SINGLE-MOLECULE STUDIES OF MECHANICAL AND HELICASE-CATALYZED DISRUPTION OF NUCLEIC ACID DUPLEXES

BY

KEVIN D. WHITLEY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biophysics and Computational Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Associate Professor Yann R. Chemla, Chair
Professor Oleksii Aksimentiev
Professor Paul R. Selvin
Professor Taekjip Ha, Johns Hopkins University

Abstract

Nucleic acids (*e.g.* DNA, RNA) are subjected to numerous twisting, bending, and stretching forces within cells, and enzymes process them in a variety of ways. The behavior of nucleic acids in response to applied forces and enzymatic activity is therefore necessary for a fundamental understanding of biology. Furthermore, a detailed knowledge of nucleic acids and the enzymes that process them has fueled advances in bio- and nano-technology. In this thesis, we focus on two main systems: the elastic behavior of ultrashort nucleic acids and the activity of *E. coli* UvrD helicase.

First, we use a hybrid instrument combining high-resolution optical tweezers with single-fluorophore sensitivity to observe the hybridization of ultrashort (<15 nt) DNA and RNA oligonucleotides under tension, one molecule at a time. We quantify the effect of tension on the rates of hybridization, and in doing so determine the elastic behavior of the transition state for the reaction. We then investigate the elasticity of the ultrashort oligonucleotides by observing the change in extension that takes place during hybridization. Our results enable us to produce a model describing the shear-induced fraying of base-pairs in a nucleic acid duplex.

We then use similar single-molecule techniques to characterize *E. coli* UvrD helicase. First, we investigate UvrD's stepping dynamics by directly observing individual motor steps of the protein. Then, we examine the factors influencing the ability of UvrD to switch between unzipping and re-zipping behaviors. Finally, we place UvrD in its biological context by observing the effect of its interactions with an accessory protein in DNA mismatch repair.

To my parents Linda and Scott
and my sister Susan

Acknowledgments

I would first like to thank Yann Chemla for being such an excellent advisor over the past six years. His guidance has molded me into the scientist that I am today. Although I came from a biochemistry background with little experience in physics, he taught me how to think like a physicist. Furthermore, his hands-on training in instrumentation helped me numerous times when I was aligning optics. I am always grateful for the time he spends carefully reading and editing my papers, and for the advice he gives during practice presentations.

I would also like to thank the members of my prelim/thesis committees for reading my documents and offering helpful suggestions: Prof. Oleksii Aksimentiev, Prof. Sua Myong (Johns Hopkins), Prof. Taekjip Ha (Johns Hopkins), and Prof. Paul Selvin.

I also thank Matt Comstock, who was a postdoc in our lab before taking a faculty position at Michigan State University. As a postdoc, he worked closely with me during my early years in the lab to show me how to perform experiments and analyze data. Even after moving to Michigan he was a valuable collaborator on nearly every project on which I worked.

Next I would like to thank all my other past and present labmates: Rustem Khafizov, Taejin Lance Min, Markita Landry, Zhi Qi, Patrick Mears, Kiran Girdhar, Sukrit Suksombat, Vishal Kottadiel, Barbara Stekas, Tanya Perlova, Isaac Li, Roshni Bano, Ruopei Feng, Alice Troitskaia, Monika Makurath, and Aniket Ravan. They have been wonderful colleagues and friends who have always been supportive. I will miss our conversations over lunch, nasty afternoon coffee, and colloquium cookies.

I also thank Prof. Timothy Lohman (Washington University School of Medicine) and his lab for their collaboration on the work in the latter half of this dissertation. Lohman lab provided us with all the protein I used in Chapters 4, 5, and 6, and our discussions continue to yield insight into these biological systems.

I would like to thank Cindy Dodds, the administrative coordinator for the biophysics program, for her help and advice in navigating the requirements of the program.

I also extend my thanks to Prof. Bob Clegg, who was the director of the biophysics program when I began. He went out of his way to help me locate a lab during my first year. Without him, I may never have had the opportunity to join Chemla lab. I also learned so much about optics from him by taking his Optical Spectroscopy course and being a TA for his Optical Microscopy course.

Next, I thank the Center for the Physics of Living Cells (CPLC) for such a great collaborative environment here in Illinois, and also for providing me with the opportunity to present my research at a conference in Germany (iPoLS 2014). I am also grateful to Susan Flanegin and Sandra Patterson for being such helpful administrators of the Center.

I also thank the funding agencies that supported my research assistantship. My work was supported by the National Science Foundation (NSF MCB 09-52442 CAREER to Y.R.C.) and the National Institutes of Health (R21 RR025341 and R01 GM120353, both to Y.R.C.).

I would also like to thank the members of the Cusanovich lab at the University of Arizona in which I did my undergraduate research: Prof. Michael Cusanovich, Dr. Terry Meyer, Dr. John Kyndt, and John Fitch. This lab was where I first began my research career, and the hands-on experience I received from John Kyndt and John Fitch was invaluable.

I also extend my thanks to the members of the Graduate Christian Fellowship (GCF) for being great friends and for fostering such helpful discussions about how to be a Christian that takes both faith and science seriously. I have benefited tremendously from the conversations I have had with my friends in this fellowship that will continue to serve me throughout my career.

Finally, I would like to thank my parents and sister for enabling my scientific curiosity throughout my life. I still remember going to the St. Louis Science Center with them many times when I was young and seeing (and being afraid of) the giant animated dinosaurs. My family has always been supportive of my career choice of scientific research, and has continued to support me throughout my seven years of grad school.

Table of Contents

Appendices

Chapter 1

Introduction

This thesis has two major parts: the first (**Chapters 2** and **3**) involves a detailed investigation of the behavior of ultrashort nucleic acids under an applied force, and the second (**Chapters 4**, **5**, and **6**) describes the characterization of *E. coli* UvrD helicase. Since each chapter has its own substantial background devoted to it, in this first chapter we provide a broad introduction to both separate parts, and the methods used throughout this thesis to study them.

1.1. Nucleic acid elasticity

Nucleic acids (NAs; *e.g.* DNA, RNA) are subjected to numerous twisting, bending, and stretching forces within living systems (2), and their elastic behavior is an important determinant in the function of many cellular processes (3-5). Researchers have recently exploited the way in which NAs stretch and rupture in response to force to design ever-more sophisticated NA-based nanostructures (6) and nanodevices (7,8). Quantifying the way in which NAs behave under applied force can shed some light on these cellular processes, as well as contribute to the development of more sophisticated NA-based nanotechnology. The elastic behaviors of long NA strands have been studied in increasing detail, yet those of short strands remain incompletely understood. In **Chapter 2** we will look at the role of NA elasticity in the hybridization reaction. Then in **Chapter 3** we will investigate the elasticity of NA strands at the shortest length scales, which we refer to as 'ultrashort'.

1.2. Dynamics of *E. coli* UvrD helicase

Helicases are vectorial enzymes that couple the chemical hydrolysis of nucleotide triphosphates (NTPs) to mechanical work of unzipping nucleic acids. They perform essential roles in cellular pathways involving nucleic acid processing, including DNA replication, repair, recombination, transcription, ribosome biogenesis, translation, and splicing (9). Due to their ubiquity, defects in some helicases are linked to cancer and other genetic diseases (10-12). In addition to their medical relevance, helicases are also appealing to study as model systems for more broadly understanding NTPase-coupled motor proteins (9,13).

Helicases are grouped into six superfamilies (SFs): two include non-hexameric helicases (SF1 and SF2) and four include hexameric ones (**Figure 1.1**) (SF3 through SF6) (1,14,15). The non-hexameric helicases comprise the largest two superfamilies, and display high structural similarity in their catalytic cores. Within each of these superfamilies there are smaller families, which share even higher structural homology. Each family is further classified by the directionality with which its members translocate along nucleic acids. For example, SF1A helicases like the UvrD/Rep family translocate 3’ to 5’, but SF1B helicases like the Pif1-like family translocate 5’ to 3’ (1).

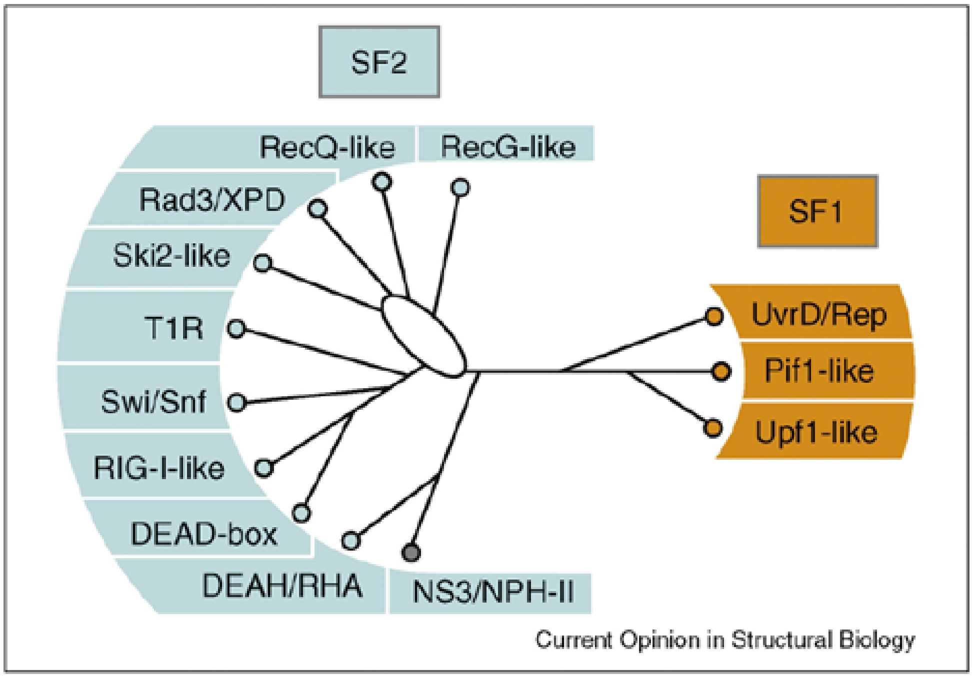


Figure 1.1: Superfamilies 1 and 2. Figure reproduced from Fairman-Williams et al. (2010) (1) with permission from Elsevier.

The focus of this thesis will be on the UvrD/Rep family, which encompasses the structurally-related proteins UvrD, Rep, PcrA, and Srs2 (1). These helicases are involved in a wide variety of cellular systems; *E. coli* UvrD alone is involved in methyl-directed mismatch repair (16), nucleotide excision repair (17), transcription-coupled repair (18), reversal of replication forks (19,20), replication of some plasmids (21), and removing RecA filaments from ssDNA to prevent unwanted recombination from occurring (22-24).

Many previous studies have demonstrated that helicases of the UvrD/Rep family exhibit a rich diversity of behaviors. Rep, PcrA, UvrD, and Srs2 have all been found to translocate along single-stranded (ss) DNA for a certain distance, and then snap back to their original positions, only to translocate along the DNA again (25-28). This behavior is termed ‘repetitive shuttling’ or ‘repetitive looping’, and was proposed to be necessary for clearing stalled DNA replication forks of toxic recombination intermediates. In addition to repetitive shuttling, UvrD has also been shown to transition from unzipping to re-zipping behavior by switching to the opposite strand of the duplex and translocating away from the fork junction (29).

E. coli UvrD is a prototypical representative of SF1 helicases that is used as a model system for understanding nucleic acid-processing motor proteins (13). In **Chapter 4** we will observe the individual motor steps of UvrD and provide a model to describe the mechanism by which it unzips DNA. Then in **Chapter 5** we investigate the ability of both UvrD and a related helicase to switch between unzipping and re-zipping DNA. Finally, in **Chapter 6** we place UvrD into its biological context by examining its interactions with an accessory protein involved in DNA repair.

1.3. Single-molecule methods to study biomolecules

1.3.1. Optical tweezers

Single-molecule force spectroscopy (SMFS) is a powerful technique for investigating biomolecular systems (30,31) such as protein folding (32,33), cytoskeletal motors (34-36), and nucleic acid-processing motors (37). The three most commonly-used SMFS techniques are atomic force microscopy (AFM), magnetic tweezers, and optical tweezers. Although each is interesting to discuss, we here limit our discussion to optical tweezers, which will be the primary technique used throughout this thesis. Below we describe the principles by which optical tweezers work, as well as common design aspects.

Optical tweezers (a.k.a. optical traps) function by trapping micron-sized dielectric objects (*e.g.* polystyrene beads) near the focus of a laser beam. It may not be obvious at first how a focused beam of light can trap something, but this becomes more intuitive if one considers the transfer of momentum from light to the object. Light carries momentum, and the direction of this momentum changes when the light interacts with an object (*e.g.* by reflection or refraction). By Newton's third law, this change in momentum (*i.e.* force, since $\boldsymbol{F} = d\boldsymbol{p}/dt$) must be accompanied by an equal and opposite change in momentum (force) by the object itself.

The force exerted on the object by incident light can be separated into two components: a scattering force and a gradient force (**Figure 1.2**). The scattering force is the most familiar to physicists, and results from both specular reflection and absorption at the surface of the object. This can be thought of as a 'photon fire hose' that pushes the object along the direction of light propagation (38) (**Figure 1.2a**). The gradient force is less familiar; it results from refraction as light passes through the object. The transfer of momentum in this case results in the object being pushed laterally toward the center of the intensity gradient (**Figure 1.2b**). However, in order to trap an object stably in three dimensions, a force in the axial direction must be applied to counteract the scattering force pushing the object along the direction of light propagation. This is accomplished by creating a steep intensity gradient by tightly focusing the beam of light using a high-numerical aperture objective lens (**Figure 1.2c**). In this case the axial component of the gradient force pulls the object back toward the focus of light, although due to the scattering force the object will rest at equilibrium somewhat downstream of the focus. Any displacement x of the object from this equilibrium position results in a Hookean restoring force $F = -\kappa x$, where κ is a spring constant proportional to the light intensity.

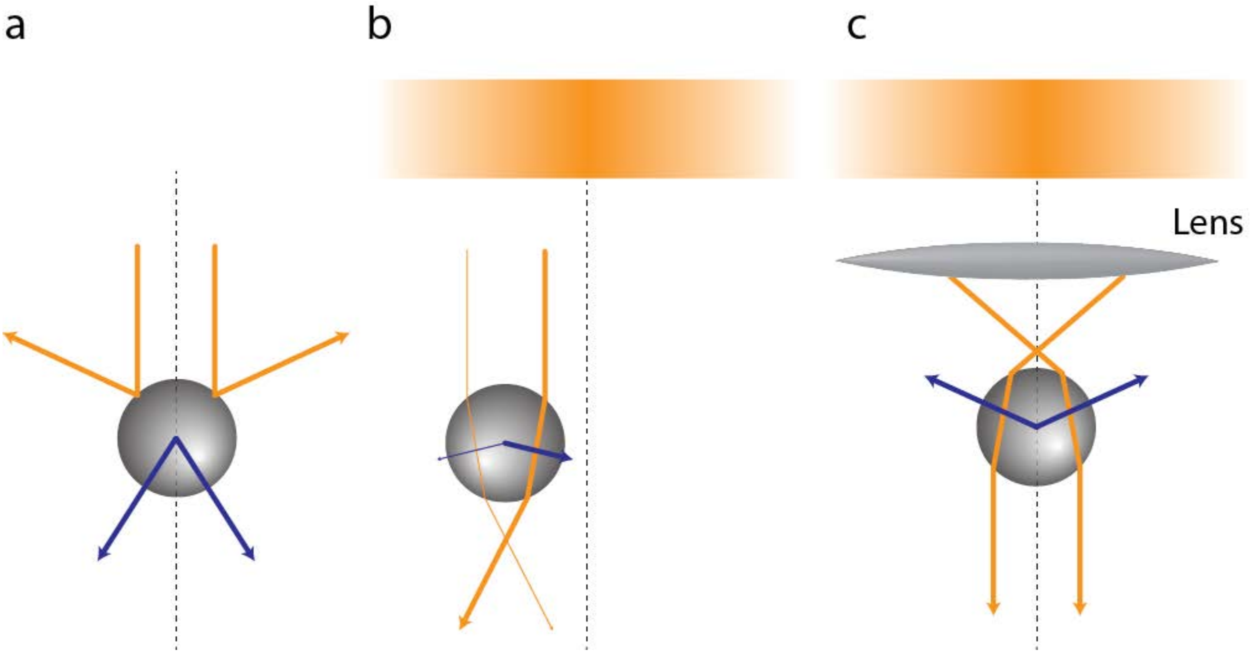


Figure 1.2: Ray approximation of optical trapping. Beads (*grey spheres*) experience forces (*blue arrows*) in response to momentum transfer from a laser light source (*orange arrows*). (**a**) The scattering force results from specular reflection off the bead surface, causing a force that pushes the bead further downstream. (**b**) A laser beam with a Gaussian intensity profile (*orange bar*) creates a radially-symmetric force that pulls the bead toward the center of the intensity gradient. (**c**) A high numerical-aperture lens focuses the light from a Gaussian beam to a tight spot, which exerts a force on the bead that pulls it toward the focus of the light. Since it is also subject to some scattering force pushing it forward, the bead rests at equilibrium slightly downstream of the laser focus.

Measuring the displacement of a trapped object from its equilibrium position can enable precise quantitative characterization of a system of interest. There exist several methods for measuring such displacement, such as video-based detection (39), direct object imaging (40), and back focal-plane interferometry (41,42). Back focal-plane interferometry relies on the interference pattern produced by the forward-scattered light of the object and the non-scattered light of the trap (**Figure 1.3**). Any displacement of the object from its equilibrium position creates a change in the interference pattern, which is imaged onto a position-sensitive photodetector. The detector readout can be converted into the displacement of the object from its equilibrium position (in nanometers) and the force on the object (in piconewtons) through a proper calibration. There exist many such calibration methods in the literature, but we use the standard method of recording the Brownian motion of a trapped object to acquire the conversion factors $\alpha_{x,y}$ between bead position and detector output (in nm/V) and the trap stiffnesses $\kappa_{x,y}$ (in pN/nm) (43).

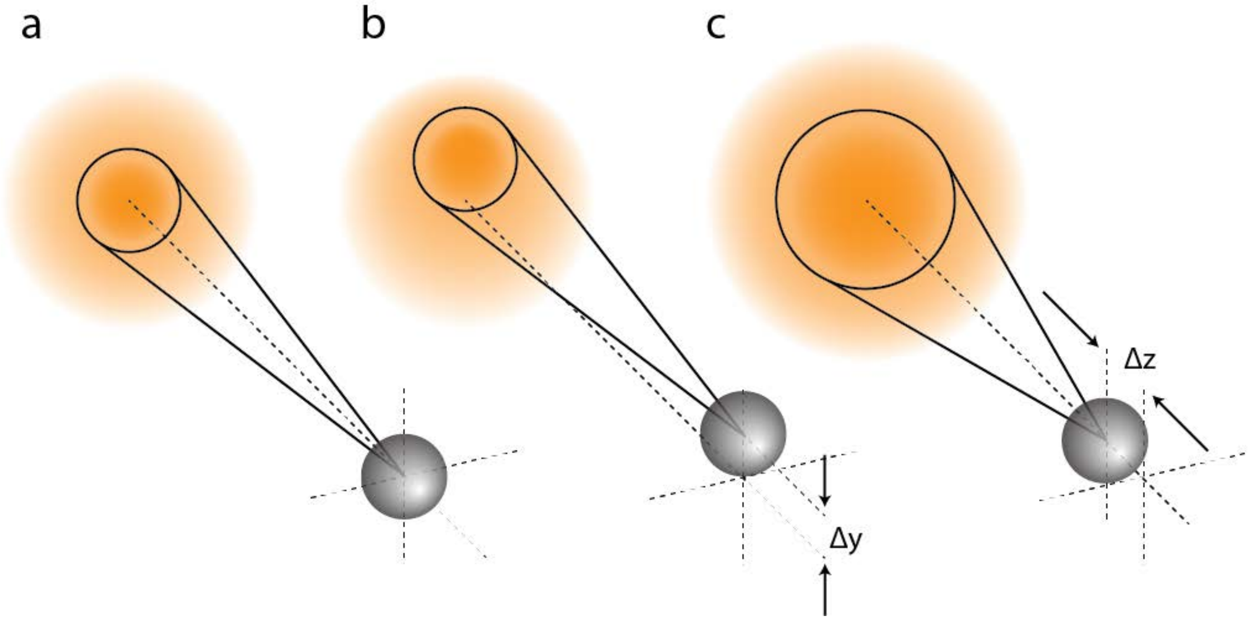


Figure 1.3: Detection of bead displacement by back focal-plane interferometry. Each panel depicts a trapped bead (*grey sphere*) and its interference pattern incident on a position-sensitive photodetector (*orange*) when the bead is (**a**) at its equilibrium position, (**b**) displaced laterally by Δy, and (**c**) displaced axially by Δz.

More recently, advances in instrument design have produced high-resolution optical tweezers that are capable of detecting sub-nanometer displacements. One of the factors contributing noise to optical tweezers systems is the random drift of the sample stage, and so some designs have focused on actively stabilizing this stage (44), while others have used dual-trap designs that enable trapping far away from drifting surfaces (45-47). In such dual-trap setups the two traps are formed from a single laser source, and so the two traps follow an almost identical optical path. Because of this, any noise coming from fluctuations in the beam path can be largely removed by differential detection of the two traps (**Figure 1.4**). Separating a single laser source into two traps is accomplished either by splitting the beam by polarization (45,47) or by 'time-sharing' the traps using an acousto-optic deflector (AOD) to deflect the beam rapidly between two positions (48). An acousto-optic deflector operates by producing a diffraction grating inside a crystal in response to an input radio frequency sound wave. In the polarization design, the position of one of the beams can be steered using a mirror mounted on a piezoelectric stage, and in the time-sharing design the beams can be steered by changing the angle at which the AOD deflects the incident light beam. Throughout this thesis we will perform experiments using both dual-trap designs: polarization-based and time-shared.

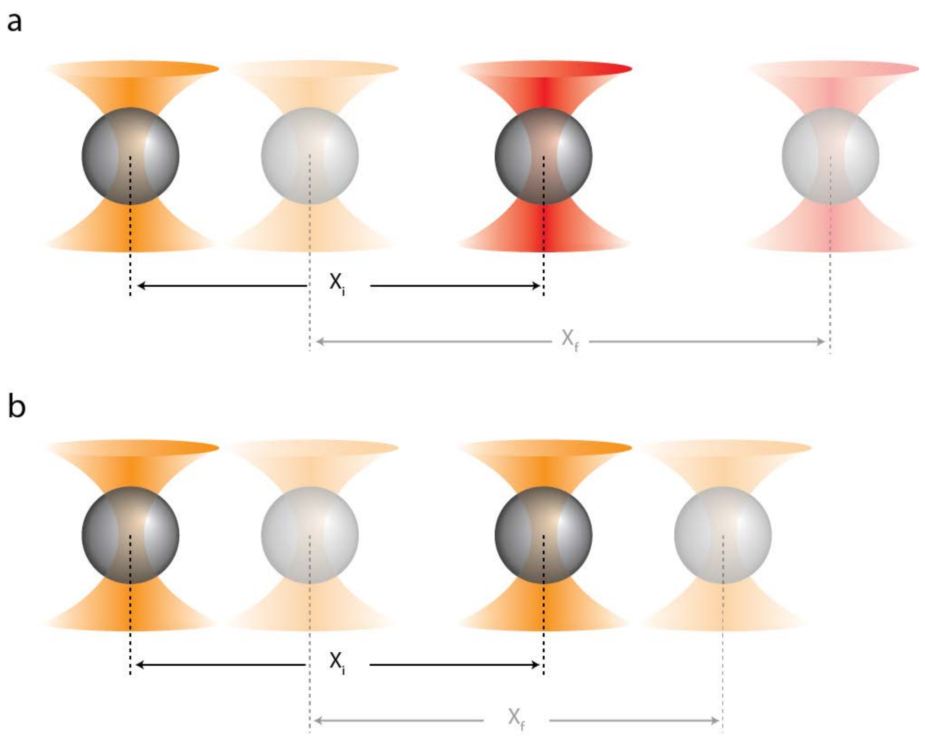


Figure 1.4: High-resolution optical tweezers through differential detections of dual traps. Schematics depicting beads (grey spheres) held in two separate dual-trap setups where the beam(s) forming the traps experience a shift due to fluctuations in the optical path. Final states (post-fluctuation) are depicted as transparent. (**a**) A dual-trap setup in which each trap is formed from a different laser. A fluctuation in the optical path of one or both beams causes a change in the trap separation ($X_i \neq X_f$). (**b**) A setup in which both traps are formed from the same laser. A fluctuation in the optical path causes an equal displacement of both traps, causing no change in trap separation ($X_i = X_f$).

1.3.2. Fluorescence microscopy

Single-molecule fluorescence assays have become the method of choice for investigating many biological systems due to their versatility and simplicity. A detailed knowledge of fluorophore photophysics and clever experimental design have enabled a wealth of various techniques to provide information on biological systems. Here for the sake of brevity we limit our discussion to two frequently-used techniques exploiting fluorophore photophysics that will be used later in this thesis: photobleaching and Förster resonance energy transfer (FRET).

The ability of fluorophores to permanently enter a dark, non-emitting state (termed ‘photobleaching’) is often a nuisance for performing fluorescence experiments, but has been employed for both single-molecule assays and super-high-resolution imaging (49-51). A common single-molecule assay that takes advantage of photobleaching is the ‘counting assay’, whereby the stoichiometry of a multi-subunit biomolecular complex is determined (**Figure 1.5**) (52). The assay is performed by first labeling each subunit with a single fluorophore, allowing the labeled subunits to assemble into their native complex, and then illuminating the complex with a fluorescence excitation laser. When each fluorophore photobleaches, a sudden drop in intensity appears in the fluorescence emission signal. After all fluorophores have photobleached, the number of subunits in the complex can be inferred from the number of down-steps in the fluorescence trace.

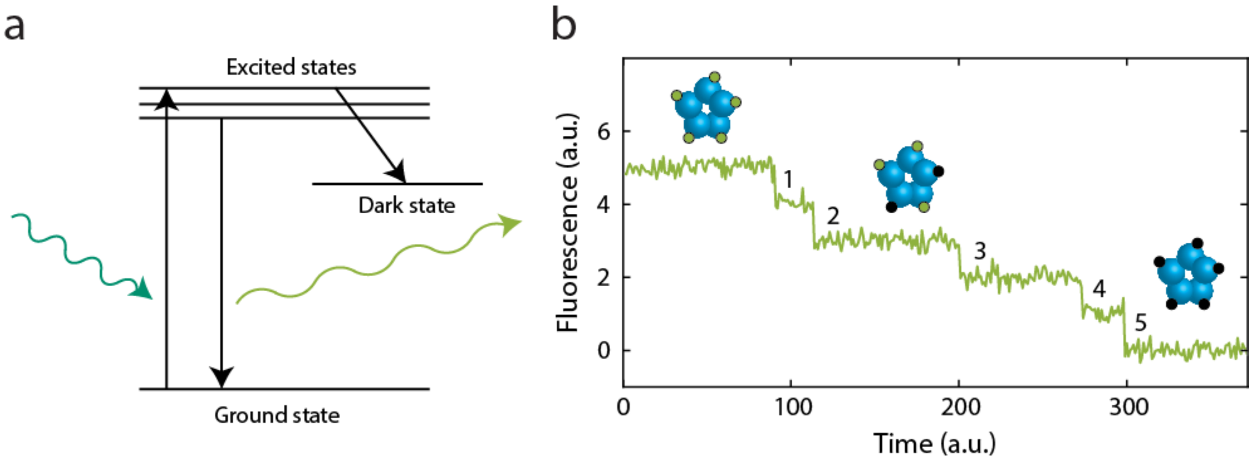


Figure 1.5: Demonstration of photobleaching to determine biomolecule stoichiometry. (**a**) Schematic Jablonski diagram depicting fluorescence and photobleaching. An incident photon (*teal*) excites a system from a ground state to an excited state. The system decays to a lower excited state due to interaction with the environment, then emits a photon of lower energy (*yellow-green*) as fluorescence. In some cases, the excited system can enter a permanently non-emitting dark state (photobleaching). (**b**) Simulation of photobleaching to determine biomolecule stoichiometry. A pentameric biomolecule (*blue spheres*) has a fluorophore (*green disks*) attached to each monomer. The system emits fluorescence in response to excitation by a laser source, and the fluorophores enter the dark, non-emitting state stochastically (photobleached fluorophores: *black disks*). When all fluorophores have photobleached, the number of down-steps reveals the number of monomers. Here only three of the six photobleached states are depicted schematically.

An excited fluorophore is also capable of transferring its energy to a second nearby fluorophore in a phenomenon known as Förster resonance energy transfer (FRET). FRET is used in an ever-increasing variety of single-molecule techniques, and has provided a vast wealth of information on biomolecular dynamics (53,54). The energy transfer phenomenon relies on the interaction of two fluorophores: one (the “donor”) absorbs a photon of light and enters an excited state, then transfers the energy via dipole-dipole interactions to the second fluorophore (the “acceptor”). The acceptor subsequently drops to its ground state, emitting a photon with a wavelength distinctly different from that of the donor (55,56). The efficiency of this energy transfer depends on the sixth power of distance between the fluorophores (**Figure 1.6**):

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \tag{1.1}$$

where r is the distance between the fluorophores, and R_0 is the Förster radius, which is the distance at which 50% of the energy is transferred. Since the efficiency of this energy transfer is distance-dependent, FRET can be thought of as a 'nanoscopic ruler' between two fluorophores. To investigate the intramolecular dynamics of a single molecule, these two fluorophores can be attached to two separate domains within the same biomolecule. The dynamics of the biomolecule can then be observed by the FRET efficiency, measured by the relative intensity of acceptor fluorescence emission to the total intensity from both fluorophores:

$$E = \frac{I_A}{I_A + I_D} \tag{1.2}$$

where I_A and I_D are the fluorescence emission intensities of the acceptor and donor molecules, respectively.

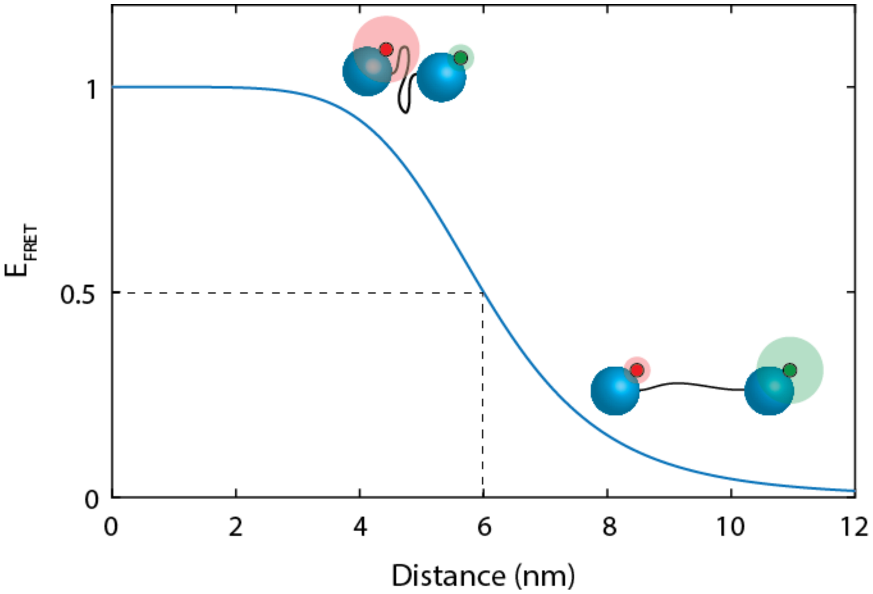


Figure 1.6: Depiction of FRET efficiency as a function of distance. Two domains of a biomolecule (*blue spheres*) are connected by a long flexible linker (*black*), and each is labeled with a different fluorophore (donor: *green disk*, acceptor: *red disk*). The donor fluorophore is directly excited by a laser, and transfers some of its energy to the acceptor in a distance-dependent process. When the domains are close together, the efficiency of energy transfer from the donor to acceptor fluorophore is high, and most emitted photons come from the acceptor. When the domains are far apart, the efficiency of energy transfer is low, and most emitted photons come from the donor. In this figure, R_o is assumed to be 60 Å.

*1.3.3. Fleezers: **fl**uorescence + optical tw**eezers***

Optical tweezers and fluorescence techniques are powerful techniques to probe biomolecular systems, yet each has significant limitations. Optical tweezers are ill-equipped to capture the three-dimensional dynamics of biomolecules, since they typically project all movement onto a single axis, along the direction of applied force (57). Fluorescence techniques do not allow biomolecules to be easily manipulated, and are not as well-suited to probing sparsely-populated states of a system's energy landscape, in contrast to force-based approaches.

It has long been recognized that a hybrid setup, combining fluorescence with optical tweezers (or 'fleezers'), could overcome the limitations of either separate technique, but significant technical challenges have prevented their implementation until recently. Early attempts to design such a setup revealed that fluorophores photobleach abnormally quickly when exposed to the combination of trap (near-IR) and fluorescence excitation (visible) lasers due to a two-photon absorption process (58). This two-photon process occurs when a fluorophore absorbs a fluorescence excitation photon followed by a near-IR optical trap photon, then drops to a dark, non-emitting state. One solution to this problem is to separate the trap and excitation lasers far enough to prevent two-photon absorption (59). A disadvantage to this approach, however, is that separating two traps by an adequate distance typically necessitates long, compliant molecules that serve as a poor transducers of mechanical signals, and thus result in poor spatial resolution. An additional solution to the two-photon absorption problem is to separate the trap and excitation lasers in *time* rather than *space* by rapidly interlacing the two light sources so that neither is on at the same time (60).

To design a *high-resolution* fleezers setup, this concept of interlacing trap and excitation light sources can be combined with a dual-trap design using the 'time-sharing' approach described above. In such a high-resolution setup, the interlacing cycle involves rapidly switching between three components: two traps and the fluorescence excitation (61-63) (**Figure 1.7**). This design enables simultaneous detection of single-molecule fluorescence and mechanical displacements with sub-nm resolution. We will primarily describe experiments using this instrument throughout this thesis.

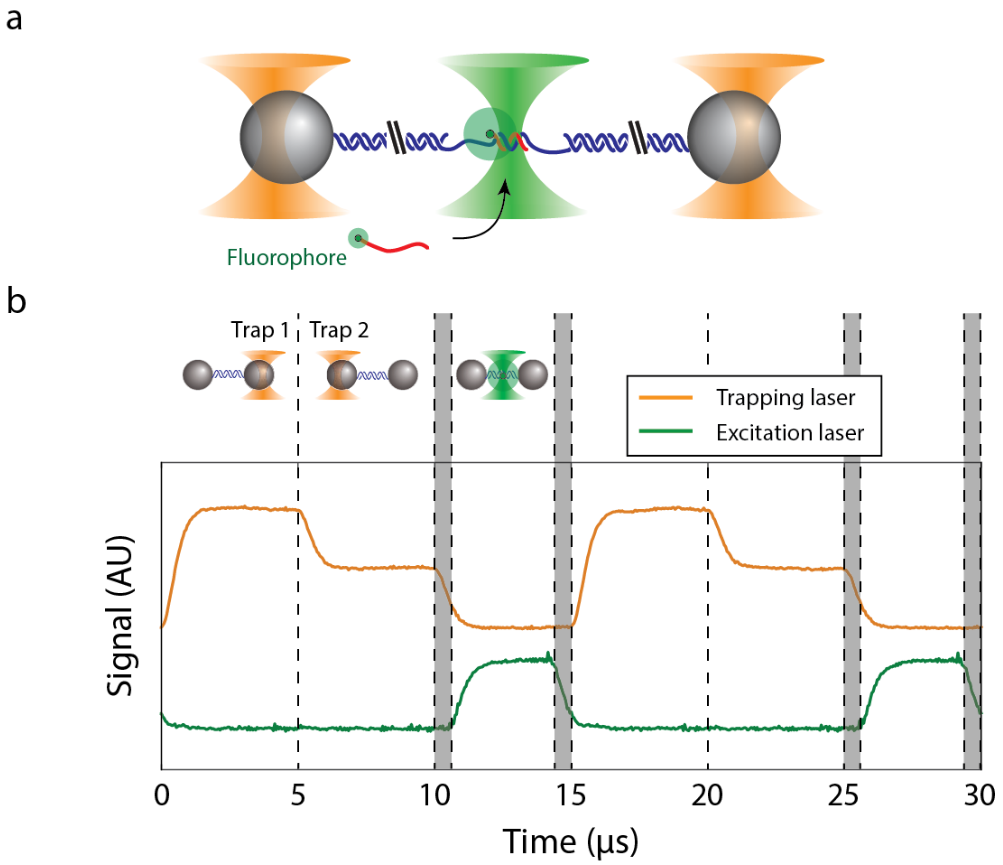


Figure 1.7: Schematic of high-resolution fleezers operation. (**a**) Schematic of an experiment using high-resolution fleezers setup. Two beads (*grey spheres*) are held in optical traps (*orange cones*) tethered together by a DNA molecule (*blue*). An oligonucleotide (*red*) with an attached fluorophore (*green disk*) binds to a complementary single-stranded region of the tethered DNA molecule, where the fluorophore is excited by a fluorescence excitation laser (*green cone*). (**b**) The interlacing cycle. The two traps (*orange*), and the fluorescence excitation laser (*green*) are switched on and off in 5 μs intervals (*black dotted lines*). A short time delay (625 ns; *grey shaded regions*) between the traps and the excitation laser is included to decrease the probability of two-photon photobleaching. In this figure, the two traps are set to different intensities for clarity.

Chapter 2

Elasticity of the transition state for oligonucleotide hybridization*

In this chapter we will examine the kinetics of the hybridization reaction under an applied force. We will use the fleezers instrument described in **Chapter 1** to observe individual oligonucleotides binding to and unbinding from a complementary DNA strand held under tension. We will quantify how the application of force affects melting and annealing rates, and from this determine the elastic properties of the transition state for the hybridization reaction. We conclude by constructing an energy landscape for the reaction and discuss the nature of the transition state.

---

*This work in this chapter has been published as:

Whitley, K. D., Comstock, M.J., and Chemla, Y.R. (2017). "Elasticity of the transition state for oligonucleotide hybridization." Nucleic Acids Research 45(2): 547-555.

2.1. Background

2.1.1. Nucleic acid hybridization in biology

The hybridization of nucleic acids (NA), in which two complementary strands base-pair to form a duplex, is one of the most fundamental processes in biology. It is prominent in a wide variety of cellular NA transactions such as replication, transcription, translation, homologous recombination, telomere extension, gene regulation, and cellular immunity. Since there are many examples, we limit this discussion to a few specific examples of the role of hybridization in gene regulation.

After (or during) transcription, mRNA is subject to a number of possible hybridization events that regulate expression of the genes encoded within it. In particular, small non-coding RNAs such as microRNAs (miRNAs) and small interfering RNAs (siRNAs) are used by cells to regulate the expression of many genes post-transcriptionally, including >60% of human protein-coding genes (64,65). The target search process of these molecules inside cells is highly dynamic, and the rates of binding and unbinding between two strands may be specific to the regulation of each gene (66). Many of these small RNA strands function as “guides” for Argonaute proteins to target mRNAs for repression. Studies have shown that base-pairing between the guide RNAs and the target mRNA requires complementarity between a “seed region” of ~6 nucleotides (nt) of the guides (67,68). This interaction is not limited only to RNA-RNA interactions; some prokaryotic Argonautes have also revealed guide DNA strands that target either RNA (69-71) or DNA (72).

2.1.2. Hybridization in biotechnology

The NA hybridization reaction has been exploited to develop a vast array of biotechnologies, including polymerase chain reaction (PCR) (73), DNA microarrays (74), and fluorescence *in situ* hybridization (FISH) (75). One of the most rapidly-emerging examples of this is the CRISPR-Cas9 system. The CRISPR system is the adaptive immune system of bacteria, with a complex of Cas proteins used to locate and destroy invading viral DNA (76). Cas9 is an endonuclease that binds a "guide" RNA that allows it to target a DNA sequence for cleavage. This has dramatically facilitated genome engineering, opening up myriad possibilities in biotechnology and medicine (77), and research continues to elucidate the mechanism by which Cas9 functions. In particular, many recent studies on this enzyme have investigated the kinetics of locating the target DNA sequence (78-80). Structural studies have also revealed a 10-nt "seed" sequence similar to that used by Argonaute proteins that is essential for base-pairing with the target DNA (81).

2.1.3. Hybridization in nanotechnology

NA hybridization has also found significant uses in nanotechnology. In the past ~25 years researchers have developed increasingly sophisticated NA-based nanodevices and nanomachines for a variety of functions. One example relevant to the present discussion is the tension gauge tether (TGT), a nanodevice designed to measure the forces exerted by cells on their environments by rupturing (*i.e.* melting) at a particular force (7). Another large set of examples are DNA nanomachines that use the energy of hybridization as "fuel" to perform useful work. These nanomachines have become increasingly sophisticated, from nanoscopic tweezers (82) to bipedal walkers (83-85) to molecular logic circuits (86,87). However, these machines all suffer from slow reaction rates brought about by the slow strand-displacement reaction, whereby an "invader" strand replaces an "incumbent" by first binding a short "toehold" and then competing with the incumbent strand for base-pairs until the incumbent is totally displaced (88). Computational studies have shown ways in which the reaction rate could be increased, for example by programming the energy landscapes of the strand-displacement reaction by strategically introducing mismatches (89) or by distorting the energy landscape of the incumbent strand by applying a shearing force (8).

2.1.4. Past studies of hybridization kinetics

The thermodynamics of the hybridization reaction have been worked out in great detail — to the point that melting temperatures can be predicted to within 2ºC (90-92) — but the kinetics are far less understood due to their dependence on infrequently-visited transitional structures. The transition between states has traditionally been viewed as a two-step process: a rate-limiting step whereby two single NA strands form a “nucleus” of a few adjacent base-pairs, and then a fast step resulting from base-by-base zipping of the strands to the final duplex state (93). The hybridization of short oligonucleotides is a convenient model for the hybridization reaction since these systems display two-state “all-or-none” behavior (94-96), allowing for more direct determination of elementary parameters (97).

The kinetics of any two-state reaction is governed by the transition state that is visited as the system moves from one state to the other. The identity of the transition state for hybridization is often thought of as a state in which a few base-pairs have formed (*i.e.* a “nucleated duplex”). However, several earlier studies suggested that there is likely a pre-equilibrium step prior to nucleation (95,96), which may come from the alignment of the two strands (98). Consistent with a pre-nucleation step, recent biochemical studies have shown that the transition state has no H-bonding and little to no base-stacking (99). Such a rate-limiting strand-alignment step is further supported by FCS measurements suggesting that only one out of every 100 to 1000 diffusional encounters between single strands results in duplex formation (100). Presumably, two complementary single strands must encounter one another in the correct orientation and register in order to form a successful duplex. The probability of two NA strands aligning to one another depends on their elastic properties, which can be investigated with single-molecule techniques.

Single-molecule force spectroscopy has proven to be a powerful tool to study transition states for biomolecular reactions such as NA hybridization. The application of force modulates reaction rates (101-104) and projects the entire reaction onto a well-defined coordinate (57,105), allowing researchers to obtain information on the distances from the equilibrium states to the transition state. This has been accomplished for DNA and RNA hairpin folding by pulling open the hairpins along a force axis orthogonal to the hairpins stems (102,106), and for DNA duplexes by pulling along the helical axis itself (101,107).

In the following study, we use the fleezers instrument described in **Section 1.3.3** to measure the rates at which short (≤12 nt) DNA or RNA oligonucleotides bind to and unbind from a complementary DNA strand held under constant tension (**Figure 2.1a**) (108). Our investigation of the hybridization reaction as a function of oligonucleotide length and force, over different ionic strengths, and for DNA-DNA vs. RNA-DNA duplexes paints a single, coherent picture for the energy landscape of the reaction. Using a generalized model for force-dependent rate constants, we find that the transition state is elastically similar to, though stiffer and more compact than, single-stranded NA. This finding suggests a mechanism in which the hybridizing oligonucleotides must adopt pre-ordered single-stranded structures for hybridization to proceed. We provide a simple model for the elasticity of the transition state that can be used to predict the kinetics of oligonucleotide hybridization under force.

2.2. Measuring the force-dependence of hybridization rate constants and equilibrium free energies

2.2.1. DNA constructs and probe design

All oligonucleotides used in this study were purchased from Integrated DNA Technology (IDT), and are listed in **Tables A.1** and **A.3**. The detailed protocol for synthesizing this DNA construct can be found in **Appendix A.1.1**. The DNA constructs in this study (**Figure 2.1a**) contained a ssDNA site of 9, 14, or 19 nt flanked by two long dsDNA "handles", and were made by ligating together three segments: left handle ("LH", 1.5 kb), insert ("Ins") and right handle ("RH", 1.7 kb). LH was synthesized from PCR amplification of the pBR322 plasmid (New England Biolabs) using a reverse primer containing biotin conjugated to the 5' end. The PCR product was digested with PspGI, leaving a 5' overhang. RH was similarly synthesized by PCR amplification of lambda phage DNA (NEB) using a reverse primer containing digoxigenin conjugated to the 5' end. This PCR product was then digested with TspRI (NEB), leaving a 3' overhang. RH and LH were then ligated to Ins, which contained a 5' phosphate, using T4 ligase (NEB). The insert consisted of a 9-nt binding site complementary to the 9-nt oligonucleotide probe used, flanked by $(dT)_5$ spacers on both sides ("2Sp insert"), one side ("1Sp insert") or neither side ("0Sp insert"). The DNA and RNA probes longer than 9 nt were extended with dA or A bases, respectively, to hybridize with the spacers of the 2Sp construct. A few exceptions were the alternate sequences ("seq2", "rnaSeq2") that bound to a unique insert ("seq2 insert"). Most probes used had a single Cy3 fluorophore conjugated to the 3' phosphate, with a few exceptions: one had a Cy3 fluorophore conjugated internally to a dT base ("9merIntCy3") to remove any steric hindrance with the handles upon binding the 0Sp construct, and two had no fluorophores at all to test the effect of the labels themselves ("9mer" and "rna10mer").

2.2.2. Trap + fluorescence assay

The rates of hybridization were measured using a high-resolution optical trap combined with a single-molecule confocal fluorescence microscope, as described previously (61). All data were collected in a buffer containing 10-30 nM probe, 100 mM Tris, 100 mM NaCl, 0, 2, or 20 mM $MgCl_2$, an oxygen scavenging system to increase both tether and fluorophore lifetimes (1% glucose, 1 mg/mL glucose oxidase (Sigma-Aldrich), 0.13 mg/mL catalase (EMD)) (55,109), and a triplet-state quencher to prevent fluorophore blinking (1 mg/ml Trolox (Sigma-Aldrich)) (110). Within a flow chamber containing this buffer (61,111), a streptavidin- and anti-digoxigenin-coated microsphere (Spherotech) were trapped and tethered together *in situ* by the DNA construct containing the single-stranded hybridization site for the probe under investigation (**Figure 2.1a**). To confirm the proper behavior of the DNA construct, a force-extension (F-x) curve was taken for each tether formed and fitted to the extensible worm-like chain (XWLC) model. The construct was held at a constant tension using active force-feedback (112). The binding/unbinding of probes was observed by both the fluorescence signal increase/decrease from the Cy3 fluorophores attached to the probes and the change in separation between the two traps as they were moved to maintain a constant tether tension (**Figure 2.1a**). For each tethered construct, data were collected at multiple forces to eliminate any artifacts arising from variations between individual tethered molecules.

2.2.3. Obtaining rate constants from fleezers assay

Fluorescence data were saved at 100 Hz, and events were located in fluorescence trajectories by using an appropriate threshold to identify the minima and maxima of the derivative of the fluorescence signal. Since the shorter probes had such short lifetimes, the sliding window over which the derivative was calculated to locate events was necessarily smaller for shorter probes. The sliding window was 0.15 s for 7-mers, 0.5 s for 8-mers, and 1 s for all other probe lengths. The lifetimes of each state (bound, τ_b; unbound, τ_u) were measured from the fluorescence signal as a function of tension on the tethered DNA strand and for different probe lengths (ℓ = 7-12 nt) (**Figure 2.1b**). Lifetimes measured by observing the change in molecular extension of the tethered DNA agreed within error of those measured by fluorescence.

For each condition used (probe, force, and buffer condition), a distribution of lifetimes was obtained. In order to avoid complications from histogram bin sizes when fitting the distributions, we instead calculated the survival probability of the state in question (bound or unbound) as

$$p(t) = 1 - \frac{1}{N}\int_0^t n(t')dt', \tag{2.1}$$

(**Figures 2.1c** and **2.1d**) where $p(t)$ is the probability that the probe is still in this state at time t, N is the total number of events, and $n(t')$ is the number of events in an interval dt'. For each probe length, the measured bound- and unbound-state lifetimes followed single exponential distributions (**Figures 2.1c** and **2.1d**, respectively), indicating a single rate-limiting step.

The survival probabilities were fitted to single exponential functions to obtain the first-order unbinding rate constants $k_{off} = \langle \tau_b \rangle^{-1}$ and the second-order binding rate constants $k_{on} = \left(\langle \tau_u \rangle [oligo]\right)^{-1}$, where $[oligo]$ is the concentration of oligonucleotide probe. Due to the finite size of the sliding window used for calculating the derivative of the fluorescence signal, there was a lower bound in the calculated survival probabilities (*i.e.* $p(t < t_{lb}) = 1$ where t_{lb} is the lower bound). To record the lifetime of an individual event, two extrema in the fluorescence signal derivative were located: one for binding and one for unbinding. Therefore, any events shorter than twice the sliding window are inaccurately located (*i.e.* t_{lb} = 2x sliding window). To account for this, we do not consider any events shorter than t_{lb}, and the distributions were fit to an exponential function shifted by t_{lb}:

$$p(t) = e^{-k(t - t_{lb})} \quad , \tag{2.2}$$

where k is the rate constant of the transition in question.

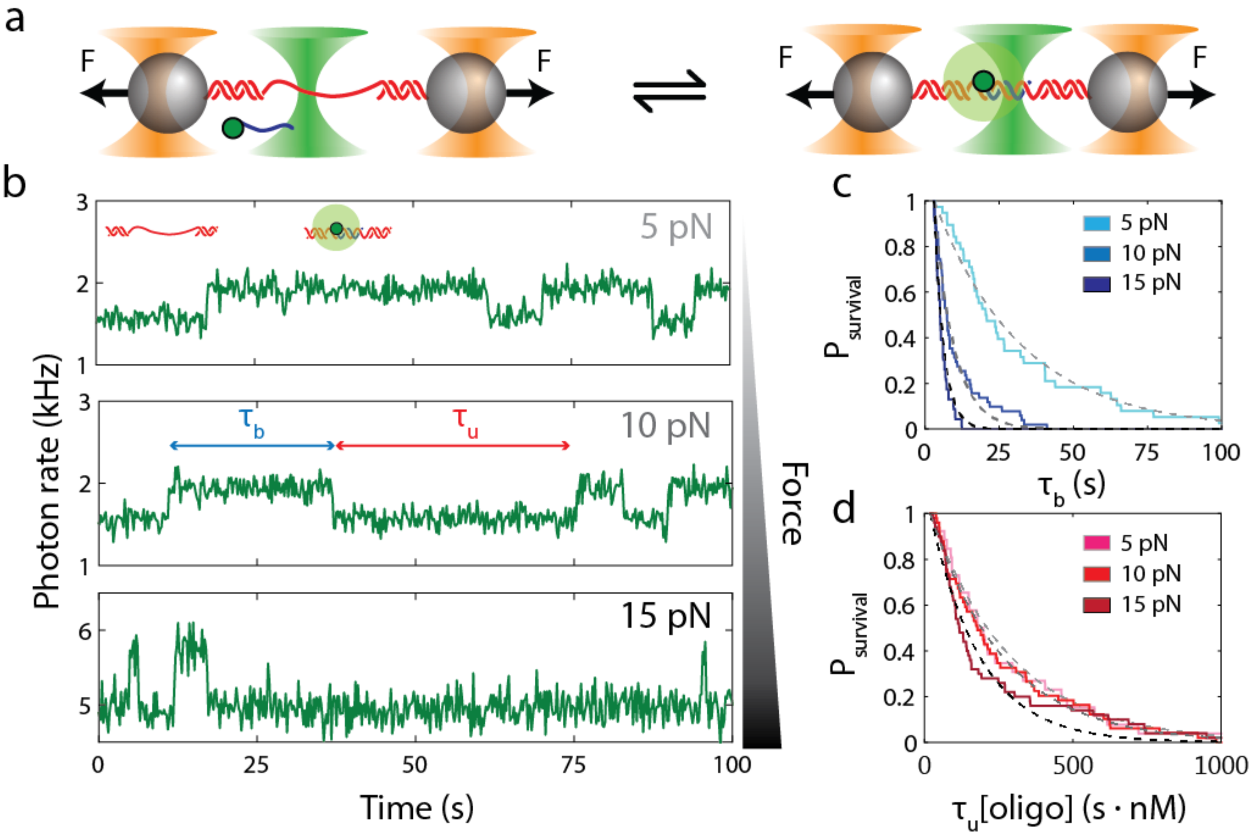


Figure 2.1: Measurement of single-oligonucleotide hybridization kinetics under force. (**a**) Schematic of the hybridization assay (not to scale). An engineered DNA molecule (*red*) containing a short, central ssDNA region flanked by long dsDNA handles is held under constant force by polystyrene beads (*grey spheres*) held in optical traps (*orange cones*). A fluorescence excitation laser (*green cone*) is focused on the central ssDNA region. Short oligonucleotides (*blue*) labeled with a Cy3 fluorophore at the 3' end (*green disk*) bind and unbind to the complementary ssDNA sequence in the center of the tethered DNA. The binding and unbinding is observed by the fluorescence emitted from the attached fluorophores. (**b**) Representative time traces showing 10-nt probes binding and unbinding a DNA construct held under three constant forces (5, 10, and 15 pN, ± 0.02 pN each). The lifetimes of the oligonucleotide bound states, τ_b, and the unbound states, τ_u, are measured from the increase and decrease of the fluorescence intensity. (**c, d**) The survival probabilities of the bound and unbound states over time for the three forces are shown. The probabilities are fitted to a single exponential function (dotted lines) for each force. Because the binding reaction displays second-order kinetics, the survival probabilities of the unbound states is plotted vs. $\tau_u[oligo]$.

2.2.4. Rate constants show a non-exponential dependence on force

Figures 2.2b and **2.2c** show the effect of tension on the unbinding and binding rates, respectively, with each color representing a different oligonucleotide length. Unbinding rates vary strongly as a function of length and force, whereas binding rates appear independent of length and show a weak but detectable dependence on force. According to Bell's model (113), the rate constants for transitions between bound and unbound states should depend exponentially on force, as $k \sim e^{F\Delta x^{\ddagger}/k_BT}$, where $\Delta x^{\ddagger}$ is the distance between the initial state and transition state along the mechanical reaction coordinate, k_B is the Boltzmann constant, and T is the absolute temperature. Since the unbinding rate constants depend much more strongly on force than those for binding, it follows that the transition state is "closer" to the unbound state than to the bound state. However, the rates do not strictly follow an exponential force dependence expected from Bell's model. The non-exponential behavior becomes apparent in the unbinding rates at lower forces (F = 1-8 pN; **Figure 2.2b**). A deviation from exponential force dependence indicates that the distance from the bound state to the transition state of the reaction $\Delta x^{\ddagger}$ along the pulling coordinate does not remain constant as force is applied (114,115). The need for a model allowing distances along the pulling coordinate to depend on force is also made clear by plotting the standard-state equilibrium free energy difference between bound and unbound states, obtained from the ratio of binding and unbinding rate constants, $\Delta G°(F) = -k_BT \cdot \ln\left(k_{off}/k_{on}\right)$, (with k_{on} in $M^{-1}s^{-1}$) vs. force (**Figure 2.2d**).

To test the generality of our findings, we repeated these measurements not only at high ionic strength (high Mg^{2+}) but also with an RNA hybridizing probe in place of DNA (**Figures 2.3** and **2.4**). The high ionic strength also allowed us to extend our measurements to $\ell = 7$ nt by reducing the rate of unbinding (**Figure 2.5**). The non-exponential behavior at lower forces is reproducibly observed in all conditions assayed.

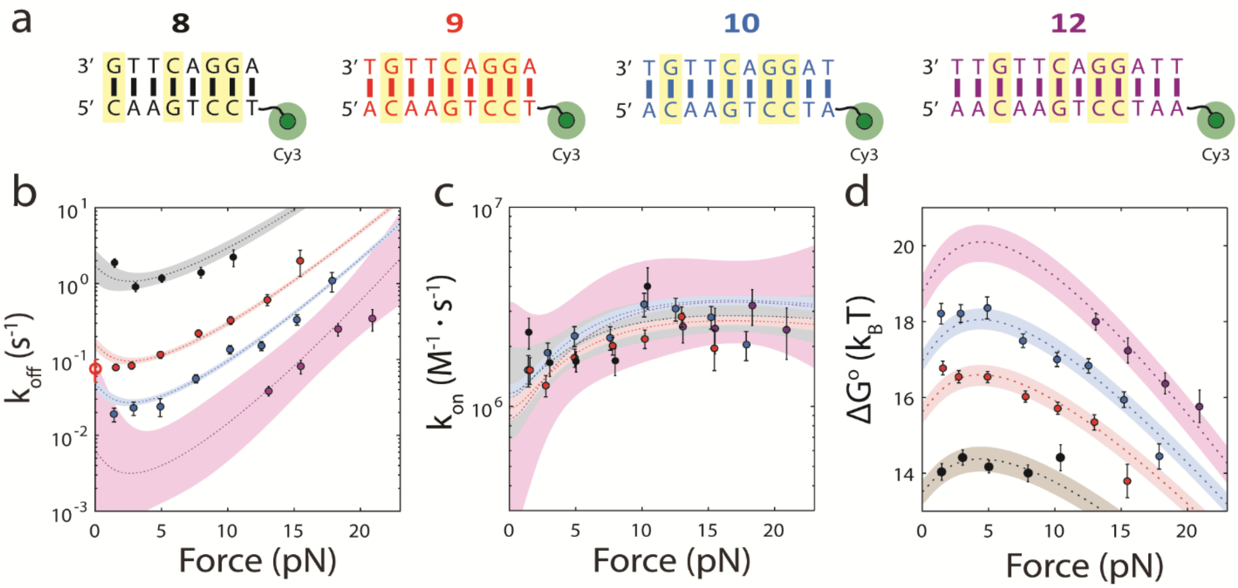


Figure 2.2: Force-dependence of oligonucleotide hybridization kinetics and thermodynamics. (**a**) The four oligonucleotides used, bound to their complementary sequences on the DNA construct (GC pairs highlighted). (**b**, **c**) Force-dependence of the unbinding (k_{off}) and binding (k_{on}) rate constants for each probe length. The dotted lines show the force-dependent model (Eqs. (2.3)-(2.5)) using parameters obtained from the globally fitted data (**Table 2.1**). Open circle: measured zero-force unbinding rate constant from (116). (**d**) Force-dependence of the equilibrium free energy ΔG^o between bound and unbound states for each probe length. The dotted lines show the force-dependent model (Eq. (2.6)) using parameters from the literature and those determined empirically (see **Appendix A.6**). Shaded regions represent 95% confidence intervals. Error bars represent s.e.m.

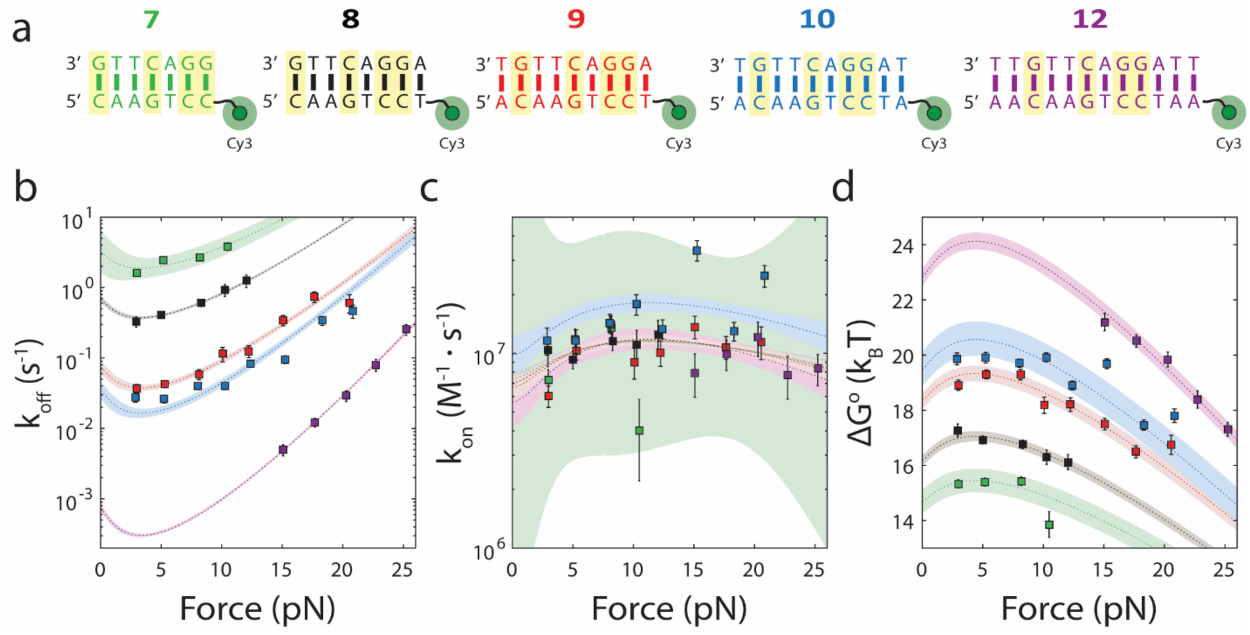


Figure 2.3: Force-dependent hybridization kinetics and thermodynamics of DNA at high [Mg^{2+}]. (**a**) The five DNA oligonucleotides used in this experiment (ℓ = 7, 8, 9, 10, 12 nt), bound to their complementary sequences on the

Figure 2.3 (continued): DNA construct (GC pairs highlighted). All experiments were performed in the same buffer as that described in **Section 2.2.2**, but with an additional 20 mM Mg^{2+}. (**b**, **c**) Force-dependent unbinding (k_{off}) and binding (k_{on}) rate constants for each probe length. The dotted lines show the force-dependent model (Eq. (2.3)-(2.5)) globally fitted to the data (fitted values in **Table 2.1**). (**d**) Force-dependence of the equilibrium free energy ΔG^{o} for each probe length. The dotted lines show fits to the force-dependent model (Eq. (2.6)). Shaded regions represent 95% confidence intervals. Error bars represent s.e.m.

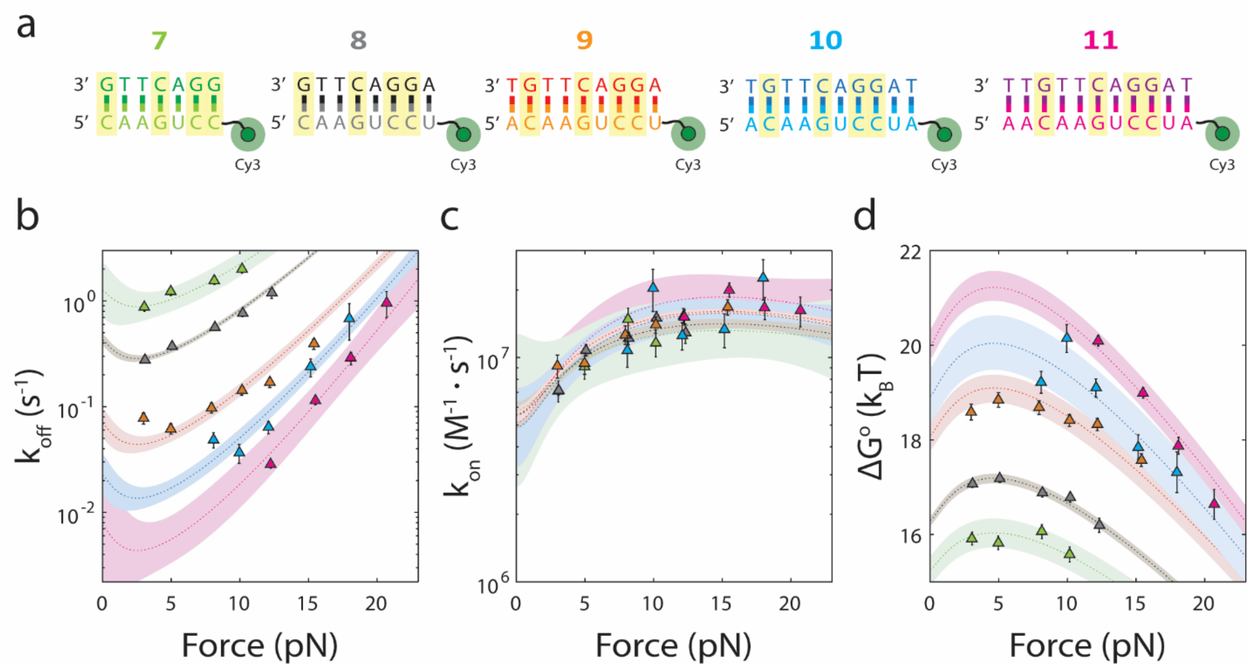


Figure 2.4: Force-dependence of RNA-DNA hybridization kinetics and thermodynamics. (**a**) The five RNA oligonucleotides used in this experiment (*light colors*; ℓ = 7-11 nt), bound to their complementary sequences on the DNA construct (*dark colors*; GC pairs highlighted). All experiments were performed in the same buffer as that described in **Section 2.2.2**, but with an additional 20 mM Mg^{2+}. (**b, c**) Force-dependent unbinding (k_{off}) and binding (k_{on}) rate constants for each probe length. The dotted lines show the force-dependent model (Eq. (2.3)-(2.5)) globally fitted to the data (fitted values in **Table 2.1**). (**d**) Force-dependence of the equilibrium free energy ΔG^{o} for each probe length. The dotted lines show the force-dependent model (Eq. (2.6)). Due to the uncertainty in the elastic parameters for RNA-DNA hybrids, these equilibrium free energy data were fitted to the persistence length and helix rise of the double-stranded state, yielding $P_{RNA\text{-}DNA} = 100 \pm 90$ nm and $h_{RNA\text{-}DNA} = 0.33 \pm 0.01$ nm/bp, respectively. Shaded regions represent 95% confidence intervals. Error bars represent s.e.m.

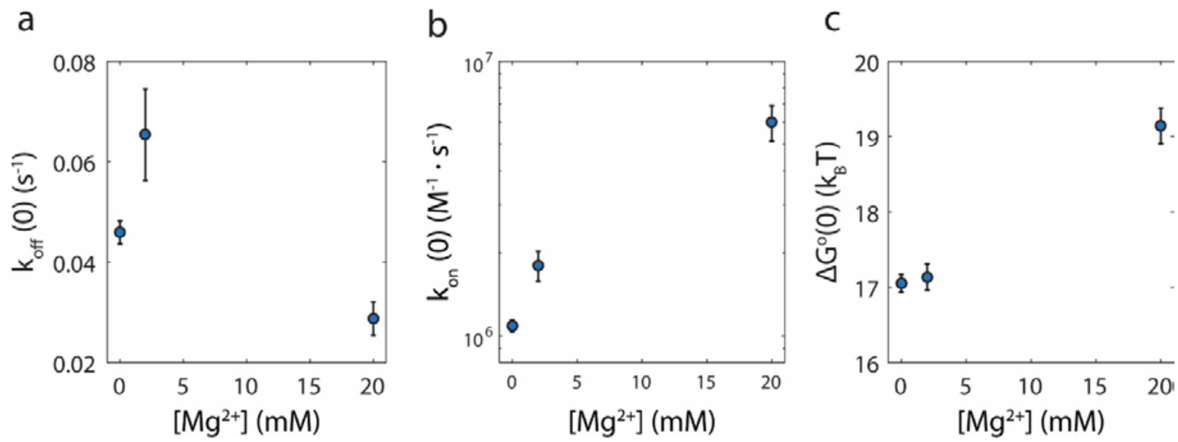


Figure 2.5: Effect of [Mg^{2+}]. (**a, b,** and **c**) [Mg^{2+}]-dependence of the rate constants and equilibrium free energies of the 10-nt probe under zero force, obtained by fitting the force-dependent data as done in Eq. (2.3)-(2.6). Error bars represent s.e.m.

2.2.5. Quantification of photobleaching

Dye photobleaching is one factor that could lead to inaccurate determination of rate constants, since it is indistinguishable from probe unbinding (although extension changes can also be used to identify unbinding events, these are difficult to detect at low forces (1-7 pN) where probe bound lifetimes are longest). To quantify the effect of photobleaching on our measured rate constants, we determined the mean number of photons detected before loss of fluorescence $\langle N_{photons} \rangle$ for each set of data, and then compared them to an experimentally determined estimate of the maximum total number of photons N_{tot} detected from these probes before photobleaching (**Figure 2.6**).

N_{tot} was measured using an experimental laminar flow sample chamber described previously (see **Appendix A.5**) (111). The chamber has two flow channels that do not mix (**Figure 2.6a**). One channel contained the longest probe used in this study (12 nt), while the other contained a blank buffer solution. A tether was first formed in the blank solution, then incubated in the top channel for ~15 s to load a 12-nt probe with the excitation laser remaining OFF. The tether was then moved back into the blank channel under low force (~3-5 pN), where the excitation laser was then turned ON. The tether was held in this channel until loss of fluorescence signal, and the cycle was repeated. The number of photons detected was calculated by integrating the signal (photon rate) over time, after subtracting the background. Under the experimental conditions (3-5 pN, $\ell = 12$ nt), the probe lifetime is expected to be >300 s, significantly longer than the measured fluorescence signal duration (mean = 103 s under our normal excitation intensity 5 μW). Thus, loss of fluorescence in these measurements likely results from photobleaching rather than probe unbinding. The mean number of photons emitted was determined to be $\langle N_{tot} \rangle = (40 \pm 7)\times 10^3$ photons. Since some unbinding events may also be mixed in to the data set, this value represents a lower bound on the total number of photons emitted until photobleaching. Regardless, all values of <$N_{photons}$> for our experimental data set fall well below this bound (**Figure 2.6c**), indicating that the effect of photobleaching is not significant across the range of conditions assayed.

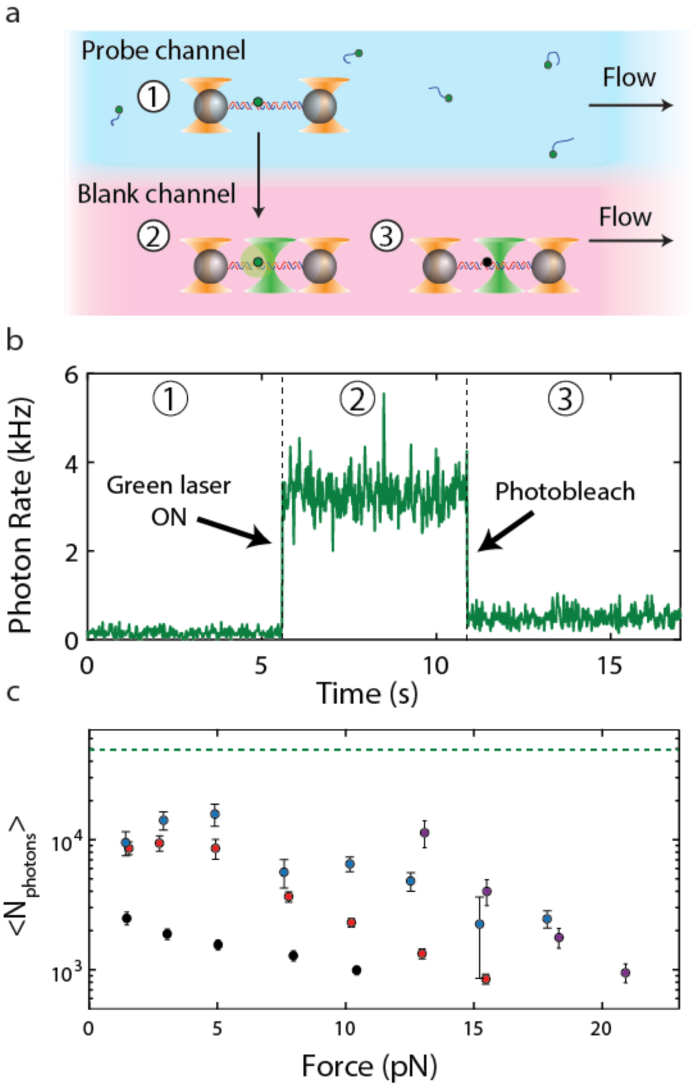


Figure 2.6: Effect of photobleaching. (**a**) Schematic depicting single-fluorophore photobleaching assay. The assay is performed in a laminar flow cell with two un-mixing streams, one containing blank buffer (*bottom*; *red*) and one containing 5 nM labeled probe (*top*; *blue*). (1) After forming a tether in the blank buffer, the construct is moved into the channel containing labeled probes, and incubated for ~15 s at low force (~5 pN) to allow a probe to bind. (2) The construct is then moved into the channel containing blank buffer, and the excitation laser is turned ON. (3) Loss of fluorescence occurs when the dye labeling the probe photobleaches. The mean number of photons detected before signal loss (from either dissociation or photobleaching), <$N_{photons}$>, gives a lower bound on the total number of photons detected before photobleaching, N_{tot}. (**b**) Representative time trace of the assay in (**a**). (**c**) <$N_{photons}$> calculated for all data shown in **Figure 2.2b**. Each color represents a different probe length (*black*: 8 nt, *red*: 9 nt, *blue*: 10 nt, *purple*: 12 nt). *Green dotted line*: lower bound on N_{tot} from experiment in (**a**).

2.2.6. Effect of fluorescent dye on kinetics

In order to determine the effect of the 3'-Cy3 label on the unbinding rates, we devised an experiment to compare directly the rates of labeled and unlabeled probes. To detect binding and unbinding of an unlabeled probe, we used the extension change of the DNA tether, relying on the fact that at most forces dsDNA and ssDNA have different end-to-end extensions. At forces $\gtrsim$7 pN the assay is sufficiently sensitive that changes in extension alone reliably identify binding and unbinding events. We mixed equal amounts of labeled (9merCy3; see **Table A.3**) and unlabeled (9mer; see **Table A.3**) probes together in a "half-and-half" experiment, so that the signals of both probes could be seen in the same time trace. An example data trace is shown in **Figure 2.7a**.

A different method was devised to locate events from extension time traces compared to that used to analyze fluorescence-based measurements. We performed a paired *t* test on the DNA extension data, comparing the sample distributions of two 3-s sliding time windows prior to and following each time point and calculating the p-value. Events were located by time points exceeding a p-value threshold set for each time trace. The threshold was validated from the detection of binding and unbinding events of labeled probes.

We performed this procedure for both the 9-nt DNA probe and the 10-nt RNA probe (**Figures 2.7b** and **c**). The 3'-Cy3 appeared to have either no effect on kinetics or a slightly stabilizing effect (compare labeled and unlabeled probes for the 10-nt RNA probe in **Figure 2.7c**). This may be consistent with studies showing that an attached 5'-Cy3 label has a stabilizing effect on the bound state through stacking interactions, although 3' dyes have weaker effects (117,118).

Furthermore, we investigated the effect of the dye linkage itself. For this we used a probe with a 9-carbon linker inserted between the 3' end of the oligonucleotide and the Cy3 dye ("9mer9linkCy3"; see **Table A.3**), and also a probe with a Cy3 dye attached internally to a dT base ("9merIntCy3"; see **Table A.3**). The internal label increased k_{off} nearly 10-fold (**Figure 2.7b**), which is consistent with studies showing that internally-attached Cy3 dyes are usually destabilizing (118).

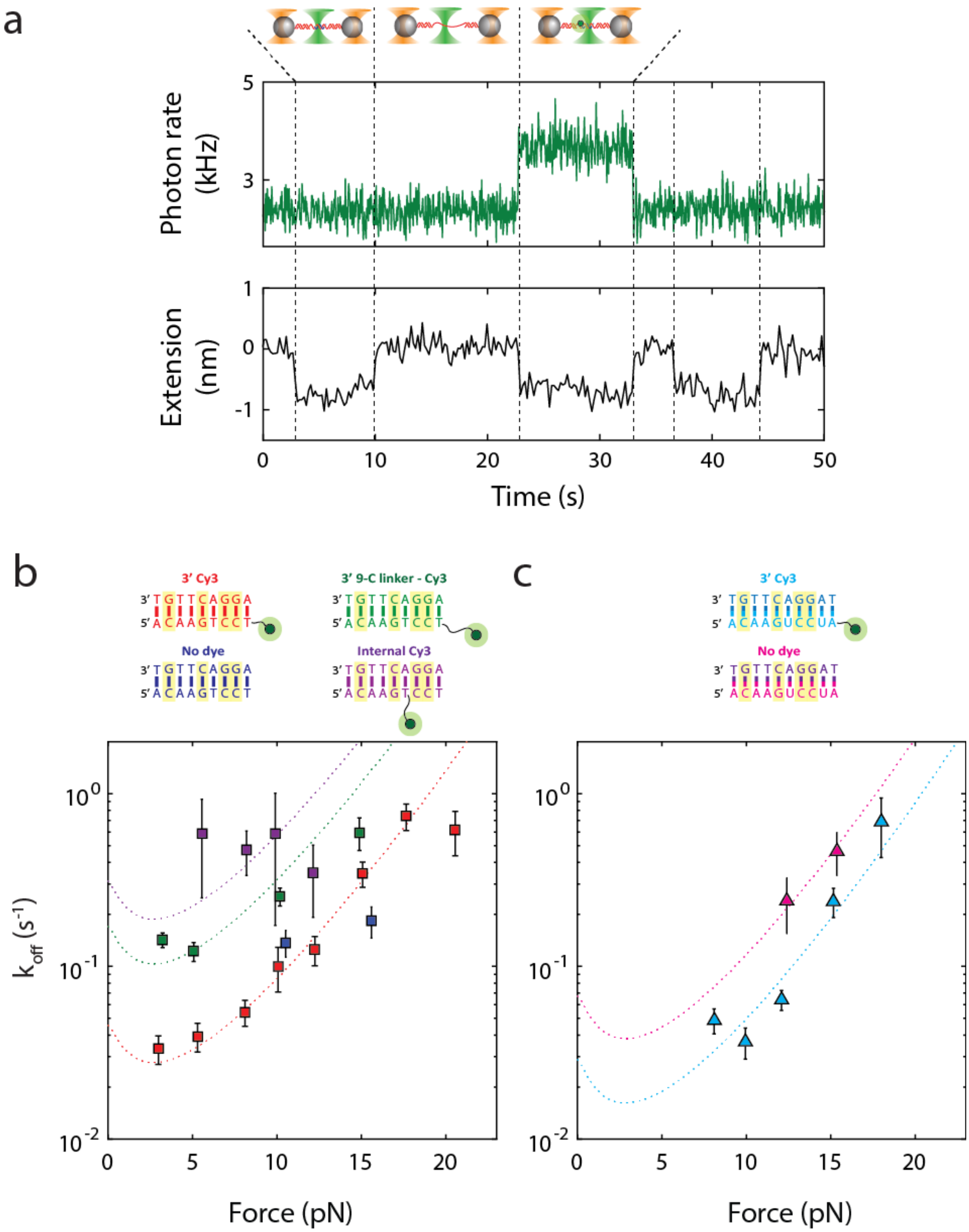


Figure 2.7: Effect of fluorescent dye. (**a**) Representative time trace depicting a "half-and-half" experiment, where equal concentrations (5 nM) of labeled (9merCy3; **Table A.3**) and unlabeled (9mer; **Table A.3**) probes were together in solution, and binding/unbinding events to a tether under 15 pN tension were detected by the change in extension of

Figure 2.7 (continued): the DNA construct upon probe hybridization. The first event (~3 s) shows an unlabeled probe binding, detected as a decrease in tether extension without any increase in fluorescence signal. The unlabeled probe unbinds (~10 s), resulting in an increase in tether extension. A second event (~22 s) shows a labeled probe binding, detected as a simultaneous increase in fluorescence signal and decrease in tether extension. Measurements were performed in 20 mM Mg^{2+} buffer. (**b**) Force-dependent unbinding rate constants (k_{off}) the 9-nt DNA probe dye variants in 20 mM Mg^{2+} buffer. *Top*: Probe dye variants where the position (or presence) of the Cy3 dye (*green disk*) is varied. (**c**) Force-dependent unbinding rate constants (k_{off}) the 10-nt RNA probe dye variants in 20 mM Mg^{2+} buffer. *Top*: RNA probe dye variants. Dotted lines represent fits of each probe to zero-force unbinding rate using Eq. (2.3)-(2.4) with values from **Table 2.1**. Error bars represent s.e.m.

2.3. Modeling the force-dependence of hybridization rate constants and free energies

2.3.1. Rate and equilibrium data fitting

To interpret the data, we adopted a generalized kinetic model in which distances are allowed to be force dependent. The force-dependent unbinding rate constant can be described by an Arrhenius-like equation:

$$k_{off}(F) = k_0 e^{-\Delta G^{\ddagger}(F)/k_B T} \tag{2.3}$$

where $\Delta G^{\ddagger}(F)$ is the force-dependent activation energy for unbinding, and k_0 is the attempt rate. $\Delta G^{\ddagger}(F)$ can be expressed as (115,119):

$$\Delta G^{\ddagger}(F) = \Delta G^{\ddagger}(0) - \ell \cdot \int_0^F \left(x^{\ddagger}(f) - x_b(f) \right) df \tag{2.4}$$

where ℓ is the length of the oligonucleotide, and $x^{\ddagger}(F)$ and $x_b(F)$ are the force-dependent extensions of the transition state and bound state per base pair, respectively. In the limit that the extension difference $\Delta x^{\ddagger} \equiv x^{\ddagger} - x_b$ is force-independent, Eq. (2.3) and (2.4) reduce to Bell's equation. Likewise, assuming a two-state reaction with a single barrier, the binding rate can be written as:

$$k_{on}(F) = k_0 e^{-(\Delta G^{\circ}(F) + \Delta G^{\ddagger}(F))/k_B T}, \tag{2.5}$$

in which $\Delta G^{\circ}(F)$ is given by

$$\Delta G^{\circ}(F) = \Delta G^{\circ}(0) - \ell \cdot \int_0^F x_u(f) - x_b(f) df, \tag{2.6}$$

where $x_u(F)$ is the force-dependent extension of the unbound state per base pair and $\Delta G^{\circ}(0)$ is the standard-state hybridization free energy at zero force.

2.3.2. Modeling the force-dependence of equilibrium free energies

Application of Eq. (2.3)-(2.6) to the data requires us to model the force dependences, *i.e.* elastic properties, of the bound, unbound, and transition states through the functions $x_b(F)$, $x_u(F)$, and $x^{\ddagger}(F)$, respectively. The bound and unbound states correspond to double- and single-stranded DNA (**Figure 2.1a**). For simplicity, we model $x_b(F)$ with the worm-like chain (WLC) model, using a persistence length $P_b = 53$ nm and helix rise $h_b = 3.4$ Å/bp, as reported for dsDNA (120,121). For $x_u(F)$, we determined the parameters empirically as described in **Appendix A.6** using the WLC and snake-like chain (SLC) models (122). Again for simplicity we here use the WLC model for $x_u(F)$, although using the SLC model provides very similar results. We validated our models for $x_b(F)$ and $x_u(F)$ against the force-dependent equilibrium free energy difference $\Delta G°(F)$. As shown in **Figure 2.2d**, the data are fitted well to Eq. (2.6) with no adjustable parameters other than the zero-force free energy for each probe length, $\Delta G°(0)$. Moreover, the values for $\Delta G°(0)$ are in very good agreement with those reported for the equilibrium hybridization free energies of the specific oligonucleotide sequences we used (**Figure 2.8**) (90,92). This good agreement also shows that the poly-dT spacers and dsDNA handles in our construct have a minimal effect on the oligonucleotide hybridization reaction.

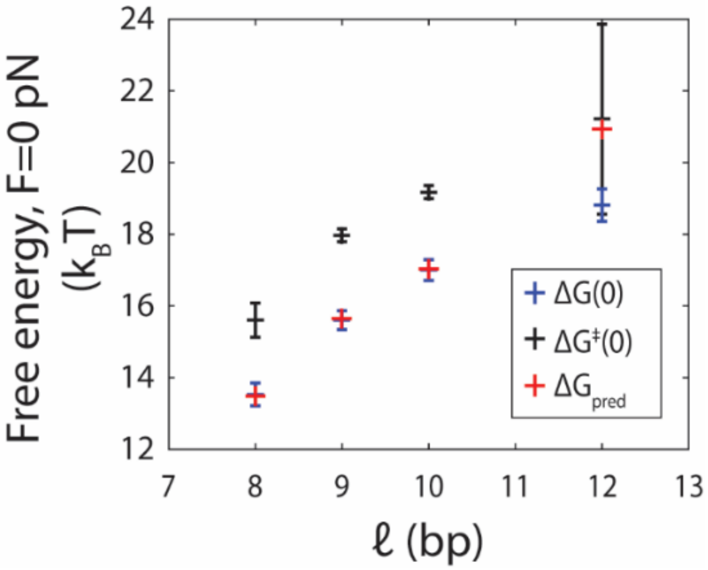


Figure 2.8: Comparison of measured zero-force free energies to predicted values. The zero-force equilibrium free energies $\Delta G(0)$ (*blue crosses*) are determined from fits to the data in **Figure 2.2d**. The zero-force activation free energies $\Delta G^{\ddagger}(0)$ are calculated from the zero-force unbinding rate constants (**Figure 2.2b**) according to the Arrhenius equation: $k_{off}(0) = k_0 e^{-\Delta G^{\ddagger}(0)/k_BT}$, where k_0 is the attempt frequency. We assume a value of $k_0 = 10^7$ s^{-1} based on literature values (96,123), and the calculated $\Delta G^{\ddagger}(0)$ are shown (*black crosses*). Experimentally determined free energies were compared to predicted values from literature (*red crosses*). We calculated the predicted free energies from nearest-

Figure 2.8 (continued): neighbor base pairing energies measured by Huguet et al. (92) and initiation energies determined by Santa Lucia (90), correcting for our experimental conditions (T = 22 °C, monovalent salt concentration $[\mathrm{Mon}^+] = [\mathrm{Na}^+] + [\mathrm{Tris}^+]$ = 150 mM). Error bars represent 95% c.i.

2.3.3. Modeling the force-dependence of rate constants

We next considered $x^{\ddagger}(F)$. A number of models have been proposed for the hybridization transition state (100,101,107), but none have been verified directly. In one candidate model (101), the transition state corresponds to a nucleated duplex where a few native base-pairs have formed. To test this nucleated duplex model, we fitted the unbinding and binding rates globally using a force-dependent extension in Eq. (2.3)-(2.5) in which n native base-pairs (dsDNA) are formed and $\ell - n$ remain unpaired (ssDNA) (**Figure 2.9**). In this case, the extension of the transition state would be a linear combination of the bound and unbound states:

$$X^{\ddagger}(F) = x_b(F)\cdot n + x_u(F)\cdot(\ell - n) \tag{2.7}$$

The global fitting was performed using a minimization routine that supplied "guesses" for n and all zero-force rates for the different probe lengths (for example, the data in **Figures 2.2b** and **2.2c** had four values for $k_{off}(0)$ and four values for $k_{on}(0)$ for the four probe lengths) that found the values minimizing the sum of squares of the residuals. In order to obtain error bars for these values, the minimization routine itself was resampled using jackknife resampling to produce distributions of values for each parameter. We found fits to the data in **Figures 2.2b** and **2.2c** to be poor (reduced chi-squared $\chi^2_{\mathrm{red}} = 6.1$), failing to reproduce the force dependence of the binding rates and yielding $n = 0$ native base pairs, corresponding to a purely ssDNA transition state, a model proposed by Ho et al. (107). We thus considered another model where the transition state behaves as a homogeneous polymer with unique elastic parameters. For simplicity, we assumed that the transition state extension scales linearly with oligonucleotide length over the narrow range investigated (8-12 nt). Globally fitting the binding and unbinding rate constant data (**Figures 2.2b** and **2.2c**) to this model yielded better fits (reduced chi-squared $\chi^2_{\mathrm{red}} = 2.9$) with values of $P^{\ddagger} = 2.6 \pm 0.2$ nm and $h^{\ddagger} = 0.54 \pm 0.01$ nm/nt (**Table 2.1**). Although these values are similar to those reported for ssDNA (106,124-126), they differ from those used to describe the unbound state in our data, indicating that the transition state is close in form to ssDNA, except stiffer (higher P) and more compact (lower h). This difference is reflected in the weak force-dependence of the binding rate data (**Figure 2.2c**).

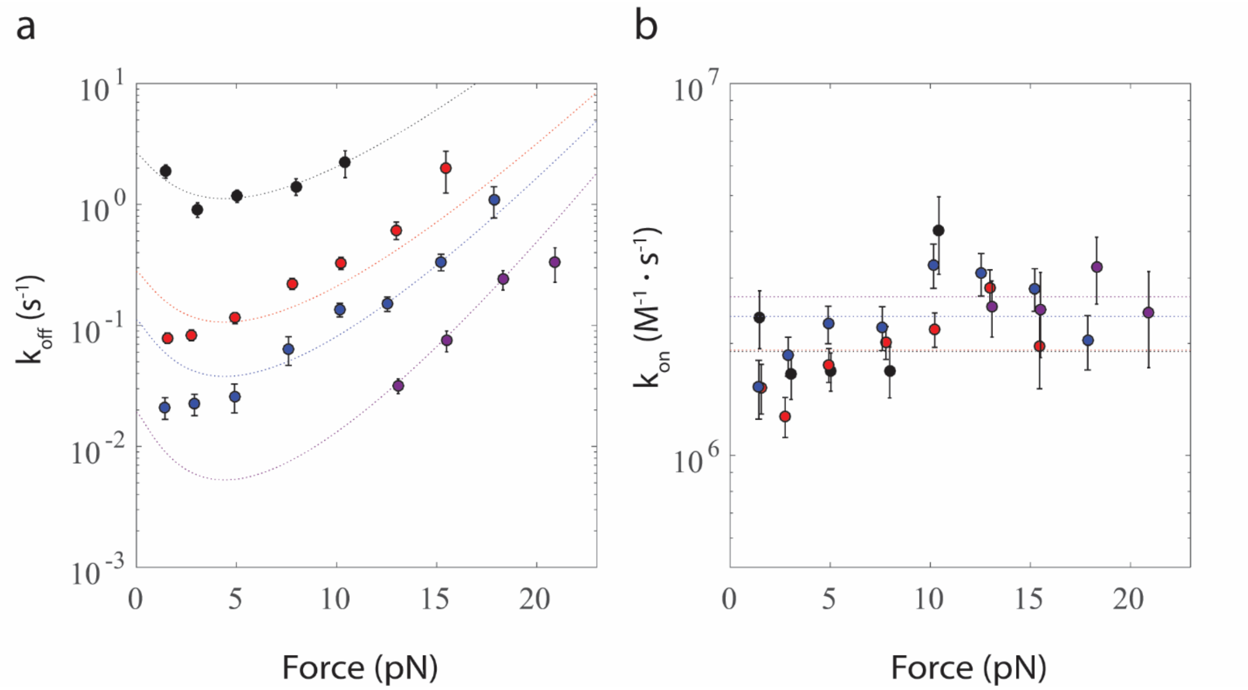


Figure 2.9: Rate constant data fitted to "nucleated duplex" model for transition state. (**a, b**) Force-dependent unbinding (k_{off}) and binding (k_{on}) rate constants for each probe length (*black*: 8 nt, *red*: 9 nt, *blue*: 10 nt, *purple*: 12 nt). The dotted lines show the force-dependent model (Eq. (2.3)-(2.5)) globally fitted to the number of native base-pairs formed in the transition state (Eq. (2.7)).

	DNA, 0 mM Mg^{2+}	DNA, 20 mM Mg^{2+}	RNA-DNA, 20 mM Mg^{2+}
$P^{\ddagger}$ (nm)	2.6 ± 0.2	1.92 ± 0.08	2.7 ± 0.5
$h^{\ddagger}$ (nm/nt)	0.54 ± 0.01	0.54 ± 0.01	0.53 ± 0.01
P_u (nm)	1.32 ± 0.07	1.07 ± 0.05	1.07 ± 0.05

Table 2.1. Obtained parameters from fitting of data.

To confirm the above model, we devised an alternate method of quantifying the transition state force dependence. As done previously for RNA hairpins (114,127), we solved Eq. (2.3)-(2.6) for the transition state extension to we obtained an expression to extract $x^{\ddagger}(F)$ directly from our unbinding rate constant data and the validated model for the bound state extension $x_b(F)$:

$$x^{\ddagger}(F) = \frac{k_B T}{\ell}\left(\frac{\partial \ln k_{off}}{\partial F}\right) + x_b(F)\,, \tag{2.8}$$

and a similar expression in terms of the binding rate k_{on} and unbound state extension $x_u(F)$. This exercise allows us to visualize directly the transition state force dependence and also validates the assumption made above that the transition state extension scales linearly with probe length over the range assayed (**Figure**

2.10a). **Figure 2.10b** shows the force-extension curve, $x^{\ddagger}(F)$, that results when scaling all transition state extensions by length and averaging them together. Although $x^{\ddagger}(F)$ lies close to ssDNA, it is better fit by the homogeneous polymer model described above and used in the global fits.

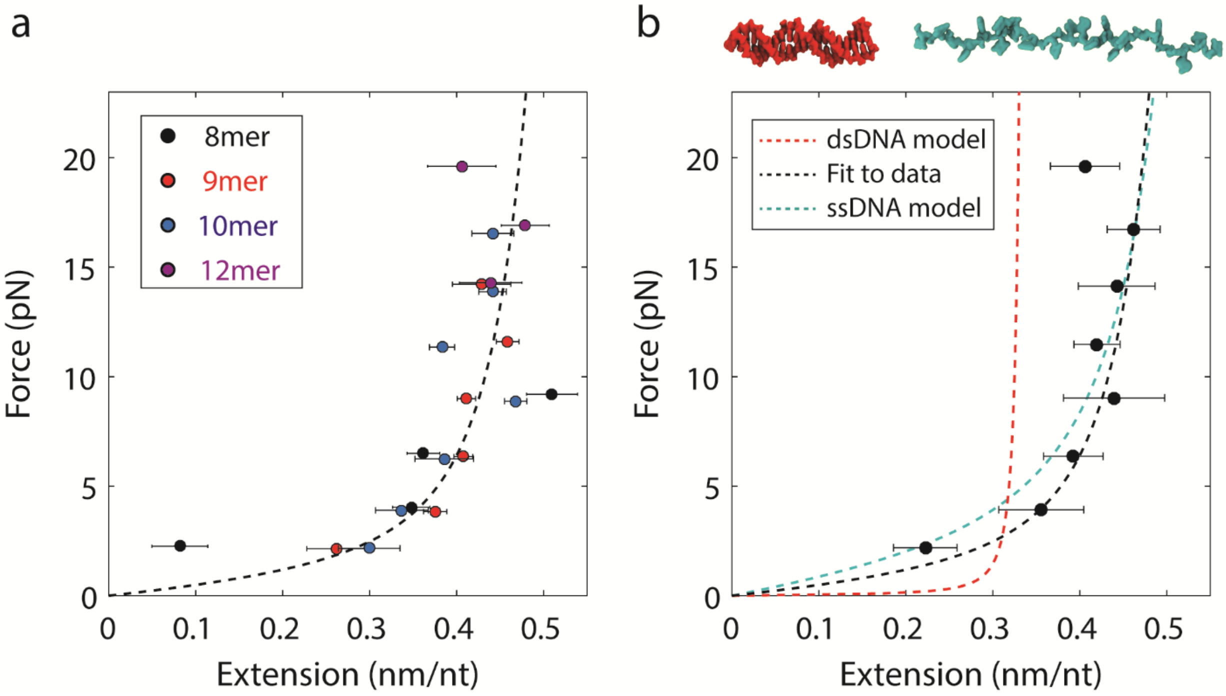


Figure 2.10: Force-extension curves of the transition state. (**a**) The end-to-end extension of the transition state for each probe (calculated from Eq. (2.8)) scaled by length, plotted vs. force. (**b**) The end-to-end extension of the transition state, as calculated from Eq. (2.8). The model for the transition state using the fitted parameters for $P^{\ddagger}$ and $h^{\ddagger}$ (*black dotted line*) is plotted alongside the models for dsDNA (*red dotted line*) and ssDNA (*cyan dotted line*) for comparison.

2.4. Elasticity of the transition state

2.4.1. Comparison to previous studies

The kinetics of any two-state reaction are determined by the transition state that is visited when passing from one state to the other. Here we have characterized the transition state for nucleic acid hybridization by measuring the kinetics of the reaction under force applied to the target strand. Across the range of conditions assayed, our analysis indicates a transition state that is similar to ssDNA. At first glance, our findings appear to stand in contrast to those in prior studies. Strunz et al. (101) extracted the rates of oligonucleotide melting from force-ramp measurements where force was applied across opposite 5’ ends of short duplexes. Fitting the force-dependent melting rates to Bell’s equation, they determined that the distance from the bound state to the transition state, $\Delta x^{\ddagger}$, was a constant ~1 Å per bp of the duplex. However, as discussed above, Bell’s model can be inappropriate in many cases despite its common use. Our binding and unbinding rates show clear non-exponential behavior over our force range (**Figures 2.2b** and **2.2c**), which indicates that $\Delta x^{\ddagger}$ is not constant. If we consider only forces >10 pN such as those assayed in Strunz et al., our measured unbinding rate constants display approximate single-exponential behavior (**Figure 2.2b**). Fits of the high-force data to Bell's equation provides us with a linear relation of $\Delta x^{\ddagger} = 1.4 \pm 0.2$ Å/bp, similar to that of Strunz et al. and in reasonable agreement with our prior estimate also using Bell's model (61). Our generalized model, which allows for force-dependent distances to the transition state, is demonstrably better at fitting the data over the entire range of forces assayed, but remains consistent with the results of Strunz et al. in the high force range. Nevertheless, our findings illustrate the caution needed in interpreting parameters extracted from Bell's model.

Our model approximates Bell's equation in the high force range and captures the deviations from exponential force dependence in the rates k_{off} and k_{on} observed at lower forces (<10 pN). One question our results raise is how the rates behave at and near zero force. Our model predicts a "roll-over" in k_{off} below ~1.5 pN (**Figures 2.2b**, **2.3b**, and **2.4b**), suggesting that low forces would stabilize the DNA duplex, a result that has been discussed theoretically (107,128). The lack of data below ~1.5 pN in our experiments makes it difficult to determine this behavior directly. Single-molecule FRET measurements of hybridization performed by Cisse et al. (116) at zero force using an identical sequence to the 9-nt probe used here and under similar experimental conditions yield a k_{off} ranging between 0.05 to 0.1 s^{-1}, which compares well to our 9-nt probe data (Open red circle in **Figure 2.2b**). However, the FRET result suggests that the force-dependence of the unbinding rate constants vanishes below 1.5 pN, in disagreement with our model. It is possible that the hybridization reaction coordinate no longer projects well onto the pulling coordinate (*i.e.* the end-to-end extension) at these forces, resulting in force-independent rates (57). Assays more sensitive to this low-force range will be necessary to quantify this behavior accurately.

2.4.2. Structure of the transition state

An important question is the nature of the transition state. As shown in our analysis above, our data are best described by a simple homogeneous worm-like polymer. Some studies have proposed that the hybridizing strands must first be "pre-aligned" or "pre-ordered" prior to base pairing (**Figure 2.11a**) (98,100,129). Such "pre-ordering" may be reflected in the persistence length for the transition state, which is larger than that of the unbound state. This means that the transition state is slightly stiffer than the unbound state. We speculate that some amount of single-stranded base-stacking or electrostatic stiffening from the increased number of backbone phosphate charges as two strands approach one another could lead to an increased stiffness (130).

This "pre-aligning" mechanism is supported by the observation that annealing is not diffusion-limited (129) and that only one in 100-1000 diffusional encounters successfully produces a duplex (100). The binding rates we measure, extrapolated to zero force, range between 10^{6}–10^{7} $M^{-1}\cdot s^{-1}$ (across the range of

Mg^{2+} concentrations assayed; **Figure 2.5**), consistent with this picture (**Figure 2.2c**). Presumably, the low success rate arises because the strands must be in the correct register and alignment with respect to one another to anneal successfully. Although this configuration is entropically unfavorable, we can use a simple, polymer elasticity-based to estimate the probability that a 5-13 nt oligonucleotide in solution is extended and aligned with the strand under tension, (**Appendix B.1**).

Irrespective of the detailed transition state structure, it is important to note that our assumption that the transition state extension scales linearly with probe length can only hold over short lengths, since it is unreasonable to expect long strands (>>10 nt) to pre-order before annealing. We expect any structural properties unique to the transition state to be confined to a short DNA segment comparable to the longest probe length (12 nt) (**Figure 2.11a**). We can use our simple estimate of the probability that a short ssDNA length is aligned with the strand under tension above to place a bound on the maximum length that can be pre-aligned. Based on this model, probabilities that a segment $\gtrsim$13 nt is pre-aligned to the tethered strand must be $\lesssim 10^{-3}$. This provides evidence that the unique structural properties of the transition state we report are confined to lengths within the range of those used in this study (8-12 nt).

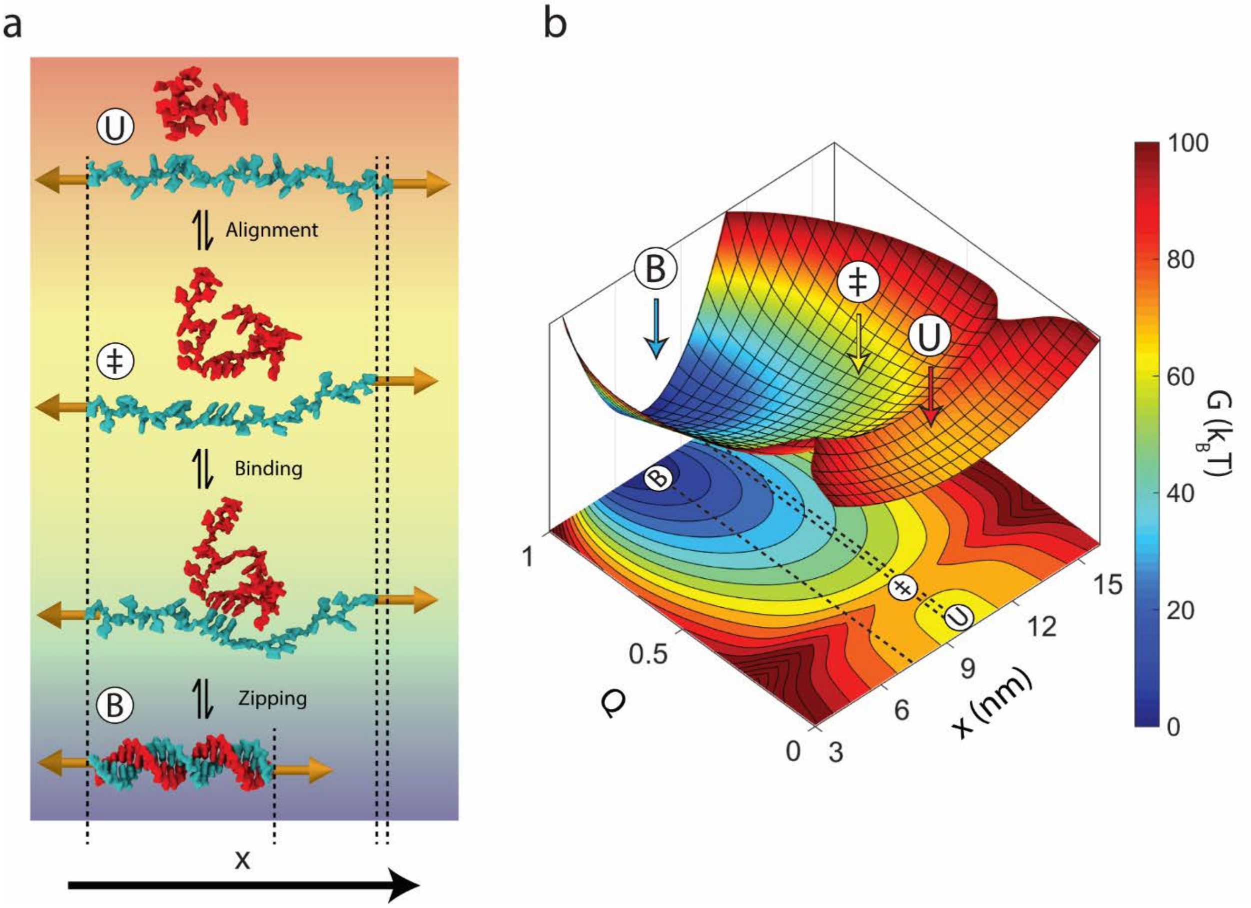


Figure 2.11: Model for nucleic acid hybridization. (**a**) Schematic depicting nucleic acid hybridization of a 22-nt oligonucleotide (*red*) to a complementary strand (*cyan*) under a force of 15 pN. In the unbound state (U), the strand

Figure 2.11 (continued): under tension encounters a random-coil oligonucleotide. Most encounters between these two strands do not result in duplex formation because they are not aligned properly with respect to one another. When both strands transiently form a short stretch of aligned nucleotides (‡), they are prepared to bind to one another. The two strands then bind and rapidly zip together to form the bound-state duplex (B). The double-stranded DNA in this schematic was created using VMD (131) using PDB entry 1BNA. (**b**) Model energy landscape corresponding to the schematic in (**a)** at 15 pN. Two reaction coordinates are shown: the end-to-end extension of the strand held under tension, x, and the fraction of native duplex contacts formed, Q. Dotted lines on the x-Q projection are shown to clarify the end-to-end extension of each state.

2.4.3. Constructing an energy landscape for the hybridization reaction

Using our generalized model for the transition state, we can construct a schematic energy landscape that encapsulates the hybridization reaction (**Figure 2.11b**). The extensions of the bound, unbound, and transition states are calculated from the elasticity models described above (**Table 2.1**). We note that there are forces at which the equilibrium states have the same extension as the transition state (bound, ~3 pN; unbound, ~12 pN; **Figures 2.2b, 2.2c,** and **2.10b**), making the activation barrier disappear in the extension coordinate. This behavior implies that extension is a poor coordinate at these forces and that a simple 1D model of the energy landscape is insufficient (57). It is therefore helpful to introduce a second coordinate, Q, representing the number of native contacts formed to illustrate the hybridization reaction (132). The free energy difference and extension differences between the bound and unbound states ΔG^{o} are taken directly from our data (**Figure 2.2d**), whereas Q is schematic. The total free energy difference between the bound state and the transition state $\Delta G^{\ddagger}$ was calculated from the zero-force unbinding rate constants, assuming an intrinsic rate constant $k_0 = 10^7$ s^{-1}, in accordance with literature values (96,106,123) (**Figure 2.8**).

2.4.4. Biological implications

A model of the transition state as a pre-ordered, single-stranded nucleic acid is attractive as it may help explain the mechanism of protein-mediated hybridization in the cell. Oligonucleotides do not hybridize very quickly on their own, but a protein may speed up the reaction by pre-ordering a single-stranded nucleic acid for binding a target. There are several noteworthy examples of such proteins: crystal structures of DNA-degrading CRISPR-Cas9 and RNA-silencing Argonaute reveal a short stretch of “seed” ssNA (RNA for Cas9 and human Argonaute-2, DNA for some bacterial Argonautes) pre-ordered in preparation for binding their targets (69,81,133,134). Our results suggest that this convergent “seed” mechanism may not only increase the specificity of the guide for its target, but also the rate at which they bind by mimicking the transition state for hybridization.

2.4.5. Applications to nanotechnology

We also expect our results to help in the design and modeling of DNA nanotechnology. For example, recently developed DNA-based force sensors (7) would benefit from a detailed model for force-dependent melting rates, which has been lacking to-date (119). Nucleic acid-based nanomachines are another example. Most nanomachines rely on strand-displacement, where an invading strand displaces an incumbent strand (88). This process is generally very slow, but kinetic control may be possible through application of force on the incumbent strand to increase its off-rate (8). We thus anticipate our model for the force- and length-dependence of hybridization rate constant to be useful in predicting and designing force-sensitive nucleic acid-based nanostructures.

Chapter 3

Elasticity of ultrashort oligonucleotides

In this chapter we will discuss the elastic behavior of ultrashort nucleic acids under an applied force. To measure the elasticity at such a short length scale, we will use the assay described in **Chapter 2** to observe the change in extension of a tethered DNA molecule when a short (~10 nt) oligonucleotide hybridizes with a complementary sequence on the DNA tether. We will show that this change in extension does not follow behavior predicted by common polymer models, and that this deviation is the result of edge effects. We then explain our results with a model that describes shear-induced base-fraying at the ends of a nucleic acid duplex.

3.1. Background

3.1.1. Nucleic acid elasticity in biological systems

Nucleic acids are subjected to numerous stretching, bending, and twisting forces within cells, and their elastic behavior is an important determinant in the function of many enzymes. For example, gene expression in eukaryotes depends on the local flexibility of DNA that is wrapped tightly around nucleosomes, which depends on both sequence and chemical modification of bases (3,135). NAs are also subjected to stretching forces by enzymes involved in diverse cellular systems such as replication (4), transcription (136,137), translation (5,138,139), and chromatin remodeling (140). In some cases the tension within nucleic acids causes enzymes to switch between different behaviors, as when T7 DNA polymerase switches from polymerization to base excision activities above a certain force (4).

3.1.2. Nucleic acid elasticity in nanotechnology

Detailed knowledge of the structural and elastic properties of NAs has fueled the dramatic expansion in NA-based nanotechnology over the past ~25 years. The advent of techniques such as DNA origami have enabled the design of increasingly-complex nanostructures (141,142). DNA origami is a technique by which structures are formed from a long, single-stranded DNA “scaffold” (typically the genomic DNA from the M13 phage) and short, single-stranded DNA “staples.” This technique has enabled the synthesis of 2D shapes as diverse as squares, stars, smiley faces, and a map of the Americas (143). Shortly after the development of this technique, researchers expanded it to three dimensions to create structures like tetrahedra (144), icosahedra (145,146), boxes with lids that can be opened with “keys” (147), and “nanoflasks” (148). Some nanostructures have the potential to be used as drug delivery systems to target specific cell types like cancer cells by releasing a cargo in response to signals within that specific cell type, such as miRNAs (149), protein antigens (150), or small-molecule secondary messengers (151). Relevant to our present discussion, the ability to incorporate local stretching and compressing forces has further expanded the toolkit for engineering more complicated structures (6).

Structural information of DNA has also been used to design nanodevices for performing useful tasks. Various nanoscopic machine components have been constructed using DNA origami for investigating biological systems, such as rotational and linear joints, crank-sliders (152,153), valves (154), calipers (155), actuators (156), and force clamps (157). One group of DNA nanodevices especially relevant to the present discussion are force-sensors termed Tension Gauge Tethers (TGTs), which are short DNA duplexes designed to rupture at a particular force. These sensors have been used to measure the forces required to activate mechanosensitive receptors on cell surfaces (7,158).

In addition to DNA, a detailed understanding of the structural properties of RNA-DNA hybrids could be useful for NA nanotechnology. RNA nanostructures have been expressed *in vivo*, and these structures were used to increase hydrogen gas biosynthesis in bacteria 11- to 48-fold (159). It is conceivable that RNA-DNA hybrid nanostructures could also be assembled *in vivo*, using short RNA strands (like miRNAs) as ‘staples’ and a transfected ssDNA phage genome (like the commonly used M13mp18) as a ‘scaffold’. However, the design of NA nanostructures depends on a detailed knowledge of RNA-DNA structural properties, which remain poorly understood.

3.1.3. Previous studies of nucleic acid elasticity

Over the past 25 years the elastic properties of long NA duplexes have been revealed in increasing detail by manipulating individual strands using single-molecule techniques. Early studies using magnetic tweezers to apply stretching forces to long strands of dsDNA discovered that these polymers are well-described by the worm-like chain (WLC) model (121,160), which describes a polymer as a semiflexible rod characterized by its “persistence length.” Persistence length is a measure of stiffness that corresponds to the length at which a polymer behaves as a rigid rod, which for dsDNA was found to be ~50 nm (~150 bp) (121). However, subsequent experiments revealed that this model is not wholly adequate over a larger range of stretching forces, as it does not account for the ability of real polymers like dsDNA to extend beyond their contour lengths (120,161,162). These studies found that a modified version of the WLC model called the “extensible worm-like chain” (XWLC) (163) better described the behavior of dsDNA under force. This updated model describes the elastic behavior of dsDNA well until higher (>40 pN) forces, where newer models are required to account for more complex phenomena like twist-stretch coupling (164-166) and the so-called ‘overstretching’ transition (161,162,166-168).

The elastic behavior of long strands of ssDNA has proven more complicated than that of dsDNA. Unlike its double-stranded counterpart, the elasticity of ssDNA depends strongly on nucleotide sequence and environmental conditions like ionic strength (130). Because of its inherent variation, a wide range of models and elastic parameters have been reported in the literature (for a discussion of the many models and parameters, see (169)). More recently, the elastic behavior of ssDNA has been described with using the “snake-like chain” (SLC) model that accounts for locally-interacting domains (122,170), and long-distance polyelectrolyte effects (170-173).

Long NA strands are well-described by ideal polymer models like WLC and SLC, but it remains unclear how well these models describe short strands. Several experimental studies have observed that short strands of dsDNA (~100 bp) are far more flexible than predicted by the WLC model (174-177), and alternative models to explain this behavior, such as the “sub-elastic chain” model, have been proposed (176). However, experimental artifacts may be responsible for the apparent increase in flexibility (178-181). Alternative experimental designs and computational studies have instead found that the flexibility of short strands of dsDNA is well-described by the “kinkable worm-like chain” model, which is a WLC model modified to account for the finite probability of DNA to form “kinks,” likely caused by opening of isolated base-pairs (178-180,182,183).

Although numerous studies have investigated the elastic behaviors of short and long NA strands, very few have examined those of ultrashort (<20 bp) strands under force. One prevailing model by de Gennes treats an ultrashort duplex as a network of harmonic springs (184). According to this model, a shearing force produces a distortion of the base-pairs near the ends of the duplex, and base-pairs break if they are displaced too far from their equilibrium positions. This model has been used to describe the observed dynamics of duplex rupture under a shearing force (185,186).

Here, we used the hybridization assay described in **Section 2.2** (61) to investigate the elastic behavior of ultrashort (≤12 nt) DNA and RNA oligonucleotides under tension by observing the change in molecular extension that takes place during the hybridization reaction (187). Our investigation reveals that NA elasticity is well-described by the standard extensible worm-like chain (XWLC) model under low forces (<10 pN), but deviates significantly from it at higher forces (>10 pN). We demonstrate that this high-force deviation is the result of free duplex ends, and can be reduced or abolished by decreasing the number of free ends. Our findings suggest that edge effects play a significant role in the flexibility of short NA strands. We present a simple model to describe the force-dependent extension of short NAs with free ends, and compare this to existing models in the literature. Our findings may provide insights into the role of stretching forces in biological systems, as well as inform the design of NA-based nanodevices.

3.2. Measuring the force-dependent extension difference between single- and double-stranded nucleic acids

3.2.1. Measuring extension changes

The change in extension between single- and double-stranded NAs during hybridization was measured using the assay described in **Section 2.2** (**Figure 3.1a**). In addition to recording the fluorescence signal, the trap signal was recorded at 267 Hz to observe the change in extension upon probe binding and unbinding (**Figure 3.1b**). Binding and unbinding events were located using the fluorescence signal as described (**Section 2.2.3**). The difference in extension between bound and unbound states was measured by taking the difference of the trap positions before and after the stepwise event, located by fluorescence, averaged over 0.5 s both before and after the event. The binding and unbinding events were equal in magnitude to one another, within error (**Figure 3.1c**), and so the data for binding and unbinding were combined together for subsequent analysis.

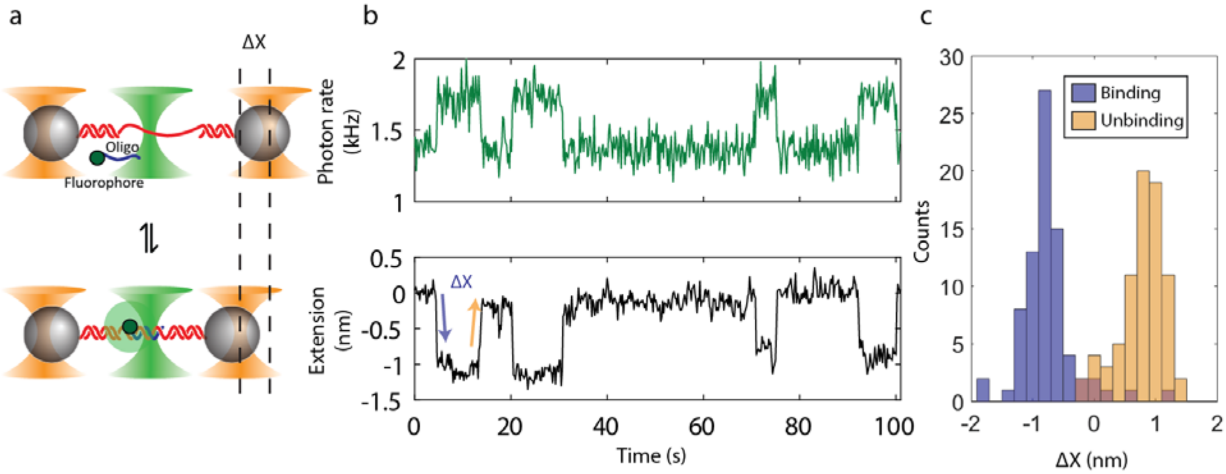


Figure 3.1: Measurement of extension difference of single- and double-stranded states of ultrashort oligonucleotides under force. (**a**) Schematic of the hybridization assay (not to scale). An engineered DNA molecule (*red*) containing a short, central ssDNA region flanked by long dsDNA handles is held under constant force by polystyrene beads (*grey spheres*) held in optical traps (*orange cones*). A fluorescence excitation laser (*green cone*) is focused on the central ssDNA region. Short oligonucleotides (*blue*) labeled with a Cy3 fluorophore at the 3’ end (*green disk*) bind and unbind to the complementary ssDNA sequence in the center of the tethered DNA. The binding and unbinding is observed by the fluorescence emitted from the attached fluorophores and the change in separation between the optical traps. (**b**) Representative time trace showing 10-nt DNA probes binding and unbinding a DNA

Figure 3.1 (continued): construct held under constant force (12.4 pN). The extension difference of the single-stranded state and the double-stranded state, ΔX, is measured from the stepwise increase or decrease of the trap separation. (**c**) Histogram of the recorded extension differences for the 10-nt probes using the hybridization assay.

3.3. The elasticity of ultrashort oligonucleotides reveals edge effects

3.3.1. Changes in extension don't follow predicted behavior

Figure 3.2 shows the effect of tension on the extension change for all probes, scaled by probe length $\Delta x = \Delta X/\ell$. We compared these measured values to the extension change expected for long polymers, *i.e.* $\Delta x_{model}(F) = x_{ds}(F) - x_{ss}(F)$, where $x_{ds}(F)$ is the extension of the double-stranded state per base pair and $x_{ss}(F)$ is the extension of the single-stranded state per nucleotide, using models described in **Appendix A.6**. Although the measured Δx agree very well with this simple model at low forces (<10 pN), they increasingly deviate from it at higher forces (>10 pN).

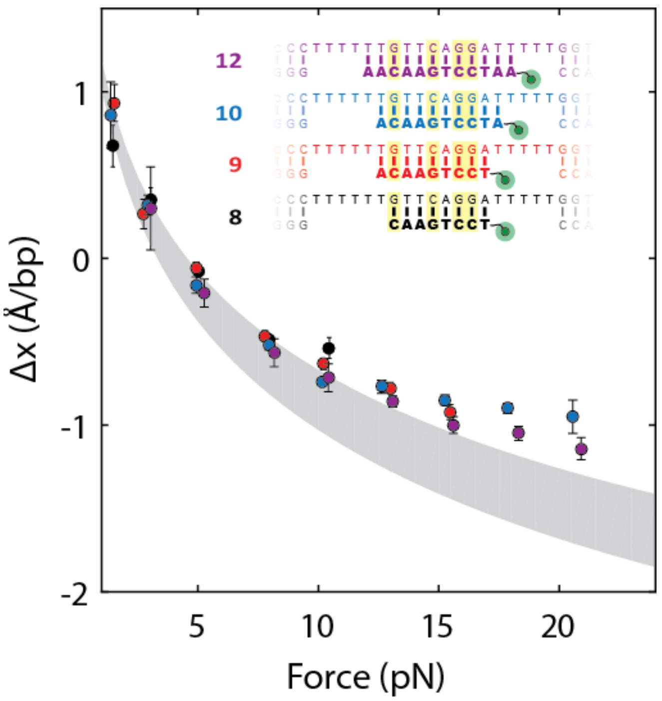


Figure 3.2: Comparison of measured extension changes to predicted polymer models. Extension changes of DNA constructs when probes bind their complementary sequences, scaled by probe length (extension changes from both binding and unbinding events are combined for each data point; error bars denote s.e.m.). The grey shaded region shows the force-dependent model $\Delta x_{model}(F) = x_{ds}(F) - x_{ss}(F)$ using the XWLC model for $x_{ds}(F)$ and the SLC model for $x_{ss}(F)$ (see **Appendix A.6**). *Inset*: the four oligonucleotide probes used in this study (*bold*), bound to their complementary sequences on the DNA construct (*non-bold*). GC pairs are highlighted. Each oligonucleotide has a Cy3 fluorophore conjugated to its 3' end (*green disks*).

3.3.2. Effect of electrostatics

Since NAs can be strongly affected by the ions in their environments, we first investigated the role of electrostatics in the measured extension changes. We repeated the above experiments under conditions of higher ionic strengths (2 and 20 mM [Mg^{2+}]; **Figure 3.3**), finding that the extension differences at lower forces were significantly affected by the presence of Mg^{2+}, although the deviation from the long-polymer model at higher forces was unchanged. The observed change in behavior at low force is most likely caused by the single-stranded state, which is strongly affected by the presence of multivalent ions (170,188). We found that the data collected in the presence of Mg^{2+} was better described using the XWLC model for ssDNA using parameters described in **Appendix A.6**, although this did not account for the deviation at higher forces.

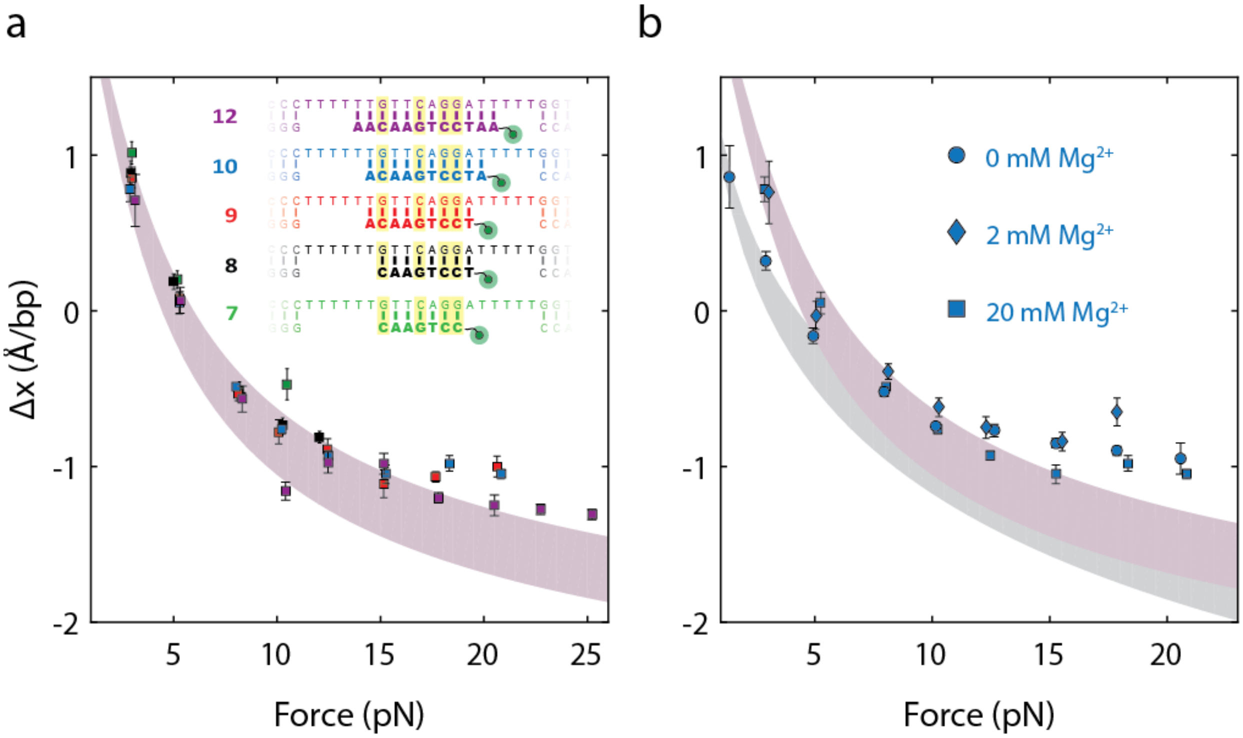


Figure 3.3: Effect of [Mg^{2+}] on measured extension changes. (**a**) Extension change of DNA constructs in 20 mM Mg^{2+} when probes bind their complementary sequences, scaled by probe length (extension changes from both binding and unbinding events are combined for each data point; error bars denote s.e.m.). The pink shaded region shows the force-dependent model $\Delta x_{model}(F) = x_{ds}(F) - x_{ss}(F)$ using the XWLC model for $x_{ds}(F)$ and the XWLC model for $x_{ss}(F)$ (see **Appendix A.6**). *Inset*: the five oligonucleotide probes used in this study (*bold*), bound to their complementary sequences on the DNA construct (*non-bold*). GC pairs are highlighted. Each oligonucleotide has a Cy3 fluorophore conjugated to its 3' end (*green disks*). (**b**) The dependence of measured extension changes on [Mg^{2+}] for the 10-nt DNA probe. The force-dependent model is shown for two different polymer models of ssDNA (*Grey shaded region*: using SLC model. *Pink shaded region*: using XWLC model).

3.3.3. RNA-DNA hybrids show larger deviation from predicted behavior

We next considered if the extension change observed for DNA duplex formation extended to other NAs. We thus replaced the DNA oligonucleotide probe with RNA and repeated our measurements for RNA-DNA hybrid formation. The resulting extension differences showed more complicated behavior than that of DNA duplex formation (**Figure 3.4**). For the 7- and 8-nt RNA probes investigated, the extension difference did not agree with the predicted model across any range of forces (see **Section 3.4**), while the 9-, 10-, and 11-nt RNA probes agreed with the predicted model at low forces, but not high forces. The longer probes thus demonstrate qualitative agreement with the extension differences from DNA duplex formation, although the high-force deviation is considerably larger for the RNA-DNA system.

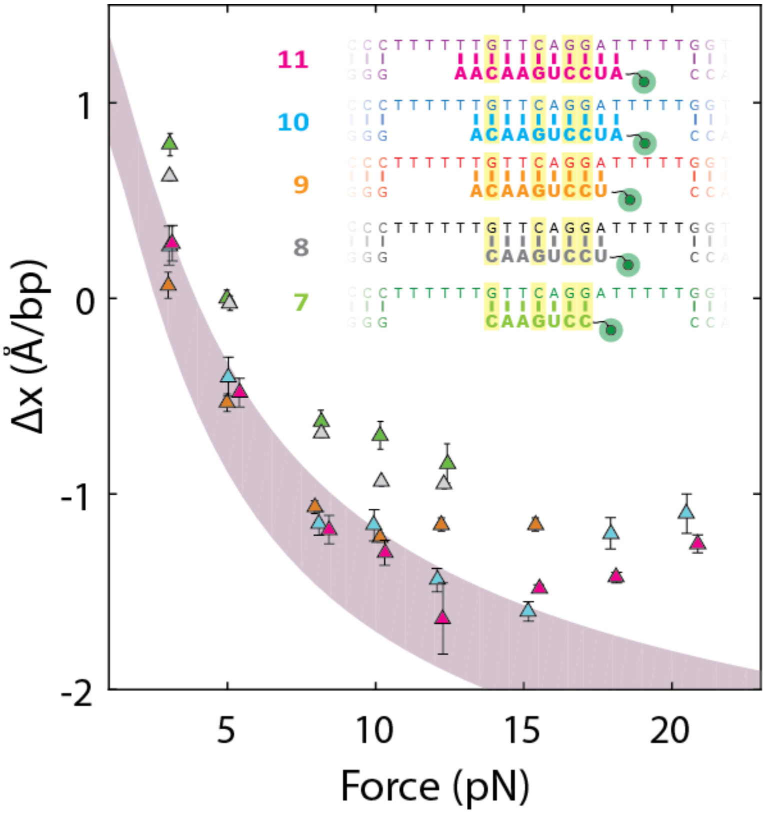


Figure 3.4: Extension changes of RNA-DNA hybridization. Extension change of DNA constructs in 20 mM Mg^{2+} when RNA probes bind their complementary sequences, scaled by probe length (extension changes from both binding and unbinding events are combined for each data point; error bars denote s.e.m.). The pink shaded region shows the force-dependent model $\Delta x_{model}(F) = x_{ds}(F) - x_{ss}(F)$ using the models for $x_{ds}(F)$ and $x_{ss}(F)$ described in **Appendix A.6**. *Inset*: the five RNA probes used in this study (*bold, light colors*), bound to their complementary sequences on the DNA construct (*non-bold, dark colors*). GC pairs are highlighted. Each oligonucleotide has a Cy3 fluorophore conjugated to its 3' end (*green disks*).

3.3.4. Effect of fluorescent dye

We considered that the attached 3’ fluorescent dye may be responsible for the observed high-force deviations. In order to determine the effect of the 3’-Cy3 label on the measured extension changes, we used the assay described in **Section 2.2.6** to compare directly the extension changes of labeled and unlabeled probes. An example data trace is shown in **Figure 3.5**. The extension changes measured from labeled and unlabeled probes agreed well with one another.

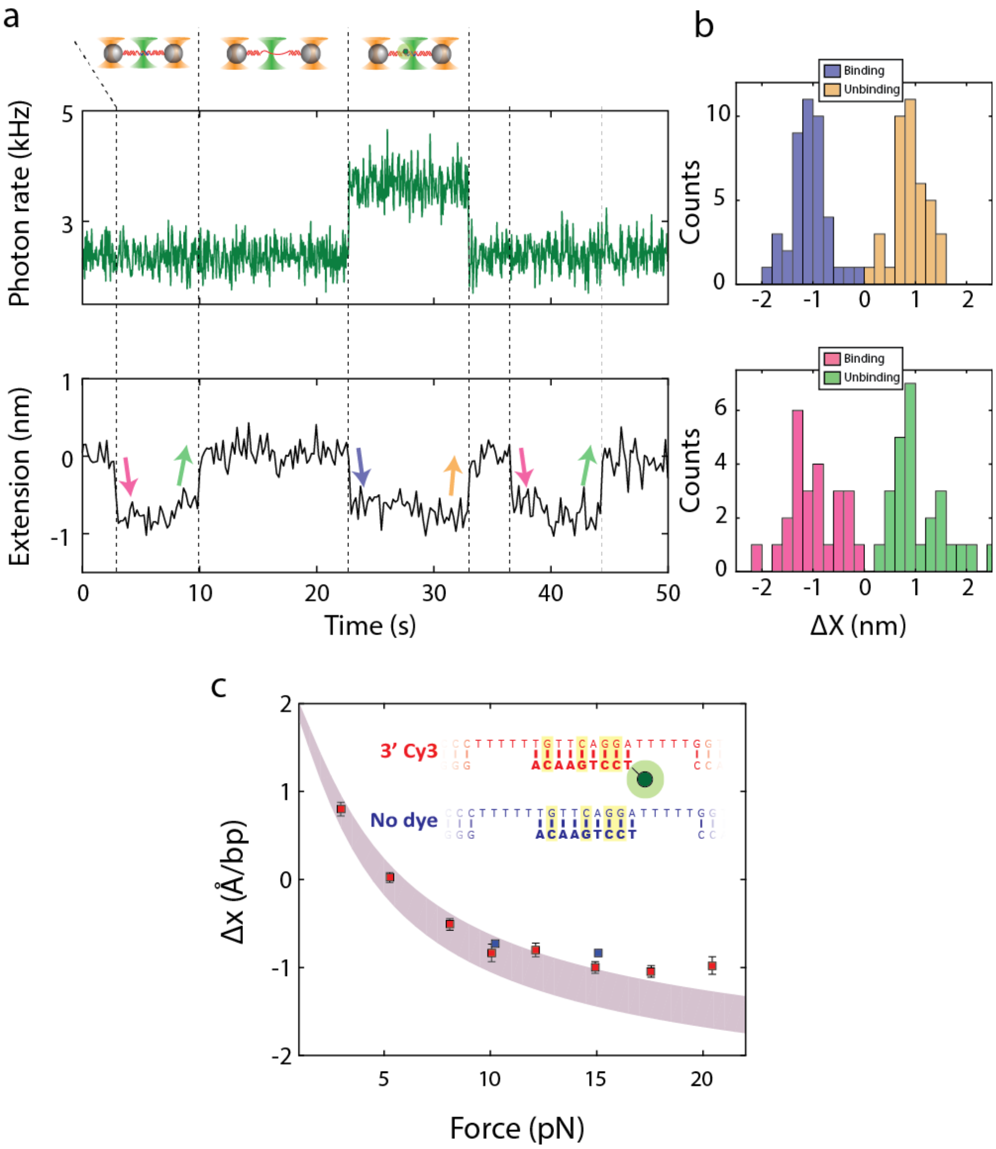


Figure 3.5: Effect of 3’-Cy3 dye on measured extension changes. (**a**) Representative time trace as shown in **Figure 2.7** depicting “half-and-half” experiment, where labeled and unlabeled probes bind and unbind the DNA construct. (**b**) Histograms of labeled (*top*) and unlabeled (*bottom*) extension change events for the half-and-half experiment using 9merCy3 and 9mer. (**c**) Length-scaled extension change for 9merCy3 (*red*) and 9mer (*blue*) probes with force-dependent model (*pink shaded region*; see **Appendix A.6**).

3.3.5. *Effect of sequence*

We considered that the high-force deviations may be due to edge effects of the short duplexes formed by the oligonucleotides bound to the tethered ssDNA, such as fraying of terminal base pairs under tension. We thus repeated the above measurements using oligonucleotide sequences whose weak terminal AT pairs were replaced with stronger GC pairs (**Figure 3.6**). None of the alternate sequences assayed demonstrated significantly different behavior from those assayed previously.

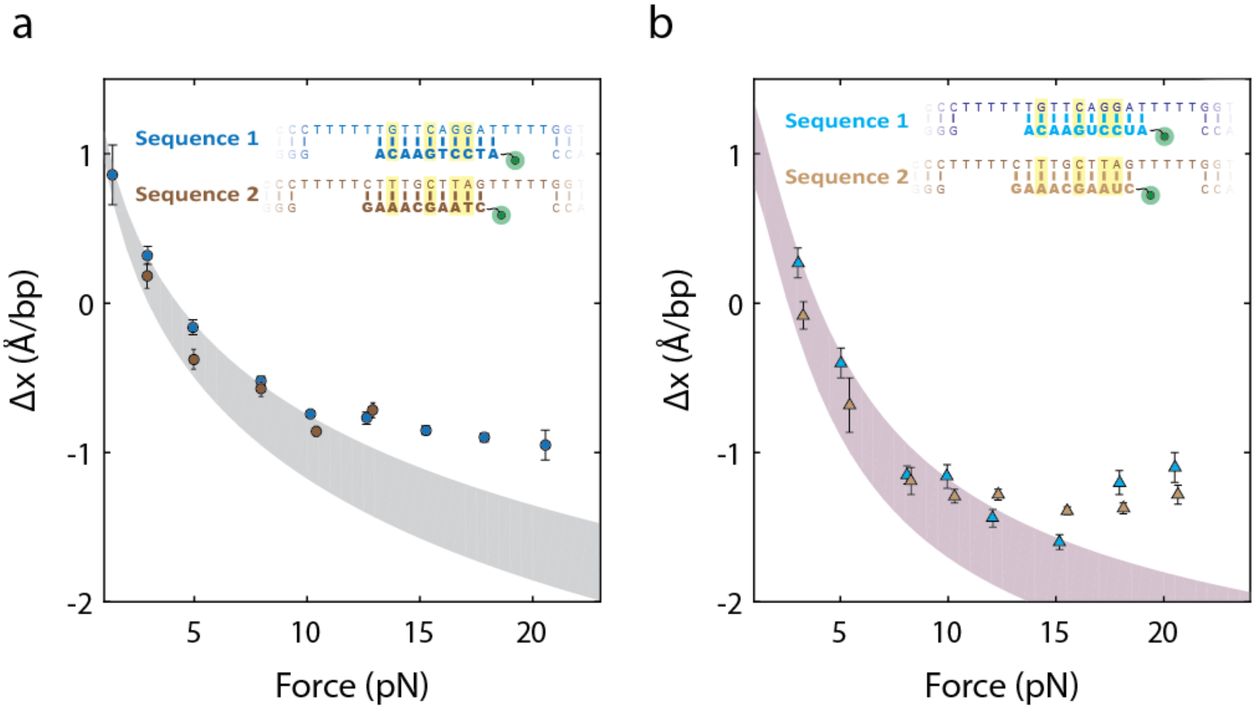


Figure 3.6: Effect of sequence. (**a**) Extension changes of DNA constructs when probes of two different sequences (*inset; bold letters*) bind their complementary sequences (*non-bold letters*), scaled by probe length (extension changes from both binding and unbinding events are combined for each data point; error bars denote s.e.m.). The grey shaded region shows the force-dependent model $\Delta x_{model}(F) = x_{ds}(F) - x_{ss}(F)$ using the model described in **Appendix A.6**. (**b**) Measured extension changes for an alternate RNA probe sequence binding a DNA construct with complementary sequence. The two probe sequences have the same GC content, but Sequence 2 has GC pairs at the ends. The pink shaded region shows the force-dependent model $\Delta x_{model}(F) = x_{ds}(F) - x_{ss}(F)$ using the model described in **Appendix A.6**.

3.3.6. *Effect of free ends*

A commonality in the measurements above is the absence of neighboring base pairs at the edges of the bound probes in our construct design (**Figure 3.2 inset**). Removing one or both of the dT 'spacers' flanking the probe binding site had a significant effect on the deviation between data and model. Data from binding of a 9-nt probe on variants of the DNA substrate lacking one ("1Sp") and both spacers ("0Sp", **Figure 3.7 inset**; **Table A.1**) displayed none of the high-force deviation observed from the construct with both spacers ("2Sp"; **Figure 3.7**), instead showing a change in extension well in-line with that predicted by the model. (For the 0Sp measurement, we used a 9-nt DNA probe that had an internally-attached dye ("9merIntCy3", **Table A.3**) to prevent steric hindrance with the neighboring 'handles'). These results demonstrate that the deviation from the predicted elastic behavior at high force is strongly affected by the terminal base pairs of the hybridized probe.

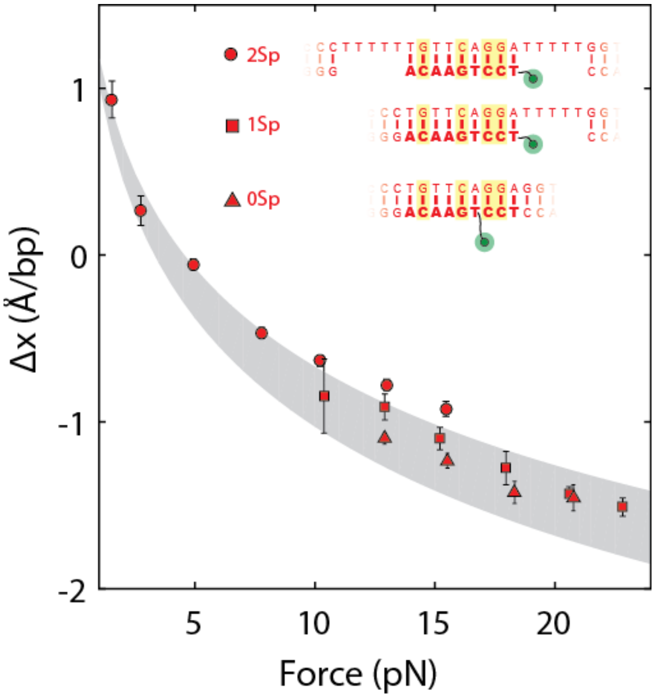


Figure 3.7: Effect of free ends on extension difference between single- and double-stranded states. Extension changes of DNA constructs when 9-nt probes binds their complementary sequences, scaled by probe length (extension changes from both binding and unbinding events are combined for each data point; error bars denote s.e.m.). The grey shaded region shows the force-dependent model $\Delta x_{model}(F) = x_{ds}(F) - x_{ss}(F)$ using the XWLC model for $x_{ds}(F)$ and the SLC model for $x_{ss}(F)$ (see **Appendix A.6**). *Inset*: the 9-nt probe (*bold*) bound to the three DNA constructs used in these experiments (*non-bold*). GC pairs are highlighted. The 9-nt probes used for binding the 2Sp and 1Sp constructs had a Cy3 fluorophore (*green disks*) conjugated to their 3' ends, while the 9-nt probe used for binding the 0Sp construct had a Cy3 fluorophore conjugated to an internal dT base to avoid steric clashes with the handles.

Our findings above indicate deviations from the long-polymer expectation at forces exceeding 10 pN. Because deviations from this model vary based on the oligonucleotide probe (*i.e.* DNA vs. RNA) while the tethered construct remained the same, the error must lie in our model of elasticity of the duplex formed during probe hybridization. The results with constructs lacking one or both spacers suggest that edge effects from terminal base pairs contribute greatly to the deviation. To model this effect, we consider that each duplex edge has a different force-extension behavior $x_e(F)$ compared to the internal portion of the duplex, which we assume to follow a long-polymer model $x_{ds}(F)$. We must account for such edge effects not only at each end of the hybridized probe but also at any other ds-ssDNA junction found on the tethered molecule (**Figure 3.8**). Thus, the extension of the unbound (unhybridized) state, $X_u(F)$, is given by:

$$X_u(F) = (N_h - 2n_e)x_{ds}(F) + 2n_e x_e(F) + (\ell + \theta\ell_{sp})x_{ss}(F) \tag{3.1}$$

where N_h is the number of base pairs of the dsDNA handles, n_e is the number of base pairs that comprise the edge regions with different elastic properties, ℓ is the length of probe binding site, ℓ_{sp} the spacer length, and $\theta = 0, 1, 2$ is the number of spacers flanking the binding site (**Figure 3.8**). $x_{ds}(F)$, $x_e(F)$, and $x_{ss}(F)$ are the extensions of 1 base pair of internal dsDNA, edge dsDNA, and 1 nucleotide of ssDNA, respectively. Upon probe binding, the bound (hybridized) state, $X_b(F)$, is given by:

$$X_b(F) = (N_h + \ell - 2\theta n_e)x_{ds}(F) + 2\theta n_e x_e(F) + \theta\ell_{sp}x_{ss}(F) \tag{3.2}$$

Thus, the measured extension change is:

$$\Delta X = X_b(F) - X_u(F) = (\ell - 2(\theta - 1)n_e)x_{ds}(F) + 2(\theta - 1)n_e x_e(F) - \ell x_{ss}(F) \tag{3.3}$$

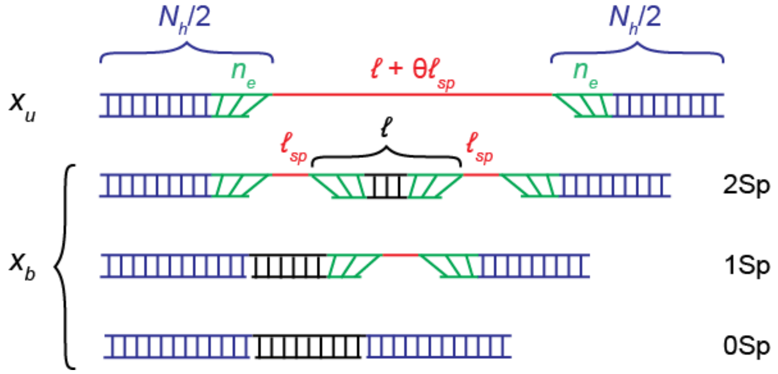


Figure 3.8: Free edges in bound and unbound states. Schematic of the construct variants used in this study. *Top* (x_u): unbound-state construct. *Bottom three* (x_b): bound-state constructs (2Sp, 1Sp, and 0Sp) with probe bound. *Blue*: dsDNA handles. *Green:* Edges. *Red:* Probe binding site and dT spacers. *Black:* Bound probe.

It is instructive to plot the deviation between the measured extension change and the long-polymer model (**Figure 3.9**). According to Eq. (3.3) above, the deviation should equal the following simple expression:

$$\begin{aligned}\Delta X-\Delta X_{model} &=\left(X_b(F)-X_u(F)\right)-\ell\left(x_{ds}(F)-x_{ss}(F)\right)\\ &=2(\theta-1)n_e\left(x_e(F)-x_{ds}(F)\right)\end{aligned} \tag{3.4}$$

Eq. (3.4) predicts that the $\theta = 1$-spacer construct should not deviate from the long-polymer model while the $\theta = 2$- and 0-spacer constructs must deviate from this model in opposite directions by the same magnitude. This is precisely what we observe (**Figure 3.9b**). The reason for this behavior is simple: for $\theta = 1$ the same number of duplex edges are present before and after probe binding, whereas this number changes by +2 in the case of $\theta = 2$ spacers, and –2 in the case of $\theta = 0$ spacers (**Figure 3.9b**).

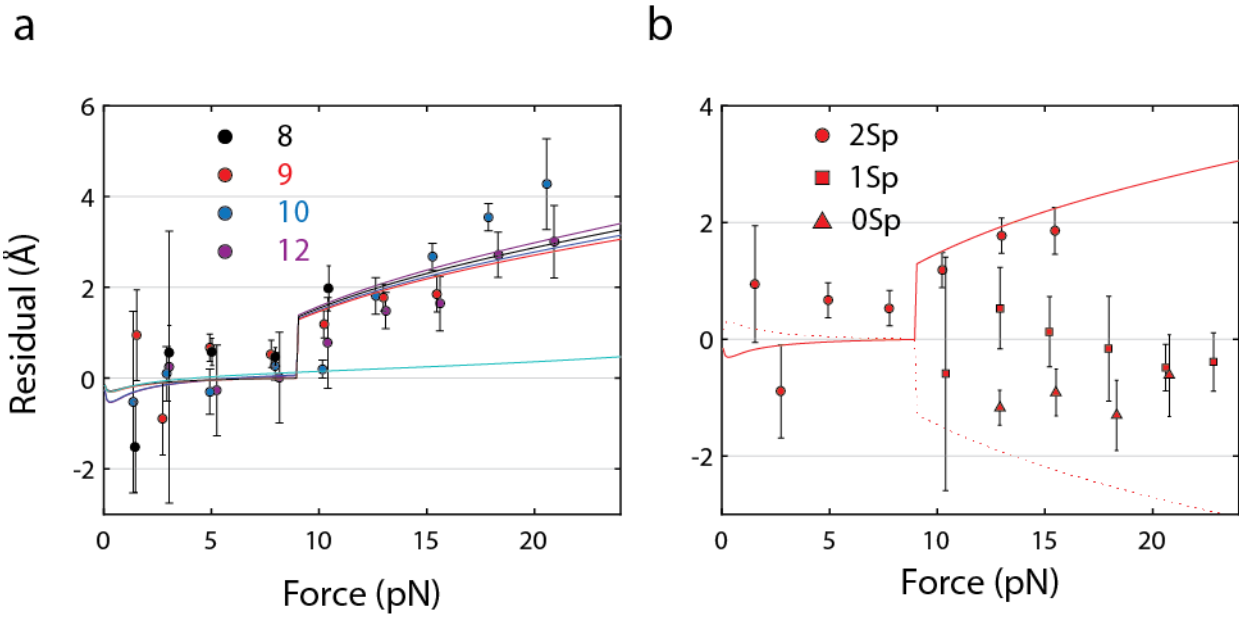


Figure 3.9: Deviation of measured extension changes from long-polymer model. (**a**) Residuals of extension change data for probe variants, using optimal model parameters for subtraction (see **Appendix A.6**). *Cyan line*: Fraying model with no shearing term. *Black, red, blue, and purple lines*: Models for shear-induced base-fraying for corresponding probe lengths with globally-fitted parameters. (**b**) Residuals of extension change data for construct variants. *Solid red line*: Model for 2Sp construct. *Dotted red line*: Model for 0Sp construct.

3.3.7. Modeling the effect of free ends

What edge effects could lead to different force vs. extension behavior? One possibility is simply that the terminal base-pairs of the duplex increasingly fray under force. We developed a simple statistical model similar to that of Gross et al. (166) in which the duplex can have a number of base-pairs n_e thermally frayed from its ends. The free energy of a duplex with n_e frayed base-pairs is a sum of three terms:

$$g_{tot}(F)=(\ell-n_e)g_{ds}(F)+n_e g_{ss}(F)+\sum_{i=1}^{\ell-n_e-1} g_{bp}^{(i)} \tag{3.5}$$

Here the first term represents the elastic energy of the double-stranded (*i.e.* un-frayed) portion of the duplex $\ell - n_e$ in length and where $g_{ds}(F)$ is the energy of stretching a single base-pair to force F calculated using the XWLC model, the second term represents the elastic energy of the single-stranded (*i.e.* frayed) portion n_e nucleotides in length, where $g_{ss}(F)$ is the energy of stretching a single nucleotide to force F calculated using the SLC model, and the third term sums over the $\ell - n_e - 1$ nearest-neighbor base-pairing energies $g_{bp}^{(i)}$ taken from literature values (90,92).

The expected deviation from the long-polymer model is given by Eq. (3.4) with $x_e(F) = x_{ss}(F)$ and $n_e = \langle n_e(F) \rangle$ the average number of frayed base-pairs at force F. Comparing this model to our results, fraying fails to capture the magnitude of the deviation observed (**Figure 3.9a**) because the difference in elastic energies of the double- and single-stranded states is not comparable to the base-pairing energies until a force of ~60 pN (166), much higher than the forces assayed.

To generate the larger deviations observed, a fraying model must include other contributions destabilizing the edge base-pairs. In our experimental configuration, tension on the tethered strand not only stretches each strand of the duplex but also generates a shearing force. We considered the possibility that the terminal base-pairs of the duplex are destabilized by the applied shearing force. This effect was first considered, albeit in a different geometry, by de Gennes in a “ladder model” (184), in which shearing force distorts base-pairs at the ends of the duplex. The ladder model describes a short double-stranded DNA molecule as a network of harmonic springs between neighboring bases along each strand (with spring constant Q) and between base-pairs across the two strands (with spring constant R). Shearing creates a distortion in edge base-pairs which propagates a characteristic distance $\chi^{-1} = \sqrt{Q/2R}$ into the duplex. This model also includes the possibility of base-pairs fraying at the ends by treating the R springs as ‘brittle’ bonds: they behave as simple harmonic oscillators until a force threshold f_1, at which point they abruptly break.

We derived the energy of shearing according to the de Gennes model for our (3’-5’) pulling orientation (see **Appendix B.2**) and included this in the energy term in Eq. (3.5). To this we also included a breakage threshold: when the force on an R spring exceeds a threshold f_1, the base-pair breaks. For our model we assume a fixed Q =1620 pN/nm based on the empirically-determined values for the stretch modulus (120) (see **Appendix B.3**). The data was globally fitted to this new shear-induced base-fraying model using a minimization routine that found the values for R and f_1 that minimized the chi-squared value (**Figure 3.9**). Our fits yielded $R = 3100 \pm 800$ pN/nm and $f_1 = 2.1 \pm 0.1$ pN (errors from jackknife resampling). The fitted value of R gives $\chi^{-1} = 0.51$, implying that the distortion in edge base-pairs does not propagate far into the duplex. This result stands in contrast to a value of $\chi^{-1} = 6.8$ determined from a previous study that measured the force required to rupture short duplexes (185).

The sudden breakage of base-pairs assumed in this model can be viewed as an approximation to the shear-induced fraying at the edges. A more physical model would take into account the escape-over-a-barrier problem for the edge base-pairs rather than an abrupt breakage threshold (119).

3.4. RNA-DNA hybrids have variable helical structure

The measured extension changes of the RNA probes binding to their complementary DNA sequences showed two distinct behaviors (**Figure 3.4**). We compared the data to the XWLC model using a range of persistence lengths $10 < P_{RNA\text{-}DNA} < 22$ nm from literature (189) and helix rise $h_{RNA\text{-}DNA} = 2.9$ Å/bp, characteristic of an A-form helix (190). The longer probes (9-11 nt) agreed well with this model at low forces, but deviated from it significantly at higher forces (>10 pN). The shorter (7- and 8-nt) RNA probes, however, did not agree with the predicted model at any forces.

We noticed that these shorter probes compared favorably to the XWLC model using a helix rise of $h_{RNA\text{-}DNA} = 3.4$ Å/bp, which is characteristic of a B-form helix like dsDNA (**Figure 3.10a**). We performed a modified orcinol assay that confirmed that these probes were RNA rather than DNA (191) (**Figure 3.10b**). This assay was performed by pre-treating a DNA or RNA sample with strong acid, then adding an orcinol reagent. Orcinol reacts more strongly with RNA than DNA, and produces a compound that can be detected spectroscopically. Our data therefore suggest that the RNA-DNA hybrids investigated here fall into two discrete sets: B-form helices (7 and 8 nt probes) and A-form helices (9-11 nt probes). Previous studies have shown that RNA-DNA hybrids can adopt either helical conformation (192,193), although the determinants of helicity remain poorly known. It appears that the helicity of these hybrids depends on the purine content of the strands, where a high purine content in the RNA strand tends to favor an A-form duplex, and a high pyrimidine content tends to favor a B-form duplex (192,193). Consistent with the literature, the longer RNA probes investigated above have a higher purine content (44-55%) than the shorter ones (40-43%).

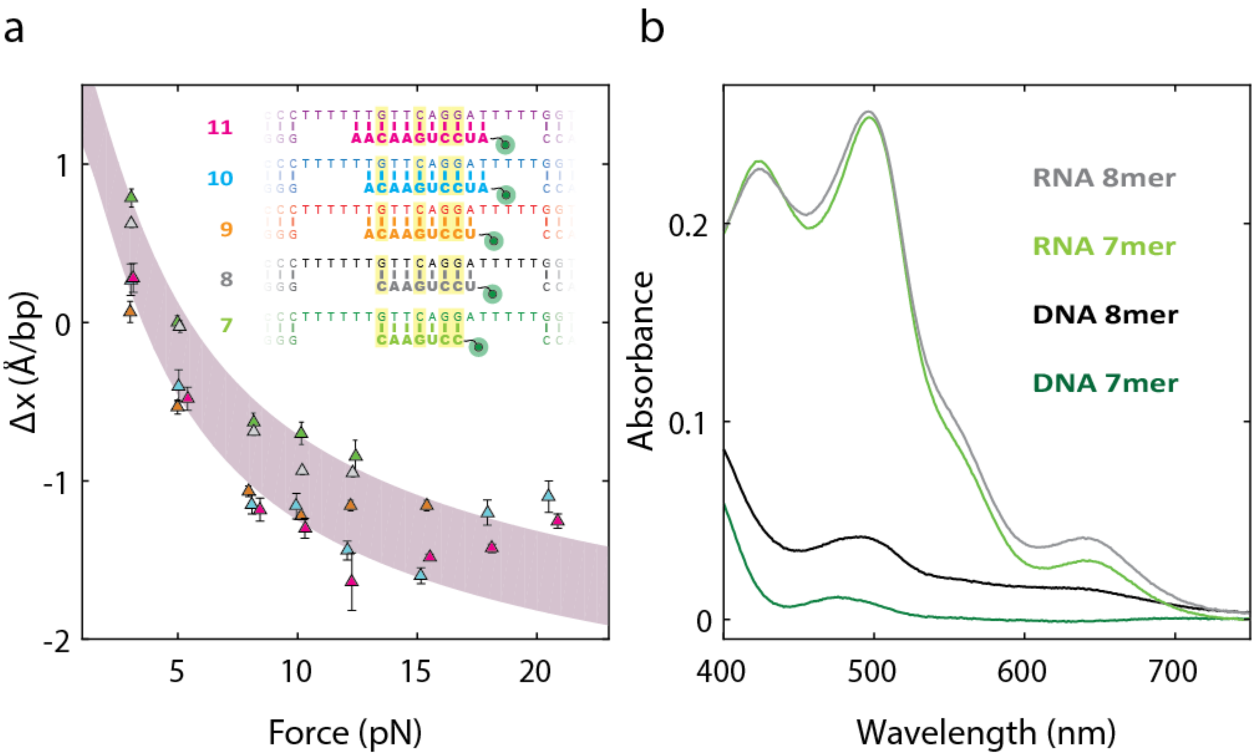


Figure 3.10: Helicity of RNA-DNA hybrids. (**a**) Extension changes as in **Figure 3.4** using the XWLC model (*pink shaded region*) with $10 < P_{RNA\text{-}DNA} < 22$ nm and $h_{RNA\text{-}DNA} = 3.4$ Å/bp shown for comparison. (**b**) Spectra from orcinol assay comparing the 7- and 8-nt DNA probes (*dark green and black, respectively*) to the 7- and 8-nt RNA probes (*light green and grey, respectively*).

Chapter 4

Stepping dynamics of *E. coli* UvrD helicase

In this chapter we will present a preliminary study investigating the mechanism by which *E. coli* UvrD helicase unzips and re-zips DNA. We first describe an assay to measure helicase activity using a DNA hairpin tethered between optical traps, then use this assay to directly observe individual motor steps of UvrD, and measure both the sizes of the steps and their kinetics. We integrate these results with observations from previous studies to produce a mechanism for the stepping behavior of UvrD.

4.1. Background

4.1.1. Stepping dynamics of motor proteins

Motor proteins participate in many cellular systems, from replicating DNA to trafficking vesicles. Nearly every aspect of nucleic acid metabolism involves motor proteins, including RNA polymerases (46), helicases (111,194), ribosomes (139), and viral genome packaging proteins (195). The mechanisms of many of these molecular motors have been elucidated in great detail through the use of single-molecule techniques to directly observe the events in the stepping process that cannot be determined using bulk biochemical methods (37).

The stepping behavior of helicases in particular has revealed rich diversity of molecular motor mechanisms. While some helicases have been observed to unzip DNA one base-pair at a time with simple kinetics, others have shown more complicated behavior. The SF2 helicase XPD was shown to have a motor step size of 1 bp for unzipping DNA, and the distribution of dwell times between each step followed a single exponential function, implying a single rate-limiting event (111). This led to a simple model whereby XPD unzips 1 bp per ATP molecule hydrolyzed. However, the structurally-related SF2 helicase from hepatitis C virus (HCV), NS3, was found to exhibit far more complicated behavior. Initial studies showed that NS3 unzipped RNA in discrete steps of 11 bp, but that these large steps consisted of rapid smaller sub-steps of 3.6 bp (196). However, further experiments revealed that the distribution of dwell times for these steps followed a gamma function, which implied an underlying Poissonian process with several rate-limiting events. The shape of the distribution led to the conclusion that 3-4 events exist within each 3.6 bp sub-step, implying that these sub-steps are not the most elementary steps in the mechanical cycle (197). Instead, Myong et al. proposed a “spring-loaded” mechanism for NS3 whereby 3-4 bp are unzipped rapidly, with one ATP molecule hydrolyzed per base-pair, and the accumulated tension within the protein motor core causes all 3-4 nucleotides to be released at once. The release of unzipped nucleotides was later found to occur asynchronously with the unzipping reaction itself, causing individual nucleotides from the two strands to be released separately (198).

4.1.2. Stepping dynamics of UvrD

The stepping behavior of UvrD has been investigated by a variety of techniques, although there remains poor agreement over quantities like the motor step size. Early bulk kinetic studies showed that the step size is 4-5 bp for unzipping dsDNA (199) or 4-5 nt for translocating along ssDNA (200,201), and single-molecule studies have also inferred a step size of 4-6 bp based on the analysis of noise during the unzipping process (29,202). In contrast to these dynamic experiments, a much smaller step size of 1 bp was inferred from crystal structures showing UvrD in multiple conformational states during the ATPase cycle (203). However, certain behaviors of helicases like pausing and slipping backward can complicate the analysis of many measurements, in some cases leading to an overestimate of the step size (204). UvrD in particular shows complex behaviors that can potentially affect analysis of stepping kinetics. Past studies have shown that it can transition from unzipping to re-zipping the DNA substrate by switching to the opposite side of the duplex and translocating away from the fork junction, allowing DNA to re-zip in its wake (29).

It is possible that the different step sizes measured between structural and other studies can be reconciled. For example, it is possible that the elementary event of the motor process is 1 base-pair unzipped per ATP molecule hydrolyzed, but that there is a slower kinetic event where the protein actually translocates forward, as in the spring-loaded mechanisms proposed for NS3 (197) and T7 gp4 (194). Some evidence for this comes from an investigation into the ATP-coupling stoichiometry of UvrD that demonstrated that the protein hydrolyzes one ATP molecule per nucleotide translocated along ssDNA (204). Such a mechanism would reconcile the ATP-induced 1 bp step seen in crystal structures (203) with the larger 4-6 bp step sizes observed from kinetic measurements (29,199-202).

4.2. Observing helicase activity using hairpin assay

4.2.1. DNA hairpin construct

To measure helicase activity at single-base-pair resolution we used a hairpin assay described previously (196). In this assay, a single helicase is loaded onto the fork junction of a NA hairpin in a controlled manner, and given ATP to unzip it. The advantage of using a hairpin is that it provides an amplified mechanical signal: when a helicase unzips 1 bp from the hairpin stem, 2 nt of ssDNA are released and extend the tethered molecule (196). For example, at a force of 10 pN, each base-pair unzipped generates a 0.8-nm extension change signal.

The hairpin construct was adapted from a previous protocol (111), and consisted of three DNA fragments that are ligated together: a ‘Hairpin’ (HP) flanked by a ‘Right Handle’ (RH) and a ‘Left Handle’ (LH) that serve as functionalized linkers that attach to the trapped beads. The detailed protocol for synthesis of these constructs can be found in **Appendix A.1.2**. Briefly, RH is made from a 1.5-kb PCR-amplified section of the pBR322 plasmid using a 5’-digoxigenin-modified reverse primer and a forward primer containing one abasic site and a long 5’ overhang (**Table A.2**). The digoxigenin moiety is used to link this end of the construct to an anti-digoxigenin-coated bead. The overhang consists of a poly-dT loading site of varying length (10, 19, or 38 nt, depending on the specific experiment) for helicase binding immediately adjacent to the abasic site, followed by 29 nt that anneal to a complementary sequence in HP. LH is synthesized from a different PCR-amplified section of pBR322 as described in **Section 2.2.1**. HP is a single long oligonucleotide containing the complementary sequence to the LH overhang on its 5’ end, followed by the complementary sequence to the RH overhang and a 153-nt self-complementary sequence (**Table A.2**). When self-annealed and ligated to LH and RH, HP makes an 89-bp hairpin stem capped by a $(dT)_4$ tetraloop.

When force is applied to our hairpin constructs (**Figure 4.1a**), they display three behaviors. 1) At lower forces (0-15 pN), the molecule extends as the dsDNA handles and ssDNA loading site are stretched. 2) Near 15 pN, the hairpin is mechanically unzipped by the applied force, causing an abrupt extension of the molecule that follows a sequence-dependent pattern. 3) Following the rupture of the hairpin, the molecule then continues to extend gradually as before, although with more ssDNA available from the opened hairpin. We model the force-extension behavior of both the dsDNA and ssDNA portions with the XWLC model, using P_{ds} = 53 nm, h_{ds} = 0.34 nm/bp, S_{ds} = 1100 pN, P_{ss} = 1 nm, h_{ss} = 0.59 nm/nt, S_{ss} = 1000 pN (112).

4.2.2. Helicase assay

Sample flow chamber. To ensure that no new UvrD molecules can bind the DNA substrate after hairpin unzipping is initiated, we used laminar flow sample chambers that spatially segregated the solution containing protein from the solution containing ATP (111) (**Figure 4.1b**). The chambers had a single central channel in which the two separate streams smoothly merged to create a sharp, un-mixing interface. The lower stream contained ATP (10 μM) but no UvrD while the upper stream contained UvrD (8-10 nM) but no ATP. When moving a DNA tether across the interface, solution exchange typically occurred within ~2 s. To prevent adsorption of protein to the glass surfaces of the coverslips, we passivated the flow chambers with polyethylene glycol (PEG) using a protocol modified from (205) (see **Appendix A.5**).

Experimental conditions. Both fluid streams in the central channel contained 35 mM Tris (pH = 8.0), 20 mM NaCl, 5 mM $MgCl_2$, 2% glycerol, and an oxygen scavenging system to increase the lifetimes of the tethers and fluorophores (55,109) (1.2% glucose, either 1 mg/mL glucose oxidase or 0.29 mg/mL pyranose oxidase, and 0.13 mg/mL catalase; see **Appendix A.3**). All measurements were carried out at 22°C.

Dipping assay. Within the lower channel of the laminar flow chamber, we tethered two beads together *in situ* using the hairpin construct. To confirm the proper behavior of the construct, we took a force-extension (F-x) curve for each tether and fitted it to the XWLC models for the 'closed' and 'open' hairpin states (**Figure 4.1a**). The tethered DNA molecule was then held at a constant force below the rupture force of the hairpin (5-14 pN) using an active force-feedback system (112). The tensed strand was then moved into the channel containing UvrD to load protein onto the short ssDNA loading site (**Figure 4.1b**). After incubating in the UvrD channel for ~15 s to load protein, the tether was then moved back into the channel containing ATP to observe helicase activity.

The activity of the protein was observed by the motion of the beads in response to the release of hairpin DNA (**Figure 4.1c**). As the helicase unwinds the DNA hairpin, the released ssDNA extends the tethered molecule. Under a constant force, this "slack" is taken up by moving the traps further apart. To determine the number of base pairs that are unwound, we convert the trap displacement (in nm) to base pairs (bp) unwound by dividing by the extension of the 2 released nucleotides per base-pair unwound at the force of the measurement. For calculating the extension of ssDNA per nucleotide, we used the XWLC model with the parameters described in **Section 4.1.1**.

***Stepping assay*:** To observe individual motor steps of UvrD directly, we performed the dipping assay using unlabeled, wild-type protein. The dipping assay, along with a DNA hairpin construct with a short (10 dT) protein loading site (**Table A.2**) ensured that only monomers of UvrD would load. Stepping data was collected with the hairpin typically held under a force of 9-15 pN. We repeated this assay for multiple concentrations of ATP (0.5, 1, 2.5, 5, and 10 μM) to investigate the role of ATP in stepping kinetics and thus provide information on UvrD's mechanochemical cycle. Data was collected at either 100 Hz (0.5 to 5 μM ATP data) or 267 Hz (10 μM ATP data), and this raw data was used for subsequent analysis. In order to observe individual steps directly, we selected low-noise segments of the collected time traces for further analysis.

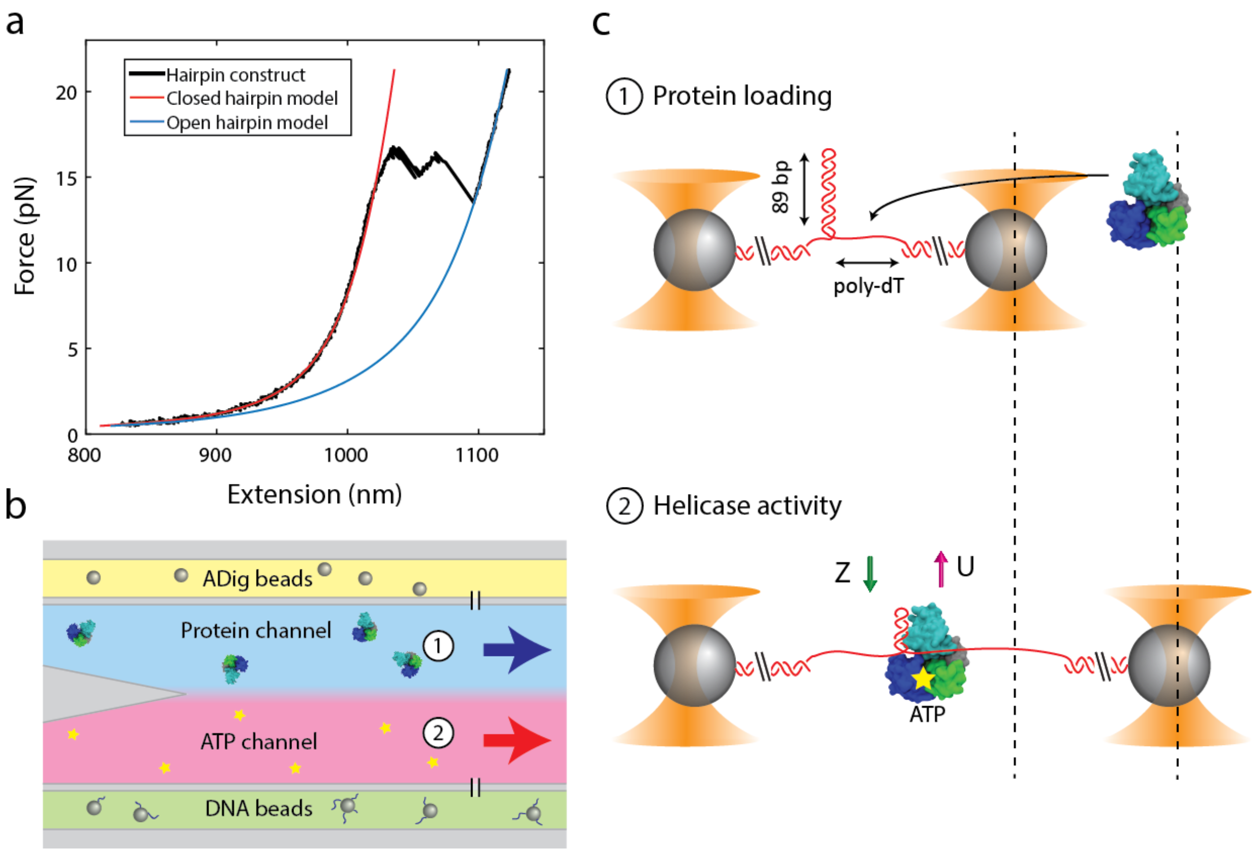


Figure 4.1: Helicase assay using hairpin construct. (**a**) Representative pulling curve of hairpin construct. At lower forces (<15 pN), the extension of the molecule smoothly increases as force is applied. Near 15 pN, the hairpin abruptly ruptures, producing an increase in extension with a sequence-dependent rupture pattern. At higher forces (>15 pN), the extension of the molecule increases smoothly again, now extended by the mechanically-unzipped ssDNA. The 'closed hairpin' and 'open hairpin' are both modeled using XWLC (*red* and *blue*, respectively). (**b**) Schematic of flow chamber used for *in situ* nucleoprotein complex assembly. The flow chamber consists of three channels. The top (*yellow*) and bottom (*green*) channels contain anti-digoxigenin (ADig) and DNA-coated streptavidin (DNA) beads, respectively, and the central measurement channel consists of two parallel laminar flow streams containing 10 nM protein (*blue*) and 10 µM ATP (*red*). During an experiment, we first form a tether between an ADig bead and a DNA bead in the ATP channel and take an F-x curve to assure DNA hairpin quality. We then move the nascent tether to the protein channel (1) and incubate for 15 s to load UvrD, and then finally move to the ATP channel (2) to record helicase activity. (**c**) Schematic of experimental setup. A DNA hairpin (*red*) is tethered between two micron-sized polystyrene beads (*grey spheres*) held in optical traps (*orange cones*). In the protein channel (1), UvrD (*blue, green, grey, and cyan*) is loaded onto a poly-dT ssDNA loading site. After moving into the lower channel (2), UvrD uses ATP (*yellow star*) to unzip (U) and re-zip (Z) the DNA hairpin, causing motion of the trap.

4.3. Direct observation of UvrD stepping behavior

4.3.1. Motor step size of UvrD

The resolution of our optical tweezers assay allowed us to observe the motor steps of UvrD directly during unzipping and re-zipping activities (**Figure 4.2a**). A pairwise distribution analysis showed a predominant step size of ~3 bp across all ATP concentrations (**Figure 4.2b**). For more information on the stepping behavior, we located individual motor steps of UvrD by fitting our selected segments to an algorithm developed by Kerssemakers et al. (206) (red lines, **Figure 4.2a**). The resulting histograms revealed two distinct distributions: one for forward steps, and one for backward steps that were consistent for all [ATP] (**Figure 4.3**). Both distributions have a mean of ± ~3 bp (forward: 3.13 ± 1.23 bp, backward: -3.35 ± 1.56 bp; errors are s.d.), corroborating the value obtained from pairwise distance analysis. It is interesting that the distribution for backward steps is peaked near 3 bp, as this seems to suggest that UvrD actively steps backward by 3 bp rather than passively sliding backward by a random amount as seen in other helicases (111). An active backward step is consistent with UvrD's strand-switching behavior, where the protein translocates away from the fork junction in an ATP-dependent process (29).

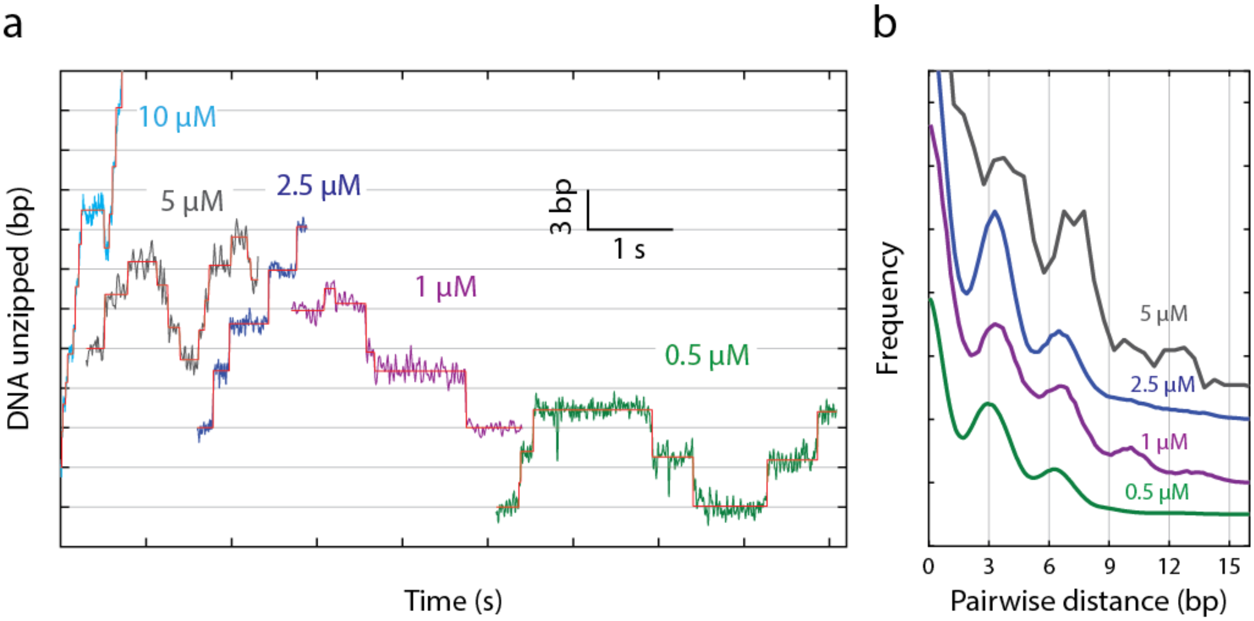


Figure 4.2: ATP-dependent stepping behavior of UvrD. (**a**) Representative time traces showing forward and backward steps of UvrD at different [ATP]. Red lines: result of fitting steps to time traces (206). Data collected with DNA construct under 10-15 pN force. (**b**) Pairwise distributions across [ATP]. Colors correspond to that of the traces in (**a**). Pairwise distribution for 10 μM data not shown due to high noise.

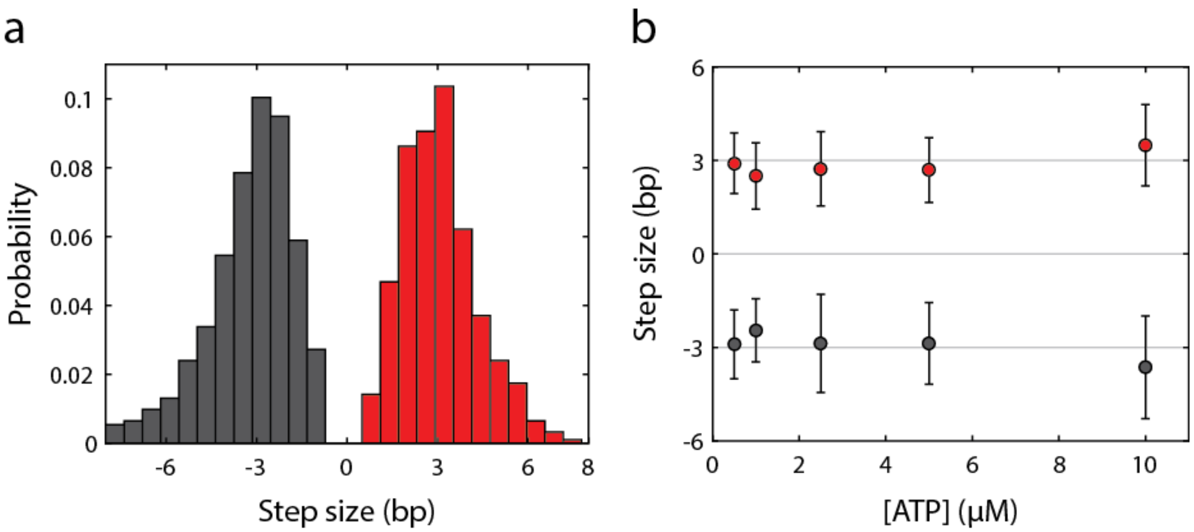


Figure 4.3: Step size of UvrD from step-fitting algorithm. (**a**) Histogram of step sizes obtained step-fitting algorithm for all [ATP] (*red*: forward, N = 416 steps; *grey*: backward, N = 423 steps). (**b**) Dependence of step size on [ATP] (*red circles*: forward steps; *grey circles*: backward steps). Error bars are S.D.

4.3.2. Dwell times

For more information on the mechanism of UvrD unzipping, we analyzed the amount of time the helicase dwells before taking a step. The distributions of dwell times are shown in **Figure 4.4a**. The mean dwell times followed Michaelis-Menten dependence on [ATP] (**Figure 4.4b**) with v_{max} = 17 ± 2 steps s^{-1} and K_m = 4.5 ± 0.8 μM. This v_{max} value agrees well with the 18.6 ± 1.3 steps s^{-1} from one single-turnover bulk study at saturating [ATP] (199), but not with the 95 ± 3 s^{-1} from a different bulk assay (29).

The shape of the dwell time distributions provides further information about the kinetic events taking place within each dwell. We quantify this using the ratio of the squared mean to the variance, $n_{min} = \frac{\mu^2}{\sigma^2}$, which is the inverse of the "randomness parameter" introduced by Block and co-workers (207,208). This parameter comes from a comparison between the predicted mean and variance of the gamma distribution ($\mu = n_{min}\theta$ and $\sigma^2 = n_{min}\theta^2$, respectively), and provides us with the minimum number of "hidden" kinetic events present within each dwell. A value of unity implies that the dwell time distribution follows a single exponential function, and that a single rate-limiting event governs the stepping process. In contrast, a value greater than unity tells us that the dwell time distribution follows a gamma function, and multiple hidden Poissonian steps exist with similar rate constants.

Our data reveals that n_{min} begins close to unity at low [ATP] (**Figure 4.4c**), but then increases to a value of ~1.5 near 5 μM ATP. At low [ATP] we measure an n_{min} of <1, which may result from false events that are shorter than our temporal resolution, skewing the distribution. A value of unity implies a single kinetic event that would likely correspond to binding of an ATP molecule from solution. At higher [ATP], we measure an n_{min} of ~1.5, which suggests multiple kinetic events. At saturating [ATP] we expect that protein translocation will become the rate-limiting event instead of ATP binding, which will likely cause n_{min} to drop to a lower value.

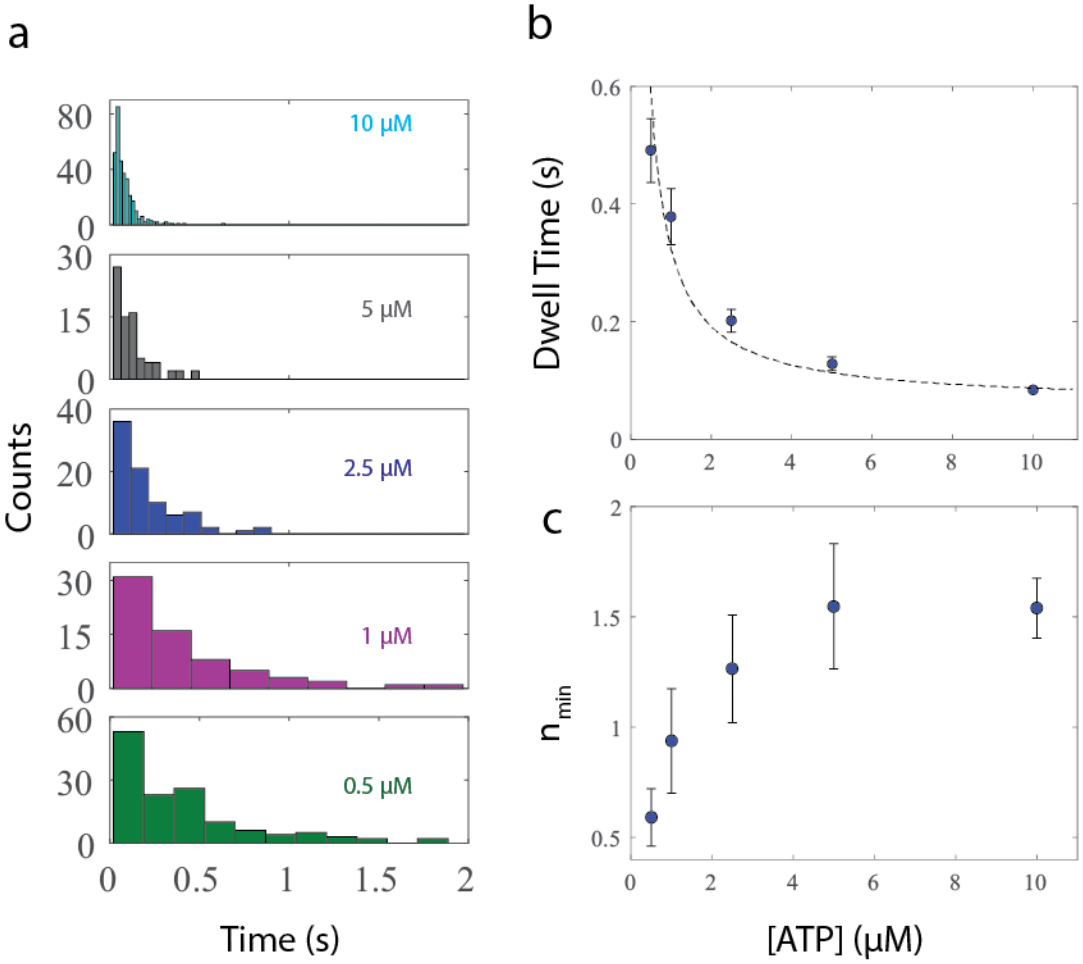


Figure 4.4: Step dwell times. (**a**) Histograms of dwell times across [ATP]. Colors correspond to those in **Figure 4.2**. (**b**) Effect of [ATP] on dwell time before taking a step. *Black dotted line*: fit to Michaelis-Menten kinetics. (**c**) Number of "hidden" kinetic steps across [ATP]. Error bars represent s.e.m.

4.4. Sub-steps of UvrD

The large standard deviations observed for the step sizes (1.23 and 1.56 bp for forward and backward steps, respectively) was significantly larger than the error associated with individual steps (0.47 bp, averaged across all steps), suggesting that the spread of step sizes was not solely due to experimental noise.

It is interesting to note that a very similar step size (3.6 bp) and standard deviation (1.3 bp) were observed for the SF2 helicase NS3 (196), and these distributions were later found to be comprised of a hierarchy of step sizes (198). Consistent with this picture, we observed numerous instances in which UvrD appeared to take smaller or larger steps (**Figure 4.5a**). To determine whether these constitute discrete sub-steps or are simply the result of measurement noise, we constructed a modified histogram using a kernel density estimator (**Figure 4.5c**) using a Gaussian kernel for each step size. The probability density for each step was calculated using $p = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-s)^2}{2\sigma^2} \right]$, where s is the size of the step and σ is its standard error. By this method we can observe the presence of multiple peaks below 3 bp for both forward and backward steps, suggesting that 3 bp is not the elemental step size of UvrD. The peaks in the distribution sometimes appear at integral numbers of base-pairs, while some peaks appear at half-integral values (*e.g.* 2.5 μM ATP case in **Figure 4.5c**). Measuring a half-integral step size is possible in our assay if only one of the two nucleotides in a base-pair is released after being unzipped. That is, UvrD may unzip a base-pair, but then sequester one of the nucleotides while releasing the other, as observed for NS3 (198).

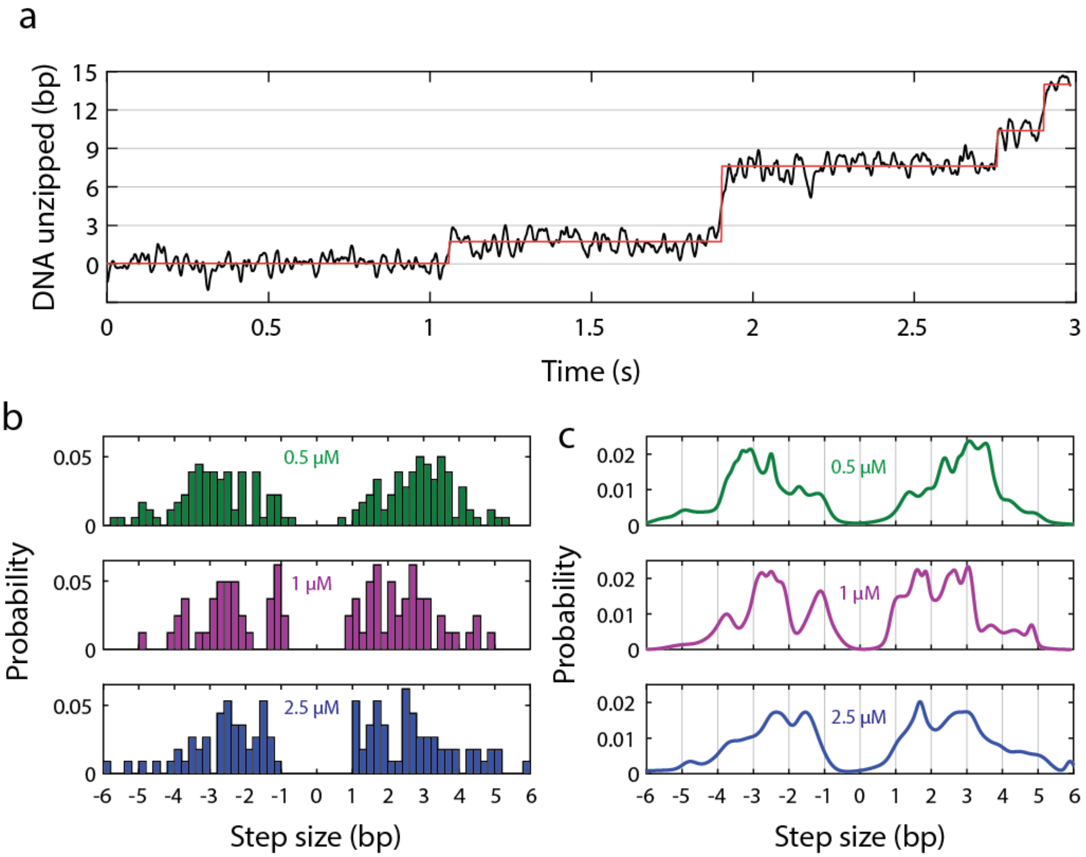


Figure 4.5: Sub-steps of UvrD. (**a**) Time trace (at 2.5 μM ATP) showing multiple step sizes, overlaid with fitted steps (*red*). (**b**) Histograms of step sizes for 0.5, 1, and 2.5 μM ATP. (**c**) Kernel density plots corresponding to the histograms in (**b**).

We considered that the kinetic processes underlying large motor steps may be different from that of small ones. If the larger motor steps of UvrD are comprised of sub-steps, each large step may contain multiple kinetic transitions within it. We therefore investigated the dwell-times for motor step sizes (**Figure 4.6a**). Our analysis reveals a slight correlation between the two: larger step sizes tend to be preceded by longer dwells. This is consistent with the hypothesis that there are hidden kinetic transitions within larger steps.

The dwell-time distributions for small and large steps demonstrate different dependence on [ATP]. We separated the dwell times that preceded small (|s| < 2 bp) and large steps (3 < |s| < 5 bp) and looked at the number of hidden kinetic events, n_{min}, for each [ATP] (**Figure 4.6b** and **c**). For small forward steps, n_{min} appears to remain constant at 1 across [ATP], implying a single rate-limiting event governs the process, and that [ATP] does not influence the number of these that take place (**Figure 4.6b**). Since the dwell times for small steps depend on [ATP] (**Figure 4.6c**), this suggests that the rate-limiting event is binding of ATP from solution. In contrast, n_{min} for large forward steps is not constant across [ATP], but reaches a peak at intermediate concentrations. At limiting [ATP] the rate-limiting event is likely binding of ATP molecules themselves, and at saturating [ATP] the rate-limiting event is likely the translocation of the protein. At intermediate [ATP], the rates of ATP binding and protein translocation become equally rate-limiting. Since n_{min} appears to reach a maximum near 3, this could mean that completion of one large step requires binding and hydrolysis of 2 or 3 ATP molecules, followed by protein translocation.

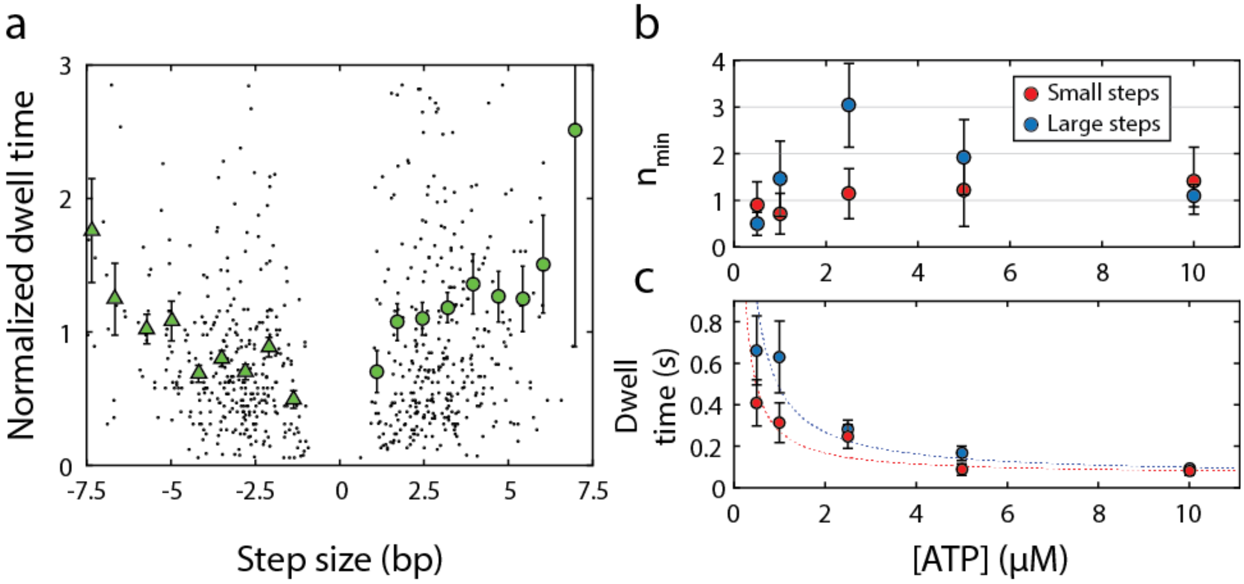


Figure 4.6: Effect of step size on dwell times. (**a**) Dependence of step size on dwell times. Dwell times for each [ATP] are normalized by their respective means (*black dots*: dwell times for individual steps; *green circles*: means of individual dwells). (**b**) [ATP]-dependence of of n_{min} for small (*red*) and large (*blue*) forward steps. (**c**) [ATP]-dependence of dwell times for small and large forward steps. Error bars for all panels are s.e.m.

4.5. Stepping mechanism

Put together, our results paint a picture for the stepping mechanism of UvrD. Small forward steps are governed by a single ATP-dependent kinetic event, which likely corresponds to 1 ATP molecule per base-pair unzipped. In contrast, large forward steps are governed by ~3 kinetic events at intermediate concentrations of ATP, which could result from 2-3 ATP binding and hydrolysis events followed by a protein translocation event that releases the unzipped base-pairs. At saturating [ATP], the release of unzipped nucleotides may become rate-limiting, such that the steps involving breakage of base-pairs no longer contribute to the observed dwell times. Such an ATP-coupling stoichiometry is consistent with bulk and structural studies (203,204). This mechanism is similar to the “spring-loaded” model proposed previously for the SF2 helicase HCV NS3 (197). NS3 appeared to unzip RNA in 3-bp increments with 3 kinetic events, which was proposed to result from the build-up of tension within the protein’s two RecA-like motor domains. The structure of the helicase revealed a tryptophan residue within these domains that may act as an anchor to stop translocation of the unzipped nucleotides until sufficient build-up of tension forces it to release them. It was noted that other helicases, notably UvrD, also have an aromatic residue in this position (tyrosine for UvrD) that may serve the same function (197).

However, this spring-loaded model does not account for the behavior of large steps at low [ATP]. We observe a single rate-limiting event for these circumstances, yet this seems to suggest that a single ATP binding and hydrolysis event is necessary to unzip multiple base-pairs. A possible solution to this is that the unzipping of base-pairs and release of nucleotides are decoupled from one another. UvrD may not release nucleotides immediately after they have been unzipped, but may instead release several after a later round of unzipping. This model was also proposed for NS3 helicase (198). NS3 appeared to unzip RNA in 0.5-bp increments, and release the nascent nucleotides asynchronously. An apparent 0.5-bp step size was proposed to arise from one of the two nascent nucleotides being released while the other was sequestered by the protein, to be released at a later round. UvrD may also sequester nascent nucleotides, although it is unclear how it may do so. A patch of the protein on the 2A domain near the fork junction contains a higher density of positively-charged amino acids, and so this may serve as a secondary binding site for unzipped nucleotides. This strand-sequestration mechanism is depicted schematically in **Figure 4.7**.

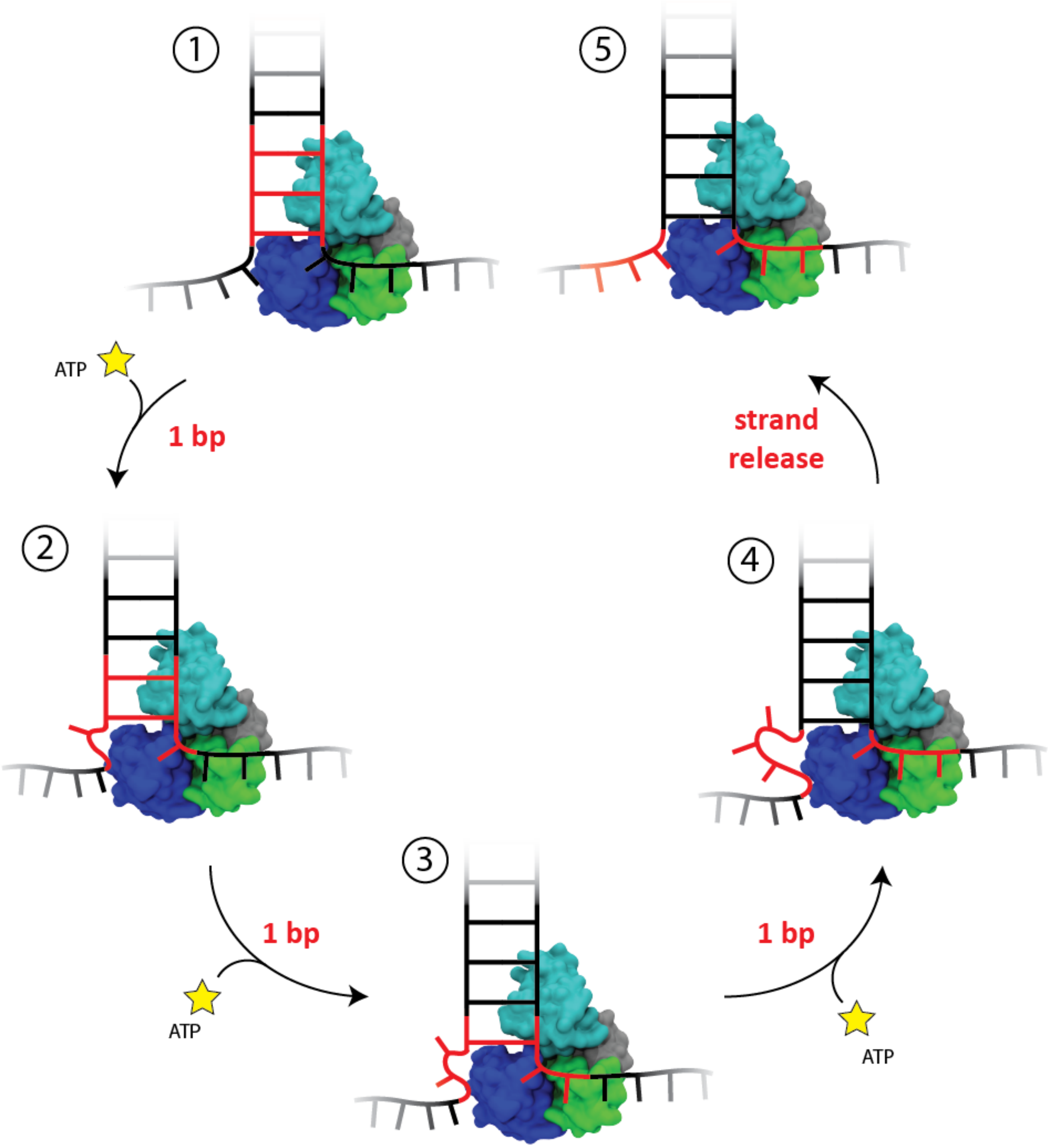


Figure 4.7: Model of stepping mechanism by asynchronous strand release. By this model, UvrD unzips one base-pair per ATP molecule hydrolyzed (**1** to **4**), but does not release the strands immediately. Nucleotides from the displaced strand remain bound to part of the protein, and are released in a manner asynchronous with unzipping (**4** to **5**). In this example, three nucleotides are released in the last step, leading to a measured step size of 1.5 bp.

Chapter 5

Determinants of strand-switching behavior in superfamily 1 helicases*

In this chapter we will investigate the strand-switching behavior of SF1 helicases, using the structurally-related helicases UvrD and Rep from *E. coli*. Much of the work in this chapter was part of a published study that was done in collaboration between our lab and those of Profs. Taekjip Ha (University of Illinois) and Timothy M. Lohman (Washington University School of Medicine) that investigated key structure-function relationships in UvrD related to its strand-switching behavior. The primary work was done by Dr. Matthew J. Comstock, and so I will not focus on the main conclusions of that study. Instead, the main conclusions from that work will form the background for this chapter, and we will probe deeper into UvrD (and Rep) strand-switching behavior in an attempt to answer the question: what factors influence UvrD to switch from one behavior (unzipping or re-zipping) to the other? We begin with my contribution to the aforementioned publication, examining what happens when UvrD encounters energetic barriers to unzipping by altering the base-pair stability in the DNA substrate. Then, we provide preliminary data on the role of the critical 2B domain using a mutant of Rep that lacks 2B altogether. Finally, we present data showing the effect of having free ssDNA on the 5' side of the fork junction on the transition from re-zipping to unzipping.

---

*Some of the work in this chapter has been published as:

Comstock, M. J., Whitley, K.D., Jia, H., Sokoloski, J., Lohman, T.M., Ha, T., and Chemla, Y.R. (2015). "Direct observation of structure-function relationship in a nucleic acid-processing enzyme." Science **348**(6232): 352-354.

5.1. Background

5.1.1. Effect of oligomerization on strand-switching

Many monomeric helicases are known to form functional self-interactions (9) that modify their activities. The two SF1 helicases already discussed, UvrD and Rep, are believed to form functional dimers or higher-order oligomers, and this oligomerization appears to be critical for their activity (209,210). While a UvrD monomer is known to translocate along single-stranded DNA (ssDNA) in a 3'-5' direction (28,200,201,211), studies have shown that duplex DNA unzipping requires at least a UvrD dimer (200,209,212-215).

Our previous study using optical traps to monitor unzipping of a hairpin observed two distinct behaviors under conditions in which either monomer or dimer activity could be observed, termed ‘frustrated’ (<20 bp unzipped) and ‘long-distance’ (>20 bp unzipped) (112). During frustrated activity, UvrD rapidly switched between unzipping and re-zipping behaviors (**Figure 5.1A**, lower panel), and this typically repeated many times before UvrD dissociation. In contrast, UvrD motion during long-distance activity was far less repetitive, and switching between unzipping and re-zipping behaviors occurred more slowly (**Figure 5.1B**, lower panel).

This study also revealed that the two distinct behaviors were correlated with protein stoichiometry (monomer vs. dimer) by performing a fluorescence counting experiment (**see Section 1.3.2**) using the fleezers instrument (**Figure 5.1C inset**) (112). In this assay, helicase activity was monitored with the optical traps while protein stoichiometry was determined by counting the number of photobleaching steps from labeled UvrD monomers. This assay revealed that monomers tended to exhibit frustrated activity, while dimers tended to display long-distance activity. This was consistent with previous studies showing that UvrD dimers are required for long-distance unzipping (209,214), and it also demonstrated that monomers are capable of unzipping a limited amount of DNA under tension. These results suggest that dimerization may result in long-distance unzipping activity by limiting the rate of switching from unzipping to re-zipping behaviors.

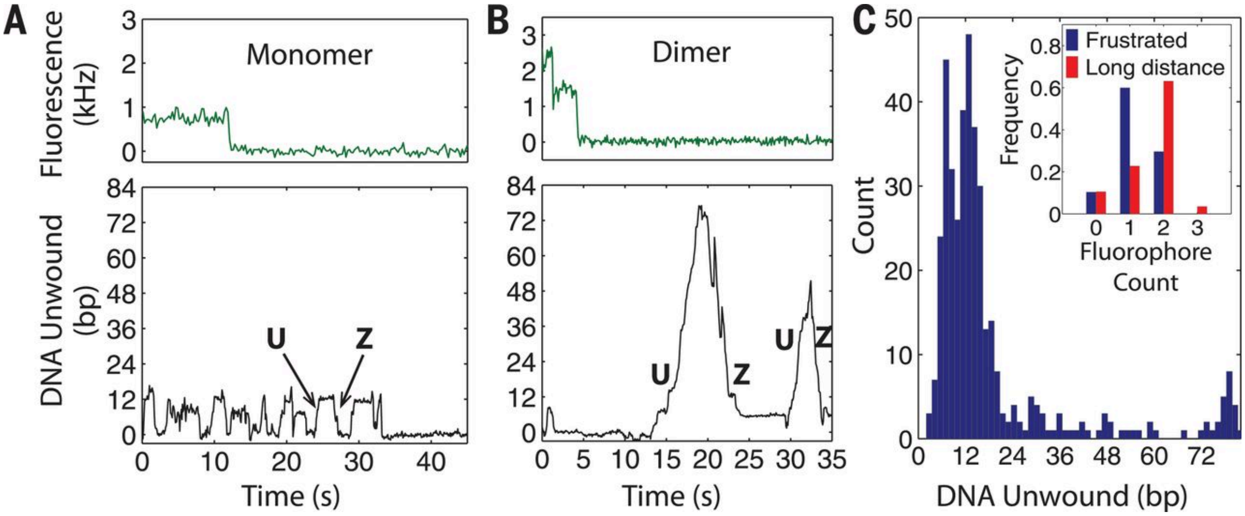


Figure 5.1: Effect of stoichiometry on UvrD unzipping behavior. Representative time traces of unzipping activity for a UvrD monomer (**A**) and dimer (**B**). Upper panels: Fluorescence photobleaching from a monomer (**A**) and a dimer (**B**). Lower panels: Simultaneous measurements of hairpin unzipping (U) and re-zipping (Z). The UvrD monomer **Figure 5.1 (continued):** displays frustrated unzipping (**A**), whereas the dimer displays long-distance unzipping (**B**). (**C**) Histogram of the maximum number of base-pairs unzipped per attempt, showing frustrated (<20 bp) and long-distance (>20 bp) unzipping. Inset: Distribution of fluorophore count for frustrated (*blue*) or long-distance (*red*)

unzipping attempts. All data in this figure were collected and analyzed by Matthew J. Comstock. Reproduced with permission from (112).

5.1.2. Role of the 2B sub-domain in SF1 helicases

All SF1 helicases share a similar structure consisting of four subdomains (**Figure 5.2**). The helicase motor core is comprised of the RecA-like 1A and 2A subdomains that bind ssDNA and contain a site for binding ATP. The duplex is contacted upstream of the fork junction by the 1B and 2B subdomains, which appear to have regulatory roles (203,216-219). Some structural and biochemical studies have shown that swiveling of the 2B domain is coupled to ATP binding and hydrolysis, providing evidence that the 2B domain plays an active role in the translocation and unzipping mechanism of UvrD (203,219). However, mutants of Rep lacking the 2B domain (RepΔ2B) actually unzip more DNA than wild-type (220,221), suggesting instead that this domain is not required for unzipping activity, but rather has an auto-inhibitory role.

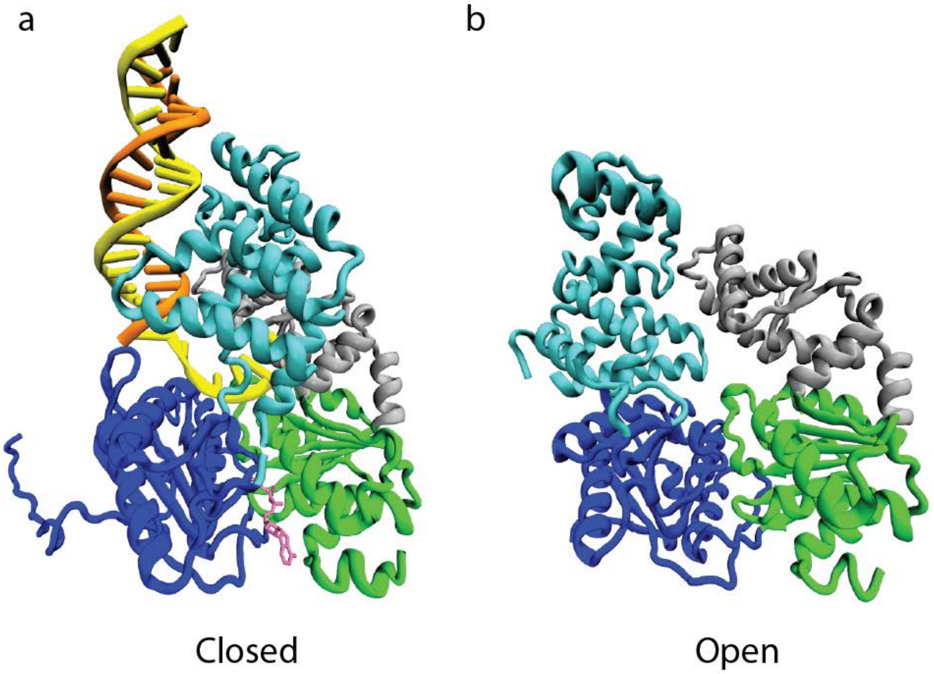


Figure 5.2: Structure and conformations of UvrD. (**a**) Crystal structure of UvrD in the 'closed' state, bound to a DNA fork junction (PDB entry 2IS4). Colors represent different subdomains: 1A (*green*), 2A (*blue*), 1B (*grey*) and 2B (*cyan*). The non-hydrolyzable ATP analog AMPPNP is shown in pink between the 1A and 2A subdomains. (**b**) Crystal structure of UvrD in the 'open' state (PDB entry 3LFU). Figure rendered using VMD (131).

Previous studies have shown that the 2B subdomain of UvrD (203,219), Rep (216), and PcrA (217) can exhibit two orientations—'open' and 'closed' relative to the other domains (**Figure 5.2**). Speculation has arisen over which of the two corresponding helicase conformations is the actively unzipping form. Evidence for the open conformation being the active state came from a mutant of UvrD containing a 2-amino acid substitution between the 2B and 1B subdomains (both Asp → Ala) that was super-processive. The mutations were hypothesized to weaken the interaction between the two domains, thus forcing the protein to adopt a more open conformation (222). In contrast, evidence for the closed conformation being the active state came from a study using engineered versions of Rep that locked it into either the open (RepY) or closed (RepX) conformation by chemical crosslinking. This revealed that the permanently-closed form was super-processive with a higher stall force than any known helicase, while the permanently-open form showed only modest processivity (223).

Our previous study also provided evidence that the closed conformation is the actively unzipping form (112). In this study, optical trapping was combined with single-molecule FRET to observe the conformational state and helicase activity of UvrD simultaneously. This assay was performed by labeling two subdomains of UvrD (1B and 2B) with a FRET donor-acceptor pair such that a low FRET efficiency corresponded to the ‘open’ state while a high FRET efficiency corresponded to the ‘closed’ state (**Figure 5.3A**). The unzipping/re-zipping activities were then observed by the motion of the traps while conformational state was simultaneously observed by the FRET efficiency (**Figure 5.3B** and **C**). This assay revealed that UvrD is in the ‘closed’ state when it unzips DNA, and is in the ‘open’ state while DNA is re-zipped (**Figure 5.3D**).

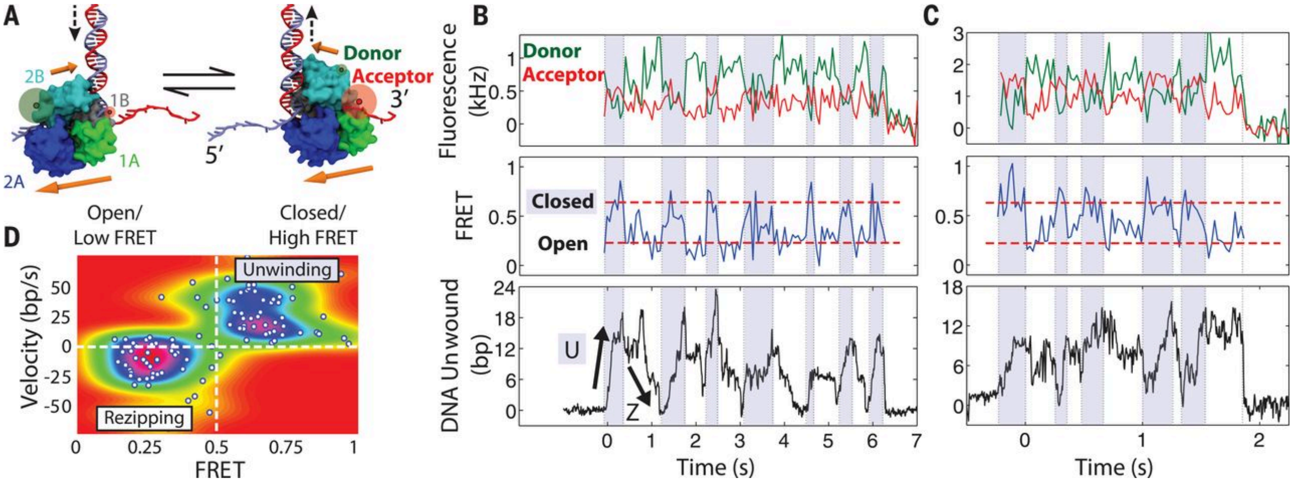


Figure 5.3: Simultaneous observation of UvrD helicase activity and conformational state. (**A**) Location of donor and acceptor fluorophores for FRET measurement and model of UvrD conformational switching. Upper (and lower) orange arrows denote 2B domain orientation. (**B** and **C**) Representative time traces of monomeric UvrD conformation and activity. Upper panels: Donor (green) and acceptor (red) fluorescence intensity. Middle panels: Corresponding

Figure 5.3 (continued): FRET efficiency showing UvrD reversibly switching between open (low FRET) and closed (high FRET) conformations (*dashed red lines*). Shaded and unshaded areas denote high- and low-FRET intervals, respectively. Lower panels: Simultaneous measurements of unzipping (U) and re-zipping (Z) of the DNA hairpin. (**D**)

Correlation between UvrD activity and conformation. The mean FRET efficiency and mean UvrD velocity determined over each time interval are plotted. The color map represents the probability distribution of FRET state and activity. All data in this figure were collected and analyzed by Matthew J. Comstock. From (112). Reproduced with permission from AAAS.

This finding led to a simple model to explain the correlation between the 2B domain orientation and UvrD unzipping/re-zipping activity (**Figure 5.3A**). In this model, the 2B domain acts as an anchor to hold UvrD at the fork junction while the 1A-2A motor domains switch from one ssDNA strand to the one on the opposite side of the duplex. In the 'closed' state, the motor domains are on the 3' side of the junction, and translocate into the fork junction, resulting in duplex unzipping. In the 'open' state, the motor domains are on the 5' side of the junction, and translocate away from the fork junction, allowing the duplex to re-zip in its wake. By this model the 2B domain facilitates strand-switching behavior, although it is unclear what determines the time or location of such switching.

In this chapter, we will investigate the determinants of UvrD strand-switching. We first demonstrate that the helicase tends to switch from unzipping to re-zipping when it encounters an energetic barrier due to high base-pair stability. We also probe deeper into the role of the 2B domain by observing the behavior of a Rep mutant that lacks this domain altogether. Finally, we examine the transition from re-zipping to unzipping when ssDNA is freely available on the 5' side of the fork junction while UvrD translocates away from the junction.

5.2. Effect of DNA base-pair stability

5.2.1. Helicase assay

All of the experiments performed in this section used hairpin constructs with a 19 dT protein loading site (see **Appendix A.1.2**). We have found from stoichiometry measurements that this poly-dT length favors monomer loading. All experiments in this section were performed using the dipping assay as described in **Section 4.2.2** to prevent new monomers of UvrD from loading after unzipping is initiated.

Trap-FRET assay: In one set of experiments we used a mutant UvrD with a FRET donor-acceptor pair to observe the protein's conformational state (see **Appendix A.7.1**). Since this assay involved fluorescence, we added a triplet-state quencher to our buffer to prevent fluorophore blinking (1 mg/ml Trolox; see **Appendix A.4**) (110). To observe fluorescence during the dipping assay, we turned on the fluorescence excitation laser when the protein-DNA complex entered the ATP channel. The resulting fluorescence and trap signals were recorded simultaneously. Data from the optical traps was saved at 66 kHz and boxcar averaged to 267 Hz. Fluorescence data were initially saved at 10 ms, and further integrated for analysis and plotting to 20-50 ms. To determine the protein conformational state, we computed the FRET efficiency as

$E = I_A / (I_A + I_D)$, where I_A and I_D are the fluorescence intensities of the acceptor and donor molecules, respectively, after subtracting off the background fluorescence of each.

5.2.2. Destabilizing the DNA hairpin with applied force

In order to determine the effect of base-pair stability of the DNA substrate on UvrD strand-switching behavior, we first quantified the effect of force, which destabilizes the DNA substrate by lowering the barrier to base-pair unzipping. To do this, we measured the distance unzipped by UvrD as a function of applied tension, ranging from 4.5 to 13.5 pN. **Figure 5.4b** shows the number of consecutive base-pairs unzipped for individual UvrD molecules before a reversal in direction as a function of tension. We observed no effect of tension on unzipping activity below 11 pN (**Figure 5.4b**). The number of consecutive base-pairs unzipped remained constant and frustrated unzipping was almost always detected in this force range. Below 4.5 pN, our assay did not have the resolution to detect short-distance frustrated unzipping (~15 bp) reliably. At low tensions, the destabilizing effect of force on the hairpin base pairs is minimal. For instance, at 5 pN the applied force contributes only ~10% of the average base-pairing free energy (**Appendix B.4**). This is consistent with the frustrated behavior of UvrD monomers observed from fluorescence counting experiments (**Section 5.1.1**).

When force was increased above 12 pN, near the hairpin opening threshold of ~15 pN, we found an increased likelihood of observing long-distance unzipping with UvrD. The same monomer could be made to switch from frustrated to long-distance unzipping activity by suddenly increasing the tension to this range (**Figure 5.4a**). We confirmed that this strand-switching behavior corresponded to the conformational change of a monomer by performing the simultaneous trap-FRET experiment described above (**Section 5.1.2**) under high force (~13.5 pN; **Figure 5.4c**, **d**, and **e**). At forces ≥12 pN, the destabilizing effect of tension is more significant as the applied force contributes ≥50% of the average base-pairing free energy (**Appendix B.4**). These results therefore demonstrate that monomers of UvrD can unzip long distances before switching strands if the energetic barrier to unzipping is lowered. This in turn implies that high energetic barriers increase the probability of switching from unzipping to re-zipping.

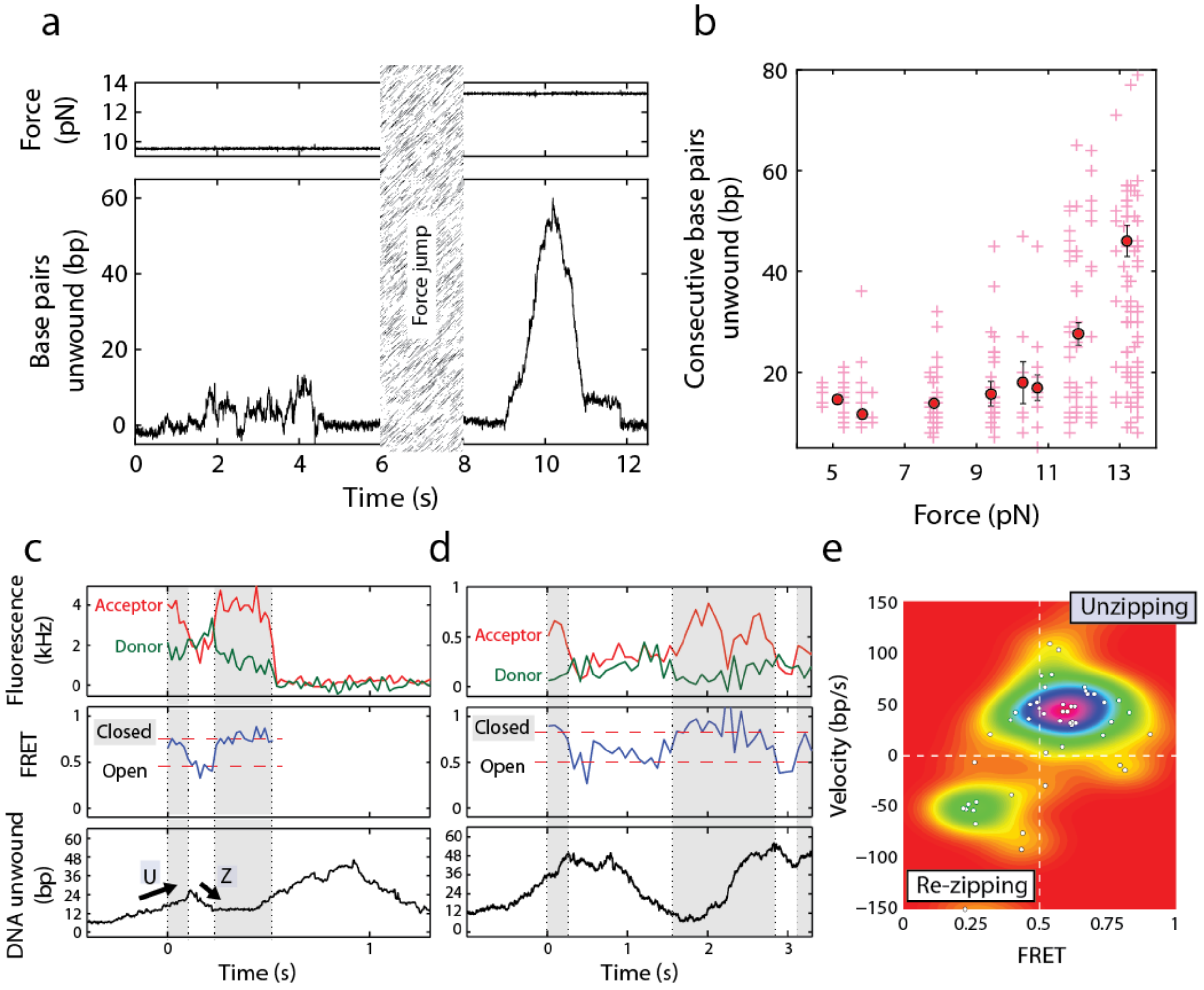


Figure 5.4: Effect of force on distance unzipped before strand-switching. (**a**) Example trace of UvrD switching from frustrated to long-distance unwinding (lower panel) upon increased tension (upper panel). (**b**) Positions along hairpin sequence where UvrD reversed direction vs. tether tension. Open symbols represent mean reversal positions (error bars s.e.m.). (**c** and **d**) Representative time traces of monomeric UvrD conformation and activity under high force conditions (~13.5 pN). Upper panels: Donor (*green*) and acceptor (*red*) fluorescence intensity. Middle panels: Corresponding FRET efficiency showing UvrD reversibly switching between open (low FRET) and closed (high FRET) conformations (*dashed red lines*). Shaded and unshaded areas denote high- and low-FRET intervals, respectively. Lower panels: Simultaneous measurements of unzipping (U) and re-zipping (Z) of the DNA hairpin. (**e**) Correlation between UvrD activity and conformation under high force conditions. The mean FRET efficiency and mean UvrD velocity determined over each time interval are plotted. The color map represents the probability distribution of FRET state and activity. From (112). Modified with permission from AAAS.

5.2.3. Effect of hairpin sequence

More insight can be gained if we consider that the energetic barrier to unzipping is not constant across the hairpin, but varies with sequence. Regions of higher GC content have higher base-pair stability, and will present a larger barrier for UvrD monomers to unzip. We therefore measured the distance unzipped by UvrD using two different hairpin sequences (**Figure 5.5a**), and compared this to the probability $P_{closed}(p,F)$ that the hairpin remain closed at a given position p under tension F (**Appendix B.4**), similar to that described by Johnson et al. (224) and Qi et al. (111) . This probability essentially represents the energetic barrier due to sequence in a given region of the hairpin. As seen in **Figure 5.5c-f**, UvrD has a higher probability of switching strands in regions of high base-pair stability, although the average distance unzipped in a single attempt appeared to be independent of the hairpin sequence (**Figure 5.5b**). This is consistent with the experiments performed under force, showing that switches from unzipping to re-zipping tend to occur in regions of high base-pair stability.

Put together, these results support a simple model for strand-switching based on a kinetic competition between unzipping (stepping forward) and switching strands. At any given position along the hairpin, UvrD may either take a step forward or switch to the opposite strand to begin re-zipping. When the barrier for stepping forward is lowered, either by the application of force or by encountering an AT-rich region, the helicase is more likely to unzip more base-pairs before a strand-switching event. In addition to lowering the barrier to unzipping, long-distance activity could conceivably be achieved by raising the barrier to strand-switching. Since strand-switching appears to be associated with the orientation of the 2B domain, influencing this orientation may decrease strand-switching and lead to long-distance unzipping. This was observed for RepX helicase (223), and it is possible that dimerization of UvrD influences the orientation of 2B as well. In the next section we explore the consequences of removing the 2B domain entirely.

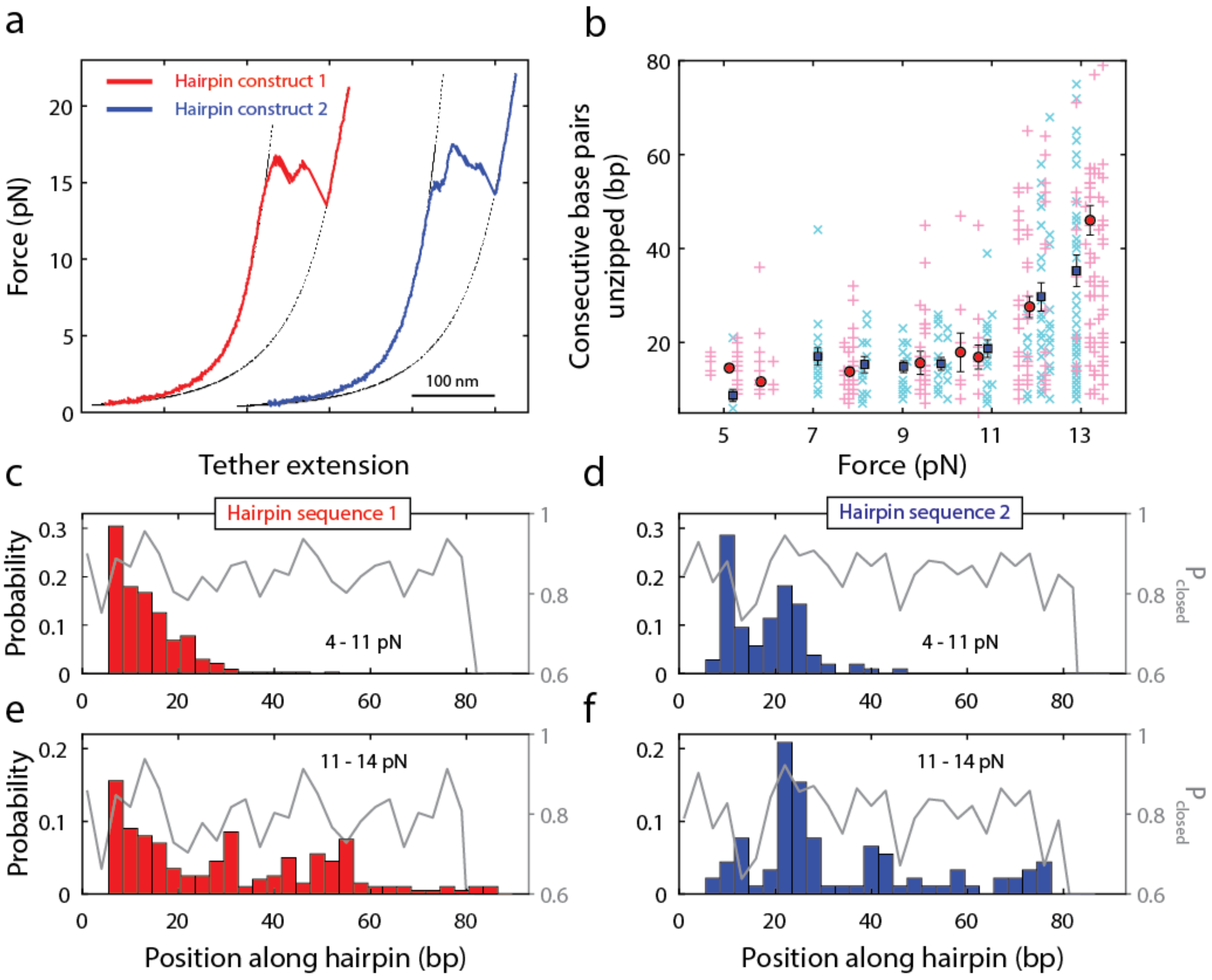


Figure 5.5: Effect of sequence on position of strand-switching. (**a**) Representative force-extension curves of hairpin sequences 1 (*red*) and 2 (*blue*) in experimental buffer. Black dotted lines represent the expected force-extension behaviors for the fully zipped and unzipped DNA hairpins. Hairpin sequence 2 reverses all A-T base-pairs in hairpin sequence 1 to G-C and vice versa. (**b**) Positions along hairpin sequence where UvrD reversed direction vs. tether tension (sequence 1: *magenta* '+', sequence 2: *cyan* 'x'). Open symbols represent mean reversal positions (sequence 1: *red circles*, sequence 2: *blue squares*) (error bars are s.e.m.). (**c**) - (**f**) Distribution of individual unwinding reversal events of (**b**): hairpin sequence 1, <11 pN (**c**) and >11 pN (**e**) and hairpin sequence 2, <11 pN (**d**) and >11 pN (**f**). The computed probability P_{closed} (224) of the hairpin remaining closed at a given sequence position is overlaid in grey for each sequence and force range. Figure modified from (112).

5.3. Removal of the 2B domain

More insight into the role of the 2B domain in strand-switching can be obtained by creating a mutant that lacks the domain altogether. Unfortunately, attempts to produce a 2B deletion mutant of UvrD (UvrDΔ2B) have been unsuccessful, possibly because an unregulated UvrD helicase is lethal to *E. coli*

(221). However, a 2B deletion mutant of the structurally homologous Rep helicase (RepΔ2B) has been successfully expressed (221), and so we here compare the activities of wild-type Rep (wtRep) and RepΔ2B. Here, we present preliminary data on the helicase activities of these two proteins using the hairpin assay.

5.3.1. wtRep and RepΔ2B assays

Both proteins were expressed and purified by the laboratory of Prof. Timothy M. Lohman as described previously (221). The DNA substrate used was the hairpin sequence 1 with either 19 dT or 38 dT protein loading site (**Table A.2**), and the buffer in which the experiments were performed was the same as that used for wtUvrD (see **Section 4.2.2**). For this preliminary investigation of helicase activity we performed the assay under multiple-turnover conditions – that is, new molecules of Rep were allowed to bind the substrate after unzipping was initiated. To accomplish this we used the same flow chamber described in **Section 4.2.2**, except both protein (6-9 nM) and ATP (1 mM) were present together in the top channel, while the bottom channel contained blank buffer. Rather than ‘dipping’ a formed DNA tether into the top channel to load protein and then moving back into the bottom to record helicase activity, we moved the newly-formed tether into the top channel only one time and then recorded activity.

5.3.2. Comparison of wtRep and RepΔ2B activities

Figure 5.6 depicts the helicase activities of both wtRep and RepΔ2B, showing striking differences. Both helicases appeared to unzip DNA in discrete ‘bursts’, but the bursts of each protein demonstrated very different characteristics. The wtRep bursts displayed frustrated unzipping (<20 bp), yet the RepΔ2B bursts nearly always resulted in the unzipping of the entire hairpin stem (~89 bp). We cannot infer the number of Rep monomers present on the DNA substrate during each burst, although these observations are consistent with previous studies showing that wtRep is incapable of unzipping DNA as a monomer (210) unless its 2B domain has been removed (220). Interestingly, RepΔ2B did not proceed to unzip the dsDNA handle adjacent to the hairpin after reaching the 5’ end of the nascent ssDNA.

The time between each unzipping burst (the ‘off’ lifetime) was measured from the time the hairpin fully re-annealed (returned to 0 bp) until the next increase in tether extension. We then converted these times into $P_{survival}$ (**Figure 5.6e** and **f**), as done in **Section 2.2.3**. The ‘off’ lifetime for wtRep was well-fit to a single exponential function with rate constant $k_{on} = 3.6\times10^{-3} \pm 1\times10^{-4}$ nM^{-1} s^{-1}, implying that a single rate-limiting event governs the process. This may correspond to the arrival of a monomer of Rep to the bare fork junction, or it may correspond to a second monomer joining an inactive one already at the junction to form a dimer. Our rate constant is more consistent with the rate constant for dimerization measured from bulk studies (210). The same ‘off’ lifetime of RepΔ2B, however, was poorly fit to a single exponential function, but well-fit to a that of a double exponential, with rate constants $k_1 = 7 \pm 1$ s^{-1} and $k_2 = 4.8\times10^{-2} \pm$

4×10^{-3} s^{-1}. The first rate likely corresponds to the repetitive unzipping of the hairpin by a single Rep helicase still bound to the junction and the second to binding of a new Rep monomer from solution. This would make the 'binding' rate constant $k_2 = k_{on} = 7.1\times10^{-3} \pm 6\times10^{-4}$ nM^{-1} s^{-1}, similar to that of wtRep. The rate constant k_1 likely corresponds to repetitive unzipping behavior of single RepΔ2B molecules, suggesting that Rep is capable of switching strands at the base of the hairpin stem even without the 2B domain.

The removal of the 2B domain appears to have a drastic effect on the strand-switching rate. During each burst, wtRep exhibited many reversals in direction mid-hairpin (**Figure 5.6a, b**), likely corresponding to strand-switching events. In contrast, RepΔ2B displayed very few such reversals mid-hairpin (~14% of 70 recorded unzipping attempts; **Figure 5.6c, d**), unzipping through the top of the hairpin nearly every time (when the hairpin is fully unzipped, the protein can translocate along the bare ssDNA to the other side of the hairpin without 'switching strands'). This is consistent with the finding that the 2B domain facilitates strand-switching behavior (112).

Although RepΔ2B displayed few reversals in the middle of the hairpin, it often returned to unzipping behavior after reaching the base of the hairpin stem (**Figure 5.6d**). Such behavior was likely the result of switching back to the original strand after encountering the dsDNA handle on the 5' side of the hairpin. This highlights a significant difference in strand-switching behavior on our hairpin construct: in the middle of the hairpin there is available ssDNA on both the 3' and 5' sides of the fork junction, but at the base of the hairpin there is no ssDNA available on the 5' side. This may alter helicase behavior at the base of the hairpin by forcing the helicase to switch to the 3' side. We next investigate the effect of including a 5' ssDNA overhang at the base of the hairpin stem.

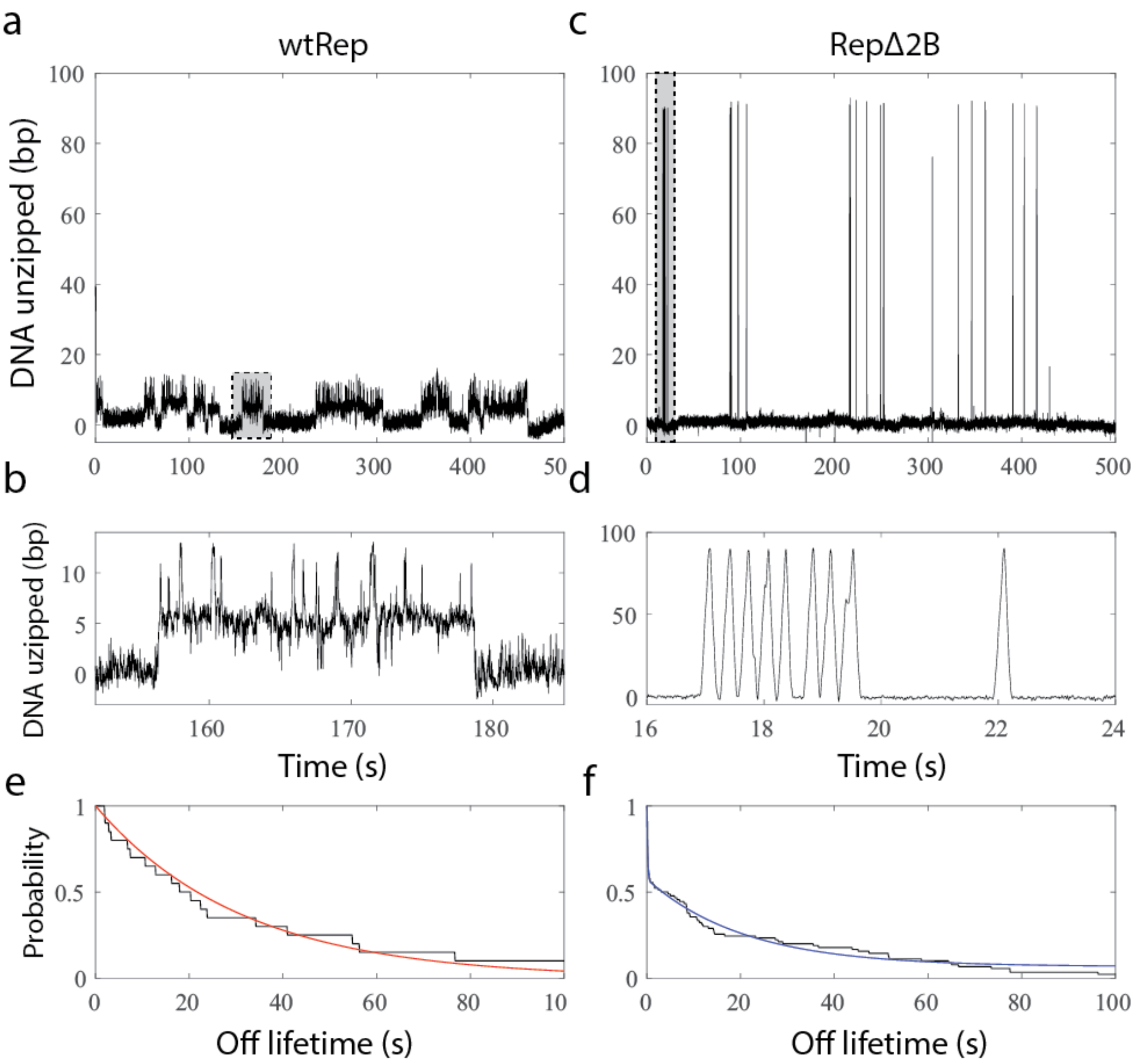


Figure 5.6: Comparison of wtRep and RepΔ2B activities. (**a**) wtRep displays discrete bursts of activity in which a limited number of base-pairs (~15 bp) are unzipped. (**b**) Zoom-in of a single unzipping burst (*grey shaded region* in (**a**)), demonstrating multiple mid-hairpin reversals in behavior. (**c**) RepΔ2B likewise displays discrete bursts of activity, unzipping the entire hairpin stem (~89 bp) nearly every time. (**d**) Zoom-in of a single unzipping burst of RepΔ2B (*grey shaded region* in (**c**)), demonstrating few mid-hairpin reversals, but multiple reversals at the base of the hairpin. (**e**) and (**f**) Distributions of lengths of time between bursts ('off' lifetimes) by wtRep and RepΔ2B.

5.4. Effect of 5' ssDNA overhang on switching from re-zipping to unzipping

Our discussion of strand-switching so far has primarily focused on the factors that influence the transition from unzipping to re-zipping. However, factors that influence the transition from re-zipping back to unzipping may be common in the biological systems in which UvrD and Rep participate. For example, during replication restart these helicases may encounter Okazaki fragments on the lagging strand as they translocate away from the fork junction during re-zipping behavior. Single-molecule studies have shown that Rep tends to release some DNA and 'snap back' to its original position when confronted with additional duplexes like this (25).

To investigate the role of the 5' side of the duplex on the strand-switching behavior of UvrD, we used three separate hairpin constructs where the length of available ssDNA on the 5' side of the hairpin was varied. The constructs had poly-dT overhangs on the 5' side of the hairpin (0 dT, 10 dT, and 38 dT; see **Table A.2**) in addition to the 3' poly-dT protein loading site. Using these constructs, we performed the dipping assay with UvrD as described in **Section 4.2.2**. **Figure 5.7a** shows characteristic traces of wtUvrD unzipping three different hairpin constructs, illustrating that the number of unzipping attempts per molecule decreased as the 5' overhang was extended (**Figure 5.7b**). This decrease may suggest that UvrD translocates away from the fork junction after switching to the 5' side of the hairpin, thus becoming 'trapped' far away from the 3' side when the 5' overhang is too long. It may also suggest that collision with a blockade such as duplex DNA influences the transition from re-zipping to unzipping. This latter case may be related to the behavior observed for Rep (25).

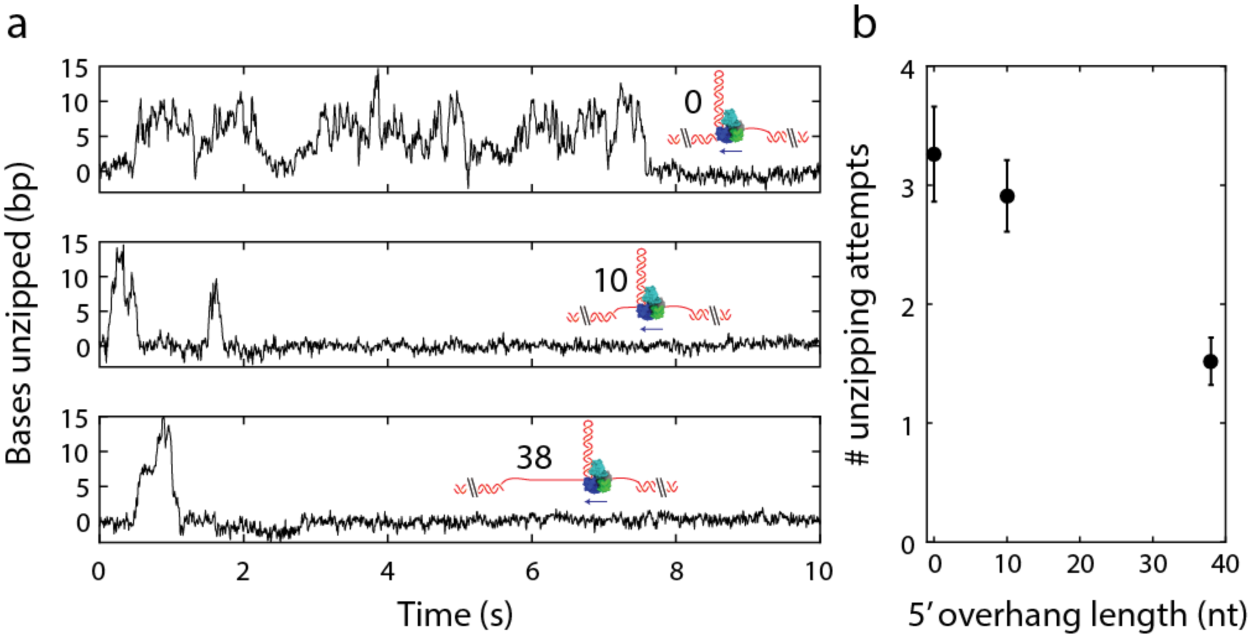


Figure 5.7: Repetitive unzipping behavior of UvrD on hairpin constructs with 5' dT overhangs. (**a**) Examples of repetitive unzipping behavior by UvrD on three different DNA hairpin constructs. Insets depict different hairpin constructs (*red*) with UvrD (*blue, green, cyan, and grey*) translocating 3' to 5'. With a 0 dT overhang on the 5' side of the hairpin, UvrD displays several attempts to unzip the hairpin after returning to the base of the hairpin. With longer 5' overhangs, the number of attempts to unzip decreases significantly. (**b**) Number of unzipping attempts by UvrD as a function of 5' overhang length. An unzipping attempt is only considered when the hairpin begins unzipping from 0 bp.

Chapter 6

Interaction between UvrD and MutL

In this chapter we place UvrD in its biological context by investigating its interaction with the DNA mismatch-repair protein MutL. We begin by describing methyl-directed mismatch repair in *E. coli*, emphasizing the role that MutL plays in stimulating UvrD to unzip hundreds of base-pairs. We then present preliminary data to elucidate the mechanism by which this stimulation occurs.

6.1. Background

6.1.1. Mismatch repair overview

In order to survive and pass on their genetic material to offspring, cells must preserve the integrity of their genomes. However, genomes are routinely damaged by external agents such as ultraviolet light and oxidative agents, and processes within the cell such as replication often introduce errors. Failure to correct damage can result in the introduction of deleterious mutations during the next cycle of DNA replication. When the ability of cells to repair DNA is disrupted, mutations can begin to accumulate in their genomes; in human cells, this often results in cancer. Since cells must repair up to 200,000 base pairs of DNA per day (225), it is unsurprising that there are numerous pathways within cells to mediate this damage. Each repair mechanism is responsible for correcting different types of damage. In *E. coli*, two noteworthy examples are methyl-directed mismatch repair (MMR), which primarily corrects mismatches in DNA that have arisen from errors in replication, and nucleotide excision repair (NER), which repairs a wide variety of damages (226). The detailed mechanisms of these processes are not well understood, despite their importance in DNA metabolism and relevance to medicine. About a fourth of all sporadic cancers have defects in MMR (227). In particular, genetic defects in MMR are linked to Lynch syndrome (a.k.a. hereditary non-polyposis colon cancer) (228), a genetic disease responsible for ~3-4% of all colorectal cancers, though another ~10% of colon tumors result from epigenetic silencing of MMR genes (229).

Mismatch repair is primarily used to correct errors arising from replication. MMR follows shortly after replication to correct errors that are either non-Watson-Crick base pairs or insertion-deletion regions, which arise due to slippage of the newly synthesized strand relative to the template (229). The mispairs that are most often bypassed by DNA polymerase are those that cause little distortion to the double helix, as the proofreading exonuclease on the polymerase usually removes the others (229). The MMR system follows to correct such errors, but it must do so quickly – the system must be able to distinguish the newly synthesized strand from the template strand, and only repair the former. In *E. coli*, MMR is methyl-directed: strand discrimination is accomplished by exploiting the fact that the daughter strand is transiently unmethylated. Dam methylase, which methylates d(GATC) sites on the daughter strand, lags behind replication by ~2 min (229). It remains unclear how strand discrimination is accomplished in eukaryotes, which do not methylate their genomes, or even in most other prokaryotes, which do not rely on methylation for this process (227). Since several MMR proteins interact with the β clamp in prokaryotes and its homolog PCNA in eukaryotes, which are oriented with respect to new/template strands, it has been suggested that this interaction plays the role that methylation does in *E. coli* (229). Interestingly, *E. coli* MMR proteins also interact with the β clamp (230), although the reason for this is unknown (231).

Mismatch repair in *E. coli* requires four overall steps (**Figure 6.1**) that are primarily carried out by the Mut proteins: MutS (recognition), MutL (strand discrimination), MutH, UvrD, exonuclease (strand removal), DNA polymerase III, and ligase (resynthesis). MutS is a homodimer that diffuses along double-stranded DNA (dsDNA), rotating along the helix contour until it reaches the mispaired site, where it changes conformation to a sliding clamp (232,233). Strand discrimination is accomplished by MutL, which binds to and diffuses with the MutS sliding clamp (232,233). Upon reaching a methylated d(GATC) site on the parent strand, MutL stimulates the endonuclease activity of MutH, which nicks the daughter strand at a single site. This nicked site can be on either the 5’ or 3’ end of the new strand relative to the mismatch, and can be located more than 1000 bp away due to the spacing of methyl groups (234). Strand removal begins when MutL loads UvrD (a.k.a. MutU) onto the DNA at the nicked site (235). After being loaded by MutL, UvrD unwinds DNA past the mismatch, followed by one of four exonucleases (ExoI, ExoVII, ExoX, or RecJ), which digests the unwound ssDNA of the daughter strand after stimulation by MutL (236). DNA polymerase III then re-synthesizes the digested strand, and DNA ligase completes the process by sealing the nick left by the polymerase.

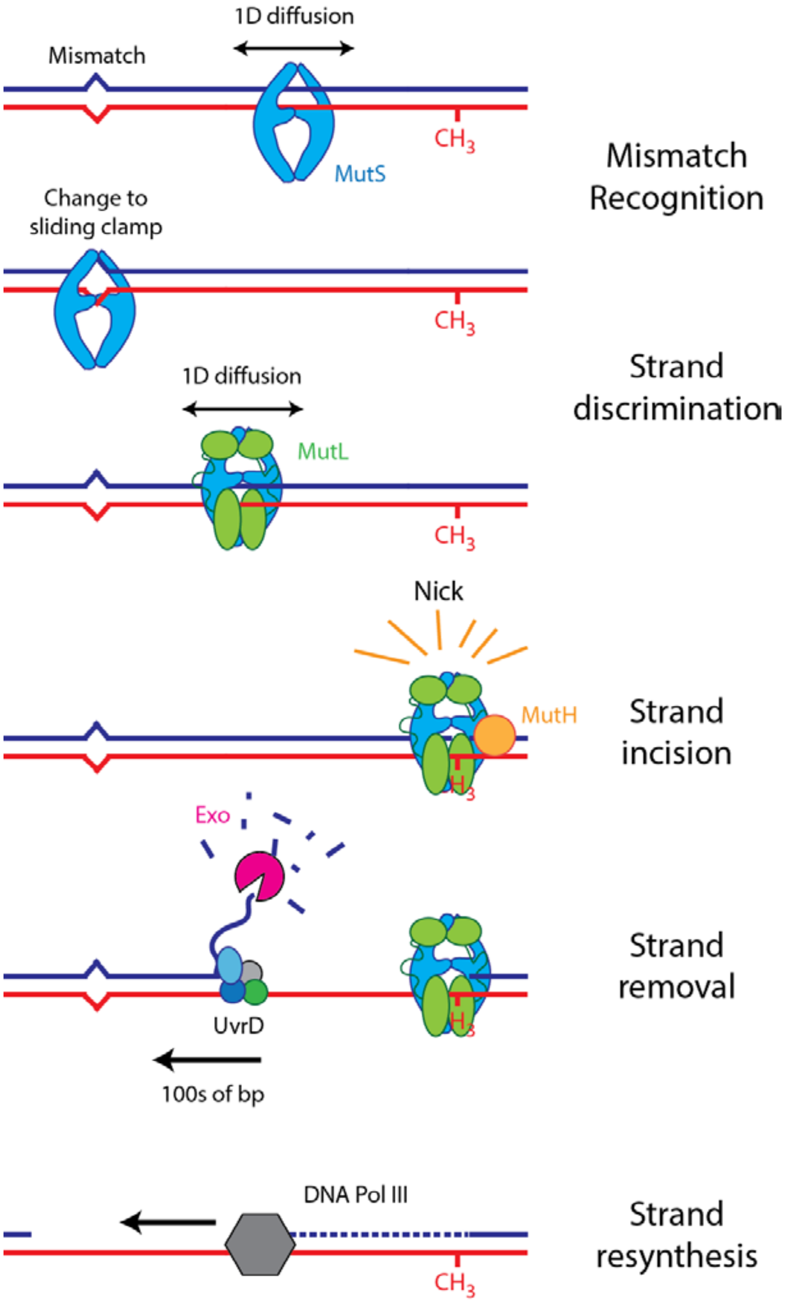


Figure 6.1: Overview of methyl-directed mismatch repair in *E. coli*. MutS (*blue*) diffuses along DNA behind the replication fork until it finds a mismatched site. MutS then undergoes a conformational change to a sliding clamp state, releases the mismatch, and MutL (*green*) binds and diffuses along with it. Upon locating a methylated d(GATC) site on the parent strand (*red*), MutL stimulates the endonuclease activity of MutH (*orange*) to produce a nick in the daughter strand (*blue*). MutL then stimulates UvrD (*blue, cyan, green, and grey*) helicase to unzip the mismatched segment as exonuclease (*pink pacman*) digests the daughter strand. Finally, DNA polymerase III (*grey hexagon*) re-synthesizes the strand, which is finally sealed by ligase (*not shown*).

6.1.2. MutL-UvrD interactions

MutL is the central player in MMR, connecting most of the major steps together by acting as a "molecular matchmaker" (237). MutL is involved not only in MMR, but also in a wide variety of other cellular systems as far ranging as apoptosis and somatic hypermutation (238). In prokaryotes, MutL forms a homodimer where each monomer contains two constitutively-dimerized C-terminal domains that interact with UvrD (239) and two N-terminal domains with weak ATPase activity (240,241) that interact with the MutS sliding clamp (233,242). The behavior of MutL within MMR has not been fully established, or even the number of MutL molecules involved – *in vivo* imaging has recently shown that multiple MutL molecules are present near a mismatch for every MutS monomer (243).

The interaction between MutL and UvrD is critical to successful mismatch repair. Since a mismatch can be more than 1000 bp away from the nearest methylation site (234), the MMR system must be capable of excising and resynthesizing very long tracts of DNA. In order to accomplish this, the helicase activity of UvrD is stimulated more than 10-fold by MutL (16) to unzip in the direction of the mismatch (235). This interaction is very specific: MutL stimulates UvrD, but not Rep (16), despite substantial sequence homology between the two helicases. It has further been discovered that MutL requires bound ATP (or a non-hydrolyzable analog) to stimulate UvrD helicase (and other MMR proteins) (244,245), although it is unclear at what point MutL hydrolyzes its bound ATP. A mutant of MutL that is incapable of hydrolyzing ATP (240) was unexpectedly observed to increase UvrD activity more than wild-type MutL, suggesting that the hydrolysis of ATP is required for a later regulatory role (245).

The mechanism by which MutL stimulates UvrD remains unknown. It has been proposed that this stimulation is based on UvrD stoichiometry. Three hypotheses exist concerning the number of UvrD monomers loaded by MutL: (1) MutL functions by continually loading UvrD monomers at the nicked site, as suggested by some single-turnover bulk experiments (246). (2) MutL stabilizes a dimer or higher order oligomer (244). This is consistent with the known high processivity of a UvrD dimer (112,209,214). (3) MutL activates the helicase activity of a monomer, possibly by acting as a processivity clamp (247).

6.1.3. Comparison with the role of UvrD in nucleotide excision repair

While MMR is primarily involved in correcting replication errors, NER acts on a variety of DNA damage types (226). In *E. coli*, damage is detected by UvrA, which locates sites of damage through an unknown mechanism, though it may involve bending DNA to look for soft spots. Upon locating a potential damage site, UvrB is recruited to verify the damage, and UvrC is then recruited to incise the DNA in two places, 3' and 5' of the damage, in contrast to the single incision made in MMR (248). UvrD is then recruited, although it is unknown how it binds, and with what stoichiometry (226). Although MutL is the only repair protein known to form specific interactions with UvrD, UvrAB is known to stimulate UvrD helicase (and not Rep helicase) activity ~4-fold at nicked duplexes (249).

Although involved in both repair pathways, the requirements for UvrD in NER are very different from those in MMR. Firstly, in NER the protein UvrC makes nicks on either side of the mismatch, which has been shown to increase UvrD unwinding efficiency (250). Secondly, the lengths of DNA excised in each case is dramatically different: in NER only ~12 bp must be unwound, compared to the >1000 bp in MMR.

6.2. MutL stimulates UvrD under multiple-turnover UvrD conditions

6.2.1. MutL-UvrD interaction assay

Protein expression. MutL was provided by the laboratory of Prof. Timothy M. Lohman. It was expressed and purified as described previously (244). Details can be found in **Appendix A.7.2**.

Gapped and hairpin DNA substrates. To ensure a long track length for processive unzipping by UvrD, the DNA substrate used was a gapped construct rather than a hairpin. This was synthesized using the same primers and protocol as in **Section 2.2.1**, but with an insert containing a 35 dT single-stranded region (**Table A.1**) to facilitate loading of multiple proteins.

Serial assays. The buffer used in these assays consisted of 10 mM Tris (pH = 8.0), 20 mM NaCl, 1 mM $MgCl_2$, 2% glycerol, 1 mM β-mercaptoethanol (BME), and the oxygen scavenging system to prevent tether damage (1.2% glucose, 0.29 mg/mL pyranose oxidase, 0.13 mg/mL catalase; see **Appendix A.3**). Phosphate buffer was not used since it may interfere with the ATPase activity of UvrD, although BME was still included to prevent MutL aggregation. For all assays performed here we used 10 nM MutL, 10 nM UvrD, and 10 μM ATP.

Data analysis. When the assay was performed with the gapped DNA substrate, the helicase activity of UvrD was determined by converting the trap displacement signal to numbers of base-pairs unzipped. This conversion is not the same as for the hairpin DNA substrate. When a single base-pair of a hairpin is unzipped, two single-stranded nucleotides are released along the axis of force, resulting in a trap displacement of $2x_{ss}$, where x_{ss} is the extension of ssDNA per nucleotide. In contrast, when a single base-pair of the gapped DNA substrate is unzipped, one double-stranded base-pair along the force axis converts to one single-stranded nucleotide along the force axis, resulting in an extension change of $x_{ss} - x_{ds}$, where x_{ds} is the extension of dsDNA per base-pair. For both x_{ss} and x_{ds} we used the XWLC model with parameters described in **Section 4.2.1**.

As described in **Section 5.1.1**, we observed two distinct behaviors of UvrD in these experiments: those that unzipped short distances, and those that unzipped long distances. However, the ‘track’ available for

UvrD to unzip in these assays using the gapped DNA construct (~1500 bp) was much longer than that of the hairpin construct used previously (89 bp), and in many cases we observed the unzipping of hundreds of base-pairs. For subsequent analysis we therefore distinguish between ‘short-distance’ (<300 bp) and ‘super-long-distance’ (>300 bp) unzipping behaviors.

6.2.2. The presence of MutL significantly increases the processive unzipping of UvrD

We performed a series of assays using the laminar flow chamber, varying the contents of each stream to create different turnover conditions for each protein. We first created conditions whereby new monomers of UvrD can load onto the DNA substrate after unzipping has been initiated (multiple-turnover conditions) by first loading the bottom stream with blank buffer, and the top stream with UvrD and ATP (**Figure 6.2a**). After forming a tether in the bottom channel, we translated the DNA substrate into the top channel and recorded helicase activity. The pink trace in **Figure 6.2c** shows a representative time trace depicting the number of base-pairs unzipped by UvrD under these conditions.

We then repeated this assay in the presence of 10 nM MutL, so that new monomers of both UvrD and MutL can load onto the DNA substrate after unzipping is initiated (**Figure 6.2b**). The addition of MutL had a dramatic effect on the number of base-pairs unzipped by UvrD, enabling the protein to unzip hundreds of base-pairs before tether breakage (**Figure 6.2c, black trace**). In some cases we recorded >1500 bp unzipped, which is the length of the dsDNA ‘track’ on which we expect UvrD to move (the ‘Left Handle’, see **Appendix A.1.1**) due to its 3’ to 5’ directionality. This could conceivably occur if UvrD unzipped the other dsDNA handle as well, despite the absence of a 3’ ssDNA tail for protein loading. However, at this time we also cannot rule out an error in our analysis for the large numbers observed. Subsequent experiments and analysis will be necessary to provide more quantitative results.

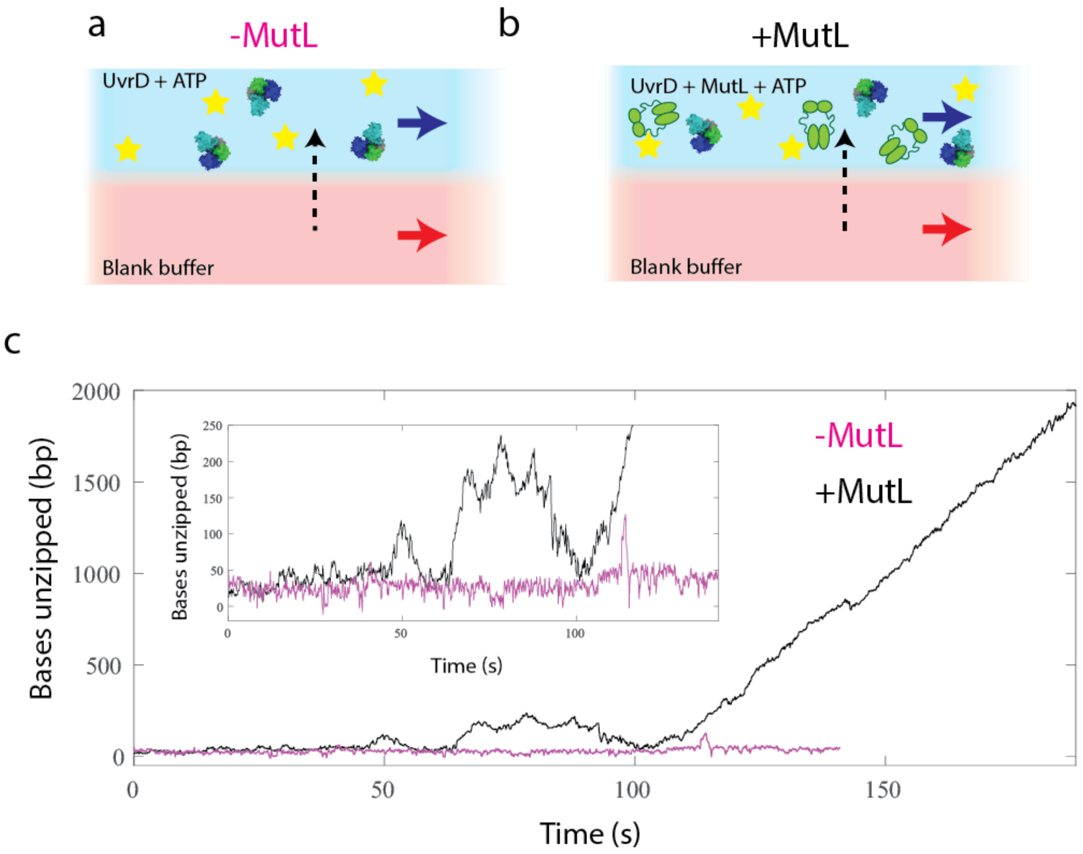


Figure 6.2: Effect of MutL addition on UvrD under multiple-turnover UvrD and MutL conditions. (**a**) Schematic of assay under multiple-turnover conditions for UvrD. The top channel (*blue*) contains UvrD (*blue, green, grey, and cyan*) and ATP (*yellow stars*), while the bottom channel (*red*) contains blank solution. Upon forming a tether in the bottom channel, the DNA tether is moved into the top channel, where helicase activity is recorded. (**b**) Schematic of assay under multiple-turnover conditions for both UvrD and MutL. The top channel contains UvrD, MutL (*green*), and ATP, while the bottom channel contains blank solution. (**c**) Representative time traces of UvrD unzipping in the absence (*pink*) and presence (*black*) of MutL. The inset shows a zoomed-in version of the trace.

6.2.3. Enhancement of UvrD unzipping by MutL requires multiple-turnover UvrD conditions

We next examined the effect of turnover conditions on processive unzipping by UvrD. We first created multiple-turnover conditions for UvrD with single-turnover conditions for MutL. That is, new UvrD monomers may load onto the DNA substrate after unzipping is initiated, but no new monomers of MutL may do so. To perform this assay we loaded MutL with ATP into the bottom channel, and UvrD with ATP into the top channel (**Figure 6.3a**). After forming a tether in the bottom channel, we waited ~15 s to load MutL, then translated the nucleoprotein complex into the top channel and recorded helicase activity. The unzipping behavior observed (**Figure 6.3c, black trace**) was very similar to that under multiple-turnover conditions for MutL, suggesting that new MutL monomers are not needed after unzipping is initiated. This is consistent with bulk studies demonstrating that MutL is not involved in the steady-state part of the reaction (246).

Finally, we reversed the above assay to create single-turnover conditions for UvrD, while maintaining multiple-turnover conditions for MutL. That is, new MutL monomers may load onto the DNA substrate after unzipping is initiated, but no new monomers of UvrD may do so. This assay was performed by loading MutL with ATP into the bottom channel, and UvrD alone (*without* ATP) into the top channel (**Figure 6.3b**). After forming a tether in the bottom channel, we waited ~15 s to load MutL, then translated the nucleoprotein complex into the top stream. We next incubated the tether in the top stream for ~15 s to load UvrD, then translated the tether back into the bottom stream to initiate unzipping. We repeated this procedure numerous times for a single tether to increase throughput. We observed very limited unzipping behavior from UvrD under these conditions (**Figure 6.3c, pink trace**). Even after repeated dips into the top channel to load UvrD monomers, no attempts resulted in super-long-distance unzipping. This finding therefore suggests that the enhancement of unzipping by MutL requires multiple-turnover UvrD conditions.

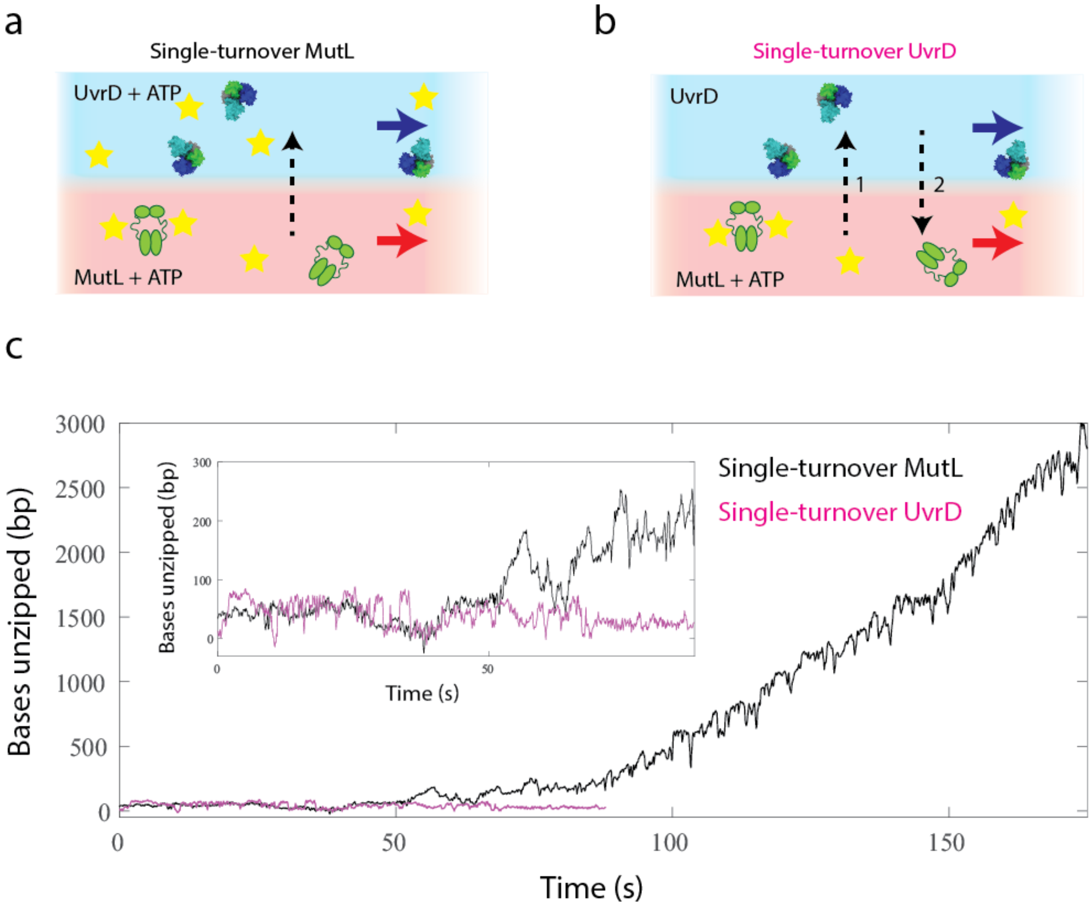


Figure 6.3: Effect of turnover conditions on MutL-enhanced UvrD processivity. (**a**) Schematic of assay under multiple-turnover conditions of UvrD and single-turnover conditions of MutL. The top channel (*blue*) contains UvrD (*blue, green, grey, and cyan*) and ATP (*yellow stars*), while the bottom channel (*red*) contains MutL (*green*) and ATP.

Figure 6.3 (continued): Upon forming a tether in the bottom channel, the DNA tether is moved into the top channel, where helicase activity is recorded. (**b**) Schematic of assay under single-turnover conditions for UvrD and multiple-turnover conditions for MutL. The top channel contains UvrD without ATP, while the bottom channel contains MutL and ATP. Upon forming a tether in the bottom channel and incubating for ~15 s, the DNA tether is moved into the top channel (1). After incubating in the top channel for ~15 s, the nucleoprotein complex is translated back into the bottom channel (2), where helicase activity is recorded. (**c**) Representative time traces of UvrD unzipping under both conditions (*Black*: single-turnover MutL conditions; *Pink*: single-turnover UvrD conditions).

6.2.4. Mechanism of MutL-stimulated UvrD unzipping activity

The results of the serial assays described above are summarized in **Figure 6.4**. The addition of MutL to the UvrD + ATP channel significantly increased the fraction of long-distance traces. However, the presence of MutL alone was insufficient to cause the increase in processive traces. The ability of new UvrD monomers to bind the DNA substrate was also a major determinant of processivity, although the ability of new MutL monomers to bind did not have a significant effect.

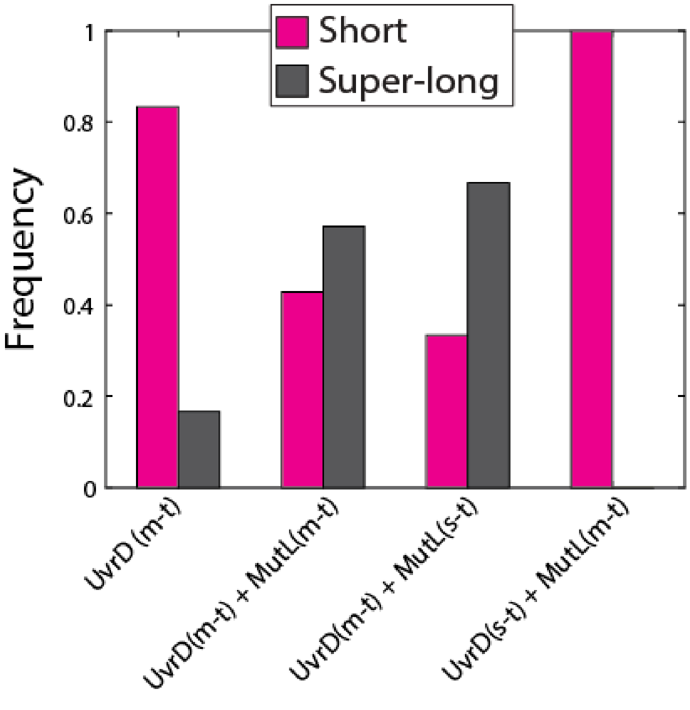


Figure 6.4: Effect of MutL on UvrD processivity. Number of total base-pairs unzipped under each assay condition. m-t: multiple turnover. s-t: single-turnover. m-t UvrD (N = 6), m-t UvrD + m-t MutL (N = 7), m-t UvrD + s-t MutL (N = 3), s-t UvrD + m-t MutL (N = 8).

Our assays suggest that MutL-enhanced processive unzipping of a DNA substrate requires additional UvrD monomers to bind after unzipping is initiated. Our results are consistent with a model whereby MutL functions by continually loading UvrD monomers. However, since single-turnover conditions for UvrD failed to produce super-long-distance unzipping behavior, our results are not consistent with a model in which MutL acts as a processivity factor for a monomer of UvrD, nor one in in which MutL acts by loading a single UvrD oligomer from solution.

6.2.5. Future directions: counting UvrD monomers during MutL-enhanced activity

The distinguishing feature of the competing models for MutL-UvrD interaction is primarily the number of UvrD monomers present during both initiation of unzipping and steady-state unzipping behavior. A more direct test of these models can be performed by simultaneously measuring helicase activity while counting the monomers of UvrD present on the substrate, as described in **Section 5.1.1** for UvrD alone. If MutL continually loads monomers, then the rate r_{on} at which monomers load onto the DNA substrate in the presence of MutL will likely be higher than that in the absence of MutL. Preliminary data taken under multiple turnover conditions for both proteins using unlabeled dual-mutant UvrD (mutations for attaching fluorophores; **Appendix A.7.1**) showed similar super-long-distance unzipping as wild-type. This is promising, as it demonstrates that the mutations required to attach fluorophores do not significantly disrupt the interaction between UvrD and MutL.

Appendices

Appendix A

Protocols for performing experiments*

*Many of the protocols described here were first published in one of two book chapters:

Whitley, K. D., Comstock, M.J., and Chemla, Y.R. (2017). High-Resolution "Fleezers": Dual-Trap Optical Tweezers Combined with Single-Molecule Fluorescence Detection. Methods in Molecular Biology. A. Gennerich, Springer. **1486:** 183-256.

Whitley, K. D., Comstock, M.J., and Chemla, Y.R. (2017). High-Resolution Optical Tweezers Combined With Single-Molecule Confocal Microscopy. Methods in Enzymology. Y. R. Chemla and M. Spies, Elsevier. **582:** 137-169.

A.1. DNA constructs

*A.1.1. Gapped constructs**

* This protocol is modified from Whitley et al. (2017) (62) with permission from Springer.

Here, we synthesize a DNA construct consisting of a short segment of ssDNA flanked by long dsDNA "handles" (**Figure A.1**). All oligonucleotides used to assemble this construct can be purchased from Integrated DNA Technology (IDT) and are listed in **Table A.1**. We synthesize three DNA segments separately and then ligate them together: left handle ("LH", 1.5 kb), insert ("Insert", 9- to 35-nt long) and right handle ("RH", 1.7 kb). LH and RH are made by PCR amplification of sections of the pBR322 plasmid and λ phage DNA using forward primers functionalized with 5' biotin and 5' digoxigenin, respectively (**Figure A.1**).

Synthesis and purification of LH and RH:

1. For PCR synthesis of LH, mix 35 μL of nuclease-free water, 5 μL of forward primer (10 μM concentration), 5 μL of reverse primer (10 μM), 2 μL of pBR322 template DNA (10 ng/μL; NEB), 3 μL DMSO, and 50 μL 2x Phusion HF Master Mix (NEB) for a final volume of 100 μL.
2. For PCR synthesis of RH, mix 35 μL of nuclease-free water, 5 μL of forward primer (10 μM concentration), 5 μL of reverse primer (10 μM), 2 μL of λ DNA (10 ng/μL; NEB), 3 μL DMSO, and 50 μL 2x Phusion HF Master Mix for a final volume of 100 μL.

3. Run PCR on both reaction mixes, using the following program: (1) 98 °C for 30 s, (2) 98 °C for 10 s, (3) 59 °C for 10 s, (4) 72 °C for 33 s, (5) repeat steps 2-4 30x, (6) 72 °C for 5 min, (7) 4 °C forever.
4. Purify PCR products following the QIAquick PCR purification kit "spin protocol". Add 30 µL of elution buffer instead of 50 µL for a more concentrated solution.

Digestion of LH and RH with restriction enzymes:

These digestions are performed to produce 5' and 3' overhangs of the LH and RH, respectively, which will eventually be used to base-pair to Insert.

5. Add to 30 µL of LH: 2 µL of PspGI restriction enzyme (NEB) and 3.5 µL of CutSmart 10x buffer.
6. Add to 30 µL of RH: 2 µL of TspRI restriction enzyme (NEB) and 3.5 µL of CutSmart 10x buffer.
7. Incubate the LH reaction mix at 75 °C for 1 hr and the RH reaction mix at 65 °C for 1 hr. For convenience, we suggest using a PCR thermal cycler with a temperature gradient of 65-75 °C, so both reactions can be done simultaneously. We program the PCR cycler to 4 °C after 1 hr to stop the reaction.

Removal of 5' phosphate from LH:

The 5' phosphate of LH is removed to prevent self-ligation. We have observed that this step decreases the amount of off-ligation products.

8. Add the following to the ~35 µL of digested LH: 4 µL 10x Antarctic Phosphatase buffer (NEB), 1 µL Antarctic Phosphatase (5 units; NEB).
9. Incubate the new LH reaction mix at 37^{o}C for 30 min, then 80^{o}C for 2 min to inactivate the phosphatase.
10. Purify the digested LH and RH using the QIAquick PCR purification kit as before.

Ligation of LH to Insert:

The ligation is done in two steps in order to avoid gel purification. We first ligate LH to Insert using an excess of Insert. Since the Insert is short (~30 bp), it can be easily removed by a PCR cleanup kit as if it were a primer. We then ligate the LH+Insert product to RH using an excess of RH. Since we later incubate the fully constructed DNA with streptavidin beads (not anti-digoxigenin beads), and the biotin moiety is on the LH, any excess RH will not bind the beads. This removes the need for any further purification.

11. Measure the concentrations of LH and RH using a Nanodrop, which requires only 1 µL of each solution.

12. Add 4 µL of 10x T4 DNA ligase buffer and 4 µL of T4 DNA ligase (NEB) to the entire ~30 µL volume of LH. Based on the concentration measured in the previous step, add a 10x excess of Insert. Add nuclease-free water up to a final volume of 40 µL.
13. Ligate LH to Insert at RT (~22 °C) for 1 hr, and then heat to 65 °C for 15 min to inactivate T4 ligase.
14. Purify ligation product with a PCR cleanup kit following the QIAquick spin protocol, adding 30 µL of elution buffer.

Ligation of LH+Insert to RH:

15. Make a solution of this ligation product (LH + insert) and a 2x excess of RH with a final volume of 32 µL, based on the concentrations measured in **step 11**. Add 4 µL of 10x T4 DNA ligase buffer and 4 µL of T4 DNA ligase to this for a final volume of 40 µL.
16. Ligate RH to LH + Insert at RT (~22 °C) for 1 hr, and then heat to 65 °C for 15 min to inactivate T4 ligase.

Verification of ligation product (optional):

17. Mix 1 µL of ligation product with 5 µL nuclease-free water and 1 µL 6x gel loading dye.
18. Place a pre-cast 1% agarose gel (in TBE; Bio-Rad) with ethidium bromide in an electrophoresis cell, and cover it with 0.5x TBE.
19. To this gel, add 5 µL of a 1 kb DNA ladder to lane 1, then 5 µL of the ligation product to lane 2.
20. Run at 70 V for ~1 hr.
21. Image using UV lamp to verify the size of the PCR products.

Primer name	**Sequence (IDT format, 5' to 3')**
LH forward primer	/5Biosg/ TGAAGTGGTGGCCTAACTACG
LH reverse primer	CAAGCCTATGCCTACAGCAT
RH forward primer	/5DigN/GGGCAAACCAAGACAGCTAA
RH reverse primer	CGTTTTCCCGAAAAGCCAGAA
Insert name	**Sequence (IDT format, 5' to 3')**
2Sp insert	/5Phos/ CCTGG TTTTT AGGACTTGT TTTTT CCCACTGGC
1Sp insert	/5Phos/ CCTGG TTTTT AGGACTTGT CCCACTGGC
0Sp insert	/5Phos/ CCTGG AGGACTTGT CCCACTGGC
Spacer-only insert	/5Phos/ CCTGG TTTTT TTTTT CCCACTGGC

Insert name	**Sequence (IDT format, 5’ to 3’)**
seq2 insert	/5Phos/ CCTGG TTTTT GATTCGTTTC TTTTT CCCACTGGC
35 dT insert	/5Phos/ CCTGG TTTTT TTTTT TTTTT TTTTT TTTTT TTTTT TTTTT CCCACTGGC

Table A.1: Oligonucleotides for synthesizing gapped constructs. Sequences are listed in IDT format. Poly-dT spacers are in blue, and oligonucleotide binding sites are in red.

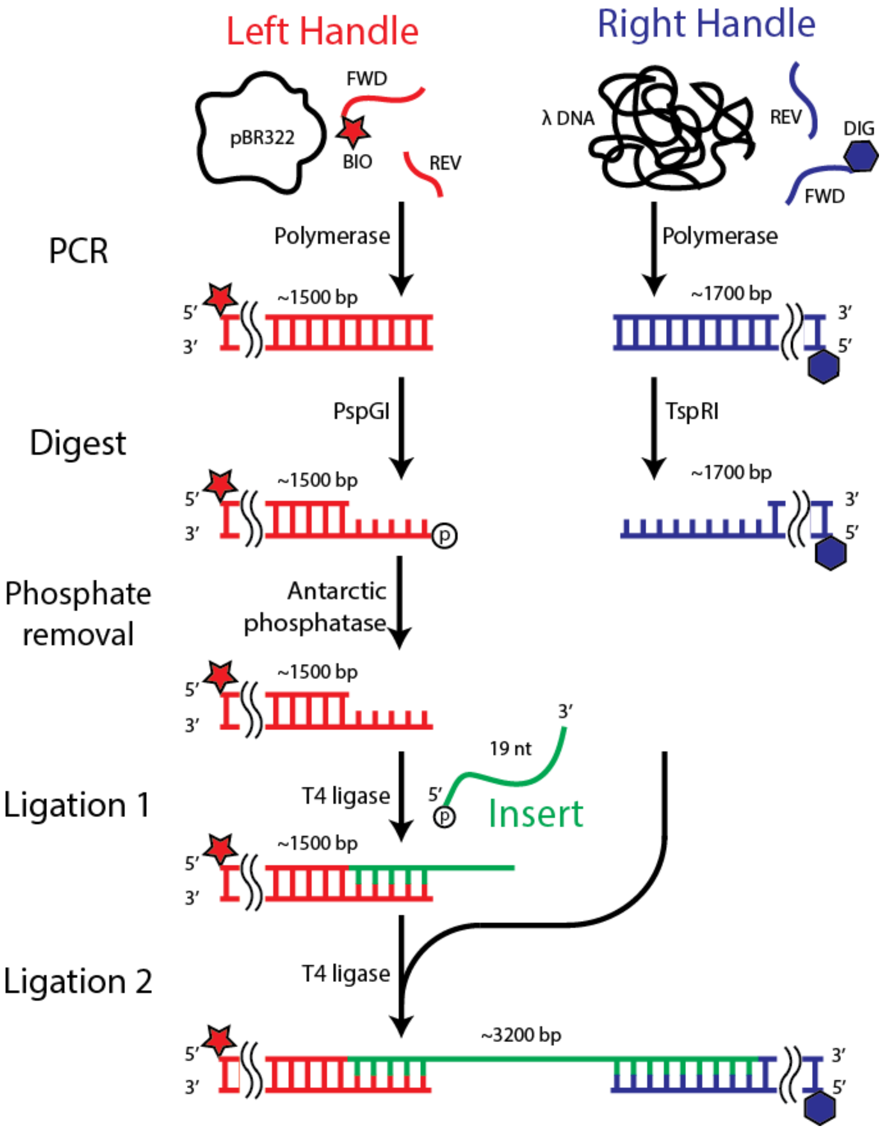


Figure A.1: Synthesis of gapped construct. Schematic depicting the major steps involved in preparing the DNA construct. The handles are first prepared by PCR of template DNA (pBR322 and λ DNA) using primers (FWD and REV) with either a biotin moiety (BIO, Left Handle) or a digoxigenin moiety (DIG, Right Handle) attached to the 5’ ends of the FWD primers. The two handles are digested by restriction enzymes (PspGI and TspRI) to produce 5’ and 3’ overhangs. The digested Left Handle is first treated with phosphatase to remove its 5’ phosphate, then ligated using

Figure A.1 (continued): T4 ligase to the short Insert containing a phosphate group on its 5’ end. Finally, the Right Handle is ligated to this product (Left Handle + Insert) to produce the final construct. Figure reproduced from (62) with permission from Springer.

*A.1.2. Hairpin constructs**

* This protocol is reprinted from Whitley et al. (2017) (63) with permission from Elsevier.

The DNA hairpin construct consists of three DNA fragments that are ligated together (**Figure A.2**): a ‘Hairpin’ (HP) flanked by a ‘Right Handle’ (RH) and a ‘Left Handle’ (LH) that serve as functionalized linkers that attach to the trapped beads. RH is made from a 1.5-kb PCR-amplified section of the pBR322 plasmid using a 5’-digoxigenin-modified forward primer and a reverse primer containing one abasic site and a long 5’ overhang (**Table A.2**). The digoxigenin moiety is used to link this end of the construct to an anti-digoxigenin-coated bead. The overhang consists of a poly-dT loading site for helicase binding immediately adjacent to the abasic site, followed by 29 nt that anneal to a complementary sequence in HP. LH is synthesized from a different PCR-amplified section of pBR322 using a 5’-biotin-modified primer. The biotin moiety allows this end of the construct to bind to a streptavidin-coated bead. HP is a single long oligonucleotide containing the complementary sequence to the LH overhang on its 5’ end, followed by the complementary sequence to the RH overhang and a 153-nt self-complementary sequence (**Table A.2**). When self-annealed and ligated to LH and RH, HP makes an 89-bp hairpin stem capped by a $(dT)_4$ tetraloop.

Synthesis and purification of LH and RH:

1. For PCR synthesis of LH, mix 35 µL of nuclease-free water, 5 µL of forward primer (10 µM concentration), 5 µL of reverse primer (10 µM), 2 µL of pBR322 template DNA (10 ng/µL; NEB), 3 µL DMSO, and 50 µL 2x Phusion HF Master Mix (NEB) for a final volume of 100 µL.
2. For PCR synthesis of RH, mix 35 µL of nuclease-free water, 5 µL of forward primer (10 µM concentration), 5 µL of reverse primer (10 µM), 2 µL of pBR322 template DNA (10 ng/µL), 3 µL DMSO, and 50 µL 2x Phusion HF Master Mix for a final volume of 100 µL.
3. Run PCR on both reaction mixes. We use the following program: (1) 98ºC for 30 s, (2) 98ºC for 10 s, (3) 59ºC for 10 s, (4) 72ºC for 33 s, (5) repeat steps 2-4 30x, (6) 72ºC for 5 min, (7) 4ºC forever.
4. Purify PCR products following the QIAquick PCR purification kit “spin protocol” (Qiagen). Add 30 µL of elution buffer instead of 50 µL for a more concentrated solution. These PCR products can be verified by gel electrophoresis.

Digestion of LH and removal of the 5'-phosphate:

5. Add the following to 30 µL of LH: 2 µL of PspGI restriction enzyme (20 units total; NEB) and 3.5 µL of CutSmart 10x buffer (NEB).
6. Incubate the LH reaction mix at 75ºC for 1 hr.
7. Add the following to the ~35 µL of digested LH: 4 µL 10x Antarctic Phosphatase buffer (NEB), 1 µL Antarctic Phosphatase (5 units; NEB).
8. Incubate the new LH reaction mix at 37ºC for 30 min, then 80ºC for 2 min to inactivate the phosphatase.
9. Purify the digested LH using the QIAquick PCR purification kit as before.

Ligation of LH and RH to HP:

10. Measure the concentrations of LH and RH using a Nanodrop UV-Vis spectrophotometer (Thermo Scientific), which requires only 1 µL of each solution.
11. Mix the three DNA components (LH, RH, and HP) in an equimolar ratio (1:1:1) to a final volume of 32 µL.
12. Add 4 µL of 10x T4 DNA ligase buffer (NEB) and 4 µL of T4 DNA ligase (1600 units; NEB) for a final volume of 40 µL.
13. Ligate at RT (~22ºC) for 1 hr, and then heat to 65ºC for 15 min to inactivate T4 ligase.
14. Run the final product on a 1% agarose gel with no ethidium bromide for ~70 min.
15. Incubate the gel in a solution of 50 mL 0.5x TBE and 15 µL 10,000x GelGreen (Biotium) for ~20 min.
16. Image the gel using Dark Reader Transilluminator (Clare Chemical Research) and cut out the appropriate band (~3.4 kb) using a clean razor blade. Place this gel slice in a previously weighed 1.5 mL tube.
17. Purify the final ligated construct from the agarose gel slices using QIAEX II gel extraction kit (Qiagen). Add 30 µL of elution buffer instead of 50 µL for a more concentrated solution.

Primer name	**Sequence (IDT format, 5' to 3')**
LH FWD primer	/5Biosg/ TGAAGTGGTGGCCTAACTACG
LH REV primer	CAAGCCTATGCCTACAGCAT
RH FWD primer, sequence 1 (3' 10 dT)	/5Phos/ TTGAAATACCGACCGCTCAGCTATCAGCC TTTTTTTTTT /idSp/ CTCTGACACATGCAGCTCCC
RH FWD primer, sequence 1 (3' 19 dT)	/5Phos/ TTGAAATACCGACCGCTCAGCTATCAGCC TTTTTTTTTTTTTTTTTTT /idSp/ CTCTGACACATGCAGCTCCC

Primer name	**Sequence (IDT format, 5’ to 3’)**
RH FWD primer, sequence 2 (3’ 19 dT)	/5Phos/ GGTCCCGCAATCAATAGACTAGCGACTAA TTTTTTTTTTTTTTTTTTT /idSp/ CTCTGACACATGCAGCTCCC
RH FWD primer, sequence 1 (3’ 38 dT)	/5Phos/ TTGAAATACCGACCGCTCAGCTATCAGCC TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT /idSp/ CTCTGACACATGCAGCTCCC
RH REV primer	/5DigN/ CAACAACGTTGCGCAAACT
Hairpin insert name	**Sequence (IDT format, 5’ to 3’)**
Hairpin insert, sequence 1	/5Phos/ CCTGG GGCTGATAGCTGAGCGGTCGGTATTTCAA AAGTCAACGTACTGATCACGCTGGATCCTAGAGTCAACGTACTG ATCACGCTGGATCCTA TTTT TAGGATCCAGCGTGATCAGTACGTT GACTCTAGGATCCAGCGTGATCAGTACGTTGACTT
Hairpin insert, sequence 2	/5Phos/ CCTGG TTAGTCGCTAGTCTATTGATTGCGGGACC CCTGACCATGCAGTCGACATAGTTCGAAGCTCTGACCATGCAGT CGACATAGTTCGAAGC TTTT GCTTCGAACTATGTCGACTGCATGG TCAGAGCTTCGAACTATGTCGACTGCATGGTCAGG
Hairpin insert, sequence 1, 10 nt 5’ overhang	/5Phos/ CCTGG TTTTTTTTTT GGCTGATAGCTGAGCGGTCGGTATTTC AAAAGTCAA CGTACTGATCACGCTGGATCCTAGAGTCAACGTACT GATCACGCTGGATCCTA TTTT TAGGATCCAGCGTGATCAGTACGT TGACTCTAGGATCCAGCGTGATCAGTACGTTGACTT
Hairpin insert, sequence 1, 38 nt 5’ overhang	/5Phos/ CCTGG TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT GGCTGATAGCTGAGCGGTCGGTATTTCAA AAGTCAACGTACTGA TCACGCTGGATCCTAGAGTCAACGTACTGATCACGCTGGATCCTA TTTT TAGGATCCAGCGTGATCAGTACGTTGACTCTAGGATCCAGC GTGATCAGTACGTTGACTT

Table A.2: Oligonucleotides for synthesizing hairpin constructs. Protein loading sites are in blue, primer/hairpin complementary sequences are in red, and the 4 dT loop is in green. Sequences are in IDT format.

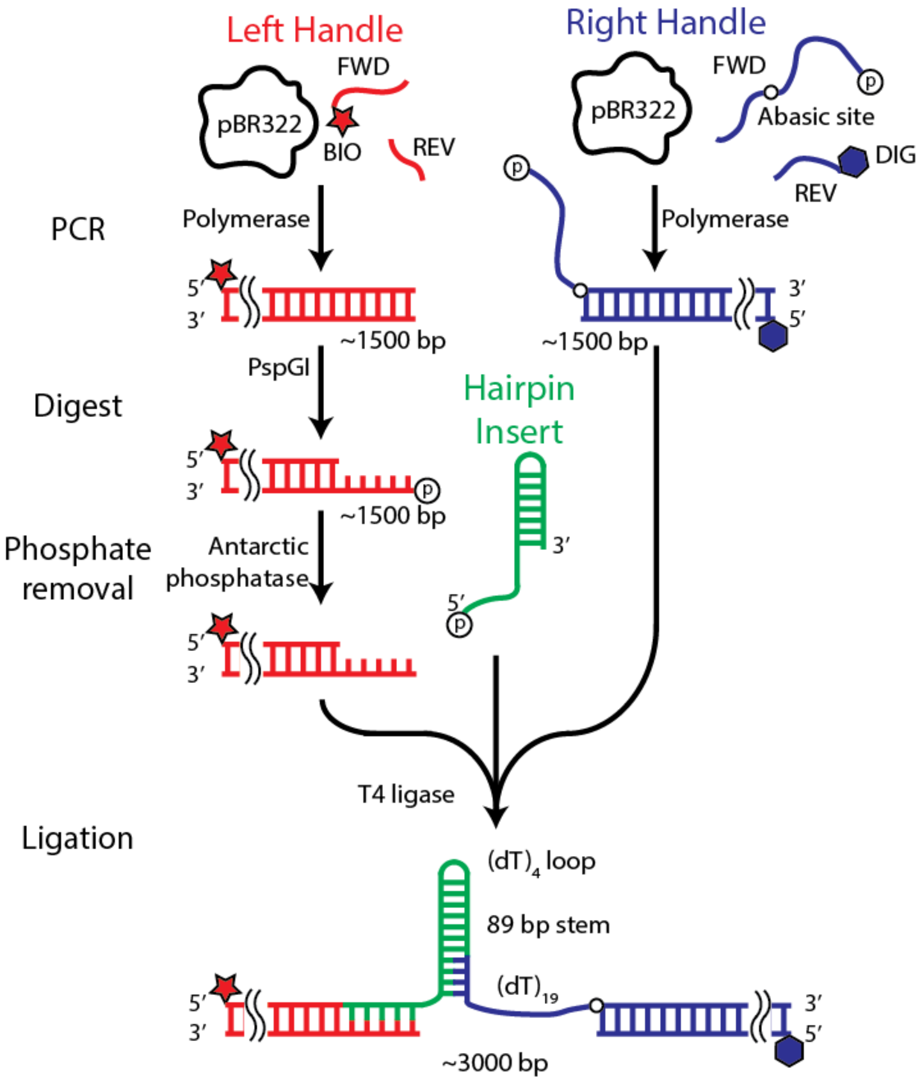


Figure A.2: Synthesis of hairpin construct. Schematic depicting the major steps involved in preparing the DNA construct. Figure reproduced from (63) with permission from Elsevier.

A.2. Oligonucleotide probes for hybridization experiments

Probe name	**Sequence (IDT format, 5' to 3')**
7merCy3	CAAGTCC /3Cy3Sp/
8merCy3	CAAGTCCT /3Cy3Sp/
9merCy3	ACAAGTCCT /3Cy3Sp/
10merCy3	ACAAGTCCTA /3Cy3Sp/
Probe name	**Sequence (IDT format 5' to 3')**

12merCy3	AACAAGTCCTAA /3Cy3Sp/
9merIntCy3	ACAAG /iCy3N/ CCT
9mer9linkCy3	ACAAGTCCT /iSp9/ /3Cy3Sp/
9mer	ACAAGTCCT
seq2	GAAACGAATC /3Cy3Sp/
rna7merCy3	rCrArArGrUrCrC /3Cy3Sp/
rna8merCy3	rCrArArGrUrCrCrU /3Cy3Sp/
rna9merCy3	rArCrArArGrUrCrCrU /3Cy3Sp/
rna10merCy3	rArCrArArGrUrCrCrUrA /3Cy3Sp/
rna11merCy3	rArArCrArArGrUrCrCrUrA /3Cy3Sp/
rna10mer	rArCrArArGrUrCrCrUrA
rnaSeq2	rGrArArArCrGrArArUrC /3Cy3Sp/

Table A.3: Oligonucleotide probes for hybridizing to gapped constructs in **Chapters 2** and **3**. Complementary sequences are in inserts in **Table A.1**.

A.3. Oxygen scavenger systems*

*This protocol is modified from Whitley et al. (2017) (62) with permission from Springer.

Reactive oxygen species such as singlet oxygen can cause fluorophore bleaching and tether breakage (55,251). It is therefore critical that we remove oxygen from any solutions where experiments will be performed. This is typically done using enzyme cocktails such as GOx (glucose oxidase, catalase, and glucose) or POx (pyranose oxidase, catalase, and glucose). In the experiments described above, we used both systems, although we have found that the POx system offers several advantages over GOx. Firstly, GOx causes a drop in pH over the course of experiments, requiring higher concentrations of buffering agents, whereas POx does not (252). Secondly, we have found that POx can be stored at 4^{o}C for several months without significant decrease in activity, while GOx begins to form insoluble aggregates over the course of 1-2 weeks when stored in the same way.

This protocol describes how to prepare both oxygen scavenging systems. The full oxygen scavenging system includes one of these two enzymes plus catalase and glucose. For long-term storage, glucose is left out of the mixture.

1. Dissolve 20 mg of glucose oxidase (for GOx; Sigma-Aldrich) or 5.8 mg of pyranose oxidase (for POx; Sigma-Aldrich) into 180 μL of T50 buffer.
2. Dissolve 13 mg of catalase (EMD) into 200 μL T50 buffer. This is a 10x stock of catalase.

3. Add 20 μL of 10x catalase to the glucose oxidase/pyranose oxidase solution, and mix well.
4. Centrifuge the solution at 18,000 rcf for 1 min to spin down bubbles and any undissolved material.
5. Centrifuge solution through a 0.22-μm centrifuge filter at 18,000 rcf for 1 min.
6. Store at 4 ºC for short-term use (weeks to months), or aliquot and flash freeze in liquid nitrogen and store at -20 ºC for long-term storage (typically months).

A.4. Anti-blinking system*

* This protocol is modified from Whitley et al. (2017) (62) with permission from Springer.

We use Trolox (TX) as a triplet-state quencher to reduce fluorophore blinking and photobleaching (253).

1. Add ~8.5 mL of water and 50 μL of 1 M NaOH to 10 mL conical tube. The high pH helps the Trolox dissolve.
2. Add 10 mg Trolox powder (Sigma-Aldrich).
3. Wrap tube in foil and rotate for ~1 hr to dissolve.
4. Add Tris up to desired concentration.
5. Add water to a total volume of 10 mL.
6. Filter the buffer through a 0.22-μm syringe filter.
7. Store the Trolox solution in the dark at 4°C.

A.5. Flow chamber*

*This protocol is modified from Whitley et al. (2017) (62) with permission from Springer.

Sample chambers consisted of two No. 1 microscope coverslips that sandwich a layer of sealing film into which the channels have been cut (**Figure A.3**). The top and bottom channels usually contain anti-digoxigenin and streptavidin beads, respectively, and are each shunted to the central channel by thin glass capillaries. The central channel consists of two inlets that smoothly merge to form a single channel with laminar flow. Since the flow is *laminar*, the two streams do not mix but rather maintain a narrow, stationary interface limited only by diffusion (254).

Once assembled, we mount the sample chamber on a custom-built U-shaped aluminum bracket (**Figure A.3**). Two acrylic mating pieces press the chamber against the two edges of the bracket, aligning holes in the bracket with the inlet and outlet holes of the front coverslip surface. Short pieces of Tygon silicone tubing are fed through set screws; these are screwed into the sample mount, pressing the silicone tubing end onto the inlet and outlet holes of the chamber, forming a seal. Polyethylene tubing is fed into the exposed silicone tubing end and used to flow solutions into the chamber.

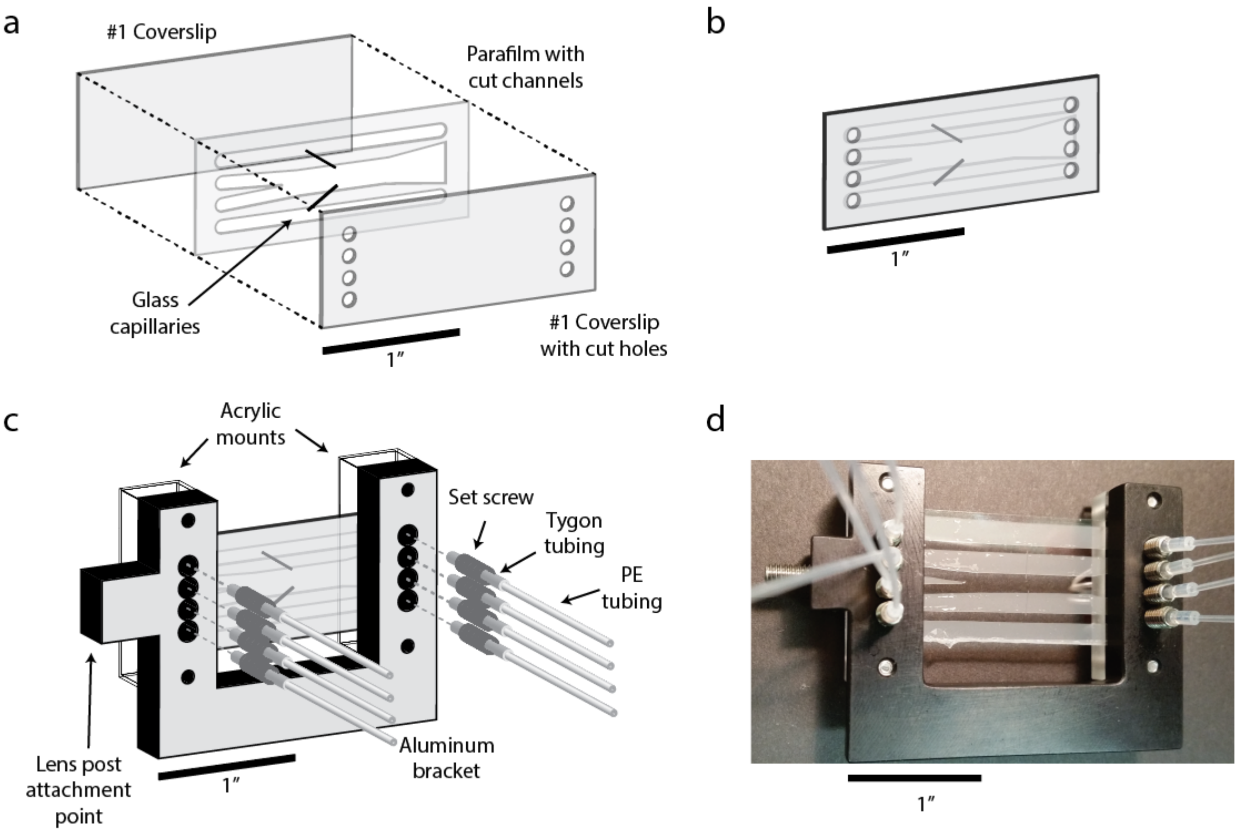


Figure A.3: Sample flow chamber assembly. (**a**) Expanded view of the "Parafilm sandwich" that comprises the chamber. A piece of Parafilm with flow channels cut into it is placed on a coverslip with eight holes cut into it. Two glass capillaries span the Parafilm to connect the bottom and top channels to the large central channel. A coverslip with no holes is then placed on top of the Parafilm to form the assembled chamber. (**b**) A fully assembled flow chamber. (**c**) A flow chamber mounted on an anodized aluminum bracket, held in place by two acrylic mounts. Four holes on either side of the mount are aligned with the holes of the coverslip. A short length of Tygon tubing is threaded through a set screw, and a longer stretch of polyethylene (PE) tubing is inserted into the Tygon tubing. Eight threaded set screws are prepared and screwed into the eight holes in the aluminum bracket to serve as inlet and outlet channels for the flow chamber. (**d**) Photograph of an assembled and mounted flow chamber. Figure reproduced from (62) with permission from Springer.

Some biomolecules will tend to adsorb to the surfaces of the chamber. Adsorption may be particularly important if the molecules are fluorescently labeled, as this may increase fluorescence background. To keep molecules in solution, chambers should be passivated with polyethylene glycol (PEG) (112), using a protocol adapted from (255). **Steps 1-5** should follow if the chambers are to be passivated with PEG. Otherwise these steps may be skipped:

1. Sonicate the coverslips in acetone for 30 min.

2. Rinse the coverslips with water 3x to remove excess acetone.
3. Sonicate the coverslips in 3 M KOH for 20 min.
4. Rinse with water 5x to remove excess KOH.
5. Rinse once with methanol, and dry with nitrogen gas.

The following steps describe cutting channels into Parafilm using a laser engraver (Universal Laser Systems):

6. Cut a 4” × 4” square of Parafilm and place it taut over a frame.
7. Make a blueprint of the flow channels as shown in **Figure A.3**.
8. Set the following settings for the laser engraver: Power 9%, Speed 10%, PPI 500, Z-Axis 4”.
9. Connect the exhaust from the laser engraver to the portable dust collector motor, and the output of this into a fume hood. Turn on the motor.
10. Run the engraver program and print one copy.
11. After engraving is finished, use tweezers to pull off carefully any loose segments of Parafilm that may be still attached to the channel structure.

The following steps describe how to assemble a flow chamber:

12. Use tweezers to carefully lay down the cut Parafilm on the coverslip with holes, aligning the Parafilm channels with the holes (**Figure A.3**). It is important to attempt to do this correctly the first time, as moving the Parafilm around tends to leave residue on the coverslips.
13. Use a fresh, clean razor blade to cut two glass capillaries to a length such that they will span the Parafilm between the outer channels and the inner channel.
14. Use tweezers to place the two capillaries at the appropriate positions on the Parafilm.
15. Lay the uncut coverslip on top of the Parafilm to form a “Parafilm sandwich” between the two coverslips (see **Figure A.3**).
16. Preheat a hot plate and a ~0.5 kg weight before use to ~100-130 ^{o}C.
17. Melt the Parafilm by laying the assembled chamber on the hot plate between two Kimwipes (to prevent melted Parafilm from sticking to the hot plate) with the weight on top of it. Wait ~2-3 min for the Parafilm to melt. It is important that the weight is well-centered to apply a uniform pressure over the chamber, otherwise one side will become flatter than another.
18. Take the chamber off the hot plate and allow it to cool for ~1 min.
19. Mount the chamber onto its bracket, lining up the chamber holes with the bracket holes (see **Figure A.3**). Gently screw the mounts into place to secure the chamber position.

20. Cut a sufficient length of Tygon tubing for all input/output interfaces (typically 5/8” per input/output), cutting one end of it diagonally into a point to ease the next step.
21. Thread the Tygon tubing into the set screw, leaving sufficient space between the tubing ends.
22. Carefully cut the Tygon tubing between the set screws using a new razor blade, making sure the cut is straight.
23. Cut a sufficient length of PE tubing (7-10” typically), and push it into the Tygon tubing.
24. Gently screw the set screws into the chamber bracket until the Tygon tubing is flush up against the chamber.
25. Insert 26-gauge 3/8” intradermal bevel needles to the input PE tubing.

Steps 26-39 should follow if the chambers are to be passivated with PEG. Otherwise these steps may be skipped:

26. Sonicate a 125 mL Erlenmeyer flask in 1 M KOH.
27. Rinse the flask with water 3x to remove excess KOH.
28. Sonicate the flask in methanol for 20 min.
29. Rinse the flask with methanol and dry with nitrogen gas.
30. In the flask, make a solution of 1% (v/v) N-(2-aminoethyl)-3-aminopropyltrimethoxysilane (United Chemical Technologies) and 5% (v/v) acetic acid in methanol.
31. Assemble the flow chamber, then flow 1 mL of the amino-silane mixture through the channels.
32. Allow the amino-silane mixture to incubate for 15 min.
33. Flow another 1 mL of the mixture, and incubate for 15 min.
34. Flush out the amino-silane mixture with 1 mL methanol.
35. Flush out the methanol with 3 mL water.
36. Make a solution of 25% (w/v) methoxy-PEG-succinimidyl valerate (Laysan Bio) in 0.1 M sodium bicarbonate.
37. Flow ~200 μL of this PEG solution into each channel.
38. Allow the chamber to incubate for ~4 hours, then rinse with copious amounts of water and dry with nitrogen gas.
39. Store PEG chambers at 4°C in the dark.

A.6. Polymer models for dsDNA, ssDNA, and RNA-DNA hybrids in hybridization experiments

To model dsDNA we used the extensible worm-like chain (XWLC) model, with parameters taken from the literature: helix rise h_{ds} = 3.4 Å/bp, persistence length P_{ds} = 53 nm (121), and stretch modulus S_{ds} = 1100 pN (120). For RNA-DNA hybrids, these values are less established. A previous optical tweezers study found that the contour lengths of RNA-DNA hybrids were consistent with an A-helix (189), and so we used a helix rise $h_{RNA\text{-}DNA}$ = 2.9 Å/bp that is characteristic of A-helices (190). This same study found a range of persistence lengths, and so we use the full range reported: $10 < P_{RNA\text{-}DNA} < 22$ nm (189). Additionally, we assume a value of $S_{RNA\text{-}DNA}$ = 1100 pN.

A range of values for the elastic parameters of ssDNA have been reported in the literature for a variety of experimental conditions (106,124). We determined a range of parameters directly for our experimental conditions, following the procedure described in (256) and (108). We designed a DNA construct identical to the one used in the extension change experiments (made using “2Sp insert”; **Table A.1**), but lacking the probe binding site (made using “spacer-only insert”; **Table A.1**). The 2Sp construct contained a 19-nt single-stranded region, whereas the one lacking the binding site contained a 10-nt single-stranded region. We collected a set of force-extension curves for each construct in a solution containing no complementary oligonucleotides and averaged each set separately. We then subtracted the two averaged curves to produce the force-extension curve of the 9-nt binding site alone (**Figure A.4**).

We then sought to fit this “difference” curve to an appropriate polymer model. For all data collected under conditions of 0 mM Mg^{2+} we used the snake-like chain (SLC) model (122). Unfortunately, the subtraction of the two F-x curves leaves us with a variable offset due to variation in bead diameters and other factors, and so we cannot directly fit the SLC model to obtain the free parameter L_o, the extension of the tether at 20 pN. However, since the SLC model and extensible worm-like chain (XWLC) model are nearly identical at this force, we instead fitted the F-x difference curve to the XWLC model for fixed values of contour length per nucleotide h_u within a range reported in literature ($5.9 < h_u < 6.3$ Å/nt) (106,124-126) and stretch modulus (800 pN) (161) to obtain a range of persistence lengths P_u ($1.67 < P_u < 1.99$ nm). We then used these values to calculate a range of values for L_o. The parameters then used for the SLC model were f_c = 0.9 pN, γ = 0.6, L_c/L_o = 0.35, and $4.23 < L_o < 4.68$ nm. The SLC model using these parameters was found to fit the F-x difference curve better than the XWLC model (χ^2(SLC) = 68.2, χ^2(XWLC) = 80.0 over force range $1 < F < 22$ pN).

For all data collected under conditions of 20 mM Mg^{2+} (**Figures 2.3** and **2.4**), we separately collected F-x curves of the two different DNA constructs in this buffer. We then fitted the resulting “difference” F-x curve to the XWLC model using fixed values of h_u ($5.9 < h_u < 6.3$ Å/nt) and stretch modulus (800 pN) to obtain a range of P_u (**Figure A.4b**). The fitted range of persistence lengths was then $1.02 < P_u < 1.17$ nm.

Conditions	**Model**	***P*** **(nm)**	***h*** **(Å /nt)**	L_o **(nm)**
dsDNA, 0 mM Mg^{2+}	XWLC	53	3.4	
dsDNA, 20 mM Mg^{2+}	XWLC	53	3.4	
RNA-DNA, 20 mM Mg^{2+}	XWLC	10 - 20	2.9	
ssDNA, 0 mM Mg^{2+}	SLC			4.23 – 4.68
ssDNA, 20 mM Mg^{2+}	XWLC	1.02 – 1.17	5.9 – 6.3	

Table A.4: Polymer models and parameters used for various conditions.

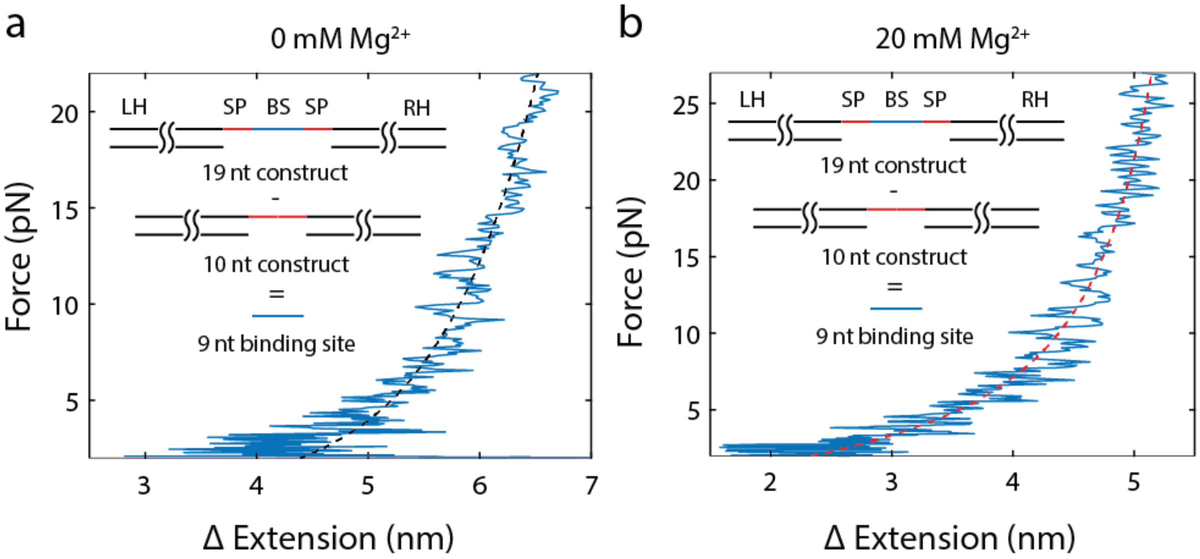


Figure A.4: Calculated and fitted force-extension curves of the 9-nt binding site. (**a**) F-x curve of the 9-nt binding site in 0 mM Mg^{2+} buffer. Inset, top: 19-nt gapped construct used in this study. LH: Left Handle, SP: Spacer, BS: Binding Site, RH: Right Handle. Middle: 10-nt gapped construct containing only spacers, with no binding site. Force-extension curves were collected for each construct (19-nt construct: $N = 8$; 10-nt construct: $N = 16$) in the absence of complementary oligonucleotides, averaged together, and then subtracted from one another to produce the F-x curve of the 9-nt binding site only (*blue curve*). The SLC model is shown (*black dotted line*) using parameters $f_c = 0.9$ pN, $\gamma = 0.6$, $L_c/L_o = 0.35$, and $L_o = 4.23$ nm. The model is fitted to an arbitrary extension offset resulting from bead size variation and other factors. (**b**) F-x curve of the 9-nt binding site in 20 mM Mg^{2+} buffer (*blue curve*). F-x curves were collected for each construct (19-nt construct: $N = 30$; 10-nt construct: $N = 17$) in the absence of complementary oligonucleotides. The XWLC model is shown (*red dotted line*) using $h_{ss} = 0.60$ nm/nt and $P_{ss} = 1.07$ nm. The model is fitted to an arbitrary extension offset resulting from bead size variation and other factors.

A.7. Protein expression, purification, and labeling

*A.7.1. FRET-labeled UvrD**

*This section is modified from Comstock et al. (2015) (112) with permission from AAAS.

Proteins were expressed, purified, and labeled by the laboratory of Prof. Timothy M. Lohman as described in (219). Briefly, in order to label the protein with a FRET donor-acceptor pair, UvrDΔCys mutants were first constructed, replacing all six native cysteine residues (C52, C181, C322, C350, C441, and C640) with serine. In order to label UvrD with two fluorophores using maleimide chemistry, two Cys residues were then reintroduced at the desired positions (A100 and A473) by substituting Cys for Ala, yielding the mutant 6×His-UvrDΔCys[A100C,A473C] (hereafter referred to as dual-mutant UvrD or dmUvrD). Each mutated position was exposed on the surface of UvrD to enable convenient fluorescent labeling.

After expression and purification, dmUvrD was labeled stochastically with an equimolar mixture of AlexaFluor555 (donor) and AlexaFluor647 (acceptor) maleimides. The average labeling efficiencies per protein were 85% for AlexaFluor555 and 72% for AlexaFluor647. The ssDNA-stimulated steady-state ATPase activity was reduced by about 25% for dmUvrD compared to wtUvrD (219). The ensemble average translocation rate of dmUvrD (155 ± 10 nt/s) was slightly slower than that of wtUvrD (191 ± 3 nt/s) (219). DNA unwinding velocities of dmUvrD was within 10% of those of wtUvrD as measured using the optical trap hairpin assay.

A.7.2. Wild-type MutL

MutL was provided by the laboratory of Prof. Timothy M. Lohman. It was expressed and purified as described previously (244). The protein was stored in 50 mM potassium phosphate, 50 mM KCl, 0.1 mM EDTA, and 1 mM β-mercaptoethanol (BME). The presence of both phosphate and BME were needed to prevent protein aggregation, as observed by analytical ultracentrifugation experiments (244).

Appendix B

Modeling

B.1. Probability of spontaneous alignment of oligonucleotides*

* This section is reprinted from Whitley et al. (2017) (108).

In our model, hybridization proceeds through a "pre-alignment" of the two complementary DNA strands. Here, we are interested in estimating the probability that an oligonucleotide of length ℓ free in solution can be pre-aligned to a strand stretched under tension, as in our experiments. We used the inextensible WLC model to estimate the molecular extension of the free oligonucleotide. Integrating the Marko-Siggia interpolation formula (257) for the WLC model, we obtain an analytical expression for the free energy of stretching the oligonucleotide to a fractional extension $\xi \equiv X / L$:

$$G(\xi, L) = \frac{k_B T}{P} L \left(\frac{\xi^2}{4(1-\xi)} + \frac{\xi^2}{2} \right) \tag{B.1}$$

where L is the contour length and P is the persistence length. For ssDNA, we took $L = \ell h$ with h = 0.6 nm/nt and P = 1.32 nm as described in **Section 2.3.1**.

From this expression, the probability that an oligonucleotide of contour length L is extended to a fractional extension ξ_0 or beyond is given by:

$$p(L, \xi \geq \xi_0) = \frac{1}{Z(L)} \int_0^{2\pi} d\phi \int_0^{\theta_0} d\theta \sin\theta \int_{\xi_0}^{1} d\xi\, \xi^2 e^{-G(\xi,L)/k_B T}, \tag{B.2}$$

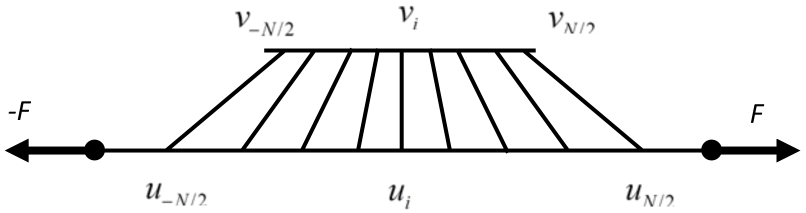
where $Z(L)$ is the partition function,

$$Z(L) = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta \int_0^{1} d\xi\, \xi^2 e^{-G(\xi,L)/k_B T}, \tag{B.3}$$

and the polar and azimuthal angles θ and ϕ are defined relative to axis along which tension to the tethered strand is applied. In Eq. (B.2), the angle θ_0 in the integration accounts for the fact that the pre-aligned oligonucleotide must not only be extended but also oriented parallel to the tethered strand to within some tolerance angle θ_0.

We evaluated the probability $p(L, \xi \geq \xi_0)$ for different contour lengths $L = \ell h$, given reasonable choices for ξ_0 and θ_0. We selected $\xi_0 \approx 0.7$, which corresponds to the fractional extension of ssDNA at a force of F = 10 pN, and $\theta_0 \approx 30°$, which is roughly the variation in orientation angle of the tethered ssDNA expected at the same force. The probability is not significantly affected by the exact choice of parameter values. FCS measurements have shown that only 1 in 100–1000 diffusional attempts produces a successful duplex (100), consistent with the binding rates we measure. Thus, we determined the range of contour lengths L that would generate $p(L, \xi \geq \xi_0)$ between 10^{-2} and 10^{-3}. This calculation yields lengths ranging between $\ell = 5$ and 13 nt, in line with the oligonucleotide probe lengths assayed in our experiments. The upper bound of this range also provides an estimate for the maximum probe length that can pre-align at a reasonable frequency to the tethered strand.

B.2. Derivation of shearing energy for 3’-5’ pulling orientation



Beginning with a Hamiltonian for our shearing energy:

$$H = \sum_{i=-N/2}^{N/2-1} \frac{1}{2} Q\left[\left(u_{i+1}-u_i\right)^2 + \left(v_{i+1}-v_i\right)^2\right] + \sum_{i=-N/2}^{N/2} \frac{1}{2} R\left(u_i - v_i\right)^2 - F \cdot (u_{N/2} - u_{-N/2}) \tag{B.4}$$

where the $u_{-N/2}$ is the part distinguishing our pulling geometry (3’-5’) from others (5’-5’ or 3’-3’) (184,258). Q is the spring constant for stretching bases along the backbone, and R is the spring constant for stretching base pairs across the duplex.

Defining

$$y_i \equiv \frac{u_i - v_i}{\sqrt{2}} \tag{B.5}$$

and

$$x_i \equiv \frac{u_i + v_i}{\sqrt{2}} \tag{B.6}$$

as done by Prakash and Singh (258), we get

$$H=\sum_{i=-N/2}^{N/2}\frac{1}{2}Q\Big[\left(x_{i+1}-x_i\right)^2+\left(y_{i+1}-y_i\right)^2\Big]+\sum_{i=-N/2}^{N/2}Ry_i^{\ 2}-\frac{F}{\sqrt{2}}\cdot\Big[(x_{N/2}+y_{N/2})-\left(x_{-N/2}+y_{-N/2}\right)\Big]$$

(B.7)

which separates into

$$H_x=\sum_{i=-N/2}^{N/2}\frac{1}{2}Q\left(x_{i+1}-x_i\right)^2-\frac{F}{\sqrt{2}}\left(x_{N/2}-x_{-N/2}\right) \tag{B.8}$$

and

$$H_y=\sum_{i=-N/2}^{N/2}\frac{1}{2}Q\left(y_{i+1}-y_i\right)^2+\sum_{i=-N/2}^{N/2}Ry_i^{\ 2}-\frac{F}{\sqrt{2}}\left(y_{N/2}-y_{-N/2}\right) \tag{B.9}$$

Solving H_x for equilibrium conditions:

$$\frac{\partial H_x}{\partial x_i}=-Q\left(x_{i-1}-2x_i+x_{i+1}\right)=0 \tag{B.10}$$

and taking this to the continuum limit gives

$$\frac{\partial H_x}{\partial x_i}=-Q\frac{d^2x}{di^2}=0 \tag{B.11}$$

so

$$x=c_3i+c_4 \tag{B.12}$$

and since $x_0 = 0$, $c_4 = 0$.

Differentiating then with respect to the last base pair (N/2), we get

$$\begin{aligned}\frac{\partial H_x}{\partial x_{N/2}}&=Q\left(x_{N/2}-x_{N/2-1}\right)-\frac{F}{\sqrt{2}}\\&=Q\left(c_3\frac{N}{2}-c_3\left(\frac{N}{2}-1\right)\right)-\frac{F}{\sqrt{2}}\\&=Qc_3-\frac{F}{\sqrt{2}}=0\end{aligned} \tag{B.13}$$

so

$$c_3=\frac{F}{Q\sqrt{2}} \tag{B.14}$$

and hence

$$x_i = \frac{Fi}{Q\sqrt{2}} \tag{B.15}$$

For y_i, we do the same thing:

$$\frac{\partial H_y}{\partial y_i} = -Q\left(y_{i-1} - 2y_i + y_{i+1}\right) + 2Ry_i = 0 \tag{B.16}$$

again taking this to the continuum limit,

$$\frac{\partial H_y}{\partial y_i} = -Q\frac{d^2 y}{di^2} + 2Ry = 0 \tag{B.17}$$

the solution of which is a combination of exponentials:

$$y_i = c_1 e^{\chi i} + c_2 e^{-\chi i} \tag{B.18}$$

where

$$\chi \equiv \sqrt{\frac{2R}{Q}} \tag{B.19}$$

In our pulling geometry, the displacement of a base pair at $i = 0$ is 0, since it is in the middle. So,

$$y_0 = c_1 + c_2 = 0 \tag{B.20}$$

and

$$y_i = c_1\left(e^{\chi i} - e^{-\chi i}\right) = 2c_1 \sinh(\chi i) \tag{B.21}$$

Then to get c_1 we again differentiate with respect to the last base pair:

$$\begin{aligned} \frac{\partial H_y}{\partial y_{N/2}} &= Q\left(y_{N/2} - y_{N/2-1}\right) + 2Ry_{N/2} - \frac{F}{\sqrt{2}} \\ &= Q\left[2c_1 \sinh\left(\frac{\chi N}{2}\right) - 2c_1 \sinh\left(\chi\left(\frac{N}{2} - 1\right)\right)\right] + 2R\left[2c_1 \sinh\left(\frac{\chi N}{2}\right)\right] - \frac{F}{\sqrt{2}} \\ &= 0 \end{aligned} \tag{B.22}$$

then making use of the identity $\sinh(a-b) = \sinh(a)\cosh(b) - \cosh(a)\sinh(b)$,

$$\begin{aligned} \frac{\partial H_y}{\partial y_{N/2}} = Q&\left[2c_1 \sinh\left(\frac{\chi N}{2}\right) - 2c_1\left[\sinh\left(\frac{\chi N}{2}\right)\cosh\left(\chi\right) - \cosh\left(\frac{\chi N}{2}\right)\sinh\left(\chi\right)\right]\right] \\ &+2R\left[2c_1 \sinh\left(\frac{\chi N}{2}\right)\right] - \frac{F}{\sqrt{2}} \end{aligned} \tag{B.23}$$

Using a small angle approximation for χ,

$$\begin{aligned}\frac{\partial H_y}{\partial y_{N/2}} &= Q\left[2c_1\sinh\left(\frac{\chi N}{2}\right)-2c_1\left[\sinh\left(\frac{\chi N}{2}\right)-\chi\cosh\left(\frac{\chi N}{2}\right)\right]\right]+4Rc_1\sinh\left(\frac{\chi N}{2}\right)-\frac{F}{\sqrt{2}}\\ &= 2c_1Q\chi\cosh\left(\frac{\chi N}{2}\right)+4Rc_1\sinh\left(\frac{\chi N}{2}\right)-\frac{F}{\sqrt{2}}\\ &= 0\end{aligned}$$

(B.24)

$$\frac{c_1Q\chi\cosh\left(\frac{\chi N}{2}\right)}{2R\sinh\left(\frac{\chi N}{2}\right)}+c_1=\frac{F}{\sqrt{2}}\frac{1}{4R\sinh\left(\frac{\chi N}{2}\right)} \tag{B.25}$$

$$c_1\chi^{-1}\coth\left(\frac{\chi N}{2}\right)+c_1=\frac{F}{\sqrt{2}}\frac{1}{4R\sinh\left(\frac{\chi N}{2}\right)} \tag{B.26}$$

So,

$$c_1=\frac{F}{\sqrt{2}}\frac{1}{4R\sinh\left(\frac{\chi N}{2}\right)\left[\chi^{-1}\coth\left(\frac{\chi N}{2}\right)+1\right]} \tag{B.27}$$

Then,

$$y_i=\frac{F}{\sqrt{2}}\frac{2\sinh(\chi i)}{4R\sinh\left(\frac{\chi N}{2}\right)\left[\chi^{-1}\coth\left(\frac{\chi N}{2}\right)+1\right]} \tag{B.28}$$

Then plugging these back into our original Hamiltonians,

$$\begin{aligned}H_x &= \sum_{i=-N/2}^{N/2-1}\frac{1}{2}Q\left(\frac{F(i+1)}{Q\sqrt{2}}-\frac{Fi}{Q\sqrt{2}}\right)^2-\frac{F}{\sqrt{2}}\left(\frac{F(N/2)}{Q\sqrt{2}}-\frac{F(-N/2)}{Q\sqrt{2}}\right)\\ &= \sum_{i=-N/2}^{N/2-1}\frac{1}{2}Q\left(\frac{F}{Q\sqrt{2}}\right)^2-\frac{F^2N}{2Q}\\ &= \frac{F^2(N-1)}{4Q}-\frac{F^2N}{2Q}\\ &= -\frac{F^2(N+1)}{4Q}\end{aligned} \tag{B.29}$$

which is the energy of holding *N* spring in series at a force *F*.

$$\begin{aligned} H_y &= \sum_{i=-N/2}^{N/2} \frac{1}{2} Q\left[2c_1 \sinh(\chi(i+1)) - 2c_1 \sinh(\chi i)\right]^2 + \sum_{i=-N/2}^{N/2} R\left[2c_1 \sinh(\chi i)\right]^2 - \frac{F}{\sqrt{2}}\left[2c_1 \sinh\left(\frac{\chi N}{2}\right) - 2c_1 \sinh\left(-\frac{\chi N}{2}\right)\right] \\ &= \sum_{i=-N/2}^{N/2} 2Qc_1{}^2\left[\sinh(\chi i + \chi) - \sinh(\chi i)\right]^2 + \sum_{i=-N/2}^{N/2} 4Rc_1{}^2 \sinh^2(\chi i) - \frac{4F}{\sqrt{2}} c_1 \sinh\left(\frac{\chi N}{2}\right) \\ &= \sum_{i=-N/2}^{N/2} 2Qc_1{}^2\left[\sinh(\chi i)\cosh(\chi) + \cosh(\chi i)\sinh(\chi) - \sinh(\chi i)\right]^2 + \sum_{i=-N/2}^{N/2} 4Rc_1{}^2 \sinh^2(\chi i) - \frac{4F}{\sqrt{2}} c_1 \sinh\left(\frac{\chi N}{2}\right) \end{aligned}$$

(B.30)

Again using a small angle approximation for χ,

$$\begin{aligned} &= \sum_{i=-N/2}^{N/2} 2Qc_1{}^2\left[\sinh(\chi i) + \chi\cosh(\chi i) - \sinh(\chi i)\right]^2 + \sum_{i=-N/2}^{N/2} 4Rc_1{}^2 \sinh^2(\chi i) - \frac{4F}{\sqrt{2}} c_1 \sinh\left(\frac{\chi N}{2}\right) \\ &= \sum_{i=-N/2}^{N/2} 2Qc_1{}^2\chi^2 \cosh^2(\chi i) + \sum_{i=-N/2}^{N/2} 4Rc_1{}^2 \sinh^2(\chi i) - \frac{4F}{\sqrt{2}} c_1 \sinh\left(\frac{\chi N}{2}\right) \\ &= 2c_1{}^2 \sum_{i=-N/2}^{N/2}\left[Q\chi^2 \cosh^2(\chi i) + 2R\sinh^2(\chi i)\right] - \frac{4F}{\sqrt{2}} c_1 \sinh\left(\frac{\chi N}{2}\right) \end{aligned}$$

(B.31)

Since $\chi^2 = \dfrac{2R}{Q}$,

$$H_y = 2c_1{}^2 \sum_{i=-N/2}^{N/2}\left[2\,\mathrm{R}\cosh^2(\chi i) + 2R\sinh^2(\chi i)\right] - \frac{4F}{\sqrt{2}} c_1 \sinh\left(\frac{\chi N}{2}\right) \qquad \text{(B.32)}$$

and since $\cosh^2(x) + \sinh^2(x) = \cosh(2x)$,

$$H_y = 4Rc_1{}^2 \sum_{i=-N/2}^{N/2} \cosh(2\chi i) - \frac{4F}{\sqrt{2}} c_1 \sinh\left(\frac{\chi N}{2}\right) \qquad \text{(B.33)}$$

or more explicitly,

$$H_y = \frac{F^2 \displaystyle\sum_{i=-N/2}^{N/2} \cosh(2\chi i)}{8R\sinh^2\left(\dfrac{\chi N}{2}\right)\left[\chi^{-1}\coth\left(\dfrac{\chi N}{2}\right) + 1\right]^2} - \frac{F^2}{2R\left[\chi^{-1}\coth\left(\dfrac{\chi N}{2}\right) + 1\right]} \qquad \text{(B.34)}$$

finally providing us with

$$H = \frac{F^2 \sum_{i=-N/2}^{N/2} \cosh(2\chi i)}{8R\sinh^2\left(\frac{\chi N}{2}\right)\left[\chi^{-1}\coth\left(\frac{\chi N}{2}\right)+1\right]^2} - \frac{F^2}{2R\left[\chi^{-1}\coth\left(\frac{\chi N}{2}\right)+1\right]} - \frac{F^2 N}{4Q} \tag{B.35}$$

The second and third terms do not depend on base-pair position, and are merely energy offsets that drop out later when we compute expectation values for extension and number of base-pairs frayed.

B.3. Relation of stretch modulus to spring constant Q

The stretch modulus K is related to Young's modulus E via

$$K = AE \tag{B.36}$$

where A is area. Young's modulus is defined as

$$\frac{F}{A} = E\left(\frac{\Delta X}{X_o}\right) \tag{B.37}$$

where ΔX is the strain and X_o is the original length of the full rod. Combining these two expressions gives

$$K = F\left(\frac{X_o}{\Delta X}\right) \tag{B.38}$$

In the long-polymer limit where we can ignore shearing effects, we treat each base-pair as a harmonic spring with constant Q:

$$F = Q\Delta x \tag{B.39}$$

where Δx is the displacement per base pair (to be distinguished from ΔX above, which is strain of the entire rod). For a long polymer this means that we have two parallel strands of N springs in series, giving us an overall relation

$$F = Q_{eff}\Delta X = \frac{2Q}{N}\Delta X \tag{B.40}$$

Plugging this into our expression for K, we have finally

$$K = \frac{2Q}{N}X_o \tag{B.41}$$

Since X_o is the unstressed contour length of the rod, this is

$$X_o = Nh_{ds} \tag{B.42}$$

where h_{ds} is the helix rise of dsDNA. So, we get a final relation of

$$K = 2Qh_{ds} \tag{B.43}$$

and with a literature stretch modulus value of $K = 1100$ pN (120) and $h_{ds} = 0.34$ nm/bp, we then get $Q = 1620$ pN/nm.

B.4. Probability of spontaneous hairpin opening*

*This section is modified from Comstock et al. (2015) (112) with permission from AAAS.

All measurements using hairpins were taken with DNA under tension by the optical traps. Force was applied in a direction perpendicular to the DNA hairpin strands and can reduce the stability of the base-pairs at the base of the hairpin duplex. The destabilizing effect of force can be understood by writing the free energy of opening *n* base-pairs of the *N* base-pair long hairpin:

$$\Delta G_{tot}(n,F) = \sum_{i=n}^{N} \Delta G_{bp}(i) - 2n\int_{0}^{F} x(F')dF' \tag{B.44}$$

The first term represents the base-pairing energies for base pairs *n* to *N* and the second is the energy of stretching 2*n* released nucleotides at force *F* (224). As force increases, the second, force-dependent energy term overcomes the base-pairing energy, and the hairpin base-pairs unzip spontaneously.

To determine the effect of DNA sequence on UvrD activity, we calculated the probability $P_{closed}(p,F)$ the hairpin remains closed at a given position *p* under tension *F*, similar to the method of Johnson et al. (224). This quantity was calculated by summing over all possible unzipped states from *p* to the end of the hairpin stem:

$$P_{closed}(p,F) = \frac{1}{1 + \sum_{n=p+1}^{N} \exp\left(-\frac{\Delta G_{tot}(n,F)}{k_B T}\right)} \tag{B.45}$$

where *n* is the number of base-pairs opened upstream of *p*, *N* is the total number of base-pairs in the hairpin stem, $\Delta G_{tot}(n,F)$ is the energy associated with releasing *n* base-pairs upstream of the fork junction, k_B is Boltzmann's constant, and *T* is temperature. The free energy of opening base-pairs 1 to *n* is given by Eq. (B.44), which is the sum of two terms: the base-pairing energies for base-pairs *n* to *N* and the energy of stretching 2*n* released nucleotides at force *F*. The base-pairing free energies ΔG_{bp} were obtained from the measured 10 nearest-neighbor and 1 loop free energies as described in Huguet et al. (92), and were adjusted to our buffer conditions using an 'effective' monovalent ion concentration

$\left[Mon^{+}\right]=\frac{1}{2}\left[Tris^{+}\right]+\left[Na^{+}\right]+\beta\sqrt{\left[Mg^{2+}\right]}$ (259), where previous work has shown β~8 for our experimental conditions (111).

References:

1. Fairman-Williams, M.E., Guenther, U.-P. and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology*, **20**, 313-324.
2. Cozzarelli, N.R., Cost, G.J., Nöllmann, M., Viard, T. and Stray, J.E. (2006) Giant proteins that move DNA: bullies of the genomic playground. *Nature Reviews Molecular Cell Biology*, **7**, 580-588.
3. Ngo, T.T.M., Zhang, Q., Zhou, R., Yodh, J.G. and Ha, T. (2015) Asymmetric Unwrapping of Nucleosomes under Tension Directed by DNA Local Flexibility. *Cell*, **160**, 1135-1144.
4. Wuite, G.J., Smith, S.B., Young, M., Keller, D. and Bustamante, C. (2000) Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature*, **404**, 103-106.
5. Qu, X., Wen, J.-d., Lancaster, L., Noller, H.F., Bustamante, C. and Tinoco, I. (2011) The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature*, **475**, 118-121.
6. Dietz, H., Douglas, S.M. and Shih, W.M. (2009) Folding DNA into twisted and curved nanoscale shapes. *Science*, **325**, 725-730.
7. Wang, X. and Ha, T. (2013) Defining single molecular forces required to activate integrin and notch signaling. *Science*, **340**, 991-994.
8. Ouldridge, T.E., Hoare, R.L., Louis, A.A., Doye, J.P.K., Bath, J. and Turberfield, A.J. (2013) Optimizing DNA nanotechnology through coarse-grained modeling: A two-footed DNA walker. *ACS Nano*, **7**, 2479-2490.
9. Patel, S.S. and Donmez, I. (2006) Mechanisms of helicases. *Journal of Biological Chemistry*, **281**, 18265-18268.
10. Sung, P., Bailly, V., Weber, C., Thompson, L.H., Prakash, L. and Prakash, S. (1993) Human xeroderma pigmentosum group D gene encodes a DNA helicase. *Nature*, **365**, 852-855.
11. Ellis, N.A., Groden, J., Ye, T.-Z., Straughen, J., Lennon, D.J., Ciocci, S., Proytcheva, M. and German, J. (1995) The Bloom's Syndrome Gene Product Is Homologous to RecQ Helicases. *Cell*, **83**, 655-666.
12. Gray, M.D., Shen, J.-C., Kamath-Loeb, A.S., Blank, A., Sopher, B.L., Martin, G.M., Oshima, J. and Loeb, L.A. (1997) The Werner syndrome protein is a DNA helicase. *Nature Genetics*, **17**, 100-103.
13. Yang, W. (2010) Lessons learned from UvrD helicase: mechanism for directional movement. *Annual Review of Biophysics*, **39**, 367-385.
14. Gorbalenya, E. and Koonin, E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology*, **3**, 419-429.
15. Berger, J.M. (2008) SnapShot: Nucleic Acid Helicases and Translocases. *Cell*, **134**, 888-889.
16. Yamaguchi, M., Dao, V. and Modrich, P. (1998) MutS and MutL Activate DNA Helicase II in a Mismatch-dependent Manner. *Journal of Biological Chemistry*, **273**, 9197-9201.
17. Husain, I., Van Houten, B., Thomas, D.C., Abdel-Monem, M. and Sancar, A. (1985) Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease. *Proceedings of the National Academy of Sciencees of the United States of America*, **82**, 6774-6778.
18. Epshtein, V., Kamarthapu, V., McGary, K., Svetlov, V., Ueberheide, B. and Proshkin, S. (2014) UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature*, **505**, 372-377.
19. Flores, M.J., Bidnenko, V. and Michel, B. (2004) The DNA repair helicase UvrD is essential for replication fork reversal in replication mutants. *EMBO reports*, **5**, 983-988.
20. Payne, B.T.I., van Knippenberg, I.C., Bell, H., Filipe, S.R., Sherratt, D.J. and Mcglynn, P. (2006) Replication fork blockage by transcription factor-DNA complexes in Escherichia coli. *Nucleic Acids Research*, **34**, 5194-5202.

21. Bruand, C. and Ehrlich, S.D. (2000) UvrD-dependent replication of rolling-circle plasmids in Escherichia coli. *Molecular Microbiology*, **35**, 204-210.
22. Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F. and Petit, M.-A. (2005) UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli. *The EMBO Journal*, **24**, 180-189.
23. Lestini, R. and Michel, B. (2007) UvrD controls the access of recombination proteins to blocked replication forks. *The EMBO Journal*, **26**, 3804-3814.
24. Petrova, V., Chen, S.H., Molzberger, E.T., Tomko, E.J., Chitteni-Pattu, S., Jia, H., Ordabayev, Y., Lohman, T.M. and Cox, M.M. (2015) Active displacement of RecA filaments by UvrD translocase activity. *Nucleic Acids Research*, **43**, 1-17.
25. Myong, S., Rasnik, I., Joo, C., Lohman, T.M. and Ha, T. (2005) Repetitive shuttling of a motor protein on DNA. *Nature*, **437**, 1321-1325.
26. Park, J., Myong, S., Niedziela-Majka, A., Lee, K.S., Yu, J., Lohman, T.M. and Ha, T. (2010) PcrA helicase dismantles RecA filaments by reeling in DNA in uniform steps. *Cell*, **142**, 544-555.
27. Qiu, Y., Antony, E., Doganay, S., Koh, H.R., Lohman, T.M. and Myong, S. (2013) Srs2 prevents Rad51 filament formation by repetitive motion on DNA. *Nature Communications*, **4**, 2281.
28. Tomko, E.J., Jia, H., Park, J., Maluf, N.K., Ha, T. and Lohman, T.M. (2010) 5'-Single-stranded/duplex DNA junctions are loading sites for E. coli UvrD translocase. *The EMBO Journal*, **29**, 3826-3839.
29. Dessinges, M.-N., Lionnet, T., Xi, X.G., Bensimon, D. and Croquette, V. (2004) Single-molecule assay reveals strand switching and enhanced processivity of UvrD. *Proceedings of the National Academy of Sciencees of the United States of America*, **101**, 6439-6444.
30. Greenleaf, W.J., Woodside, M.T. and Block, S.M. (2007) High-Resolution, Single-Molecule Measurements of Biomolecular Motion. *Annual Review of Biophysics and Biomolecular Structure*, **36**, 171-190.
31. Neuman, K.C. and Nagy, A. (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods*, **5**, 491-505.
32. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. and Gaub, H.E. (1997) Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science*, **276**, 1109-1112.
33. Woodside, M.T. and Block, S.M. (2014) Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annual Review of Biophysics*, **43**, 19-39.
34. Block, S.M., Goldstein, L.S.B. and Schnapp, B.J. (1990) Bead movement by single kinesin molecules studied with optical tweezers. *Nature*, **348**, 348-352.
35. Svoboda, K. and Block, S.M. (1994) Force and Velocity Measured for Single Kinesin Molecules. *Cell*, **77**, 773-784.
36. Finer, J.T., Simmons, R.M. and Spudich, J.A. (1993) Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature*, **368**, 113-119.
37. Chemla, Y.R. (2010) Revealing the base pair stepping dynamics of nucleic acid motor proteins with optical traps. *Physical chemistry chemical physics*, **12**, 3080-3095.
38. Neuman, K.C. and Block, S.M. (2004) Optical trapping. *Review of Scientific Instruments*, **75**, 2787-2809.
39. Gosse, C. and Croquette, V. (2002) Magnetic Tweezers: Micromanipulation and Force Measurement at the Molecular Level. *Biophysical Journal*, **82**, 3314-3329.
40. Molloy, J.E., Burns, J.E., Sparrow, J.C., Tregear, R.T., Kendrick-jones, J. and White, D.C.S. (1995) Single-Molecule Mechanics of Heavy Meromyosin and S1 Interacting with Rabbit or Drosophila Actins Using Optical Tweezers. *Biophysical Journal*, **68**, 298-303.
41. Gittes, F. and Schmidt, C.F. (1998) Interference model for back-focal-plane displacement detection in optical tweezers. *Optics Letters*, **23**, 7-9.

42. Pralle, A., Prummer, M., Florin, E.-L., Stelzer, E.H.K. and Horber, J.K. (1999) Three-Dimensional High-Resolution Particle Tracking for Optical Tweezers by Forward Scattered Light. *Microscopy Research and Technique*, **44**, 378-386.
43. Berg-Sørensen, K. and Flyvbjerg, H. (2004) Power spectrum analysis for optical tweezers. *Review of Scientific Instruments*, **75**, 594-612.
44. Carter, A.R., King, G.M., Ulrich, T.A., Halsey, W., Alchenberger, D. and Perkins, T.T. (2007) Stabilization of an optical microscope to 0.1 nm in three dimensions. *Applied Optics*, **46**, 421-427.
45. Moffitt, J.R., Chemla, Y.R., Izhaky, D. and Bustamante, C. (2006) Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proceedings of the National Academy of Sciencees of the United States of America*, **103**, 9006-9011.
46. Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R. and Block, S.M. (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature*, **438**, 460-465.
47. Shaevitz, J.W., Abbondanzieri, E.A., Landick, R. and Block, S.M. (2003) Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*, **426**.
48. Visscher, K., Gross, S.P. and Block, S.M. (1996) Construction of Multiple-Beam Optical Traps with Nanometer-Resolution Position Sensing. *IEEE Journal of Selected Topics in Quantum Electronics*, **2**, 1066-1076.
49. Gordon, M.P., Ha, T. and Selvin, P.R. (2004) Single-molecule high-resolution imaging with photobleaching. *Proceedings of the National Academy of Sciencees of the United States of America*, **101**, 6462-6465.
50. Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-schwartz, J. and Hess, H.F. (2006) Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, **313**, 1642-1646.
51. Rust, M.J., Bates, M. and Zhuang, X. (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, **3**, 793-795.
52. Ulbrich, M.H. and Isacoff, E.Y. (2007) Subunit counting in membrane-bound proteins. *Nature Methods*, **4**, 319-321.
53. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. and Ha, T. (2008) Advances in single-molecule fluorescence methods for molecular biology. *Annual Review of Biochemistry*, **77**, 51-76.
54. Hwang, L.C., Hohlbein, J., Holden, S.J. and Kapanidis, A.N. (2009) In Hinterdorfer, P. and van Oijen, A. (eds.), *Handbook of Single-Molecule Biophysics*. Springer.
55. Ha, T. (2001) Single-molecule fluorescence resonance energy transfer. *Methods*, **25**, 78-86.
56. Roy, R., Hohng, S. and Ha, T. (2008) A practical guide to single molecule FRET. *Nature Methods*, **5**, 507-516.
57. Dudko, O., Graham, T. and Best, R. (2011) Locating the Barrier for Folding of Single Molecules under an External Force. *Physical Review Letters*, **107**, 15-18.
58. van Dijk, M.A., Kapitein, L.C., van Mameren, J., Schmidt, C.F. and Peterman, E.J.G. (2004) Combining optical trapping and single-molecule fluorescence spectroscopy: enhanced photobleaching of fluorophores. *The Journal of Physical Chemistry B*, **108**, 6479-6484.
59. Hohng, S., Zhou, R., Nahas, M.K., Yu, J., Schulten, K., Lilley, D.M.J. and Ha, T. (2007) Fluorescence-Force Spectroscopy Maps Two-Dimensional Reaction Landscape of the Holliday Junction. *Science*, **318**, 279-284.
60. Brau, R.R., Tarsa, P.B., Ferrer, J.M., Lee, P. and Lang, M.J. (2006) Interlaced optical force-fluorescence measurements for single molecule biophysics. *Biophysical Journal*, **91**, 1069-1077.
61. Comstock, M.J., Ha, T. and Chemla, Y.R. (2011) Ultrahigh-resolution optical trap with single-fluorophore sensitivity. *Nature Methods*, **8**, 335/340.
62. Whitley, K.D., Comstock, M.J. and Chemla, Y.R. (2017) In Gennerich, A. (ed.), *Methods in Molecular Biology*. Springer, Vol. 1486, pp. 183-256.

63. Whitley, K.D., Comstock, M.J. and Chemla, Y.R. (2017) In Chemla, Y. R. and Spies, M. (eds.), *Methods in Enzymology*. Elsevier, Vol. 582, pp. 137-169.
64. Friedman, R.C., Farh, K.K.-h., Burge, C.B. and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, **19**, 92-105.
65. Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *cell*, **120**, 15-20.
66. Fei, J., Singh, D., Zhang, Q., Park, S., Balasubramanian, D., Golding, I., Vanderpool, C.K. and Ha, T. (2015) Determination of in vivo target search kinetics of regulatory noncoding RNA. *Science*, **347**, 1371-1374.
67. Lambert, N.J., Gu, S.G. and Zahler, A.M. (2011) The conformation of microRNA seed regions in native microRNPs is prearranged for presentation to mRNA targets. *Nucleic Acids Research*, **39**, 4827-4835.
68. Lewis, B.P., Shih, I.-h., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of Mammalian MicroRNA Targets. *Cell*, **115**, 787-798.
69. Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T. and Patel, D.J. (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, **4566**, 921-926.
70. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. and Patel, D.J. (2008) Structure of the guide-strand-containing argonaute silencing complex. *Nature*, **456**, 209-213.
71. Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T. and Patel, D.J. (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Molecular Cell*, **2005**, 3.
72. Swarts, D.C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.J. *et al.* (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature*, **507**, 258-261.
73. Erlich, H.A., Gelfand, D. and Sninsky, J.J. (1991) Recent Advances in the Polymerase Chain Reaction. *Science*, **252**, 1643-1651.
74. Stoughton, R.B. (2005) Applications of DNA microarrays in biology. *Annual Review of Biochemistry*, **74**, 53-82.
75. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. and Tyagi, S. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, **5**, 877-879.
76. Wiedenheft, B., Sternberg, S.H. and Doudna, J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, **482**, 331-338.
77. Barrangou, R. and Doudna, J.A. (2016) Applications of CRISPR technologies in research and beyond. *Nature Biotechnology*, **34**, 933-941.
78. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62-67.
79. Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V. and Seidel, R. (2014) Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proceedings of the National Academy of Sciencees of the United States of America*, **111**, 9798-9803.
80. Knight, S.C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L.B., Bosanac, L., Zhang, E.T., El Beheiry, M., Masson, J.-B., Dahan, M. *et al.* (2015) Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*, **350**, 823-826.
81. Jiang, F., Zhou, K., Ma, L., Gressel, S. and Doudna, J.A. (2015) A Cas9 – guide RNA complex preorganized for target DNA recognition. *Science*, **348**, 1477-1481.
82. Yurke, B., Turber, A.J., Jr, A.P.M., Simmel, F.C. and Neumann, J.L. (2000) A DNA-fuelled molecular machine made of DNA. *Nature*, **406**, 605-608.

83. Sherman, W.B. and Seeman, N.C. (2004) A Precisely Controlled DNA Biped Walking Device. *Nano Letters*, **4**, 1203-1207.
84. Omabegho, T., Sha, R. and Seeman, N.C. (2009) A Bipedal DNA Brownian Motor with Coordinated Legs. *Science*, **324**, 67-71.
85. Tomov, T.E., Tsukanov, R., Liber, M., Masoud, R., Plavner, N. and Nir, E. (2013) Rational design of DNA motors: fuel optimization through single-molecule fluorescence. *Journal of the American Chemical Society*, **135**, 11953-11941.
86. Seelig, G., Soloveichik, D., Zhang, D.Y. and Winfree, E. (2006) Enzyme-free nucleic acid logic circuits. *Science*, **314**, 1585-1588.
87. Qian, L. and Winfree, E. (2011) Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science*, **332**, 1196-1201.
88. Srinivas, N., Ouldridge, T.E., Sulc, P., Schaeffer, J.M., Yurke, B. and Louis, A.A. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Research*, **41**, 10641-10658.
89. Machinek, R.R.F., Ouldridge, T.E., Haley, N.E.C., Bath, J. and Turberfield, A.J. (2014) Programmable energy landscapes for kinetic control of DNA strand displacement. *Nature Communications*, **5**, 5324.
90. SantaLucia, J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciencees of the United States of America*, **95**, 1460-1465.
91. Owczarzy, R., You, Y., Moreira, B.G., Manthey, J.A., Huang, L., Behlke, M.A. and Walder, J.A. (2004) Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry*, **43**, 3537-3554.
92. Huguet, J.M., Bizarro, C.V., Forns, N., Smith, S.B., Bustamante, C. and Ritort, F. (2010) Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proceedings of the National Academy of Sciencees of the United States of America*, **107**, 15431-15436.
93. Ross, P.D. and Sturtevant, J.M. (1960) The Kinetics of Double Helix Formation From Polyriboadenylic Acid and Polyribouridylic Acid. *Proceedings of the National Academy of Sciencees of the United States of America*, **46**, 1360-1365.
94. Eigen, M. and Pörschke, D. (1970) Co-operative Non-enzymic Base Recognition I. Thermodynamics of the Helix-Coil Transition of Oligoriboadenylic Acids at Acidic pH. *Journal of Molecular Biology*, **53**, 123-141.
95. Pörschke, D. and Eigen, M. (1971) Co-operative Non-enzymic Base Recognition III. Kinetics of the Helix-Coil Transition of the Oligoribouridylic . Oligoriboadenylic Acid System and of Oligoriboadenylic Acid alone at Acidic pH. *Journal of Molecular Biology*, **62**, 361-381.
96. Craig, M.E., Crothers, D.M. and Doty, P. (1971) Relaxation Kinetics of Dimer self Complementary Oligonucleotides. *Journal of Molecular Biology*, **62**, 383-401.
97. Pörschke, D. (1974) Model calculations on the kinetics of oligonucleotide double helix coil transitions. Evidence for a fast chain sliding reaction. *Biophysical Chemistry*, **2**, 83-96.
98. Manning, G.S. (1976) On the application of polyelectrolyte limiting laws to the helix-coil transition of DNA. V. Ionic effects on renaturation kinetics. *Biopolymers*, **15**, 1333-1343.
99. Rauzan, B., McMichael, E., Cave, R., Sevcik, L.R., Ostrosky, K., Whitman, E., Stegemann, R., Sinclair, A.L., Serra, M.J. and Deckert, A.A. (2013) Kinetics and thermodynamics of DNA, RNA, and hybrid duplex formation. *Biochemistry*, **52**, 765-772.
100. Dupuis, N.F., Holmstrom, E.D. and Nesbitt, D.J. (2013) Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophysical Journal*, **105**, 756-766.

101. Strunz, T., Oroszlan, K., Schäfer, R. and Güntherodt, H.J. (1999) Dynamic force spectroscopy of single DNA molecules. *Proceedings of the National Academy of Sciencees of the United States of America*, **96**, 11277-11282.
102. Liphardt, J., Onoa, B., Smith, S.B., Tinoco, I. and Bustamante, C. (2001) Reversible unfolding of single RNA molecules by mechanical force. *Science*, **292**, 733-737.
103. Woodside, M.T., Anthony, P.C., Behnke-Parks, W.M., Larizadeh, K., Herschlag, D. and Block, S.M. (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, **314**, 1001-1004.
104. Neupane, K., Ritchie, D.B., Yu, H., Foster, D.a.N., Wang, F. and Woodside, M.T. (2012) Transition Path Times for Nucleic Acid Folding Determined from Energy-Landscape Analysis of Single-Molecule Trajectories. *Physical Review Letters*, **109**, 1-5.
105. Neupane, K., Manuel, A.P., Lambert, J. and Woodside, M.T. (2015) Transition-path probability as a test of reaction-coordinate quality reveals DNA hairpin folding is a one-dimensional diffusive process. *Journal of Physical Chemistry Letters*, **6**, 1005-1010.
106. Woodside, M.T., Behnke-Parks, W.M., Larizadeh, K., Travers, K., Herschlag, D. and Block, S.M. (2006) Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proceedings of the National Academy of Sciencees of the United States of America*, **103**, 6190-6195.
107. Ho, D., Zimmermann, J.L., Dehmelt, F.a., Steinbach, U., Erdmann, M., Severin, P., Falter, K. and Gaub, H.E. (2009) Force-driven separation of short double-stranded DNA. *Biophysical Journal*, **97**, 3158-3167.
108. Whitley, K.D., Comstock, M.J. and Chemla, Y.R. (2017) Elasticity of the transition state for oligonucleotide hybridization. *Nucleic Acids Research*, **45**, 547-555.
109. Landry, M.P., McCall, P.M., Qi, Z. and Chemla, Y.R. (2009) Characterization of photoactivated singlet oxygen damage in single-molecule optical trap experiments. *Biophysical Journal*, **97**, 2128-2136.
110. Rasnik, I., Mckinney, S.A. and Ha, T. (2006) Nonblinking and long- lasting single-molecule fluorescence imaging. *Nature Methods*, **3**, 891-893.
111. Qi, Z., Pugh, R.a., Spies, M. and Chemla, Y.R. (2013) Sequence-dependent base pair stepping dynamics in XPD helicase unwinding. *eLife*, **2**, e00334.
112. Comstock, M.J., Whitley, K.D., Jia, H., Sokoloski, J., Lohman, T.M., Ha, T. and Chemla, Y.R. (2015) Direct observation of structure-function relationship in a nucleic acid – processing enzyme. *Science*, **348**, 352-354.
113. Bell, G.I. (1978) Models for the specific adhesion of cells to cells. *Science*, **200**, 618-627.
114. Manosas, M., Collin, D. and Ritort, F. (2006) Force-dependent fragility in RNA hairpins. *Physical Review Letters*, **96**, 218301.
115. Dudko, O.K., Hummer, G. and Szabo, A. (2008) Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciencees of the United States of America*, **105**, 15755-15760.
116. Cisse, I.I., Kim, H. and Ha, T. (2012) A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nature Structural and Molecular Biology*, **19**, 623-628.
117. Moreira, B.G., You, Y., Behlke, M.a. and Owczarzy, R. (2005) Effects of fluorescent dyes, quenchers, and dangling ends on DNA duplex stability. *Biochemicaal and Biophysical Research Communications*, **327**, 473-484.
118. Moreira, B.G., You, Y. and Owczarzy, R. (2015) Cy3 and Cy5 dyes attached to oligonucleotide terminus stabilize DNA duplexes: Predictive thermodynamic model. *Biophysical Chemistry*, **198**, 36-44.

119. Mosayebi, M., Louis, A.A., Doye, J.P.K. and Ouldridge, T.E. (2015) Force-Induced Rupture of a DNA Duplex: From Fundamentals to Force Sensors. *ACS Nano*, **9**, 11993-12003.

120. Wang, M.D., Yin, H., Landick, R., Gelles, J. and Block, S.M. (1997) Stretching DNA with optical tweezers. *Biophysical Journal*, **72**, 1335-1346.

121. Bustamante, C., Marko, J., Siggia, E. and Smith, S.B. (1994) Entropic Elasticity of lambda-phage DNA. *Science*, **265**, 1599-1600.

122. Saleh, O., McIntosh, D., Pincus, P. and Ribeck, N. (2009) Nonlinear Low-Force Elasticity of Single-Stranded DNA Molecules. *Physical Review Letters*, **102**, 068301.

123. Pörschke, D. (1974) A direct measurement of the unzippering rate of a nucleic acid double helix. *Biophysical Chemistry*, **2**, 97-101.

124. Murphy, M.C., Rasnik, I., Cheng, W., Lohman, T.M. and Ha, T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophysical Journal*, **86**, 2530-2537.

125. Kuznetsov, S.V., Shen, Y., Benight, a.S. and Ansari, A. (2001) A semiflexible polymer model applied to loop formation in DNA hairpins. *Biophysical Journal*, **81**, 2864-2875.

126. Rivetti, C., Walker, C. and Bustamante, C. (1998) Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *Journal of Molecular Biology*, **280**, 41-59.

127. Bizarro, C.V., Alemany, A. and Ritort, F. (2012) Non-specific binding of Na + and Mg 2+ to RNA determined by force spectroscopy methods. *Nucleic Acids Research*, **40**, 6922-6935.

128. Rouzina, I. and Bloomfield, V.A. (2001) Force-induced melting of the DNA double helix 1. Thermodynamic analysis. *Biophysical Journal*, **80**, 882-883.

129. Wetmur, J.G. and Davidson, N. (1968) Kinetics of Renaturation of DNA. *Journal of Molecular Biology*, **31**, 349-370.

130. McIntosh, D.B., Duggan, G., Gouil, Q. and Saleh, O.a. (2014) Sequence-dependent elasticity and electrostatics of single-stranded DNA: Signatures of base-stacking. *Biophysical Journal*, **106**, 659-666.

131. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, **14**, 33-38.

132. Suzuki, Y. and Dudko, O. (2010) Single-Molecule Rupture Dynamics on Multidimensional Landscapes. *Physical Review Letters*, **104**, 048101.

133. Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D.C., van der Oost, J., Patel, D.J. and Wang, Y. (2013) Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proceedings of the National Academy of Sciencees of the United States of America*, **111**, 652-657.

134. Schirle, N.T., Sheu-Gruttadauria, J. and MacRae, I.J. (2014) Structural basis for microRNA targeting. *Science*, **346**, 608-613.

135. Ngo, T.T.M., Yoo, J., Dai, Q., Zhang, Q., He, C., Aksimentiev, A. and Ha, T. (2016) Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nature Communications*, **7**, 10813.

136. Wang, M.D., Schnitzer, M.J., Yin, H., Landick, R., Gelles, J. and Block, S.M. (1998) Force and velocity measured for single molecules of RNA polymerase. *Science*, **282**, 902-907.

137. Galburt, E.A., Grill, S.W., Wiedmann, A., Lubkowska, L., Choy, J., Nogales, E., Kashlev, M. and Bustamante, C. (2007) Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature*, **446**, 820-823.

138. Liu, T., Kaplan, A., Alexander, L., Yan, S., Wen, J.-d., Lancaster, L., Wickersham, C.E., Fredrick, K., Noller, H., Tinoco, I. *et al.* (2014) Direct measurement of the mechanical work during translocation by the ribosome. *eLife*, **3**, e03406.

139. Wen, J.-D., Lancaster, L., Hodges, C., Zeri, A.-C., Yoshimura, S.H., Noller, H.F., Bustamante, C. and Tinoco, I. (2008) Following translation by single ribosomes one codon at a time. *Nature*, **452**, 598-603.
140. Sirinakis, G., Clapier, C.R., Gao, Y., Viswanathan, R., Cairns, B.R. and Zhang, Y. (2011) The RSC chromatin remodelling ATPase translocates DNA with high force and small step size. *The EMBO Journal*, **30**, 2364-2372.
141. Zhang, F., Nangreave, J., Liu, Y. and Yan, H. (2014) Structural DNA nanotechnology: state of the art and future perspective. *Journal of the American Chemical Society*, **136**, 11198-11211.
142. Jones, M.R., Seeman, N.C. and Mirkin, C.A. (2015) Programmable materials and the nature of the DNA bond. *Science*, **347**, 1260901.
143. Rothemund, P.W.K. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature*, **440**, 297-302.
144. Goodman, R.P., Heilemann, M., Doose, S., Erben, C.M., Kapanidis, A.N. and Turberfield, A.J. (2008) Reconfigurable, braced, three-dimensional DNA nanostructures. *Nature Nanotechnology*, **3**, 93-96.
145. Douglas, S.M., Dietz, H., Liedl, T., Hogberg, B., Graf, F. and Shih, W.M. (2009) Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, **459**, 414-418.
146. Zhang, C., Su, M., He, Y., Zhao, X., Fang, P.-a., Ribbe, A.E., Jiang, W. and Mao, C. (2008) Conformational flexibility facilitates self-assembly of complex DNA nanostructures. *Proceedings of the National Academy of Sciencees of the United States of America*, **105**, 10665-10669.
147. Andersen, E.S., Dong, M., Nielsen, M.M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M.M., Sander, B., Stark, H., Oliveira, C.L.P. *et al.* (2009) Self-assembly of a nanoscale DNA box with a controllable lid. *Nature*, **459**, 73-76.
148. Han, D., Pal, S., Nangreave, J., Deng, Z., Liu, Y. and Yan, H. (2011) DNA Origami with Complex Curvatures in Three-Dimensional Space. *Science*, **332**, 342-346.
149. Zadegan, R.M., Jepsen, M.D.E., Thomsen, K.E., Okholm, A.H., Schaffert, D.H., Andersen, E.S., Birkedal, V. and Kjems, J. (2012) Construction of a 4 Zeptoliters Switchable 3D DNA Box Origami. *ACS Nano*, **6**, 10050-10053.
150. Douglas, S.M., Bachelet, I. and Church, G.M. (2012) A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science*, **335**, 831-834.
151. Banerjee, A., Bhatia, D., Saminathan, A., Chakraborty, S., Kar, S. and Krishnan, Y. (2013) Controlled Release of Encapsulated Cargo from a DNA Icosahedron using a Chemical Trigger. *Angewandte Chemie-International Edition*, **52**, 6854-6857.
152. Castro, C.E., Su, H.-J., Marras, A.E., Zhou, L. and Johnson, J. (2015) Mechanical design of DNA nanostructures. *Nanoscale*, **7**, 5913-5921.
153. Marras, A.E., Zhou, L., Su, H.-j. and Castro, C.E. (2015) Programmable motion of DNA origami mechanisms. *Proceedings of the National Academy of Sciencees of the United States of America*, **112**, 713-718.
154. Burns, J.R., Seifert, A., Fertig, N. and Howorka, S. (2015) A biomimetic DNA-based channel for the ligand-controlled transport of charged molecular cargo across a biological membrane. *Nature Nanotechnology*, **11**, 152-156.
155. Le, J.V., Luo, Y., Darcy, M.A., Lucas, C.R., Goodwin, M.F., Poirier, M.G. and Castro, C.E. (2016) Probing Nucleosome Stability with a DNA Origami Nanocaliper. *ACS Nano*, **10**, 7073-7084.
156. Ke, Y., Meyer, T., Shih, W.M. and Bellot, G. (2016) Regulation at a distance of biomolecular interactions using a DNA origami nanoactuator. *Nature Communications*, **7**, 10935.
157. Nickels, P.C., Wünsch, B., Holzmeister, P., Bae, W., Kneer, L.M., Grohmann, D., Tinnefeld, P. and Liedl, T. (2016) Molecular force spectroscopy with a DNA origami – based nanoscopic force clamp. *Science*, **354**, 305-307.

158. Wang, X., Rahil, Z., Li, I.T.S., Chowdhury, F., Leckband, D.E., Chemla, Y.R. and Ha, T. (2016) Constructing modular and universal single molecule tension sensor using protein G to study mechano-sensitive receptors. *Scientific Reports*, **6**, 21584.

159. Delebecque, C.J., Lindner, A.B., Silver, P.A. and Aldaye, F.A. (2011) Organization of intracellular reactions with rationally designed RNA assemblies. *Science*, **333**, 470-474.

160. Smith, S.B., Finzi, L. and Bustamante, C. (1992) Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, **258**, 1122-1126.

161. Smith, S.B., Cui, Y. and Bustamante, C. (1996) Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science*, **271**, 795-799.

162. Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J.-l., Chatenay, D. and Caront, F. (1996) DNA : An Extensible Molecule. *Science*, **271**, 792-794.

163. Odijk, T. (1995) Stiff Chains and Filaments under Tension. *Macromolecules*, **28**, 7016-7018.

164. Gore, J., Bryant, Z., Nöllmann, M., Le, M.U., Cozzarelli, N.R. and Bustamante, C. (2006) DNA overwinds when stretched. *Nature*, **442**, 836-839.

165. Lionnet, T., Joubaud, S., Lavery, R., Bensimon, D. and Croquette, V. (2006) Wringing Out DNA. *Physical Review Letters*, **96**, 178102.

166. Gross, P., Laurens, N., Oddershede, L.B., Bockelmann, U., Peterman, E.J.G. and Wuite, G.J.L. (2011) Quantifying how DNA stretches, melts and changes twist under tension. *Nature Physics*, **7**, 731-736.

167. Rief, M., Clausen-schaumann, H. and Gaub, H.E. (1999) Sequence-dependent mechanics of single DNA molecules. *Nature Structural Biology*, **6**, 346-349.

168. van Mameren, J., Gross, P., Farge, G., Hooijman, P., Modesti, M., Falkenberg, M., Wuite, G.J.L. and Peterman, E.J.G. (2009) Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *Proceedings of the National Academy of Sciencees of the United States of America*, **106**, 18231-18236.

169. Camunas-Soler, J., Ribezzi-Crivellari, M. and Ritort, F. (2016) Elastic Properties of Nucleic Acids by Single-Molecule Force Spectroscopy. *Annual Review of Biophysics*, **45**, 65-84.

170. McIntosh, D., Ribeck, N. and Saleh, O. (2009) Detailed scaling analysis of low-force polyelectrolyte elasticity. *Physical Review E*, **80**, 041803.

171. Dittmore, A., McIntosh, D., Halliday, S. and Saleh, O. (2011) Single-Molecule Elasticity Measurements of the Onset of Excluded Volume in Poly(Ethylene Glycol). *Physical Review Letters*, **107**, 148301.

172. Stevens, M.J., McIntosh, D.B. and Saleh, O.a. (2012) Simulations of Stretching a Strong, Flexible Polyelectrolyte. *Macromolecules*, **45**, 5757-5765.

173. Toan, N.M. and Thirumalai, D. (2012) On the origin of the unusual behavior in the stretching of single-stranded DNA. *The Journal of Chemical Physics*, **136**, 235103.

174. Cloutier, T.E. and Widom, J. (2005) DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proceedings of the National Academy of Sciencees of the United States of America*, **102**, 3645-3650.

175. Vafabakhsh, R. and Ha, T. (2012) Extreme Bendability of DNA Less than 100 Base Pairs Long Revealed by Single-Molecule Cyclization. *Science*, **337**, 1097-1101.

176. Wiggins, P., van der Heijden, T., Moreno-Herrero, F., Spakowitz, A., Phillips, R., Widom, J., Dekker, C. and Nelson, P. (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotechnology*, **1**, 137-141.

177. Mathew-Fenn, R.S., Das, R. and Harbury, P.A.B. (2008) Remeasuring the double helix. *Science*, **322**, 446-449.

178. Vologodskii, A. and Frank-Kamenetskii, M.D. (2013) Strong bending of the DNA double helix. *Nucleic Acids Research*, **41**, 6785-6792.

179. Du, Q., Smith, C., Shiffeldrim, N., Vologodskaia, M. and Vologodskii, A. (2005) Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proceedings of the National Academy of Sciencees of the United States of America*, **102**, 5398-5402.
180. Le, T.T. and Kim, H.D. (2014) Probing the elastic limit of DNA bending. *Nucleic Acids Research*, **42**, 10786-10794.
181. Mastroianni, A.J., Sivak, D.A., Geissler, P.L. and Alivisatos, A.P. (2009) Probing the conformational distributions of subpersistence length DNA. *Biophysical Journal*, **97**, 1408-1417.
182. Mazur, A. (2007) Wormlike Chain Theory and Bending of Short DNA. *Physical Review Letters*, **98**, 218102.
183. Mazur, A.K. and Maaloum, M. (2014) Atomic force microscopy study of DNA flexibility on short length scales: Smooth bending versus kinking. *Nucleic Acids Research*, **42**, 14006-14012.
184. de Gennes, P.G. (2001) Maximum pull out force on DNA hybrids. *Comtes Rendus de l'Academie des Sciences-Series IV*, **2147**, 1505-1508.
185. Hatch, K., Danilowicz, C., Coljee, V. and Prentiss, M. (2008) Demonstration that the shear force required to separate short double-stranded DNA does not increase significantly with sequence length for sequences longer than 25 base pairs. *Physical Review E*, **78**, 011920.
186. Danilowicz, C., Hatch, K., Conover, A., Ducas, T., Gunaratne, R., Coljee, V. and Prentiss, M. (2010) Study of force induced melting of dsDNA as a function of length and conformation. *Journal of Physics: Condensed Matter*, **22**, 414106.
187. Whitley, K.D., Comstock, M.J. and Chemla, Y.R. (in preparation).
188. McIntosh, D. and Saleh, O. (2011) Salt Species-Dependent Electrostatic Effects on ssDNA Elasticity. *Macromolecules*, **44**, 2328-2333.
189. Wen, J.-D., Manosas, M., Li, P.T.X., Smith, S.B., Bustamante, C., Ritort, F. and Tinoco, I. (2007) Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophysical Journal*, **92**, 2996-3009.
190. Horton, N.C. and Finzel, B.C. (1996) The structure of an RNA/DNA hybrid: a substrate of the ribonuclease activity of HIV-1 reverse transcriptase. *Journal of Molecular Biology*, **264**, 521-533.
191. Almog, R. and Shirey, T.L. (1978) A modified orcinol test for the specific determination of RNA. *Analytical biochemistry*, **91**, 130-137.
192. Xiong, Y. and Sundaralingam, M. (1998) Crystal structure and conformation of a DNA-RNA hybrid duplex with a polypurine RNA strand: d(TTCTTBr5CTTC)-r(GAAGAAGAA). *Structure*, **6**, 1493-1501.
193. Shaw, N.N. and Arya, D.P. (2008) Recognition of the unique structure of DNA:RNA hybrids. *Biochimie*, **90**, 1026-1039.
194. Syed, S., Pandey, M., Patel, S.S. and Ha, T. (2014) Single-Molecule Fluorescence Reveals the Unwinding Stepping Mechanism of Replicative Helicase. *Cell Reports*, **6**, 1037-1045.
195. Moffitt, J.R., Chemla, Y.R., Aathavan, K., Grimes, S., Jardine, P.J., Anderson, D.L. and Bustamante, C. (2009) Intersubunit coordination in a homomeric ring ATPase. *Nature*, **457**, 446-450.
196. Dumont, S., Cheng, W., Serebrov, V., Beran, R.K., Tinoco, I., Pyle, A.M. and Bustamante, C. (2006) RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature*, **439**, 105-108.
197. Myong, S., Bruno, M.M., Pyle, A.M. and Ha, T. (2007) Spring-Loaded Mechanism of DNA Unwinding by Hepatitis C Virus NS3 Helicase. *Science*, **317**, 513-516.
198. Cheng, W., Arunajadai, S.G., Moffitt, J.R., Tinoco, I. and Bustamante, C. (2011) Single-Base Pair Unwinding and Asynchronous RNA Release by the Hepatitis C Virus NS3 Helicase. *Science*, **333**, 1746-1749.
199. Ali, J.A. and Lohman, T.M. (1997) Kinetic measurement of the step size of DNA unwinding by Escherichia coli UvrD helicase. *Science*, **275**, 377-380.

200. Fischer, C.J., Maluf, N.K. and Lohman, T.M. (2004) Mechanism of ATP-dependent translocation of E. coli UvrD monomers along single-stranded DNA. *Journal of Molecular Biology*, **344**, 1287-1309.
201. Tomko, E.J., Fischer, C.J., Niedziela-Majka, A. and Lohman, T.M. (2007) A Nonuniform Stepping Mechanism for E. coli UvrD Monomer Translocation along Single-Stranded DNA. *Molecular Cell*, **26**, 335-347.
202. Sun, B., Wei, K.-J., Zhang, B., Zhang, X.-H., Dou, S.-X., Li, M. and Xi, X.G. (2008) Impediment of E. coli UvrD by DNA-destabilizing force reveals a strained-inchworm mechanism of DNA unwinding. *The EMBO Journal*, **27**, 3279-3287.
203. Lee, J.Y. and Yang, W. (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell*, **127**, 1349-1360.
204. Tomko, E.J., Fischer, C.J. and Lohman, T.M. (2012) Single stranded DNA translocation of E. coli UvrD monomer is tightly coupled to ATP hydrolysis. *Journal of Molecular Biology*, **418**, 32-46.
205. Ha, T., Rasnik, I., Cheng, W., Babcock, H.P., Gauss, G.H., Lohman, T.M. and Chu, S. (2002) Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature*, **419**, 638-641.
206. Kerssemakers, J.W.J., Munteanu, E.L., Laan, L., Noetzel, T.L., Janson, M.E. and Dogterom, M. (2006) Assembly dynamics of microtubules at molecular resolution. *Nature*, **442**, 709-712.
207. Svoboda, K., Mitra, P.P. and Block, S.M. (1994) Fluctuation analysis of motor protein movement and single enzyme kinetics. *Proceedings of the National Academy of Sciencees of the United States of America*, **91**, 11782-11786.
208. Schnitzer, M.J. and Block, S.M. (1995) Statistical Kinetics of Processive Enzymes. *Cold Spring Harbor Symposia on Quantitative Biology*, **LX**, 793-802.
209. Maluf, N.K., Fischer, C.J. and Lohman, T.M. (2003) A dimer of Escherichia coli UvrD is the active form of the helicase in vitro. *Journal of Molecular Biology*, **325**, 913-935.
210. Cheng, W., Hsieh, J., Brendza, K.M. and Lohman, T.M. (2001) E. coli Rep oligomers are required to initiate DNA unwinding in vitro. *Journal of Molecular Biology*, **310**, 327-350.
211. Matson, S.W. (1986) Escherichia coli Helicase II (uvrD Gene Product) Translocates Unidirectionally in a 3' to 5' Direction. *Journal of Biological Chemistry*, **261**, 10169-10175.
212. Maluf, N.K., Ali, J.A. and Lohman, T.M. (2003) Kinetic mechanism for formation of the active, dimeric UvrD helicase-DNA complex. *Journal of Biological Chemistry*, **278**, 31930-31940.
213. Lohman, T.M., Tomko, E.J. and Wu, C.G. (2008) Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nature reviews: Molecular cell biology*, **9**, 391-401.
214. Lee, K.S., Balci, H., Jia, H., Lohman, T.M. and Ha, T. (2013) Direct imaging of single UvrD helicase dynamics on long single-stranded DNA. *Nature Communications*, **4**, 1878.
215. Yokota, H., Chujo, Y.A. and Harada, Y. (2013) Single-molecule imaging of the oligomer formation of the nonhexameric Escherichia coli UvrD helicase. *Biophysical Journal*, **104**, 924-933.
216. Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M. and Waksman, G. (1997) Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. *Cell*, **90**, 635-647.
217. Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S. and Wigley, D.B. (1999) Crystal Structures of Complexes of PcrA DNA Helicase with a DNA Substrate Indicate an Inchworm Mechanism. *Cell*, **97**, 75-84.
218. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and Mechanism of Helicases and Nucleic Acid Translocases. *Annual Review of Biochemistry*, **76**, 23-50.
219. Jia, H., Korolev, S., Niedziela-Majka, A., Maluf, N.K., Gauss, G.H., Myong, S., Ha, T., Waksman, G. and Lohman, T.M. (2011) Rotations of the 2B sub-domain of E. coli UvrD helicase/translocase coupled to nucleotide and DNA binding. *Journal of Molecular Biology*, **411**, 633-648.
220. Brendza, K.M., Cheng, W., Fischer, C.J., Chesnik, M.A., Niedziela-Majka, A. and Lohman, T.M. (2005) Autoinhibition of Escherichia coli Rep monomer helicase activity by its 2B subdomain.

Proceedings of the National Academy of Sciencees of the United States of America, **102**, 10076-10081.

221. Cheng, W., Brendza, K.M., Gauss, G.H., Korolev, S., Waksman, G. and Lohman, T.M. (2002) The 2B domain of the Escherichia coli Rep protein is not required for DNA helicase activity. *Proceedings of the National Academy of Sciencees of the United States of America*, **99**, 16006-16011.
222. Meiners, M.J., Tahmaseb, K. and Matson, S.W. (2014) The UvrD303 hyper-helicase exhibits increased processivity. *Journal of Biological Chemistry*, **289**, 17100-17110.
223. Arslan, S., Khafizov, R., Thomas, C.D., Chemla, Y.R. and Ha, T. (2015) Engineering of a superhelicase through conformational control. *Science*, **348**, 344-347.
224. Johnson, D.S., Bai, L., Smith, B.Y., Patel, S.S. and Wang, M.D. (2007) Single-Molecule Studies Reveal Dynamics of DNA Unwinding by the Ring-Shaped T7 Helicase. *Cell*, **129**, 1299-1309.
225. Spies, M. (2013) There and back again: new single-molecule insights in the motion of DNA repair proteins. *Current Opinion in Structural Biology*, **23**, 154-160.
226. Kisker, C., Kuper, J. and Van Houten, B. (2013) Prokaryotic nucleotide excision repair. *Cold Spring Harbor Perspectives in Biology*, **5**, a012591.
227. Guarné, A. (2012) The functions of MutL in mismatch repair: the power of multitasking. *Progress in molecular biology and translational science*, **110**, 41-70.
228. Sijmons, R.H. and Hofstra, R.M.W. (2016) Clinical aspects of hereditary DNA Mismatch repair gene mutations. *DNA Repair*, **38**, 155-162.
229. Jiricny, J. (2013) Postreplicative mismatch repair. *Cold Spring Harbor Perspectives in Biology*, **5**, a012633.
230. López de Saro, F.J., Marinus, M.G., Modrich, P. and O'Donnell, M. (2006) The beta sliding clamp binds to multiple sites within MutL and MutS. *Journal of Biological Chemistry*, **281**, 14340-14349.
231. Pluciennik, A., Burdett, V., Lukianova, O., O'Donnell, M. and Modrich, P. (2009) Involvement of the beta clamp in methyl-directed mismatch repair in vitro. *Journal of Biological Chemistry*, **284**, 32782-32791.
232. Gorman, J., Wang, F., Redding, S., Plys, A.J., Fazio, T., Wind, S., Alani, E. and Greene, E.C. (2012) Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proceedings of the National Academy of Sciencees of the United States of America*, **109**, 3074-3083.
233. Groothuizen, F.S., Winkler, I., Cristovao, M., Fish, A., Winterwerp, H.H.K., Reumer, A., Marx, A.D., Hermans, N., Nicholls, R.A., Murshudov, G.N. *et al.* (2015) MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA. *eLife*, **4**, e06744.
234. Lahue, R.S., Au, K.G. and Modrich, P. (1989) DNA mismatch correction in a defined system. *Science*, **245**, 160-164.
235. Dao, V. and Modrich, P. (1998) Mismatch-, MutS-, MutL- and Helicase II-dependent Unwinding from the Single-strand Break of an Incised Heteroduplex from the Single-strand Break of an Incised Heteroduplex. *Journal of Biological Chemistry*, **273**, 9202-9207.
236. Cheng, F., Hou, J., Chen, Y.-Y., Zhou, Y., Zhang, H.-T., Bi, L.-J. and Zhang, X.-E. (2010) Functional interaction between MutL and 3'-5' exonuclease X in Escherichia coli. *Archives of Biochemistry and Biophysics*, **502**, 39-43.
237. Sancar, A. and Hearst, J.E. (1993) Molecular matchmakers. *Science*, **256**, 1415-1420.
238. Polosina, Y.Y. and Cupples, C.G. (2010) Wot the 'L-Does MutL do? *Mutation Research*, **705**, 228-238.
239. Hall, M.C., Jordan, J.R. and Matson, S.W. (1998) Evidence for a physical interaction between the Escherichia coli methyl-directed mismatch repair proteins MutL and UvrD. *The EMBO Journal*, **17**, 1535-1541.

240. Ban, C., Junop, M. and Yang, W. (1999) Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell*, **97**, 85-97.

241. Spampinato, C. and Modrich, P. (2000) The MutL ATPase Is Required for Mismatch Repair. *Journal of Biological Chemistry*, **275**, 9863-9869.

242. Winkler, I., Marx, A.D., Lariviere, D., Heinze, R.J., Cristovao, M., Reumer, A., Curth, U., Sixma, T.K. and Friedhoff, P. (2011) Chemical trapping of the dynamic MutS-MutL complex formed in DNA mismatch repair in Escherichia coli. *Journal of Biological Chemistry*, **286**, 17326-17337.

243. Elez, M., Radman, M. and Matic, I. (2012) Stoichiometry of MutS and MutL at unrepaired mismatches in vivo suggests a mechanism of repair. *Nucleic Acids Research*, **40**, 3929-3938.

244. Niedziela-Majka, A., Maluf, N.K., Antony, E. and Lohman, T.M. (2011) Self-assembly of Escherichia coli MutL and its complexes with DNA. *Biochemistry*, **50**, 7868-7880.

245. Robertson, A.B., Pattishall, S.R., Gibbons, E.a. and Matson, S.W. (2006) MutL-catalyzed ATP hydrolysis is required at a post-UvrD loading step in methyl-directed mismatch repair. *Journal of Biological Chemistry*, **281**, 19949-19959.

246. Mechanic, L.E., Frankel, B.A. and Matson, S.W. (2000) Escherichia coli MutL loads DNA helicase II onto DNA. *Journal of Biological Chemistry*, **275**, 38337-38346.

247. Acharya, S., Foster, P.L., Brooks, P. and Fishel, R. (2003) The Coordinated Functions of the E . coli MutS and MutL Proteins in Mismatch Repair. *Molecular Cell*, **12**, 233-246.

248. Grilley, M., Griffith, J. and Modrich, P. (1993) Bidirectional Excision in Methyl-directed Mismatch Repair. *Journal of Biological Chemistry*, **268**, 11830-11837.

249. Atkinson, J., Guy, C.P., Cadman, C.J., Moolenaar, G.F., Goosen, N. and McGlynn, P. (2009) Stimulation of UvrD helicase by UvrAB. *Journal of Biological Chemistry*, **284**, 9612-9623.

250. Hermans, N., Laffeber, C., Cristovão, M., Artola-Borán, M., Mardenborough, Y., Ikpa, P., Jaddoe, A., Winterwerp, H.H.K., Wyman, C., Jiricny, J. *et al.* (2016) Dual daughter strand incision is processive and increases the efficiency of DNA mismatch repair. *Nucleic Acids Research*, **44**, 6770-6786.

251. Landry, M.P., McCall, P.M., Qi, Z. and Chemla, Y.R. (2009) Characterization of photoactivated singlet oxygen damage in single-molecule optical trap experiments. *Biophysical Journal*, **97**, 2128-2136.

252. Swoboda, M., Cheng, H., Brugger, D., Haltrich, D., Plumere, N. and Schlierf, M. (2012) Enzymatic Oxygen Scavenging for Photostability without pH Drop in single-molecule experiments. *ACS Nano*, **6**, 6364-6369.

253. Rasnik, I., McKinney, S.A. and Ha, T. (2006) Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat Methods*, **3**, 891-893.

254. Brewer, L.R. and Bianco, P.R. (2008) Laminar flow cells for single-molecule studies of DNA-protein interactions. *Nature Methods*, **5**, 517-525.

255. Ha, T., Rasnik, I., Cheng, W., Babcock, H.P., Gauss, G.H., Lohman, T.M. and Chu, S. (2002) Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature*, **419**, 638-641.

256. Suksombat, S., Khafizov, R., Kozlov, A.G., Lohman, T.M. and Chemla, Y.R. (2015) Structural dynamics of E. coli single-stranded DNA binding protein reveal DNA wrapping and unwrapping pathways. *eLife*, **4**, e08193.

257. Marko, J.F. and Siggia, E.D. (1995) Stretching DNA. *Macromolecules*, **28**, 8759-8770.

258. Prakash, S. and Singh, Y. (2011) Shear unzipping of double-stranded DNA. *Physical Review E*, **84**, 031905.

259. Owczarzy, R., Moreira, B.G., You, Y., Behlke, M.A. and Walder, J.A. (2008) Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry*, **47**, 5336-5353.