Exploring Biomolecular Interactions Through Single-Molecule Force Spectroscopy and Computational Simulation

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493410">http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493410</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Exploring biomolecular interactions through single-molecule force spectroscopy and computer simulation

A dissertation presented
by
Darren Yang
to
The Harvard John A. Paulson School of Engineering and Applied Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Applied Physics

Harvard University
Cambridge, Massachusetts
February 2016
Exploring bio-molecular interactions through single-molecule force spectroscopy and computational simulation

Abstract

Molecular interactions between cellular components such as proteins and nucleic acids govern the fundamental processes of living systems. Technological advancements in the past decade have allowed the characterization of these molecular interactions at the single-molecule level with high temporal and spatial resolution. Simultaneously, progress in computer simulation has enabled theoretical research at the atomistic level, assisting in the interpretation of experimental results. This thesis combines single-molecule force spectroscopy and simulation to explore inter- and intra-molecular interactions. Specifically, we investigate the interaction between RecA and DNA to elucidate the underlying molecular mechanism of the DNA homologous recombination process. We also evaluate the stability of the von Willebrand Factor (vWF) A2 domain to determine the molecular origins of von Willebrand Diseases (vWD). This thesis also describes the development and application of a new single-molecule technique that combines the centrifuge force microscope (CFM) with DNA self-assembled mechanical switches to enable massively parallel repeating force measurements of molecular interactions.
Prior Publication ...................................................................................................................... viii
Acknowledgments ................................................................................................................ viii

Chapter 1 Introduction ............................................................................................................. 1
  1.1 Background ..................................................................................................................... 1
  1.2 Chapter Summary ........................................................................................................... 3

Chapter 2 Complementary strand relocation may play vital roles in RecA-based homology
recognition ............................................................................................................................. 8
  2.1 Synopsis ......................................................................................................................... 8
  2.2 Introduction ................................................................................................................... 9
  2.3 Materials and Methods ............................................................................................... 17
  2.4 Result .......................................................................................................................... 19
  2.5 Discussion .................................................................................................................... 28

Chapter 3 Integrating multi-scale data on homologous recombination into a new
recognition mechanism based on simulations of the RecA-ssDNA/dsDNA structure 38
  3.1 Synopsis ......................................................................................................................... 38
  3.2 Introduction ................................................................................................................... 39
  3.3 Methods ....................................................................................................................... 49
  3.4 Result .......................................................................................................................... 54
  3.5 Discussion .................................................................................................................... 66

Chapter 4 Structural Investigation of the von Willebrand Disease A2 R1597W mutation by
Molecular Dynamics Simulation ............................................................................................ 77
  4.1 Introduction ................................................................................................................... 77
  4.2 Method ........................................................................................................................ 82
  4.3 Result .......................................................................................................................... 85
  4.4 Discussion .................................................................................................................... 97
Chapter 5 Multiplexed single-molecule force spectroscopy using a centrifuge        103
  5.1 Synapsis.................................................................................................................. 103
  5.2 Introduction .............................................................................................................. 104
  5.3 Result and Discussion .............................................................................................. 105
  5.4 Material and Methods.............................................................................................. 114
Appendix 1...................................................................................................................... 125
   A1.1 Supplementary Notes for Chapter 2 ................................................................. 125
Appendix 2...................................................................................................................... 141
   A2.1 Supplementary Notes for Chapter 3 ................................................................. 141
Appendix 3...................................................................................................................... 158
   A3.1 Supplementary Notes for Chapter 4 ................................................................. 158
Appendix 4...................................................................................................................... 162
   A4.1 Supplementary Notes for Chapter 5 ................................................................. 162
REFERENCES .................................................................................................................. 175
Prior Publication

This thesis is largely based on the following publication:

Chapter 2:

Chapter 3:

Chapter 4:
Darren Yang, Nathan E. Hudson, Wesley P. Wong. “Structural Investigation of the von Willebrand Disease A2 R1597W mutation by Molecular Dynamics Simulation.” In Preparation

Chapter 5:

The author has also contributed to:


Julea Vlassakis, Efraim Feinstein, Darren Yang, Antoine Tilloy, Dominic Weiller, Julian Kates-Harbeck, Vincent Coljee, and Mara Prentiss. "Tension on dsDNA bound to

¹ Contributed equally
Acknowledgments

First and foremost, I would like to express and my deepest gratitude to my research advisors Professor Mara Prentiss and Wesley Wong for their encouragement, support, and guidance throughout my Ph.D. I would like to thank Professor Efthimios Kaxiras and Neel Joshi for their valuable time and assistance as a thesis committee member. In addition, I would also like to thank Professor Daniel Needleman for his supervision and guidance.

I would like to thank all the current and previous members of Prentiss and Wong lab who have provided scientific advice, technical help, and support. I have had the great pleasure of working with Claudia Danilowicz, Efraim Feinstein and Alex Peacock-Villada from the Prentiss lab; as well as, Ken Halvorsen, Ted Feldman, Yan Jiang, Mounir Koussa, Andrew Ward, Serkan Cabi, and Clinton Hansen from the Wong lab. I also would like to thank Chantal Prévost, Hongxia Fu, and Nathan Hudson for incredibly fruitful collaborations. Outside of the laboratory, I thank my longtime friends Steven and Jennifer Shimizu for participating in many adventures from San Diego to Boston.

Last but not least, I would like to express my love and gratitude to my family. I would not be where I am today without them as an unwavering source of love, support, and encouragement. I will forever be indebted for my mother Shujen Chang who dedicated her life for my sister and me. I would also like to thank my late uncle Albert Donabedian who was always there for me and my family. Finally, gratitude to my fiancée Weiru Kuo for her love and support.
Chapter 1

Introduction

1.1 Background

Recent developments in the single-molecule force spectroscopy (SMFS) techniques have allowed the direct investigation of the role of force in biomolecular processes by capturing rare and short-lived events that typically get averaged out in traditional ensemble measurements [1-4]. Specifically, instruments such as optical tweezers [5, 6], atomic force microscopes (AFM) [7-9], magnetic tweezers [10-12], and centrifuge force microscopes [13] provide the ability to both apply mechanical forces on biological molecules and measure the displacement at high temporal and spatial resolution. For example, one of the earliest single-molecule force measurements is the force-extension relationship of nucleic acid [14-17], which seeded the investigation of molecular motors and enzymes that regulate DNA and RNA [17-21]. Single-molecule manipulation techniques also evolved to study the strength of molecular interactions, which provide the force-dependent lifetimes and
energetic landscapes [22, 23] of interactions such as platelet adhesion in hemostasis [24] and antigen-antibody binding [25]. Beside the development of SMFS, molecular dynamics simulation have also become an important tool for elucidating the relationships between biomolecular structures, dynamics and function [26, 27]. In molecular dynamics simulations, interactions between atoms are described by parametrized molecular mechanics force-fields [28-31], and trajectories of the atoms are obtained by integrating Newton’s equation of motion[32]. The time evolution of proteins conformations observed from the simulations that provides information on processes such as enzyme activity [33], molecular recognition and association [34], and protein folding [35].

This thesis combines single-molecule force spectroscopy and molecular dynamics simulations to understand the molecular mechanism of the DNA homologous recombination process promoted by RecA and regulation of the vascular multimeric protein von Willebrand factor (vWF). Chapter 2 and 3 describe the research results, under the guidance of Professor Mara Prentiss, on the DNA-RecA system. Chapter 4 presents work with Professor Wesley Wong on the vWF system. Lastly, Chapter 5 summarizes the development of a high-throughput and cost-effective single-molecule force spectroscopy instrument that combines the centrifuge force microscope (CFM) with DNA self-assembled mechanical switches to enable massively-parallel repeating force measurements of molecular interactions.
1.2 Chapter Summary

Chapter 2: Complementary strand relocation may play vital roles in RecA-based homology recognition

RecA family proteins play an important role in homology recognition and strand exchange, but many features of the structure and function of these proteins are not yet known. The crystal structure of the RecA/single-stranded DNA complexes in the initial homology searching state, and the structure of the protein/heteroduplex DNA complex were determined [36]. However, no structural information on the third DNA strand that participated in the homology recognition/strand exchange process is available at the molecular level. Though the crystal structure was a tremendous breakthrough, it left substantial unanswered questions about the relationship between structure and function. For example, the crystal structure demonstrated that the initial homology searching state and the final post-strand exchange state are virtually identical; however, strand exchange proceeds spontaneously in the absence of hydrolysis, so it must be free energetically favorable despite the almost identical nature of the initial and final structures. Furthermore, it was unclear how sequence stringency is enforced in the transition from the searching state to the final post-strand exchange product.

The results presented in Chapter 2 suggests that the homology recognition/strand exchange process includes an intermediate structure with the complementary strand in a different position than the one it occupies in the final post-strand exchange product. Homology recognition and strand exchange occur with the complementary strand in this intermediate position. After strand exchange, the complementary strand backbone
relocates to its final post-strand exchange position, and the relocation of the complementary strand limits the strand exchange rate.

The experimental and computational studies in Chapter 2 provides a new paradigm for RecA based homology recognition where homology stringency is not determined only by the free energy difference due to mismatched Watson-Crick pairing. Rather, homology stringency is enforced by sequence dependent structural changes that transfer bound DNA from unstably bound states to stably bound states. Homology stringency requires that the transfer rate from the intermediate state to the final state must be slow enough to allow short regions of accidental homology which have strand exchanged to unbind before the strand exchange product is transferred to the very stable final post-strand exchange state.

Chapter 3: Integrating multi-scale data on homologous recombination into a new recognition mechanism based on simulations of the RecA-ssDNA/dsDNA structure

The mechanism of sequence recognition and strand exchange of genomic DNA, characteristic of homologous recombination (HR), has remained an enigma for the last three decades in spite of intensive work from various research areas. Notably, the question of how, from a structural point of view, genomic sequences with over $10^5$ base pairs can be searched in the order of minutes and with high stringency is of fundamental importance. Very recently it was established, based on experimental as well as theoretical studies, that the process follows two very different stages: a first, very rapid stage where most of the mismatched sequences are discriminated and a slower stage. During the second slower stage, strand exchange progresses at a steady rate and where accidental mismatches that have passed the first stage are eliminated.
The research in Chapter 3 explores a molecular mechanism which fulfills the speed and stringency requirement associated to the first stage of the recognition/strand exchange process. The work is divided in two main parts: first, a construction part that proposes an intermediate structure of association between DNA and the active filament of recombination, and the second part explores the stability and time evolution of these intermediates using molecular dynamics simulations. From the modeling and simulations, we observed spontaneous base pairing exchange and known structural features of the reaction product that had not been incorporated in the model. Finally, the simulations revealed unanticipated information that provides plausible clues for the mechanism.

Lastly, the study in Chapter 3 is an example of an interdisciplinary approach integrating detailed structural and functional information obtained via modeling approaches with physical models of homology search. The resulting overall model proposed structural details of the homology search mechanism which are supported by a large body of experimental and theoretical evidence.

**Chapter 4: Dynamically multiplexed force spectroscopy using a centrifuge**

In Chapter 4, we describe new technique that combines the centrifuge force microscope (CFM) with DNA self-assembled mechanical switches (DNA nanoswitch) for new experimental platform with the potential to both accelerate and democratize single-molecule force spectroscopy. This new design of CFM enables high-throughput single-molecule force measurements to be made using the benchtop centrifuge which removes the barriers of cost, throughput, technical difficulty, and strict infrastructure requirements. The
DNA nanoswitches increase the reliability of single-molecule measurements by enabling automated discrimination between good rupture data and anchor rupture events, non-specific interactions and multiple bonds. As demonstration, massively parallel studies of both bond rupture and molecular deformation under force were carried out. By taking advantage of the portability and temperature control ability of the bench-top centrifuge, temperature-dependent measurements were also made.

More importantly, by combining massively parallel spatial multiplexing enabled by the CFM with temporal multiplexing enabled by DNA nanoswitches, we can now repeatedly interrogate an ensemble of molecules, while observing their behavior at the single molecule level. As shown in Chapter 4, we collected almost 2000 statistics from a single sample (more than the total number of statistics in almost any published single-molecule force spectroscopy paper). For the first time, we created an ensemble of single-molecule force rupture histograms, showing that one can characterize the unique properties of each molecule in a population through repeated interrogation. We performed per-molecule averaging to reduce noise without losing the unique properties of each molecule, which provides the ability to discriminate different populations of molecules even below the intrinsic thermal noise limit—in analogy with super-resolution imaging through the use of per-fluorophore localization to resolve details below the Abbe diffraction limit.

Chapter 5: Structural Investigation of the von Willebrand Disease A2 R1597W mutation by Molecular Dynamics Simulation

von Willebrand factor (vWF) is a plasma glycoprotein that serves as a ligand for platelet adhesion and aggregation in hemostasis and circulates in the form of concatenated
multimers. Regulation of vWF synthesis and catabolism is critical for maintaining a healthy hemostatic potential. The vWF multimer size distribution is carefully controlled by metalloprotease ADAMTS13, which cleaves the vWF A2 domain between residues Y1605 and M1606. A2 domain cleavage is stimulated by hydrodynamic force in blood flow, which unfolds the A2 domain and exposes the proteolysis site. Von Willebrand Disease (vWD) type 2A is characterized by a lack of large multimers in the vasculature, and is often linked to mutations in the A2 domain. Here, using one of the most common vWD type 2A mutations, R1597W, we characterize the structural mechanisms of this disease using equilibrium and steered molecular dynamics simulations. We also evaluate the effect of calcium in the A2 calcium-binding loop. Simulations show that the R1597W mutation disrupts the native wild-type interactions that effect the binding of the associated calcium ion. Moreover, the measured tensile force from the force-induce unfolding simulation confirmed that R1597W mutation destabilized the A2 domain, hence leading to increase of the ADAMTS13 cleavage rate of vWF. The detailed structural model at the molecular level we present here is informative for future therapeutic development.

Appendix 1: Supplemental notes and figures for Chapter 2.

Appendix 2: Supplemental notes and figures for Chapter 3.

Appendix 3: Supplemental notes and figures for Chapter 4.

Appendix 4: Supplemental notes and figures for Chapter 5.
Chapter 2

Complementary strand relocation may play vital roles in RecA-based homology recognition

2.1 Synopsis

RecA-family proteins mediate homologous recombination and recombinational DNA repair through homology search and strand exchange. Initially, the protein forms a filament with the incoming single-stranded DNA (ssDNA) bound in site I. The RecA–ssDNA filament then binds double-stranded DNA (dsDNA) in site II. Non-homologous dsDNA rapidly unbinds, whereas homologous dsDNA undergoes strand exchange yielding heteroduplex dsDNA in site I and the leftover outgoing strand in site II. We show that applying force to the ends of the complementary strand significantly retards strand
exchange, whereas applying the same force to the outgoing strand does not. We also show that crystallographically determined binding site locations require an intermediate structure in addition to the initial and final structures. Furthermore, we demonstrate that the characteristic dsDNA extension rates due to strand exchange and free RecA binding are the same, suggesting that relocation of the complementary strand from its position in the intermediate structure to its position in the final structure limits both rates. Finally, we propose that homology recognition is governed by transitions to and from the intermediate structure, where the transitions depend on differential extension in the dsDNA. This differential extension drives strand exchange forward for homologs and increases the free energy penalty for strand exchange of non-homologs.

2.2 Introduction

Homologous recombination (HR) plays an important role in meiosis and DNA damage repair [37-39]. During HR, an incoming single-stranded DNA (ssDNA) molecule bound to recombinase protein molecules, such as RecA, searches for a sequence matched double-stranded DNA (dsDNA) molecule [40, 41]. Once a sequence-matched dsDNA is found, the incoming ssDNA displaces one of the strands in the dsDNA (outgoing strand) and Watson–Crick pairs with the other strand (complementary strand), yielding heteroduplex dsDNA in site I [42, 43] and an unpaired outgoing strand. The crystal structure of the searching state with ssDNA in site I is known [44]. The final structure of the heteroduplex dsDNA in site I and some residues associated with the final position of the outgoing strand are also known [44]. Figure 2.1 illustrates the structure of the heteroduplex in the final
post-strand exchange state and the binding sites associated with the final position of the outgoing strand [45]. The structure of the incoming ssDNA in the searching state is almost identical to the structure of the dsDNA in the final post-strand exchange state. In both cases, the DNA consists of base triplets with nearly B-form spacing, where the base triplets are separated by backbone extensions that are much larger than the equilibrium extension for B-form dsDNA [44]. This non-uniform spacing results in an average extension along the filament axis of \(\sim 1.5\times\) the B-form extension.

Figure 2.1: Proposed schematics of the strand exchange process. (A) Representation of the side view of the X-ray structure of the dsDNA heteroduplex in the final post-strand exchange state with incoming and complementary strands indicated by the cyan and red stick renderings. Orange, cyan and red arrows indicate the positions of the corresponding phosphates on the outgoing, incoming, and complementary strands in the final post-strand exchange structure. The VMD [45] renderings of RecA crystal structure 3CMX [44] show site II residues Arg226 (pink), Arg243 (yellow) and Lys245 (magenta) with charged nitrogen atoms (blue). The outgoing strand position was calculated by minimizing the energy of the interaction with those residues. The grey arrow points at the red plus sign.
indicating the proposed position of the corresponding complementary strand phosphate in the proposed intermediate structure. (B) Bottom view of the same structure. Circles correspond to the radii occupied by the phosphates. (C) Information as (B), but only one set of the corresponding phosphates is shown. Final state phosphate positions for the incoming, complementary, and outgoing strands are shown with filled colored circles. Paired bases shown as lines. The complementary strand position in the proposed intermediate structure is shown by the grey circle. (D) Same as (B) with circles and base pairs added to show radial positions of the corresponding phosphate groups and repositioning of the complementary strand.

An outstanding question has been what drives strand exchange forward given that the Watson–Crick pairing is the same before and after strand exchange, yet the process must be free energetically favorable since it occurs in the absence of hydrolysis. Another outstanding question is whether strand exchange is simply a two-step process in which the Watson–Crick pairing of the complementary strand bases is transferred from the outgoing strand to the incoming strand by base flipping, or whether additional steps occur. Many experiments have probed for additional intermediate states [46-50], but the existence, number, nature and roles of intermediate states remain controversial. The existence of an additional step involving a structure other than the known initial and final states might allow strand exchange to result in a significant change in the DNA–protein interaction that could produce the free energy reduction required to drive strand exchange forward.
The X-ray structure of the filament shows that the incoming strand backbone is located \( \sim 1 \) nm from the center of the helical RecA filament, whereas the binding sites associated with the outgoing strand are \( \sim 2 \) nm from the center of the RecA filament, as shown in Figure 2.1. As a result of this difference in separation from the center of the RecA filament, the total linear extension of the outgoing strand backbone is significantly longer than the total linear extension of the incoming strand. If the base pairs in the outgoing strand are organized in B-form triplets separated by rises, and strand exchange is accomplished by base flipping, then the bases in the outgoing strand must also be arranged in nearly B-form triplets separated by large rises; however, in the rises the linear extension of the outgoing strand backbone must exceed the linear extension of the incoming strand backbone since the outgoing strand is located farther from the center of the RecA protein helix, as shown in Figure 2.1. A similar effect appears in a spiral staircase with a hollow center: the railing on the outside of the staircase has a longer linear extension than the railing on the inside, even though the steps connecting them are perpendicular to the axis of the helix. In metal spiral staircases, the longer outer railing simply includes more metal than the inner railing. In contrast, in DNA molecules all three backbones contain the same number of phosphates and all of the molecules are extended beyond their B-form lengths; therefore, larger extensions can result in larger mechanical stress because the bonds in the more extended backbones are more deformed. For backbones with strong direct interactions with the protein, the protein–DNA interactions may be so free energetically favorable that they more than compensate for stress in the rises; however, if direct DNA–protein interactions are weak, mechanical stress due to large rises may be free energetically unfavorable.
In the final post-strand exchange state, the complementary strand is attached to the RecA–ssDNA filament dominantly via its Watson–Crick pairing with the incoming strand [44]; therefore, large rises between bases in the incoming strand would place great mechanical stress on the complementary strand bases unless the protein structure provides additional mechanical support for those rises. If the complementary strand is paired with the even more highly extended outgoing strand, then the stress on the complementary strand will be even larger than the stress on the complementary strand paired with the incoming strand. Thus, strand exchange may reduce the stress on the complementary strand bases and the complementary strand backbone, which would reduce the mechanical energy of the system. In sum, strand exchange might be free energetically favorable because it reduces the tension on the complementary strand by transferring the base pair interaction from the very highly extended outgoing strand backbone to the somewhat less extended incoming strand backbone.

Seminal strand exchange models proposed that the transition between the homology searching state and the final strand exchange state could be achieved by flipping the bases in the complementary strand while the positions of all three DNA backbones remained approximately fixed [40]; however, the X-ray structures show that in the final post-strand exchange state, the separation between the outgoing and complementary strands is too large to allow for Watson–Crick pairing (Figure 2.1A–C and Figure 2.2A). Earlier FRET results had suggested that the complementary strand backbones reposition after the complementary strand base pairing is transferred from the outgoing strand to the incoming strand [46]. The suggestion that the incoming and outgoing strands do not reposition during strand exchange is sensible since the incoming and outgoing strands are
attached strongly through direct contacts between the backbone and the protein [44] (Figure S1.1). Thus, it is reasonable to assume that the complete strand exchange process involves the relocation of the complementary strand backbone from a position where its bases can flip between pairing with the outgoing strand and pairing with the incoming strand to the final post-strand exchange position, known from the X-ray structure. Since changing the position of the complementary strand backbone from the proposed initial position to its known final position requires such a large relocation of the backbone, it is plausible that the time required to relocate the backbone might represent the rate-limiting step in strand exchange. In the final post-strand exchange structure portions of the L1 and L2 loops are located in the rises [44]. Thus, in the final post-strand exchange state, the protein may provide mechanical support for the rises as well as additional free energetically favorable interactions [44], which make the final repositioning of the complementary strand free energetically favorable.
**Figure 2.2:** Illustrations of the RecA–DNA structures. (A–C) show the approximate positions of the outgoing (orange), complementary (red) and incoming (cyan) backbones, where the base pairs are located within the shaded regions. (A), (B) and (C) show the searching state, the intermediate post-strand exchange state, and the final post-strand exchange state respectively. (D–F) have the same backbone color code as (A). They show a schematic representation of the stress on the base pairs, where purple indicates high stress, lavender indicates moderate stress, and black indicates low stress. (D), (E) and (F) show the searching state, the intermediate post-strand exchange state, and the final post-strand exchange state, respectively.

Earlier experiments had suggested that the structure of dsDNA in the final post-strand exchange state is similar to the structure of dsDNA in a RecA filament created by the dsDNA binding to RecA that is free in solution [51]. If the incoming strand backbone maintains its position throughout strand exchange, then the X-ray structure suggests that the binding of free RecA to dsDNA would occur in a conformation similar to the conformation of the dsDNA in the strand exchange process immediately after the pairing interaction of the complementary strand bases has shifted from the outgoing strand to the incoming strand. If this were true, the binding of free RecA to dsDNA would involve the same complementary strand relocation that occurs during strand exchange after the Watson–Crick pairing of the complementary strand has been transferred from the outgoing strand to the incoming strand. Thus, if free RecA binding is not diffusion limited, then the extension rates due to free RecA binding should be the same as the extension rate due to
strand exchange because both would be limited by the time required to relocate the complementary strand backbone.

In this work we investigate the strand exchange process in which the base pairing of the complementary strand is transferred from the outgoing strand to the incoming strand by considering the interaction between RecA–ssDNA filaments and homologous dsDNA held under tension by an external force. Since the dsDNA is homologous to the ssDNA, the Watson–Crick pairing of the complementary strand can be transferred from the outgoing strand to the incoming strand. Thus, we study the possibility that strand exchange is free energetically favorable because it reduces the mechanical stress on the complementary strand bases due to the transfer of Watson–Crick pairing from the highly extended outgoing strand to the less extended incoming strand. Such a reduction would occur if the complementary strand were bound to the RecA filament dominantly through its Watson–Crick pairing since the bases that connect the complementary strand would experience great mechanical stress from either of its pairing partners, which have extensions that greatly exceed the B-form length.

We probe for effects of tension associated with the differential extension of the complementary strand and its Watson–Crick pairing partners by applying force to the ends of homologous dsDNA during strand exchange. If a reduction in the differential extension between the complementary strand and its pairing partners drives strand exchange forward, then using an external force to reduce that differential extension should slow strand exchange. Thus, pulling on the 3′-ends of the complementary strand would slow strand exchange because it decreases the differential extension between the complementary strand
and its pairing partners, whereas pulling on the 3′5′-ends of the outgoing strand should not affect the strand exchange rate because it does not reduce the differential extension. Finally, we performed experiments where we investigated both the extension rates during strand exchange and during free RecA binding to dsDNA as a function of the force applied to the ends of the dsDNA to determine whether the relocation of the complementary strand could represent the rate-limiting step in strand exchange.

2.3 Materials and Methods

Sample preparations

Double-stranded λ DNA molecules (New England Biolabs) were modified by hybridizing and ligating biotinylated oligonucleotides yielding 3′3′-labeled dsDNA. Ligation steps were done in the presence of a thermostable DNA Ligase (Ampligase, Epicentre, Madison, WI, USA). The oligonucleotides at both ends included a ssDNA tail [(dT)–(biotin-dT)x] to allow free rotation of the bonds. After each modification step was completed, the dsDNA sample was washed three times using Amicon YM-100 filters (Millipore, USA) and 70 mM Tris buffer pH 7.6. The final concentration was determined by the absorbance at 260 nm.

The RecA–ssDNA filaments were prepared by mixing 3 µM ssDNA (5 kb) with 1 µM RecA, 1 mM ATPγS and 0.2 µM SSB in RecA buffer (70 mM Tris–HCl, 10 mM MgCl2, and 5 mM dithiothreitol, pH 7.6). The ssDNA was previously prepared by amplifying a 5-kb fragment using λ-phage as a template where one of the primers was 5′-
phosphorylated, and purified using a Macherey–Nagel kit. The dsDNA PCR fragment was subsequently incubated with λ-exonuclease enzyme (NEB) at 37°C for 30 min, and the resulting ssDNA was further purified using a Qiagen kit.

**Experimental conditions and data analysis**

For experiments where free RecA binds to dsDNA, an aliquot of dsDNA in RecA buffer, 1 µM RecA (New England Biolabs), 1 mM ATPγS and the beads were placed in a square micro-cell. Initially, a force of ∼70 pN was briefly applied to achieve overstretching. After overstretching, the force was quickly reduced to between 15 and 36 pN. Strand exchange experiments were performed with and without initial overstretching, and the results in both cases were the same. In strand exchange experiments, free RecA was replaced by a 5 µl aliquot of the filament preparation in a 50 µl final volume.

The experiments that presented different concentrations of Mg(II) were done by initially incubating all the reagents in 10 mM Mg(II) and finally diluting to each Mg(II) concentration to allow for the initial binding of the filaments in the homolog region and subsequently following strand exchange at several Mg(II) concentrations.

The temperature of the square capillary was varied from 22 to 37 °C using a thermoelectric cooler placed on top of the aluminum mount holding the square capillary. A temperature sensor close to the capillary channel provides feedback for the stabilization loop controlling the thermoelectric cooler.
Experiments were carried out using our magnetic tweezers set up [52]. Data analysis was performed using scripts custom-written in Matlab (www.mathworks.com) as described in our previous work [53].

2.4 Result

We perform single-molecule experiments to measure the changes in dsDNA extension as a function of time in the presence of either free RecA or RecA–ssDNA filaments (Figure S1.2). The total extension of each dsDNA molecule is followed at various constant forces, with values from 15 to 36 pN. Given that the binding of dsDNA to RecA extends the average length of the dsDNA by 0.51 nm per bound base pair triplet, measurements of changes in the dsDNA extension can be used to calculate the binding rate. A detailed description of the algorithms used to determine the binding rates was presented elsewhere [53].

For strand exchange experiments, we mix and briefly incubate the RecA–ssDNA filaments with dsDNA at zero force and subsequently apply a constant force to monitor the change in dsDNA extension due to strand exchange. Initial force-induced nucleation is not required for experiments using RecA–ssDNA filaments since these filaments spontaneously pair with homologous dsDNA. In contrast, for the experiments where free RecA binds to dsDNA, nucleation is required before RecA can begin to polymerize along a dsDNA molecule. Since dsDNA in the post-strand exchange complex is bound in site I of RecA, properties of this complex are often probed via measurements of direct binding.
of free RecA to dsDNA, which is believed to result in dsDNA bound to site I [51]. Thus, in free RecA-binding experiments nucleation is achieved by briefly applying a force of \( \sim 70 \) pN to each dsDNA molecule before starting the extension measurement [53-56].

Histograms of the slopes of the extension curves for primary site binding experiments usually show clear periodic peaks corresponding to integer multiples of a characteristic elongation rate [53], whereas histograms for the strand exchange experiment usually show one dominant peak that may sometimes be accompanied by a few additional peaks that are integer multiples of the dominant peak (Figure 2.3A). In the free RecA experiments the integer multiples may correspond to multiple nucleation sites [53].
**Figure 2.3:** Effect of force applied to different ends of dsDNA λ-phage during strand exchange and free RecA binding. (A) Elongation rate histogram for strand exchange in ATPγS at 30°C; peak 1 shows the characteristic rate of 211.8 bp/min (0.60 nm/s) whereas peak 0 corresponds to molecules that were followed but showed no change in extension. (B) Extension rates as a function of temperature for different 3′5′- and 3′3′-pulling techniques. Arrhenius plot of single-molecule extension rates as a function of temperature for free RecA binding and strand exchange in ATPγS and 10 mM MgCl₂ with 3′5′ and 3′3′ pulling techniques. Strand exchange rates in bulk experiments and no external force [57] (grey line); free RecA binding while pulling 3′5′ from the complementary strand (red plus...
signs); free RecA binding while pulling 3′5′ from the outgoing strand (blue × symbols). Strand exchange rate while pulling 3′5′ from the complementary strand (red squares) and while pulling the other 3′5′ strand with the alternative filament complementary to the pulled strand (red diamond). Strand exchange rate while pulling 3′5′ from the outgoing strand (blue circles); strand exchange rate while pulling 3′3′ from the outgoing strand nearest the filament (grey triangle) and strand exchange rate while pulling 3′3′ from the complementary strand nearest the filament (grey upside-down triangle). Error bars: confidence intervals. (C) Schematic representation of the effect of force applied to different ends of the dsDNA constructs during strand exchange experiments. (i, ii and iii) dsDNA pulled from 3′5′-ends with stressed and unstressed base pairs shown in magenta and blue, respectively. (iv and v) dsDNA pulled from the 3′3′-ends with stressed and unstressed base pairs shown in magenta and blue, respectively. The grey and black ssDNA correspond to filaments complementary to opposite strands of the dsDNA. In the representation of strand exchange in the first row, the RecA molecules were omitted for better clarity. The lavender band indicates the GC-rich end in λ-phage dsDNA. The ellipses in the second row indicate RecA monomers with Site I and Site II shown in grey and purple, respectively. The outgoing, complementary, and incoming strands are shown in orange, red and cyan, respectively. The effect of pulling the 3′5′-ends of the complementary strand, 3′5′-ends of the outgoing strand, and 3′3′-ends of the dsDNA is represented. The symbols under each figure correspond to the symbols in (B).

In the experiments that studied whether or not strand exchange was affected by applying a force to single-stranded tails at the ends of ~50-kb λ-phage dsDNA molecules
during strand exchange (Figure 2.3B), we used three different dsDNA constructs and two different ssDNA filaments (Figure 2.3C). The sequence of λ-phage is more GC-rich at one end than the other, so the two strands of λ-phage dsDNA can be unambiguously identified by specifying the direction with respect to the GC-rich end (indicated Figure 2.3C by the lavender band). In Figure 2.3C, the black dsDNA backbone corresponds to the 5′3′-strand starting from the GC-rich end and the grey dsDNA backbone indicates the 3′5′-strand starting from the GC-rich end. The two ssDNA filaments were prepared from a 5-kb fragment amplified from the GC-rich end of λ-phage. After treatment with λ-exonuclease, ssDNA is obtained. Depending on the strand chosen to interact with the exonuclease, the resulting ssDNA is complementary to either the 5′3′-strand or the 3′5′'(Figure S1.3). The filament complementary to the 5′3′-strand is shown in grey, and the filament complementary to the 3′5′-strand is shown in black. These schematics illustrate that pulling on the 3′5′-ends of the complementary strand reduces the differential tension between the complementary and outgoing strands. In contrast, pulling on the 3′3′-ends or pulling on the 3′5′-ends of the outgoing strand does not reduce the differential tension between the complementary and outgoing strands.

Figure 2.3B shows the rate at which dsDNA is extended by strand exchange. The solid line indicates the strand exchange rate observed in previous work when no force was applied to the ends of the dsDNA [57]. The symbols in the figure correspond to results where the dsDNA is pulled along a single backbone, as well as results where dsDNA was pulled from the 3′3′-ends. The blue circles show the results when the force is applied to the outgoing strand, as determined by the filaments shown in grey, whereas the red squares show the results when the force is applied to the complementary strand, as determined by
the filaments shown in grey. The red diamond corresponds to the case when the strand pulled is complementary to the RecA–ssDNA black filament (Figure 2.3C iii). Thus, though Figure 2.3C i and iii both correspond to pulling on the strand that is complementary to the ssDNA in the filament, they represent opposite strands of the dsDNA molecule. Similarly, though Figure 2.3C ii and iii correspond to pulling on the same physical strand of the dsDNA, they correspond, respectively, to the outgoing and complementary strands for the RecA–ssDNA filaments used in each of those experiments. The figure also shows that when pulling on the 3′3′-ends (grey triangles) or pulling on the 3′5′-ends of the outgoing strand, the observed strand exchange rate is the same as the rate that is observed in the absence of force. In contrast, pulling on the 3′5′-ends of the complementary strand significantly reduces the observed strand exchange rate. This result is true whether the filament is complementary to the sequence at the 3′-end (black) or the 5′-end (grey) of the dsDNA.

We conducted these experiments to allow us to discriminate between the following possibilities: (1) the results depend on whether the pulled strand was complementary to the RecA–ssDNA filament, (2) the results depend on the physical strand being pulled, (3) the results depend on the sequence in the filament and (4) the results depend on whether the pulled dsDNA end nearest the filament was 3′ or 5′. If (1) were true, then cases represented in Figure 2.3C i and iii would be the same, whereas Figure 2.3C ii, iv and v would be different. If (2) were true, then Figure 2.3C ii and iii would be the same, whereas Figure 2.3C i would be different. If (3) were true, then Figure 2.3C i and ii would be the same, whereas Figure 2.3C iii would be different. Finally if (4) were true, then Figure 2.3C ii, iii, iv and v would be the same and Figure 2.3C i would be different.
In sum, the results show that the strand exchange rate is reduced if and only if a force between 15 and 36 pN is applied preferentially to the dsDNA that is complementary to the ssDNA in the searching filament. In all other cases, the strand exchange rates are independent of force for forces <36 pN. At sufficiently low forces, the strand exchange rate for 3′5′ pulling on the complementary strand must return to the rate observed in the absence of force, but forces <5 pN were inaccessible to us in our experiment because the low dsDNA tension increased the uncertainty in the bead position due to Brownian motion making the extension rate determinations too difficult.

Furthermore, we compared the strand exchange rates and the rates for free RecA binding. The red plus signs and blue × symbols in Figure 2.3B show the rate at which dsDNA is extended by the binding of free RecA. The force-independent extension rates for free RecA binding are independent of which strand is being pulled and similar to the force independent extension rates obtained during strand exchange when the 3′3′-ends were pulled, shown as grey triangles. The extension rates observed when pulling on the 3′5′-ends of the outgoing strand are also similar.

In addition, Figure 2.4A shows an Arrhenius log plot of the characteristic elongation rates as a function of 1000/Temperature for dsDNA during strand exchange and during free RecA binding. The red circles show new results for free RecA binding to dsDNA under tension in a buffer containing ATPγS. New results for strand exchange obtained while force was applied to the 3′3′-ends of dsDNA molecules in the presence of either ATPγS or a mixture of ATPγS and ADP are shown by the black outlined red triangles and squares, respectively. For comparison, previous strand exchange rates obtained earlier
by other groups are also shown in the figure. The line corresponds to the best fit results from earlier bulk experiments in the absence of force [57]. The rest of the points correspond to single-molecule experiments [58-60]. The grey diamonds show zero force results in a buffer containing ATP or ATPγS [59]; the grey squares show results in a buffer containing either ATP or ATPγS when ~20–1000 bp were bound to the filament and a 0.5 pN force was applied to both termini at both ends of the dsDNA [60]. Finally, the grey triangle corresponds to single-molecule experiments measured in a buffer containing ATP, where the strand exchange rate was determined by measuring the decrease in extension due to the unbinding of heteroduplex dsDNA at the lagging end of the strand exchange window.

**Figure 2.4:** Comparison of the measured dsDNA elongation rates due to strand exchange and the binding of free RecA to dsDNA. (A) Arrhenius plot: strand exchange rates in bulk experiments and no external force [57] (black line); single-molecule strand exchange rates in ATP or ATPγS: grey diamonds [59] and grey squares [60]; dsDNA release rate at the back of the strand exchange window in ATP: grey triangle [58]; new results for strand
exchange in ATPγS (red triangles), ATPγS–ADP (outlined-red square), and free RecA binding (red circles). Error bars: confidence intervals. (B) Extension rates as a function of MgCl₂ concentration. Log of single-molecule dsDNA extension rates (nm/s) as a function of MgCl₂ concentration (mM) for both free RecA binding (navy diamonds) and strand exchange (green triangles) in ATPγS at 22°C. MgCl₂ concentrations varied from 0.1 to 10 mM. The dependence of the rates of free RecA binding on MgCl₂ concentration is fit by a logarithmic trend-line (black line).

Figure 2.4A shows that the strand exchange rates are insensitive to hydrolysis, applied force and the amount of dsDNA bound to the filament. Earlier work has shown that the elongation rate due to the binding of free RecA is insensitive to the applied force, free RecA concentration, hydrolysis and ATP concentration [53]. Thus, the process or processes that limit the rates must not depend strongly on any of these factors.

We further studied the effect of salt concentration dependence of the rate at which dsDNA is extended by strand exchange and the rate at which dsDNA is extended by the binding of free RecA for the cases where the dsDNA is pulled from the 3′3′-ends (Figure 2.4B). Both rates increase as a function of MgCl₂ concentration. Similar results were obtained at 24°C. These data provide additional support for the contention that the rate-limiting step in strand exchange is the same as the rate-limiting step in the extension of dsDNA due to free RecA binding from solution. Though increasing MgCl₂ concentration increases the observed extension rate, earlier results have shown that increasing the NaCl concentration decreases the observed extension rate for free RecA binding [61] and strand
exchange [62]. Together these results suggest that the characteristic extension rate is not a simple function of Debye screening, but rather depends on details of the interaction between the salts and the RecA–DNA complex.

2.5 Discussion

When the dsDNA is in the intermediate state, pairing the complementary strand bases with either the outgoing strand or the incoming strand results in significant tension on the base pairs due to the differential extension of the complementary strand and its pairing partners. One would expect that pulling on the complementary strand would make strand exchange less favorable because it would reduce the free energy advantage gained by strand exchange. In contrast, pulling on the outgoing strand would not make the transition less favorable. Similarly, applying a uniform tension on both strands should not alter the strand exchange rate. Finally, pulling on any set of dsDNA ends should have no effect on the binding of free RecA since that process does not involve strand exchange.

The results shown in Figure 2.3B are completely consistent with the proposal that the differential extension between the incoming strand and the outgoing strand drives strand exchange forward when the dsDNA is homologous to the ssDNA in the RecA–ssDNA filament. In particular, Figure 2.3B shows that the extension rates due to free RecA binding are independent of force. Similarly, the extension rates due to strand exchange are independent of force, except for the case where a force in excess of 15 pN is applied to the ends of the complementary strand. In that case, the strand exchange rate is significantly reduced. This reduction occurs whether the ssDNA is homologous to the 3’- or the 5’-end. Similarly, the results do not depend on which dsDNA strand was physically pulled, but on
whether or not the pulled strand is complementary to the ssDNA filaments used in the experiments. Thus, the consistent differences shown in Figure 2.3B must not be the result of some artifact associated with the preparation of the dsDNA construct, or with some peculiarity of one of the two filaments. These results are also consistent with earlier experimental results suggesting that the differential tension between the complementary and outgoing strands drives the unbinding of non-homologous dsDNA in the homology searching state [52].

The experimental results shown in Figure 2.3 suggest that the samples do not contain significant internal nicks; otherwise, the observed results would be independent of the ends to which the force is applied because the nicks would redistribute the stress between the two strands (see Appendix 1 section). Similarly, the finite stiffness of the backbone might redistribute the tension between the dsDNA strands so that the tension on both backbones was the same except at short regions near the end of the molecules (see Appendix 1 section and Figure S1.4). In this case, force-dependent effects would depend only on the terminus pulled at the end nearest the filament since the terminus pulled at the other end would have no effect on the force distribution near the filament. In this case, the situation illustrated in Figure 2.3C ii–v would produce the same result, whereas Figure 2.3C i would produce a different result. In contrast, the results show that force-dependent effects are not determined by the choice of terminus at the filament end, rather they depend on whether or not the force is applied to both ends of the strand that is complementary to the ssDNA in the filament. Thus, the results show that when the dsDNA is pulled from the 3′3′-ends, the tension on both backbones extends fairly uniformly along the length of the
molecule, whereas for 3′5′ pulling the tension over the entire length of the molecule is largely confined to the single backbone being pulled, as illustrated in Figure 2.3C.

We propose that the differential extension between the three DNA strands not only drives strand exchange forward for perfect homologs but also plays a vital role in decreasing the probability that partially homologous dsDNA will remain in the post-strand exchange state of the intermediate structure. Theory suggests that the differential extension in the searching state is so large that thermal energy fluctuations are insufficient to bind >~15 bp in the searching state, which represents less than one helical turn of the filament [63]. Thus, the state shown in Figure 2.2A will not occur. For perfect homologs, strand exchange reduces the stress on the dsDNA and allows more base pairs to bind, but non-homologs simply unbind. In the intermediate state, the differential extension of the complementary and incoming strands results in a non-linearity in the free energy as a function of the number of bound triplets because the complementary strand backbone physically connects neighboring triplets. The physically connected backbone redistributes stress along all of the base pairs in the filament. As a result, the stress on any given base pair depends on the binding of all of the other base pairs in the filament [63]; consequently, a mismatched triplet in the strand exchange state is not only unfavorable because of the loss of its Watson–Crick pairing but also unfavorable because the mismatch does not relieve any of the stress on the other bases in the strand exchange state. In contrast, the strand exchange of a sequence-matched triplet reduces the stress on other strand exchanged triplets, which substantially reduces the total free energy of the system [63]. Thus, in the intermediate strand exchange state the free energy difference between a homologous triplet and a non-homologous triplet can be much larger than the Watson–Crick pairing difference.
because the strand exchange of matched triplets reduces the stress on other homologous triplets in the intermediate strand exchange state.

A possible sequence of steps in the strand exchange of a perfect homolog is shown in Figure S1.5. For sequences <30 bp, the stress due to the differential extension is distributed across all of the base pair triplets. Under these conditions, the presence of a mismatch in the final state increases the stress in all the other triplets. Thus, the strand exchange of subsequent triplets beyond a mismatched triplet can be quite free energetically unfavorable, even if the subsequent triplet is perfectly sequence-matched to the corresponding triplets in the incoming strand. We note that previous experiments have shown that extending past a mismatch can be unfavorable [64].

The discussion above assumed that the base flipping that tests for homology recognition occurs in an intermediate state where the position of the complementary strand differs from its known position in the final post-strand exchange structure. This proposition is supported by the similarity between the dsDNA extension rate due to strand exchange and the dsDNA extension rate due to the binding of free RecA to dsDNA, as we discuss below.

Figure 2.4A shows that for a range of temperatures, the dsDNA extension rates due to free RecA binding from solution are the same as the extension rates due to strand exchange, while pulling on the outgoing strand. Figure 2.4B shows that the rates as a function of MgCl\(_2\) concentration are also similar suggesting that the two processes have a common rate-limiting step (Figure 2.3B). Given that binding of RecA to dsDNA increases dsDNA extension in both strand exchange and free RecA polymerization along dsDNA,
one might suggest that dsDNA extension is the common rate-limiting step; however, we propose that this is unlikely for the following reasons: (i) Earlier measurements that divided the strand exchange process into steps have suggested that the sequence independent dsDNA extension occurs much more rapidly than subsequent sequence dependent steps [49]. (ii) Each homology search attempt requires extension [49, 52], and the measured in \textit{vitro} homology searching rate exceeds $\sim 50 \text{ bp/s}$ [65] or $100 \text{ bp/s}$ [40]. (iii) \textit{In vivo}, the search rate must also exceed $100 \text{ bp/s}$ in order for the homology search to occur on a biologically relevant timescale. (iv) The observed rates for free RecA binding and strand exchange are both insensitive to force, but theory suggests that if the rate depended on extending the dsDNA [54, 55], then the observed extension rates should depend on force, like the force dependence of the nucleation rate for free RecA [54].

Since transition steps that do not occur in both strand exchange and free RecA binding cannot provide the common rate-limiting step, we can eliminate several processes that might have limited either individual process. Unlike strand exchange, the binding of free RecA from solution does not involve base flipping, homology recognition or repositioning of the outgoing strand. Similarly, unlike free RecA binding, extension of the strand exchange product does not involve free protein diffusion or assembly of protein monomer interfaces from interactions with a free monomer.

Additional support for the contention that the strand exchange rate is not limited by the homology search process is provided by earlier experimental work that studied strand exchange in bulk reactions [66]. These bulk reaction experiments showed that the process, which limits the strand exchange rate, occurs after the homology searching process is
complete where the search process includes the dsDNA binding, extending and the base flipping to test for homology.

Having ruled out all of the possible rate-limiting factors detailed above, we propose that the repositioning of the complementary strand is the rate-limiting step in both free RecA binding and strand exchange. If in free RecA binding experiments one strand of the dsDNA initially binds to the strong contacts in site I while the other strand is pulled toward the positive charges in the direction of site II, then the pull orients the bases. The interaction between the L2 loop and the bases may also position Phe203 in the rise between the triplets in the intermediate state, providing a functional role for Phe203, which does not appear to have a functional role in the initial or final states. The resulting RecA–dsDNA structure would be similar to the post-strand exchange intermediate (Figure 2.5A, Appendix 1 section, Figures S1.6 and Figure S1.7). If this is true, the repositioning of the complementary strand could be the rate-limiting step for free RecA binding and strand exchange (Figure 2.5B).
Figure 2.5: Proposed steps in strand exchange and free RecA binding. (A) Schematic of the steps involved in the proposed strand exchange process. Grey areas show regions occupied by the protein, excluding the L2 loop: (i) incoming strand (cyan), in the post-strand exchange state. (ii) dsDNA outgoing strand (orange circle) bound into Site II; complementary strand (red filled circle) in proposed intermediate position. (iii) Complementary strand bases and incoming strand bases rotate in search of homology, where an L2 loop may rotate with the incoming strand. (iv) Strand exchanged state that may benefit from an interaction with the L2 loop if the bases are homologous. (v) The heteroduplex dsDNA rotates to the final X-ray structure position, possibly accompanied by the L2 loop. (B) Schematic of possible steps in the binding of free RecA to dsDNA: (i) an additional free RecA binds to the dsDNA in the intermediate state, possibly accompanied
by an interaction with the L2 loop. (ii) dsDNA (red circle) rotates to the final state shown by the x-ray structure, possibly accompanied by the L2 loop.

In sum, we propose that the final step in strand exchange is the transition from the intermediate state to the final state, which requires the repositioning of the complementary strand. In the final state, the differential extension between the incoming and complementary strands is slightly lower than the differential extension in the intermediate state. In addition, for perfect homologs, interactions with the L1 and L2 loops may provide additional mechanical support for the rises which is absent in the intermediate state (see Appendix 1 section, Figures S1.6, Figure S1.7 and Figure S1.8). If the linear term in the free energy is favorable for the intermediate post-strand exchange state, but the non-linear term is favorable for the final state because of the reduced stress on the base pairs due dominantly to support from the protein, then the transition from the intermediate state to the final state will not become free energetically favorable until a sufficient number of contiguous homologous bases have undergone strand exchange [63]. If the required number were less than six triplets or 18 bases [67], then the non-linear difference in the free energy due to the differential extension would provide yet another significant enhancement in sequence stringency in comparison with a system that relied only on the Watson–Crick pairing energy of each individual base, since the correct sequence matching of all of the contiguous bases is required to move forward with strand exchange.

In conclusion, we speculate that the intermediate structure represented in Figure 2.5A iv may play the following roles in homology recognition: (i) The existence of an
intermediate structure increases homology stringency because the total probability that the system will make a transition from the initial state to the final state is the product of each of the intermediate transition probabilities [63]. (ii) If in the intermediate structure, the DNA tension is between the tension in the initial and final states, then dsDNA tension in the searching and intermediate states can be large enough to enforce accurate sequential kinetic proofreading, while the final state tension is low enough that adding dsDNA to site II remains free energetically favorable as long as almost all of the bound base pairs are in the final state and the total number of bound base pairs <~80 [63]. (iii) For homologous triplets, strand exchange is driven forward by the reduction in dsDNA tension that occurs as a result of the differential extension of the backbones. (iv) The extension difference between the complementary strand and its pairing partners amplifies the free energy penalty due to the strand exchange of non-homologous bases so the penalty may be much larger than the Watson–Crick pairing penalty. (v) If incorrectly paired heteroduplex bases cannot stack with minimal free energy, then the transition from the intermediate state to the final state may be more favorable for homologs than for heterologs because homologs have less stress on the base pairs and more amino acid residue interactions. (vi) The free energy penalty for a sequence mismatch may be large enough to make the transition from the intermediate state to the final state so unfavorable that it does not occur before strand exchange spontaneously reverses and the dsDNA unbinds. This may explain why if <~21 bp are exchanged, encountering a base pair mismatch allows strand exchange to be fully reversed. In contrast, if >30 bp are exchanged, strand exchange cannot be reversed [64, 67-69]. A detailed discussion of these points is presented in Appendix 1 section.
The type of kinetic proofreading system described in the previous paragraph represents a departure from Hopfield’s original proposal for kinetic proofreading [64, 70-72]. In that original proposal, heterologs initially unbind faster than homologs, and the system makes a sequence independent irreversible transition to a tightly bound state after a time T. This article, however, considers a system where the initial unbinding rates for homologs and heterologs are the same, but a series of fully reversible sequence dependent intermediate transitions to more tightly bound states provides homology recognition. Thus, this article may provide insights about intermediate structures in the homology recognition–strand exchange process and their possible roles in fully reversible kinetic proofreading. Similar models may apply to other natural systems. Finally, since RecA family protein-based homology recognition is fully reversible it may provide a new paradigm for artificial self-assembly of nanoscale.
Chapter 3

Integrating multi-scale data on homologous recombination into a new recognition mechanism based on simulations of the RecA-ssDNA/dsDNA structure

3.1 Synopsis

RecA protein is the prototypical recombinase. Members of the recombinase family can accurately repair double strand breaks in DNA. They also provide crucial links between pairs of sister chromatids in eukaryotic meiosis. A very broad outline of how these proteins align homologous sequences and promote DNA strand exchange has long been known, as are the crystal structures of the RecA-DNA pre- and postsynaptic complex; however, little
is known about the homology searching conformations and the details of how DNA in bacterial genomes is rapidly searched until homologous alignment is achieved. By integrating a physical model of recognition to new modeling work based on docking exploration and molecular dynamics simulation, we present a detailed structure/function model of homology recognition that reconciles extremely quick searching with the efficient and stringent formation of stable strand exchange products and which is consistent with a vast body of previously unexplained experimental results.

3.2 Introduction

Homologous genetic recombination is a prescribed and necessary part of the DNA metabolism of every free-living organism. Recombination can accurately repair DNA double strand breaks, provides crucial links between pairs of sister chromatids in eukaryotic meiosis, and contributes in smaller ways to a host of additional cellular requirements. All of this is centered on the function of RecA-class recombinases and their capacity to catalyze (a) an alignment of homologous sequences in one single-stranded DNA (ssDNA) and another double-stranded DNA (dsDNA) and (b) the transfer of one strand of DNA from the duplex to the initiating ssDNA leading to the formation of a stable heteroduplex if the ssDNA and dsDNA are homologous. The latter process is referred to as strand exchange and is illustrated schematically in Figure 3.1.
RecA protein, found in essentially all bacteria, is the prototypical recombinase. Homologs exist in other organisms, including RadA in archaea and Rad51 and Dmc1 in eukaryotes. An outline of the mechanism by which these proteins align homologous sequences and promote DNA strand exchange is in hand [37, 38, 73], and RecA will be used to illustrate it. First RecA protein forms a presynaptic active filament by polymerizing onto an initiating ssDNA. The resulting filament is a right-handed helix, with six RecA subunits per turn and three nucleotides of DNA bound per subunit with the initiating ssDNA bound deep within the helical groove of the filament, in a location referred to as the primary DNA-binding site (site I) (Figure 3.2A and Figure 3.2B). The filament then interrogates any nearby dsDNA to find a homologous match. During the search, the dsDNA enters the helical filament groove and is initially (usually transiently) tethered. Once sequences are aligned, the sampling required to sense homology involves transient destabilization of a short region within the dsDNA and limited base-flipping within that

**Figure 3.1:** Schematic of the homology searching process. The initiating, complementary, and outgoing strands are shown in orange, purple, and cyan, respectively. RecA protein monomers are excluded for simplicity.
short region [49, 74]. If the flipped bases in the complementary strand match the corresponding bases in the initiating ssDNA, the pairing of the complementary strand substrate is transferred to the initiating ssDNA resulting in the formation of the product duplex, while the complementary strand’s original pairing partner, the outgoing strand, is displaced (Figure 3.1). The beginning and endpoints of this process, the bound initiating ssDNA, and the product dsDNA bound to the site I have been structurally defined through crystallography [44] (Figure 3.2C-D). In both cases, the bound DNA strands are extended to about 1.5 times the B-form dsDNA length and present a very specific conformation, where the extension is not uniformly distributed; instead, the bases are grouped into stacked nearly B-form groups of three consecutive base pairs (triplets) that are separated by large rises. The B-form triplets are compatible with the overall extension and untwisting because the extension and untwisting occur dominantly in the rises between the triplets.
Experimental work suggests that strand exchange is divided into stages characterized by very different strange exchange rates. In particular, single molecule measurements of the strand exchange rate indicating that strand exchange slows drastically after a very rapid initial interaction [50]. The very rapid early stages of homology recognition are very difficult to capture experimentally [49, 50]; however, both bulk [75] and single molecule experiments [58, 76] have captured strand exchange after the initial rapid interaction. Those experiments indicate that after the initial rapid interaction, strand exchange progresses at a characteristic rate that is insensitive to dsDNA length and ATP
hydrolysis, but dependent on temperature. At 37°C the characteristic rate is \( \sim 6 \text{ bp/s} \), whereas at 22°C it is \( \sim 2 \text{ bp/sec} \) [58, 75, 76].

The division of homology recognition into stages where the first stage tests \( \sim 8 \text{ bp} \) is consistent with measurements suggesting that the strand exchange product remains highly unstable until the length of the product extends to 8-9 consecutive base pairs, after which there is a sudden increase in the stability of the product in both RecA [77] and Rad51 [78]. This greatly increased stabilization may result from a transition to a metastable conformation whose strand exchange rate is much slower than the strand exchange rate for the initial interaction. Both RecA [77] and Rad51 [78] experiments have shown that after the initial rapid interaction each strand exchanged triplet increases the stability of the strand exchange product until an asymptotic binding value is approached when 15-20 contiguous homologous bp occupy the strand exchange conformation. The homology recognition and strand exchange process that leads to the formation of 20 bp strand exchange products can occur without ATP hydrolysis [79]; however, the unbinding of the heteroduplex product dsDNA from the protein requires ATP hydrolysis [58].

Seminal early work suggested that RecA might use an initial very stringent homology test of \( \sim 8 \text{ bp} \), that is followed by a less stringent testing stage [47]. Recent experimental work indicates that an initial \( \sim 8 \text{ bp} \) test that accepts \( \sim 1 \) mismatch in 8 bases is followed by a less stringent testing stage that allows homology recognition to pass over non-homologous triplets within otherwise homologous sequences [80]. The progression of strand exchange through mismatched triplets in otherwise homologous sequences was also observed in previous work [81].
Experimental and theoretical studies suggested that dsDNA tension could provide a physical mechanism that divides the search into stages since the tension creates an unfavorable binding energy contribution that increases non-linearly as a function of the number of base pairs bound in the positively charged helical track within the filament groove (Figure 3.2B) known as the secondary DNA-binding site (site II) [52, 82]. That previous work proposed that favorable interactions between the dsDNA and the presynaptic filament drives rapid binding of 2-3 base pair triplets to site II. The binding of two or more triplets to site II extends the outgoing strand more than the complementary strand, creating a mechanical stress on the base pairs joining the complementary strand to the outgoing strand backbone. That stress destabilizes the base pairing in the dsDNA, leading to rapid base flipping and homology testing of ~ 8 bp. In this model, dsDNA tension prevents more bases from binding to site II unless the ~ 8 bp that are tested initially make a homology dependent transition to a metastable conformation. If the initial test is passed, the system makes a transition to a metastable conformation [82].

Recent theoretical work [83, 84] suggests that a two stage system could overcome the general speed stability paradox that arises because stringency requires correct interactions to be strongly bound while speed requires that nearly correct interactions be easily reversed. More importantly, the theoretical work suggests that the two stage system could search more than $10^5$ bp in less than 1000 seconds, while non-homologous pairings efficiently form stable strand exchange products [84, 85]. In the system, the searching speed is provided by the initial rapid test that rejects ~ 95% of all pairings before they progress to the slower testing stage. Only rare pairings that pass that initial test would progress to the second slower testing stage that is characterized by the slow strand
exchange rate [84, 85]. This is consistent with recent experimental studies showing that 20-nt sequences presenting less than 15 matches (which represent the large majority of accidental matches in vivo) do not lead to the formation of any detectable metastable intermediate during strand exchange, while rarer accidental sequence matches lead to the formation of metastable strand exchange products before being rejected [80].

The early stage of recognition during which strand exchange progresses rapidly encompasses much of the mystery of homologous recombination: what structural features allow the search to form the correct homologous pairing in much less than an hour? In principle, molecular modeling should be an appropriate method to fill the structural gap between the beginning and endpoints of the recognition and strand exchange process, while integrating the large experimental information that has been accumulated along the last decades. However, modeling such a large and complex system remains a formidable challenge for current computational capacities. Early studies considered only nucleic acids without including interactions with the protein. Those studies successfully explored how changes in DNA properties due to RecA-induced stretching distortion may contribute to the mechanism [86-88]. However, the whole nucleofilament system must be considered in order to understand which structural features in the DNA-filament association drives strand exchange forward for homologous sequences while rapidly halting the process in case heterologous sequences are encountered.

Recently, Saladin et al. [89] have investigated possible association geometries at early stages of the recombination process using a multi-scale strategy consisting of coarse-grained extensive docking search, interactive simulations of selected docking geometries and atomic molecular dynamics simulations. This work showed that combined state-of-
the-art modeling methods can now tackle very complex systems that are out of reach of classical modeling methods [90]. Due to the high dimensionality of the systems under study, such strategies gain in efficiency if the search space can be restricted using available experimental information such as low resolution data or information obtained at the residue level. In the case of polymeric systems such as the active filament, the multiplicity of identical residues in similar environments within each monomer makes it difficult to restrict the search space based on the proximity to given residues: one particular interacting residue with functional role cannot be easily distinguished from the non-interacting instances of that same residue. If in addition the system is dynamic like in recombination, it is generally not known in which step of the reaction a given residue is involved. Models of the recombination system must therefore be guided by other considerations than the residue information, which can nevertheless be used for a posteriori validation.

In their study, Saladin et al. resolved to look for direct interactions between the bases of the searched dsDNA and the initiating ssDNA as early contact points between the dsDNA and the filament. In this scheme, the strength of dsDNA binding to the filament would directly depend on sequence homology. The study concluded that such mode of direct interaction involving short segments of three consecutive bp is possible provided that the DNA structure adopts a conformation curved towards the major groove upon binding the filament and that at least one long RecA loop (L2 loop, residue 198 to 206, shown as yellow in Figure 3.2B) is displaced [89]. It is compatible with a propagation of the reaction where groups of three dsDNA base pairs would successively unroll and bind the corresponding ssDNA bases via homology-dependent Watson-Crick interactions,
while displacing one L2 loop at a time. Stabilization of such system would increase linearly when more homologous triplets bind and decrease if at least one bp does not match.

This model has many desirable features; however, the timescale suggested by the model is slow since at least one L2 loop needs to be displaced at each search event involving three bp, so it is unlikely that millions of search events take place in less than an hour. However, in the context of a two-stage process discussed above, the 3 bp-progression characterizing the Saladin model may be compatible with the second (slow) phase of strand exchange progression, where mismatches that have passed the rapid initial testing are eliminated.

In the present work, we explore the geometries that may characterize the first rapid stage by investigating an alternative route to sequence probing and strand exchange. Given the steric configuration of the filament groove [44] the route via site II is in fact the only alternative to the direct base contact previously explored. The modeling followed the outlines of the Saladin et al. former work with, however, some noticeable additional challenges since no structure at the atomic level is available for the stretched and unwound distorted DNA structure induced by its binding to site II [49] (this structure necessarily differs from the endpoint DNA structure resulting from strand exchange).

The initial search for geometries of dsDNA association to the filament therefore needed to simultaneously search for favorable docking geometries and generate DNA structural distortions that would optimize the association. Yet, the few published flexible docking methods that account for DNA flexibility during the docking search either limit the DNA deformations to curvature [91] or explicitly rely on experimental information at the residue level to restrict the search space [92]. Since neither approach could be used
here, we developed a multi-step strategy for this part of the modeling, which can be related to the approach by Banitt and Wolfson as it includes the docking of DNA fragments, but which takes advantage of the helical periodicity of the system and allows any type of distortion. More generally, the search was guided by principles issued from recent theoretical studies of recognition at the genomic scale [84, 85] as well as previous modeling work on distorted DNA [87, 93-99]. Finally, when submitted to all-atom molecular dynamics simulations in solvated environment, the resulting model spontaneously gave way to pairing exchange. Importantly, without any guiding, the structure assumed by the strand exchange bases corresponds to the structure known from crystallography. This spontaneous evolution towards the known structure of the strand exchange product provides substantial support for the validity of the modeling that precedes it.

We will end by discussing functional implications of structural features revealed in this work, some of which we mention briefly here. Importantly, the model suggests that the tension resulting from the extension of the dsDNA bound to site II naturally divides the search into stages, where the first stage rapidly tests ~ 8 bp. Furthermore, the unanticipated iterative flipping of successive base pair duplets provides previously unexpected advantages in searching speed. Finally, the work suggests a possible origin for the large increase in stability that is observed after 8 contiguous bp have undergone strand exchange.
3.3 Methods

Overall strategy

The construction of intermediate states of DNA association to the nucleofilament active for recombination followed a combination of one or several modeling steps described below:

(i) Scanning the possible modes of DNA/filament association via docking simulations; when the association was expected to induce distortions in the DNA structure, the scanning step was refined either by using different (curved) forms of the DNA or by docking small DNA fragments, followed by reconstruction into a complete DNA;

(ii) Constructing DNA structures either with helical symmetry or where different DNA regions interact in different ways with distinct regions of the RecA filament, using the PTools/Heligeom computing tools and restrained energy minimization;

(iii) Refining the modeling results using short all atom, fully flexible molecular dynamics (MD) simulations;

(iv) Simulating the dynamic evolution of the refined model during longer MD trajectories.

Docking simulations

Rigid body docking of B-form DNA to the filament (receptor) was performed using either Autodock Vina [100] or the ATTRACT program [93, 101] in its PTools/PyAttract implementation [89, 102]. These simulations explored the possible mode of association of oligonucleotides to the filament as a gateway to filament incorporation. In both cases, the B-DNA structure (ligand) was docked on a protein nucleofilament composed of five RecA
monomers bound to ssDNA (PDB code 3CMW) [44] (receptor). Autodock Vina considers
the association partners in atomic representation and uses docking scoring function
developed based on X-Score [103]. For this simulation, the whole filament surface was
selected as putative binding pocket on the receptor surface. The ATTRACT docking
simulations in coarse-grained representation followed the same protocol as detailed in
reference [89]; however, the simulations done here include the L2 loops that had been
omitted in that previous work. The two simulations converged towards DNA binding in
the filament groove as described in the Results section.

Flexible docking approaches were used to model the DNA incorporated in the
filament since the stretching/unwinding distortions induced by its binding to the filament
were not structurally elucidated at the atomic level. The docked DNA was either a single-
stranded or a double-stranded DNA with sequence homologous to that of the RecA-bound
single strand in the crystal structure (i.e., (dT.dA)n). Three different protocols were used
to explore the possible interaction sites, and two protocols were used for subsequent
construction. Using different protocols aimed at avoiding any possible dependency on the
flexible docking method, given that the docked structure presents uncommon distortions
and length-dependent internal stress.

Two out of the three docking protocols started by a rigid body docking run as
detailed above. In the ATTRACT protocol, the docked ligand was a curved (dT.dA)30
DNA built from the structure of SRY-bound DNA (PDB code 1HRY) (described in
reference [104]), and the simulations targeted the whole filament surface, again in the
presence of the L2 loops. The best interacting patches on the protein surface were situated
in a cleft between the L2 loop and the filament interior, containing the positively charged
protein residues K198, K216, R222, R226, R227, R243 and K245 associated to site II in the literature [44]. In the Autodock Vina protocol, the docked ligand was a short DNA fragment composed of a 3-base-pair (bp) long nucleotide phosphate-deoxyribose, and the simulations targeted the putative site II basic patches described above. A third protocol explored the positioning of a preliminary stretched-unwound single-stranded DNA in the site II cleft using interactive simulations with the BioSpring simulation motor described in [89], with the help of a haptic device. In this case, both the filament and the DNA were considered in coarse grained representation, and internal flexibility of the DNA and selected protein regions (L2 loops and residues lining site II, i.e. R143, R145, K226, K227) was accounted for using an augmented spring network [89]. Each of the three protocols resulted in identifying the best interacting regions on the protein surface and the best interacting trinucleotides along the DNA structure; the local geometries associated to the most favorable interaction energies were used for further modeling.

Modeling the DNA bound to the identified sites followed two main protocols. The first one consisted in reconstructing full single- or doubled-stranded DNA structures starting from the protein-bound trinucleotides selected in the previous step, taking into account the stoichiometry of association of three nucleotides per RecA monomer. More precisely, these geometries were replicated at each equivalent region along the filament, following its helical symmetry, and the DNA trinucleotide single- or double-stranded fragments were linked together using energy minimization of the phosphodiester backbone geometry under harmonic restraints on the interacting phosphate atoms to maintain the interactions with their protein residue partners. This was followed by short MD simulations (see below) and annealing of a second strand when the starting DNA fragment was single-
stranded. Alternatively, a whole (dT.dA)9 DNA oligomer obtained by truncating the site I-RecA-bound dsDNA structure in the 3CMX, already presenting 50% stretching and 40% unwinding deformations with respect to B-DNA, was fitted to the site II patches identified in the first step. To this aim, harmonic restraints were applied to the phosphate atoms of thymine residues 1, 2, 4, 5, 7, 8 during energy minimization with the Jumna software [93] to radially pull these phosphate atoms to the favorable positions identified in site II under constant helical symmetry conditions (i.e., conserving the global stretching and winding characteristic of the binding to the site I). This process was applied to DNA in the absence of the protein to capture the contribution of the DNA internal mechanics to the protein-DNA association geometry.

**Molecular dynamics simulations**

Molecular dynamics (MD) simulations were performed using the NAMD 2.9 package [105] with the CHARMM 27 force field including CMAP correction [106]. Solvation of the simulated structures used TIP3P water model placed within a 125 Å x 120 Å x 170 Å box, subject to periodic boundary condition. Physiological concentration of Na\(^+\) and Cl\(^-\) ions of 0.15 mol/L were added to each simulation to maintain electro-neutrality (Figure S2.1A). As a whole, the simulated systems contained over 260,000 atoms. We used 2-fs time steps, and the bond lengths involving hydrogen atoms were constrained using the SHAKE method. We utilized particle-mesh Ewald method for long-range electrostatics calculation, and van der Waals interactions were smoothly switched off at 10–12 Å by a force-switching function. The temperature and pressure were maintained along the simulation using a Langevin dynamics scheme and a Nosé-Hoover-Langevin piston,
respectively. We performed MD simulations on several RecA/ssDNA/dsDNA complexes resulting from the docking and construction steps (see above). These structures were first energy minimized using 50,000 conjugate gradient energy minimization steps followed by slow heating from 30 K to 300 K over 500 ps. During these preparation steps, the α-carbon and phosphate atoms of the protein and nucleotide were restrained using a soft harmonic potential with a spring constant of 0.5 kcal.mol\(^{-1}\).A\(^{-2}\). The system was further allowed to equilibrate for 5 ns in the NPT ensemble with pressure of 1 bar and temperature of 300 K. During the equilibration process only the α-carbon atoms of the terminal RecA monomers and phosphate atoms of the three nucleic acid strands were restrained with a harmonic potential with slowly decreasing spring constant from 0.5 to 0.05 kcal.mol\(^{-1}\).A\(^{-2}\). In the production simulations, the system was maintained under NPT ensemble conditions, and harmonic restraints of 0.05 kcal.mol\(^{-1}\).A\(^{-2}\) only applied to the α-carbon of terminal RecA monomers and phosphate atoms of the terminal nucleotide of each nucleic acid strand. The simulation durations varied between 10 and 25 ns. In one case, an accelerated dynamics protocol was used after the first 10 ns of classical MD [107]. For this phase of the simulation, a dual boost mode was used. The threshold energy \(E\) (kcal.mol\(^{-1}\)) and the acceleration factor \(α\) were respectively set to 21678, 1946 and -775729, 56772 for the dihedral potential and the total minus dihedral potential which were calculated as suggested by Markwick et. al. [108]. The root mean square deviation (RMSD) of the backbone and the energies during the heating, equilibration and production of the simulation are shown in Figure S2.1B.
Construction with PTools/Heligeom

Model construction was performed at different stages of the study; for example, to replicate DNA fragments during the flexible docking construction stage, when joining DNA stretches that interact in a different way with different regions of the filament or when issuing intermediate structures for the movie animation. Many of these DNA structure manipulations or helical transformations were performed using the PTools/Heligeom library [109].

3.4 Result

In this work, we modeled intermediate species along the early process that goes from non-specific DNA uptake by the filament to the very first pairing exchange events. These events entirely take place during the first, very rapid stage of HR. We divide the overall DNA/RecA structural changes along that early process into four major conformational classes: (1) presynaptic active filament, (2) bound B-form dsDNA, (3) conformations with dsDNA distortions stabilized by interactions with site II, and (4) postsynaptic filament with the complementary strand paired with the initiating strand bound to site I.

The structure of the presynaptic active filament (class 1), which consists of the RecA subunit polymerized onto the initiating ssDNA, was determined through crystallography [44] (Figure 3.2B and Figure 3.2D). The initiating ssDNA is bound deeply in the site I in the region where the L1 and L2 loops meet. Its structure, along with that of
the protein filament, is almost identical in the presynaptic and the postsynaptic filaments [44], as Figure 3.2 illustrates.

**Non-specific binding of B-form dsDNA to the RecA filament**

The structure of the non-specific binding intermediate with unaltered DNA structure (class 2) was modeled as the result of rigid body docking of B-DNA to about one turn of the presynaptic active filament using either the Autodock Vina in all-atom representation or the ATTRACT program in coarse-grained representation (see Methods). The two protocols converged towards structures where the DNA is partially inserted in the filament groove and where at least four DNA phosphates strongly interact with residues K280, K282, K286, and K302 through the formation of salt bridges. In these structures, which remained stably assembled during 20 ns of MD simulations (Figure S2.2 and Figure 3.2), the interacting phosphates are distributed in the two DNA strands while the basic protein residues belong to two consecutive or near-consecutive C-terminal domains (CTD) of RecA. Two examples of such structure are shown in Figure S2.3.

**Non-specific dsDNA binding to the filament site II**

The presence of bulky L2 loops folded upon the ssDNA in site I of the RecA filament (residues 198-206 shown in yellow in Figure 3.2B) only offers two possibilities for the dsDNA alignment in parallel with the ssDNA, either on the ssDNA side of the L2 loops or on the opposite side (Figure S2.4). The first possibility has been explored in former theoretical work [89] which concluded that at least one L2 loop needs to be displaced for each dsDNA triplet to reach alignment with the incoming strand. In the present work, we explore the second route. The C-terminal side of the L2 loops forms a strongly
electronegative cleft when intersecting with the protein core (Figure S2.5), which can accommodate a DNA single strand as was confirmed here using interactive docking simulations with the BioSpring software (see Methods section). This cleft, lined by clusters of basic protein residues R226, R227, R243 and K245, most of which are well-conserved among bacterial RecA proteins [110], was previously proposed to be the secondary DNA-binding site (site II) of the protein [44]. In what follows, we will assimilate the cleft and its clusters of basic residues to site II. Indeed, the three flexible docking protocols with Autodock Vina, ATTRACT, or BioSpring (see Methods) that we used to model single-stranded or double-stranded DNA in site II in the presence of the L2 loops converged towards these clusters of basic residues. From the results of the docking simulations, we modeled the structure of DNA in site II as described in Methods.

In the resulting structure shown in Figure 3.3 (see also Figure S2.5), each two consecutive phosphates from outgoing strand triplets strongly bind a cluster of basic residues R226, R227, R243 and K245 of each RecA monomer. Both the outgoing and the complementary strands directly contact many L2 loop residues including F203, and to a lesser extent M202, which notably intercalate in the rises between triplets supporting the extension.
**Figure 3.3:** A detailed view of the active filament bound to dsDNA stabilized by interactions with RecA C-terminal domain and site II residues. The initiating, complementary, and outgoing strands are shown in orange, purple, and cyan, respectively. The lysine residues K280, K282, K286, and K302 are shown in green. K232 is shown in silver. The secondary DNA-binding site residues R226, R227, R243, and K245 are shown in red, and the L2 loop reside M202 and F203 is shown in yellow. The black arrows point to the site II residues at the beginning and end of the extended and untwisted region. F203 intercalates the bound dsDNA, and some simulations showed M202 can intercalate as well (see also Figure S2.5).
The dsDNA in site II presents exactly the same helical characteristics as DNA bound to site I, with an average axially projected rise of ~ 5.1 Å and an average twist of ~ 20° per base pair step, consistent with the overall helical characteristics of the filament (Figure S2.6). It ensures that the corresponding bases between initiating (in site I), complementary, and outgoing strands (in site II) can remain in registration, as suggested earlier [76]. The DNA in site II is, however, much more stressed than the DNA in site I since the phosphodiester backbone of the outgoing strand is radially displaced from the helix axis by ~ 15 Å, whereas the phosphate backbone of the initiating strand is only displaced by ~ 9 Å (Figure S2.6). Thus, the average distance between consecutive phosphate groups for the outgoing strand DNA bound to site II is ~ 2 Å longer than the ~ 5 Å separation for DNA in site I. Indeed, the phosphate extension of the outgoing strand DNA in site II approaches the maximum possible extension of the backbone [111]. Such large extension prevents more than two of the three bases within a triplet from stacking as in B-form dsDNA, and, indeed, all MD simulations run on the system indicated constant evolution of the patterns of intra- or inter-strand base stacking, base-F203 stacking, and base pairing during the trajectories. The dynamic evolution of site II-bound dsDNA will be further described below.

**Junction between the dsDNA and the B-DNA fragment**

The direction of the B-form dsDNA bound to the open filament groove substantially differs from that of the dsDNA bound to site II parallel to the helical axis. As a result, the junction between these two DNA regions can be expected to be bent. Similar expectation arises from structural considerations, as the junction between the stretched/unwound DNA
distortions inside the filament and the B-form dsDNA at the filament entry is expected to produce a kinked conformation.

Indeed, previous work revealed a mechanical relationship between the type of global DNA distortion found in the RecA filament and the local distortion induced by DNA minor groove binding proteins such as SRY, TBP or HU [97, 112]. Applying pulling restraints to the 3′-extremities of the DNA region that contacts the protein either gave rise to the bending/untwisting distortions observed in case of single protein binding or to the stretching/untwisting distortions observed in the RecA filament [88, 98]. For this family of distorted DNA structures, kinks (and therefore stacking interruption) appear at junctions between DNA regions that shifted to the distorted state and the rest of the structure in B-form dsDNA. When the structural transition concerns the whole DNA structure like in the RecA case, the stacking interruptions turn into intercalation-like base pair separation.

Quite interestingly, recent investigation on the intercalation pathway of daunomycin to the DNA double helix identified a kinked DNA state as a metastable intermediate between a straight B-form state with daunomycin bound in the DNA minor groove and the intercalated state [99]. In the case of dsDNA incorporation into the RecA filament, we propose that both types of stacking interruptions co-exist along the same DNA, either as a kink at the junction between the B-form dsDNA and the distorted regions, or as intercalation sites within the distorted region. This further suggests that a mechanism analogous to that of daunomycin intercalation may be a key component of the process of DNA extension in site II, as illustrated in Supplementary movie (scene 3). Experimental support for the presence of a sharp bend at the junction between the DNA outside the filament and the DNA inside the filament is provided by earlier experiments that imaged
Rad51 mediated strand exchange and found that DNA incorporation into the filament during strand exchange is accompanied by a sudden change of direction, with measured angular values close to 90° [113].

Based on these observations, we modeled the junction region between the CTD-bound B-form dsDNA region and site II-bound stretched-unwound DNA region starting from the structure of DNA bound to the SRY protein [114]. This DNA was first anchored to the filament interior using docking simulations as reported in Methods. The most favorable docking regions involved the bent region on the DNA side and a region largely overlapping the site II region and partially a CTD on the protein side. Specifically, the MD relaxed structure presented salt bridges between the basic protein residues R243, R245, K232, R226, R227 and DNA phosphates from both strands and on both sides of the curved region (Figure 3.3 and Figure S2.5). Additionally, the aromatic ring of a phenylalanine residue F203 interacts in the DNA minor groove at the level of the kink, thus stabilizing the associated stacking interruption.

**Dynamic evolution of site II-bound dsDNA**

We carried out extensive all atom MD simulations of dsDNA structures bound to one turn of the RecA-ssDNA filament in an explicit solvent environment (see Methods). Since the evolution of MD simulations depends on details of the starting structures, we performed several simulations with different starting structures, either restricted to the dsDNA region bound to site II or including the dsDNA B-form region bound to the CTDs and the junction region, with a number of bound triplets varying between one and five.
Some robust features of the homology search process emerge from the comparison of these MD simulations. First of all, all DNA strands remained bound to the protein during the course of the MD simulations extending over more than 20 ns, despite the large DNA distortions. Furthermore, if the structure contained B-form regions, the B-form dsDNA showed little structural evolution regardless of the number of triplets bound to site II (Figure S2.2); however, whether or not B-form tails were present, the evolution of the region of the dsDNA that bound to site II showed an evolution that depended strongly on the number of base pairs bound to site II, where the dependence is consistent with a base pair stress that increases with the number of base pairs bound to site II. Appendix 2 includes a detailed discussion of the evolution of structures in which all of the dsDNA is bound to site II. That discussion concludes that such structures are not suitable for homology recognition and strand exchange.

As mentioned above, simulations of structures in which the site II bound base pairs are joined to B-form tails are consistent with base pair tension increasing with the number of site II bound triplets (Figure S2.7). In particular, structures with only one triplet bound to site II rarely evolved. In contrast, when two triplets were bound to site II, one or two bases spontaneously flipped towards the minor groove, typically occurring within tens of nanoseconds after two consecutive triplets were bound in site II. The base flipping was followed by Watson-Crick pairing interactions between bases from the complementary strand and bases from the site I-bound ssDNA.

We decided to further explore the potential energy landscape associated with structures with two dsDNA triplets bound to site II using accelerated MD simulations [107]. This technique, where the depth of potential energy wells is lessened with no a priori
choices regarding variables or pathways, accelerates the passage of energy barriers to enhance the exploration. The results of this simulation are described below and in Figure 3.4 and Figure 3.5. The trajectory after the first 10 ns can also be seen in supplementary movie (last scene).

**Figure 3.4:** Illustration of base-pair homology testing using a duplet that is nearly in site I, while the third base of the triplet at the 3’ end of the triplet remains positioned in site II. (A) Structure of full filament with stretched and unwounded dsDNA bound to site II. The transparent and solid structures represent the conformations before and after 14 ns of simulation. (B) Structures before and after 14 ns simulation (10 ns conventional MD and 4 ns accelerated MD). (B) I. The complementary backbone in the postsynaptic filament is shown in silver. After flipping, the backbone near the flipped duplet shown in purple is nearly in postsynaptic position. (B) I. and II. Show top and schematic view of the
complementary and initiating bases at the 5’ end of a triplet before and after flipping. The transparent structures indicate the initial positions. The purple arrow indicates the relocation of the complementary strand backbone. (C) Postsynaptic dsDNA in site I; side view and top view are show in I. and II. In (C), respectively.

Interestingly, base flipping that gave way to Watson-Crick binding in site I (shown in Figure 3.4 and Figure 3.5) only occurred when two bases collectively flipped as a stacked duplet. Transient flipping of a unique base did not produce any binding in site I. Moreover, when base flipping resulted in Watson-Crick pairing interactions with the initiating strand, the base flipping events were locally accompanied by relocations of the dsDNA backbones (Figure 3.5). Relocation of the outgoing strand backbone was spatially limited but involved the transient loss of a salt bridge between the phosphate in 5’ of the flipped duplet and the corresponding R245. This salt bridge was eventually recovered after stabilization of the duplet binding in site I. Relocation of the complementary strand backbone involved large conformational changes at the duplet location, locally leading the backbone close to the position it occupies in the crystal structure of the site I-bound duplex DNA. It remained at that position until the end of the simulation (Figure 3.4B and Figure 3.5D). The flipping of duplets occurred iteratively in the 5’ to 3’ direction, starting from the site II-bound triplet closest to the kinked junction. The second duplet only started to flip once the first duplet was stably bound (Figure S2.8). Interestingly, the third base of each triplet did not flip due to steric hindrance from loop L2 (Figure 3.5A and Figure S2.4); it remained positioned in site II where it appeared to participate in a series of stacking interactions in the dsDNA together with the two bases of the outgoing strand that lost their pairing partners. It can
also be noticed that the outgoing strand partner of the 3′ base was looped out from the distorted helix in site II and firmly interacted with nearby R243. Such interactions were observed in all the triplets bound to site II and appeared to be a constant feature in the simulations, independently of the starting point. An important feature of the simulated base pairing exchange process is that the duplet that flipped towards the initiating strand remained confined in a hydrophobic environment throughout the homology testing: both the “top” base at the 5′ side and the “bottom” base in 3′ showed strong interactions with hydrophobic residues of the L2 loop, respectively with F203 and with M202, G200 and V201 (Figure 3.5A). The flipping movement can therefore be compared to sliding between two parallel air cushions, which greatly reduces the required energy when compared for example with the base flipping occurring due to thermal fluctuations. In the latter case, the flipping base not only needs to break its Watson-Crick hydrogen bonds but also loses most of its stacking interactions and gets immersed in an unfavorable aqueous environment. In the present case, the Watson-Crick bonds that needed to be broken were already greatly destabilized by the high level of stress in the site II dsDNA before the bases started to flip and the bases remained protected from water during flipping. These considerations explain why base flipping could be observed during our MD simulations only covering tens of nanoseconds, instead of ms for spontaneous flipping in isolated B-form DNA.
Figure 3.5: Trajectories for structures with 2 base pair triplets initially bound to site II. (A) I. Side view of the dsDNA. II. Rotated 90° view of I. III. Structure at a later simulation time when the duplet at the 5’ end has begun to flip. IV. Flipped duplet paired with the initiating strand. The flipping of the base at the 3’ end is sterically hindered by the L2 loop. (B) Base pairing parameters D1 and θ� are the C1’ and C1’ distance between complementary and initiating strands and C1’-C1’-N9 projection angle onto the base-flipping plane, respectively. The graphs show the histogram of D1 and θ� for a paired dsDNA in the stable postsynaptic filament (Figure S2.9). The bottom three histograms show results for a triplet bound in the site II. The left and middle one corresponds to the base-flipping process shown in (A). (C) Illustration of the separation of the outgoing strand
from the site II residues that occurs as the bases begin to flip. Top panel illustrates the initial time and bottom panel is 11 ns later when the bases are flipped, the outgoing strand is locally separated from the site II, and in the region near the flipped duplet the complementary strand is nearly in the postsynaptic conformation. Accelerated MD was carried out after 10 ns of conventional MD simulation. (D) Top diagram illustrates the backbone of the complementary progress toward the position of the postsynaptic filament. Bottom diagram shows the measured distance of the outgoing strand locally separating from site II.

3.5 Discussion

Factors supporting the starting structure

In this work, we have modeled intermediate structures in the early recognition and strand exchange processes of homologous recombination using several modeling techniques. We then evaluated starting structures using MD simulations of the evolution of the structures to determine whether they naturally lead to strand exchange processes that are compatible with homology testing. A rather complex process led us to the starting structures that we considered in the MD simulations. The unguided evolution of the starting structures with B-form dsDNA bound to lysines in the C-terminal domain that are linked to dsDNA bound to site II via kinked regions revealed both well-known structural features that were not incorporated in the model as well as new features that were unanticipated but provide insight into how the speed/stability stringency paradox can be overcome, as we will discuss in detail below.
The present work explicitly concerns the first rapid initial interactions that lead to homology testing; consequently, the speed of pairing exchange has been a major consideration when constructing the starting structure. During the construction of the starting structure, we privileged the conformations that would require as little protein movement as possible. Indeed, during the simulated base pairing exchange pathway, the protein structure almost did not vary, and the whole conformational change was concentrated on the DNA strands. Furthermore, the simulations showed that the very tense conformation of the dsDNA in site II facilitates base flipping by destabilizing its internal pairing and stacking interactions. Following its binding to site II, the dsDNA functioned like a spring whose tension increases with the number of added triplets, consistent with the observation of Danilowicz et al. [52] from single molecule pulling experiments. We propose that these intermediates correspond to the quick stage of recognition and lead to the metastable 8 nt strand exchange intermediate that has been experimentally detected [77]. In what follows we provide a detailed discussion of support for the major features of the structures shown in this work.

The most important support for the major features of the starting structures and their subsequent evolution is that the strand exchange occurred if duplets moved to their known postsynaptic positions, without any a priori guiding. That shifting included both base flipping and backbone relocation. Remarkably, base flipping and backbone relocation were reproduced at two levels of the filaments in successive triplets that had different structural environments, the first triplet being situated between the kinked junction in the 5′ direction and a site II-bound triplet in 3′ while the second triplet was situated between a partially flipped triplet in the 5′ direction and the dsDNA region that links the two site II
bound triplets to the B-form dsDNA at the 3’ end of the filament. Evidence for the flipping of a third successive duplet situated in that dsDNA linking region in 3’ was also observed, although this duplet was too close to the filament end for any end effect to be ruled out. This reproducibility suggests the simulated process relies on general mechanical properties of the distorted double helix in the starting structure that drive evolution toward the known crystal structure of the dsDNA in the postsynaptic filament.

Further support for the starting structures is provided by the sensitivity of the evolution to features of the starting structures where the sensitivity is in good agreement with known features of the RecA system. For example, the simulations suggest that the stability of the base pairing decreased with the number of triplets bound to site II, which is consistent with base tension increasing with the number of bound triplets as a result of the differential extension between the outgoing and complementary strand backbones [52]. Furthermore, B-form dsDNA ends are present in vivo, and the simulations suggest that they are required in order for the local evolution to the postsynaptic conformation. In addition, even structures with B-form tails did not show strand exchange unless there were exactly two triplets bound in site II. In that structure 8 bp are positioned to rapidly undergo strand exchange without significantly disrupting the protein structure or the structure of the dsDNA that is not bound to site II. Thus, an initial rapid test of 8 bp emerges naturally from the sensitivity of the evolution to the number of base pairs bound to site II in structures with B-form tails.

Additional support for the starting structure comes from a direct comparison of the structure itself with structural features of the RecA system that are known from experiment, but were not explicitly included in the modeling. First, the distance between the DNA
strands and the helix axis agrees with the experimentally measured distance using fluorescence energy transfer by the Singleton group (Xiao and Singleton 2002) even though attribution of the strand identities in their work differs from what we propose in the present model. Second, the model also features a sharp change in the DNA direction, consistent with observations of Rad51 strand exchange intermediates [113]. Third, the spontaneous base flipping shown in the simulations is consistent with previous suggestions that base flipping is part of the recognition process [73, 74]. Fourth, the relocation of the complementary backbone is consistent with FRET experiments showing that backbone relocation occurs during strand exchange [49]. Fifth, the flipping of a duplet that occurred while the third base from the initial triplet remained in nearly the original location is consistent with FRET data indicating that some complementary strand bases remained in the initial positions, after other complementary strand bases had already paired with the incoming strand [49, 74]. All of these known structural features of the dsDNA emerged naturally from unguided evolution of the starting structures.

The starting structures are also supported by comparing the roles played by protein residues during the simulations and the functional roles of the residues that are known from previous studies. For example, previous experiments suggested that the “gateway” to homology recognition is the CTD surface of RecA subunits [115, 116]. Specifically, Shibata and colleagues showed that CTDs, and particularly exposed regions on the CTDs bearing a cluster of lysine residues, bind dsDNA [116]. Among the CTD lysine residues, K286 and K302 are particularly well-conserved in bacterial RecA protein, and mutations of these residues impede strand exchange without significantly suppressing active filament formation. Mutations of less conserved lysine residues K280 and K282 also showed some
suppression of homologous pairing activity [115]. In addition, the intercalation of F203 into the rises in dsDNA triplets bound to site II is consistent with previous studies. In particular, proteins show some functionality when F203 is replaced by tyrosine or tryptophan [117] and conserved residue studies showed that in 8 strains F203 was replaced by the functionally similar double ring structure [118]. Finally, the dsDNA binding interactions with site II residues contribute to the untwisting of the dsDNA [119]. As explained above, this information on residues function could not be used directly due to the profusion of the identified residues along the whole filament and the incertitude about the sequence of events in which they are involved; however, the good agreement between the modeling and the residue studies provides additional support for the modeling, and our model proposes a plausible scenario for the time course of their involvement.

**Implications of the starting structures**

As discussed below, our simulations also brought new and unexpected features that not only provide structural interpretations for many observations formerly published on the system, but also offer a solution to the speed/stability/stringency paradox as it applies to searches over a bacterial genome. The most important result that had not been anticipated before the simulations was the observation that the bases flip in duplets. Though the base pairs in both site I and site II are divided into triplets, the L2 loop sterically hinders the flipping of a complete triplet, but a duplet can easily flip. Thus, the observation that the bases do not flip in triplets does not depend on the evolution of the MD simulations, it is a basic feature of the starting structures that were obtained using several independent
modeling techniques. The MD simulations merely confirmed that the duplets can indeed flip.

In addition, even without MD simulations, the starting structures suggest that during the initial stage of recognition, the dsDNA binding to the filament is dominated by interactions with dsDNA backbone phosphates. Furthermore, the B-form regions maintain the interaction between the complementary and outgoing strands even if there is no base pairing between those strands in the homology testing region. Thus, during the initial homology testing stage the nucleoprotein filament forces all three strands to remain in positions where strand exchange or reverse strand exchange can occur. Finally, base pairing in the initial strand exchange conformation can be fairly weak even for homologous pairings, while still allowing homologous pairings to remain bound long enough to progress with high probability to an irreversible strand exchange product. The influence of the B-form dsDNA ends had not been anticipated in previous works.

**Implications of the MD simulations**

Other features of the model emerged only from the MD simulations. For example, the large backbone distortion that results from the flipping of a base pair duplet had not been anticipated previous to the simulations. We note that the simulations suggest that the base pairing of the flipped duplets is much less stable than the base pairing characteristic of complete triplets in the postsynaptic conformation [80]. This difference in stability is consistent with homology testing progressing much faster in transition states containing flipped duplets than in conformations containing complete triplets in the postsynaptic
conformation. The MD simulations provided a possible structural basis for the large base pairing instability that must characterize interactions during the initial rapid testing stage.

The sequential flipping of successive duplets is another feature that had not been anticipated. Importantly, it eliminates the large entropic barrier to strand exchange that would be present if 8 bases are allowed to flip individually in any order [85]. Flipping in duplets may produce additional advantages since previous work has suggested that the presence of a mismatch will destabilize the pairing of neighboring bases [120]. Thus, the sequential duplet flipping revealed by the simulations offers significant speed and stringency advantages over other testing strategies.

The MD simulations also provide a possible mechanism for the increase in stability that occurs once 8 contiguous base pairs have undergone strand exchange. The MD simulations suggest that three successive duplets rapidly flip, leaving two unflipped bases separating the flipped duplets, as illustrated in Figure 3.6. If the two unflipped bases would flip in turn, then eight contiguous bp would assume the postsynaptic filament conformation and occupy the post strand exchange conformation, which would greatly reduce the mechanical stress on the complementary strand backbone, as illustrated in Figure 3.6. In addition, given that flipped duplets are not very stable, but structures with complete triplets in the postsynaptic conformation are highly stable [80] the flipping of the two unflipped bases would be expected to produce a great increase in the base pairing stability. None of the simulations ran long enough to observe that flipping. We anticipate that this may require some noticeable conformational change within the system, such as loop L2 displacement. Thus, results of the MD simulations are consistent with initial homology testing rapidly occurring in structures with 8 bp in a position to undergo strand exchange,
where passing that homology test would lead to a transition to a much more stable strand exchange conformation.

**Figure 3.6:** Proposed transition to the more stable configuration. The outgoing, complementary, and initiating strands are shown in cyan, purple, and orange respectively. (A) I. Schematic of the base pairing in a structure with 3 successive strand exchanged duplets. The red region indicates the position of the site II residues. (A) II. A structure showing one flipped duplet in nearly the postsynaptic conformation, followed by an unflipped duplet, followed by a duplet beginning to flip. The resulting distortion of the complementary strand backbone is visible. (B) I. Schematic of a conformation with 2 complete triplets in the very stable postsynaptic conformation. (B) II. Same as II. in (A), but with all of the bases in the postsynaptic conformation.
Finally, the MD simulations show that in those structures 8 bp can undergo strand exchange while preserving the binding of B-form dsDNA at the two ends. A return to B-form dsDNA could still maintain the binding of the dsDNA to the C-terminal domains. Given that the C-terminal domains are flexible (Figure S2.10), additional homology testing in a different registration could occur without changing the binding at the C-terminal domain. This would allow testing of ~ 4 different registrations. Simulations also suggest that binding can be transferred from domain to domain. Such transfer of dsDNA binding between C-terminal domains might be consistent with experimental results described as “sliding” [120]. These two mechanisms for testing multiple registrations without free diffusion would make homology testing much more rapid than the case where the dsDNA must freely diffuse before another registration could be tested. The detailed mechanism for testing without diffusion was another unanticipated result that emerged from the model being presented here. Speculations on possible mechanisms for simultaneously testing multiple sites [121] (see also Appendix 2) or for rejecting base pairings involving repeated sequences extending over more than 100 bp are presented in Appendix 2.

In conclusion, the present work offers the particularity to explicitly consider the intrinsically multiscale character of the HR reaction by integrating detailed structural and functional information obtained via modeling approaches with physical models of homology search (Figure 3.7). This is an unusual strategy, but we propose that the high level of integration between the different aspects of the recognition puzzle is a key element for successfully studying a complex biological system that evolves on various timescales. The features of the model are in good agreement with previous experimental results. All new features obtained during the construction of the intermediate species or revealed by
molecular dynamics simulations were systematically tested and/or incorporated in the theoretical model of recognition, leading to further improvements of this model. We are conscious that this body of evidence is not a direct proof that the proposed model elucidates the structural details of the mechanism of DNA sequence recognition and strand exchange. Nevertheless, we believe that owing to the intrinsic multi-scale and dynamic character of the process, our proposed model represents a formidable template to tailor new experiments that will finally establish the mechanism.
Figure 3.7: Overview of the proposed searching process. (A). Illustration of the known homology recognition process. The initiating, complementary, and outgoing strands are shown in orange, purple and cyan, respectively. (B) Overall structural transition during homology testing including the known structures (I and IV) and structures obtained using modeling and simulations (II and III). We propose that the structures correspond to the illustration shown above them in (A). The upper and central panels of (B) show end views of the filament, while the bottom panels show side views. The residues in the C-terminal domain are shown in pink except for lysine residues K280, K282, K286, and K302 which are shown in green. K232 is shown in silver. Site II residues R226, R227, R243, and K245 are shown in red, and the L2 loop (198-206) is shown in yellow.
Chapter 4

Structural Investigation of the von Willebrand Disease A2 R1597W mutation by Molecular Dynamics Simulation

4.1 Introduction

von Willebrand Factor (vWF) is an essential protein in hemostasis. During arterial bleeding, vWF mediates the adhesion and aggregation of platelets at the site of vascular injury to form a hemostatic platelet plug [122, 123]. As a multimeric glycoprotein, vWF is made of 240 kDa monomers concatenated through specific disulfide bonds at the N and C termini; each monomer has multiple domains with specific functions. In particular, vWF has three distinct “A” domains: A1, A2, and A3. The A1 and A3 domains have adhesive functions, with the A1 domain binding to the platelet glycoprotein-Ibα receptor (GPIbα), and the A3 domain binding to collagen of the exposed subendothelium at the injury site. Under normal flow, the vWF A domains are mostly buried and inaccessible for binding,
but at hydrodynamic stress rates above a critical value, the globular multimeric vWF transitions to an elongated conformation, exposing the necessary binding domains and activating its adhesive function [124, 125].

The primary function of A2 domain is not adhesion, but to regulate the vWF size distribution using hydrodynamic stress. Endothelial cells secrete ultra-large vWF strings anchored to the cell surface that extended to 100 to 1,000 μm long in present of flow [126]. The A2 domain can be unfolded by tensile force exerted along vWF by vascular flow, exposing the cryptic Y1605-M1606 scissile bond (Figure 4.1A and C) that is cleavable by metalloproteinase ADAMTS13. Cleavage releases the vWF strings into the circulation [126-128]. The vWF strings in the circulation are cut into multimeric units of 500 kD to >20,000 kDa through A2 cleavage. The hemostatic potential of vWF is related to the size of the multimers, such that the large vWF multimers have higher adhesion activity than the smaller multimers. Hence, excess of ultra-large vWF (UL-vWF) can contribute to thrombotic disorders [129, 130], whereas absence of the large multimers causes bleeding diathesis [122]. The balance between synthesis and force-induced ADAMTS13 proteolysis is essential for maintaining a proper size distribution of vWF with normal hemostatic potential [129]. Mutations in vWF can disrupt this balance and lead to von Willebrand disease (vWD) [131], an inheritable bleeding disorder. vWD type 2A is categorized by a lack of large multimers in the circulation, leading to increased bleeding in vascular injury. Here, we investigated the structural mechanisms underlying most common vWD Type 2A: A2 domain R1597W point mutation [132]. Previous studies demonstrated that the R1597W mutation increases the A2 domain’s susceptibility to ADAMTS13 cleavage [133], which causes size reduction and rapid clearance of large vWF in
circulation.

**Figure 4.1.** (A) Crystal structure the calcium-bound vWF A2 domain with the calcium ion shown as orange sphere. The Y1605-M1606 scissile bond located in the β4 strand is highlighted in purple. The R1597 residue located in the α3-β4 loop is shown in green. (B) Secondary structure arrangement of the A2 domain from N to C termini. The β-sheet and α-helix structure is shown in yellow and red, respectively. (C) In the folded structure, the β4 is sandwiched between the β1 and β5 and buried in the core of the protein making the Y1605-M1606 scissile bond cryptic. (D) Residue R1597 forms salt-bridge interaction with residue D1498, L1497 and S1534 and stacks with H1536 in the crystal structure. (E) Calcium ion in the calcium binding loop. The calcium interacts with six ligands to form octahedral coordination geometry, including R1597 residue and a water molecule.

A recent crystal structure showed that R1597 resides in the α3-β4 loop and participates in coordinating a calcium binding site (Figure 4.1 A, D and E) [134, 135]. The
presence of a bound calcium ion increases the stability of the wild-type A2, reducing the susceptibility of ADAMTS13 cleavage. Whereas a force-spectroscopy study also showed that R1597W mutation destabilizes the A2 domain [136], which can contribute to the increase of vWF cleavage. These results complement the bulk data measuring A2 stability and ADAMTS13 cleavage rates, which described the regulation of full length vWF size distribution [137] and early single-molecule force spectroscopy studies that demonstrated that force-induced unfolding of the A2 domain activates ADAMTS13 cleavage, and suggested that hydrodynamics stress can induce sufficient force to unfold A2 [127]. In spite of these results, there is no crystal structure of the A2 R1597W mutant, so there is limited structural insight at the molecular level on how the R1597W mutation destabilizes the A2 domain and the role of the calcium ion in physiological function. Here, we carried out extensive molecular dynamics (MD) simulations in an effort to provide these molecular insights at the atomistic level.

Molecular dynamics simulations have been applied to vWF research to explore various structural and function relationships. Early MD simulation studies of the A2 domain, without the availability crystal structure, were carried out using a computational homology model as the starting structure [138]. In the first study, the modeled A2 structure was constructed using six protein templates, and the short 10 ns MD simulation investigated structural differences between the wild-type and mutant associated with the type 2A von Willebrand disease [138]. Later, constant-velocity steered molecular dynamics (SMD) techniques were used to evaluate the unfolding force and pathway of A2 in a partially solvated [139] and fully solvated [140] modeled A2. When comparing this modeled A2 structure with the solved crystallographic structure [128], it was observed that
they were overall similar, but with significant differences in the α3-β4 loop region[141]. Recent simulations of the A2 domain show L1657I, I1628T, and E1638K mutations destabilize the α6 of A2, which can leads to type 2A VWD [142]. The crystal structure of calcium-bound A2 structure was reported with molecular dynamics analysis as well, illustrating that the calcium ion can stabilized the A2 domain [134, 135]. Other additional studies include the characterization of the A1-A2 protein-protein interaction that potentially inhibits the vWF binding to GPIbα [143], and various aspects of A1 and GPIbα binding [143-145].

In this study, we evaluate the effect of the R1597W mutation and the binding of calcium ion on the stability of A2 using explicit solvation molecular dynamics (MD) and constant velocity pulling steered molecular dynamics (SMD) simulations. We simulated four different systems: A2 wild type (WT), calcium-bound A2 wild type (WT+Ca2+), A2 R1597W mutant (R1597W), and calcium-bound A2 R1597W mutant (R1597W+Ca2+), each with more than 500 ns simulation for a total cumulative of > 2 μs. The SMD pulling velocity is 0.15 nm ns−1 with three repeats, one order magnitude slower than the current convention for system of this size. We compare the structural, energetic and stability differences between the systems to understand how R1597W affects A2 stability at the molecular level, and show the structural mechanisms behind experimental observations.
4.2 Method

**Molecular Dynamics Simulation**

Molecular dynamics (MD) simulations were performed using the NAMD 2.9 package [105] with the CHARMM 27 force field including the CMAP correction [106]. The starting crystal structures of wild type A2 domain with and without calcium were obtained from the PDB database with ID’s 3GXB [128] and 3ZQA [135], respectively. The starting structures of the mutant were made by modifying the side chain of arginine residue R1597 to tryptophan (W) using the Mutator plug-in in VMD [146] as the crystal structure wasn’t available. A short 100-step energy minimization was carried out with only the mutated R1597W residue flexible after the modification. We simulated four systems including wild-type (WT), wild-type with calcium bound (WT+Ca$^{2+}$), R1597W mutant (R1597W) and R1597W mutant with calcium bound (R1597W+Ca$^{2+}$).

**Equilibrium Molecular Dynamics Simulation**

For each simulation, a periodic boundary box was defined with the boundary faces 12 Å away from the protein surface. We used TIP3P water model to solvate the protein, and a physiological concentration of Na$^+$ and Cl$^-$ ions of 0.15 mol/L were added to each simulation to maintain electro-neutrality (Figure 4.2A). The solvated and ionized simulation box contained ~15,000 atoms. We used 2-fs time steps, and the hydrogen atom bond lengths were constrained using the SHAKE method. We utilized a particle-mesh Ewald method for long-range electrostatics calculations, and van der Waals interactions
were smoothly switched off at 10–12 Å by a force-switching function. The temperature and pressure were maintained along the simulation using a Langevin dynamics scheme and a Nosé-Hoover-Langevin piston, respectively.

Simulations were energy minimized using 1,000 conjugate gradient energy minimization steps followed by increasing the temperature from 0 to 300 K over 500 ps. During the energy minimization and heating process, the $C_\alpha$ atoms of the protein were restrained using a soft harmonic potential with a spring constant of 0.5 kcal/mol-A$^2$. The system was further allowed to equilibrate for 5 ns in the NPT ensemble with pressure of 1 bar and temperature of 300 K. For the production simulation, harmonic potential constant was turned off and the systems were maintained under NPT ensemble conditions with pressure of 1 bar and temperature of 300 K. The production duration for each of the four systems was 500 ns for a cumulative total of 2 μs. The root mean square deviation (RMSD) of the $C_\alpha$ atoms and the energies during the heating, equilibration and production of the simulation are shown in Figure 4.2B.

**Interaction Energy Calculation**

We used the Molecular Mechanics-Poisson–Boltzmann Surface Area (MM-PBSA) method [147] to compute the binding energy ($\Delta G_{\text{bind}}$) of the residue 1597 side-chain or the calcium ion with the rest of the A2 domain. The MM-PBSA calculation was carried out using g_mmpbsa plugin tool [148] implemented through the GROMACS [149] and APBS [150] software. According to the method $\Delta G_{\text{bind}}$ is computed as $\Delta G_{\text{bind}} = \Delta E_{\text{MM}^+}$
\[ \Delta G_{\text{sol}} - T\Delta S, \] where \( \Delta E_{\text{MM}} \) is the change in the molecular mechanics energy upon binding that consists of contributions from van der Waals (\( \Delta E_{\text{vdW}} \)), electrostatic (\( \Delta E_{\text{ele}} \)), internal bonded interactions (\( \Delta E_{\text{int}} \)), and \( \Delta S \) is the change in the configurational entropy upon binding (\( T \) is the temperature). \( \Delta G_{\text{sol}} \) is the change in the solvation free energy upon binding that was computed using mean field approaches as a sum of polar (\( \Delta G_{\text{sol-pol}} \)) and non-polar contributions (\( \Delta G_{\text{sol-np}} \)). \( \Delta G_{\text{sol-pol}} \) was computed using the Poisson–Boltzmann approach with the solute and solvent dielectric constants are set to 1 and 80, respectively. \( \Delta G_{\text{sol-np}} \) was computed using the solvent accessible surface area (SASA) as \( \gamma \Delta \text{SASA} \), where \( \gamma \) is the surface tension. The water probe radius and the surface tension are set to 1.4 Å and 0.0072 kcal mol\(^{-1}\) Å\(^{-2}\), respectively. The net polar and non-polar contributions to \( \Delta G_{\text{bind}} \) are then given by \( \Delta G_{\text{polar}} = \Delta G_{\text{sol-pol}} + \Delta E_{\text{ele}} \) and \( \Delta G_{\text{np}} = \Delta G_{\text{sol-np}} + \Delta E_{\text{vdW}} \), respectively. In this study we only consider the enthalpy contribution and not the configurational entropic contribution (\( T\Delta S \)) due to its high computational cost.

**Steered Molecular Dynamics**

To study the force induced unfolding pathway of the A2 domain, we carried out constant velocity steered molecular dynamics simulation (SMD) to unfold the protein by pulling on the C-terminal. We started with the 5-ns equilibrated protein simulations and then added an additional water layer in the pulling direction. We constrained the Cα atom position of the N-terminus, and the Cα of the C-terminus was attached to a spring with a stiffness of 1 kcal/mol-Å\(^2\) that was pulled at a constant velocity. The applied tensile force during the pulling simulation was calculated using the Hook’s law defined by \( F = k \Delta x \), where \( k \) is the stiffness of the attached spring and \( \Delta x \) is the spring extension.
Here, we carried out two different constant velocity pulling simulation protocols for each of the four systems. In the first protocol, the protein C-terminus was pulled at a constant velocity of 0.5 nm/ns. In the second protocol, the pulling velocity was 0.15 nm/ns with three repeats for each system. The SMD simulation was terminated when the Y1605-M1606 cleavage site and the $\beta_4$ were fully exposed to solvent.

4.3 Result

Structural and Dynamics Analysis

We performed explicit solvent MD simulation of the A2 WT and R1597W mutant with and without bound-calcium ion (see Methods and Figure 4.2A for schematic) to investigate the structural and dynamical differences in the folded state. The wild-type starting structures were obtained from the crystal structure, and the mutants were created by modifying the side-chain of the R1597 residue to tryptophan (see Methods).

We measured the backbone C$_\alpha$ atom root mean square deviation (RMSD) as a function of time (Figure 4.2B) to examine the simulation stability and backbone deviation from the wild-type crystal structure. The average RMSD of the wild-type with or without calcium during the production simulation (last 400 ns) is $\sim$ 1.6 Å. For the R1597W mutant, the average RMSD is 19% higher ($\sim$ 1.9 Å), indicating some backbone deviation from the wild-type (Figure 4.2B). The backbone deviation transitions occurred at approximately 25 ns and 110 ns for the R1597W and R1597W+Ca$^{2+}$, respectively. For the wild-type, several smaller and distinguishable RMSD transitions can be observed as well. The time-scale of
the backbone transitions that we observed was on the order of $10^2$ ns, and the 500 ns simulation time for each system is the longest continuous A2 simulation that has been conducted.

**Figure 4.2:** (A) Schematic of the simulation. The backbone of the A2 is shown as an orange cartoon. The solvation water is shown as a transparent box. The solvation sodium and chloride ions are shown as yellow and blue cyan, respectively. (B) Root mean square deviation (RMSD) of the $C_\alpha$ atoms relative to the respective wild-type structure. The overlapping dash line is the average RMSD of the last 400 ns. For the R1597W and R1597W+Ca$^{2+}$ traces, backbone deviation transitions are indicated by arrows at approximately 25 and 110 ns, respectively.

We evaluated the secondary structure at the residue level (Figure S3.1), with Figure 4.3 summarizing the propensity and differences between each system. The $\alpha$-helixes and $\beta$-sheets observed from the crystal structure were stable during the simulation in all
systems. Structure changes only occurred near the mutation and calcium binding regions. First, the R1597W mutation destabilized the small $\alpha$-helical span (residue 1597 to 1599) in the $\alpha_3$-$\beta_4$ loop (Figure 4.3B), causing an RMSD jump at the 25 ns time point (Figure 4.2B and Figure S3.1). The R1597W mutation increased the length $\alpha_3$ helix of from 10 to 15 residues (residue 1578 to 1592) compared to the WT, but this increase was diminished in the presence of the calcium (Figure 4.3B). The presence of the calcium ion reduced the $\alpha$-helical region in the $\alpha_3$-$\beta_4$ loop of the WT as well. The molecular interaction that leads to these structural differences is illustrated in Figure 4.4, which will be discussed later in detail.
Figure 4.3: (A) Residue secondary structure propensity during the last 400 ns of simulation. (B) Change in the secondary structure propensity as function of residue between the different systems. (C) Residue root mean square fluctuation (RMSF) during the last 400 ns. Data plotted in black line is the B-factor from the wild-type A2 crystal. (D) Change in the RMSF between the different systems. The green vertical line indicates the location of the residue 1597.

We calculated the $C_\alpha$ root mean square fluctuation (RMSF) for each residue to characterize the mobility of different regions (Figure 4.3C). The $\alpha$-helix residues ($\alpha_1$ to $\alpha_6$) and $\beta$-sheet ($\beta_1$ to $\beta_5$) residues were stable with an RMSF value of $< 1$ Å; the $\alpha_4$-less and $\alpha_3$-$\beta_4$ loops were the two regions that exhibited the highest mobility. The R1597W mutation further increased the mobility of the $\alpha_3$-$\beta_4$ loop, whereas the presence of calcium ion stabilized it (Figure 4.3D). The RMSF from the simulation shows reasonable agreement with the B-factor of the crystallography data (Figure 4.3C). However, the mobility of $\alpha_4$-less and $\alpha_3$-$\beta_4$ loop in the simulation was significant higher than the estimation based on the crystallography Beta-factor data (Figure S3.2). However, the intermolecular contacts between crystallographic units [128] could potentially constrain the $\alpha_4$-less and $\alpha_3$-$\beta_4$ loops flexibility [142].

Figure 4.4 shows the representative snapshots of the A2 structure to highlight the differences at the molecular level. In the early part of the WT simulation, the R1597 stacked with H1536 and formed hydrogen-bonds with L1497, S1593, S1534 and D1498 (Figure 4.4A-i). However, this network of interactions, as observed in the crystal structure,
wasn’t maintained throughout the simulation; the R1597 could also stack with F1552 and form hydrogen bonds with D1497 and F1552, respectively (Figure 4.4A-ii). Those specific interactions involving the R1597 helped form and stabilize the small helical region (residue 1597 to 1599) within the α3-β4 loop. We found that the R1597W mutation disrupted the native interactions of R1597. The mutant W1597 residue side-chain rotated to a different orientation after 25 ns, buried the hydrophobic side-chain, and transiently stacked with F1590 (Figure 4.4A-ii). This different orientation of W1597 residue, without the wild-type native interactions, eliminated the helical structure within the α3-β4 loop, increased the mobility of the region, and extended the helical content of α3 (Figure 4.3B and D).

In the case of wild-type with calcium bound (WT+Ca^{2+}), the overall structure and interaction involving R1597 and calcium observed in the crystal structure were stable throughout the 500 ns simulation (Figure 4.4C). Specifically, the calcium ion formed octahedral coordinated interactions with the backbone carbonyl of R1597 and A1600, the side-chain carboxyl of D1498 and D1596, the side-chain carbonyl of Asn1602, and an interchangeable water molecule. With the R1597W mutation, the calcium ion maintained octahedral coordination interaction only during the first 110 ns (Figure 4.4D-i). Afterward, the carbonyl of R1597 and A1600 detached from the calcium octahedral coordination geometry within 5 ns of each other (Figure 4.4D-ii), leading to a more flexible α3-β4 loop as shown in the RMSF (Figure 4.3B). This result highlighted that binding of the calcium in the octahedral coordinated form requires the R1597 residue to have the specific network of interaction as observed in the crystal structure.
Figure 4.4: Representative snapshots of the α3-β4 loop (blue), 1597 residue (green) and calcium ion (orange) of the four different systems. (A-i) Interaction of the WT R1597 residue with neighboring residues observed in the crystal structure. (A-ii) WT observed during the production simulation where the network of interaction varied from the crystal structure. (B) Structure of the R1597W mutant with the 1597 residue rotated to a different orientation and stacking with F1590. (C) WT A2 with the bound calcium ion. The interactions involving R1597 and calcium ion observed from the crystal structure were stable in the simulation. (D) Representative structure of the R1597W A2 mutant with the bound calcium ion before and after 110 ns is shown in (D-i) and (D-ii), respectively. (D-ii) C=O bond of the A1600 and R1597W residues detached from binding to the calcium ion after 100 ns.
**Interaction Energy Analysis**

We evaluated the interaction energy of the 1597 residue or the calcium with the rest of the A2 to estimate the relative stability differences between each system in the folded state. Here, we utilized the molecular mechanical and continuum solvent approach (MM-PBSA) algorithm to characterize polar and apolar interaction energies (see Methods). Due to high computational cost and slow convergence, we did not include the conformational entropy contribution in the energy calculation.

The interaction energy between the R1597 with the rest of the A2 protein is enthalpically favorable through polar interactions (Figure 4.5A). Compared to the mutant, the hydrophobic W1597 diminished nearly all the polar interactions and only marginally increased the apolar component, which led to a net loss of 75% in the interaction energy. Such significant change in the interaction energy suggested that the mutation destabilizes the A2 native state, which is consistent with the experimental observation that R1597W mutation reduced the melting temperature from 56.7 to 51.5 °C [136].

Similar to the R1597 residue, the calcium ion bound to the wild-type A2 protein through favorable polar interactions, consistent with the experimental results showing that melting temperature increase from 56.7 to 67.7 °C in the presence of calcium [151]. For the mutant, it was first anticipated that the calcium would bind more weakly to the A2 R1597W mutant since the mutation alters the structure of the calcium binding loop (Figure 4.4D-ii). Yet, the calculation shows that the mutation does not effect the binding energy, as the tryptophan eliminated the positive-positive electrostatic repulsion between the arginine and calcium. Even though the enthalpy binding energy of the mutation was
favorable, it is entropically unfavorable as calcium binding reduces the dynamical range of the $\alpha_3$-$\beta_4$ loop can explore as seen from the RMSF calculation (Figure 4.3D) providing an explanation for the marginal increase of melting temperature from 51.5 to 53.9 °C in present of calcium [136].

Figure 4.5: Binding energy analysis using MM-PBSA method. (A) Average binding energy between residue 1597 and the rest of the A2 domain calculated from the last 400 ns. The total binding energy is the sum of the polar and apolar contributions. The wild-type (WT) and the mutant (R1597W) residue are shown in blue and orange respectively. The error bars are the standard deviation. (B) Binding energy of the calcium ion Ca$^{2+}$ to the A2 domain for the wild-type and mutant are shown in gold and purple, respectively.
**Force-induced A2 Unfolding Simulations**

To characterize the stability of A2 under force, we carried out constant-velocity steered molecular dynamics (SMD) simulations for each system and compared the unfolding tensile force and structural transitions during the unfolding process. In full-length vWF, tensile force on an individual molecule pulls on the N- and C- termini of the A2 domain. To mimic this in SMD, we fixed the N-terminus of the molecule and applied force to the C-terminus. In the first set of SMD simulations with pulling rates of 0.5 nm/ns, all the A2 systems began unfolding at the C-terminus. Because the N-terminal $\beta_1$ is buried within the core of the protein, the N-terminal region maintained the majority of the native contacts and secondary structure prior to $\beta_4$ (Figure 4.1, Figures S3.2). This progression agrees with previous steered MD pulling studies of A2 domain that either used a higher pulling velocity [134, 140] or fixed the C-terminus (as opposed to fixing the N-terminus as in our study) [139, 142].

Before the cleavage site was exposed, there were three rupture peaks in the force profile that corresponded to unfolding of $\alpha_6$, $\beta_6$ then $\beta_5$ (Figure S3.3). The first peak was the undocking (pulling away from the rest of the structure) of $\alpha_6$ from A2. This peak corresponded to a relatively smaller rupture force than the others, indicating a weaker interaction. There are mutations in the $\alpha_6$ that are known to cause VWD, including E1638K, L1671I and I1628T. Previous studies have indicated that these mutations lower the undocking force of $\alpha_6$ from the rest protein by destabilizing the packing of tertiary structure [142]. Undocking the $\beta_6$ required shear-rupturing of hydrogen bonds, requiring a higher force than $\alpha_6$. Subsequent to the rupturing of the $\beta_6$, the $\alpha_5$ region undocked from
the A2 without any significant force required. Next, the β5 region also resisted unfolding in a shear configuration and the subsequent α4-less region undocked without resistance as well. The unfolding pathway of α6-β6-α5-β5-α4-less was similar across all four systems (Figure 4.6 and S3.3), except the rupture force of the β5 in the R1597W+CA2+ system was lower (Figure 4.6D and S3.3D). Because interactions with R1597W were not observed to commence this event, the difference may just be due to intrinsic thermal variation.

The unfolding of β4, which fully exposed the cleavage site, exhibited the highest rupture force. The hydrogen bounds between the β4-β1 sheets attributed to the high resistance to unfolding, with additional contributions from the upstream α3-β4 loop region that contains the 1597 residue and the bound-calcium ion. In the WT case, the hydrogen bond between R1597 and D1498 resisted the unfolding (Figure 6A and 6C panel i). The interaction between D1498 and D1597 bridged by the calcium ion also led to an additional rupture peak followed by the unfolding of β4 (Figure 6C and 6D, panel ii). In a previous study using SMD that utilized the OPLS all-atom force field with a pulling velocity of 1.25 nm/ns, the presence of calcium led to three rupture peaks after the unfolding of β4, representing the sequential breakage of the coordinated residues with the calcium ion [134] whereas we only observed one. This is likely due to differences in the molecular force fields or SMD pulling velocities.

Comparing the peak force that corresponds to the exposure of the cleavage site in each simulation, WT+Ca2+ had the highest rupture force with 1033 pN follow by R1597W (983 pN), WT (868 pN) and R1597W+Ca2+ (776 pN). Here, the presence of calcium ion increased the peak force of the WT as expected based on previous experimental
observation. Yet, the R1597W system had a higher peak force than WT, and the presence of calcium lowered the force in the R1597W+Ca²⁺ system, with both predictions qualitatively inconsistent with experimental observation. This discrepancy suggests that SMD simulation with multiple repeats and with slower pulling rates are required for accurate and statistically significant results in order to compare with experimental observation.

![Figure 4.6](image.png)

**Figure 4.6:** Tensile force profile of constant velocity pulling unfolding simulation with a 0.5 nm/ns pulling rate for the (A) WT, (B) R1597W, (C) WT+Ca²⁺, and (D) R1597W+Ca²⁺ systems. (A-i) and (B-i) show the structure of A2 at the peak force of β₄ rupture. R1597 residue resisted unfolding of β₄ by interacting with D1498. (A-ii) and (B-ii) show the structures of the peeling of β₄ and undocking of α₃ after rupture. (C-i) and (D-i) show the structures of A2 at the force peak of β₄ rupture, and how the bound calcium ion resists unfolding. (C-ii) and (D-ii) show the bridge interaction of D1498-Ca²⁺-D1596 that led to an additional rupture transition after unfolding of β₄.
We reduced the pulling velocity by a factor ~ 3.3 to 0.15 nm/ns and repeated the unfolding simulation three times for each system (Figure 4.7). At this pulling rate, the unfolding of β4 still resulted in the highest rupture force, and the presence of the calcium ion also yielded an additional rupture transition after unfolding of β4 that corresponds to breakage of the D1498-Ca²⁺-D1597 interaction (Figure 4.7A). In terms of averages, WT+Ca²⁺ had the highest peak rupture force at 1254 pN follow by WT (917 pN), R1597W+Ca²⁺ (770 pN) then R1597W (764 pN). This mechanical stability trend is positively correlated with the experimental data on the thermal stability of the A2 (r = 0.87) (Figure 4.7B) [136]. Furthermore, WT+Ca²⁺ had significantly higher rupture force as compared to all other systems (p < 0.05). The R1597W mutation did not significantly decrease the rupture force compared to the WT (p = 0.13), and the presence of calcium ion did not significantly increase the stability of the R1597W mutant (p=0.97).
4.4 Discussion

In this study, we investigated the molecular mechanisms of VWD caused by the A2 R1579W mutation. We first conducted extensive MD simulations to evaluate structural and dynamic differences between the wild-type and mutant. From our simulation, we observed R1597 residue engaged in two modes of specific interaction with the neighboring residues that stabilized the α3-β4 loop (Figure 4.4A). Even though the R1597W mutation

**Figure 4.7:** Tensile force profile of constant velocity pulling unfolding simulation with 0.15 nm/ns pulling rate. (A) Each panel corresponds to a system. The tensile profiles are overlaid with different shadings. The rupture of β4 resulted the highest peak force in the profile in all system. (B) The top panel shows the average and standard deviation of the peak rupture force in each system. The bottom panel shows the averages compare to the melting temperature of each system. The grey line is a linear fit of the data that leads to correlation coefficient of $r = 0.87$. 
did not significantly alter the overall structure in our simulation, structural changes near the mutation site were predicted (Figure 4.3). The R1597W mutation eliminated the helical structure within the $\alpha_3$-$\beta_4$ loop while increasing the helical span of $\alpha_3$, which may explain the previous observation that the mutation did not change the A2 circular dichroic spectra [136]. The binding of calcium ion also reduced the helix within the $\alpha_3$-$\beta_4$ loop. Due to lack of specific interactions with the R1597W residue, the $\alpha_3$-$\beta_4$ loop became highly flexible and coordination between the calcium ion and residues 1597 and 1600 was disrupted (Figure 4.4D).

We also focused on the impact of the R1597W mutation on the stability of A2. We first determined whether the R1597W mutation is energetically favorable (Figure 4.5). The R1597W mutation diminished the favorable polar contribution of R1597 in the native state leading to a less stable state, which explains the previously measured thermal destabilization after mutation [136]. Energy calculations also showed that the calcium ion binds to the wild-type with a favorable polar interaction, bringing A2 to a lower energy state, consistent with the significant increase in melting temperature in the presence of calcium [135, 151]. The R1597W mutation was experimentally shown to disrupt the calcium stabilization such that the melting temperature did not significantly increase in the presence of calcium [136], and our calculation also suggests the same. Specifically, even though calcium binds enthalpically favorably to the A2 mutant, the binding of calcium limited the mobility of the highly flexible $\alpha_3$-$\beta_4$ loop that is entropically unfavorable. This compensation of enthalpy and entropy reduced the calcium stabilization effect. It would be more rigorous to compare the differences in the in free energy change between folded-unfolded states ($\Delta\Delta G_{\text{unfolding}}$), but free energy calculation is still challenging. Recent effort
of combining the free energy perturbation method with approximation of unfolded state free energy has been shown to predict changes in protein stability due to mutation, but has low accuracy for mutations that resulted in change in net charge [152]. Thus, it is illuminating that simple interaction energy calculations and dynamics analysis can provide insight into the relative stability of a folded protein under different conditions and also qualitatively match experimental observations complementing other previous developed techniques [152-154].

Hydrodynamic force in circulation regulates A2 cleavage by unfolding the A2 to expose the cryptic cleavage site. Mimicking this in our SMD simulations, we found that both A2 wild-type and mutant unfolded from the C-terminus and progressed toward the N-terminus. The unfolding pathway of the wild-type and mutant were comparable until the rupture of the β4, at which point differences emerged. The unfolding of β4 was coordinated with the unraveling of the α3-β4 loop that contains the 1597 residue and calcium ion. The rupture resulted in the highest tensile force in the unfolding pathway. From our first set of SMD simulations with a pulling speed of 0.5 nm/ns and single statistics, the trend in the peak rupture, which corresponded to the unfolding of β4, did not quantitatively match with the experimental observation. We then carried out exhaustive repeats of SMD simulation with slower pulling speed (0.15 nm/ns), and the results were more consistent with the experiments. First, using thermal stability for comparison, the peak rupture force correlated positively with the melting temperature measurement (r = 0.87). The relationship between mechanical and thermal stabilities has been studied experimentally and theoretically for DNA molecules; these reports suggest that the relationship between the two variables is positively correlated [155, 156]. Here, our result shows that slow
pulling speeds and sufficient statics allow the relative thermal stability of a protein to be predicted from SMD simulation. Second, calcium had been shown to increase the resistance of the native state of the wild-type against force [134, 135], and our simulation also demonstrated that the present of bound-calcium increased the peak rupture force significantly for the wild-type A2. Furthermore, we observed marginal decreases in the peak rupture force due to mutation, which agrees with the experimental observation that the mutant unfolds at lower force [136].

A2 is known to have a mechanically stable but partially unfolded intermediate state [127, 135]. The structure of the intermediate is still uncertain, as only the contour length (i.e. number of residues) of the intermediate structure has been measured in optical tweezers experiments [127, 135]. Based on the analysis of the native structure and previous simulations, it has been suggested that the folded regions of the intermediate state correspond to the N-terminal regions of the A2, starting from the N-terminal up to, or including, the β4 strand. This model of the intermediate structure exposes the 73 amino-acid (residues D1596 to R1668) minimal substrate in the C-terminal region for ADAMTS13 [157]. Recently, one of the two optical tweezers studies that investigated the effect of calcium binding, suggested that the present of calcium ion stabilized a non-native contact intermediate state [135]. Here, the simulation could not provide the structure of a non-native intermediate state as even our slowest pulling rate was still too fast to sample a non-native contact structure. Nevertheless, the simulation certainly showed that calcium ion stabilized the N-terminal region from mechanically induced unfolding by including an additional barrier for the N-terminal region from unfolding.
The calcium also increases the refolding rate of A2, whereas the R1597W mutation reduces the refolding rate and also causes the refolding rate insensitive to calcium ion [136]. Detailed molecular mechanisms of how metal ions can affect the protein folding rates under tensile force are lacking. A2 and calmodulin [158] are some of the examples of systems for which this has been reported. Currently, the MD simulation is still limited in providing molecular details on refolding process, especially since A2 folds on the order of second time-scale. However, several key interactions that we observed might be important factors in the refolding process. The simulation showed that the calcium ion binds very strongly to the binding site residues through coordinated metal-protein interaction that can bridge residues that are distant in sequence. For example, the bridge interaction of D1498-Ca$^{2+}$-D1597 connected residues that are nearly 100 residues apart. Perhaps a calcium ion increases the folding rate by connecting charge-charge repulsive negative residues to form a stable transition state that reduces the degree of freedom in the folding pathway.

In summary, we showed that mutation induced structural changes localized around the mutation site and disrupted calcium binding. The calcium ion stabilized the wild-type A2 by forming specific coordinated interactions within the negatively charged binding site, while the mutation disrupted calcium binding to make A2 stability insensitive to the presence of calcium. These results are in good agreement with previous experimental observations, which further establish MD technique for studying bimolecular stability. Furthermore, the novel mutant structure of A2 presented here can be useful for drug development for Type 2A VWD caused by R1597W mutation. For example, $\alpha_3$-$\beta_4$ loop region of A2 is a promising target site for a drug that increasing the stability of A2. The computational methods outlined in this work or other emerging computational screening
techniques [159, 160] combined with rapid experimental validation [161, 162] can effectively select promising drug candidates. Lastly, even though the atomistic MD simulation technique has yet to be able to show how calcium ion or mutations alter refolding pathway or resolve intermediate structures that deviate from the crystal structure, it is encouraging that significant progress have been made on a smaller fast folding protein recently [163]. In the near future, it would be interesting to utilize MD simulation to study the detailed mechanism of metal ion regulation of protein refolding process to advance the understanding of protein structural, dynamical and functional relationships.
Chapter 5

Multiplexed single-molecule force spectroscopy using a centrifuge

5.1 Synapsis

We present a miniature Centrifuge Force Microscope (CFM) that repurposes a benchtop centrifuge for high-throughput single-molecule experiments with high-resolution particle tracking, a large force range, temperature control, and simple push-button operation. Incorporating DNA nanoswitches to enable repeated interrogation by force of single molecular pairs, we demonstrate increased throughput, reliability and the ability to characterize population heterogeneity. We perform spatiotemporally multiplexed experiments to collect 1863 bond rupture statistics from 538 traceable molecular pairs in a single experiment, and show that two populations of DNA zippers can be distinguished using per-molecule statistics to reduce noise.
5.2 Introduction

The ability to mechanically manipulate single molecules is leading to insights throughout biomedical research, from the action of molecular motors in replication and transcription to the role of mechanical forces in development [2-4, 164, 165]. While in principle these approaches enable the full characterization of individual molecular complexes and the study of population heterogeneity at the single-molecule level, in practice key challenges exist. The first challenge for force spectroscopy studies is the low throughput of most single-molecule approaches, which is just starting to be addressed with recently developed multiplexed methods [166-173]. Furthermore, sufficient statistics must be collected not only for the population [22], but also for each individual molecule, which can be a challenge for studying catastrophic transitions such as bond rupture [24, 174, 175]. Another challenge is the positive identification of the single-molecule interactions of interest over non-specific and multiple interactions. Finally, there is the subtle challenge of noise, both thermal and experimental, that makes distinguishing different populations of molecules with similar force properties difficult.

We have met all of these challenges with spatiotemporally multiplexed force spectroscopy, a combination of massively parallel spatial multiplexing with repeated interrogation enabled by self-assembled nanoscale devices. First, we introduce a miniature Centrifuge Force Microscope (CFM) for high-throughput single-molecule experimentation that utilizes a commercial benchtop centrifuge. Functionally similar to our first prototype device [13], an entire microscope imaging system is rotated to observe microscopic objects subjected to uniform centrifugal force (unlike earlier “spinning disk” centrifuge
microscopes [176, 177]). This inexpensive design brings new features including temperature control and high-resolution particle tracking (~2 nm). Second, we introduce a high-throughput CFM assay that integrates mechanical nanoswitches [175, 178] to provide important new functionality. The nanoswitches serve two roles—one as a molecular signature to facilitate reliable and automated analysis of large data sets, and the second to enable the repeated interrogation of each single-molecule pair, increasing throughput and enabling new measurements of heterogeneity in single-molecule experiments. By making repeated force measurements on hundreds of single-molecule complexes, we can collect multiple statistics on each molecule that comprises the population. We additionally show that by averaging multiple rupture forces on a per-molecule basis we can reduce noise to enable super-resolved force spectroscopy—the identification of different populations of molecules below the thermal force-resolution limit. Averaging allows us to reduce the spread in force distributions (averaging reduces noise by a factor of \( \sim \sqrt{N} \)) without losing information about differences between molecules. Furthermore, the rich and relatively large data sets provided by our technique could also complement other analysis techniques for statistical deconvolution [179, 180].

5.3 Result and Discussion

The redesigned CFM is a miniaturized microscope that fits into a commercial centrifuge so that force can be applied to samples while their micro- to nano-scale motions are tracked (Figure 5.1 and Figure S4.1). To accommodate the microscope components in the 400mL bucket volume, we selected components with size and weight in mind, and induced two
right-angle bends in the optical path (Figure S4.1b). To ensure accessibility of the instrument, most components are commercially available and can be easily assembled (Figure S4.1d,e and Table S4.1). A 3D-printed housing was made to encapsulate the device and to ensure a tight fit within the bucket. Transmission of camera data out of the CFM during centrifugation is accomplished by converting the camera’s gigabit Ethernet signal to a fiber-optic signal, then passing this data out of the centrifuge through a fiber rotary joint (Figure S4.1a,f). Details of the design are presented in the Online Methods.

Figure 5.1: Overview of the benchtop Centrifuge Force Microscope (CFM). (a, i) Illustration of a spinning centrifuge rotor containing the CFM in the red swing bucket; the rotation induces forces on molecules tethered between beads and the coverslip surface. The direction of force application can be set by constraining the angle of the centrifuge bucket, as described in Figure S4.5. (a, ii) Schematic of the CFM. (b) Tether extension is monitored as a function of time while force is applied. When the DNA nanoswitch is integrated, molecular transitions such as bond rupture between a receptor (R) - ligand (L) pair causes a well-defined change in tether extension, providing a distinct signature for detecting interactions between two molecules of interest.
We implemented a unique solution to measure nanometer-level extensions of tethers in the CFM by projecting tether length changes onto the X-Y plane of the coverslip, enabling these measurements to be made in a relatively simple and computationally efficient way. Taking advantage of the fact that we can precisely control the direction of force application by mechanically constraining the angle of the centrifuge bucket, we intentionally misaligned the force and imaging axes (Figure S4.2a and Supplementary Note 3.1). Using this technique, we can “tune” our tether length resolution based on the bucket angle with a range of approximately 2.5 nm (at 20°) to 12 nm (at 80°) based on our lateral tracking resolution of ~2 nm (Figure S4.2b,c). To validate and demonstrate this approach, we measured DNA force-extension for over 100 molecules simultaneously over a span of < 1 minute (Figure S4.3a), and fit each with the standard worm-like chain model. The most likely contour length and persistence length was 8.2 ± 0.2 μm and 46 ± 1 nm, respectively, in agreement with expected values [181]. We additionally performed multiplexed overstretching measurements of lambda DNA (Figure S4.3c), yielding an overstretching force of 63.5 ± 1.7 pN (mean ± SD), consistent with previous measurements at these conditions [181].

To enable robust and repeatable rupture experiments, we integrated DNA nanoswitches into the CFM. These molecular switches are designed to adopt a looped structure when the molecules of interest are interacting and a linear structure when they are not (Figure 5.2a), providing a distinct “signature” (i.e. increase in tether length) for rupture events (Figure S4.4b). We performed DNA unzipping experiments on a 29 bp DNA interaction, and used the “signature” unlooping of the nanoswitches to positively identify and discriminate valid single-molecule data from multiple tethers and non-specific
interactions (Figure 5.2a and Figure S4.5). Following this approach, unzipping force measurements were carried out in the presence and absence of magnesium ions, with hundreds of rupture statistics for each condition collected in under 30 seconds of centrifuge run time. We found that magnesium stabilizes the duplex as observed previously [182], with the average unzipping force (± the standard deviation) increasing from 10.1 ± 0.9 pN to 14.6 ± 1.1 pN with the addition of magnesium (Figure 5.2b).

We additionally performed DNA unzipping experiments at four temperatures, 4, 13, 23, and 37 °C (Figure 5.2c). Seminal works introduced temperature control into optical tweezers over ten years ago[183, 184], but even today the vast majority of single molecule pulling experiments are carried out at room temperature, and temperature control remains an active area of development [185]. The benchtop CFM has the advantages of built-in temperature control and portability to move into cold (4 °C) or warm (37 °C) rooms available to most biologists. Accurate, real-time measurements of the sample temperature during experiments were made with a wireless thermocouple embedded within the centrifuge bucket (Online Methods for full details). We found an increase in the unzipping force with decreasing temperature, in good quantitative agreement with the thermodynamic models of DNA unzipping and previous observations [186] (see also Supplementary Note S3.2).
Figure 5.2: DNA unzipping force measured using DNA nanoswitches on the CFM. (a, i) Schematic of the unzipping construct. Two complementary oligos (red and purple strands) hybridize to form a looped DNA nanoswitch. Force can unzip the two complementary strands, resulting in a measurable increase in tether length, providing a signature of DNA unzipping. (a, ii) Images of a bead tethered to the surface via a DNA nanoswitch showing the looped and unlooped states. The scale bar is 1 μm long. (a, iii) In one example of rupture force measurement, we identified 381 tethers with the DNA nanoswitch transitions signature to collect rupture forces while the remaining 673 transitions that corresponded to bead detachment and improper transitions were omitted (see also Figure S4.5 for detail). (b) Unzipping force histograms of 29 bp dsDNA measured with the DNA nanoswitch under two different buffer conditions. (c) Average unzipping force of 29 bp dsDNA under different temperatures with PBS buffer (Total n = 306), with histograms of rupture forces shown as an inset. The theoretical line is calculated using a previously described thermodynamic model [186].

Finally, we demonstrate spatiotemporally multiplexed force spectroscopy using the DNA nanoswitch assay on the CFM to repeatedly interrogate a population of molecules at
the single-molecule level (Figure 5.3a). We measured 1863 statistics of DNA unzipping from repeated pulls of 538 molecules in a single sample to demonstrate the large amounts of single-molecule force data that can be accumulated with this approach (Figure 5.3b). Even more significantly, the nanoswitches enable the unique properties of each molecule in a sample to be characterized from repeated measurements, as we demonstrate by determining a rupture-force histogram for each molecule in a sample, illuminating population heterogeneity at the single-molecule level (Figure 5.3c). Furthermore, we show that by averaging data from multiple pulls of the same molecular pair the spread in force is reduced without losing the unique characteristics of each molecule. When applied to data from a single population, this per-molecule force averaging generates a super-resolved histogram with the expected narrowing when compared to the raw histogram of all the data (Figure 5.3d). When applied to combined statistics from two populations of DNA zippers (introducing another G-C rich zipper), we show that the super-resolved histogram generated from per-molecule averaging can separate out two populations that are unresolvable from the raw histograms due to the intrinsic broadening of force that results from thermal noise and instrumental noise (Figure 5.3e).
**Figure 5.3:** Repeated rupture force measurement of single molecular pairs. (a) Protocol for repeated cycles of force application, with each cycle consisting of a linear force ramp to induce rupture and DNA nanoswitch unlooping, followed by a low force reassociation period allowing the molecular pairs to rebind. (b) DNA unzipping force histogram of 1863 rupture events collected from a total of 538 molecular pairs with 12 cycles of force application. The color from blue to red corresponds to statistics collected from each cycle. (c) Example rupture force histograms generated for individual molecular pairs. The calculated average rupture force is shown below in red. (d) Combined histogram of rupture forces from 27 molecules with 7 cycles of force rupture each (top), and histogram of the per-molecule averaged rupture force (bottom), showing a reduced width. (e) Combined histogram for two populations of DNA unzipping experiments (top), and the per-molecule averaged super-resolved histogram (bottom) that recovers the two separate populations from the mixed data.
As we have shown, together the CFM and the DNA nanoswitch assay provide powerful single-molecule capabilities, and give non-specialists unprecedented access to force spectroscopy. And even separately these technologies provide unique benefits that can make single molecule experiments better, cheaper, and faster. Our miniature CFM presented here, which integrates into standard benchtop centrifuges, has important new features including an expanded biologically-relevant dynamic force range (Figure S4.6), easily implemented temperature control between 4°C and 37°C, improved safety, lower cost (Supplementary Table 1 and Supplementary Note 3), and greater ease of use compared to the first-generation CFM. These last two points are important strides towards increasing the accessibility of single-molecule experiments for non-specialists. The total reduction in cost is particularly significant when considering the additional infrastructure typically needed for single-molecule measurements that are obviated by our approach, as the centrifuge itself provides an isolated, temperature-regulated enclosure. Additionally, our instrument can be operated without significant training, even by undergraduate researchers. Similarly, the DNA nanoswitch-based single-molecule assay has its own benefits, significantly reducing the effort to obtain reliable and accurate measurements by standardizing sample preparation, and by supplying a distinct molecular signature that enables reliable data verification and automated analysis. Furthermore, it enables repeated interrogation of each molecular complex for building up statistics of both the population and the individual molecules, revealing population heterogeneity at the single-molecule level.

There is difficulty in generalized comparisons of single molecule instruments due to the wide variety of designs, but the miniature CFM presented here (using the 2.8 µm
dynabeads) is comparable to our own optical tweezers setup[127, 175] in force range (~0.1 to 100 pN), force resolution (~3-5%), and spatial resolution (~2 nm). It is worth noting, however, that the most advanced optical tweezers push spatial resolution into the sub-Angstrom range, have force resolution of < 0.1 pN, and have temporal resolutions > 10 kHz [164] (this CFM is 15 Hz). Regarding multiplexing, magnetic tweezers have achieved up to 357 statistics in a single experiment [168], acoustic manipulation has achieved 145 statistics [166], and other multiplexing methods show great promise as well [172, 173]. However, our demonstration of 1863 validated single-molecule statistics in one experiment is to our knowledge more than in any previous work.

As the potential for single-molecule approaches to address critical problems in biology becomes increasingly clear, the availability of such simple and powerful tools that do not require months of training or hundreds of thousands of dollars becomes key. We believe the technologies presented here are an important step in that direction, enabling both a more detailed view of single-molecules as part of a population, and a more streamlined way to amass large numbers of statistics. It is not unreasonable to imagine that this approach, paired with sample scanning and/or fast force cycling could eventually enable the collection of millions of single-molecule statistics from a single sample. With these advances, previously unrealistic applications such as high-throughput single-molecule screening and diagnostics become more viable, and single-molecule analysis may start to become as ubiquitous as standard bulk assays.
5.4 Material and Methods

Centrifuge Force Microscope Instrumentation

The Centrifuge Force Microscope (CFM) consists of an optical microscope and digital image acquisition system, which is integrated into a centrifuge (Figure 5.1a and Figure S4.1). We used a refrigerated benchtop centrifuge (Thermo Scientific, Heraeus X1R) with two modifications. First, the TX-400 rotor was modified to mount a fiber optic rotary joint along the central axis (Princetel MJX). This was accomplished by removing the rotor’s central push-release mechanism (by loosening the screw on the side of the button), and threading the four existing through holes to accept 10-32 screws. An adapter was designed and installed on the rotor to hold the slip ring (Figure S4.1a,f). To enable the fiber optic to pass through the lid, the central plastic viewing window was removed. Second, the centrifuge’s control module was upgraded to enable computer control. An upgraded module was kindly provided by Thermo Fisher Scientific to support this project, and was a simple drop-in replacement.

The light microscope of the benchtop CFM was constructed using mainly Thorlabs SM1 compatible components (Figure S4.1b,d,e and Table S3.1). A red LED (Thorlabs, LED630E) threaded to a tube mount (Thorlabs, S1LEDM) served as the illumination source. A glass diffuser positioned between the LED source and sample cell provided uniform illumination across the field of view. The 25 mm diameter of the sample cell was designed to be compatible with the SM1 lens tube (see below CFM sample cell). The sample was magnified and imaged onto a CCD camera (AVT, Prosilica, GC 2450) with a 40X Olympus Plan Achromat objective (infinity corrected, 0.65 NA and 0.6 mm WD) and
Ø1" 100 mm tube lens (Thorlab, AC254-100-A). Due to the limited depth of the centrifuge TX-400 bucket (~16 cm), we shortened the length of the microscope by bending the light path 180° using a pair of turning mirrors glued to a custom turning cube (part #17 in Figure S4.1e). Moreover, a 3D-printed enclosure made of acrylonitrile butadiene styrene (ABS) was used to secure and integrate the imaging and acquisition system inside of the centrifuge bucket. The enclosure also included an open slot for a battery (SparkFun, PRT-00339), and a connected DC-to-DC step up circuit (SparkFun, PRT-08290) that served as the power source for the LED, camera, and media converter. The camera used the standard GigE Vision interface, outputting the data as a gigabit Ethernet signal. To enable live imaging, we utilized a fiber-optic rotary joint (PrinceTel, MJX) installed at the center of the centrifuge rotor (Figure S4.1f). The camera signal was converted from twisted-pair Ethernet to a fiber-optic signal by a small media converter inside of the centrifuge, transferred through the rotary joint, then converted back to a standard Ethernet signal by a second media converter connected to the acquisition computer. (IMC Networks, 855-10734 and 855-10735) (Figure S4.1a). The images collected from the camera were recorded with custom LabView software developed with the Vision Module.

We embedded a portable wireless thermocouple connector (Omega Engineering, MWTC-D-K-915) with a surface adhesive thermocouple (Omega Engineering, SA1XL-K) within the bucket that contains the CFM to measure the sample temperature. A wireless receiver (Omega Engineering, WTC-REC1-915) was used to acquire the temperature from the thermocouple connector to record the temperature in real time.
Molecular Constructs

The looped DNA nanoswitch construct was made using our previously published DNA self-assembly protocol[175, 178]. Circular M13mp18 single-stranded DNA (ssDNA) (New England Biolabs, N4040S) was linearized by hybridizing a 40 bp oligo that created a double-stranded restriction site for the BtsCI enzyme (New England Biolab, R0647S). Subsequently, a set of complementary oligos (Integrated DNA Technologies) was hybridized onto the linear ssDNA. Functionalized oligos (biotinylated and digoxigenin-modified) were hybridized onto the 3' and 5' ends of the ssDNA respectively. The hybridization was carried out with 15 nM of linearized ssDNA and 10 molar excess of the complementary oligos in 1x NEBuffer 2 with a temperature ramp from 90 to 20 °C (-1 °C/minute) in a thermocycler. After this initial hybridization two specific single-stranded regions remained, which were bridged by two partially-complimentary oligos to form the final looped construct (Error! Reference source not found.a). The sequence of the complimentary bridge oligo that formed the loop was: CTCAATATCAAAACCCTCAATCAATATCT. This secondary hybridization step was carried out at a final construct concentration of 250 pM with a 1.25 molar excess of the bridge oligos in 1x NEbuffer 2 at room temperature for 1 hour. We verified looping of the construct using gel-shift assays, single-molecule optical trap measurements, and AFM imaging (Figure S3.4). In the optical trap measurement, we applied force on the looped DNA construct via tethering between laser-trapped streptavidin and anti-digoxigenin functionalized silica beads.
For the DNA overstretching measurements in the CFM, we functionalized both ends of lambda DNA with biotin to provide strong anchorage to the streptavidin functionalized glass coverslip and bead surfaces. First, 20 μL of lambda DNA (0.28 ug/ml, Roche, 10745782001) was incubated for 20 minutes at 65°C to remove the hybridized overhangs. Subsequently a nucleotide mixture that consists of Biotin-14-dATP, Biotin-14-dCTP, dTTP and dGTP, each at 100 μM final concentration, was added to the lambda DNA solution with 0.25U/ml Klenow Fragment (New England Biolabs, M0212S). This mixture was incubated for 1 hour at 37°C. The dual-end biotin lambda DNA was purified from the excess nucleotides and enzyme using the Qiagen PCR Purification Kit.

For the parallel force-extension measurements, we made the half-length lambda DNA functionalized with digoxigenin and biotin. First the biotin-labeled full-lambda DNA construct was cut near the middle using the Xbal restriction enzyme (New England Biolab, R0145S). The resulting overhangs were functionalized with digoxigenin to produce a heterobifunctional 24 kbp construct labeled with digoxigenin on one side and biotin on the other.

CFM sample cell

The looped DNA nanoswitch construct was made using our previously published DNA self-assembly protocol [175, 178]. Circular M13mp18 single-stranded DNA (ssDNA) (New England Biolabs, N4040S) was linearized by hybridizing a 40 bp oligo that created a double-stranded restriction site for the BtsCI enzyme (New England Biolab, R0647S). Subsequently, a set of complementary oligos (Integrated DNA Technologies) was hybridized onto the linear ssDNA. Functionalized oligos (biotinylated and digoxigenin-
modified) were hybridized onto the 3’ and 5’ ends of the ssDNA respectively. The hybridization was carried out with 15 nM of linearized ssDNA and 10 molar excess of the complementary oligos in 1x NEBuffer 2 with a temperature ramp from 90 to 20 °C (-1 °C/minute) in a thermocycler. After this initial hybridization two specific single-stranded regions remained, which were bridged by two partially-complimentary oligos to form the final looped construct (Figure 5.2a). The sequence of the complimentary bridge oligo that formed the loop was: CTCAAAATATCAAACCCTCAATCAATATCT. This secondary hybridization step was carried out at a final construct concentration of 250 pM with a 1.25 molar excess of the bridge oligos in 1x NEbuffer 2 at room temperature for 1 hour. We verified looping of the construct using gel-shift assays, single-molecule optical trap measurements, and AFM imaging (Figure S4.4). In the optical trap measurement, we applied force on the looped DNA construct via tethering between laser-trapped streptavidin and anti-digoxigenin functionalized silica beads.

For the DNA overstretching measurements in the CFM, we functionalized both ends of lambda DNA with biotin to provide strong anchorage to the streptavidin functionalized glass coverslip and bead surfaces. First, 20 μL of lambda DNA (0.28 ug/ml, Roche, 10745782001) was incubated for 20 minutes at 65°C to remove the hybridized overhangs. Subsequently a nucleotide mixture that consists of Biotin-14-dATP, Biotin-14-dCTP, dTTP and dGTP, each at 100 μM final concentration, was added to the lambda DNA solution with 0.25U/ml Klenow Fragment (New England Biolabs, M0212S). This mixture was incubated for 1 hour at 37°C. The dual-end biotin lambda DNA was purified from the excess nucleotides and enzyme using the Qiagen PCR Purification Kit.
For the parallel force-extension measurements, we made the half-length lambda DNA functionalized with digoxigenin and biotin. First the biotin-labeled full-lambda DNA construct was cut near the middle using the XbaI restriction enzyme (New England Biolab, R0145S). The resulting overhangs were functionalized with digoxigenin to produce a heterobifunctional 24 kbp construct labeled with digoxigenin on one side and biotin on the other.

Data Analysis

The miniCFM sample cell was constructed using double-sided Kapton tape sandwiched between a 25 mm diameter support glass and a 19 mm diameter cover glass (Gold Seal, 3346). The 0.7 mm thick support glass was ordered from S.I. Howard Glass (D263) and two 1 mm diameter ports were drilled that served as the solution inlet and outlet. The cover and support glasses were cleaned by immersing in 100 mL a 1% (v/v) Hellmanex III solution, microwaving for 1 minute, then sonicating for 30 minutes. Subsequently, the slides were rinsed thoroughly with Millipore water then dried with nitrogen flow. A 1 mm x 7 mm rectangular flow channel was cut on the double-sided Kapton tape using a cut plotter (Graphtec). To form tethers with digoxigenin functionalized construct, we functionalized the cover glass with anti-digoxigenin using a modified version of a previously developed protocol[187]. First the cover glass was coated with a nitrocellulose solution by depositing 2 μL of amyl acetate solution with 0.2 % (m/v) dissolved nitrocellulose. We then incubated the channel with phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM phosphate buffer, pH 7.4) solution containing 100 ug/ml anti-digoxigenin (Roche, 11333089001) for 15 minutes. The channel was then washed and
further incubated with a surface passivation solution (10mg/ml Roche Blocking Reagent in PBS) for 1 hour. After the passivation step, the channel was flushed with experimental buffer then incubated with 5 pM of construct for 15 minutes. At 5 pM construct concentration, the construct was limited to an average spacing of roughly 2 μm on the surface, making formation of double tethers a rare event. After tethering the construct to the surface, the flow channel was washed with 20 μL of the experimental buffer then incubated with 15 mg/ml streptavidin beads (Invitrogen M-270). For each experiment, beads were washed excessively with the experimental buffer before loading them to the sample cell. Before loading the sample cell to the mini-CFM, the solution inlet and outlet ports were sealed with vacuum grease. The tris experimental buffer in Figure 5.2 consisted of 10 mM Tris, 30 mM NaCl at pH 7.5 with or without 10 mM MgCl₂

For the overstretching experiment, the surface tethering was strengthened by replacing digoxigenin-anti-digoxigenin with biotin-streptavidin—in other words, biotin-streptavidin interactions were used to anchor both ends of each tether. The nitrocellulose surface was functionalized by incubating it with 1 mg/ml streptavidin in PBS solution that contained 1mg/ml of Roche Blocking Reagent for 12 hour, followed by incubation with passivation solution (10mg/ml Roche Blocking Reagent in PBS buffer) for 1 hour. The channel was flushed with PBS and incubated with 5 pM of dual-biotin λ-DNA for 15 minutes before loading in the streptavidin-coated beads. Under such conditions, the density of streptavidin on the surface was sparse enough that only one end of the biotin-labelled λ-DNA bound to the surface, leaving the other biotinylated end free to bind to the streptavidin-coated bead.
Swinging Bucket angle measurement

To measure the angle of the swinging bucket relatives to the axis of rotation, we attached an oil-based marker to the bottom of the bucket, which marks the height of the bucket on the wall as the centrifuge spins (Figure S4.2d). The uncertainty of the angle measurement was based on the distance measurement error estimate of 1 mm. At the rotational speed of 300 RPM the bucket swings out to an angle of (81.4 ± 0.8)°. At a much higher speed of 1800 RPM the angle increases by 2.3%. Further increase of the rotational speed does not increase the angle beyond the error of the measurement.

Data Analysis

The loop opening signature was identified by tracking the beads’ x and y positions. As the angle of the centrifuge swing bucket was not 90° with respect to the axis of rotation, a component of the centrifugal force was directed in the x-y plane (Figure S4.5). As the rotational speed of the CFM increased, the looped DNA tethers were continuously extended. The opening of the loop was identified as a discontinuous change in extension. Each movie contained approximately 1000 beads. To track these beads, they were first identified using the Matlab function imfindcircles. A template image for each bead was stored. To identify the bead in the subsequent frame the template image was scanned in the x-y plane to find the position of maximum correlation. First the image was scanned in the x direction over a 25 pixel search region centered on the bead position from the previous frame. A 2nd order parabola was then fit to the correlation coefficient as a function of position. The position of maximum correlation was identified as the new bead position. The template image was than centered on the new x position, and the same procedure was
done in the y direction. During the course of each experiment, there was some drift in the x-y plane. This was corrected for by taking the median change in x and y for all beads being tracked from frame to frame. We found this drift correction to be sufficient for identifying the looped to un-looped transition. Transitions were identified by filtering out all bead trajectories except those that contained a discontinuous change in extension of both the correct magnitude and direction (Figure S4.5b,c). This removed false transitions that may have resulted from non-specific interactions or the formation of multiple bonds between the bead and the surface. Following this automated filtering procedure, the transition events were visually inspected, in random order, to reject any remaining erroneous transitions, which may have occurred due to particle tracking artifacts that manifest as discontinuous changes in position (i.e. particle mislabeling, overlapping beads, etc.).

For DNA unzipping experiments the number of transitions as a function of time was converted to the number of transitions as a function of force as follows: The rotational speed of the centrifuge was recorded during each movie using WinMess software (provided by Thermo Fisher Scientific, R&D) which enables communication and control with the computer. The rotational speed was then converted to force ($F$) as a function of time using the following equation:

$$F = m_{\text{bead}} R_{\text{CFM}} \omega^2$$

where $R_{\text{CFM}}$ and $\omega$ are the rotational radius and speed, respectively. To determine the effective mass of the beads ($m_{\text{bead}}$) in solution, we measured the Invitrogen M-270 bead density through sink-float analysis using aqueous sodium polytungstate solution (Sigma-Aldrich, 71913) and obtained a density of $1.61 \pm 0.02$ g/cm$^3$ that confirmed the
manufacturer’s (Life Technologies) reported value of 1.6 g/cm$^3$. The manufacturer also provided the bead diameter of our specific lot giving a mean of 2.80 um with <1.6% CV (lot #144315600). The density of the silica beads used in the DNA force extension and overstretching experiment was measured similarly using the sodium polytungstate solution, yielding a density of 1.50 ± 0.03 g/cm$^3$. The diameter of the silica beads was measured using transmission electron microscopy (TEM), yielding an average and standard deviation of 4.27 ± 0.16 um.

**Tethered Particle Motion**

We carried out tethered particle motion analysis for our tethered beads prior to the rupture force measurement. We analyzed the lateral fluctuations of each bead, calculating the root mean square of the drift-subtracted displacement and the symmetry ratio, following previously established methods [188]. Here the symmetry ratio was calculated as the square root of the ratio between the minimum and maximum eigenvalues of the covariance matrix for the in-plane displacement. The in-plane position of each bead was recorded for 10 seconds at an acquisition rate of 10 Hz.

**Repeated Cycles of Rupture Force Measurement**

Multiple rupture events were collected for each molecule in a set by repeatedly spinning the same sample multiple times. For the data presented in Figure 5.3, the sample was spun up with an effective force loading rate of 1 pN/s to a maximum force of 22 pN. The speed was then ramped down to zero rpm for approximately 1 hour to allow rebinding between each pair of molecules. Beads were identified from cycle to cycle by their positions relative
to a set of fiducial beads which were common to each cycle. For the measurement of the
two populations of molecules, we made two different DNA nanoswitch unzipping
constructs, one with 48% GC content (CACGAATTCTCTGCTCCCTTTTAACCCTAG) and one with 31% GC content
(CTCAAATATCAAAACCCTCAATCAATATCT).
Appendix 1

A1.1 Supplementary Notes for Chapter 2

Position of the outgoing and complementary strands in the proposed intermediate structure.

We first determined the position of the outgoing strand during the strand-exchange by minimizing the interaction energy of the outgoing strand with Arg226, Arg243, and Lys245. The position of the complementary strand in the proposed intermediate structure was determined by the following criterion: The base-pairs of the complementary strand can form stable hydrogen bonds with the outgoing strand and with the incoming strand upon base-flipping. This criterion can be satisfied if the positions of the backbone are 17.5 Å apart. When dsDNA enters the open region of the protein which contains site I and site II, we propose that one phosphate backbone binds to site I while the complementary phosphate backbone is attracted toward site II, which forms a surface of positive electrostatic potential (8) (Supplementary Figure S1.1). The complementary strand may also be attracted toward the net positive charge on the unstructured L2 loop since crystal structures of M.
tuberculosis (38) or M. smegmatis (39) RecA show that the L2 loops occupy the open region of the protein structure near site II (Figure S1.4). These electrostatic effects locate the complementary strand near the position occupied by the complementary strand in the strand exchange intermediate state. The L2 loop may then structure in response to the interaction with the dsDNA resulting in a complex with a structure that is similar to the intermediate state in strand exchange after the heteroduplex dsDNA has formed. If a sufficient number of DNA triplets are bound, a final rotation of the dsDNA/L2 loop complex becomes free energetically favorable. We propose that it is this last transition which is the common rate limiting step for both the binding of free RecA from solution and strand exchange.

**Detailed discussion of the proposed steps in homology recognition and strand exchange**

During the initial step in strand exchange an ssDNA molecule enters the open region of the protein and the phosphate backbone binds to site I, where the L1 and L2 loops are not structured previous to binding. The ssDNA is not bound to another strand; therefore, unlike the case where dsDNA interacts with free RecA, the bases in the incoming strand are not pulled toward site II. The unbound ssDNA bases interact with the L1 and L2 loops, forming the presynaptic filament known from X-ray crystallography. In this structure, the L1 and L2 loops largely surround the phosphate in the rises between the triplets. Arg226 and Glu207 form a salt bridge that may have a role in stabilizing the L2 loop in this position (8). Unlike the case when dsDNA is bound to site I, L2 (Ile199) and L1(Met164) do not bind because Met164 is in a different position, as shown in Supplementary Figures S1.7
and S1.8. Similarly, the indirect link between L1 and L2 due to Ser162 interacting with the complementary strand backbone is also absent when only ssDNA is bound to site I since there is no complementary strand backbone in that case.

The RecA-ssDNA complex searches for homology by binding to dsDNA to site II because of the strong electrostatic attraction between the outgoing strand phosphates and the amino acid residues in site II including Arg226 (pink), Arg243 (yellow), and Lys245 (purple) (8). Other residues such as Arg227 may also play a role. Experimental work suggests that in this first searching state, the complementary strand is attached to the RecA-ssDNA complex through Watson-Crick pairing with the outgoing strand (15). The arrival of these two DNA strands adds negative charges that may change the free energy such that the final post-strand exchange state is no longer the most free energetically favorable state for the ssDNA/L2 loop complex. For example, the presence of the negative charge may disrupt the interaction between Glu207 and Arg226. We propose that the new free energy minima can be reached by a concerted rotation of the L2 loop and the incoming strand bases, possibly due to an interaction between the negative charge in the dsDNA backbone and Lys198.

Thus, after the dsDNA has bound to site II, the incoming strand/L2 complex rotates arriving at the new free energy minimum. This new position allows flipping of the incoming bases in order to pair with the complementary strand bases, as illustrated in Figure 2.5 (main text). If the complementary strand bases pair with the incoming strand, it will be bound to the complex through Watson-Crick pairing with the incoming strand. If both the intermediate state and the final state are characterized by a free energy with a term
that is linear as a function of the number of bound triplets and a term that is quadratic, then when a small number of triplets are bound, the linear term will dominate; however, when a sufficient number are bound the quadratic term becomes dominant. Previous theoretical work has suggested that the linear term for the intermediate state is more favorable than the linear term for the final state; however, the same work suggests that the quadratic term is more favorable for the final state than for the intermediate state because of the mechanical energy stored in the base pairs of the dsDNA that is extended by binding to RecA (26). Thus, it will be free energetically favorable for the number of bound triplets in the intermediate state to increase until the quadratic term dominates. Once the quadratic term dominates, the final state will be free energetically favorable. As a result, the L2 loop/heteroduplex complex will rotate to its final post-strand exchange position. This rotation may represent commitment to strand exchange.

In the final position, hydrogen bonds link the Ser162 on L1 to L2 loops via the complementary strand. In addition, there is an interaction between L2 (Ile199) and L1 (Met164) loops that was not present when only ssDNA was bound to site I, shown in Figure S1.8. The additional bonds may make rotation back to the intermediate state much more unfavorable than it was for the initial searching state where only ssDNA was bound in site I: with only ssDNA bound to site I, the additional bonds did not have to be broken and rotating back did not increase the stress on dsDNA base pairs.

For non-homologs, the final post-strand exchange state will have a less favorable binding energy and more mechanical stress on the bases if the lack of Watson-Crick pairing distorts the structure. Such a structural distortion may make the free energy of the final
state less favorable for several reasons: 1. Watson-Crick pairing of the correctly matched bases may be distorted. 2. The interaction between L2 (Ile199) and L1(Met164) may be reduced. 3. The interaction between Ser162 and the complementary strand may be less favorable. For correctly matched AT triplets R169 also provides favorable binding energy and mechanical support that may be absent for non-homologs (8). Thus, in the final post-strand exchange state the energy difference between a bound homolog and a near homolog with one base pair mismatch may be significantly larger than the Watson-Crick pairing energy for the mismatch because the mismatch may reduce or eliminate favorable interactions that are present for the perfect homolog.

**Possible Artifacts in the end dependent pulling experiments**

i) **Internal Nicks**

In these experiments force was applied to single stranded tails at the ends of lambda phage molecules (50 kb long). The 5 kb homologous ssDNA filament matches the sequence at one end of the dsDNA molecule (Figure S1.3). In these experiments, we did some direct tests to determine that only one dsDNA molecule connected the bead to the capillary and to test for nicks near the ends of the molecule. We performed experiments where each molecule underwent an overstretching cycle before we measured the strand exchange rate. If the molecule overstretches at 65 pN, then only one dsDNA must be connecting the bead and the capillary. After overstretching, the force was reduced and the force vs. extension measurements obtained before and after overstretching were compared. Previous experimental studies have shown that nicked dsDNA shows much larger
hysteresis than un-nicked dsDNA. Any molecules showing anomalous hysteresis were rejected from the study.

We checked that overstretching did not unbind the filaments by performing additional experiments using a biotinylated homologous filament bound to a 1-μm polystyrene bead. The location of this small bead indicates binding to the homologous region, and the presence of the bead after several overstretching cycles proves that overstretching does not lead to the unbinding of the filament.

Finally, we performed experiments where we did not overstretch the dsDNA before measuring the strand exchange rate. In those experiments, we measured the strand exchange rate at an applied force of 36 pN or less. Once the strand exchange rate measurements were finished, we overstretched the dsDNA to make sure that only one dsDNA was bound between the bead and the filament. The strand exchange rates measured in experiments on dsDNA molecules that were not overstretched before the strand exchange rate was measured were the same as the results of those where we overstretched first. Thus, overstretching first did not affect the strand exchange rate.

Even if the tests for nicks that are described above were inadequate, the presence of nicks that randomly redistributed the force between the two strands of the dsDNA would make the measured results independent of the ends to which the force was applied. Extensive earlier experimental work using the same protocols applied force to either the 3'3', 3'5', or 5'5' ends of lambda phage molecules. Those experiments showed systematic differences between pulling techniques (40,41) that would not have been present had random internal nicks randomly redistributed force between the strands.
Furthermore, the experimental results shown in Figure 2.3B (main text) indicate that the results obtained when the dsDNA is pulled from the 3'3' ends or the 3'5' ends of the outgoing strand are significantly different from the results obtained when the dsDNA is pulled from the 3'5' ends of the complementary strand. These results included experiments where exactly the same dsDNA strand was pulled, but the ssDNA filament was changed. In that case, the physical strand that represented the complementary strand in one experiment became the outgoing strand in the other experiment. In this case, any defects in the dsDNA would be the same in both cases. The clear difference in these experimental results indicates that the observed reduction in strand exchange rate is a function of the force being applied preferentially to the strand that is complementary to the filament. Thus, the experimental results demonstrate that any remaining nicks are sufficiently rare that in these experiments they do not result in a random redistribution of force over both backbones.

ii) Redistribution of force between the strands due to the finite elasticity of the backbones

Seminal theoretical work by de Gennes suggested that when force is applied to the 3'3' or 5'5' ends of a dsDNA molecule the finite elasticity of the phosphate backbones would result in a redistribution of force between the backbones (42). The number of base pairs required to redistribute the force between the strands is called the de Gennes length. Experiments that measured the shear force required to melt short dsDNA molecules pulled from the 3'3' or 5'5' ends were well modeled by the theory with a de Gennes length of approximately 10 base pairs (43). In contrast, measurements of the force required to shear
long dsDNA molecules did not fit the theory (41), and suggested that the de Gennes length could extend through the entire 50 kb length of lambda phage molecules.

Figure S1.4 indicates the expected experimental results for three cases: 1. The de Gennes length is much shorter than the 5000-base long ssDNA filament 2. The de Gennes length is approximately equal to the length of the 5000-base long ssDNA filament 3. The de Gennes length is much longer than the 5000-base long ssDNA filament. As the illustrations show, if the de Gennes length were $<<$ 5000-base long, the experimental results would be completely independent of the strand to which the force is applied since the small elasticity of the backbones would result in a uniform tension on both strands throughout the length of the dsDNA that interacts with the filament. Furthermore, if the de Gennes length were comparable to the 5000-base length of the filament, then the experimental results would depend only on the strand to which the force was applied at the end nearest to the filament. The strand to which the force was applied at the other end of the molecule would have no effect. The experimental results are clearly inconsistent with the first two cases, so the de Gennes length must be long, as indicated in Figure 2.3C.
Figure S1.1: Electrical potential map of the RecA crystal structure (3CMX) (8) calculated using APBS package (44). These calculations were performed at a temperature of 300°K, solute and solvent dielectric constants of 4 and 80, respectively, and ion concentration and exclusion radius of 0.2 M and 2.0 Å, respectively. The red to blue scale is -16 to 16 kT/e. The surface potential was visualized using PyMol (http://www.pymol.org).
Figure S1.2: Strand exchange schematic. (A) Primary site binding of free RecA to tethered dsDNA. (B) Secondary site binding of RecA-ssDNA filament to tethered dsDNA, strand exchange experiment schematic.
Figure S1.3: Different pulling techniques. The filament is prepared amplifying a 5kb fragment at one end of lambda phage dsDNA (GC-rich end). The ssDNA is obtained by incubating the dsDNA fragment with lambda exonuclease. The strand to be removed is prepared with a 5' phosphorylated primer. The different constructs are obtained by modifying lambda phage samples and the labels placed to allow pulling from several ends.
Figure S1.4: Schematics of the distribution of stress between the two dsDNA backbones for de Gennes lengths: shorter, equal, and longer than the filament length. Pink color indicates base pairs under large stress. (A) The de Gennes length is much shorter than the 5 kb filament length, where for all pulling techniques the dsDNA would not be under differential tension along the length of the filament. (B) The de Gennes length is approximately equal to the filament length, where all of the constructs would be under differential tension along the length of the filament. (C) The de Gennes length is much longer than the filament length. In this case only 3’5’ pulling places the two strands under differential tension along the entire length of the filament.
**Figure S1.5:** Schematic representation of the strand exchange process. (A) Initial 9 bp (1 rise) participating in the homology search. (B) More bases added: 15 participate in the initial searching state. (C) Central base pairs in the intermediate post-strand exchange state whereas the outer bases are in the searching state. (D) Central bases are in the final state; outer bases are in the searching state, and the intermediate ones in the post-strand exchange state. (E) The number of central bases in the final state has expanded as strand exchange proceeds out from the ends. In the absence of hydrolysis, the number of central bases in the final state can increase without binding. In the presence of hydrolysis, the number is limited to ~80 bp. At that point dsDNA begins to unbind from the lagging end of the strand exchange window while new dsDNA is incorporated into the leading end. The two net rates are approximately the same, so the number of base pairs bound to the filament remains constant at approximately 80 as strand exchange progresses through the dsDNA.
**Figure S1.6:** Positions of L2 loops. L2 loops in RecA from *M. tuberculosis* (PDB: 1MO4) (38) or *M. smegmatis* (PDB: 1UBG) (39) in the absence of DNA are shown, where ice-blue is the MtRecA and orange corresponds to MsRecA. The loops are shown superimposed on the final post-strand exchange structure with dsDNA bound in site I. The position of the L2 loop in that final post-strand exchange structure is shown in light blue. The incoming strand is shown in green, and the complementary strand is shown in red.
Figure S1.7: Positions of selected amino acid residues in the final post-strand exchange structure when dsDNA is bound to site I. The incoming strand is shown in cyan, and the complementary strand is shown in red. (A) Several amino acid residues near site II are shown. (B) Rotated view of (A) and selected amino acid residues on L2 loop as well as residues that interact with the complementary strand are shown.
Figure S1.8 Comparison of Positions of residues Ile199 and Met164 in the case when ssDNA is bound to site I and when dsDNA is bound to site I. When dsDNA is bound to site I, the two interact providing a link between the L1 loop of one RecA and the L2 loop of the neighboring RecA. When only ssDNA is present in site I, Met164 assumes an orientation that does not allow this interaction.
Appendix 2

A2.1 Supplementary Notes for Chapter 3

Parallel searching may decrease search times non-linearly

Finally, recent experimental work has indicated that the active filament probes multiple dsDNA positions in parallel [121]. In such a parallel search, deep kinetic traps may barely increase searching times even if they last more than 1000x longer than the quick initial stage. Such long binding times may not greatly affect the overall searching time because the other parallel searching points may continue their rapid probing while one particular region of the active filament is stuck in a deep kinetic trap. In contrast, if the search were not parallel, the system could not test any other possible registration until the deep trap had unbound. Thus, in the presence of deep kinetic traps, a system with $N$ parallel contact points may have a search time which is much shorter than $1/N$ x the searching time for a non-parallel search.
Possible strategy for rejecting pairings involving regions of accidental homology extending over 100 bp

Bacterial genomes can contain more than 10,000 repeated sequences that extend over more than 100 bp. Pairings involving such sequences could not be eliminated using the strategies considered in the main paper. Recent theoretical work [109] and the new experimental single molecule studies discussed above suggest that in the presence of ATP hydrolysis RecA-ssDNA filaments can form superstructures such as superhelices that offer the possibility for dsDNA to be aligned in registration with the ssDNA in the filament at multiple different positions. Such superhelical structures can consist of 60 RecA monomers per (or 180 ssDNA nucleotides) per superturn [109]. In what follows we speculate about how RecA mediated homology recognition could exploit such superstructures to discriminate against pairings between long repeated sequences. An important feature of the proposed model is the change in the rate at which strand exchange occurs once the system makes the transition to the metastable conformation.

Once an 8-nucleotide test region of the active filament has found a homologous 8 base pair region in the dsDNA, a metastable conformation is formed [189]. After the transition to the metastable conformation, strand exchange proceeds at ~ 6 bp/sec, but the strand exchange product remains fairly unstable until it extends to at least 17-20 bp. We note that in vivo chromosomes display significant relative motion that is absent in vitro [190].

We speculate that if the pairing is in registration during the ~ 2 seconds required to complete the strand exchange of ~ 20 bases, the active filament will have made in-
registration contact with homologous dsDNA at multiple positions separated by 180 bp. In contrast, if the pairing involves a region of accidental homology that extends over < 180 bp, then no additional in-registration contacts will be made. If one assumes that *in vivo* the postsynaptic filament is only stable if several separate ~ 20 bp regions have undergone strand exchange, such a system would allow repeated sequence regions to be rejected even if they extended over more than 100 bp. Such a system could also greatly accelerate strand exchange since it would be starting at multiple positions.
**Figure S2.1:** Molecular dynamic simulation setup. (A) Schematic of a typical molecular dynamics simulation system. The solvation water is shown as the transparent box, and the sodium and chloride ions are shown as green and blue spheres, respectively. The surf molecule is the RecA-DNA complex. (B) RMSD and energies measured during the heating, equilibration, and production simulation.
Figure S2.2: Stable binding of B-form dsDNA to the surface of the C-terminal domain (CTD). (A) Simulation of the B-form dsDNA bound to the surface of CTD. The L2 loop, complementary, initiating, and outgoing strands are shown in yellow, purple, orange, and blue, respectively. The snapshots of the trajectories are overlapped with 3 ns simulation time gaps. (B) Phosphate position RMSD from the starting structure.
Figure S2.3: Structures obtained by docking B-form dsDNA with the presynaptic filament. The structures show that the large difference in direction between the dsDNA which is nearly parallel to the open grooves in the protein filament and the ssDNA which is aligned with the central axis of the protein strongly limits the number of bases that could undergo strand exchange. The initiating, complementary, and outgoing strands are shown in orange, purple, and cyan, respectively. The residues in the C-terminal domain 270–333 are shown in pink except for lysine residues K280, K282, K286, and K302 which are shown in green. K232 is shown in silver. The remaining residues in the protein are shown in white or ice blue. (A) Interaction where the dsDNA is bound to neighboring CTD, where the closest backbone separation between the initiating and complementary strands is 28Å. (B) Interaction where the dsDNA is bound to C-terminal domains that are separated by a RecA monomer away, allowing a minimum backbone separation of 24 Å, which only applies to
a single base. All other bases have larger separations. A structure with B-form dsDNA bound to C-terminal domains separated by two RecA domains is not possible because it would require a very large deformation of the stiff protein core; however, the binding is possible if the dsDNA bends, as shown in Figure 3 of the main text.
**Figure S2.4**: Site II of the active RecA filament. (A-B) Residue detail and electrostatics of the secondary DNA-binding site (site II). The initiating strand is shown in orange, lysine residues K280, K282, K286, and K302 of the C-terminal Domain (CTD) are shown in green, site II residues R226, R227, R243, and K245 are shown in red, and the L2 loop residues M202 and F203 are shown in yellow. The electrostatic potential calculation was carried out by using the APBS package; the potential value is colored from red to blue with the value of (-3 to 3 in units of kT/e). (C) dsDNA bound to secondary site complex. The initiating, complementary, and outgoing strands are shown in orange, purple, and cyan,
respectively. L2 loop residues 202-205 are shown in yellow and residues 197-199 are shown in pink. The remaining residues are shown in white and ice-blue indicating successive protein monomers. Multiple L2 loop residues 202-205 (shown in yellow) interact directly with the bound dsDNA.
Figure S2.5: Detailed view of bent dsDNA bound to active filament of RecA. The initiating, complementary, and outgoing strands are shown in orange, purple, and cyan, respectively. The residues in the C-terminal domain 270–333 are shown in pink except for lysine residues K280, K282, K286, and K302 which are shown in green. K232 is shown in silver. The secondary DNA-binding site residues R226, R227, R243, and K245 are shown in red, and the L2 loop (198-206) is shown in yellow. The remaining residues in the protein are shown in white or ice blue. The contacts with the C-terminal lysine residue are
approximately the same as in B-form dsDNA as shown in Figure 3 (main text). On the 5’ side of the bend, the complementary strand has a strong interaction with R243, whereas on the 3’ side of the bend the R243 in the next monomer interacts strongly with the outgoing strand. On the 3’ side of the bend, ~ 3 bp occupy locations where the backbone spacing between the incoming and complementary strands permits base flipping to produce strand exchange.
Figure S2.6: Structures of modeled secondary site bound dsDNA in active filament and site I bound dsDNA in the postsynaptic filament. (A) The initiating, complementary, and outgoing strands are shown in orange, purple, and cyan respectively. The illustrations highlight the large extension difference between the outgoing strand in the secondary site and the initiating strand in the primary site. The first structures show the base pairing within the transparent cylinders. The second structure shows the strands as solid cylinders. These representations resemble the prediction of the previous experimental work by Egelman et al.
al. [191]. Egelman proposed 3 backbone positions within the RecA filament corresponding to backbones with radii of 15Å, 10Å, and 6Å, and suggested that though all three proposed structures had a 5.1 Å rise/bp along the helical axis, the 15 Å radius structure (corresponds to the blue identical strand) had a 7.2 Å separation between phosphates, which corresponds to the maximum possible backbone extension. In contrast, the 10-Å and 6-Å radius structures were projected to have separations of only 6.1 and 5.5 Å, respectively. (B) Side view of corresponding base pair duplet is shown for the two structures. In the postsynaptic filament on the right, the complementary base position prior to base-flipping is shown by a transparent rendering. (C) Bottom view from the 3´ to 5´ end. The illustrations also show that the dynamics evolution toward postsynaptic structure is not simply a matter of base flipping: substantial complementary strand backbone relocation is also required.
Figure S2.7: Structures consisting of B-form tails attached to dsDNA untwisted and extended by binding to site II. The number of bp bound to site II increases from left to right (A) to (C). The colors and the first figure are the same as those used in Figure 3.3 and Figure S2.5.
Figure S2.8: Evolution of a complementary strand duplet flipping from pairing with the identical strand to pairing with the initiating strand. (A) Base pairing parameters $D_1$ and $\theta_1$ are the C1´ and C1´ distance between complementary and initiating strands and the C1´-C1´-N9 projection angle onto the base-flipping plane, respectively. (B) Histogram of $D_1$ and $\theta_1$ of a paired dsDNA in the stable postsynaptic filament obtained from MD simulation (see Figure S2.9). (C) Histograms and trajectories (blue to red) illustrating base-flipping parameter of six bases from three triplets bound in the secondary site. The second and third diagrams correspond to the flipped duplex as shown in the main text. The last two diagrams illustrate the early stage of flipping.
Figure S2.9: Stable binding of dsDNA in the primary site. Simulation of the postsynaptic filament with the post strand-exchange dsDNA bound to the primary site (site I). The phosphate position RMSD from the initial starting structure during production simulation is shown on the right. The average RMSD of the DNA phosphate atom is ~1.5 Å.
**Figure S2.10:** Illustration of the flexibility and range of fluctuation of the CTD. Trajectories of the CTD during the 15 ns production simulation are colored blue to red with 3 ns gap. The backbone of CTD can have over 15-Å range of motion while maintaining its internal structure.
Appendix 3

A3.1 Supplementary Notes for Chapter 4

**Figure S3.1:** Secondary structure time trajectory with 1-ns smooth RMSD plotted on top. In the R1597W system, a RMSD jump occurred which correspond to helical changes in the α3-β4
loop as indicated by the arrows. In the R1597W+Ca$^{2+}$ system, RMSD jump occurred around 110 ns which corresponds to structural changes in the calcium binding site with the detachment carbonyl group of A1600 and W1597 from the calcium ion.
Figure S3.2: A2 wild-type residues root mean square fluctuation (RMSF) compare to estimated RMSF from Beta-Factor (B). The $\alpha_{4\text{less}}$ and $\alpha_3$-$\beta_4$ loop residues are shown as green and blue points, respectively.
**Figure S3.3:** Details of the unfolding pathway of A2 under constant velocity pulling SMD simulation. The WT, R1597W, WT+Ca2+ and R1597W+Ca2+ are shown in (A), (B), (C) and (D), respectively. The top figure in each panel shows the tensile force profile with the horizontal dash line indicating the peak force. The middle figure shows the secondary structure changes during unfolding. The bottom figure shows the solvent surface area (SASA) of residue 1597 and the cleavage site (residue 1605 and 1606).
Appendix 4

A4.1 Supplementary Notes for Chapter 5

Supplementary Note A4.1

Based on the geometry as illustrated in (Figure S4.2a), extension of the molecular tether can be measured by tracking the tethered microsphere’s motion parallel to the coverslip. Specifically, changes in tether extension will appear as a lateral displacement of the bead ($\Delta L_{\text{obs}}$). The actual changes in extension ($\Delta L$) in the direction parallel to the force can be calculated based on the angle of the bucket ($\theta$) as follows:

$$\Delta L = \frac{\Delta L_{\text{obs}}}{\cos(\theta)}$$

The minimum tether length that can be measured in this way depends on both the bead size used in the experiment and the angle of the centrifuge bucket. The bead can make direct contact with the cover glass surface if the tethered length is not long enough for the bead to be pulled away from the surface. The minimum tether length ($L_{\text{min}}$) as a function of bead radius ($R_{\text{bead}}$) and bucket angle ($\theta$) is given by:
\[ L_{\text{min}} = R_{\text{bead}} \left( \frac{1}{\sin(\theta)} - 1 \right) \]

The tracking resolution of tether extension (\(\Delta L\)) and values of \(L_{\text{min}}\) in units of bead radius \((R_{\text{bead}})\) are calculated for bucket angles between 20° to 85° as shown in Figure S4.2c.

**Supplementary Note A4.2**

The unzipping force of the 29 bp DNA duplex measured here increases as the temperature decreases in reasonable quantitative agreement with simple thermodynamic theoretical predictions. Interestingly, the theoretical prediction slightly underestimates the unzipping force at low temperature (Figure 5.2c). Previously, Danilowicz et al. measured the temperature-dependent unzipping force of a 1.5 kbp long DNA, and noted an unzipping force at 15 °C of 36 pN that was 3 times larger than the predicted value of 12 pN [186]. This surprising observation was attributed to a DNA conformational change that occurs at low temperature. This underestimate in the unzipping force that results from the thermodynamic analysis was not as pronounced in our much shorter DNA sequence, suggesting that the low temperature DNA conformational change proposed to explain the earlier results may depend strongly on the length of the unzipped DNA.

**Supplementary Note A4.3**

The benchtop CFM bill of materials is summarized in Table S3.1 and the exploded view drawing of the assembly is shown in Figure S4.1e. The majority of the opto-mechanical components can be purchased from Thorlabs. Comparable components such as the lenses, objective and mirror can be purchased from other vendors as well. The
objective (Item 11), camera (Item 21), and fiber optic rotary joint (Item 23) account for most of cost of this CFM design. The camera used here is equipped with a 5 Megapixel CCD sensor with a maximum frame rate of 15 fps at full resolution. Depending on the application, alternative cameras with different resolutions and acquisition rates could be used to reduce the total cost of construction—for example, the Basler Ace CMOS camera at 15 Megapixels is under $700 and a 2 Megapixel version is under $400 (Edmund Optics, Inc.), which would reduce the total price to about $2,500. The media converter (Item 22) that converts the twisted wire Ethernet connection to an optical signal could be omitted if a camera with a 10 GigE optical fiber output is used. Data output using a compact wireless router is an alternative approach, but may result in a slower acquisition rate. Additional customized parts such as the turning mirror housing can be replaced with a Thorlabs compact cage cube system, but for added stability of the imaging path, we recommend a solid aluminum construction. Moreover, larger floor model centrifuges with 1 Liter buckets can provide sufficient bucket depth that the turning mirror is no longer needed.
Figure S4.1: Detailed design of the benchtop CFM. (a) Photograph of the benchtop CFM module inside of a centrifuge. (b) The benchtop CFM module next to a typical centrifuge bucket. The LED, sample holder, and objective are housed within the lens tubes that are mounted to the turning mirror and camera. (c) Typical field of view during a CFM force spectroscopy experiment showing thousands of surface-tethered beads that can be monitored in parallel. (d) Assembled model of the CFM. The full length of the CFM is less than 5 inches long and it is compact enough to fit into a conventional benchtop centrifuge bucket. (e) Exploded view of the CFM. The majority of the parts can be purchased and threaded for direct assembly without the need for user alignment. Supplementary Table 1 contains a list of parts with detailed purchasing information. The insert (Part 6-8) is the illustration of the sample coverslip. (f) Installation of the fiber-optic rotary joint onto the benchtop centrifuge. The computer-aided design (CAD) drawings of the CFM assembly and customized parts are included as supplemental material.
Figure S4.2: Characterization of bead x-y position tracking. (a) Illustration of tether extension measurements obtained using XY projection when the direction of force and imaging axis are intentionally misaligned. Changes in the length of molecular tethers can be calculated using the observed lateral shift in the position of a tethered bead ($\Delta L_{\text{obs}}$), and the bucket angle ($\theta$). (b) Drift-corrected x and y position of 5 um silica stuck bead as a function of time recorded in the centrifuge spinning at 2,000 RPM is shown in blue and the right panel shows the histogram of the position with normal distribution fits. Standard deviations of the fits for x and y positions are both $\sim$2 nm. The drift in x and y position as a function of time, based on the average position of 12 immobile reference beads, is shown in magenta. While the standard deviation listed here represents the particle-tracking resolution, we note that the overall accuracy with which tether lengths can be measured also depends on the intrinsic thermal fluctuations of these beads as previously described [10] (c) Top panel shows Tracking resolution of tether extension calculated as a function
of the bucket angle as illustrated in (a). The bottom panel shows the Minimum tethered length ($L_{\text{min}}$), scaled by the bead radius ($R_{\text{bead}}$), required to measure tether extension from lateral displacement, calculated here as a function of the bucket angle. (d) Illustration of the angle measurement. The angle is determined from the height of the swinging bucket while the centrifuge spins. The inserted table shows the angle at different rotational speeds.
**Figure S4.3:** Parallel DNA force-extension and overstretching measurements made with the CFM. (a) Force-extension data of half lambda DNA (24 kbp) obtained from a single sample with 113 DNA tethers using a centrifuge bucket constrained to a 20° angle (see Supplementary Figure S4.2a for detail). The top panel shows a scatter plot of the persistence length and contour length obtained from fits to the worm-like chain model performed for each tether (n=113). Histograms projecting the persistence length and contour length of the model onto the x- and y-axis, respectively, are shown with red lines indicating expected values. The bottom panel shows the force-extension curves of single-tethered DNA data filtered by persistence length and contour length (n=30). Ranges were selected from the peaks of the histograms (one bin-width on either side), yielding filtering ranges of 43-48 nm and 7.8-8.6 μm for the persistence length and contour length, respectively. The overlayed red curve represents the expected force-extension curve. (b) Multiplexed DNA overstretching measured with the benchtop CFM. The top panel shows a representative force-extension curve near the overstretching transition. The overstretching force was extracted as the half-way-point of the two overstretching transition forces and is shown as a vertical grey line between the dashed vertical lines. The
bottom panel shows a histogram of the overstRETChing force measured from a single sample (n=29) with an average and standard deviation of 63.5 ± 1.7 pN.
Figure S4.4: Verification of DNA Unzipping Nanoswitch Construct. (a) Gel electrophoresis of the DNA nanoswitch showing loop formation. Lane i is the 1 kbp extension ladder (Invitrogen, 10511-012), lane ii is a linear construct without the two complementary oligos that close the loop, and lane iii is the nanoswitch construct with the two complementary oligos that can form the looped nanoswitch. The looped DNA migrates more slowly in the gel than the linear construct, resulting in a discrete band with a higher apparent molecular weight as previously observed [175, 178]. (b) Force-extension curve of the DNA unzipping nanoswitch construct measured using optical tweezers. The blue curve corresponds to the looped construct. When forces above ~12 pN were applied, the 29 bp dsDNA that formed the loop unzipped, causing the tether length to increase to the full length of the M13 dsDNA tether. (c) AFM images of the construct. The yellow arrow indicates the putative location of the hybridized DNA zipper. The length of the scale bar is 100 nm.
Figure S4.5: DNA Nanoswitches as Molecular Signatures. (a) Experimental geometry. The centrifuge bucket is at angle, $\theta$, therefore the centrifugal force has components both perpendicular and parallel to the sample surface (x-y plane). When the bond is ruptured, the DNA Nanoswitch goes from looped (blue) to unlooped (red) experiencing a change in length $\Delta L$. This is identified by measuring the projected change in length in the xy plane $\Delta L_{\text{obs}} = \Delta L \cos \theta$. The orientation, $\phi$, of the miniCFM in the bucket, defines the direction of $\Delta L_{\text{obs}}$. (b) Images of a bead before (above), and after (below) a loop opening transition. A bead which is tethered to the surface with a single DNA nanoswitch will undergo a discontinuous change in position with a well-defined length $\Delta L_{\text{obs}}$, and direction $\phi$. (c) A
scatter plot of all contour length changes detected for all directions. Color represents data density. Only transitions within the boxed region are accepted as nanoswitch transitions. Inset, a histogram of transition forces for three different types of transitions: beads which leave the surface (yellow), beads which display discontinuous transition with (green) and without (light-blue) correct direction and magnitude. (d) A scatter plot of the symmetry ratio and root mean square displacement based on the lateral fluctuations of all tracked beads, with color representing data density. The overlayed black data points are for tethered beads that undergo validated nanoswitch transitions.
Figure S4.6: Force and loading rate range of the benchtop CFM. (a) Force as a function of rotational speed calculated using four different types of beads (1 and 2.8 um Dynabeads, and 5 and 10 um silica beads) for the benchtop CFM. Using this set of beads, the CFM is capable of applying a force range that spans eight orders of magnitude (10^4 to 10^8 pN). (b) The fastest and slowest force-loading rates of the benchtop CFM measured using the Thermo Scientific Heraeus X1R Centrifuge. The fastest ramping rate of 317 g/s, shown in blue, can correspond to a 1,300 pN/s loading rate using a large 10 um silica bead. The slowest ramping rate of 0.373 g/s, shown in red, can correspond to a 1.15 fN/s loading rate using a small 1.0 um Dynabead (Invitrogen).
Table S4.1: Parts list of the benchtop CFM including vendor information. The item numbers listed here correspond to the item numbers in Supplementary Figure S4.1. The parts highlighted in the table are discussed in detail in Supplementary Note A.4 and with suggestions for more economical substitutes.

<table>
<thead>
<tr>
<th>ITEM NO.</th>
<th>VENDOR</th>
<th>PART NUMBER</th>
<th>DESCRIPTION</th>
<th>QTY.</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Thorlabs</td>
<td>S1LEDM</td>
<td>SM1-Threaded Mount for LED</td>
<td>1</td>
<td>27.00</td>
</tr>
<tr>
<td>2</td>
<td>Thorlabs</td>
<td>SM1T1</td>
<td>SM1 (1.035”-40) Coupler</td>
<td>1</td>
<td>9.50</td>
</tr>
<tr>
<td>3</td>
<td>Thorlabs</td>
<td>SM05RR</td>
<td>Retaining Ring for diffuser</td>
<td>1</td>
<td>4.00</td>
</tr>
<tr>
<td>4</td>
<td>Thorlabs</td>
<td>DG05-220</td>
<td>Light Source Diffuser, Ø1/2” N-BK7 Ground Glass</td>
<td>1</td>
<td>12.10</td>
</tr>
<tr>
<td>5</td>
<td>Thorlabs</td>
<td>SM1A6T</td>
<td>Diffuser and sample cell mount</td>
<td>1</td>
<td>19.75</td>
</tr>
<tr>
<td>6</td>
<td>SI Howard Glass Co</td>
<td>D265</td>
<td>Ø 25 mm, 0.7 mm Thick</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>Kapton Tape</td>
<td>PPTDE-1</td>
<td>Double Sided tape for Sample Cell Assembly</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>Electron Microscopy</td>
<td>63782-01</td>
<td>Gold Seal, #1 19 mm coverglass</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>Thorlabs</td>
<td>SM1L03</td>
<td>Sample Holder, SM1 Lens Tube, 0.3” Thread Depth</td>
<td>1</td>
<td>12.16</td>
</tr>
<tr>
<td>10</td>
<td>Thorlabs</td>
<td>SM1V05</td>
<td>Focusing Ø1” SM1 Lens Tube</td>
<td>1</td>
<td>29.60</td>
</tr>
<tr>
<td>11</td>
<td>Edmund Optics</td>
<td>#86-815</td>
<td>40X Olympus Plan Achromat Objective, 0.65 NA, 0.6 mm WD</td>
<td>1</td>
<td>575.00</td>
</tr>
<tr>
<td>12</td>
<td>Thorlabs</td>
<td>SM1A3</td>
<td>Objective Adapter with External SM1 Threads and Internal RMS Threads</td>
<td>1</td>
<td>16.75</td>
</tr>
<tr>
<td>13</td>
<td>Thorlabs</td>
<td>AC254-100</td>
<td>Tube Len, f=100.0 mm, Ø1” Achromatic Doublet, ARC: 400-700 nm</td>
<td>1</td>
<td>73.40</td>
</tr>
<tr>
<td>14</td>
<td>Thorlabs</td>
<td>SM1RR</td>
<td>Tube Lens SM1 Retaining Ring</td>
<td>1</td>
<td>4.50</td>
</tr>
<tr>
<td>15</td>
<td>Thorlabs</td>
<td>SM1M20</td>
<td>Objective SM1 Lens Tube Without External Threads, 2” Long Adapter with External SM1 Threads and Internal SM05 Threads,</td>
<td>1</td>
<td>15.50</td>
</tr>
<tr>
<td>16</td>
<td>Thorlabs</td>
<td>SM1A6T</td>
<td>0.40” Thick</td>
<td>2</td>
<td>39.50</td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>-</td>
<td>Custom made turning block, aluminum</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>Thorlabs</td>
<td>PFE10-P01</td>
<td>Turning Mirror, 1” Silver Elliptical Mirror, 450 nm - 20 µm</td>
<td>2</td>
<td>123.60</td>
</tr>
<tr>
<td>19</td>
<td>Thorlabs</td>
<td>SM1NT</td>
<td>Camera SM1 (1.035”-40) Locking Ring, Ø1.25” Outer Diameter Camera Adapter with External C-Mount Threads and Internal SM1</td>
<td>1</td>
<td>6.50</td>
</tr>
<tr>
<td>20</td>
<td>Thorlabs</td>
<td>SM1A9</td>
<td>Threads</td>
<td>1</td>
<td>18.75</td>
</tr>
<tr>
<td>21</td>
<td>Allied Vision Technologies</td>
<td>GC2450</td>
<td>Sony ICX625 CCD sensor, 2448 x 2050 resolution, 15 fps, 12 bit</td>
<td>1</td>
<td>3000.00</td>
</tr>
<tr>
<td>22</td>
<td>IMC Network</td>
<td>855-10734</td>
<td>MiniMc-Gigabit Twisted Pair to Fiber Media Converter</td>
<td>1</td>
<td>260.00</td>
</tr>
<tr>
<td>23</td>
<td>PrinceTel</td>
<td>MJX</td>
<td>Fiber Optic Rotary Joint</td>
<td>1</td>
<td>750.00</td>
</tr>
</tbody>
</table>

Total 4997.61
REFERENCES


