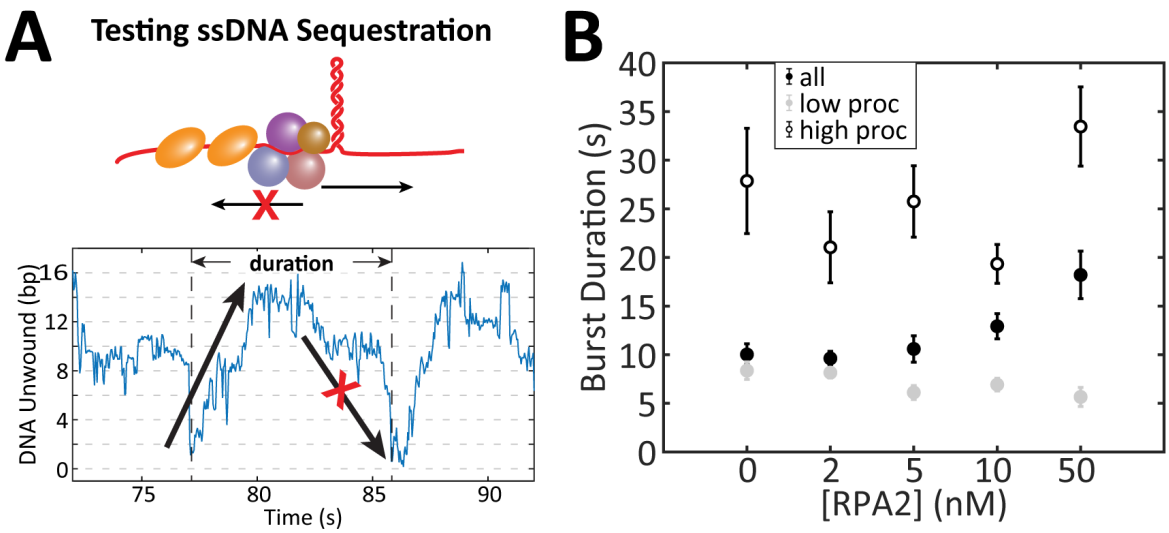

Figure 4.2 Testing ssDNA sequestration model. **A)** In the ssDNA sequestration model, RPA2 prevents XPD from back stepping and back sliding. If this were the case, since the end of each burst is set by XPD's backward motion we would expect unwinding bursts to be of longer duration as RPA2 concentration increased. **B)** However, while we find that the average burst duration increases (black circles), this is due to increasing number of highly processive bursts, which have a longer duration. The average duration of low- (gray circles) and high-processivity bursts (white circles) remains constant with RPA2 concentration. All error bars are s.e.m.

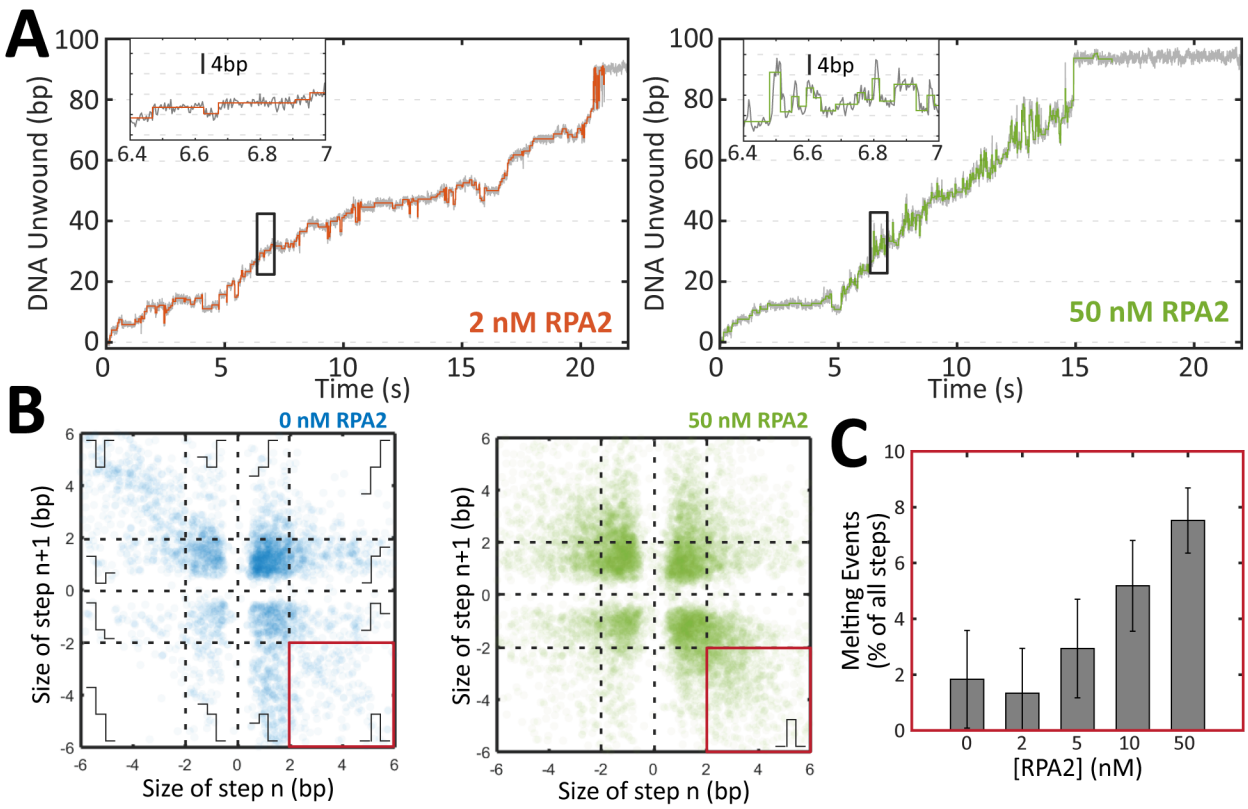

Figure 4.3 RPA2 melts the DNA hairpin during unwinding. **A)** As we increase the concentration of RPA2 in solution, we begin to see RPA2-like melting events *during* unwinding. These are short duration melting events of ~4bp. We use a step-fitting algorithm[84] to fit the unwinding data and pick out these RPA2-like events. **B)** Plotting the size of step n vs the size of step n+1 for all fitted steps, we see that in the absence of RPA2, small steps (<2 bp) make account for the majority of unwinding behavior along with some activity associated with 5 bp back-steps described in previous work[19]. For our analysis of the effect of RPA2, we focus on the bottom right corner (red outline) which corresponds to a large forward step followed by a large backward step – as we see with RPA2 melting events. We set as criteria for these events: >2 bp forward followed by <-2 bp back with <0.06 s between the two steps. **C)** While this type of event is observable even in the absence of RPA2, the probability does increase with RPA2 concentration confirming their connection to RPA2 activity.

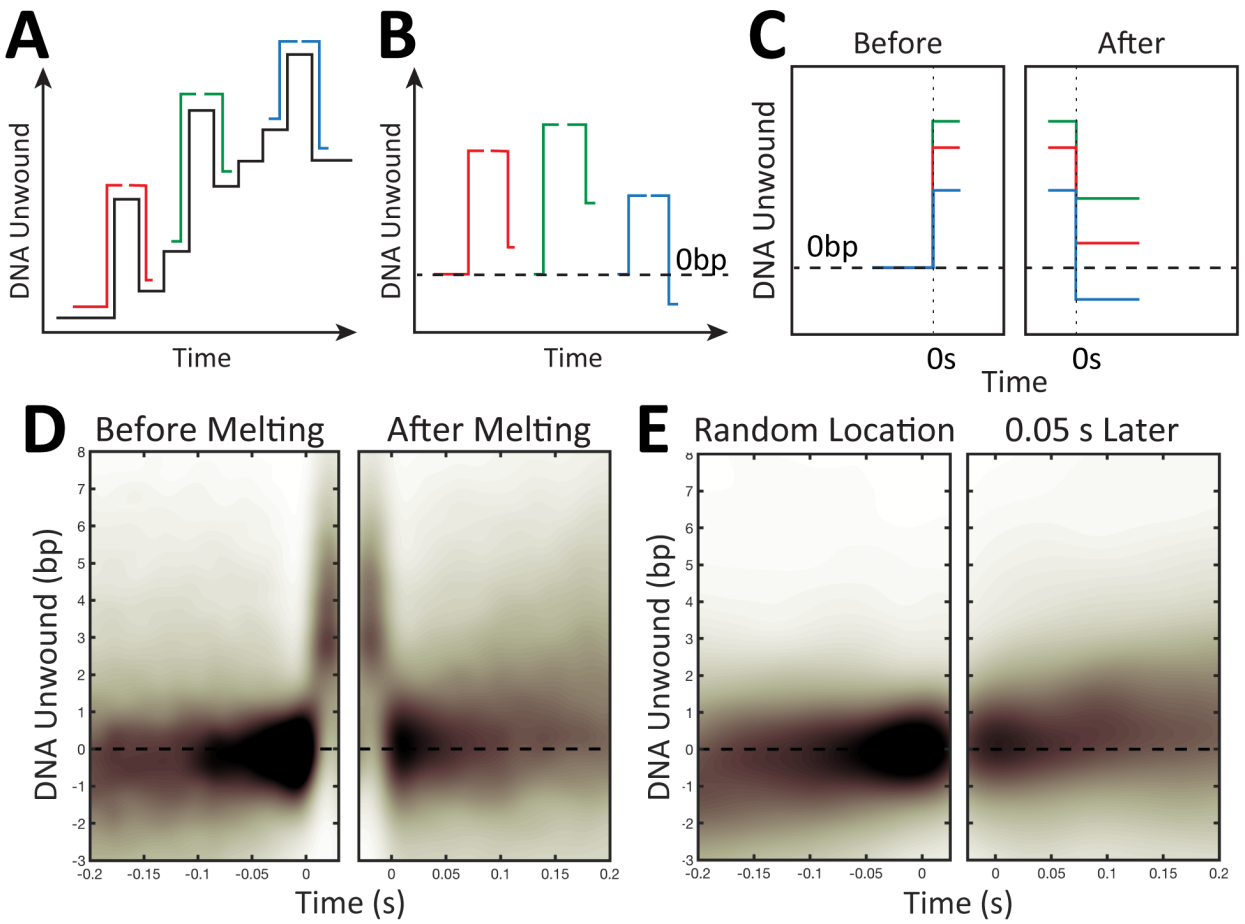

Figure 4.4 XPD makes no net progress forward after RPA2 melting events. **A)** We first identify fit steps that meet the criteria for RPA2 melting events. **B)** We then normalize each step such that it begins at 0 bp before melting. **C)** Finally, we divide the step into the melting half (before) and reannealing half (after). We independently align all the "before' steps such that melting occurs at time zero. And then all the "after" steps such that reannealing occurs at time zero. **D)** By looking at the kernel density estimation of all aligned events, we can see the effect of RPA2 on the net forward motion of XPD. The darker color represents higher probability. There is little to no net progress of XPD after melting occurs, the hairpin reanneals back to where it started. **E)** Net progress after an RPA2 melting event is indistinguishable from the net progress from a random location after 0.05 s, indicated melting events do not aid unwinding.

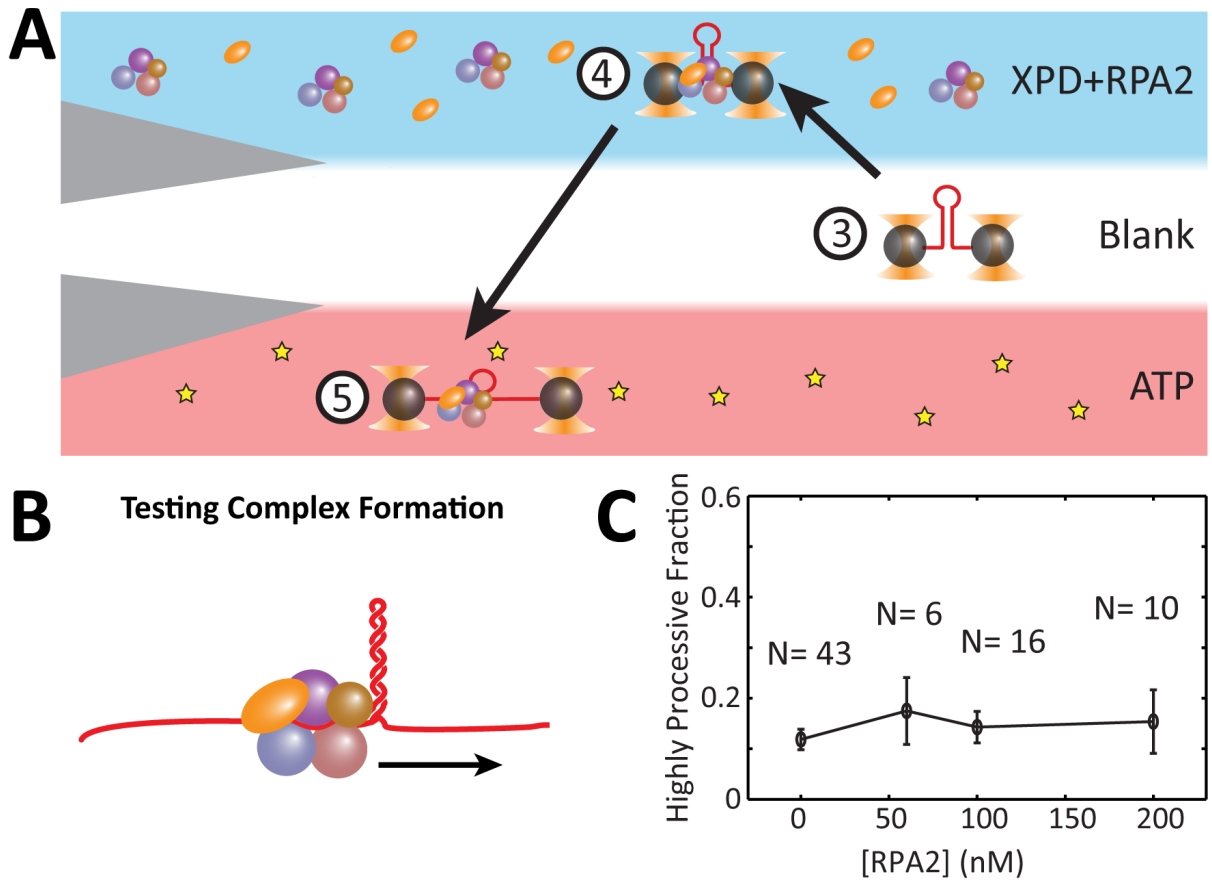

Figure 4.5 Testing the and complex formation model. **A)** To test the complex formation model, we alter our experimental set up, removing RPA2 from the ATP stream and adding it to the XPD stream. This allows us to pre-load an XPD-RPA2 complex during incubation that will begin unwinding with high processivity upon entering to the ATP stream. (Circled numbers correspond to numbered steps described in Chapter 2:.) **B)** In this model, the XPD-RPA2 complex is more processive than XPD alone. We expect an XPD-RPA2 complex will be more likely to form at high concentration of RPA2 and thus we would see more highly processive bursts at higher concentrations of RPA2. **C)** However, mean processivity does *not* change with RPA2 concentration leading us to conclude a complex was not formed. (Error bars are s.e.m. N is the number of XPD molecules tested at each concentration of RPA2.)

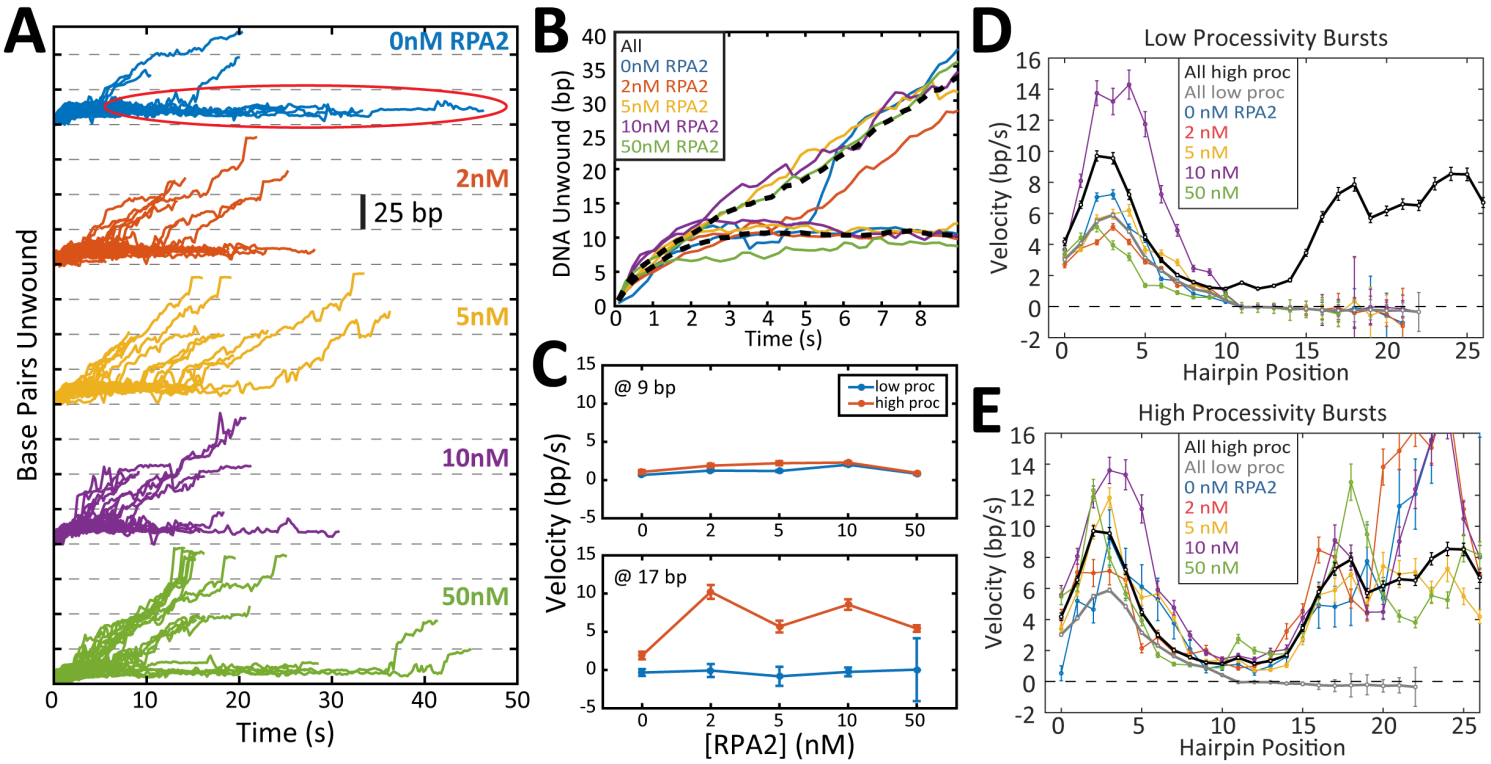

Figure 4.6 XPD unwinds in two states determined by processivity, not RPA2 concentration. **A)** The start of each burst is aligned to time $t = 0$ and bursts are grouped by RPA2 concentration. The number of bursts that reach high processivity clearly increase with RPA2 concentration. Other features remain the same at all concentrations, notably long stalling behavior near 10 bp (circled in red for 0 nM RPA2) occurs in all types of bursts. **B)** The average high and low processivity burst is plotted for each concentration (colored lines) as well as the average for all concentrations (dashed black lines). Unwinding behavior is consistent for all concentrations of RPA2, the only difference appears to be in the processivity. **C)** The average unwinding velocity at two particular hairpin locations (9 bp and 17 bp) for each RPA2 concentration and processivity type (error bars are s.e.m.). Processivity type rather than RPA2 concentration produces a difference in velocity. **D-E)** the average unwinding velocity as a function of hairpin position for low and high processivity types (respectively) at each RPA2 concentration. We observe that the velocity of highly processive bursts is higher than for low processivity bursts particularly near the 20 bp mark, but within error below the 10 bp (error bars are s.e.m.).

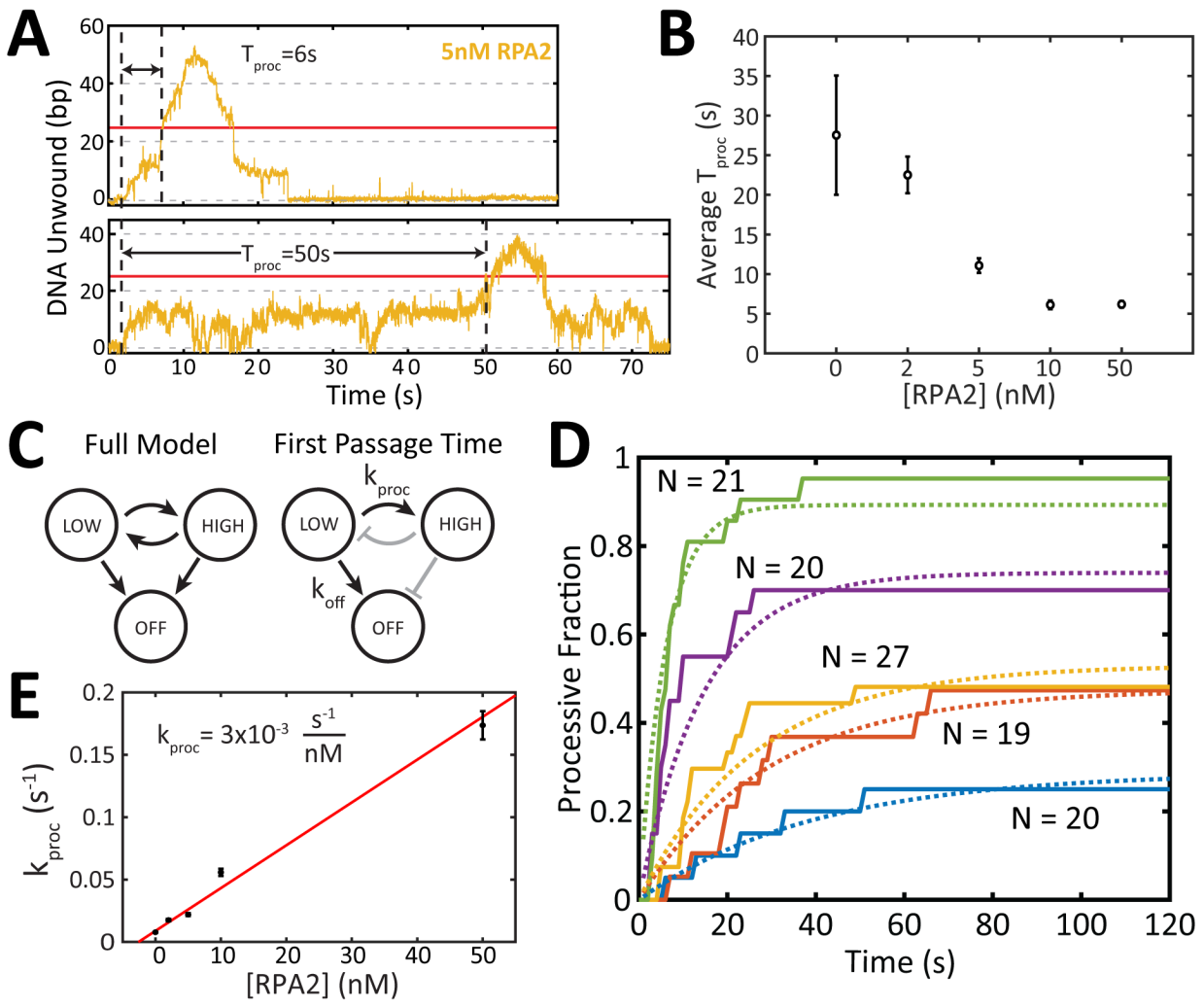

Figure 4.7 Three-state kinetic model of XPD processivity. In our model, XPD can inhabit three possible states: a low processivity state defined as unwinding less than 25 bp, a high processivity state defined as unwinding 25 bp or more, and a dissociated state where XPD becomes unbound from the DNA substrate and goes back into solution. **A)** We measure the first passage time of entering the processive state, T_{proc}, as the amount of time before XPD crosses the 25 bp threshold, if it ever does. **B)** The average value of T_{proc} decreases with RPA2 concentration (error bars are s.e.m.). **C)** In the full model, XPD switches back and forth between high and low processivity states before dissociating. To fit our data, we use the first-passage time model where XPD unwinds in a low processivity state until it dissociates or reaches high processivity. **D)** If we plot the fraction of all molecules that have reached high processivity after time t for each concentration. The curves were globally fit to Equation (4.2) each with a different value for k_{proc}, but the same value for k_{off}. **E)** The rate at which XPD enters the high processivity state, k_{proc}, increases linearly with RPA2 concentration. Error bars the 95% confidence interval from the fit in **D**.

Chapter 5: XPD Unwinding on a Uniform Hairpin Sequence

This chapter will detail the creation of a new hairpin sequence with a uniform base-pairing energy landscape and preliminary results of XPD unwinding.

5.1 Designing a new sequence

Helicases unwind dsDNA by breaking the bonds between base pairs. GC base pairs are connected by one more bond than AT base pairs and therefore require more energy to break. For this reason, helicase unwinding may depend on the DNA sequence. Previous work has shown that XPD unwinding is sequence dependent, unwinding more processively through AT-rich sequences and back-stepping more frequently in GC-rich regions[19]. Rates of XPD forward and backward stepping were related to the likelihood of base pairs thermally melting ahead of XPD defined quantitatively as P_{open}. This work showed XPD tended to stop forward progress at points in the sequence where $P_{open} \approx 0.1$. In the case of our DNA substrate, this first occurs at the 13-14 bp hairpin positions, resulting in relatively low processivity in the absence of RPA2 (mean 19 bp).

5.1.1 Definition of Popen

As defined by Johnson et al.[88], P_{open} is the probability of the DNA duplex opening thermally ahead of XPD at a given hairpin sequence position and force. If the helicase is located at hairpin position *l* (and thus *l* bp have been unwound), the probability of *m* bp melting ahead of the helicase (**Figure 5.1**) is given by the Boltzmann distribution.

$$P(l,m,F) = \frac{\exp\left(-\Delta G(l+m,F) \middle/ k_B T\right)}{\sum_{m=0}^{l_{\max}-l} \exp\left(-\Delta G(l+m,F) \middle/ k_B T\right)}$$

l_{max} is the total length of the hairpin in bp and $\Delta G(n,F)$ is the change in the free energy, relative to the fully closed state, of the DNA hairpin when held at force *F* with $n = l + m$ total base pairs open. The free energy will have contributions from the loss of base-pairing interactions, ΔG_{bp}, and from the added *2n* nucleotides of ssDNA being stretched under applied force, ΔG_{ssDNA}.

$$\Delta G(n,F) = \Delta G_{bp} + \Delta G_{ssDNA}$$
$$= \sum_{i=1}^{n} \Delta G_{bp}(i) - \int_{0}^{F} x_{ssDNA}(2n,F')dF'$$

where $\Delta G_{bp}(i)$ is free energy change due to the loss of base-pairing at the *i*th hairpin position and $x_{ssDNA}(2n,F)$ is the length of *2n* nucleotides of ssDNA at force *F*. $x_{ssDNA}(2n,F)$ is determined using the Extensible Worm-like Chain model for polymers[70]. $\Delta G_{bp}(i)$ are the salt- and sequence-dependent base-pairing free energies determined by Huguet et al.[89]. We then generalize this expression to the probability of *at least* one base pair opening thermally ahead of position *l* by summing the probabilities of all *m* ≠ 0.

$$P_{open}(l,F) = \sum_{m=1}^{l_{\max}-l} P(l,m,F) = \frac{\sum_{m=1}^{l_{\max}-l} \exp\left(-\Delta G(l+m,F) \Big/ k_B T\right)}{\sum_{m=0}^{l_{\max}-l} \exp\left(-\Delta G(l+m,F) \Big/ k_B T\right)} \quad (5.1)$$

The effect of RPA2 is difficult to decouple from the effect of sequence

It is clear from data presented in this thesis and previous work[19] that the region around the 13 bp position in this sequence presents a significant impediment to XPD's unwinding progress due to low values of P_{open}. It is also clear that RPA2 can help XPD overcome this impediment based on the significant increases in processivity observed in the presence of RPA2. However, even during processive unwinding, XPD may pause in this sequence region for long periods of time making it difficult to determine precisely how RPA2 interacts with XPD in this region. Changes in rates of forward or backward stepping and velocity are difficult to measure in this region because they are so low to begin with.

XPD's response to sequence also means that data at one sequence position cannot be directly compared to another. All analysis must take into account the effect of sequence and only compare data with similar values of P_{open}. This has the effect of decreasing sample sizes as we bin data by P_{open} or single base pair sequence position and also removes time dependence from that analysis.

Uniform P_{open} decouples unwinding behavior from hairpin position

A sequence with uniform values of P_{open} would remove sequence-dependent barriers to unwinding and decouple unwinding behavior from hairpin position. Each base pair would have a roughly equal barrier to opening and any differences seen from one step to the next could be attributed to a change in XPD's unwinding state and/or an effect of RPA2. Additionally, data at all positions along the hairpin could be combined and compared. This would increase sample sizes and making subtle changes in measures such as velocity easier to identify.

5.1.2 Generating the new sequence

We developed a simple algorithm for designing sequences with a flat unzipping energy landscape:

1. Pick a desired range of P_{open}.
2. Choose an initial sequence or generate a random sequence S
3. Calculate P_{open} for the sequence S.
4. Find positions in S with P_{open} outside of desired range. Ignore positions near the end loop of the hairpin (above the 80 bp position).
5. Change bases at positions found in (4) to generate a new sequence S*
    a. If base is A or T change to G or C with 50/50 probability
    b. If base is G or C change to A or T with 50/50 probability
6. Repeat 3-5 until all hairpin positions in S* fall within the desired range for P_{open}.

This is a brute force algorithm that would likely be time and computationally expensive for very long sequences. However, for our 89-bp hairpin sequence, the algorithm converges on a satisfactory sequence quickly (<1 min, exact value depends heavily on initial sequence). DNA is a fairly stiff molecule and the harsh bend at the end of the DNA hairpin is energetically costly to maintain. The combined effect of this bend energy and applied force cause P_{open} to rise sharply within ~10 bp of the end loop, regardless of sequence. For this reason, our analysis does not include unwinding data near the end of the hairpin, nor do we impose strict limits on P_{open} beyond the 80 bp position for the new sequence.

Check stable conformations and likelihood of secondary structures

A drawback to having a constant value of P_{open} is that it necessarily restricts the combinations of bases available for use in the final sequence. Base-pairing energies depend on nearest-neighbor interactions, thus P_{open} is calculated using the energies of pairs of base pairs – 10 combinations in all[89]. For example, $\frac{CC}{GG}$ and $\frac{CG}{GC}$ have high nearest-neighbor base-pairing energies and thus a low value of P_{open} (these energies are unlikely to be overcome due to thermal fluctuations). On the other hand, $\frac{AA}{TT}$ and $\frac{AT}{TA}$ have lower high nearest-neighbor base-pairing energies and thus give a high value of P_{open}. Tightly controlling P_{open} leads to a fairly repetitive sequence which can cause undesirable secondary structures to form in DNA. Generated sequences must be checked in a DNA folding program such as Mfold[90] to ensure that the lowest energy (and thus most stable) conformation is the desired hairpin structure. Other structures may inhibit ligation when making the final DNA construct or may even form while trapping, presenting an unpredictable substrate for unwinding experiments.

5.1.3 Force response of resulting sequence

The new sequence generated has P_{open} between 0.35 and 0.6 at every position (**Figure 5.2 A**) with 12 pN tension applied. This is a much tighter range than the original sequence for which P_{open} ranges between 0.1 and 0.85 at the same force. This more consistent thermal energy results in a flat response to applied force. As the DNA is stretched, the new hairpin opens at a near constant force of ~14 pN (**Figure 5.2 C**), compared to the original sequence where the force response varies as the hairpin opens from ~14.5-17.5 pN along its length (**Figure 5.2 B**).

5.2 Preliminary unwinding results

Preliminary data were taken of XPD unwinding this new sequence. Experiments were performed in the same manner as described in Chapters 2 and 4, the only difference being the DNA substrate. Some representative traces are plotted in **Figure 5.3 A** from 8-12 pN applied force. It is immediately clear that XPD is more processive on the new sequence. While on the previous sequence XPD rarely unwound more than 25 bp at 12 pN force, on the new sequence it regularly unwinds >40 bp at the same force. To the trained eye, unwinding bursts also appear

qualitatively different with a more uniform velocity across the entire burst and a lack of obvious stall points in the sequence. These differences will be quantified in the sections to follow.

5.2.1 More uniform unwinding velocity

Performing the same analysis as described in Section 4.5, we select the start and end times of each burst and plot the unwinding segments, grouped by applied force, for analysis (**Figure 5.3 B**). There is not an obvious distinction between different processivity states like we observed previously (**Figure 4.6**). Indeed, unwinding trajectories appear to be very consistent at each force. A possible exception is data taken at 10 pN where there is greater diversity of unwinding trajectories. More data is needed at every force to determine if this is unique effect of that particular force, consistent across all forces, or an anomaly caused by a relatively small data set. But even so, the diversity of unwinding trajectories appears to be continuous rather than split between two apparent states.

Figure 5.3 D shows the mean unwinding velocity of XPD at each base along the hairpin. The velocity is nearly constant and this is reflected in the average unwinding traces as well – each is nearly a straight line (**Figure 5.3 C**). Notably, average velocity increases between hairpin positions 20 and 30, particularly at lower forces (for higher forces, it appears to be within the variance established by positions 1-20). Again, more data is needed to determine if this is merely caused by having a smaller number of data points (reflected in the larger standard error bars) or if this is truly a change in unwinding behavior. If the latter is true, it cannot be explained by a difference in base-pairing energies in that region – this sequence was designed to remove that as factor with near-constant base-pairing energies at every hairpin position. We would need investigate alternative explanations for this unexpected result (elaborated in Section 5.3).

5.2.2 Processivity distribution reflects more random process

We find that burst processivity increases as the force on the hairpin increases, as expected given the increase in P_{open} with force. More interesting is the probability distribution of XPD processivity. Recall with our previous sequence, the distribution of burst processivity was peaked near 20 bp with a heavy, sub-exponential tail. The tail got heavier as RPA2 was added and overall

processivity increased, but the low processivity peak remained (**Figure 4.1**). We theorized that this was due to sequence effects that stalled XPD unwinding near 15 bp.

In an idealized case where XPD processivity were a purely random process and determined only by constant forward and backward stepping rates of the protein, the number of base pairs open would be Poisson-distributed. And because each burst is independent from the others, the likelihood of achieving a particular processivity in one burst would be Poisson-distributed as well. By removing sequence variations, we expected XPD processivity on the new sequence to reflect this purely random case. And indeed, the distribution of burst processivity is much closer to a Poisson distribution than the previous sequence results (**Figure 5.4 B-F**).

However, while this distribution is *closer* to a Poisson process, there is an unexpected deviation from this prediction near 30 bp. Again, there is nothing energetically special about this region, yet there is a minimum in the processivity distribution at this location. This means that if XPD reaches this point in the hairpin it never turns around there, it always unwinds through. This surprising result will shape future experiments moving forward (see Section 5.3).

5.3 Future work

Future work will primarily set out to test the two-state model of XPD processivity laid out in Chapter 4: against the new sequence data. If the model is supported, we hope to find out what precisely the underlying unwinding states are. We will also observe the effect of RPA2 on unwinding and again test the consistency of our processivity model.

5.3.1 Processivity states and the gap at 30 bp

Histograms of XPD processivity on the new sequence show a local minimum near 30 bp across all forces (**Figure 5.4**). This indicates that, rather than persistently stalling as before, XPD never stops at that position. This may be an effect of the relatively low number of bursts; more data will determine if this is the case. However, if this a real feature of XPD unwinding, the section speculates on plausible explanations with interesting implications for XPD structure and function.

In Section 4.5, we theorized the existence of two underlying processivity states in XPD. Are the two peaks in the processivity distribution simply these two processivity states manifested on the new sequence? That is certainly a likely possibility, though surprising. While the

processivity distribution on the previous sequence had a heavy tail, it was still continuous. Processive runs could be anywhere from 26 to 89 bp with no preferred stopping points between the two extremes. Why, then is this peak quite tight around 40 bp and not spanning more of the hairpin?

It could be that XPD preferentially unwinds in 20 bp "spurts" and thus we would see peaks in the distribution at each multiple of 20. A small peak at 60 bp in **Figure 5.4 B** and **E** gives some support to this idea, but why 20 bp? The scale is reasonable from a physiological standpoint as XPD's role in nucleotide excision repair requires it to unwind ~30 bp of dsDNA. But what molecular mechanism could accomplish this?

Non-specific ssDNA binding effect

As a helicase such as XPD unwinds, it binds to and moves along one strand of DNA known as the "translocating strand." The opposite strand is known as the "displaced strand" as it is separated/displaced from the translocating strand. In some cases, it is known or theorized that the displaced strand interacts with the helicase at secondary binding sites on its surface[60,91,92]. These interactions with the displaced strand have been suggested to aiding unwinding in some helicases[91,92] and hinder it in others[60].

In the case of XPD, previous work has shown the helicase prefers unwinding substrates with an ssDNA overhang on the displaced strand indicating a possible interaction[18]. Our hairpin construct in its fully closed state lacks this overhang, but as it is unwound the length of the displaced strand grows. It is possible that this 20 bp processivity is the result of the increased length of the displaced strand allowing for increased interaction with XPD. Whether this interaction aids unwinding or hinders it remains in question.

Sequence-specific effect

Frequent stopping at particular locations could also be due to sequence-specific interactions between XPD and DNA. We have already discussed XPD unwinding as being "sequence-dependent" due to its sensitivity to variation in the free energy of base pairs due to sequence. "Sequence-specific" is much more narrow term referring to interactions where a protein can recognize and preferentially interact with one particular DNA sequence. Restriction enzymes are a famous example, binding to and detecting short sequences (4-8 bp) and then

cleaving the DNA at that location[93]. EcoRI, a restriction enzyme, will cleave dsDNA at the sequence GAATTC and nowhere else[94].

It could be that we have accidentally placed a sequence that XPD recognizes into our hairpin. Physiologically, for XPD's role in nucleotide excision repair, it must readily unwind many different sequences through the entire genome. Interacting specifically with any one sequence would seem more of a hindrance than a help in this role. Still, it is certainly possible. A straightforward way to check for sequence specificity is to design another new hairpin sequence with flat P_{open} and observe XPD unwinding on it. This could be a new sequence entirely or, perhaps more instructively, a rearrangement of the current sequence such that the sequence that was at the 20 bp position is now at the 30 bp position. If the peaks in our processivity distribution move to match change locations, we would know that the local sequence is important to this behavior. If the peaks persist in their old locations, then this behavior is not related to sequence and we must look elsewhere.

5.3.2 Multiple Velocity States?

Taking our unwinding trajectories from **Figure 5.3 B** and plotting them on the same *y* axis, we begin to notice trend in the unwinding velocity of XPD (**Figure 5.5 A**). At 9 pN applied force, the velocity is consistently lower than at 11 and 12 pN. While at 10 pN, velocity seems to span the whole range. We calculated the average unwinding velocity for each burst as the maximum processivity of the burst over the total time spent unwinding to that point and plotted a histogram at each force (**Figure 5.5 B**). In this preliminary data set, unwinding at 12 pN applied force appears to be exclusively in a high velocity regime, ~6 bp/s, whereas at 9 pN the majority of unwinding seems to be ~3 bp/s. Between 10 and 11 pN force, average burst processivity spans the full range. Does applied force increase XPD velocity? is increased velocity linked to higher processivity? Further data and analysis should look into these questions, especially given the surprising velocity increase beyond 20 bp unwound.

5.3.3 Detecting State Changes During Unwinding

In the previous chapter, we presented a two-state model of XPD unwinding and speculated about what the states may be. The ultimate goal of creating this new sequence was

to produce a homogenous substrate to detect slight changes in unwinding behavior above the noise produced by a XPD's sensitivity to sequence. Thus far we have presented data on average XPD velocity at a particular hairpin position and over a full winding burst. But because XPD unwinds in single-base pair steps, velocity is an ambiguous measurement for a single molecule at small timescales. Near the timescale of a single step we will essentially see either 0 or near infinite +/- instantaneous velocity.

The best way to analyze this data is to treat XPD unwinding as discrete steps and track the dwell-times between steps over the course of unwinding. Short dwell-times correspond to high velocities and higher stepping rates (and long dwells to low). If XPD stepping can be reliably tracked, rates of forward stepping, back-stepping, etc. can be efficiently calculated and provide more insight into the effect of RPA2 on unwinding kinetics.

That is a big if though. Commonly used step-fitting algorithms work well for data with longer dwell times (e.g. t-test) or a few consistent "state" locations (e.g. Markov modelling). But XPD unwinding data can have very short dwell times between steps, sometimes just a few data points, making it ill-suited for t-tests. Markov models would require up to 90 states, one for each hairpin location plus 0 bp, to characterize unwinding the full hairpin – a huge computational expense.

Step fitting via chi-squared minimization is a good option and has been used by our group for XPD unwinding in the past[19,84]. However, it works much more reliably on short stretches of data with not too many steps. That is excellent for short, low processivity traces, but longer traces or traces with very high processivity are inconsistent; over-fit in some places and under-fit in others. We did use a chi-squared minimization algorithm to find RPA2-like melting events during unwinding in Section 4.4, but the algorithm did not find every such step nor was it consistent enough for us to confidently use the results for kinetic analysis.

Future analysis should look to improve step fitting for this data set and allow access to the wealth of information hidden in the stepping dynamics. This may require updating the existing chi-squared minimization algorithm or workflow for longer, more varied stepping data. Or possibly moving to Markov models despite the computational expense.

5.3.4 Effect of RPA2

What happens when XPD unwinds this new sequence in the presence of RPA2? We expect the processivity to increase as it did on the old sequence, but this will need to be confirmed with experiments. If processivity is increased, will it shift the entire distribution to the right or will it lengthen the distribution tail as it did on the old sequence? An especially interesting question will regard the two processivity peaks we already see in the data. Are these peaks maintained in the presence of RPA2? Does RPA2 shift the likelihood of one over the other?

By decoupling the effect of sequence, we also hope to enhance the effect of RPA2 making clearer any subtle changes in unwinding behavior and velocity. Reliable step-fitting would aid these investigations greatly. Does the rate of forward-stepping increase? Does back-stepping decrease? However, this may prove quite challenging because RPA2 is able to melt the hairpin itself. Care will have to be taken to distinguish between helicase steps and melting due to RPA2 in the final analysis.

This sequence also provides a better opportunity to investigate the effect of force on XPD-RPA2 cooperation. On the original sequence, high forces were needed to help XPD overcome sequence “roadblocks” and observe significant unwinding to measure and analyze. The new sequence does not have any roadblocks and so XPD is able to unwind significant amounts of dsDNA (~20+ bp) at a relative wide range of forces (9-13 pN). Thus, we could investigate if higher force enhances the effect of RPA2 on XPD unwinding, diminishes it, or has no effect giving added insights to possible interactions between the two proteins.

5.4 Figures

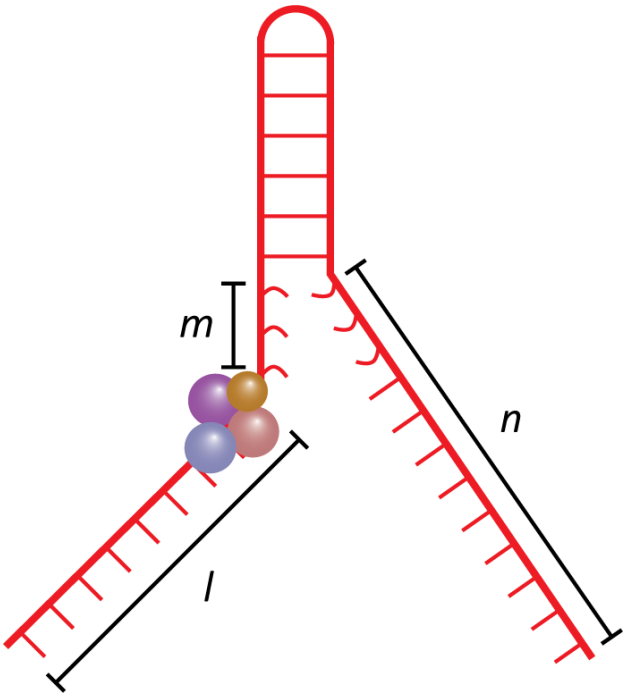

Figure 5.1 Schematic of P_{open} variables. The helicase has unwound *l* bp and is located at hairpin position *l*. *m* bp have thermally melted ahead of the helicase for a total of *n* bp of the hairpin open. In the cartoon above *l = 9* bp, *m = 3* bp, and *n = m + l = 12* bp.

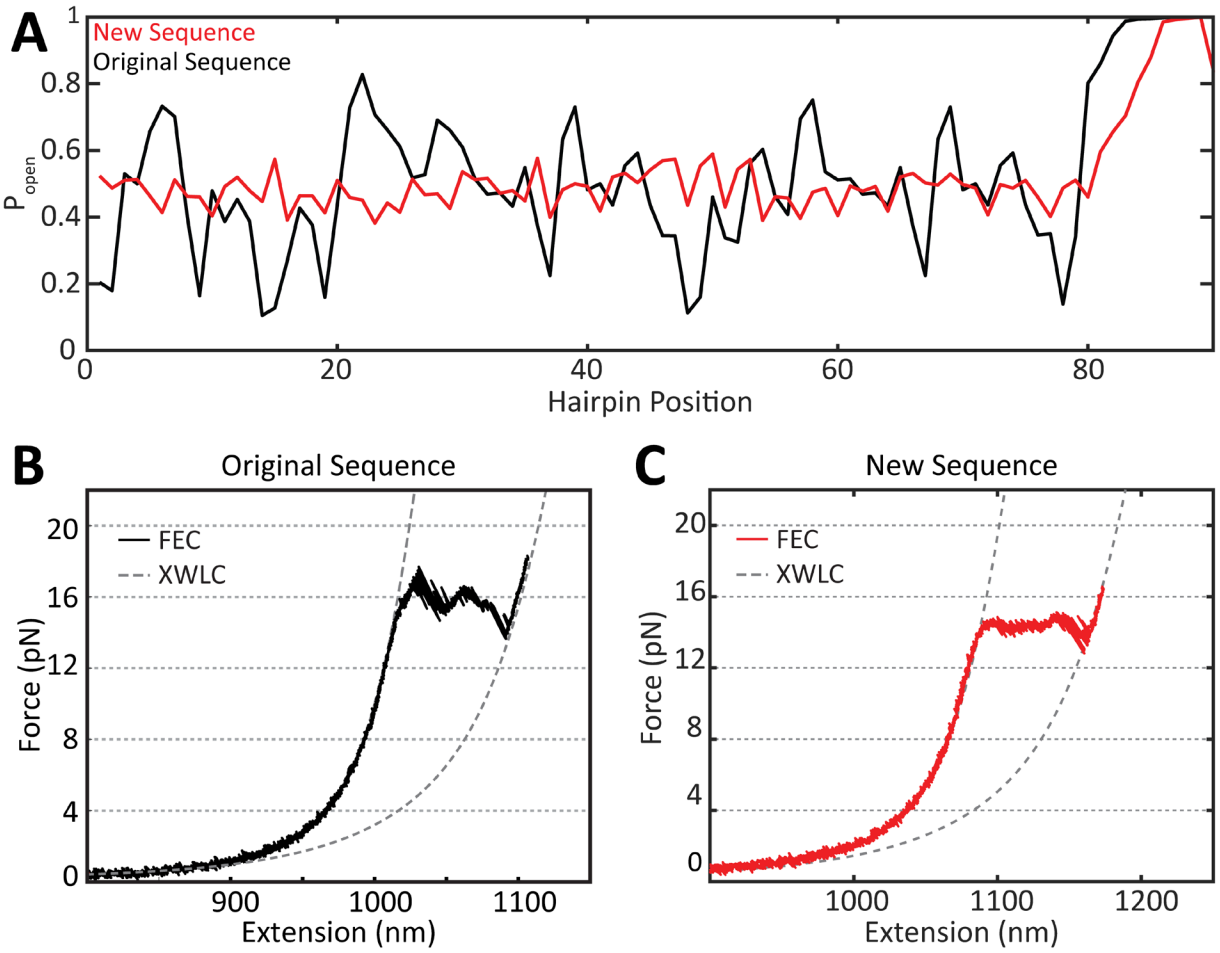

Figure 5.2 Characterizing the new hairpin sequence. **A)** When held at 12 pN tension, P_{open} of the new sequence (red) remains between 0.35 and 0.6 for the first 80 hairpin positions. This is much more consistent than the original sequence for which P_{open} ranges between 0.1 and 0.85 at 12 pN tension (black). **B)** A representative force-extension curve for the original construct. The force required to open each base of the hairpin varies with the sequence. **C)** A representative force-extension curve for the new sequence. The force remains constant as the hairpin opens due to the consistent value of P_{open}.

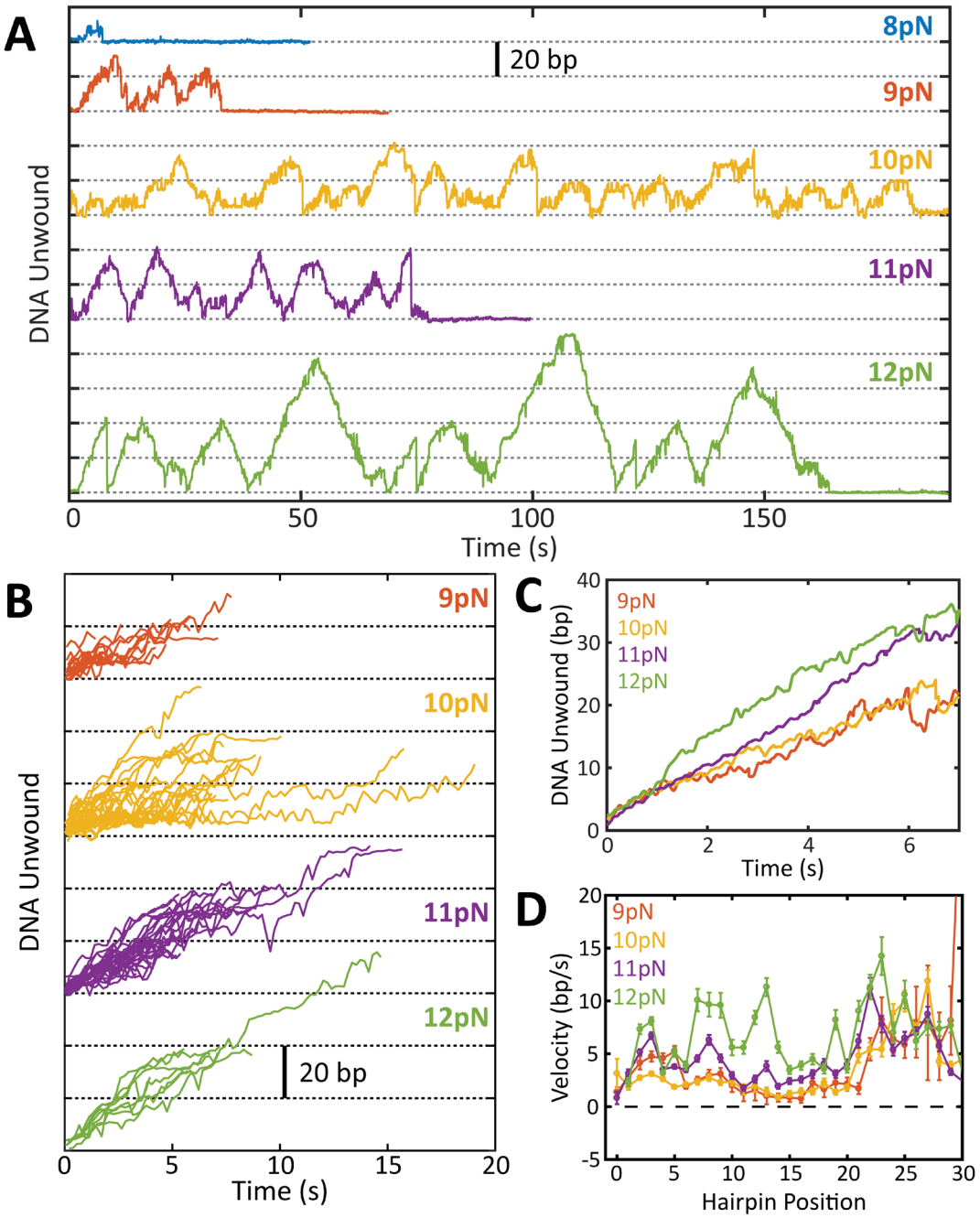

Figure 5.3 Initial data for XPD unwinding the new sequence. **A)** Example unwinding traces at varying applied force. XPD again unwinds in bursts of varying processivity. **B)** The unwinding segment of each burst aligned to begin at $t = 0$ and grouped by force. **C)** Average unwinding burst at each applied force. There does not seem to be much change in behavior when force changes nor are there obvious stall points. **D)** Average unwinding velocity is constant at each hairpin position and seemingly with applied force (error bars are s.e.m.).

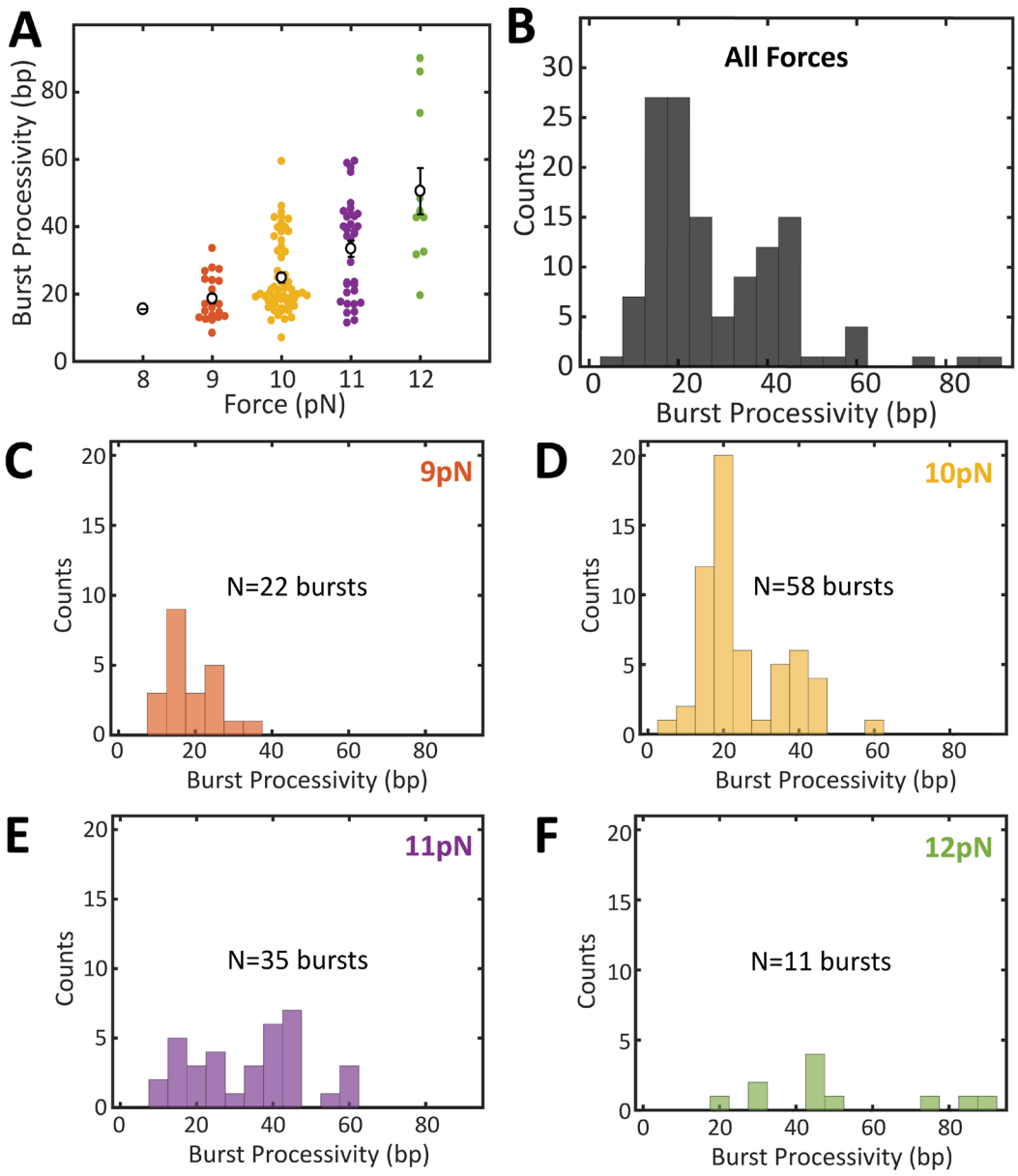

Figure 5.4 A) Processivity of each unwinding burst (colored circles) and the mean ± s.e.m. (open black circles) at each applied force. The average processivity increases with force. Histograms of XPD processivity for **B)** all forces and **C-F)** at each force. There is no clear divide between high and low processivity activity as there was with the original sequence. However, there does seem to be a persistent gap in the histogram around 30 bp. Whether this is a real feature or an effect of low numbers remains to be determined.

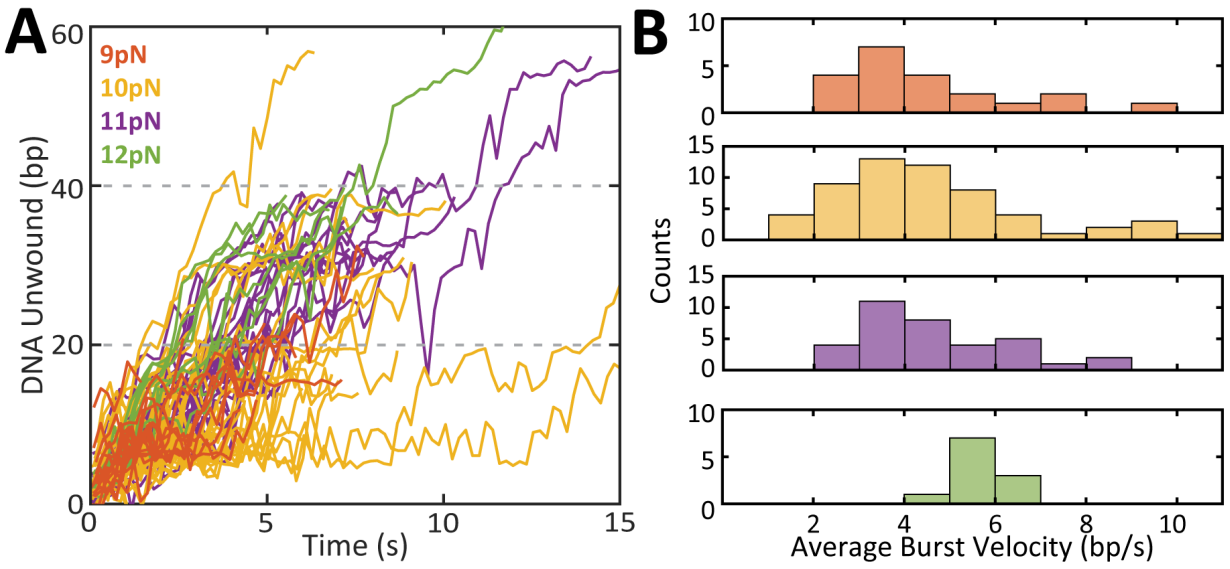

Figure 5.5 Possible existence of velocity states? **A)** All unwinding trajectories from **Figure 5.3 B** are plotted on the same axes. At 9 pN applied force (orange) the velocity is consistently lower than at 11 and 12 pN (purple and green). While at 10 pN (yellow) velocity seems to span the whole range. **B)** Histograms of average unwinding velocity during a burst are plotted for each force. Average burst velocity appears to increase with applied force.

Appendix A: Mathematical Derivations

Derivations of mathematical results used in this thesis.

A.1 Power Spectrum of Trapped Bead

Used to calibrate trap stiffness, this section will derive the result of Neuman and Block in their reviews of optical trapping[39,95], that the power spectrum of an optically trapped bead undergoing Brownian motion is

$$S_{xx}(f) = \frac{k_B T}{\pi^2 \beta (f_0^2 + f^2)} \tag{A.1}$$

We start by describing the forces on the trapped object. The total force on the object will equal the force of the trap plus a drag force plus a random force due to Brownian motion. This is Langevin's Equation plus a Hooke's law term for the force of the trap.

$$m\ddot{x}(t) = -\kappa x(t) - \beta \dot{x}(t) + \xi(t) \tag{A.2}$$

$\mathrm{x(t)}$ is the 1D bead position as a function of time.

$\dot{x}(t) = \frac{d}{dt}x(t)$ is the velocity.

κ is the trap stiffness.

β is Stokes' drag coefficient. For a spherical bead $\beta = 6\pi\eta r$.

$\xi(t)$ is a random force representing thermal noise/Brownian motion of the particle.

We assume $\langle \xi(t) \rangle = 0$, the average thermal force on the object over time is zero.

We also assume the power spectrum $S_{\xi\xi}(f) = \left|\hat{\xi}(f)\right|^2 = 4\beta k_B T$ (white noise).

The Reynolds number for an object in a fluid is the ratio of inertial forces to viscous (drag) forces. For a particle in a fluid, the Reynolds number of the particle is given by the equation

$$Re_P = \frac{\rho v r}{\eta}$$

where v and r are particle velocity and radius, and ρ and η are the density and viscosity of the fluid. For a 1μm object in water that travels ~10μm/s, $Re_P \sim 10^{-5}$. At this extremely low Reynolds number viscous forces greatly dominate inertial forces and $\beta\dot{x}(t) \gg m\ddot{x}(t)$. This allows us to ignore inertial force in our calculations and rewrite Equation (A.2)

$$m\ddot{x}(t) + \beta\dot{x}(t) = -\kappa x(t) + \xi(t)$$

$$\beta\dot{x}(t) = -\kappa x(t) + \xi(t) \tag{A.3}$$

The power spectrum of a function is defined $S_{xx}(f) = |\hat{x}(f)|^2$, where $\hat{x}(f)$ is the Fourier transform of $x(t)$. We define the Fourier transform of a function:

$$\mathcal{F}(x(t)) = \hat{x}(f) = \int_{-\infty}^{\infty} dt\, e^{-2\pi i f t} x(t)$$

Taking the Fourier transform of both sides of our force equation (A.3) we find

$$\int_{-\infty}^{\infty} dt\, e^{-2\pi i f t} \beta\dot{x}(t) = \int_{-\infty}^{\infty} dt\, e^{-2\pi i f t}(-\kappa x(t) + \xi(t))$$

$$\beta \int_{-\infty}^{\infty} dt\, e^{-2\pi i f t} \dot{x}(t) = -\kappa \int_{-\infty}^{\infty} dt\, e^{-2\pi i f t} x(t) + \int_{-\infty}^{\infty} dt\, e^{-2\pi i f t} \xi(t)$$

$$\beta\mathcal{F}(\dot{x}(t)) = -\kappa\hat{x}(f) + \hat{\xi}(f) \tag{A.4}$$

Using the derivative property of Fourier transforms

$$\mathcal{F}\big(\dot{x}(t)\big) = -2\pi i f \mathcal{F}\big(x(t)\big) = -2\pi i f \hat{x}(f)$$

Substituting the above into Equation (A.4), we get

$$-2\pi i f \beta \hat{x}(f) = -\kappa\hat{x}(f) + \hat{\xi}(f)$$

And solving for $\hat{x}(f)$

$$\hat{x}(f)\left[\frac{\kappa}{2\pi\beta} - if\right] = \frac{\hat{\xi}(f)}{2\pi\beta}$$

$$\hat{x}(f) = \frac{\hat{\xi}(f)}{2\pi\beta(f_0 - if)} \tag{A.5}$$

where $f_0 = \frac{\kappa}{2\pi\beta}$. Finally, the power spectrum for the bead's motion is

$$S_{xx}(f) = |\hat{x}(f)|^2 = \left|\frac{\hat{\xi}(f)}{2\pi\beta(f_0 - if)}\right|^2$$

$$= \frac{\left|\hat{\xi}(f)\right|^2}{|2\pi\beta(f_0 - if)|^2}$$

$$= \frac{4\beta k_B T}{4\pi^2\beta^2|f_0 - if|^2}$$

$$= \frac{4\beta k_B T}{4\pi^2\beta^2(f_0 - if)(f_0 + if)}$$

$$S_{xx}(f) = \frac{k_B T}{\pi^2 \beta (f_0^2 + f^2)}$$

Which matches the results of Neuman and Block in Equation (A.1). **Figure A.1** on the left shows the position data in the x and y directions collected by the QPD. This raw data has units of volts which is proportional to the distance measured in nm (expanded on in Sections 2.1 and A.2).

$$x^{(nm)}(t) = \alpha x^{(QPD)}(t)$$

The power spectra plotted on the right of **Figure A.1** is that of the QPD data.

$$S_{xx}^{(QPD)}(f) = \left|\hat{x}^{(QPD)}(f)\right|^2 = \left|\frac{\hat{x}^{(nm)}(f)}{\alpha}\right|^2 = \frac{S_{xx}^{(nm)}(f)}{\alpha^2} = \frac{k_B T}{\alpha^2 \pi^2 \beta (f_0^2 + f^2)} \quad \text{(A.6)}$$

Experiments are done at room temperature, so $k_B T$ is known. As is β because $\beta = 6\pi\eta r$. η is the viscosity of our experiment buffer which is approximately equal to that of water. r is the radius of the trapped bead which is provided by the manufacturer. Thus fitting the power spectrum of the QPD data to Equation (A.6) yields two parameters – the stiffness of the trap, κ, and the conversion factor from QPD voltage to nm distance, α. A third value we call the offset is also calculated as the mean of the raw QPD position signal. The use of this offset will be discussed in the next section.

A.2 DNA Tether Extension

This section will calculate the tension and end-to-end extension of a DNA tether, including offset removal, from data collected on the Chemla Lab optical tweezers instruments. We begin with a cartoon diagram of the dual trap + DNA tether to help us visualize the system. From the picture in **Figure A.2** it is fairly straightforward to write down an equation for the end-to-end DNA extension in terms of the parameters included.

$$\text{DNA extension} = \sqrt{(ext_x)^2 + (ext_y)^2} - r_A - r_B$$
$$ext_x = \Delta X_{traps} - (AX - BX)$$
$$ext_y = \Delta Y_{traps} - (AY - BY)$$

Where $\Delta X_{traps}, \Delta Y_{traps}$ are the x and y distances between the traps, AX (BX) is the x distance of the center of Bead A (B) from the center of Trap A (B), and r is the bead radius (either a known value given by the manufacturer). We will do these calculations for the set up shown in **Figure A.2** – two traps aligned to pull the DNA tether in the x direction with trap/bead A on the left and

B on the right. Be careful with signs! The bead position in x will be >0 if it is to the right of the trap center, thus in the figure $AX > 0$ and $BX < 0$. If the beads were reversed and the bead displacements had opposite signs ($AX < 0$, $BX > 0$), then we would have $BX - AX$ in parentheses instead.

A.2.1 Calculate Trap Separation in nm

On the fleezers instruments, both traps are controlled by an acousto-optic modulator (AOM) that only moves the beads in one direction, x, so there is no y component. The AOM controls the position of both traps and both traps are capable of moving, however in practice we only move one and keep the other fixed. In an AOM, an acoustic wave is transmitted through a transparent quartz crystal. The wave produces a regular pattern of density variability—and thus refractive index variability—in the crystal that acts as a diffraction grating for the incoming beam of light. Adjusting the frequency of the acoustic wave is akin to changing the spacing of a diffraction grating, increasing the deflection of the first order maxima by decreasing the frequency of the acoustic wave. Using a method known as "time-sharing," by rapidly changing the applied acoustic frequency in time, our instrument creates two traps, separated spatially by a distance proportional to the difference in the frequencies applied to the AOM (ν_A, ν_B).

$$\Delta X_{traps} = convX(\nu_B - \nu_A)$$
$$\Delta Y_{traps} = 0$$

The conversion factor has units of nm/MHz and is calibrated using a CCD video of trapped beads and a reticule.

A.2.2 Calculate Bead Positions in nm

The locations of the beads in the two traps are monitored with a quadrant photo-diode (QPD). A QPD contains four photodetectors arranged in a square that output a voltage proportional to the total intensity of light shining on them. For a beam with a Gaussian intensity profile, comparing the difference in intensity between the left and right (and/or top and bottom) detectors gives the location of the beam's intensity peak in that direction (See **Figure 2.2**).

$$V_X = (V_1 + V_4) - (V_2 + V_3)$$
$$V_Y = (V_1 + V_2) - (V_3 + V_4)$$
$$V_{SUM} = V_Z = V_1 + V_2 + V_3 + V_4$$

A particle deflected laterally from the center of the trap (x and y directions) will displace the beam of light transmitted through it in the same direction as its movement. A particle deflected axially from the center of the trap (z direction) will expand or contract the beam of light transmitted through it, potentially changing the total intensity detected. To avoid changes in V_X and V_Y due to axial motion and overall changes in intensity, we scale these values by V_{SUM}.

$$X^{(QPD)} = V_X / V_{SUM}$$
$$Y^{(QPD)} = V_Y / V_{SUM}$$

We then calculate the position of the two beads the trap using the general equation

$$AX^{(nm)} = \alpha_{AX}(AX^{(QPD)} - offset_{AX}) \quad AY^{(nm)} = \alpha_{AY}(AY^{(QPD)} - offset_{AY})$$
$$BX^{(nm)} = \alpha_{BX}(BX^{(QPD)} - offset_{BX}) \quad BX^{(nm)} = \alpha_{BY}(BY^{(QPD)} - offset_{BY})$$

Here $AX^{(QPD)}, AY^{(QPD)}, BX^{(QPD)}, BY^{(QPD)}$ are the scaled QPD voltages and $\alpha_{AX}, \alpha_{AY}, \alpha_{BX}, \alpha_{BY}$ are conversion factors calculated from the trap calibration data (see Section A.1 for details).

A.2.3 Offset Calculation

For all of our calculations we need the distance from the bead center to the trap center, but the true center of the trap is not at the QPD center, it is *offset* somewhat. The offset is then the QPD voltage when the bead is at rest and the force is 0. This can be calculated either from calibration data or from an offset scan.

Using calibration data assumes that the offset value is the same at all trap separations. The offset value is the average position of the bead in volts during trap calibration, $offset_{AX} = mean(AX^{(QPD)})$.

However, there can be crosstalk between traps that deflects the beads when the two traps are close together making the offset a function of trap separation. In that case, the bead offset is not a static value, but changes over time during an experiment as the trap separation changes. In this case, we save an "offset file" that tracks the bead position in volts over a wide range of trap separations when no tether is formed (and thus the force should be 0). We use that file to find the offset as a function of trap separation, $f(\Delta X_{traps})$, and then calculate what the offset should be at each point in the data trace from the trap separation at that point.

$$offset_{AX}(t) = f_{AX}(\Delta X_{traps}(t))$$

$$AX^{(nm)}(\mathrm{t}) = \alpha_{AX}\left(AX^{(QPD)}(\mathrm{t}) - offset_{AX}(\mathrm{t})\right)$$

A.2.4 Total DNA Extension

We now have all the data we need to calculate DNA extension.

$$DNAext_{nm} = \sqrt{\left(\Delta X_{trap} - (AX_{nm} - BX_{nm})\right)^2 + \left(\Delta Y_{trap} - (AY_{nm} - BY_{nm})\right)^2} - r_{A,nm} - r_{B,nm}$$

A.2.5 Calculate the Force on the Beads

We calculate the force on each bead from the trap using Hook's law

$$F_{A,\mathrm{x}} = -\kappa_{A,x} AX_{nm} \quad F_{A,Y} = -\kappa_{A,y} AY_{nm}$$
$$F_{B,\mathrm{x}} = -\kappa_{B,x} BX_{nm} \quad F_{B,Y} = -\kappa_{B,y} BY_{nm}$$

where κ is the spring constant calculated from the trap calibration. From this we calculate the total force on each bead and take the tension in the DNA tether to be average of the two.

$$F_A = \sqrt{{F_{A,X}}^2 + {F_{A,Y}}^2} \quad F_B = \sqrt{{F_{B,X}}^2 + {F_{B,Y}}^2}$$
$$F_{DNA} = \frac{F_A + F_B}{2}$$

A.3 Calculating Base Pairs Open

As a helicase unwinds a tethered DNA hairpin construct at constant force in a dual optical trap experiment, the length of the DNA tether increases. The change in extension is given by

$$\Delta Ext = Ext_{\mathrm{tether}} - Ext_{\mathrm{closed}}$$

where Ext_{tether} is the length of the DNA tether in nm. This is experimentally determined as calculated in Section A.2. Ext_{closed} is the expected extension of the closed hairpin construct).

$$Ext_{\mathrm{closed}} = l_{\mathrm{ds}} + l_{\mathrm{ss}} + l_{\mathrm{HP\ width}}$$

where l_{ds} is the length of dsDNA in the construct, l_{ss} is the length of ssDNA in the closed construct, and $l_{\mathrm{HP\ width}}$ is the approximate width of the closed DNA hairpin in the direction of applied force (2nm). We can rewrite the above in terms of the extensible worm-like chain model (XWLC).

$$Ext_{\mathrm{closed}} = n_{\mathrm{ds}} h_{\mathrm{ds}}\, \mathrm{XWLC}_{\mathrm{ds}}(F) + n_{\mathrm{ss}} h_{\mathrm{ss}}\, \mathrm{XWLC}_{\mathrm{ss}}(F) + 2$$

where $\mathrm{XWLC}(F)$ gives the contour length of a WLC molecule at a given force, *F*, normalized by the contour length of the chain at rest. nh gives the contour length of DNA at rest. n_{ds} is the

length of the construct's dsDNA handles in bp, h_{ds} is the helix rise of dsDNA (0.34 nm). n_{ss} is the length of the construct's ssDNA loading site in nt, h_{ss} is the helix rise of ssDNA (0.6 nm).

The increase in extension is due entirely to the creation of 2nt of ssDNA for each bp of the hairpin unwound. Thus, we can also write ΔExt in terms of the number of basepairs unwound.

$$\Delta Ext = n_{\text{unwound}} \times (\text{length of 2 nt ssDNA}) = n_{\text{unwound}} (2h_{ss}\ \text{XWLC}_{ss}(F))$$

Combining our two equations for ΔExt, we find

$$n_{\text{unwound}} = \frac{Ext_{\text{tether}} - Ext_{\text{closed}}}{2h_{ss}\ \text{XWLC}_{ss}(F)}$$

A.4 Diffusion in Sample Chamber

In this section, we will derive the equation for the concentration of a solute as a function of position in a laminar flow cell. Common solutes for biophysical experiments are proteins, nucleic acids, and small molecules like ATP. In a typical flow cell like that shown in **Figure A.3**, two solutions, one with solute (blue) and one without (white), are pumped at constant velocity through separate channels cut into parafilm. The channels converge at the tip of a parafilm divider. Because flow is only in one direction (x) with no turbulence, the only mixing that occurs is due to diffusion in the y direction. At points farther away from the parafilm tip, particles have had more time to diffuse and thus more mixing between streams and solute dilution occur. The concentration, ϕ, as a function of position in the chamber is given by the equation

$$\phi(x,y) = \frac{1}{2} C_{\max} \left[1 + erf\left(\frac{y}{2} \sqrt{\frac{v_{flow}}{Dx}} \right) \right] \tag{A.7}$$

where x and y are the distance from the tip of the parafilm divider, $C_{\max}$ is solute concentration in the upper (blue) steam, D is the diffusion coefficient of the solute, and v_{flow} is the velocity of the particles. erf is the error function, defined as

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds \tag{A.8}$$

Since the only cause of mixing is diffusion, we start with the diffusion equation. We begin the derivation with concentration as function of time, and at the end will convert from time to distance away from the parafilm tip.

$$\frac{\partial \phi(y,t)}{\partial t} = D\frac{\partial^2 \phi(y,t)}{\partial y^2} \tag{A.9}$$

We take the Fourier transform of both sides, defining the transform of the concentration

$$\hat{\phi}(k,t) = \int_{-\infty}^{\infty} e^{-2\pi iky}\phi(y,t)dy$$

By the derivative property of Fourier transforms

$$\frac{\partial^2 \hat{\phi}(k,t)}{\partial y^2} = (-2\pi ik)^2\hat{\phi}(k,t) = -4\pi^2k^2\hat{\phi}(k,t)$$

Putting this back into Equation (A.9), we get an ordinary differential equation for $\hat{\phi}(k,t)$.

$$\frac{\partial \hat{\phi}(k,t)}{\partial t} = -4\pi^2Dk^2\hat{\phi}(k,t)$$

This means the time dependent part of $\hat{\phi}(k,t)$ is an exponential and the solution can be written in the form

$$\hat{\phi}(k,t) = \hat{\phi}_0(k)e^{-4\pi^2Dk^2t}$$

where $\hat{\phi}_0(k) = \hat{\phi}(k,t=0)$ and is determined by the initial conditions of our system. Thus, we have a general solution for $\phi(y,t)$:

$$\phi(y,t) = \Im^{-1}(\hat{\phi}(k,t)) = \int_{-\infty}^{\infty} e^{-4\pi^2Dk^2t+2\pi iky}\hat{\phi}_0(k)dk \tag{A.10}$$

$$\hat{\phi}_0(k) = \int_{-\infty}^{\infty} e^{-2\pi iky}\phi_0(y)dy \tag{A.11}$$

To solve our specific case, at time $t = 0$ no diffusion has occurred yet – the concentration of solute above the parafilm tip is C_{max} and below it is 0. We can describe the initial concentration, $\phi_0(y)$, with the Heaviside step function, $H(y)$, where y is the distance above/below the parafilm tip.

$$\phi_0(y) = C_{\max}H(y) = \begin{cases} C_{\max} & y > 0 \\ C_{\max}/2 & y = 0 \\ 0 & y < 0 \end{cases} \tag{A.12}$$

Plugging this initial condition into Equation (A.11), we see that $\hat{\phi}_0(y)$ is proportional to Fourier transform of the Heaviside function, which is known.

$$\hat{\phi}_0(k) = \int_{-\infty}^{\infty} e^{-2\pi iky} C_{\max} H(y) dy = C_{\max} \hat{H}(k)$$

$$\hat{H}(k) = \frac{1}{2}\left(\delta(k) + \frac{1}{\pi ik} \right)$$

$$\hat{\phi}_0(k) = \frac{C_{\max}}{2}\left(\delta(k) + \frac{1}{\pi ik} \right)$$

Applying this result to Equation (A.10) gives

$$\begin{aligned} \phi(y,t) &= \frac{C_{\max}}{2} \int_{-\infty}^{\infty} e^{-4\pi^2 Dk^2 t + 2\pi iky} \left(\delta(k) + \frac{1}{\pi ik} \right) dk \\ &= \frac{C_{\max}}{2} \left(1 + \int_{-\infty}^{\infty} e^{-4\pi^2 Dk^2 t + 2\pi iky} \frac{dk}{\pi ik} \right) \\ &= \frac{C_{\max}}{2} \left(1 + I(y) \right) \end{aligned} \qquad \text{(A.13)}$$

where

$$I(y) = \int_{-\infty}^{\infty} e^{-4\pi^2 Dk^2 t + 2\pi iky} \frac{dk}{\pi ik}$$

The $1/i\pi k$ can be removed by differentiating both sides with respect to y.

$$\frac{\partial I(y)}{\partial y} = 2\int_{-\infty}^{\infty} e^{-4\pi^2 Dk^2 t + 2\pi iky} dk$$

We can rewrite the exponential more simply by completing the square leaving only one term with k in it instead of two.

$$\frac{\partial I(y)}{\partial y} = 2\int_{-\infty}^{\infty} e^{-(2\pi k\sqrt{Dt} + ay)^2 + a^2 y^2} dk = 2e^{a^2 y^2} \int_{-\infty}^{\infty} e^{-(2\pi k\sqrt{Dt} + ay)^2} dk$$

A change of variables gives us a Gaussian integral.

$$u = 2\pi k\sqrt{Dt} + ay$$

$$dk = \frac{du}{2\pi\sqrt{Dt}}$$

$$\frac{\partial I(y)}{\partial y} = 2e^{-a^2 y^2} \int_{-\infty}^{\infty} e^{-u^2} \frac{du}{2\pi\sqrt{Dt}} = \frac{1}{\sqrt{\pi Dt}} e^{-a^2 y^2}$$

Solving for a for to complete the square we see that

$$-(2\pi k\sqrt{Dt}+ay)^2+a^2y^2=-4\pi^2Dk^2t+2\pi iky$$

$$a=\frac{-i}{2\sqrt{Dt}}$$

Finally giving us

$$\frac{\partial I(y)}{\partial y}=\frac{1}{\sqrt{\pi Dt}}e^{\frac{-y^2}{4Dt}}$$

To calculate $I(y)$, we must integrate this expression.

$$I(y)=\int_0^y\frac{\partial I(y')}{\partial y'}dy'=\frac{1}{\sqrt{\pi Dt}}\int_0^y e^{\frac{-y'^2}{4Dt}}dy'$$

Changing variables once more

$$s=\frac{y'}{2\sqrt{Dt}}$$

$$ds=\frac{dy'}{2\sqrt{Dt}}$$

So, the integral, $I(y)$, becomes

$$I(y)=\frac{2}{\sqrt{\pi}}\int_0^{y/2\sqrt{Dt}}e^{-s^2}dv$$

This matches the definition of the error function in Equation (A.8) giving us

$$I(y)=erf\left(\frac{y}{2\sqrt{Dt}}\right)$$

Plugging this back into Equation (A.13) we have

$$\phi(y,t)=\frac{C_{\max}}{2}\left[1+erf\left(\frac{y}{2\sqrt{Dt}}\right)\right]$$

Note that t is the amount of time the particles have been diffusing and this is related to the distance away from the parafilm tip, x, by the flow velocity: $t=x/v_{flow}$. Substituting this equation in for t, we find the concentration as a function of position in the chamber.

$$\phi(x,y)=\frac{1}{2}C_{\max}\left[1+erf\left(\frac{y}{2}\sqrt{\frac{v_{flow}}{Dx}}\right)\right]\ \blacksquare$$

A.4.1 Practical Notes on Calculating v_{flow}

v_{flow} in this equation (A.7) is the particle velocity measured in distance/time with units matching those of Dx (typically mm/s). It is *not* the pump velocity set in volume/time on the syringe pumps that generates the constant flow in the chamber. To convert between the two, we need to multiply by the cross-sectional area of the chamber, $\sigma_{chamber}$ (units mm^2), and account for the fact that we have multiple channels combining into one. Most Chemla Lab chambers have 2 channels with constant flow v_{pump} (units µl/hr) combine into 1 central channel. The total flow rate in the central channel after combining will then be $2v_{pump}$. The flow rate, v_{flow} (units mm/s) is given by the equation

$$v_{flow} = \frac{2v_{pump}}{60^2 \sigma_{chamber}} \tag{A.14}$$

The size of the central channel, after streams combine, varies from chamber to chamber, but is generally ~3mm wide (top to bottom) and ~0.150 mm deep (coverslip to coverslip). Thus, a typical value for the cross-sectional area is $\sigma_{chamber} = 0.45$ mm^2. For pump rates of $v_{pump} = 100$ µl/hr, the velocity of flowing particles would be $v_{flow} = 0.1235$ mm/s.

This can also be generalized for n channels combining into one (e.g. $n_{channels}$=3 for the experiments in Chapter 4:).

$$v_{flow} = \frac{n_{channels} v_{pump}}{60^2 \sigma_{chamber}} \tag{A.15}$$

In that case, equation (A.7) for the concentration, $\phi(x, y)$, will hold across each interface separately with the above value of v_{flow}.

A.5 Figures

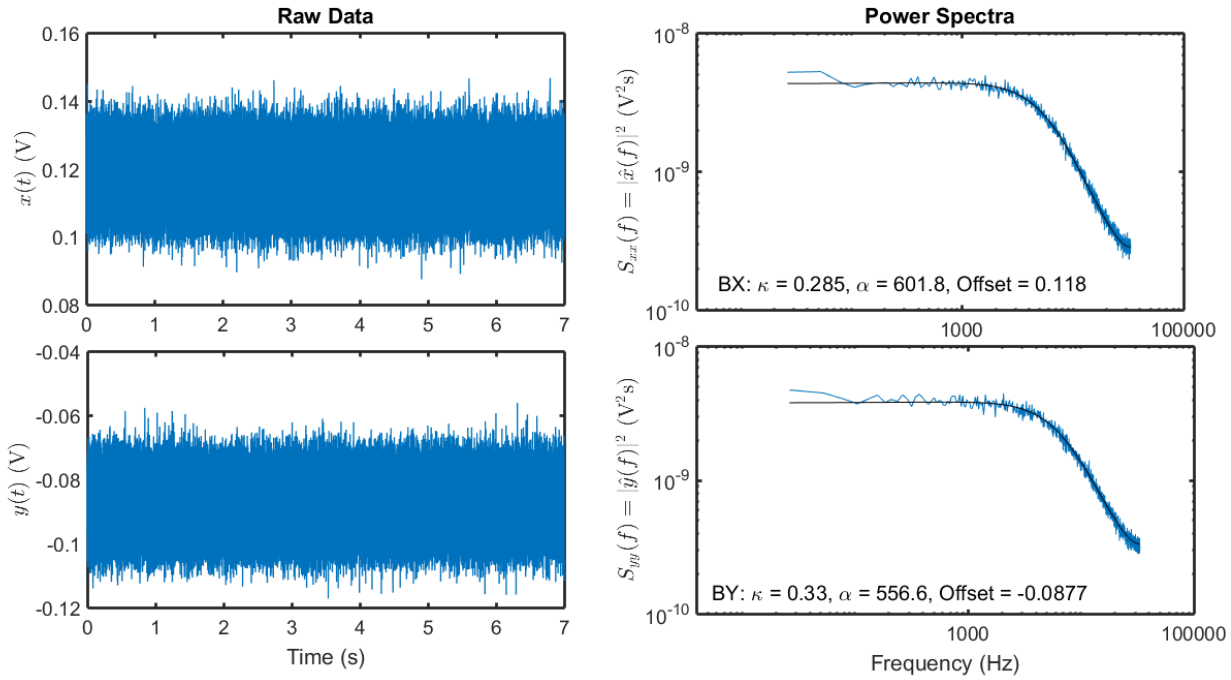

Figure A.1 Calibrating trap stiffness. Left: Raw x and y position data for a trapped bead saved from the QPD in units of volts. Right: Power spectrum of the raw data in V^2s (blue data). Fitting the spectrum to the theoretical model (black line) gives calibrated values for trap stiffness (κ), V to nm conversion (α), and the location of the bead at rest with no applied force (offset).

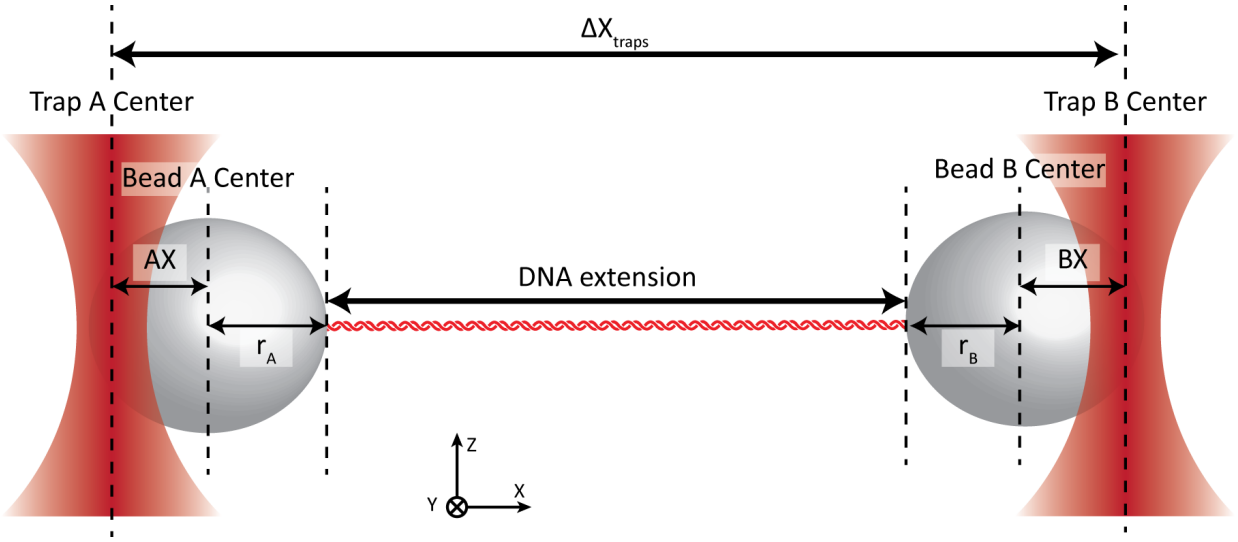

Figure A.2 Geometry of two trapped beads connected by a DNA tether. Variables used in Section A.2 are shown above. ΔX_{traps} is the distance between the centers of trap A and trap B. *AX* (*BX*) is the poition of the center of bead A (B) relative to center of trap A (B). This value has a sign indicating that the ± direction along the x-axis. In this figure, $AX > 0$ and $BX < 0$. r_A (r_B) is the radius of bead A (B).

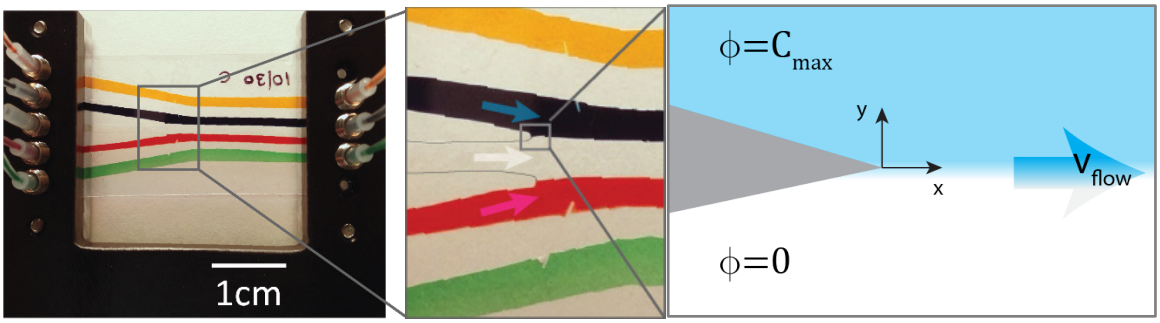

Figure A.3 Layout of a laminar flow cell. Two channels separated by parafilm meet at and combine into one larger channel. Buffer is pushed through all channels from left to right at a steady velocity, $v_{\max}$. We set the origin at the tip of the parafilm. The top channel (blue) contains concentration $\phi = C_{\max}$ of a solute. The lower channel (white) contains no solute, $\phi = 0$.

Appendix B: Protocols

B.1 Making DNA Hairpin Construct

A two-step ligation protocol for making a DNA hairpin construct. A pictorial representation of major steps in the process can be found in **Figure B.1**.

Procedure:

1) PCR Handles
    a) In two separate PCR tubes, combine the reagents for the left and right handles respectively in the amounts shown in **Table B.1** for total volume 100 µl each.
        i) Primers should be 10uM each, diluted and stored in IPTE.
        ii) Plasmid DNA template concentration should be 10 ng/µl.
        iii) Names and sequences of primers and their corresponding plasmid templates can be found in Appendix C.1.
    b) Run in thermocycler at the temperature, times, and with the number of cycles given in **Table B.2**.
    c) PCR purify Left and Right Handles separately using the QIA Quick protocol. Elute to a final volume of 30 µl. Do not do step 10.
2) Digest and Dephosphorylate Left Handle
    a) Digest LH with PSPG1
        i) Combine in PCR tube
            (1) 30 µl LH DNA
            (2) 3.5 µl NEB4 10x (or CutSmart 10x)
            (3) 2 µl PSPG1 (10 kU/ml)
        ii) Run in thermocycler at 75 C for 1 hour (no heat inactivation needed)
    b) Dephosphorylate LH with Antarctic Phosphatase
        i) Add to SAME TUBE (no purification between steps)
            (1) 2 µl Antarctic Phosphatase
            (2) 4 µl Antarctic Phosphatase Buffer
        ii) Run in thermocycler
            (1) 37 C for 15 min

(2) 70 C for 5 min (inactivate)

c) PCR purify as before the digested and dephosphorylated Left Handle with 30µl final volume

3) Ligate Right Handle + Hairpin Insert Overnight
   a) Measure concentration of Right Handle on NanoDrop if desired
      i) If unmeasured, concentration is assumed to be ~200 nM
      ii) Exact concentration not critical because we are using such a huge excess of hairpin
   b) Combine in PCR tube
      i) 30 µl Right Handle (RH)
      ii) 20x excess of Hairpin insert (HP) – ~12 µl 10 µM Hairpin
   c) Anneal 30 min at Room Temperature (RT)
   d) To annealed mixture add
      i) 6 µl 10x T4 Ligase Buffer
      ii) 3 µl T4 Ligase
      iii) 9 µl DNase free H20 or until final volume = 60 µl
   e) Run in thermocycler overnight
      i) 16 C or 16 hours (ligate)
      ii) 65 C for 15 min (deactivate)
   f) PCR Purify as before the ligation product (RH+HP) with 30 µl final volume
4) Anneal PCR Purified Product (RH+HP)
   a) On heat block, incubate RH+HP at 95 C for 2 min
   b) Remove from heat and allow to cool to Room Temperature on bench (~1hr)
5) Gel Purify RH+HP
   a) Make 1% agarose gel
      i) Add 60 ml 0.5x TBE buffer to a 150 ml Erlenmeyer flask
      ii) Add 0.6 g agarose
      iii) Microwave solution without boiling, stopping every ~30 sec to mix, until gel is dissolved (~1-1.5 min total)
      iv) Level a gel box

v) Pour in melted gel, popping any bubbles that may form
vi) Place comb – large, 8 wells total
vii) Cover entire gel box with foil to prevent moving air from causing ripples
viii) Allow to cool at RT ~1hr or until firm

b) Place gel in thing with 0.5x TBE and add to the center lanes
i) 10 µl of 1kb ladder + 2 µl Novel Juice for imaging
ii) 30 µl RH+HP product + 6 µl Novel Juice for imaging

c) Run gel at 85 V for 90 min

d) Visualize bands with blue light transluminator and cut out band at ~1.5 kbp with a razorblade. Cut away as much excess gel as possible.

e) Follow QIA Quick gel extraction protocol, skipping step 5 and eluting to a final volume of 30 µl.

6) Ligate Left Handle + RH+HP Overnight
a) Measure final concentrations of LH and purified RH+HP product on NanoDrop
b) Calculate amounts needed for a 1:1 ratio of LH:RH+HP and final volume < 51 µl
c) Combine in PCR tube
i) RH+HP (calculated quantity)
ii) LH (calculated quantity)
iii) 6 µl 10x T4 Ligase Buffer
iv) 3 µl T4 Ligase
v) DNase free H20 until final volume = 60 µl

d) Run in thermocycler overnight
i) 16 C or 16 hours (ligate)
ii) 65 C for 15 min (deactivate)

e) PCR Purify as before the ligation product (LH+RH+HP) with 30 µl final volume

7) Anneal PCR Purified Product (LH+RH+HP)
a) On heat block, incubate LH+RH+HP at 95 C for 2 min
b) Remove from heat and allow to cool to Room Temperature on bench (~1 hr)

8) Gel Purify Full Construct

a) Make 1% agarose gel as before
b) Place gel in thing with 0.5x TBE and add to the center lanes
    i) 10 µl of 1kb ladder (add 2 µl 6x loading dye if not already in it)
    ii) 30 µl LH+RH+HP product + 3 µl of 100% glycerol (or 6 µl of 6x loading dye)
c) Run gel at 85 V for 90 min
d) Stain gel in Gel Green bath for 30 min (note: final product amount too low to use Novel Juice. Gel Green is more sensitive.)
e) Destain gel in fridge for 1 hour
f) Visualize bands with blue light transluminator and cut out band at ~3 kbp with a razorblade. Cut away as much excess gel as possible.
g) Follow QIA Quick gel extraction protocol, skipping step 5 and eluting to a final volume of 30 µl.

B.2 Bead Preparation

B.2.1 Streptavidin Beads

Materials:

- 1% w/v ~0.8 µm streptavidin polystyrene beads (Spherotech, Inc.)
- Phosphate buffered saline, pH 7.0 (1x PBS) + 0.01% tween 20

Procedure:

1) Wash 40 µl streptavidin beads in 160 µl PBS+tween
2) Spin at 7,500 rpm for 1 minute
3) Remove supernatant and add 200 µl PBS+tween
4) Resuspend pellet by vortexing 10-30 sec on high
5) Repeat steps 2-4) three times.

Store in final wash at 4 C.

B.2.2 Digoxigenin Beads

Materials:

- 1% w/v ~0.8 µm Protein G polystyrene beads (Spherotech, Inc.)

- Phosphate buffered saline, pH 7.0 (1x PBS) + 0.01% tween 20
- Anti-Digoxigenin reconstituted to PBS

Procedure:

1) Wash 40 µl streptavidin beads in 160 µl PBS+tween
2) Spin at 7,500 rpm for 1 minute
3) Remove supernatant and add 200 µl PBS+tween
4) Resuspend pellet by vortexing 10-30 sec on high
5) Repeat steps 2-4) two times.
6) Add 10 µl Anti-Dig antibody reconstituted to PBS after final wash
7) Spin at 7,500 rpm for 1 minute
8) Remove supernatant and add 200 µl PBS+tween
9) Resuspend pellet by vortexing 10-30 sec on high
10) Repeat steps 7-9) three times.

Store in final wash at 4 C.

B.3 Making a Laminar Flow Cell

Steps for making a laminar flow cell for use in trapping experiments.

Materials:

- Coverslips
- Nescofilm
- Capillaries

Procedure:

1) Clean coverslips in acetone
    a) Take 3 pre-cut coverslips and 3 regular coverslips and place them in a holder. It is a good idea to clean several in case the delicate slides break during assembly.
    b) Fill holder with acetone and close.
    c) Sonicate on high for 15 minutes.
2) Rinse coverslips with distilled water and dry with lab air or a kimwipe.
3) Cut Nescofilm on Versa laser engraver in desired channel pattern.

4) Carefully align the ends of each Nescofilm channel over the holes in the pre-cut coverslip and place the film down onto the coverslip.
    a) It is best to place the film on the coverslip on the first try and not to move it. Lifting the film leaves a residue on the glass surface that cannot be removed without disassembling the chamber.
    b) Take care to evenly space the film wedges that form the 3-channel central trapping region. Each channel should be about the same width.
5) With a razor blade, cut capillary and place on top of Nescofilm spanning between the upper bead channel and the central trapping channel.
    a) Do the same for the bottom capillary.
6) Place the un-cut coverslip on top, again in one motion without lifting.
7) Put the completed chamber on a heat block under a weight to keep it flat against the block.
    a) Set the heat block on medium heat (4/10).
    b) Heat chamber in 30 sec - 1 min increments allowing the chamber to cool in between.
    c) Repeat until edges around channels are sealed with minimal bubbles in the film.

B.4 Pegyllating a Laminar Flow Cell

Protocol for making a PEG-coated trapping chambers from Matt Comstock. Used in experiments with fluorescently labeled RPA2 to prevent protein sticking to the glass chamber surface. For this protocol to succeed, cleanliness is key. Make sure all materials – tweezers, holders, etc. – are clean. Wipe tweezers with a kimwipe and acetone before using.

Materials:

- PEG
- Amino silane
- 3 M KOH
- Acetic acid
- Methanol

Procedure:

1) Prepare clean coverslips

a) Take 3 pre-cut coverslips and 3 regular coverslips and wipe them with a kimwipe and acetone to remove any visible dirt. It is a good idea to clean several in case the delicate slides break during assembly.
b) Place the coverslips in a plastic coverslip washing holder and place into a glass staining dish. Coverslips will remain in plastic holder for all cleaning steps.
c) Fill holder with acetone, enough to cover the coverslips, and sonicate for 30 min.
    i) Can clean amino-silane flask while waiting
d) Rinse coverslips with DI water 3x by filling and then emptying the holder.
e) Cover coverslips with 3 M KOH and sonicate for 20 min.
f) Rinse with DI water 5x.
g) Rinse with methanol 2x.
h) Dry with nitrogen gas keeping the coverslips in the holder.

2) Make Chambers
    a) Make chambers as usual, described in Section B.3
    b) Important! Place Nescofilm and top coverslip correctly the *first* time. The Nescofilm will leave a film on the glass when lifted, dirtying the glass.

3) Coat chamber channels in amino silane in situ.
    a) Assemble chamber in experiment bracket with fresh inlet and outlet tubing for the amino silane. Inlet tubing should be short to reduce the chance of dirt contamination or a bad reaction – ~1 inch long.
    b) Clean a 250 ml Erlenmeyer flask for amino silane
        i) Fill flask with 1 M KOH and sonicate uncovered for 10 min.
        ii) Rinse with DI water
        iii) Sonicate flask with methanol for 20 min
        iv) Rinse with methanol
        v) Dry with nitrogen gas
        vi) Cover top with parafilm.
    c) Fill clean flask with
        i) 100 ml methanol

ii) 5 ml acetic acid

iii) 1 ml aminopropyl silane (use a glass pipette tip to dispense)

d) Fill chamber with MilliQ water (in 10 ml plastic syringes with 0.2 µm filters) making sure there are no bubbles in the tubing or chamber

e) Fill chamber with amino silane mixture using 1 ml plastic syringes. Flow ~1 ml into each channel adding a few 100 µl at a time to each channel. Make sure there are no bubbles and that the entire chamber is filled.

f) Incubate on benchtop for 10 min.

g) Rinse all channels with ~1 ml methanol per channel again using 1 ml plastic syringes.

h) Rinse chamber with a few ml MilliQ water.

4) Coat chamber channels in PEG in situ.

a) Make PEG buffer

i) 10 ml MilliQ water.

ii) 84 mg $NaHCO_3$

b) Make PEG mixture. Need ~200 µl per channel.

i) For 3 central channels, combine in a 1.5 ml tube

(1) 600 µl PEG buffer

(2) 200 mg mPEG

ii) Mix by flicking until mPEG is dissolved.

iii) Centrifuge 1 min at 10k rpm to pellet any undissolved PEG.

c) Fill desired channels with PEG mixture, 200 µl per channel.

i) As for amino silane, use 1 ml plastic syringes

ii) Fill slowly and carefully, avoiding bubbles.

d) After filling the chamber with PEG mixture, insert syringe needles into outlet tubing and cap inlet and outlet tubes with luer locks. This will prevent evaporation during a long incubation.

e) Incubate chamber in a closed, dark drawer for 4 hours.

f) Rinse with several ml of MilliQ water.

g) Dry chamber with nitrogen gas and store chamber at 4 C.

5) Store and use chamber.
    a) May not perform at its best for the first few uses.
    b) Rinse well with MilliQ and dry well with nitrogen gas (~10 min at 10 psi) after each use.
    c) Store at 4 C between uses.

B.5 Protocol to Fluorescently Label RPA2

1) Change RPA2 buffer to labelling buffer
    a) Dilute stock of RPA2 in labelling buffer and measure concentration
    b) Run diluted RPA2 through P6 desalting column and measure the final concentration
2) Incubate with Cy dye
    a) Add Cy dye to RPA2 such that [dye]:[RPA2] = 10:1
    b) Incubate on rotator at room temperature for 30 minutes.
3) Terminate reaction with Tris buffer
    a) Add Tris-HCl (pH 7.5) to final concentration on 50mM.
4) Separate free dye from Cy-RPA2
    a) Run through P6 column twice
5) Measure the final concentration and check the labelling efficiency (should be ~1:1).
6) Aliquot with 30% glycerol and store at -80 C.

B.6 Calculating Protein Concentration from measured A_{280}

Measure absorbance at 280 nm (A_{280}) on the NanoDrop™, following the instrument instructions. Note: glycerol in the sample can interfere with measurement at A_{280}. You must correct for this with your blank or measure the concentration before adding glycerol. USP glycerol may not interfere, but it was not used in these experiments.

Look up the extinction coefficient, ϵ_{280}, for the protein of interest in $M^{-1}cm^{-1}$.

XPD: $\epsilon_{280,XPD} = 91300\ M^{-1}cm^{-1}$

RPA2: $\epsilon_{280,RPA2} = 21890\ M^{-1}cm^{-1}$

For unlabeled proteins:

[Protein] = A_{280}/ ϵ_{280}

For fluorescently labelled proteins:

In addition to measuring the absorption as 280 nm (A_{280}), you must also measure A_λ, the absorption of your sample at the wavelength of your fluorophore's peak absorption, λ. The concentration of the label is then given by:

[Label] = A_λ/ $\epsilon_{\lambda,label}$

Knowing the concentration of fluorophores, you can calculate the amount it is absorbing at 280 and correct for it when calculating the concentration of protein. Each label will have a correction factor, c.f., for absorbance at 280nm.

[Protein] = (A_{280}-(c.f.* A_λ))/ $\epsilon_{280,protein}$

For Cy3

λ = 552 nm

$\epsilon_{552,Cy3}$ = 150,000 $M^{-1}cm^{-1}$

c.f. = 0.08

For Cy5

λ = 650 nm

$\epsilon_{650,Cy5}$ = 170,000 $M^{-1}cm^{-1}$

c.f. = 0.05

B.7 Figures and Tables

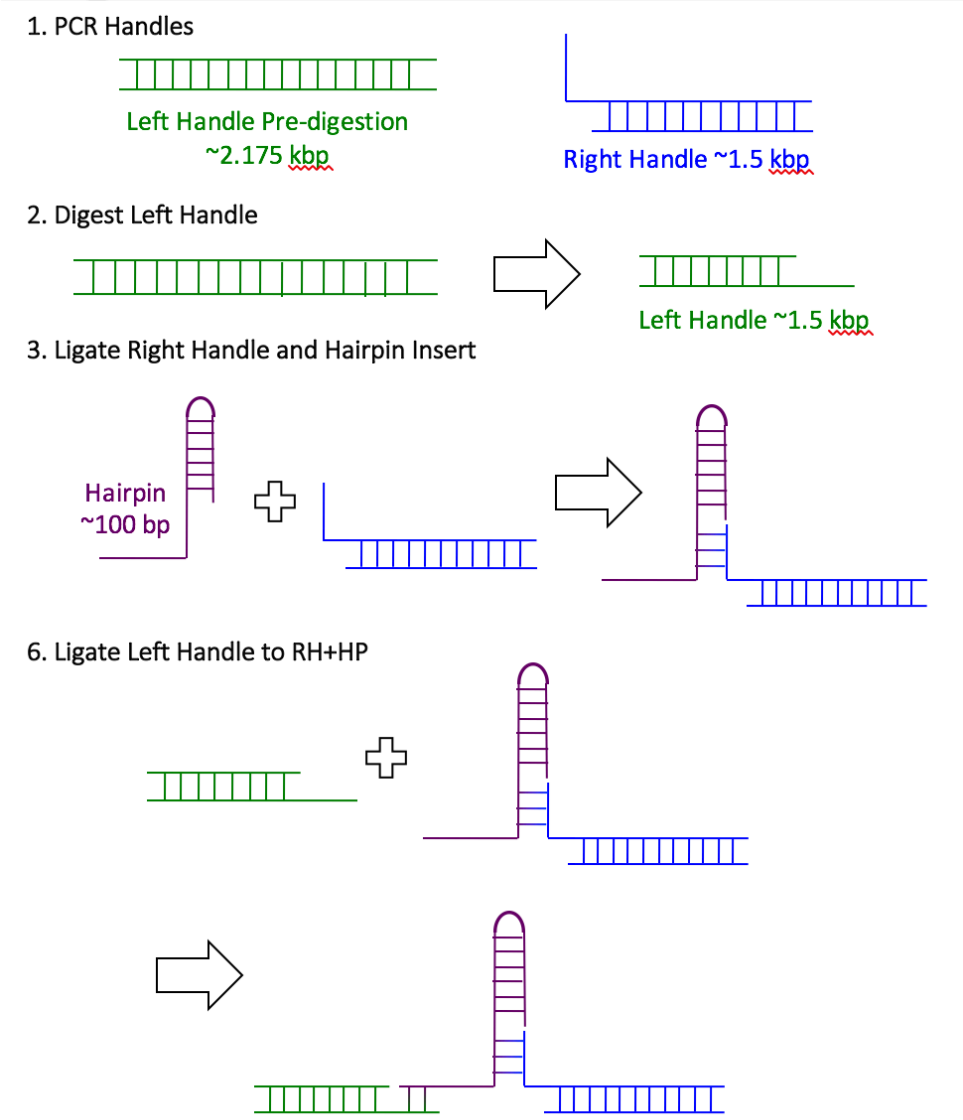

Figure B.1 Pictorial representation of DNA hairpin construction. Numbers match the steps described in Section B.1. **1)** PCR produces a dsDNA left handle and right handle with a 29 nt ssDNA overhang that will bind to hairpin insert. **2)** left handle is digested, leaving a 1.5 kbp dsDNA with a 5nt ssDNA overhang that will bind to the hairpin insert. **3)** Right handle is ligated to the hairpin insert overnight producing a connected RH+HP. **6)** Left handle is ligated to the connected RH+HP construct producing the full hairpin construct used in experiments.

Reagent Name	Amt. For Left Handle (µl)	Amt. For Right Handle (µl)
2x Phusion® Master Mix	50	50
Forward Primer	5	5
Reverse Primer	5	5
DNA template	2 (total 20ng)	2 (total 20ng)
100% DMSO	3	3
DNase free H2O	35	35
Total	100	100

Table B.1 Amounts of reagents for Handle PCR, based on instructions in the Phusion® Master Mix (NEB M0531S) manual.

Cycles	Temperature (C)	Time	Step
1	98	30s	Denature
34	98	10s	Denature
	59	30s	Anneal
	72	33s	Extension
1	72	10 min	Extension
	4	hold	Store

Table B.2 Temperatures and number of cycles for PCR in thermocycler. Times based on recommendations in the Phusion® Master Mix (NEB M0531S) manual. Annealing temperature is the melting temperature of primers given by IDT.

Appendix C: Reagents

C.1 Primers and Hairpin Inserts

Primers for PCR and hairpin inserts for the final DNA construct were ordered from Integrated DNA Technologies (IDT) with the following sequences and functional groups.

C.1.1 Original Sequence

The primer and insert sequences, originally designed by Zhi Qi, that were used to make the orignal sequence hairpin. Used in experiments in Chapters 2-4. See **Table C.1**.

C.1.2 Original Sequence +Cy3

The primer and insert sequences that were used to make the orignal sequence hairpin with a Cy3 label near the loading site. Used in experiments in Section 3.4. See **Table C.2**.

C.1.3 New Sequence

The new hairpin sequence had a very low ligation efficiency and propensity to mis-ligate into a seemingly purely double stranded construct. Thus, many iterations were made to adjust the sequence, primers, and even the plasmid DNA in attempts to increase efficiency. The primer and insert sequences that were ultimately successful and resulted in the data presented in Chapter 5: is given in **Table C.3**. Even so these primers had a strong propensity to mis-ligate and future experiments will make further adjustments to increase efficiency. Failed iterations of the hairpin insert and right handle primer are listed in **Table C.4**.

C.2 Buffer Recipes

General chemistry note: Always add powders to a smaller volume of liquid than your final amount (~60%). Once dissolved, fill the rest of the way.

C.2.1 TMS 20

Working buffer for experiments

- 100 mM Tris Base (Fisher BP152) Tris = 121.14 g/mol
- 20 mM NaCl (Sigma S3014) NaCl = 58.44 g/mol
- 3 mM $MgCl_2$ (Sigma M2670) $MgCl_2$ = 2 M solution

Check pH and add HCl until it reaches 7.6. Filter and store.

C.2.2 TMS 20 + Trolox

For fluorescence experiments: Trollox helps extend fluorophore lifetimes and prevent dye "blinking." For 10 ml final volume:

Add 50 µl 1M NaOH (Sigma 221465) to 8.5 ml H_2O in a 10 ml conical tube. This step raises the pH to dissolve the Trolox. Add 8 mg Trolox powder (Sigma 238813) to tube. Wrap the tube in foil and place on rotator for one hour until completely dissolved. Then add to the tube

- 1 ml 1M Tris-HCl pH 7.6
- 200 µl 1M NaCl (Sigma S3014)
- 15 µl 2M $MgCl_2$ (Sigma M2670)
- 0.031g DTT powder

Add H_2O until final volume reaches 10 ml. Filter and store in foil at 4 C.

C.2.3 Experiment Buffer

For 1000 µl final volume

- 960 µl TMS20 (or TMS20 + Trolox)
- 20 µl 20% glucose (Fisher D16)
- 10 µl BSA (1mg/ml) (NEB B9000S)
- 10 µl poxy

Filter with a 0.2 µm filter before adding protein or ATP to solution. For a four experiment, make at least 400 µl per channel.

C.2.4 5X TBE

Used for making running agarose gels. In 800 ml H_2O, dissolve:

- 54 g Tris Base (Fisher BP152)

- 27.5 g Boric Acid (Fisher BP168)
- 3.7 g EDTA (Fisher BP120)

Add H_2O until total volume is 1 L. Filter and store.

C.2.5 PBS

Used for storing polystyrene beads. In 800 ml H_2O, dissolve:

- 8 g NaCl (Sigma S3014)
- 0.2 g KCl (Sigma P3911)
- 1.44 g Na_2HPO_4 (Sigma S9763)
- 0.24 g KH_2PO_4 (EMD PX-1565-1)

Adjust pH to 7.4 by adding concentrated HCl. Add H_2O until total volume is 1 L. Filter and store.

C.2.6 T50

For making glucose oxidase (gloxy) and pyranose oxidase (poxy).

- 10 mM Tris Base Tris = 121.14 g/mol
- 50 mM NaCl NaCl = 58.44 g/mol

Check pH and add HCl until it reaches 7.5. Filter and store at room temperature.

C.2.7 TS200

For storing XPD and RPA2. For 10ml total volume:

- 50mM Tris-HCL (pH 7.6) 500µl s1M Tris-HCl
- 200mM NaCl 2ml 1M NaCl
- 2mM DTT 0.003g DTT (154.25 g/mol)

Fill rest of volume with MilliQ H_2O. Store TS200 in 1mL aliquots at -20 C. DTT will degrade on the order of weeks/months when stored at 4C. Degraded DTT seems to produce spectral peak ~280 nm – terrible for NanoDrop measurements. Before diluting XPD/RPA, make new TS200 or test spectrum of frozen aliquots on NanoDrop with H_2O blank.

To store XPD or RPA2, dilute with TS200 and add glycerol to 30% of final volume. Aliquot and store at -80 C for long-term storage. For short-term use (as when taking data regularly) store aliquots at -20 C.

C.2.8 ATP and Storage Buffer

Make ATP buffer. For 500 µl total volume:

- 100mM Tris-HCl (pH 7.6) 50 µl 1M Tris-HCl
- 20mM NaCl 10 µl 1M NaCl
- 440 µl MilliQ H_2O

To 500 µl ATP buffer, add 0.0138 g of ATP powder (551.14 MW) for a final concentration of 50 mM. Centrifuge filter, aliquot, and store at -20 C.

C.2.9 ATPγS and Storage Buffer

Make ATPγS buffer:

- 50mM Tris-HCl (pH 7.6)

To 36.56 µl ATPγS buffer, add 0.001 g of ATPγS powder (546.98 MW) for a final concentration of 50mM. Centrifuge filter, aliquot, and store at -20 C.

C.3 Oxygen Scavenging Systems

C.3.1 Pyranose oxidase (poxy)

In 200 µl of T50 buffer dissolve

- 5.8 mg of Pyranose Oxidase (Sigma P4234)
- 1.3 mg Catalase (EMD 219001)

Centrifuge the solution at 13,000 rpm for 5 minutes. Keep the supernatant. Centrifuge filter (Millipore UFC30GV0S) the supernatant at 13,000 rpm for 1 minute. Store at 4 C.

C.4 Tables

	Template	Oligo Name	Sequence 5’ to 3’
Right Handle Forward Primer	pBR322	HRv8_FWDdsp_ZQ_ 5phosphate_0T	/5Phos/TTG AAA TAC CGA CCG CTC AGC TAT CAG CC – /idSp/ CTC TGA CAC ATG CAG CTC CC
Right Handle Reverse Primer	pBR322	REV_3400bp_5DIG_ZQ	/5DigN/ CAA CAA CGT TGC GCA AAC T
Left Handle Forward Primer	pBR322	HLv7_FWD5Biosg_ZQ	/5Biosg/ TGA AGT GGT GGC CTA ACT ACG
Left Handle Reverse Primer	pBR322	LH_REV	CAA GCCT ATG CCT ACA GCA T
Hairpin Insert	-	Hairpinv10_Mine_ZQ_10T	/5Phos/CCT GG – T TTT TTT TTT – GGC TGA TAG CTG AGC GGT CGG TAT TTC AA – A AGT CAA CGT ACT GAT CAC GCT GGA TCC TAG AGT CAA CGT ACT GAT CAC GCT GGA TCC TAT TTT TAG GAT CCA GCG TGA TCA GTA CGT TGA CTC TAG GATCCA GCG TGA TCA GTA CGT TGA CTT

Table C.1 Oligos for constructing the original sequence hairpin construct. 10T loading site on hairpin insert is colored red. 4T loop at the end of the hairpin is colored blue.

	Template	Name	Sequence 5’ to 3’
Right Handle Forward Primer	pBR322	HRv8_FWDdsp_ZQ_ 5phosphate_0T	/5Phos/TTG AAA TAC CGA CCG CTC AGC TAT CAG CC – /idSp/ CTC TGA CAC ATG CAG CTC CC
Right Handle Reverse Primer	pBR322	REV_3400bp_5DIG_ZQ	/5DigN/ CAA CAA CGT TGC GCA AAC T
Left Handle Forward Primer	pBR322	HLv7_FWD5Biosg_ZQ	/5Biosg/ TGA AGT GGT GGC CTA ACT ACG
Left Handle Reverse Primer	pBR322	LH_REV	CAA GCCT ATG CCT ACA GCA T
Hairpin Insert	-	HPSeq1_10T-5ntCy3_50bp	/5Phos/ CCT GG /iCy3/ T TTT TTT TTT GGC TGA TAG CTG AGC GGT CGG TAT TTC AAA AGT CAA CGT ACT GAT CAC GCT TTT GCG TGA TCA GTA CGT

Table C.2 Oligos for constructing the original sequence hairpin construct with Cy3 label. 10T loading site on hairpin insert is colored red. 4T loop at the end of the hairpin is colored blue. The Cy3 dye location is colored in green.

	Template	Name	Sequence 5’ to 3’
Right Handle Forward Primer	λDNA	SSB RH REV BKS1.1	/5Phos/GA CTG TGA CTG ACA TGA GTG ACT GAG ACT /idSp/ CGT TTT CCC GAA AAG CCA GAA
Right Handle Reverse Primer	λDNA	SS RH REV	CGT TTT CCC GAA AAG CCA GAA
Left Handle Forward Primer	pBR322	HLv7_FWD5Biosg_ZQ	/5Biosg/ TGA AGT GGT GGC CTA ACT ACG
Left Handle Reverse Primer	pBR322	LH_REV	CAA GCCT ATG CCT ACA GCA T
Hairpin Insert	-	HP_BKSv1.1_10T	/5Phos/CC TGG TTT TTT TTT TAG TCT CAG TCA CTC ATG TCA GTC ACA GTC AGA GTC ATG TCT GAG TCT TGA TGA TGT CAC TGA CTG AGA CTC TGA CTC ACT GAG TCG AGC TTT TGC TCG ACT CAG TGA GTC AGA GTC TCA GTC AGT GAC ATC ATC AAG ACT CAG ACA TGA CTC TG

Table C.3 Oligos for constructing the new, flat P_{open} sequence hairpin construct. 10T loading site on hairpin insert is colored red. 4T loop at the end of the hairpin is colored blue.

Right Handle Forward Primer	pBR322	HR _BKSv1.1_0T (Unsuccessful)	/5Phos/GA CTG TGA CTG ACA TGA GTG ACT GAG ACT /idSp/CT CTG ACA CAT GCA GCT CCC
Right Handle Forward Primer	pBR322	HR_FWD_BKSv1_0T (Unsuccessful)	/5Phos/ GAC TGA GAC TGA CAT CTC AGA CTG AGA GA /idSp/ C TCT GAC ACA TGC AGC TCC C
Right Handle Forward Primer	λDNA	Lambda RH REV BKS1.0 (Unsuccessful)	/5Phos/GAC TGA GAC TGA CAT CTC AGA CTG AGA GA /idSp/ CGT TTT CCC GAA AAG CCA GAA
Hairpin Insert	-	HP_BKSv1_10T (Unsuccessful)	/5Phos/ CCT GGT TTT TTT TTT TCT CTC AGT CTG AGA TGT CAG TCT CAG TCA GAG TCT TGT CTG TGT CTT GTT GAT GTC ACT GAC TGA GAC TCT GAC TGT CTG GGT CGC GCT TTT GCG CGA CCC AGA CAG TCA GAG TCT CAG TCA GTG ACA TCA ACA AGA CAC AGA CAA GAC TCT G

Table C.4 Oligos designed constructing a new, flat P_{open} sequence hairpin construct, but were ultimately unsuccessful. 10T loading site on hairpin insert is colored red. 4T loop at the end of the hairpin is colored blue.

References

1. Pyle, A. M. Translocation and Unwinding Mechanisms of RNA and DNA Helicases. *Annu. Rev. Biophys.* **37,** 317–336 (2008).
2. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 helicases: Family matters. *Curr. Opin. Struct. Biol.* **20,** 313–324 (2010).
3. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and Mechanism of Helicases and Nucleic Acid Translocases. *Annu. Rev. Biochem.* **76,** 23–50 (2007).
4. Lohman, T. M., Tomko, E. J. & Wu, C. G. Non-hexameric DNA helicases and translocases: Mechanisms and regulation. *Nat. Rev. Mol. Cell Biol.* **9,** 391–401 (2008).
5. Egly, J. M. & Coin, F. A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor. *DNA Repair (Amst).* **10,** 714–721 (2011).
6. Holstege, F. C., van der Vliet, P. C. & Timmers, H. T. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* **15,** 1666–1677 (1996).
7. Dvir, A., Conaway, R. C. & Conaway, J. W. A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. *Proc. Natl. Acad. Sci.* **94,** 9006–9010 (1997).
8. Mitchell, J. R., Hoeijmakers, J. H. & Niedernhofer, L. J. Divide and conquer: nucleotide excision repair battles cancer and ageing. *Curr. Opin. Cell Biol.* **15,** 232–240 (2003).
9. Fan, L. *et al.* XPD Helicase Structures and Activities: Insights into the Cancer and Aging Phenotypes from XPD Mutations. *Cell* **133,** 789–800 (2008).
10. Zurita, M. & Cruz-Becerra, G. TFIIH: New discoveries regarding its mechanisms and impact on cancer treatment. *J. Cancer* **7,** 2258–2265 (2016).
11. Guo, C., Tang, T.-S. & Friedberg, E. C. SnapShot: nucleotide excision repair. *Cell* **140,** 754–4.e1 (2010).
12. Shuck, S. C., Short, E. a & Turchi, J. J. Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res.* **18,** 64–72 (2008).
13. Kuper, J. *et al.* In TFIIH, XPD Helicase Is Exclusively Devoted to DNA Repair. *PLoS Biol.* **12,** e1001954 (2014).

14. Fuss, J. O. & Tainer, J. a. XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. *DNA Repair (Amst).* **10,** 697–713 (2011).
15. White, M. F. Archaeal DNA repair: paradigms and puzzles. *Biochem. Soc. Trans.* **31,** 690–3 (2003).
16. Kelman, Z. & White, M. F. Archaeal DNA replication and repair. *Curr. Opin. Microbiol.* **8,** 669–76 (2005).
17. Farnell, D. Nucleotide Excision Repair in the Three Domains of Life. *WURJ Heal. Nat. Sci.* 1–6 (2011). doi:10.5206/wurjhns.2010-11.1
18. Pugh, R. A. *et al.* The iron-containing domain is essential in Rad3 helicases for coupling of ATP hydrolysis to DNA translocation and for targeting the helicase to the single-stranded DNA-double-stranded DNA junction. *J. Biol. Chem.* **283,** 1732–43 (2008).
19. Qi, Z., Pugh, R. A., Spies, M. & Chemla, Y. R. Sequence-dependent base pair stepping dynamics in XPD helicase unwinding. *Elife* **2,** e00334 (2013).
20. Roman, L. J., Eggleston, a K. & Kowalczykowski, S. C. Processivity of the DNA helicase activity of Escherichia coli recBCD enzyme. *J. Biol. Chem.* **267,** 4207–4214 (1992).
21. Lee, J. K. & Hurwitz, J. Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 54–9 (2001).
22. Iftode, C., Daniely, Y. & Borowiec, J. a. Replication protein A (RPA): the eukaryotic SSB. *Crit. Rev. Biochem. Mol. Biol.* **34,** 141–80 (1999).
23. Wold, M. S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66,** 61–92 (1997).
24. Haring, S. J., Mason, A. C., Binz, S. K. & Wold, M. S. Cellular Functions of Human RPA1: MULTIPLE ROLES OF DOMAINS IN REPLICATION, REPAIR, AND CHECKPOINTS. *J. Biol. Chem.* **283,** 19095–19111 (2008).
25. He, Z., Henricksen, L. A., Wold, M. S. & Ingles, C. J. RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature* **374,** 566–9 (1995).
26. Mocquet, V. *et al.* Sequential recruitment of the repair factors during NER: the role of XPG

in initiating the resynthesis step. *EMBO J.* **27,** 155–167 (2008).

27. Pugh, R. A. *et al.* Ferroplasma acidarmanus RPA2 facilitates efficient unwinding of forked DNA substrates by monomers of FacXPD helicase. *J. Mol. Biol.* **383,** 982–98 (2008).
28. Honda, M., Park, J., Pugh, R. A., Ha, T. & Spies, M. Single-Molecule Analysis Reveals Differential Effect of ssDNA-Binding Proteins on DNA Translocation by XPD Helicase. *Mol. Cell* **35,** 694–703 (2009).
29. Pugh, R. A., Honda, M. & Spies, M. Ensemble and single-molecule fluorescence-based assays to monitor DNA binding, translocation, and unwinding by iron-sulfur cluster containing helicases. *Methods* **51,** 313–21 (2010).
30. Nowak, M., Olszewski, M., Śpibida, M. & Kur, J. Characterization of single-stranded DNA-binding proteins from the psychrophilic bacteria Desulfotalea psychrophila, Flavobacterium psychrophilum, Psychrobacter arcticus, Psychrobacter cryohalolentis, Psychromonas ingrahamii, Psychroflexus torquis, and Pho. *BMC Microbiol.* **14,** 91 (2014).
31. Bartos, J. D., Willmott, L. J., Binz, S. K., Wold, M. S. & Bambara, R. A. Catalysis of strand annealing by replication protein A derives from its strand melting properties. *J. Biol. Chem.* **283,** 21758–21768 (2008).
32. Kemmerich, F. E. *et al.* Force regulated dynamics of RPA on a DNA fork. *Nucleic Acids Res.* **44,** 5837–5848 (2016).
33. Brosh, R. M. *et al.* Functional and physical interaction between WRN helicase and human replication protein A. *J. Biol. Chem.* **274,** 18341–18350 (1999).
34. Brosh, R. M. *et al.* Replication protein A physically interacts with the Bloom's syndrome protein and stimulates its helicase activity. *J. Biol. Chem.* **275,** 23500–23508 (2000).
35. Cui, S. *et al.* Analysis of the unwinding activity of the dimeric RECQ1 helicase in the presence of human replication protein A. *Nucleic Acids Res.* **32,** 2158–2170 (2004).
36. Sommers, J. a. *et al.* Novel function of the fanconi anemia group J or RECQ1 helicase to disrupt protein-DNA complexes in a replication protein A-stimulated manner. *J. Biol. Chem.* **289,** 19928–19941 (2014).
37. Gupta, R. *et al.* FANCJ (BACH1) helicase forms DNA damage inducible foci with replication protein A and interacts physically and functionally with the single-stranded DNA-binding

aprotein. *Blood* **110,** 2390–2398 (2007).

38. Suhasini, A. N. *et al.* FANCJ helicase uniquely sense oxidative base damage in either strand of duplex DNA and is stimulated by replication protein A. to unwind the damaged DNA substrate in a strand-specific manner. *J. Biol. Chem.* **284,** 18458–18470 (2009).
39. Neuman, K. C. & Block, S. M. Optical trapping. *Rev. Sci. Instrum.* **75,** 2787–809 (2004).
40. Comstock, M. J., Ha, T. & Chemla, Y. R. Ultrahigh-resolution optical trap with single-fluorophore sensitivity. *Nat. Methods* **8,** 335–40 (2011).
41. Mears, P. J., Koirala, S., Rao, C. V., Golding, I. & Chemla, Y. R. Escherichia coli swimming is robust against variations in flagellar number. *Elife* **2014,** 1–18 (2014).
42. Greber, B. J. *et al.* The cryo-electron microscopy structure of human transcription factor IIH. *Nature* **549,** 414–417 (2017).
43. Miranda Class. *Memory Alpha | Fandom powered by Wikia* Available at: http://memory-alpha.wikia.com/wiki/Miranda_class. (Accessed: 13th June 2017)
44. Tractor Beam. *Memory Alpha | Fandom powered by Wikia* Available at: http://memory-alpha.wikia.com/wiki/Tractor_beam. (Accessed: 13th June 2017)
45. Ashkin, A. Acceleration and Trapping of Particles by Radiation Pressure. *Phys. Rev. Lett.* **24,** 156–159 (1970).
46. Ashkin, A. & Dziedzic, J. M. Optical levitation by radiation pressure. *Appl. Phys. Lett.* **19,** 283–285 (1971).
47. Chu, S., Hollberg, L., Bjorkholm, J. E., Cable, A. & Ashkin, A. Physical review letters. *Phys. Rev. Lett.* **55,** 48–51 (1985).
48. Bloch, I. Ultracold quantum gases in optical lattices. *Nat. Phys.* **1,** 23–30 (2005).
49. Zhang, H. & Liu, K.-K. Optical tweezers for single cells. *J. R. Soc. Interface* **5,** 671–690 (2008).
50. Zhong, M.-C., Wei, X.-B., Zhou, J.-H., Wang, Z.-Q. & Li, Y.-M. Trapping red blood cells in living animals using optical tweezers. *Nat. Commun.* **4,** 1768 (2013).
51. Block, S. M., Goldstein, L. S. B. & Schnapp, B. J. Bead movement by single kinesin molecules studied with optical tweezers. *Nature* **348,** 348–352 (1990).
52. Comstock, M. J. *et al.* Direct observation of structure-function relationship in a nucleic acid-processing enzyme. *Science (80-.).* **348,** 352–354 (2015).

53. Chakraborty, A., Meng, C. A. & Block, S. M. in *Optical Tweezers. Methods in Molecular Biology* (ed. Arne Gennerich) **1486,** 391–409 (Humana Press, 2017).

54. Gittes, F. & Schmidt, C. F. Interference model for back-focal-plane displacement detection in optical tweezers. *Opt. Lett.* **23,** 7 (1998).

55. Moffitt, J. R., Chemla, Y. R., Izhaky, D. & Bustamante, C. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9006–9011 (2006).

56. Suksombat, S., Khafizov, R., Kozlov, A. G., Lohman, T. M. & Chemla, Y. R. Structural dynamics of E. Coli singlestranded DNA binding protein reveal DNA wrapping and unwrapping pathways. *Elife* **4,** 1–23 (2015).

57. Ngo, T. T. M., Zhang, Q., Zhou, R., Yodh, J. G. & Ha, T. Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. *Cell* **160,** 1135–1144 (2015).

58. Hegner, M., Smith, S. B. & Bustamante, C. Polymerization and mechanical properties of single RecA-DNA filaments. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 10109–10114 (1999).

59. Hodeib, S. *et al.* A mechanistic study of helicases with magnetic traps. *Protein Sci.* **26,** 1314–1336 (2017).

60. Harami, G. M. *et al.* Shuttling along DNA and directed processing of D-loops by RecQ helicase support quality control of homologous recombination. *Proc. Natl. Acad. Sci.* **114,** E466–E475 (2017).

61. Rutkauskas, M. *et al.* Directional R-loop formation by the CRISPR-cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep.* **10,** 1534–1543 (2015).

62. Charvin, G., Strick, T. R., Bensimon, D. & Croquette, V. Tracking Topoisomerase Activity at the Single-Molecule Level. *Annu. Rev. Biophys. Biomol. Struct.* **34,** 201–219 (2005).

63. Yogo, K. *et al.* Direct Observation of Strand Passage by DNA-Topoisomerase and Its Limited Processivity. *PLoS One* **7,** e34920 (2012).

64. Nugent-Glandorf, L. & Perkins, T. T. Measuring 0.1-nm motion in 1 ms in an optical microscope with differential back-focal-plane detection. *Opt. Lett.* **29,** 2611–2613 (2004).

65. Moffitt, J. R., Chemla, Y. R., Smith, S. B. & Bustamante, C. Recent advances in optical tweezers. *Annu. Rev. Biochem.* **77,** 205–28 (2008).

66. Lee, S. H. *et al.* Using fixed fiduciary markers for stage drift correction. *Opt. Express* **20,** 12177–83 (2012).

67. Gordon, E. I. A Review of Acoustooptical Deflection and Modulation Devices. *Proc. IEEE* **54,** 1391–1401 (1966).

68. 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,** 6190–5 (2006).

69. Whitley, K. D., Comstock, M. J. & Chemla, Y. R. in *Optical Tweezers. Methods in Molecular Biology* (ed. Gennerich, A.) **1486,** 183–256 (Humana Press, 2017).

70. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28,** 8759–8770 (1995).

71. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science (80-.).* **265,** 1599–1600 (1994).

72. Landry, M. P., McCall, P. M., Qi, Z. & Chemla, Y. R. Characterization of Photoactivated Singlet Oxygen Damage in Single-Molecule Optical Trap Experiments. *Biophys. J.* **97,** 2128–2136 (2009).

73. Freedman, D. & Diaconis, P. On the histogram as a density estimator:L 2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **57,** 453–476 (1981).

74. Sing, C. E., De La Cruz, M. O. & Marko, J. F. Multiple-binding-site mechanism explains concentration-dependent unbinding rates of DNA-binding proteins. *Nucleic Acids Res.* **42,** 3783–3791 (2014).

75. Lichtman, J. W. & Conchello, J. A. Fluorescence microscopy. *Nat. Methods* **2,** 910–919 (2005).

76. Kusumi, A., Tsunoyama, T. A., Hirosawa, K. M., Kasai, R. S. & Fujiwara, T. K. Tracking single molecules at work in living cells. *Nat. Chem. Biol.* **10,** 524–532 (2014).

77. Whitley, K. D., Comstock, M. J. & Chemla, Y. R. Elasticity of the transition state for oligonucleotide hybridization. *Nucleic Acids Res.* **45,** 547–555 (2017).

78. Ghoneim, M. & Spies, M. Direct Correlation of DNA Binding and Single Protein Domain Motion via Dual Illumination Fluorescence Microscopy. *Nano Lett.* **14,** 5920–5931 (2014).

79. Qi, Z. Direct Observation of Xpd Helicase Base-Pair Stepping and Regulation By Rpa2. (University of Illinois Urbana-Champaign, 2013).

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

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

82. Dale, R. E. & Eisinger, J. Intramolecular energy transfer and molecular conformation. *Proc. Nat. Acad. Sci.* **73,** 271–273 (1976).

83. McKinney, S. A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91,** 1941–51 (2006).

84. Kerssemakers, J. W. J. *et al.* Assembly dynamics of microtubules at molecular resolution. *Nature* **442,** 709–12 (2006).

85. Doherty, K. M. *et al.* Physical and functional mapping of the replication protein A interaction domain of the Werner and Bloom syndrome helicases. *J. Biol. Chem.* **280,** 29494–29505 (2005).

86. Betterton, M. D. & Jülicher, F. Opening of nucleic-acid double strands by helicases: Active versus passive opening. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **71,** 1–11 (2005).

87. Pincus, D. L., Chakrabarti, S. & Thirumalai, D. Helicase Processivity and Not the Unwinding Velocity Exhibits Universal Increase with Force. *Biophys. J.* **109,** 220–230 (2015).

88. Johnson, D. S., Bai, L., Smith, B. Y., Patel, S. S. & Wang, M. D. Single-Molecule Studies Reveal Dynamics of DNA Unwinding by the Ring-Shaped T7 Helicase. *Cell* **129,** 1299–1309 (2007).

89. Huguet, J. M. *et al.* Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15431–15436 (2010).

90. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31,** 3406–15 (2003).

91. Patel, S. S. & Picha, K. M. Structure and Function of Hexameric Helicases. *Annu. Rev. Biochem.* **69,** 651–697 (2000).

92. O'Donnell, M. E. & Li, H. The ring-shaped hexameric helicases that function at DNA replication forks. *Nat. Struct. Mol. Biol.* **25,** 122–130 (2018).

93. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G. & Murrayy, N. E. Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Res.* **42,** 3–19

(2014).

94. Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.* **43,** D298–D299 (2015).

95. Svoboda, K. & Block, M. Biological Applications of Optical Forces. *Annu. Rev. Biophys. Biomol. Struct.* **23,** 247–285 (1994).