Characterizing Excluded Strand DNA Interactions with Hexameric Helicases and Determining Roles in Unwinding Mechanisms

by

Sean M. Carney

Bachelor of Science, Gwynedd-Mercy University, 2011

Submitted to the Graduate Faculty of the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016
DNA replication is an essential process for all living organisms, and errors in this process can lead to genetic mutations and disease. An assembly of protein machinery, termed the replisome, coordinates enzymatic activities at the replication fork. The DNA helicase is the heart of the replisome, unwinding double-strand DNA at the head of the progressing replisome and providing single-strand templates for DNA polymerases. Replicative helicases are composed of six subunits, and arranged in a ring-like structure where ATP hydrolysis events provide the energy to translocate upon and unwind the DNA. The mechanism of helicase unwinding has been widely studied, but there are still many aspects that remain unknown. It is generally thought that these helicases encircle one strand of DNA while the other is excluded from the central channel of the helicase. Our lab has previously identified an interaction between the excluded strand and the helicase exterior that was important for unwinding in the archaeal MCM helicase. The steric exclusion model of replicative helicase unwinding was expanded to include the excluded strand interactions in this newly proposed steric exclusion and wrapping (SEW) model. Here, we present work that expands on the SEW model by revealing that the bacterial DnaB and mitochondrial Twinkle replicative helicases also interact with the excluded strand. We have also developed a new single-molecule FRET analysis program to characterize these excluded strand interactions. Although the excluded strand interaction is seen in multiple replicative helicases, we propose distinct roles for the interaction based on functional assays and known differences in replisome architecture across the various organisms. We have also begun to characterize the
helicase-excluded strand wrapping interaction in the presence of other replisome components, namely the single-strand binding (SSB) protein. We further characterized the archaeal SSB protein from *Sulfolobus solfataricus* and provide evidence for a novel DNA-helicase-SSB ternary complex. Overall, this thesis makes significant contributions to the understanding of replicative helicase unwinding mechanisms by expanding upon the current steric exclusion and wrapping model and introduces a novel single-molecule FRET analysis program that we anticipate will be adopted and utilized by others in the field.
# TABLE OF CONTENTS

PREFACE.................................................................................................................................................... XVI

1.0 INTRODUCTION....................................................................................................................................... 1

1.1 DNA REPLICATION .............................................................................................................................. 1

1.2 DNA HELICASES................................................................................................................................... 5

1.2.1 Helicase Core Domains: RecA-like and AAA+................................................................................. 6

1.2.2 Superfamilies 1-2 ............................................................................................................................. 8

1.2.3 Superfamilies 3-6: The Hexameric Helicases................................................................................. 10

1.3 HEXAMERIC HELICASE UNWINDING: MECHANISMS AND MODELS......................................................... 11

1.3.1 Hexameric Helicase Structure........................................................................................................ 11

1.3.2 Hexameric NTPase Activity........................................................................................................... 13

1.3.3 Hexameric Helicase Unwinding Models....................................................................................... 15

1.3.4 Impact and Applications............................................................................................................... 17

1.4 THE EXCLUDED STRAND INTERACTION: DISCUSSING IN VIVO ROLES.................................................. 18

1.4.1 The Steric Exclusion and Wrapping (SEW) Model of Unwinding.................................................. 18

1.4.2 Impact of the Excluded Strand in Unwinding............................................................................... 20

1.4.2.1 Excluded Strand as a ‘Molecular Ratchet’.............................................................................. 20
2.4.1.1 Single-Molecule Fluorescence Energy Transfer ........................................ 44
2.4.1.2 Explicit Probability and Rate Transition (ExPRT) Plots ........................................ 47
2.4.2 EcDnaB Interacts with the Excluded Strand ................................................... 51
2.4.3 EcDnaB and SsoMCM Wrap the Excluded Strand Similarly .................................. 54
2.4.4 Conservered Surface Mutants of EcDnaB Alter Excluded Strand Wrapping ................................................................. 60
2.4.5 Surface Mutants Enhance DNA Unwinding Activity ......................................... 66
2.4.6 Basis of K180A and R328A/R329A Unwinding Hyperactivity ............................ 69
2.4.7 ssDNA Translocation of Surface Mutants ......................................................... 72
2.5 DISCUSSION ....................................................................................................... 82
2.5.1 EcDnaB and SsoMCM Interact Similarly with the Excluded Strand ................. 82
2.5.2 Probing EcDnaB Excluded Strand Binding Path via Mutational Analysis ............. 83
2.5.3 Disrupting Excluded Strand Interactions Stimulates EcDnaB Unwinding .......... 84
2.5.4 Comparison with Previous Studies on EcDnaB Unwinding Mechanism ............. 86
2.5.5 Potential In Vivo Roles of Excluded Strand Interactions .................................. 87
2.6 ACKNOWLEDGEMENTS ................................................................................. 88
2.7 FUNDING .......................................................................................................... 89
2.8 CONTRIBUTIONS ............................................................................................ 89
3.0 MITOCHONDRIAL TWINKLE HELICASE EXCLUDED STRAND WRAPPING ................................................................. 90
3.1 SUMMARY ....................................................................................................... 90
3.2  INTRODUCTION ............................................................................................. 90
3.3  MATERIALS AND METHODS............................................................................ 92
  3.3.1  Materials......................................................................................................... 92
  3.3.2  Single-molecule FRET—based DNA binding and unwinding measurements. ............................................................................................................ 93
  3.3.3  Single-molecule FRET data analysis for DNA fork binding and unwinding assays. ........................................................................................................................................... 94
3.4  RESULTS ........................................................................................................... 95
  3.4.1  Twinkle dynamically interacts with both DNA strands......................... 95
  3.4.2  DNA unwinding and reannealing by Twinkle measured by smFRET..... 99
3.5  DISCUSSION................................................................................................... 100
  3.5.1  DNA Unwinding and Reannealing by Twinkle........................................ 101
  3.5.2  Interactions Between Twinkle and the Excluded Strand....................... 101
3.6  ACKNOWLEDGEMENTS................................................................................. 102
3.7  AUTHOR CONTRIBUTIONS............................................................................. 103
4.0  CHARACTERIZING SS OSSB BINDING AND INTERACTION WITH THE HELICASE-DNA FORK COMPLEX................................................................. 104
  4.1  SUMMARY ...................................................................................................... 104
  4.2  INTRODUCTION ........................................................................................... 105
  4.3  MATERIALS AND METHODS........................................................................ 108
    4.3.1  DNA Substrates........................................................................................... 108
    4.3.2  Protein Purifications................................................................................ 108
      4.3.2.1  SsoSSB......................................................................................... 108
4.3.2.2  *SsoMCM* ........................................................................................................ 109
4.3.2.3  *E.coli SSB* .................................................................................................. 109
4.3.2.4  T4 gp32 ........................................................................................................ 110

4.3.3  Atomic Force Microscopy ............................................................................... 110
4.3.4  Fluorescence Anisotropy .............................................................................. 112
4.3.5  Single-Molecule FRET Data Collection ................................................... 112
4.3.6  Single-Molecule FRET Data Analysis ....................................................... 113

4.4  RESULTS ............................................................................................................. 114
4.4.1  Determining the Oligomeric State of Free *SsoSSB* ................................... 114
4.4.2  Determining the Binding Affinity of SSB Proteins................................. 116
4.4.3  Probing the Binding Modes of SSBs to ssDNA using smFRET. .............. 118
4.4.4  Characterizing ssDNA Binding Mode of *SsoSSB* ................................... 120
4.4.5  Probing the Binding Modes of *SsoSSB* to fork DNA using smFRET.... 123
4.4.6  Identifying a DNA-Helicase-SSB Tertiary Complex ............................... 128

4.5  DISCUSSION ...................................................................................................... 133
4.5.1  Atomic Force Microscopy Study of *SsoSSB* Oligomeric State ............. 133
4.5.2  Determining the Basis of Varying *SsoSSB* ssDNA Binding Affinities.... 134
4.5.3  The Novel Binding Mode of *SsoSSB* ...................................................... 136
4.5.4  *SsoSSB* Fork Binding. ............................................................................... 139
4.5.5  Possible DNA-Helicase-SSB Tertiary Complexes................................. 140
4.5.6  Fork Binding Assays in the Context of Helicase Function....................... 142

4.6  CONCLUSION .................................................................................................. 143
4.7  CONTRIBUTIONS ............................................................................................. 144
LIST OF TABLES

Table 1-1: DNA Replication Proteins Across the Three Domains of Life ........................................ 2
Table 2-1: DNA Sequences .............................................................................................................. 34
Table 2-2: Comparison of Methods Used to Validate the Hexameric Helicase SEW Model ...... 43
Table 2-3: EcDnaB Kinetic Parameters ........................................................................................... 69
Table 2-4: EcDnaB ssDNA Translocation ...................................................................................... 80
Table 3-1: DNA Substrate Sequences ............................................................................................. 93
Table 4-1: DNA Substrates .............................................................................................................. 108
Table A-1: DNA Sequences ........................................................................................................... 154
LIST OF FIGURES

Figure 1-1: Cartoon Model of *E.coli* Replisome. ........................................................................... 4
Figure 1-2: RecA and AAA⁺ Folds................................................................................................. 7
Figure 1-3: Classification of helicases and translocases................................................................. 9
Figure 1-4: Structural Conservation of Hexameric Helicases ...................................................... 12
Figure 1-5: Models of NTP Hydrolysis by Hexameric Helicases. ................................................. 14
Figure 1-6: Models of DNA Unwinding........................................................................................ 19
Figure 1-7: SEW models showing the impact of the excluded strand on unwinding.................. 30
Figure 2-1: SDS-PAGE Analysis of *Ec*DnaB Purifications. ........................................................ 36
Figure 2-2: smFRET Experimental Setup..................................................................................... 45
Figure 2-3: smFRET Analysis of WT and Mutant *Sso*SSB bound to DNA............................... 49
Figure 2-4: Single-molecule FRET monitoring of *Ec*DnaB binding to DNA fork substrates...... 52
Figure 2-5: Titration of *Ec*DnaB onto 30/30 Fork ................................................................. 53
Figure 2-6: Comparison of the similar excluded strand interactions of *Sso*MCM and *Ec*DnaB  55
Figure 2-7: ExPRT plots of *Ec*DnaB and *Sso*MCM on DNA. ................................................. 56
Figure 2-8: Comparison of smFRET analysis methods............................................................... 57
Figure 2-9: Dwell Time Survival Analysis of *Ec*DnaB on 40/30................................................. 58
Figure 2-10: Amino acid sequence alignments of DnaB helicases............................................. 60
Figure 4-4: SSB Binding.................................................................................................................. 118
Figure 4-5: smFRET Binding Studies of EcSSB and T4 gp32.................................................. 120
Figure 4-6: smFRET ssDNA Binding Assays. .............................................................................. 122
Figure 4-7: SsoSSB Binding smFRET Traces.............................................................................. 123
Figure 4-8: SsoSSB Titration onto DNA Fork Substrate.......................................................... 124
Figure 4-9: ExPRT Analysis of SsoSSB Titration onto 30/30 Fork........................................... 126
Figure 4-10: ExPRT Plots of SsoSSB Titration onto the 30/50 Fork......................................... 127
Figure 4-11: Titration of SsoSSB onto Helicase - DNA Complex........................................... 129
Figure 4-12: ExPRT Analysis of SsoSSB onto the SsoMCM - 30/30 Complex....................... 131
Figure 4-13: ExPRT Analysis of SsoSSB Titrated onto the SsoMCM - 30/50 Complex........ 132
Figure 4-14: Binding Constant vs. DNA Length................................................................. 136
Figure 4-15: Cartoon Models of ssDNA Binding Modes......................................................... 138
Figure A-1: Cocrystals of hexameric helicases and nucleic acid substrates..................... 153
Figure A-2: Helicases induced compression of the encircled strand................................. 158
Figure A-3: Rise per base measured by smFRET and crystallography............................. 159
This thesis is the result of several years of work on projects that I have grown to feel some ownership of, but the credit for this work belongs to many.

The members of my committee have dedicated many hours to the forward progression of these projects as well as my own growth as a scientist. I am grateful for the valuable constructive feedback they have provided over the last few years. In particular, I would like to thank Michael Trakselis, who invited me into his lab and provided a fertile environment to learn and to grow as a scientist and as a person. Michael has been an excellent example of a rigorous and productive scientist as well as a thoughtful and supportive mentor. His excitement for the projects in the lab is contagious, and I am grateful that he has pushed me to think more creatively about answering scientific questions. I would also like to thank Sanford Leuba, a collaborator turned mentor who welcomed me into his laboratory and who has provided very valuable advice and the sometimes hard to hear critiques that have allowed me to develop a greater understanding of and appreciation for science and to mature as a person. Both of my mentors have taken the care to introduce me to the greater scientific community, which has been one of the most exciting and cherished aspects of my graduate career. They also gave me the freedom to explore in the laboratory. This led to mistakes, failures, disappointments, learning, discoveries, and understanding. I want to thank them for having confidence in me.
The work presented here would not have been possible without the help and support of several labs. The labs of Ben Van Houten, Alex Deiters, and Seth Childers have been extremely generous with their time, resources, and expertise. I was fortunate enough to work with great collaborators on these projects including Bob Brosh and his group at the NIH and Ben Van Houten and Muwen Kong at the University of Pittsburgh. I am also indebted to Trakselis and Leuba lab members, particularly Brian Graham, Rob Bauer, and Beth Jeffries who were a pleasure to work with and all great biochemists and molecular biologists to learn from, and to Heather McFarland who made significant contributions to the DnaB project. Also, hours of data collection in the darkroom would have been unbearable without Grant Schauer, who was patient while I learned and who was always a source of good conversation. I would also like to thank the Molecular Biophysics and Structural Biology Program as well as all the faculty and staff involved. My fellow classmates and MBSB students have also played a pivotal role. I would like to thank Muwen Kong and Abhishek Mandal in particular, who have acted as my tutors, gadget gurus, and most importantly, my good friends. Thank you to all my friends who have shaped my life in ways that I am truly grateful for, including Brian Eichler and Jon Bumstead among others.

I am very fortunate to have had great teachers and mentors throughout my life, including James Ferris, Michelle Kulp McEliece, Felicia Cosaro-Barbieri, John Pascal, and Craig Peebles. Their teaching, advice, and examples have brought me to this point, and will continue to serve as sources of inspiration and guidance throughout my life.

All of the opportunities I have had throughout my life I owe to my parents, Mary Ann and Dennis, who provided for my education, taught each of their children to perform to the best of their abilities regardless of the task, and encouraged us to work relentlessly when pursuing an endeavor, even in the face of disappointment and defeat. To Mom and Dad, I love you. Thank
you for all of your love and support. To my brothers, Bryan and Irv, and sister, Clare, thank you
for being a constant source of enjoyment and support in my life. I love you. To my maternal
grandparents, Carol and Eugene Doyle, and paternal grandparents, Ellen and Jack Carney, thank
you for providing such strong examples of sacrifice and love. I love you.

Lastly and most importantly, I want to thank Caitlin, who has been my greatest source of
support, encouragement, and inspiration over the past several years. She is truly an amazing
person, and I am incredibly grateful that she is in my life. Caitlin, your unwaivering love is the
source of my happiness, and your generosity, patience, and kindness have allowed me to
persevere. I love you.

As a symbol of my appreciation, I dedicate this work to all of those mentioned above as
well as others who have contributed to my development and the progression of these projects.
1.0 INTRODUCTION

1.1 DNA REPLICATION

Living organisms require the transmission of genetic material from one cell to another when they grow, divide and reproduce. Several studies conducted during the mid-twentieth century demonstrated that DNA is the molecule responsible for carrying and transmitting genetic information from one cell or organism to another [1-3]. Around the same time, the structure of DNA was determined [4-7], and the model proposed by James Watson and Francis Crick provided important explanations for how DNA can act as the genetic material within the cell [8, 9]. The structure consists of the nucleotide bases adenine (A) and cytosine (C) that hydrogen bond with thymine (T) and guanine (G) respectively, within the context of an antiparallel double helical sugar phosphate backbone. Watson and Crick published a follow-up paper to their initial report of the double helical structure of DNA where they provided additional commentary and discussion of the model in the context of DNA being the genetic material of cells [8]. They note that the double-stranded DNA (dsDNA) likely must be unwound before replication can occur and a new complimentary strand is polymerized against an existing template strand whose nucleotide bases are exposed and available for hydrogen bonding. They speculate about potential

---

1 The material from section 1.4 was reprinted/adapted from Methods, In press, S.M. Carney, M. A. Trakselis, The excluded DNA strand is SEW important for hexameric helicase unwinding. (2016), with permission from Elsevier. Reprinted from The Lancet, In press, S.M. Carney, M. A. Trakselis, The excluded DNA strand is SEW important for hexameric helicase unwinding. (2016), with permission from Elsevier.
protein involvement in dsDNA unwinding and subsequent polymerization. Since then, more than sixty years of research has confirmed many of their predictions, including a dsDNA unwinding enzyme, now known as the helicase.

Organisms across all domains of life require helicases to unwind the dsDNA as well as many other conserved components that mediate DNA replication. The core components of the replisome from organisms across the three domains of life are listed in Table 1.1. In general, DNA replication must be initiated at one or several origin sites within the genome by origin recognition proteins. These proteins and complexes interact with the replication origin site in a way that promotes the recruitment of helicase loading proteins and helicases [10-12]. Replicative helicases are ring-shaped hexamers that must be loaded onto or assembled around DNA before becoming active and separating the DNA strands [13-15]. Loading proteins mediate the loading and assembly of hexameric helicases at the origin of replication [10, 12, 16]. An array of replication proteins are then subsequently recruited to and assemble with the now loaded helicase to form the active replisome complex.

Table 1-1: DNA Replication Proteins Across the Three Domains of Life.

<table>
<thead>
<tr>
<th>Replication Role or Factor</th>
<th>Bacteria: E.coli</th>
<th>Archaea: Sulfolobus solfataricus</th>
<th>Eukaryota: Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin Binding</td>
<td>DnaA</td>
<td>Orc1/Cdc6</td>
<td>ORC</td>
</tr>
<tr>
<td>Helicase Loader</td>
<td>DnaC</td>
<td>Orc1/Cdc6</td>
<td>Cdc6, Cdt1</td>
</tr>
<tr>
<td>Helicase</td>
<td>DnaB</td>
<td>MCM</td>
<td>CMG (Cdc45, MCM2-7, GINS)</td>
</tr>
<tr>
<td>Primase</td>
<td>DnaG</td>
<td>DnaG, PriSL</td>
<td>PriSL - pol α complex</td>
</tr>
<tr>
<td>Polymerase</td>
<td>pol III</td>
<td>B-family polymerases</td>
<td>Pol δ, Pol ε</td>
</tr>
<tr>
<td>Clamp</td>
<td>β-clamp</td>
<td>PCNA</td>
<td>PCNA</td>
</tr>
<tr>
<td>Clamp Loader</td>
<td>γ-complex</td>
<td>RFC</td>
<td>RFC</td>
</tr>
<tr>
<td>Single-strand Binding Protein</td>
<td>SSB</td>
<td>SSB, RPA</td>
<td>RPA</td>
</tr>
<tr>
<td>RNA Primer Processing</td>
<td>RNase H, pol I</td>
<td>RNase H, FEN1</td>
<td>RNase H, FEN1, Dna2</td>
</tr>
<tr>
<td>Ligase</td>
<td>DNA Ligase</td>
<td>DNA Ligase</td>
<td>DNA Ligase</td>
</tr>
</tbody>
</table>
Once the helicase is loaded, one of the first members of the replisome that is recruited is the DNA primase. The primase must synthesize an RNA primer on the leading strand because DNA polymerases generally cannot synthesize DNA *de novo*. They require a primer several ribonucleotides in length to initiate chain elongation [17, 18]. Once an RNA primer has been synthesized on the template strand, the DNA polymerase extends the nucleotide primer chain opposite the template strand in the 5’- 3’ direction with deoxynucleotides [19]. These DNA polymerases are associated with ring-shaped processivity clamps that slide on DNA to increase the speed and processivity of the DNA polymerase’s activity. Similar to helicases, these processivity clamps also require specific factors or complexes to open and localize them to the correct sites within the replisome [20]. Leading strand replication as described above is somewhat simple compared to the events that occur to mediate lagging strand replication.

Because DNA polymerases only elongate new strands in the 5’- 3’ direction, synthesis on the lagging strand is discontinuous. Lagging strand synthesis occurs by producing discontinuous segments of ssDNA called Okazaki fragments [21, 22]. This places additional requirements on the processing of products on the lagging strand. Similar to the leading strand, the lagging strand must be primed before DNA polymerase can act on it. After an RNA primer is synthesized on the lagging strand, the DNA polymerase elongates in the opposite direction that the helicase and replication fork are moving [23]. This leaves intermittent segments of ssDNA that are vulnerable to nuclease activity and unwanted interactions. Single-strand binding proteins (SSBs) bind to and protect regions of ssDNA during lagging strand replication until a DNA polymerase can generate a complimentary strand to the lagging strand template [18, 23]. Because the primers consist of RNA, they must be replaced with DNA during processing events that occur after initial Okazaki fragment synthesis. Nucleases will remove these intermittent stretches of RNA that now exist,
and a DNA polymerase will fill in these gaps [19, 23]. To complete Okazaki fragment maturation, a DNA ligase covalently links the DNA that replaced the RNA primer to the adjacent DNA chain that was originally synthesized by DNA polymerase as an extension of the RNA primer [18, 23]. A cartoon model of the *E.coli* replisome progressing along a DNA fork is shown in Figure 1.1. While this figure may provide an idea of general replisome architecture, it should be noted that events at the replication fork are dynamic and involve many transiently associated factors in addition to those members of the replisome listed [24-26].

![Cartoon Model of E.coli Replisome.](image)
Although all of the members of the replisome described above are critical for efficient DNA replication in all domains of life, the focus of this thesis is primarily on the mechanism of DNA helicases. The hexameric helicase can be considered to be the heart of the replisome, being the first member recruited to the replication origin and the enzyme largely responsible for the forward progression and speed of the replication fork. The structure, function, and mechanism of helicases’ actions are of interest because of the evolutionary insight they provide, their potential exploitation for diagnostics and therapeutics, and their potential use in biotechnology applications.

## 1.2 DNA HELICASES

In the 1970s, some of the first helicases identified were TraI and DnaB helicases from *E.coli* [27-30]. Since then, hundreds of RNA and DNA helicase have been identified and characterized from all domains of life as well as viruses. They are ubiquitous and conserved throughout the bacteria, archaea, and eukaryotes, and their study has provided valuable evolutionary insights [31-35]. DNA helicases are critical for most aspects of nucleic acid metabolism, and the mutation or loss of helicase function often results in genome instability [36-39].

Helicases can be defined as proteins containing conserved signature motifs that mediate the binding and hydrolysis of nucleotide triphosphates (NTPs), which provides the mechanical power required to unwind or translocate upon a nucleic acid substrate [15, 40, 41]. As increasing numbers of helicases are being identified and characterized, they are organized into...
superfamilies based on the identity and positioning of the conserved helicase motifs, their unwinding polarities, translocation specificities, and their oligomeric states.

1.2.1 Helicase Core Domains: RecA-like and AAA⁺

Helicases couple the hydrolysis of NTPs to dsDNA unwinding or translocation. All helicases contain an ASCE (Additional Strand Catalytic E) core domain, which falls into the P-loop class of NTPases. The distinguishing feature of the ASCE group is that they contain an additional strand between the Walker A and Walker B motifs [15, 42]. The Walker A motif \([G-X_4-GK(S/T)]\) interacts with the phosphate group of ATP via a lysine residue on a glycine-rich loop. The Walker B motif \([RK-X_4-G-X_4—LhhhhD]\) contains a conserved aspartate required for coordinating a divalent cation. Additionally, each ASCE fold domain contains a conserved polar residue at the end of strand 4 (Sensor-1), which is situated between the Walker A and Walker B motifs. This polar residue is thought to act as a sensor of the nucleotide state [15, 43, 44]. Two other features of ASCE domains are a conserved glutamate residue responsible for activating a water molecule to generate a nucleophile for the hydrolysis reaction and an arginine finger that often interacts with the phosphate of the nucleotide bound at an adjacent site [15, 43, 44]. However, the positioning of these two conserved residues differs among ASCE domain folds found throughout nature as shown in Figure 1.2.
Figure 1-2: RecA and AAA+ Folds
Schematics of the P-loop ATPases. RecA and AAA+ folds are shown in the top and bottom respectively. Beta sheets are shown in blue and numbered while alpha helices are shown in orange and lettered. The positions of phosphate sensors (purple), Walker A (green), Walker B (yellow), arginine finger (blue), and the catalytic glutamate residue (red) are highlighted in each schematic.

Subdivisions of the ASCE group of P-loop NTPases include the RecA-like and AAA+ NTPases. Within the RecA-like group, the conserved catalytic glutamate is between strand 2 and an additional α-helix and β-strand that are absent in AAA+. The arginine finger in RecA-like ASCE domains is located at the end of an additional β-hairpin at the C-terminal end of the fold. AAA+ NTPase folds contain the catalytic glutamate within the Walker B motif, and the arginine finger is situated between α-helix D and β-strand 5. These domains also contain an additional α-helix at the N-terminus of the fold. At the C-terminus of the core ASCE folds of AAA+ domains is small helical domain, which contains a second sensor motif (Sensor-2) involved in binding to nucleotide and coupling its state to the protein’s conformation [15, 43, 45]. These two motifs play a large role in determining evolutionary relationships among helicases as well as in organizing helicases into superfamilies.
1.2.2 Superfamilies 1-2

Helicases in superfamily 1 (SF1) and superfamily 2 (SF2) make up the majority of helicases found in nature. SF1 and SF2 helicases act as either monomers or dimers, and each monomeric subunit contains two RecA-like folds as shown in Figure 1.3A-B [42]. SF1 helicases can be further divided into SF1A and SF1B based on the helicase polarity being either 3’- 5’ or 5’- 3’ respectively. These different polarities are represented in Figure 1.3D, along with the α and β distinctions that correspond to translocation on single or double-stranded DNA respectively. Representative members of SF1A are Rep, PcrA, and UvrD helicases. Members of SF2 include RecD, NS3, and Dda. SF1 helicases participate in a wide variety of cellular processes such as DNA repair, replication, recombination, telomere maintenance, and Okazaki fragment processing among others [46]. SF1 helicases interact with DNA substrates largely using base-stacking interactions. However, due to differences in motifs III and IV of the helicase core domain, SF2 helicases utilize electrostatic contacts with the phosphodiester backbone of DNA to mediate interactions. Motif III is involved in DNA binding in SF1 helicases, but not in SF2 helicases. Motif IV makes contacts with the DNA backbone only in SF2 helicases [46-49]. SF2 helicases include those from the subfamilies of RNA helicases [50] and RecQ helicases [51]. Like SF1 helicases, those from SF2 helicases participate in a wide variety of activities, and contain members that translocate from both 3’- 5’ and 5’- 3’ [42, 49].
Figure 1-3: Classification of helicases and translocases.
Reprinted from [42]. A) This classification is based largely on the work of Gorbalenya & Koonin [40]. The name of one member of each of the six superfamilies, which is used as a structural example in the text, is given in parentheses. The “core domains” and the positions of the signature motifs therein are shown for each class of helicase. Note that the precise position of each motif is based on the example family member and is representative for the whole family. Motifs colored yellow represent universal structural elements in all helicases. The positions and functions of accessory domains in each example protein are also shown, but in contrast to the core domains, these are specific to each protein, and their presence, function, and precise location within different members of the same superfamily vary widely. (B, C) Representative core structures. Universal structural elements involved in the binding and hydrolysis of NTP, and the coupling of this activity to conformational changes are shown in yellow. B) The SF1 and SF2 enzymes contain a monomeric core formed from the tandem repeat of a RecA-like fold. The N- and C-terminal RecA-like domains are shown. An NTP analogue (black) is bound at the interface of the core domains. Motifs 1 and 2, related to the Walker A and B motifs, are located on the N-core side of the cleft. Motif 6, which contains an arginine finger residue, is contributed by the C-core domain. This representative structure is the core of PcrA helicase from SF1. Note that these core domains constitute the minimal translocation motor. C) SF3-6 enzymes contain a core that consists of six individual RecA- or AAA'-like domains (red) arranged in a ring. Six nucleotide-binding pockets are present, one at each domain interface, and four are occupied with NTP analogues (black). As in the SF1/SF2 enzymes, conserved elements for the binding and hydrolysis of NTP related to the Walker A and B motifs are located on the opposite side of the cleft compared to the conserved arginine finger residues. This representative structure is of T7 gene 4 protein from SF4. D) Nomenclature for subfamilies is based on translocation directionality [3'-5' (A) or 5'-3' (B)] and whether the nucleic acid substrate is single (α) or double stranded (β). The strand along which translocation takes place is depicted in purple. Ribbon diagrams in this and subsequent figures were created with PyMOL (http://pymol.sf.net) unless stated otherwise.
1.2.3 Superfamilies 3-6: The Hexameric Helicases

While SF1 and SF2 helicases are either monomeric or dimeric helicases, SF3-6 are hexameric helicases, consisting of six units that make up a functional enzyme as shown in Figure 1.3C. Helicases from all four of these superfamilies form toroidal rings that encircle one of both DNA strands upon loading. As many of these hexamers are involved in the replication of entire genomes, this mode of interaction is generally thought to enhance processivity. SF3 and SF6 helicases contain AAA⁺ folds and translocate 3’- 5’ [15, 42]. SF3 helicases include the hexameric helicase found in the genomes of many small DNA and RNA viruses. Well-studied SF3 members include LTag and E1 helicases from the simian virus 40 (SV40) and the papilloma virus respectively [15, 52-54]. Features specific to SF3 helicases include the addition of a C-motif in addition to the Walker A and B motifs that exist in all helicases. Also, SF3 helicases contain a ori DNA-binding domain, as many of these helicases are responsible for origin recognition in addition to dsDNA unwinding [42]. SF6 helicases also translocate 3’- 5’ and contain AAA⁺ folds that mediate the binding and hydrolysis of NTPs. The eukaryotic MCM2-7 and archaeal MCM replicative helicases belong to SF6 [55, 56]. However the eukaryotic MCM2-7 requires additional factors, Cdc45 and GINS, and posttranslational modifications to catalyze dsDNA unwinding [57], which is why it is also referred to as the CMG helicase (Cdc45, MCM2-7, GINS helicase). The prokaryotic RuvB dsDNA translocase is also a member of SF6. RuvB is involved in processing holiday junctions created during recombination [15, 42, 58, 59].

Unlike SF3 and SF6, SF4 and SF5 helicases contain RecA-like core domains that bind and hydrolyze NTPs to provide the mechanical work needed to translocate 5’- 3’ during dsDNA unwinding. SF4 consists of bacteriophage replicative helicases, such as T7 gp4 and T4 gp41, as well as the bacterial DnaB and DnaB-like helicases [15, 42]. These are some of the most widely
studied helicases largely due to their role within model systems used to study DNA replication. The mitochondrial helicase, Twinkle, also belongs to SF4, and its mechanism is of great interest due to several genetic disorders caused by mutations in the gene coding for Twinkle [60-62]. It is common for SF4 helicases to be strongly associated with primases, and in some cases they are expressed as a fused polypeptide [63, 64]. All SF4 helicases contain five conserved motifs, and although the Rho helicase is closely related, it has been classified as a SF5 helicase. Rho is the bacterial transcription terminator, and the difference in classification is based on the significantly different sequence of Rho [15, 42].

1.3 HEXAMERIC HELICASE UNWINDING: MECHANISMS AND MODELS

1.3.1 Hexameric Helicase Structure

High-resolution X-ray and electron microscopy structures have provided valuable insights into hexameric helicase binding and unwinding mechanisms. Despite observations of quaternary states other than hexamers, such as heptamers, helical filaments, and double hexamers [63, 65-67], these helicases display ring-shaped hexameric symmetry when complexed with DNA substrates [52, 68-70]. In most cases, assembly of the hexamers from six monomers or three dimers requires the presence of Mg$^{2+}$ and in some cases the addition of NTPs [71-74]. Each monomer within the hexamer consists of a N-terminal domain and a C-terminal domain. Structural data has shown that these hexamers are generally organized into two tiers, the N-terminal and C-terminal tier. In all cases except E1, the C-terminal domain is positioned towards the duplex side of the fork, while the N-terminal domain is positioned away from the duplex. These features are
highlighted in Figure 1.4 with several examples of structures from various superfamilies [14, 52, 53, 63, 70, 75, 76]. The N-terminal domains of many helicases have been shown to mediate critical interactions with primases and other members of the replisome [77, 78].

Figure 1-4: Structural Conservation of Hexameric Helicases
Reprinted from [14]. Hexameric helicases are shown from different domains (RecA or AAA+) and superfamilies (SFs) with associated unwinding polarity and references. View from the N-terminal domain (NTD) rotated 90° to visualize the lateral length from the C-terminal domain (CTD) to the NTD.

Another feature of these hexameric structures worth noting is the vertical position of the subunits in relation to one another within the ring. Some of the observed structures display a planar hexamer ring, while others form a spiral or staircase structure. A possible reason for this difference is their mechanism of binding to and translocating along DNA. Hexameric helicases encircle and bind to their substrates through nucleic acid binding loops. Based on the co-crystal structures of Rho and E1 with their nucleic acid substrates, it is proposed that cycles of dNTP binding, hydrolysis, and release cause the binding loops to change position vertically within the central channel of the hexamer [52, 68]. However, the global structure of the hexamer remains largely planar throughout this process. A different mechanism of binding and translocating has been proposed for the bacterial DnaB helicase. The structure shown in Figure 1.4 displays a spiral staircase structure, where subunits are raised vertically above the previous subunit as
organized around the hexamer [70]. The authors note that the DNA binding loops responsible for interacting with the substrate within the central channel are significantly shorter and likely less mobile than those reported for E1 and Rho. Therefore, they proposed that DnaB relies on the vertical migration of entire subunits of the hexamer as the helicase tracks along DNA during rounds of NTP binding, hydrolysis, and release. This mechanism would explain the spiral staircase structure they observe. However the dynamics and kinetic details of these processes need to be characterized. It is also possible that unwinding mechanisms may involve a hybrid between the two discussed above.

1.3.2 Hexameric NTPase Activity

The driving force behind translocation and unwinding by any helicase is the hydrolysis of NTPs. The hexameric helicases contain their signature folds, including the Walker A and Walker B motifs, at the interfaces between adjacent subunits. One subunit contributes the glycine-rich P-loop, conserved polar residue, conserved acidic group, while the neighboring subunit contributes the arginine finger to that site. This coupling between subunits at the NTP binding and hydrolysis site allows for the coupling of NTP hydrolysis to conformational change across subunits [15, 79].
There has been some discussion concerning the coordination of NTP hydrolysis by the six sites within each hexamer. Several models have been put forward including a concerted model (Figure 1.5A) where all sites hydrolyze their bound NTPs simultaneously. A stochastic model (Figure 1.5B) where there is no coordination of hydrolysis between sites has also been proposed. Two rotary models have also been considered. The first is the symmetric rotary model (Figure 1.5C) where the state of the NTP sites around the hexamer would be empty, ADP+Pi, ATP, empty, ADP+Pi, ATP. The asymmetric rotary model (Figure 1.5D) has one site as the active hydrolysis center at any one time, and after the first site fires, the neighboring subunit would fire next, going around the ring hydrolyzing one site at a time [15, 80]. The structures of El and Rho bound to their nucleic acid substrates display radial asymmetry of NTPase sites. Along with the positioning of each nucleic acid-binding loop at a unique vertical position to
create a spiral staircase structure in the presence of a nucleic acid substrate, this supports the asymmetric rotary model of dNTP hydrolysis [52, 68, 81]. The rotary model of dNTP hydrolysis by hexameric helicases also plays a role in determining the translocation polarity of the helicase. E1 is a SF3 3’- 5’ helicase while Rho belongs to SF5 and translocates 5’- 3’. Despite these differences, they both bind to DNA forks with the same orientation with respect to their N- and C-terminal domains. Their opposing translocation polarity has been attributed to their asymmetric rotary hydrolyses that fire in opposing directions [15, 80], a concept which is supported by the opposing asymmetry of NTPase sites observed in the crystal structures of each [52, 68].

1.3.3 Hexameric Helicase Unwinding Models

Helicases use NTP hydrolysis to couple conformational changes to translocation and separation of nucleic acid substrates. Studies on a wide range of hexameric helicases have given rise to several models of how hexamers effectively separate dsDNA or dsRNA [82]. A majority of these models were aimed at explaining the unwinding mechanism of the eukaryotic MCM2-7 [82]. These models of unwinding included a rotary pump model, where two groups of MCM2-7 complexes load at sites distant from the replication fork and coordinately pump dsDNA in opposing directions towards this point, and the severely underwound DNA would cause strand separation [83]. Another proposed model was based on work done on the LTag helicase from the SV40 virus. This model suggests a double hexamer as the functional unit, and the two hexamers pump dsDNA into their channels from opposing directions and the resulting single-strands are extruded through side channels in the double hexamer structure [84-90]. The ploughshare model is proposed to work by the hexamer translocating along dsDNA and a region of the hexamer or
another interacting protein separates the dsDNA upon its exit from the inner channel [82]. However, continual studies point towards a general steric exclusion mechanism of dsDNA separation for hexameric helicases, including LTag and MCM2-7 [14, 91, 92]. The steric exclusion model of dsDNA unwinding places one ssDNA within the central channel of the hexamer ring, as seen in the co-crystal structures of Rho, E1, and DnaB [52, 68, 70]. The other strand of ssDNA is excluded from the central channel of the hexamer. The helicases’ ability to translocate and pull on the encircled strand results in the dsDNA being separated into two strands since the excluded strand does not have access to the central channel.

Most studies of helicases and unwinding mechanisms have focused on the interaction between the hexamer and the encircled strand that the helicase translocates upon. However, several studies have suggested that an interaction between the outer surface of the helicase and the excluded strand may play a role in the loading or unwinding activities of the helicase [93-95]. Recently, the Trakselis group identified binding sites for the excluded strand on the outer surface of the archaeal MCM helicase from *Sulfolobus solfataricus* [96, 97]. This interaction was demonstrated using hydrogen-deuterium exchange, nuclease protection assays, and single-molecule FRET. Mutants that disrupted this interaction were created, and these mutants showed significant deficiencies in unwinding activity. Based on the results, our group proposed the steric exclusion and wrapping (SEW) model of hexameric helicase unwinding. This model amends the widely accepted steric exclusion model to include the contributions of interactions with the excluded strand to unwinding activity, which were previously largely ignored.
1.3.4 Impact and Applications

Helicases and their mechanisms of unwinding are widely studied, and due to their importance in a wide variety of cellular processes, multiple helicases have been pursued as diagnostic markers or therapeutic drug targets. Human MCM proteins have been used as markers of several cancer types that express higher levels of these helicase proteins [98-100]. These efforts are directed towards less invasive and more sensitive detection of cancers. Inhibitors of the eukaryotic MCM2-7 helicase are currently being developed and studied [101]. Such inhibitors might be used as therapeutics for treating cancers, where cells are replicating much more rapidly without normal regulation. Human RecQ helicases are also being targeted as potential cancer treatment strategies [102, 103]. These helicases are heavily involved in DNA repair and maintenance, and inhibiting them within cells that are already deficient in some DNA repair or maintenance process has a synthetic lethality effect [103, 104]. Since bacteria also require helicase activity to grow and divide, there has been an effort to develop inhibitors of the bacterial replicative helicase, DnaB, that would act as antibiotics [103, 105]. Similarly, viral helicases have been targeted with inhibitors designed as antiviral agents. Inhibitors of the helicase-primase complex of the herpesvirus have entered clinical trials, and are a promising new type of antiviral [106-108]. New antibacterial and antiviral targets are constantly being pursued due to the appearance of drug-resistant mutations, and helicases remain a largely untapped target. Helicases are also applicable to biotechnologies, such as nanopore sequencing, that require nucleic acid translocases [109, 110]. It is important to continue studying the functions and mechanisms of helicases, as they will likely continue to be targeted for diagnostics and therapeutics and utilized in biotechnology applications and are of fundamental importance to all forms of life.
1.4 THE EXCLUDED STRAND INTERACTION: DISCUSSING IN VIVO ROLES

1.4.1 The Steric Exclusion and Wrapping (SEW) Model of Unwinding.

The steric exclusion (SE) model of unwinding has been accepted for decades to explain the unwinding action of not only hexameric helicases, but also of monomeric and dimeric helicases. One limitation of the SE model is that it generally ignores any contribution of the excluded strand in the unwinding mechanism. Interactions with the excluded strand have been shown previously with hexameric helicases [111], but it was only recently that their role in the DNA unwinding mechanism has been revealed [96, 97]. Using MCM helicase from *Sulfolobus solfataricus*, our group identified an interaction between the excluded strand and the helicase surface. Mutational analysis showed that this interaction is critical for unwinding. Based on these findings, we proposed the steric exclusion and wrapping (SEW) model of unwinding, where the SE model is expanded to include contributions from the excluded strand to unwinding activities. A cartoon representation of these two models is displayed in Figure 1.6. The conservation of this mechanism is now of interest and currently being investigated.
Figure 1-6: Models of DNA Unwinding.
Structural models for hexameric helicase DNA unwinding include steric exclusion (SE) or steric exclusion and wrapping (SEW) where the nontranslocating strand makes contact with the exterior surface (orange).

Although, it has been reported previously that the ssDNA is bound in the central channel and not wrapped around the DnaB hexamer [112, 113], we hypothesize that external surface binding of ssDNA is not thermodynamically stable when the primary central binding site is available. Only after encircling the 5’-strand of fork DNA would the excluded 3’-strand be conformationally favored for exterior surface binding. In support, binding of a second ssDNA strand to T7 gp4 and EcDnaB helicases has been measured, but with lower affinity [93, 114]. We expect the SEW model to be conserved across a wider range of helicases, and for the excluded strand to play a role in the loading and unwinding activities of a variety of helicases. In this section I will discuss the potential roles that interactions with the excluded or non-translocating strand could play in a cellular context based on current findings within the literature.
1.4.2 Impact of the Excluded Strand in Unwinding

1.4.2.1 Excluded Strand as a ‘Molecular Ratchet’

Our group and others have found that the excluded (nontranslocating) strand interacts with the exterior surface of a number of helicases. The remaining question is why? How does the excluded strand contribute to the dsDNA unwinding mechanism? Generally, we have found that surface-exposed and conserved positively charged residues define a path on the exterior surface [96, 97, 115]. Mutation of individual residues alters the DNA binding path and perturbs DNA unwinding. Specifically for SsoMCM, a 3’-5’ helicase, mutation of positively charged residues along the proposed binding path reduces the unwinding efficiency 2-10 fold. For this helicase, loss of contact with the excluded strand makes dsDNA unwinding more futile. We hypothesize that maintaining a grip on both ssDNA strands allows for better destabilization of the hydrogen bonding within the duplex for efficient unwinding. Specific contacts within the central channel will propel the helicase forward along DNA, coupled with ATP hydrolysis steps [116-119]. However, when contact with the excluded strand is lost, the helicase can slip backwards after an ATP hydrolysis cycle, effectively rezipping the duplex region, and ineffectively unwinding DNA. Therefore, the excluded strand interaction acts as a ‘molecular ratchet’ to promote SsoMCM unwinding in a forward direction and prevent slippage backwards (Figure 1.7A).

1.4.2.2 Excluded Strand as a ‘Molecular Brake’

Similarly, exterior excluded strand contacts are present for the SF4 5’-3 DNA helicase, EcDnaB [115], however, the effect on DNA unwinding may be opposite to that of SsoMCM. Mutations of exterior positively charged residues on EcDnaB resulted in increases in ssDNA translocation and dsDNA unwinding efficiencies. It is intriguing that the excluded strand contact may provide a
‘molecular brake’ to control the unwinding rate. It is also possible that a greater force is applied to the encircled strand by the EcDnaB motor domain to destabilize the duplex and that the excluded strand modulates that force. A greater force may be necessary thermodynamically at mesophilic temperatures for EcDnaB compared to thermophilic temperatures for SsoMCM where higher temperature can aid in destabilization of the duplex. For the exterior mutants of EcDnaB that show faster and more efficient unwinding, loss of contacts with the excluded strand releases the imposed molecular restrictions rendering the unwinding helicase ungoverned (Figure 1.7B).

In addition to EcDnaB, other SF4 helicases, T7 gp4 and T4 gp41, have also been proposed to have an excluded-strand interaction based on unwinding rates that varied with the length of excluded-strand [120, 121]. The T7 gp4-excluded-strand interaction was disrupted using hybrid DNA:morpholino substrates. Morpholinos carry no charge on their backbone, and would neutralize any electrostatic interaction a protein has with the phosphate sugar backbone. Unwinding studies demonstrated that the T7 gp4 unwinding rate and amplitude were greatly enhanced when the morpholino was on the lagging strand [122]. This result is consistent with what has been observed for DnaB specifically that disrupting the electrostatic-based excluded-strand interaction greatly stimulates unwinding activity. It seems that for this SF4 helicase, the excluded-strand interaction acts to regulate the DNA unwinding rate, opposite to that for SsoMCM, despite the similarities of the interactions.

1.4.2.3 Nonhexameric Helicases that Engage Both Strands

Even nonhexameric helicases that exist primarily as monomers or dimers have been shown to make contact with both DNA strands. SF1 helicases are comprised of at least two RecA-like domains, and are characterized by a conserved arrangement of several core motifs that are
involved in ATP hydrolysis, DNA-binding, and ssDNA translocation. SF1A helicases translocate from 3’ – 5’, and SF1B helicases translocate from 5’ – 3’ [42, 123]. The best known example of SF1 helicases engaging both DNA strands is the bacterial recombination dependent helicase/nuclease RecBCD [124-126]. The RecB subunit is a SF1A helicase, and the RecD subunit is a SF1B helicase allowing unidirectional translocation along dsDNA with opposing polarities acting on each strand. The combination of domain motors allows engagement of both DNA strands to efficiently unwind the duplex in search of a Chi recognition site for recombination.

Other model SF1A helicases include Rep, PcrA, and UvrD and are involved in rolling circle replication, replication fork progression and repair of various DNA damage events [127-129]. These homologous helicases unwind DNA through an open to closed transition gripping the DNA in an inchworm stepping mechanism [130-135]. The ‘closed’ form of the helicase corresponds to robust helicase activity, while the ‘open’ conformation is associated with a strand-switching event and translocation on the opposite strand effectively reannealing the duplex. This observed strand switching suggests that interactions between the helicase and DNA go beyond solely interacting with the translocating strand. The helicase must also interact with the duplex region or the nontranslocating strand in order to anchor the helicase through the flexible 2B domain to switch and prevent dissociation.

The eukaryotic SF1B Pif1 helicase has been implicated in a number of activities including Okazaki fragment processing [136, 137], repair of double-stranded DNA breaks [138-140] and telomere maintenance [141-143]. Pif1 has a preference for unwinding DNA:RNA over DNA:DNA duplexes due to enhanced processivity [144]. Additionally, measured dissociation rates for single-stranded tailed duplex substrates were ~2-4 fold greater than those for measured
for forked duplexes suggesting that an interaction with the nontranslocating strand was responsible. The increased processivity on forked DNA:RNA substrates could have implications for Pif1’s role inhibiting telomerase by removing the telomerase complex. Reduced processivity on dsDNA tailed substrates may be important for Pif1 binding to double-strand breaks and subsequently recruiting other repair proteins. Similarly, the monomeric SF1B T4 bacteriophage Dda helicase involved in DNA replication initiation during T4 phage infection [123, 145, 146] has been shown to interact with both the translocating and nontranslocating strands to enhance unwinding of DNA fork substrates [147, 148].

Like SF1 helicases, SF2 helicases are characterized by a distinctive organization of ATPase motifs, however, SF2 helicases generally interact with their nucleic acid substrates mostly through the phosphodiester backbone, while SF1 helicases mainly utilize base stacking to mediate interactions [46, 49, 123]. The BLM helicase belongs to the RecQ group of SF2 helicases and mutations in this protein result in genomic instability, sunlight sensitivity and the early onset of cancers. BLM is thought to function in homologous recombination, and its loss increases sister chromatid exchange [49]. Interestingly, human BLM also exhibits strand-switching. Single-molecule magnetic tweezing experiments show that BLM unwinding is interrupted by short rezipping events [149]. BLM unwinds 3’ – 5’, but intermittently the two motor domains pause and loosen their interaction with the translocating strand. An accessory RQC domain maintains contact with the nontranslocating strand while BLM ‘slips’ backwards as the DNA fork rezips. A study of RecQ2 and RecQ3 from Arabidopsis thaliana also showed that these RecQ helicases also exhibit rewinding events by interacting with both the translocating and nontranslocating strands simultaneously [150].
The hepatitis virus C (HCV) NS3 (SF2) helicase processes the viral ssRNA genome, and it has also been proposed to interact with the host’s DNA genome [151, 152]. NS3 directionally unwinds substrates with a 3’-overhang in the absence of ATP, but not 5’-tailed or blunt substrates [153]. It is proposed that NS3 diffuses on the 3’-tail until it encounters and engages the duplex junction where local basepair melting is trapped through an interaction with the translocating strand as well as interactions with the newly displaced 5’-nontranslocating strand. These additional interactions would provide specificity for forward unwinding and allow for additional NS3 molecules to load behind the leading and engaged helicase for more efficient strand separation. The dual interactions with both DNA strands measured for various SF1 and SF2 helicases may be required for the strand switching and unwinding activities of these helicases similar to hexameric helicases.

1.4.3 Sensing DNA Damage

In addition to the apparent effects that interactions with the displaced strand have on functions such as strand switching, unwinding, and rewinding, interactions with the nontranslocating strand have been shown to detect DNA damage directly. Helicases perform important functions in a variety of DNA repair processes including nucleotide excision repair (NER), base excision repair (BER), homologous recombination (HR), and mismatch repair (MMR). Deficiencies or mutations in these helicases are clinically associated with genome instabilities in the forms of heightened predisposition to cancers and age-related genetic conditions such as Fanconi anemia, Werner syndrome, Bloom’s syndrome, and Xeroderma pigmentosum among others [36-38, 154]. Recently, there have been multiple reports identifying an interaction between a DNA repair
helicase and the nontranslocating strand which may have strong implications in the sensing of DNA damage (Figure 1.7C).

FANCJ is one of the 15 or so genes associated directly with Fanconi anemia (FA) that is primarily involved in recognizing and repairing interstrand DNA crosslinks that impede DNA replication processes. FANCJ is also known to directly interact with BRCA1 to resolve G-quadruplex structures [37, 155-158]. A recent study has shown that the unwinding activity of FANCJ is completely disrupted when a single polyglycol linker is included in the sugar phosphate backbone of either the translocating or nontranslocating strand [159], which suggests that FANCJ can sense the DNA backbone chemistry of both strands, which is crucial for unwinding. However, when three adjacent abasic sites were included in either strand, unwinding was only partially inhibited when the abasic sites were in the translocating strand. This result is consistent with a later study where FANCJ unwinding was only sensitive to a cyclopurine lesion in the translocating strand, and not in the nontranslocating strand [160]. Interestingly, FANCJ unwinding is also hindered by a thymine glycol site on either strand, but it is not sensitive to 8-oxoguanine on either strand [161]. Collectively, these data suggest a lesion-specific and strand-specific damage recognition mechanism (including the nontranslocating strand) by FANCJ that could direct various post-recognition processing pathways (Figure 1.7C).

Similar specific damage and strand recognition properties also exist for the XPD (SF2) helicase involved in eukaryotic NER [154]. The structure and unwinding kinetics of the archaeal homologue, XPD from *Thermoplasma acidophilum* (TaXPD) have been widely studied [162-166]. TaXPD unwinds DNA substrates containing either a fluorescein-modified thymine or a cyclobutane pyrimidine dimer (CPD). Although XPD stalls when the fluorescein adduct is located on the translocating strand, it recognizes and stalls at CPD sites only when they are
located on the nontranslocating strand [167]. This finding suggests that the helicase must interact with both strands and monitors ssDNA for specific DNA damage on each strand. These distinct modes of DNA damage searching and recognition may play a role in mediating distinct responses from downstream repair machinery.

The ability of SF4 hexameric replication helicases to unwind over damaged DNA has also been tested. EcDnaB showed no decrease in unwinding when a cyclopurine or thymine glycol was located on either the translocating or nontranslocating strand [160, 161]. However, the related SF4 mitochondrial Twinkle helicase showed a decrease in unwinding when acting upon DNA substrates with a thymine glycol site in the translocating strand. This effect was not seen when the thymine glycol was present only on the nontranslocating strand. DNA substrates with a cyclopurine either in the translocating strand or nontranslocating strand were also tested. In these cases, Twinkle was able to unwind the substrate with cyclopurine in the translocating strand, but not when the damage was on the displaced strand [168]. This result is another example where different mechanisms of damage recognition may lead to distinct downstream processing (Figure 1.7C).

1.4.4 Role of the Excluded Strand in the Unwindosome

Interactions with the nontranslocating strand have been shown for replicative helicases from all three domains of life, SsoMCM, EcDnaB and human Twinkle. The replication of genetic material is a complex and dynamic process, involving many protein components and interactions with various roles. Until recently, the displaced strand had only been considered a passive component at the replication fork. However, it seems that the excluded-strand serves a variety of functions that differ depending on the helicase being studied. At the fork, immediate strand
separation will be maintained though interactions both within the central channel and with the exterior surface. However, the extent of the SEW interaction of the excluded strand with the exterior of the helicase may be modulated by other interacting proteins within the replisome.

In eukaryotes, the MCM2-7 helicase is in contact with additional subunits, Cdc45 and GINS, to form the CMG complex or unwindosome [10, 57, 169-171]. Although direct interactions of the excluded strand with the exterior surface of MCM2-7 have not been tested, an interaction of the excluded strand with Cdc45 has been visualized using EM [172, 173]. Interestingly, CMG unwinding was stalled when molecular roadblocks were placed on the excluded lagging strand [92]. The authors note that the presence of a bulky lesion on the excluded strand could cause stalling if the outer surface of CMG interacts closely with the lagging strand. A recently reported secondary channel or groove has been characterized when the MCM2-MCM5 ‘gate’ is closed that could potentially accommodate the excluded strand [173]. When closed, this creates a second channel bound by the outer surface of MCM2-7 and the inner surfaces of GINS and Cdc45. It is possible that during active unwinding by CMG, the leading strand is restricted to the interior channel of MCM2-7, while the lagging strand is separated and fed into this secondary exterior channel. This scenario would provide for strand sequestration and prevent reannealing, and it may also provide a mechanism to hand-off the lagging strand to the DNA primase for priming. In archaea and SV40, similar to the bacterial systems, the DNA primase has been shown to directly interact with the helicase to mediate hand-off of the displaced lagging strand for efficient priming [174-176]. Interestingly, an EM structure of the minimal eukaryotic replisome, places Pol ε in front of the CMG complex contacting the leading strand, while the lagging strand traverses the length of the MCM2-7 complex to contact the Pol-α-primase complex afterwards for priming [177]. The architecture of the eukaryotic replisome is
still being revealed, and the preliminary studies are pointing towards replication fork organization and dynamics that are more complex than bacterial and viral model systems [24, 177-179]. However, based on the various roles and interactions between the nontranslocating strand and the CMG helicase, it would not be surprising to find that such an interaction is important in coordinating function at the replication fork. Clearly, higher resolution data is required to determine the precise paths of both strands through the MCM2-7 helicase and within the eukaryotic replisome for better understanding.

It has been reported that dsDNA unwinding by T7 gp4 helicase is greatly stimulated by the presence of the DNA polymerase placed very close to the duplex junction so that little to no ssDNA is exposed [180, 181]. However, in the absence of the polymerase or when the DNA polymerase stalls at a lesion, T7 gp4 can continue unwinding at a slower rate possibly engaging the excluded strand as a molecular brake [182]. During rapid synthesis, SEW interactions with the external surface are prevented by the polymerase, thus stimulating coupled unwinding and replication in T7. A similar model has been proposed for the coupling of T4 gp41 helicase and gp43 polymerase and E. coli DnaB and Pol III at the replication fork [183, 184]. For these 5’-3’ lagging strand translocating helicases, the excluded leading strand would be disengaged from the helicase in favor of promoting synthesis within the leading strand DNA polymerase. The binding and unbinding of the excluded strand to the exterior of the helicase may regulate the speed of unwinding and replication by coupling helicase activity to the polymerase (Figure 1.7B).

As an example, incorporation of abasic DNA lesions on either the leading or lagging strand causes the uncoupling of leading and lagging strand synthesis in E. coli. Leading strand synthesis is halted, while lagging strand synthesis continues in ~2/3 of cases for up to 1kb past the point of uncoupling at a slower rate [185]. Studies of the E.coli replisome have also shown
that the speed of replication fork migration and helicase unwinding is regulated by the other components of the replisome such as primases and polymerases. For example, inclusion of the DnaG primase reduces the processivity of leading strand synthesis upon interaction with DnaB, preventing uncoupling of leading and lagging strand synthesis when priming is rate limiting [186]. Furthermore, inclusion of translesion polymerases, Pol II and Pol IV, slows the DnaB helicase compared to solely the leading strand polymerase, Pol III [187]. It may be that more contacts with the excluded strand on the exterior of DnaB slows the replisome to allow time for DNA polymerase switching and translesion synthesis. It is possible that the excluded strand interaction takes place after the leading strand polymerase becomes uncoupled, and acts to slow the rest of the replisome to a stop by acting as an electrostatic brake to unwinding (Figure 1.7B). This model would reduce the amount of vulnerable ssDNA produced by helicase unwinding, stimulating proper repair before continuing with normal replication and releasing the brake.

The excluded strand interaction may also provide a platform for the accessory helicase, Rep, to bind and interact to restart stalled replication forks. Rep unwinds 3’ – 5’, opposite to the polarity of DnaB helicase and translocates on the leading strand instead. Rep and DnaB physically interact and unwind forks cooperatively [188]. An important finding was that there must be ssDNA available on the leading strand for Rep to interact and subsequently unwind cooperatively with DnaB. This ssDNA may be interacting with DnaB’s exterior surface until assembly of Rep on the leading strand stimulates fork restart. These studies suggest that the unwinding activity of DnaB is a highly regulated component of the replisome and that the excluded strand interaction can control the speed of the replisome to direct dynamic processes at the fork.
At the heart of the replisome is the hexameric DNA helicase. Assembly of the hexameric helicase at origins controls the initiation of replication, and the rate of unwinding controls elongation, both proposed to require interactions with the excluded strand. Independent of unwinding polarity, dynamic interactions with the excluded nontranslocating strand have now been shown to control the DNA unwinding rate, sensing of DNA damage and modulating interactions with other proteins within the replisome (Figure 1.7). Higher resolution studies are required to continue mapping ssDNA interactions on both the interior and exterior of DNA helicases. As the excluded strand has now been revealed to have an almost equal or greater importance than the translocating strand, future experiments will need to test contributions of both strands within the replisome to determine and differentiate interactions directing DNA unwinding and sensing.
2.0 EXCLUDED STRAND WRAPPING REGULATES DNAB HELICASE ACTIVITY

2.1 SUMMARY

The steric exclusion model for hexameric DNA helicase unwinding has focused primarily on specific contacts with the encircled strand of DNA within the central channel. However, interactions with the excluded DNA strand on the exterior surface of hexameric helicases have been shown to be important for DNA unwinding. Using single molecule FRET experiments, we show that replicative hexameric helicases, EcDnaB and SsoMCM, have almost identical interactions and dynamics with the excluded strand even though they translocate with opposite unwinding polarities. Exterior mutations on EcDnaB alter the DNA binding path and dynamics as shown using novel explicit probability and rate transition (ExPRT) plots. Furthermore, these exterior mutations generally increase the DNA unwinding rate for EcDnaB by either altering the conformation of hexamer and/or releasing the restriction imposed by excluded strand binding. These results expand on the existing model for hexameric helicase unwinding to include contributions of the excluded strand to modulate the rate of DNA unwinding.
2.2 INTRODUCTION

Hexameric helicases are structurally conserved toroidal enzyme complexes capable of translocating and separating double-stranded DNA (dsDNA) into two single-strands (ssDNA) providing templates for DNA replication. They utilize the inherent energy from ATP hydrolysis to translocate along an encircled strand physically displacing the excluded strand. The translocation polarity and corresponding unwinding polarity for hexameric helicase differs depending on the helicase families and organization of the conserved folds [68]. SF4 helicases from bacteria and associated phages (T4 and T7) include RecA-like folds and have 5’-3’ unwinding polarity, translocating on the lagging strand. SF6 helicases from archaea and eukaryotes have AAA+ folds and 3’-5’ unwinding polarity, translocating on the leading strand [14]. Although these two well-studied helicase families have globally conserved structural features, their amino acid sequences, structural folds, and unwinding polarities are not.

Previously, we identified an interaction on the external surface of the archaeal (AAA+) MCM helicase with the excluded strand that both protects and stabilizes the complex in a forward unwinding mode [96, 189]. This interaction expanded the well-accepted steric exclusion (SE) model of unwinding to include contributions of the excluded strand in the mechanism. This new unwinding model was termed steric exclusion and wrapping (SEW) [96]. Recently, interactions with the excluded strand have been uncovered from a variety of hexameric helicase complexes in addition to archaeal MCM including: E1 [190], SV40 LargeT [191], T7gp4 [192], EcDnaB [193], [194], and the eukaryotic Cdc45/MCM2-7/GINS (CMG) complex [173, 195]. It is hypothesized that external interactions with the excluded strand will not only protect ssDNA, but also stabilize the helicase/DNA complex, and modulate the unwinding rate.
Currently, various X-ray and EM structures of hexameric DnaB (with and without DNA or accessory proteins) show the hexamer as either a closed ring [77, 196, 197] or a split lock washer [70, 198]. Loading of the DnaB helicase hexamer is accomplished through conformational changes induced by the helicase loader, DnaC, which traps a ring opened spiral state competent for loading. Closing of the DnaB ring around the lagging strand stimulates DnaC dissociation concomitant with primase (DnaG) binding. Although discrete conformational steps in the DnaB loading process are being resolved at the structural level, the mechanism and precise contacts for maintaining an open DNA bubble as well as the contacts with DNA during unwinding are not known.

In this study, we examined whether there was a strict requirement for an excluded strand interaction with the external surface for *E. coli* DnaB (*EcDnaB*) hexameric helicase unwinding. smFRET experiments were employed to measure external interactions of *EcDnaB* with the 3’-strand. The absolute FRET states, transition probabilities, and dwell times were strikingly similar between *EcDnaB* and *SsoMCM* even though they reside in different superfamilies and have opposite unwinding polarities. Novel explicit probability and rate transition (ExPRT) plots were developed to easily compare differences in FRET dynamics between helicases, mutants, and DNA substrates. Notably, conserved exterior electrostatic SEW mutations on *EcDnaB* disrupt and alter the excluded strand interaction paths. SEW mutants of *EcDnaB* either increase the loading onto DNA in the absence of *EcDnaC* or increase the dsDNA unwinding rate compared to wild-type. In all, interactions on the exterior surface of bacterial *EcDnaB* with the excluded strand exist as proposed in the SEW model and act to regulate the rate of DNA unwinding.
2.3 MATERIALS AND METHODS

2.3.1 Materials

Oligonucleotides used (Table 2.1) were purchased from IDT (Coralville, IA). Fluorescently labeled DNA was HPLC purified (IDT), and non-labeled oligos were gel purified [199]. *Sso*MCM was purified as previously described and reported as hexamer concentrations [96]. TEV was purified as described [200]. AMPPNP was from Sigma-Aldrich (St. Louis, MO). All other chemicals were analytical grade or better.

Table 2-1: DNA Sequences

<table>
<thead>
<tr>
<th>DNA</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA14F</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA14</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA15</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA43</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA44</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA111</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA116</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA30F</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA40F</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA50F</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA75F</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA39F</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>DNA39D</td>
<td>5’-TCGATGAGACAGCCTGATGTAATCCGCATACGGGTTGACGTAGGAGGTGTGGAAGGAGAGGTTG</td>
</tr>
<tr>
<td>K180A For</td>
<td>5’-CAAGAGGACGCGGCCGCAAGGATGGTGG</td>
</tr>
<tr>
<td>K180A Rev</td>
<td>5’-CAAGAGGACGCGGCCGCAAGGATGGTGG</td>
</tr>
</tbody>
</table>
| R74A For | 5’-CGACACCTGACAGACCCATCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA

5’ – Cy5, 3’ – Cy3, F – fluorescein, B – Biotin, D – Dabcyl
2.3.2 Cloning and Purification of EcDnaB

The R74A, R164A, K180A, and R328A/R329A mutations of EcDnaB were cloned by overlap extension from pET11b-EcDnaB. The DNA primers are listed in Table 2.1. Mutations were confirmed by the DNA sequencing facility at the University of Pittsburgh. Mutants and wild-type EcDnaB were expressed in Rosetta 2 cells (EMD Millipore, Billerica, MA) using autoinduction [201] or by induction with 0.1 mM IPTG. Cells were pelleted, resuspended in DnaB lysis buffer [10 % sucrose, 50 mM Tris-HCL (pH 7.5), 50 mM NaCl, 5 mM dithiothreitol (DTT)], and lysed using lysozyme and sonication. Ammonium sulfate was added to the resulting supernatant at 0.2 g/mL, pelleted, and then resuspended in DnaB buffer A [10 % glycerol, 0.1 mM EDTA, 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM DTT]. The supernatant was purified using an AKTA Prime FPLC equipped with a HiTrap MonoQ column (GE Healthsciences, Sunnyvale, CA) and eluted with a stepwise gradient of DnaB buffer A with 500 mM NaCl followed by a similar procedure using a HiTrap Heparin column (GE Healthsciences, Sunnyvale, CA). The purified fractions were combined and applied to a Superdex S-200 26/60 gel filtration column (GE Healthsciences, Sunnyvale, CA) with Buffer C [50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM DTT] to isolate the hexamer. An extinction coefficient (185,000 cm⁻¹ M⁻¹) was used to quantify the fractions containing purified hexameric DnaB [74]. All concentrations for DnaB are indicated as hexamer throughout.
Figure 2-1: SDS-PAGE Analysis of EcDnaB Purifications.
The migration of equimolar amounts of the wild-type and mutant EcDnaBs are shown on a SDS-PAGE gel. K180A I and K180A II correspond to two separate purifications of that mutant. Protein ruler plus marker was also run on the gel, and the corresponding molecular weights of each band within the marker are shown for reference.

2.3.3 Purification of EcDnaC

pET28b-DnaC was transformed into C43(DE3) cells (Lucigen, Middleton, WI) and overexpressed using 0.5 mM IPTG at 37 °C for 4 hours. Cells were spun down, resuspended in DnaC lysis buffer [50 mM HEPES (pH 7.5), 500 mM KCl, 10% glycerol, 30 mM imidazole, 10 mM MgCl₂, 0.01 mM ATP, 1 mM BME], and lysed using a french press (Thermo Fisher, Waltham, MA). The clarified supernatant was loaded onto a HisTrap HP column and eluted with a linear gradient to 300 mM imidazole using an AKTA Pure FPLC (GE Healthsciences, Sunnyvale, CA). Fractions were combined and loaded onto an amylose resin column (NEB, Ipswich, MA), washed, and eluted with 10 mM maltose. Fractions were concentrated using Amicon Ultra-30 spin column (EMD Millipore, Billerica, MA), and ATP concentration was
increased to 0.1 mM, before adding 2-3 mg TEV for overnight digestion at 4 °C. Overnight digestion was reapplied to the HisTrap HP column, and the flow-through was collected. Cleaved EcDnaC was then applied to a HiLoad S-200 26/600 gel filtration column (GE Healthsciences, Sunnyvale, CA) in S200 Buffer [50 mM Hepes (pH 7.5), 500 mM KCl, 10% glycerol, 10 mM MgCl₂, 0.1 mM ATP, 1 mM BME]. Concentration was determined by a standard curve using the Bradford reagent (Sigma-Aldrich, St. Louis, MO).

2.3.4 Single-Molecule Fluorescence Resonance Energy Transfer

DNA substrates (Table 2.1) labeled with Cy3 and Cy5 fluorophores were immobilized on a pegylated quartz slide utilizing biotin-streptavidin interactions [202]. A prism-based total internal reflection microscope was used to collect all smFRET data [203, 204]. A 532 nm diode laser was used to excite Cy3 fluorophores, and subsequent Cy3 and Cy5 emission signals were separated by a 610 nm dichroic longpass mirror, a 580/40 nm bandpass filter, and a 660 nm longpass filter. An EM-CCD iXon camera (Andor, Belfast, UK) was used to image the signals. Data was acquired at 10 fps for ten or more regions with each region containing 50 – 250 molecules in the presence of an oxygen scavenging solution [1 mg/mL glucose oxidase, 0.4 % (w/v) D-glucose, 0.04 mg/mL catalase, and 2 mM trolox]. DnaB (250 nM) was added and given a five-minute equilibration period. All single-molecule experiments were performed in 25 mM Tris Acetate (pH 7.5), 125 mM potassium acetate, and 10 mM magnesium acetate as described previously [96].

Single-pair FRET signals were identified by fitting individual regions of signal intensity to a 2D Gaussian and measuring the goodness of fit. These peaks were corrected for thermal drift.
and local background intensity [205, 206]. The resulting signal was used to calculate the apparent FRET efficiency, $E_{app}$, according to

$$E_{app} = \frac{I_A}{I_A + I_D}$$

in which $I_A$ and $I_D$ are the intensity of the acceptor and donor signals respectively. The relation between FRET efficiency and the distance between the dyes can be described by:

$$E = \frac{1}{1 + (\frac{r}{R_0})^6}$$

where $r$ is the distance between the dyes and $R_0$ is the Förster distance for the FRET pair, or distance between the dyes where half of the energy is transferred.

### 2.3.5 Single-Molecule FRET Data Analysis and ExPRT Plots

Data analysis and visualization were performed using manually selected single-molecule traces that displayed anti-correlation between the donor and acceptor fluorophores and single-step fluorophore photobleaching. Traces collected under identical experimental conditions were stitched together and fit to ideal states via Hidden Markov Modeling using the vbFRET software package [207]. Stitched traces were fit to a given number of states based on those states being more than $E_{app} = 0.1$ apart from one another, and the variation of one state not overlapping with another. Traces were then unstitched and fed into the ExPRT analysis program. This MATLAB executable program produces transition plots where the markers are sized based on the probability of transition occurring within an observed single-molecule trace and colored based on the dwell time(s) of the state preceding the transition. Each transition is analyzed by fitting a survival curve of dwell times to single and double exponential decays. If the $R^2$ value of the double exponential fit was greater than 0.970 and increased from the $R^2$ value of the single
exponential fit by more than 0.015, then the rates produced by the double exponential fitting were used to create a concentric marker, with two colors each representing a dwell time based on the rates. Otherwise, the average of all dwell times for a given transition was used to determine the color of the marker. Dwell times that had values of more than one log greater than the scale are indicated by the color black. Only dwell times that were both preceded and followed by transitions were taken into account. Stitched data as fit by vbFRET were also analyzed by the POKIT analysis program, and the resulting plot contains a legend for the ranges of rates and probabilities [208]. The data were also fit using the HaMMY program, allowing the program to fit the data to up to five states. The output of the HaMMY program was subsequently analyzed by the Transition Density Plot program [209].

2.3.6 EcDnaB Structural Homology Model

Global sequence alignments were performed using ClustalW2 analysis (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). The homology model of EcDnaB was created by threading the alignment on to the structure of Geobacillus stearothermophilus DnaB (PDBID: 4ESV) [70] using Swiss-MODEL [210].

2.3.7 Gel Based Helicase Unwinding Assays

Helicase assays were assembled in 50 mM HEPES (pH 8.0), 10 mM Mg(OAc)₂, 5 mM DTT, 0.2 mg/mL BSA at 37 °C. Final reaction concentrations of EcDnaB and forked DNA (DNA14F/DNA15) were 3 μM and 15 nM respectively. Reactions were initiated by adding 5 mM ATP and then quenched using an equal volume of quench solution (50% glycerol, 1% SDS,
100 mM EDTA pH 8.0, 300 nM DNA14). Reactions were kept on ice until loading and resolved on 14-20% (29:1 acryl:bisacryl), 8 M urea, and 1x TBE gels. The gels were imaged on a Typhoon 9400 (GE Healthsciences, Sunnyvale, CA), and the fraction unwound was calculated using ImageQuant software.

### 2.3.8 ATPase Assay

DnaB variants (350 nM) were incubated in the absence and presence of 4 μM forked DNA (DNA14/DNA15) as described previously [119]. Briefly, 25 μL reactions were incubated at 37 °C for 5 min in unwinding buffer [50 mM HEPES (pH 8.0), 10 mM Mg(OAc)₂, 5 mM DTT, 0.2 mg/mL BSA], and 1 mM ATP with trace amounts of ³²P-ATP was added to initiate the reaction. Samples were quenched at 2, 5, 10, and 15 min after initiation in equal volumes of 0.7 M formic acid. A total of 1 μL of quenched reaction was spotted on Millipore TLC PEI Cellulose F, allowed to dry, resolved in 0.6 M potassium phosphate (pH 3.5), phosphorimaged, and quantified for the linear ATPase rate (pmol/min).

### 2.3.9 Fluorescence Anisotropy

The EcDnaB binding assays were performed in 100 mM NaCl, 5 mM MgCl₂, 50 mM Tris-HCl (pH 8.1), 1 mM AMP-PNP, and 10% glycerol at room temperature. The DNA14F/DNA15 fork was used at a final concentration of 10 nM. Fluorescence anisotropy measurements were made using a Cary Eclipse Spectrophotometer (Agilent, Santa Clara, CA) at each concentration after a five-minute incubation after protein was added. Excitation was 485 nm and anisotropy was monitored at 520 nm with a two second integration time. The reported dissociation constants...


\((K_d)\) were calculated by plotting the average of three or more data sets and fitting to the following binding cooperativity equation:

\[
Y = \frac{A_{\text{max}} \times [\text{DnaB}]^n}{(K_d)^n + [\text{DnaB}]^n}
\] (3)

in which \(Y\) is the measured anisotropy, \(A_{\text{max}}\) is the maximal anisotropy and \(n\) is the Hill coefficient using the Kaleidagraph (Synergy Software, v 3.5).

2.3.10 ssDNA Translocation Assays

Stopped-flow fluorescence experiments were performed on an Applied Photophysics SX.20MV (Leatherhead, UK) in fluorescence mode equipped with a temperature controlled water bath chamber at 22 °C. 400 nM 3'-fluorescein labeled DNA (DNA30, DNA40, DNA50 or DNA75) was mixed with 175 nM EcDnaB hexamer as described in the Figure legends. Reactions were performed in a reaction buffer (20 mM Tris-HCl pH 7.5, 100 mM potassium acetate, 50 mM magnesium acetate) containing no nucleotide, 5 mM ATP, or 5 mM AMPPNP as indicated. The samples were excited at 485 nm, and a 515 nm cutoff filter was used to collect 4,000 oversampled data points. The excitation path length was 2 mm. At least eight traces were averaged for each individual experiment and multiple experiments were average to provide the rates and amplitudes. The resulting averaged traces were fit to exponential equations using the supplied software according to the following equation:

\[
F = A_1 \times e^{-k_1 t} + A_2 \times e^{-k_2 t}
\] (4)

in which \(F\) is the fluorescence intensity, \(A\) is the maximal amplitude, \(k\) are the observed rate constants with subscripts (1 & 2) designating fits to single or double exponentials, \(t\) is time, and \(C\) is an offset constant.
2.3.11 Fluorescence DNA Unwinding Assays

The EcDnaB/EcDnaC unwinding assays were performed in 50 mM HEPES (pH 8.0), 10 mM Mg(OAc)$_2$, 5 mM DTT, 0.2 mg/mL BSA at 37 °C and initiated with ATP. Final concentrations of EcDnaB, EcDnaC, and the DNA fork (DNA39F/DNA39D) were 250 nM, 1.125 μM, and 30 nM, respectively. Data was collected on a Cary Eclipse Spectrophotometer with a 0.1 second integration time. Excitation was at 485 nm and fluorescence data was collected at 520 nm. Time dependent changes from an EcDnaC only control was subtracted from all experiments that included EcDnaC. Data was averaged from three independent experiments and fit to the following equation:

$$F = A - \frac{A - A_0}{1 + (k_1 \times t)^n}$$

in which $F$ is the fluorescence intensity, $A$ is the maximal amplitude, $A_0$ is the minimal amplitude, $k_1$ is the observed rate constant, $t$ is time, and $n$ is a term to control the point of inflection, to calculate the rates. Other equations including multiple exponential and other sigmoidal growth equations with included lag phase parameters were also fitted, but Equation 4 provided the best fit of the data and allowed a direct comparison of the individual rates ($k_i$).
2.4 RESULTS

2.4.1 Developing Methods to Characterize Excluded Strand Interactions

To determine whether exterior interactions of helicases with the nontranslocating DNA strand exist, a variety of qualitative and quantitative biochemical and biophysical experiments can be performed. Both stable and dynamic binding of the nontranslocating strand to the exterior surface may aid in DNA unwinding, and assays are needed to differentiate strands and quantitate specificities. Precise detection of nontranslocating strand binding coupled with mutagenesis can unequivocally determine whether the excluded strand plays any role in DNA unwinding and stabilization for DNA helicases. Table 2.2 summarizes several methods that could be used to identify and characterize such an interaction. However, we employ smFRET here due to the advantages single-molecule methods provide as discussed below.

<table>
<thead>
<tr>
<th>Method</th>
<th>Experimental Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA Footprinting</td>
<td>Identifying specific regions and lengths of each strand of DNA protected upon binding to the helicase.</td>
</tr>
<tr>
<td>DNA Crosslinking</td>
<td>Captures both transient and stable covalent protein-DNA complexes for analyses of strand specificities and amino acid identification.</td>
</tr>
<tr>
<td>HDX-MS</td>
<td>Global unbiased measurement of DNA binding to the helicase in solution without perturbations.</td>
</tr>
<tr>
<td>smFRET</td>
<td>Determines populations of distance-based DNA conformations and their changes upon wild-type or mutant helicase binding.</td>
</tr>
<tr>
<td>ExPRT Analyses</td>
<td>Allows for easy visualization of conformational transitions, binding dynamics, and transition rates of fluorescently labeled DNA strands between two or more experimental conditions.</td>
</tr>
</tbody>
</table>

2 The material from section 2.4.1 was reprinted/adapted from Methods, In press, S.M. Carney, M. A. Trakselis, The excluded DNA strand is SEW important for hexameric helicase unwinding. (2016), with permission from Elsevier. Reprinted from The Lancet, In press, S.M. Carney, M. A. Trakselis, The excluded DNA strand is SEW important for hexameric helicase unwinding. (2016), with permission from Elsevier.
### 2.4.1.1 Single-Molecule Fluorescence Energy Transfer

Even though techniques such as footprinting and crosslinking can be used to identify interactions between helicases and DNA, they do not provide information on the dynamics and heterogeneity of those interactions. Single-molecule FRET (smFRET) is a technique that can yield detailed information concerning the dynamics of an interaction, and it has been used in several studies to characterize the interactions between hexameric helicases and the excluded nontranslocating strand during unwinding [96, 115, 168, 189]. Generally, smFRET utilizes a flow cell where the inside surfaces have been pegylated to reduce non-specific binding to the surface. A small fraction of these PEG molecules are biotinylated so that biotin-streptavidin interactions can immobilize the protein, RNA, or DNA of interest Figure 2.2B. The flow cell is imaged using a total internal reflectance fluorescence (TIRF) microscope, and a diode laser is used to excite the donor dye Figure 2.2A. Subsequent donor and acceptor fluorescent signals are imaged with dual channel emission optics Figure 2.2C. This allows for the quantification of corresponding donor and acceptor dye signals from each single molecule imaged over time Figure 2.2D. Quantification of the donor and acceptor signals allow the FRET efficiency (E) to be calculated using Equation 1 in the materials and methods section. The relation between FRET efficiency and the distance between the dyes can be determined using Equation 2 in the materials and methods section. Details concerning specific protocols for performing smFRET experiments can be found in reviews dedicated to the practical and theoretical considerations of smFRET experiments [202, 203, 211].
Figure 2-2: smFRET Experimental Setup.
A) TIRF microscopy. A 532 nm diode laser is focused and directed through a prism setting on top of the flow cell. Emitted light is captured by the objective, and separated based on wavelength before entering the CCD camera. B) smFRET flow cell design. The flow cell is coated in polyethylene glycol (PEG) to prevent non-specific binding. A small fraction of these immobilized PEG molecules are biotinylated. Streptavidin is used to sandwich biotinylated DNA to the biotinylated PEG surface. The DNA substrates are model fork structures where a duplex region branches into two poly-deoxythymidylates. The 3’ and 5’ ends of the DNA fork are labeled with Cy3 and Cy5 respectively. C) CCD imaging. The separated donor and acceptor signals are detected by the CCD camera. Two resulting channels show superimposable images of the Cy3 signal in the donor channel and the Cy5 signal in the acceptor channel. D) Single-molecule time trace. Top Panel: Corresponding donor and acceptor intensity can be followed over time as shown by the outlined circles in C). Each single pair (e.g. arrows) can be plotted over time to produce a single-molecule fluorescence intensity time trace. Bottom Panel: The FRET efficiency is calculated and plotted over time (blue) and fit to ideal states (red) to aid in downstream analysis.
The DNA substrates used in the helicase-interaction assays are designed as model forks, where a duplex region is immobilized via biotin-streptavidin interactions to the pegylated flow cell surface and two single-strand arms of the fork project outwards away from the surface Figure 2.2B. To avoid spurious base pairing of the single-strand fork arms, the arms for the fork consist solely of oligo-deoxythymidylates. The termini of the fork arms are labeled with the Cy3-Cy5 FRET pair. The energy transferred from the Cy3 donor dye to the Cy5 acceptor dye is a function of the distance between the dyes as shown in Equation 2. In the absence of any protein, the FRET signal is low due to the ends of the fork arms not being within close proximity, as can be seen in the histogram corresponding to the DNA3050 fork substrate alone Figure 2.3A. When introducing a hexameric helicase that encircles one of the fork arms in its central channel as it loads onto the fork substrate, one might expect to see a further decrease in the FRET signal due to the hexamer acting as a wedge that would drive the two arms of the fork further apart as in the SE model (Figure 1.6). However, the first time this assay was implemented to study the interaction with the \textit{Sso}MCM helicase, an increase in FRET was observed upon helicase binding [189]. Performing a similar experiment, we see consistent results (Figure 2.3A), where the wild-type (WT) \textit{Sso}MCM helicase induces a high FRET state on the DNA3050 fork. An increase in FRET efficiency is consistent with the excluded strand physically interacting with the outer surface of the hexamer so that the Cy3 – Cy5 pair are brought within close proximity. Surface mutations along the proposed path of interaction with the excluded-strand were made (\textit{Sso}MCM K323D/R440D) and the resulting histogram profile shows a broad distribution of FRET signals, suggesting a disruption of the wrapping interaction seen for the wild type. These residues have been previously suggested to mediate the excluded strand wrapping interaction in the case of \textit{Sso}MCM [96] and point to an electrostatic-based wrapping interaction consistent with the SEW
model. Although a histogram of the FRET signal can show differences in the FRET state(s) being sampled, it does not characterize or quantify the dynamics of the interaction.

2.4.1.2 Explicit Probability and Rate Transition (ExPRT) Plots.

smFRET allows for the monitoring the distances between single Cy3 – Cy5 pairs on single molecules in real time. Therefore, single events such as binding interactions, conformational changes, and protein translocation can be observed explicitly. Analyses that quantify the probabilities and rates of these events can be used to provide a detailed characterization of the dynamics being observed. There are several programs available to perform analyses on smFRET data [208, 209, 212, 213]. These programs rely on using Hidden Markov modeling programs, such as HaMMY and vbFRET to fit data to ideal states [207, 209]. The analysis programs then use the ideal state fits to the raw data to generate transition plots, where the initial FRET value (the FRET state preceding the transition) is on the x-axis, and the final FRET value (the FRET state transitioned into) is on the y-axis. For example, the Transition Density Plot (TDP) program created by the Ha laboratory uses ideal state fits to each individual trace and visualizes all transitions from all traces as heat map [209, 212]. Heterogeneity within the transitions can be visualized and qualitative comparison of transition probability can be visualized, but features that provide explicit quantification of the data directly on the plot are lacking [209, 212]. Another analysis program is the Probability and Kinetically Indexed Transition (POKIT) Plot program developed by the Walter laboratory [208, 213]. This program relies on fitting all traces as one stitched trace, which allows for the analysis of the transition rate and probability between the number of ideal states fitted to the stitched trace. The POKIT program utilizes color and concentric markers to plot transitions into bins of probabilities and rates. Each single marker and its corresponding x,y coordinate together represent a unique transition. However, plotting the
data into bins diminishes the power of single-molecule methods to determine the explicit parameters such as probability and rates based on single-molecule events. There are limitations and shortcomings for all analysis programs. The currently available analyses either cannot visualize both probability and kinetic information on a single plot or rely on binning data into ranges of values in order to do so. We sought to overcome some of these limitations by creating our own smFRET analysis program termed Explicit Probability and Rate Transition (ExPRT) plots.
Figure 2-3: smFRET Analysis of WT and Mutant SsoSSB bound to DNA.
A) A histogram of all single-molecule FRET data collected for the DNA3050 substrate alone, with WT SsoMCM, and SsoMCM (K323D/R440D). B) and C) show the ExPRT plots for WT SsoMCM and SsoMCM (K323D/R440D), respectively. D) The legend for the ExPRT plots: Transition Probability and Dwell Times. smFRET time traces for E) WT and F) K323D/R440D showing Cy3 (green) and Cy5 (red) intensities and FRET efficiencies (blue). The red line is a fit to ideal FRET states.
The ExPRT analysis program is a Matlab executable program that produces ExPRT plots from raw smFRET values. The ExPRT program follows an initial data processing and analysis approach similar to the POKIT program, where the raw data that is collected must be stitched together and fit to ideal states using software such as vbFRET [207]. That ideal stitched trace is then unstitched to avoid analyzing ‘false transitions’ and to generate single-trace statistics. The unstitched traces can then be fed into the ExPRT program. Figure 2.3E-F show examples of unstitched traces for WT SsoMCM and SsoMCM (K323D/R440D) on the DNA3050 fork. The upper panel shows the raw Cy3 donor and Cy5 acceptor intensities, and the lower panel shows the FRET efficiency calculated from the raw data. The overlaid red line is the ideal state fit to calculated FRET values. The program goes through each trace and collects all transitions that occur in addition to the dwell time of each state between transitions. Once all traces have been analyzed, the ExPRT program generates the transition plots (Figure 2.3B-C). Each transition is plotted as a marker based on its initial FRET state on the x-axis and the final FRET state on the y-axis. The size of the circle corresponds to the probability of that transition occurring within a measured single trace. We define probability as the fraction of analyzed traces that exhibit the given transition at least once. The color of the filled circle corresponds to the time spent in the initial state before transitioning (Figure 2.3D). Although the ExPRT plots resemble POKIT plots in that they both use circular markers to denote transitions, ExPRT plots use and visualize the explicit probabilities and dwell time values rather than displaying the data in bins. In addition, the ExPRT program also plots the dwell times of each transition in the form of a survival curve, and fits the subsequent curve to single and double exponential decay functions. The user is then able to determine which fit is appropriate based on the differences in R² values between fits. If a double exponential decay is selected, then the resulting marker on the ExPRT plot will be
concentric, containing two colors where each color corresponds to the dwell time based on the rates given by the exponential fit. For $\text{SsoMCM (K323D/R440D)}$, there is an increase in the number of FRET states, an increase in the probability of a transition, and a decrease in the dwell times in each state compared with WT (Figure 2.3B-C). In all, this mutation is responsible for a large increase in excluded strand binding dynamics.

The ExPRT program illustrates all observed dynamics on a single plot, which allows the reader to more easily gain greater insight into the measured interactions. When altering the length of the excluded strand or when studying mutants designed to disrupt the interaction, the ExPRT plots capture changes in FRET states, transitions, and kinetics. ExPRT analysis is the only single-molecule FRET analysis program that produces transition plots that simultaneously display explicit transitions, probability of transitions, and dwell times on a single plot. Moreover, the ExPRT program tests for and can display multiple rates that govern a transition. Visualization of the data is a critical aspect of analysis, and interpretation and comparison between data sets necessitates concise representation of the data. We believe that the ExPRT plots provide the most detailed description of dynamics on a single plot, which is advantageous for comparison between experimental conditions and interpretation of results. For these reasons, we believe that the ExPRT program may be a preferable method of analyzing and displaying data compared to other smFRET analysis programs [208, 209, 212, 213].

2.4.2 $\text{EcDnaB Interacts with the Excluded Strand}$

In order to determine whether the SEW model for hexameric helicase unwinding [96] is also employed by opposite polarity 5’-3’ helicases, we examined the interaction of the excluded 3’-
strand with *Ec*DnaB using a smFRET DNA wrapping assay. We used three separate model fork substrates, 30/30 (DNA43/DNA44), 40/30 (DNA111/DNA44), and 50/30 (DNA116/DNA44), that are composed of an 18 bp duplex with 30 (dT) on the lagging strand and 30, 40, and 50 (dT) on the 3’-leading strand, respectively (Figure 2.4A). The smFRET traces for the DNA substrates were collected and histograms were generated (Figure 2.4B). DNA forks alone result in low FRET signals from Cy3 and Cy5 on the termini of the fork arms not being in close proximity. Additional nucleotides on the leading strand arm of the fork decrease the FRET signal even further as expected.

Figure 2-4: Single-molecule FRET monitoring of *Ec*DnaB binding to DNA fork substrates. Cartoon models of the DNA forks with a static 30 base encircled 5’ strand and variable excluded-strand 3’ arm lengths (30, 40, and 50 nt) shown in blue, green and red, respectively. (B) Histograms of the FRET signal from the DNA fork substrates alone, colored to match (A). (C) Histograms of the three DNA substrates in the presence of 250 nM wild-type *Ec*DnaB.
Addition of EcDnaB to each of these substrates shifts the populations to various high FRET states bringing the Cy3 dye in close proximity to Cy5 (Figure 2.4C). As previous work has shown that EcDnaB encircles the 5’-strand [112], the occurrence of a high FRET state is consistent with an interaction of the excluded Cy3 3’-strand on the external surface of EcDnaB as described in our SEW model. A titration of EcDnaB onto 30/30 showed little to no variation in the resulting histogram profiles, suggesting that only one hexamer can be accommodated by the fork substrate over a large concentration range (Figure 2.5). EcDnaB loaded onto the 40/30 substrate produced an almost exclusively high FRET state (>0.9); the 50/30 fork produced a bimodal distribution of high (>0.9) and medium (~0.5) FRET states; and the 30/30 fork yielded a bimodal distribution of two high FRET (0.8 and >0.9) states in the presence of EcDnaB. The interaction with the excluded strand can differ depending on its length or external binding paths, but in all cases, the induced high FRET states correspond to a tightly wrapped excluded strand that places the Cy3 dye in close proximity to the Cy5 dye on the encircled strand.

Figure 2-5: Titration of EcDnaB onto 30/30 Fork.
The histogram profiles from a titration of WT EcDnaB onto the 30/30 fork substrate from 50 nM to 8 μM (hexamer) is shown. 30/30 alone exhibits low FRET (shown in dark blue). Adding WT EcDnaB shifts the FRET signal to higher FRET values in all cases, while subsequently increasing the WT EcDnaB concentration further does not significantly alter the histogram profile.
2.4.3 EcDnaB and SsoMCM Wrap the Excluded Strand Similarly

The bimodal distribution observed for EcDnaB on 30/30 fork substrate was very similar to the distribution produced by the archaeal MCM helicase on the same substrate (Figure 2.6). The single-molecule traces for EcDnaB and SsoMCM exhibited similar dynamics between two high FRET states (~0.95 and ~0.8) (Figure 2.6B-C). In order to more easily probe and compare the dynamics of the smFRET traces, we have developed a novel analysis and method of visualization termed the Explicit Probability and Rate Transition (ExPRT) plots as discussed above. ExPRT plots simultaneously visualize FRET states, explicit transition probabilities, and dwell times on a single plot. Several other methods of analysis and visualization of smFRET dynamics have been utilized in the past [208, 209]; however, they generally lack the ability to visualize all three parameters simultaneously or rely on binning data into pre-defined ranges of parameters.

ExPRT analyses and plots of the smFRET data for EcDnaB and SsoMCM bound to the 30/30 DNA fork can more easily and globally compare separate experimental situations (Figure 2.7). The positions of the circular markers correspond to transitions between specific FRET states: initial FRET state on the x-axis and the final FRET state on the y-axis. The size and color of each marker correspond to the probability of that transition occurring within a measured trace, and the average dwell time of the state preceding the transition, respectively (Figure 2.7C).
Figure 2-6: Comparison of the similar excluded strand interactions of SsoMCM and EcDnaB.
A) Shows the overlaid histograms of both SsoMCM and EcDnaB on the 30/30 fork. Representative single-molecule traces for B) EcDnaB and C) SsoMCM on the 30/30 DNA template. The top panels shows the Cy3 (green) and Cy5 (red) signals. The bottom panels show the corresponding FRET signal (blue) with overlaid ideal states (red) for each trace as fit by vbFRET (see Materials and Methods).
Figure 2-7: ExPRT plots of EcDnaB and SsoMCM on DNA.
ExPRT plots for A) EcDnaB (250 nM) and B) SsoMCM (1.3 μM), respectively, on 30/30 fork DNA. The number of states and traces fit by the data is in the upper left hand corner for each plot. The size (probability) and color (dwell time) of the markers can be compared to the legend shown in C).
Figure 2-8: Comparison of smFRET analysis methods.
The smFRET data set from WT EcDnaB on 30/30 was analyzed using previously established analysis methods, and subsequently compared to the new ExPRT analysis. The data was fit using HaMMY, and subsequently analyzed and visualized by the Transition Density Plot program to produce the plot shown in A). The TDP program visualizes transition probability as a heat map, with brighter colors corresponding to greater probabilities. Separately, the data were stitched together and fit using vbFRET. The fit data was then analyzed and visualized using the POKIT analysis program. The resulting plot and corresponding legend are shown in B). The same data was then subjected to analysis by the ExPRT program, and the corresponding plot and legend are shown in C).
Figure 2-9: Dwell Time Survival Analysis of EcDnaB on 40/30.
Dwell times for the transition from E=0.97 to E=0.75 for WT EcDnaB on 40/30 plotted as a survival curve (blue bars) with the \textbf{A}) single or \textbf{B}) double exponential fit overlaid in red. Each plot also contains an inset of the ExPRT plot where the data was analyzed with the 0.97 to 0.75 transition as a single or double exponential fit.

A comparison of analyses between established programs, the HaMMY and TDP programs [209] as well as the POKIT program [208], and the ExPRT program is shown in Figure 2.8A-C. Each program analyzed and visualized data corresponding to EcDnaB bound to the 30/30 substrate. Each plot illustrates that the transitions between FRET states of ~0.8 and ~0.95 are the most frequent. The Transition Density Plot Program analysis works on a trace-by-trace basis, and is able to reveal heterogeneities in the transition data that can be missed by programs that work on stitched datasets such as the POKIT and ExPRT programs. However, despite the TDP program’s ability to gather probability and rate values, these values are not directly visualized by the resulting plot. The POKIT program bins the probabilities and rates of each transition into user-defined ranges and produces plots that allow for some level of quantitative comparison between experimental conditions. However, these plots fail to display explicit probability and rate values. Determining explicit transitions, probabilities, and rates is an advantage of single-molecule methods that allows for extensive insights into the dynamics and kinetics of molecular interactions and enzymatic activities. The ExPRT analysis program
determines explicit values and maintains the ability to visualize them directly. This program allows users and readers to easily make comparisons between datasets on the most detailed level. ExpRT plots can also determine if one or two rates govern a particular transition. An example transition comparison fit to either a single or double exponential dwell time survival curve shows that the double exponential is more appropriate with a larger $R^2$ value (Figure 2.9). Double exponential fits are represented by two concentric circles as the marker in the ExPRT plot, where the two colors represent the individual dwell times (Figure 2.9B). Therefore, the ExPRT plots provide a powerful new method for investigation and comparison of the probability and kinetics of smFRET dynamics.

Specifically for EcDnaB compared to SsoMCM on the 30/30 fork substrate, there are strikingly similar FRET states and transitions on the ExPRT plots (Figure 2.7). Both enzymes have a reversible transition of the excluded strand between two high FRET states (~0.8 and ~0.95) that is exhibited by ~70% of the molecules analyzed. For both EcDnaB and SsoMCM, we see a preference for the ~0.95 FRET state, indicated by the longer dwell times measured for that state (shades of orange vs. green/yellow). In addition, both data sets exhibit a reversible lower probability transition between each of high FRET states and a medium FRET state of ~0.55. Altogether, it is surprising that notable similarities in the FRET states, transitions, and dwell times exist for EcDnaB and SsoMCM on the 30/30 fork even though these two helicases belong to different superfamilies with low sequence homology, exist in different domains of life (bacteria vs. archaea), and have opposite unwinding polarities.
2.4.4 Conservered Surface Mutants of EcDnaB Alter Excluded Strand Wrapping

To determine whether the interaction of the excluded strand has effects on the unwinding ability of EcDnaB, several conserved surface-exposed mutations were created (Figure 2.10). Electrostatic interactions were found to mediate the wrapping interaction in the case of SsoMCM, and titrating increasing amounts of salt onto the 30/30 WT EcDnaB complex resulted in increased dynamics (shorter dwell times), suggesting that EcDnaB also utilizes electrostatic interactions to mediate wrapping (Figure 2.11). Based on a homology model for EcDnaB, four surface mutants (R74A, R164A, K180A, and R328A/R329A) that exist in positively charged electrostatic patches (Figure 2.12A-B) were cloned, expressed, and purified. All mutants were consistent with forming a hexamer as the major peak after gel filtration.

![Figure 2-10: Amino acid sequence alignments of DnaB helicases.](http://www.ebi.ac.uk/Tools/clustalw2)

A) Multiple amino acid alignment of DnaB helicases using CLUSTAL W2. Identical (*), similar (:), and somewhat similar (.) residues are indicated. ECOLI - Escherichia coli strain (K-12); GEOSE - Geobacillus stearothermophilus; ECOBD - Escherichia coli strain (BL21-DE3); BARGA - Bartonella grahamii (strain as4aup) SALTY - Salmonella typhimurium; HELPY - Helicobacter pylori strain (26695); ECO57 - Escherichia coli O157:H7; SHIFL - Shigella flexneri; TREPS - Treponema pallidum SS14.

B) The phylogenetic tree was created using phylot (http://phylot.biobyte.de).
Figure 2-11: ExPRT plots of salt titration onto the 30/30-EcDnaB complex.
A) ExPRT plot for WT EcDnaB on the 30/30 substrate. B-D) ExPRT plots for the same complex in the presence of increasing NaCl concentrations.
Figure 2-12: ExPRT plots of WT EcDnaB and selected mutants bound to different fork lengths. Position of the mutations mapped onto the homology model for EcDnaB A) colored and B) electrostatic surfaces. ExPRT plots comparing dwell times and probability of transition for C-E) WT and F) K180A, G) R74A, and H) R328A/R329A bound to 30/30, 40/30, and 50/30 substrates, respectively.

Figure 2-13: Example smFRET kinetic traces. Comparison of A-C) WT and D) K180A, E) R74A, and F) R328A/R329A EcDnaB FRET efficiencies as a function of time on the 30/30, 40/30 and 50/30 forks, respectively. The calculated FRET values (blue) are overlaid with the ideal state fits (red).
smFRET wrapping assays were performed for each mutant on each of the three fork templates and analyzed and compared using traditional histograms and newer ExPRT plots (Figure 2.14). A selection of ExPRT plots for various mutants are compared to WT EcDnaB on the same fork substrate (Figure 2.12). Several important differences between the hexamer-excluded strand interactions as well as the dynamics of those interactions are highlighted by the ExPRT plots. For example, EcDnaB (K180A) with the 30/30 fork produces similar FRET states and dynamics when compared to the wild-type EcDnaB; however, there are five states compared to three (Figure 2.12C & F). Example traces for individual molecules for wild-type compared to K180A are shown in Figure 2.13A & D. A greater number of FRET states and transitions are indicative of a less stable interaction between the exterior surface of the helicase and the excluded strand leading to alternative binding positions. Similarly, the R164A mutant also samples a greater number of states than wild-type, especially on the 40/30 and 50/30 substrates (Figure 2.14M & R). These results suggest that residues K180 and R164 contribute to but do not solely mediate the helicase-excluded strand interactions that give rise to the FRET states we observe for the wild-type.
Figure 2-14: Histograms and ExPRT Plots of WT and EcDnaB and mutants bound to DNA forks. Histograms (A-E) report the population of molecules as a function of FRET states on DNA forks with a 30-base 5’-strand and a 30-base (blue), 40-base (green), or 50-base (red) 3’-strand for WT, R74A, R164A, K180A, and R328A/R329A. Yellow, blue, and red regions highlight low, medium, and high FRET populations, respectively. Corresponding ExPRT plots are shown for (F-J) 30/30, (K-O) 40/30, and (P-T) 50/30 forks for each of the respective EcDnaB helicases.

EcDnaB (R74A) on the longer fork substrates displays similar dynamics between two FRET states compared to WT (Figure 2.14G, L, Q vs. F, K, P). Wild-type EcDnaB yields more traces with transitions than R74A does, and R74A produces histograms and ExPRT plots that lack the high FRET state (~0.95) that we see in all other data sets (Figure 2.14). The absence of that highest FRET state for R74A across all substrates tested and the presence of that same highest FRET state for all other EcDnaB constructs with all three substrates suggests that R74 is necessary for that particular FRET state. R74 is positioned close to the top of the hexamer, projecting outwards at three separate locations within the hexamer structure. The excluded strand
interaction at R74 would be consistent with the dyes being in extremely close proximity. Therefore, the R74A mutant provides insight into the excluded strand binding path on the exterior surface that transverses the entire lateral length of the exterior surface.

The EcDnaB (R328A/R329A) mutant shows extreme differences in the binding and dynamics on the longer DNA strands compared to wild-type (Figure 2.14E, J, O, T). Wild-type EcDnaB on the 50/30 fork shows a small fraction of traces that transition between high and medium FRET states (Figure 2.12E). In comparison, EcDnaB (R328A/R329A) produces almost entirely medium FRET states that are very dynamic, with a large number of transitions and relatively short dwell times indicative of severe destabilization of binding (Figure 2.12H & Figure 2.13F). The R328A/R329A mutation exists around the ‘waist’ of the hexamer, where the N-terminal domain sits atop the C-terminal domain. In contrast to wild-type, very little high FRET signal from the R328A/R329A mutant on the 30/50 substrate is observed. These results indicate that R328 and R329 may be required to stabilize longer excluded strands (40 and 50(dT)) around the waist of the helicase and mediate interactions with other regions of the hexamer.

Altogether, these mutants alter the interaction between the excluded strand and the exterior of the helicase to varying degrees. The amount of destabilization or altered external DNA binding paths can depend on excluded strand length and provide information on the contacts all along the lateral length of the hexamer. As emphasized above, these results are similar to those obtained previously for SsoMCM on fork DNA substrates; where it was determined that disruption of the wrapping interaction caused greatly reduced unwinding abilities.
2.4.5 Surface Mutants Enhance DNA Unwinding Activity

Gel-based fluorescent DNA unwinding assays were performed to determine whether these DnaB SEW mutants have any effect on EcDnaB’s activity. Figure 2.16A shows a representative six-minute time point. However, quantification of the unwinding rates occurred over multiple time points for each mutant (Figure 2.16B & Table 2.3). All of the mutants had increased unwinding rates compared to the wild-type EcDnaB. Specifically, R74A and R164A have greater than 3-fold and 6-fold increases, respectively, over the wild-type; while K180A and R328A/R329A have greater than 20-fold increases in unwinding activity.
Figure 2-16: Helicase unwinding and ATPase assays.

A) Representative gel for 6 minute time point is shown for EcoDnaB (WT and mutants) unwinding assays performed on a fluorescein labeled 30/30 fork DNA substrate and B) quantified over multiple time points. Throughout, data for EcoDnaB constructs are consistently colored (WT - red; R74A – ochre; R164A – orange; K180A – green; R328A/R329A – blue). C) Quantification of the ATP hydrolysis rate in the absence (diagonal hash) and presence of fork DNA (DNA14/DNA15) (solid).

Although all the mutations occur on the external surface, increases in DNA unwinding could potentially be correlated with increased ATP hydrolysis by the mutants. ATPase assays were performed and the rates quantified in the absence and presence of DNA (Figure 2.16C &
Table 2.3). R74A and R164A had similar basal ATPase rates compared to wild-type EcDnaB, but K180A and R328A/R329A had ~5-6 fold increased rates. The addition of DNA stimulated the ATPase rate of wild-type, R74A, and R164A ~15-20 fold above that of helicase alone. Interestingly, although K180A and R328A/R329A had increases in unwinding rates, they do not show significant stimulation of ATPase rates in the presence of DNA (Figure 2.16B-C).

**Figure 2-17: Quantification of Fork DNA Binding by EcDnaB.**
Normalized fluorescence anisotropy for WT (○ red), (b) R74A (□ ochre), (c) R164A (◊ orange), (d) K180A (△ green), and (e) R328A/R329A (▽ cyan) EcDnaB binding to fork DNA (DNA14F/DNA15) (inset). Error bars represent the standard error from three independent titrations. Lines are the fits to Equation 2, and the $K_d$ values reported in Table 2.3.

No stimulation in ATPase rate with DNA is sometimes indicative of a perturbation in DNA binding. Therefore, fluorescence anisotropy experiments were performed to quantify the binding affinity. Although there are subtle differences in binding affinity, all mutants have similar $K_a$s to that of wild-type (Figure 2.17 and Table 2.3). Upon closer inspection, R74A and R164A have slightly larger $K_d$ values indicative of the potential roles of these residues in binding
the excluded strand. The binding curves for K180A and R328A/R329A saturate at lower absolute change in anisotropy than the WT, R74A, and R164A. This may be due to oligomeric species smaller than hexamer binding the fork, the mutants being unable to load multiple hexamers onto the fork while the others can, or a decrease in helicase mobility when bound to the substrate. All EcDnaB constructs eluted from a gel filtration column at volumes corresponding to hexameric species and a titration of WT EcDnaB onto a forked substrate showed that only one hexamer bound a fork substrate, even at high concentrations (Figure 2.5). Therefore, we propose that these two mutants are less mobile on the DNA substrate. It has been shown that increased protein mobility on a DNA substrate leads to larger changes in anisotropy [204].

**Table 2-3: EcDnaB Kinetic Parameters.**

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>R74A</th>
<th>R164A</th>
<th>K180A</th>
<th>RR328AA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>dsDNA Unwinding (nM min⁻¹)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-DNA</td>
<td>28 ± 2</td>
<td>53 ± 15</td>
<td>199 ± 15</td>
<td>191 ± 15</td>
<td></td>
</tr>
<tr>
<td>+DNA</td>
<td>0.02 ± 0.01</td>
<td>0.03 ± 0.00</td>
<td>0.04 ± 0.00</td>
<td>0.12 ± 0.00</td>
<td>0.10 ± 0.00</td>
</tr>
<tr>
<td>ATPase rate (pmol min⁻¹)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-DNA</td>
<td>0.40 ± 0.01</td>
<td>0.42 ± 0.02</td>
<td>0.56 ± 0.01</td>
<td>0.13 ± 0.01</td>
<td>0.17 ± 0.01</td>
</tr>
<tr>
<td>+DNA</td>
<td>0.19 ± 0.01</td>
<td>0.03 ± 0.01</td>
<td>0.24 ± 0.01</td>
<td>0.10 ± 0.01</td>
<td>0.10 ± 0.01</td>
</tr>
<tr>
<td><strong>Fluor. dsDNA Unwinding</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-EcDnaC (min⁻¹)</td>
<td>0.12</td>
<td>0.17</td>
<td>0.14</td>
<td>0.82</td>
<td>0.33</td>
</tr>
<tr>
<td>+EcDnaC (min⁻¹)</td>
<td>1</td>
<td>0.97</td>
<td>1.00</td>
<td>0.75</td>
<td>0.35</td>
</tr>
<tr>
<td>Fract. Amp. Change⁴</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.39</td>
<td>1</td>
</tr>
<tr>
<td>DNA Binding Kd (µM)</td>
<td>1.52 ± 0.04</td>
<td>2.03 ± 0.03</td>
<td>2.51 ± 0.04</td>
<td>1.42 ± 0.06</td>
<td>1.74 ± 0.10</td>
</tr>
<tr>
<td>ssDNA Translocation (s⁻¹)⁵</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k₁</td>
<td>7.0 ± 0.1</td>
<td>0.04 ± 0.09</td>
<td>4.0 ± 0.1</td>
<td>4.5 ± 0.5</td>
<td>4.6 ± 0.6</td>
</tr>
<tr>
<td>k₂</td>
<td>0.36 ± 0.01</td>
<td>0.17 ± 0.02</td>
<td>0.04 ± 0.01</td>
<td>0.73 ± 0.03</td>
<td>0.93 ± 0.18</td>
</tr>
<tr>
<td>Fract. Amp. Change⁶</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.39</td>
<td>1</td>
</tr>
</tbody>
</table>

¹Calculated from gel based assay time courses. Errors represent the standard error from at least three independent experiments. ²Observed rates measured using DNA50 in stopped flow ssDNA translocation assays. ³Normalized to R328A/R329A maximal amplitude change. ⁴Normalized to WT EcDnaB/EcDnaC maximal amplitude change.

2.4.6 **Basis of K180A and R328A/R329A Unwinding Hyperactivity**

The bacterial helicase loader, EcDnaC, physically interacts with EcDnaB to induce structural changes leading to opening of the hexamer, loading onto a single-strand DNA, and stimulation of translocation and unwinding [198]. As the K180A and R328A/R329A mutants have increased...
unwinding rates, we hypothesized that these mutations may have altered EcDnaB into a more hyperactive conformation. To test this possibility, we performed fluorescence dsDNA unwinding experiments in both the absence and presence of the EcDnaC loader (Figure 2.18). A DNA fork was used that has a 6-FAM dye and a dabcyl quencher at the duplex end, where upon strand separation, the fluorescence is sensitized.

Figure 2-18: Fluorescence Unwinding Assays.
Averaged data from fluorescent DNA unwinding experiments for A) EcDnaB (250 nM): WT (red), R74A (ochre), R164A (orange), K180A (green), and R328A/R329A (cyan) or B) EcDnaB/EcDnaC and illustrated in the schematic. Background time dependent fluorescence changes from EcDnaC alone is subtracted from all traces. The black lines represent the fits to Equation 4. Comparison of the C) unwinding rates and D) the normalized amplitude changes in the absence (diagonal hashed) and presence (solid) of EcDnaC.
In the absence of EcDnaC, only K180A and R328A/R329A exhibit measureable unwinding activity (Figure 2.18A). A more significant lag phase exists for K180A compared to R328A/R329A, which may indicate a structural conformational change is needed for K180A to activate unwinding. Although the exponential rate is 2-fold faster for R328AA/R329A, the amplitude change is 4-5 fold greater for K180A (Table 2.3 & Fig. 2.14 A, C, & D). When EcDnaC is included in the assay, wild-type and the EcDnaB mutants (R74A & R164A) that had limited activity in the absence of EcDnaC have large increases in their rate and amplitudes (Figure 2.18B-D & Figure 2.19). For wild-type and R74A EcDnaB unwinding, there is a 10-fold increase in unwinding (Figure 2.18B-C & Table 2.3). The unwinding rate for R164A in the absence and presence of EcDnaC is similar, but the amplitude has increased dramatically with EcDnaC (Figure 2.18C-D). The appreciable lag phase with EcDnaB has been attributed previously to EcDnaC release after loading [198] and is also consistent in these studies. For K180A and R328A/R329A, neither the rate nor the amplitude are significantly affected by addition of EcDnaC. K180A and R328A/R329A are hyperactive helicases on their own, whereas, the other mutants (R74A & R164A) behave similar to wild-type where significant enhancements in unwinding occur in the presence of EcDnaC. These results are supportive of mutations, K180A and R328A/R329A, preinducing a structural change within the EcDnaB hexamer that promotes loading and unwinding in the absence of the helicase loader, EcDnaC.
Figure 2-19: Fluorescent DNA unwinding Comparison +/- EcDnaC.
Averaged data from fluorescent DNA (DNA39F/DNA39D) unwinding experiments for (a) WT (red), (b) R74A (ochre), (c) R164A (orange), (d) K180A (green), and (e) R328A/R329A (cyan) EcDnaB (250 nM) in the absence and presence of EcDnaC (1.125 μM). Background time dependent fluorescence changes from EcDnaC alone are subtracted from all traces. The black lines represent the fits to Equation 4. Rates and relative amplitudes are reported in Table 1.

2.4.7 ssDNA Translocation of Surface Mutants

The increased unwinding activity, especially for the K180A and R328A/R329A mutants, could be attributed to either an alteration in the wrapping of the excluded strand to control unwinding speed or a change the intrinsic translocation speed of these mutants on ssDNA. In order to directly monitor ssDNA translocation rates, stopped-flow fluorescence quenching experiments were performed (Figure 2.20). These presteady-state assays have been used to quantify rates of translocation and step sizes for a variety of monomeric and dimeric helicases [132, 214], but they have never been applied to hexameric helicases.
Figure 2-20: Presteady-state Fluorescence ssDNA Translocation Assays

A) Averaged presteady-state stopped flow data for wild-type and mutant EcDnaB ssDNA translocation assays. 400 nM DNA50 was rapidly mixed with 175 nM DnaB and 5 mM ATP at 22 °C. 

B) Comparison of the ssDNA (DNA50) translocation rate and amplitude for EcDnaB (R328A/R329A) with no nucleotide, AMPNP, ATP, or ATP/DnaC (1 μM). 

C) ssDNA translocation of EcDnaB (R328A/R329A) on different length DNA (30, 40, 50, or 75 bases) The black lines represent the fits to Equation 3. 

D) Fluorescent quenching is measured upon translocation to the 3’ end as indicated in the schematic. All assay conditions and concentrations are described in the Materials and Methods.
When a single-strand of 5’-fluorescein-labeled DNA was rapidly mixed with a given *Ec*DnaB construct, there is an ATP dependent stimulation in the fluorescence change. The helicase binds a single-strand of DNA and translocates towards the 3’ end of the oligo where the fluorescein fluorescence becomes quenched. Measured ssDNA translocation rates and amplitudes are greatly stimulated for both K180A and R328A/R329A (Figure 2.20A and Table 2.3). Although R164A has a faster apparent rate of change, the normalized fluorescence intensity change is only 0.03 compared to R328A/R329A. Neither WT nor R74A have any measurable change in fluorescence correlating with no measurable translocation propensity on ssDNA alone. The experimental setup is important, as preincubation of *Ec*DnaB with DNA and initiation with ATP shows little changes in fluorescence presumambly because the fluorescence is prequenched in the bound state by molecules predistributed along the DNA (Figure 2.21). ATP hydrolysis is also required for the large change in fluorescence as experiments with no nucleotide or the nonhydrolyzable analog, AMPPNP, have very small changes in fluorescence (Figure 2.20B & Figure 2.22). Therefore, the stopped-flow data suggests that the hyperactivity displayed by K180A and R328A/R329A is either because of an intrinsic change in the structure or conformation of *Ec*DnaB or a more efficient loading and encircling of DNA that enables faster translocation and unwinding.
Figure 2-21: Stopped-Flow ssDNA Translocation Order of Addition.
A) Schematic of the stopped-flow experiments for preincubation of DnaB (R328A/R329A) and DNA50 (1) or direct addition of DnaB/ATP (2). B) Measured stopped-flow fluorescence changes observed for schemes (1) and (2). The black lines represent the fits to Equation 3.
Figure 2-22: Fluorescent ssDNA Translocation of DnaB is Dependent on ATP Hydrolysis.

Fluorescent ssDNA Translocation of DnaB is Dependent on ATP Hydrolysis. Averaged presteady-state stopped flow data for A) WT and mutant: B) R74A C) R164A and D) K180A EcDnaB ssDNA translocation assays. 400 nM DNA50 was rapidly mixed with 175 nM DnaB and 5 mM nucleotide (AMPPNP or ATP) at 22 °C. The black lines represent the fits to Equation 3. Fluorescent quenching is measured upon translocation to the 3’ end as in Figure 2.20. All assay conditions and concentrations are described in the Materials and Methods.
The ssDNA length dependence on the observed rate has been used previously to globally fit the data to a nt/s translocation rate [132, 214]. This calculation requires that only a single turnover processive reaction is monitored to prevent contributions of rebinding impacting the observed rates. However, when we attempted to include heparin over a broad range of concentrations or excess salmon sperm ssDNA (data not shown) with ATP in the second syringe to trap EcDnaB molecules that have fallen off ssDNA during translocation, the fluorescence change became muted in all cases (Figure 2.23). Again, prebinding of EcDnaB to DNA prequenches the fluorescence before the experiment can be initiated not allowing for true single turnover rates to be measured.

Figure 2-23: ssDNA Translocation with Heparin as a Trap.
A) Schematic of the ssDNA translocation stopped-flow experiment where 175 nM EcDnaB (R328A/R329A) and 400 nM DNA50 were preincubated before initiating with 5 mM ATP and various concentrations of heparin trap: 4 mg/mL (red), 1.5 mg/mL (orange), and 0.1 mg/mL (green) B). As a control and experiment was also performed with DNA50 only and 0.1 mg/mL heparin (cyan).
However, presteady-state ssDNA translocation experiments performed with different length DNA templates do have length-dependent changes in the observed rates and amplitudes (Figure 2.20C & Figure 2.24). The largest effects were again seen with K180A and R328A/R329A. Generally, the shorter substrates (DNA30 and DNA40) showed double exponential decreases in fluorescence. Longer substrates (DNA50 and DNA75) had more complex changes in fluorescence and include a single exponential decrease followed by a single exponential increase. The difference seems to be related to the length of the DNA such that *EcDnaB* molecules that dissociate from the longer templates take longer to reengage the 5’ end and translocate allowing for a fluorescence increase to be detected prior to steady-state. When the concentration of K180A or R328/R329A was changed while keeping the DNA50 concentration constant (200 nM), the first exponential rate was not significantly affected (Figure 2.25), suggesting that *k_1* corresponds to a step after binding, most likely translocation. The multiple exponentials as well as not being able to effectively simulate single turnover conditions make comparisons of the translocation rates difficult. Therefore, the observed traces are a combination of individual rate constants for translocation, off, and reengagement.
Figure 2-24: ssDNA Translocation of DnaB Mutants with Varying Length ssDNA.
Averaged presteady-state stopped flow ssDNA translocation data for A) WT and mutant: B) R74A, C) R164A, and D) K180A EcDnaB as a function of DNA length (DNA30, DNA40, DNA50, DNA75). 400 nM DNA was rapidly mixed with 175 nM DnaB and 5 mM ATP. The black lines represent the fits to Equation 3. Fluorescent quenching is measured upon translocation to the 3’ end as in Figure 2.20.
For R328A/R329A on DNA30, the slower apparent initial exponential decrease ($k_1$) and larger change in amplitude compared to DNA40, DNA50 and DNA75 is most similar to a single turnover reaction (Figure 2.20C & Table 2.4). However, the presence of $k_2$ in all reactions complicates the interpretation as this rate is a combination of multiple individual rate constants. The faster apparent observed rates ($k_i$) as the DNA length is increased (DNA40, 50, 75) is an interesting observation that can most likely be explained by greater contributions of binding along the length of longer substrates, especially those greater than 40 bases. There may also be a greater contribution of dissociation ($k_{off}$) on the longer substrate that prematurely ends the first exponential decrease. For K180A and R328A/R329A, the estimated ssDNA translocation rates are ~60-160 nts/sec at 22 °C but could be as high as 230-400 nts/sec if loading onto shorter length templates is restrictive. These rates are consistent with previous rate measurements of unwinding after loading [111, 215].

Table 2-4: EcDnaB ssDNA Translocation.

<table>
<thead>
<tr>
<th>Observed Rates (s⁻¹)</th>
<th>ssDNA Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td>K180A $k_1$</td>
<td>2.0 ± 0.2</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.40 ± 0.12</td>
</tr>
<tr>
<td>R328A/R329A $k_1$</td>
<td>2.3 ± 0.2</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.56 ± 0.10</td>
</tr>
</tbody>
</table>

$^1$Errors represent the standard error from at least three independent stopped flow ssDNA translocation experiments.
Figure 2-25: DnaB Concentration Dependence on Translocation.

Averaged presteady-state stopped flow ssDNA translocation data for A) K180A and B) R328A/R329A on DNA50 at different concentrations (as indicated). Rate (k1), for K180A are 4.7, 4.6, and 4.5 s⁻¹, at 47, 94, and 188 nM, respectively. Rate (k1) for R328A/R329A are 4.8, 4.5, 3.9, and 4.8 s⁻¹, at 47, 94, 141, and 188 nM, respectively. The black lines represent the fits to Equation 3.
2.5 DISCUSSION

2.5.1 EcDnaB and SsoMCM Interact Similarly with the Excluded Strand

Although hexameric DNA replication helicases have global structural conservation, their amino acid sequences are not conserved allowing for the classification of these helicases into different superfamilies (SF). We showed previously that the 5’ excluded strand makes important external surface interactions that aid in the mechanism of unwinding for the SF6 archaeal MCM helicase (3’-5’) to develop the steric exclusion and wraping model (SEW) of unwinding [96]. Here, we show that the bacterial replication helicase, EcDnaB, with opposite unwinding polarity (5’-3’) and of a different family (SF4), makes almost identical contacts with the excluded 3’-strand (Figure 2.7). In order to better visualize and directly compare both the states and dynamics of the interactions, we developed novel smFRET analyses that utilize ExPRT plots. ExPRT plots improve on existing analyses by simultaneously comparing transition, probabilities, and dwell times on a single plot. ExPRT plots can even represent more than one transition rate between states if applicable and can be applied to a variety of smFRET studies. Finally, some external surface mutations used to test the SEW model for EcDnaB have revealed unexpected stimulation in DNA translocation and unwinding that we can correlate with more efficient loading onto DNA in the absence of the loader, EcDnaC. The combined results highlight the importance and conservation of the SEW model for hexameric helicase unwinding of DNA and implicate external surface residues in controlling the conformational transition required for loading and activating the EcDnaB helicase.

It is striking that the absolute FRET states, transition probabilities, and dwell times are extremely similar between SsoMCM and EcDnaB hexameric helicases bound to fork DNA. In
both cases, the large increase in FRET observed is consistent with encircling of the translocating strand and exclusion of the other. For DnaB, it is widely held that the hexamer will encircle the 5’-strand, but that multiple hexamers could load on opposing strands in opposite orientations [94, 112, 114, 216]. The consequence of loading two hexamers would ultimately separate the strands further decreasing the FRET because of the size of the hexamers, which is opposite to what we observe in the smFRET experiments even at high concentrations. Therefore, EcDnaB binding of DNA includes both the encircling of the 5’-strand and the exclusion and external interaction of the 3’-strand.

2.5.2 Probing EcDnaB Excluded Strand Binding Path via Mutational Analysis

Without an appropriate DNA-bound crystal structure of EcDnaB, we had to infer binding positions for the excluded 3’-strand based on amino acid homology and electrostatics from crystal structures that represented a closed ring [77, 196] or a split lock washer structure [70]. Because ssDNA was contained in the central channel of the split lock washer structure, we used this conformation as a primary model to test interactions with the excluded strand. To test the specificity of this external interaction, conserved mutations were found to both disrupt and alter the path for binding. In particular, R74 was found to be important for stable interactions at the N-terminal tier, giving rise to the highest FRET state (~0.95). R164 becomes more important as the length of the excluded strand increases and upon mutation gives rise to many more FRET states with increased dynamics. K180, R328, and R329 all seem to be influential in wrapping longer ssDNA substrates at the waist; where upon mutation, the values shift from a medium (~0.8) to a lower (~0.5) FRET state consistent with decreased wrapping. In all cases, mutations of these external SEW residues increased the number of FRET states or altered the dwell times and
transition probabilities. A representation of the EcDnaB-excluded strand interaction is shown in Figure 2.26. It is likely that a large network of residues, including those that were mutated, coordinate the interaction with the excluded strand.

![Figure 2-26: Representative Model of EcDnaB Excluded Strand Interaction.](image)

A homology model of *E.coli* DnaB is shown with each subunit of the hexamer colored separately. Mutated residues are colored in blue. A potential interaction mode and exterior binding path with DNA is represented as hashed black lines, with Cy3 and Cy5 dyes on the 5’ and 3’ fork ends.

2.5.3 Disrupting Excluded Strand Interactions Stimulates EcDnaB Unwinding

Previously, mutation of external positively charged residues on *Sso*MCM reduced DNA binding presumably through a slippage mechanism where the mutant helicase was unable to stabilize forward unwinding steps [96]. Conversely for EcDnaB, mutation of external positively charged residues along the lateral length of the hexamer generally increased translocation and unwinding
rates. The enzymatic effects of the mutations generally correlate with two classes of results. The more intriguing class of mutations (R180A and R328A/R329A) seem to allow EcDnaB to exist in a structural conformation that promotes DNA loading and hyperactivity on its own. Previously, the R328A/R329A mutant was investigated for its potential role in a leucine zipper motif [217], although later discounted [218], and also found not to have DNA-stimulated ATPase activity. In fact, interactions with EcDnaC do not stimulate the unwinding or translocation of these already active EcDnaB mutants to the same degree as with WT. It is also possible that mutation of K180 and R328/R329 have disrupted inter-subunit salt bridges that destabilize adjacent contacts somewhat and make it easier to crack open the ring. This effect would be modest as the oligomeric state of these mutants is still consistent with a hexamer. Therefore, we suggest that these waist mutations do more than just alter binding of the excluded stand and additionally destabilize subunit interactions and hexamer symmetry contributing to more facile DNA loading and unwinding. This conformational switch is normally induced through the action of EcDnaC which remodels the complex into a split-ring spiral confirmation that is more competent for loading and subsequent unwinding [198, 219]. However, these mutants show an increased propensity for translocation and unwinding on their own. This propensity would suggest that they exist in a hybrid state between a closed hexameric ring and an EcDnaC activated split-ring conformation that may also aid in subunit shifts that are proposed to occur during unwinding [70].

The other two mutations (R74A and R164A) also show stimulated unwinding activity both in the absence and presence of EcDnaC compared to wild-type. As R74A and R164A alter both the highest FRET state and contribute to more binding dynamics than wild-type, they become more representative of the effects of excluded strand disruption on DNA unwinding. For
these second class of mutations, increased DNA unwinding is still observed both in the absence and presence of EcDnaC, highlighting the role of excluded strand interactions in modulating the rate of unwinding.

2.5.4 Comparison with Previous Studies on EcDnaB Unwinding Mechanism

Previous single-molecule work has detailed the force contributions of each DNA strand to unwinding by EcDnaB using either a hairpin or fork substrate [215]. In that study, it was concluded that the unwinding rate was controlled by both force-induced destabilization of the duplex as well as interactions of the excluded strand with the exterior surface. The main apparent discrepancy between our work and theirs is that when the excluded strand is sequestered because of constraints in the hairpin assay, the rate is slower than when it is allowed to interact with the exterior surface in the fork assay. This would imply that contacts of the excluded strand with the external surface of EcDnaB increase unwinding; whereas we show that specific external contacts restrain unwinding. However, it is probable that the force applied to the excluded strand in the fork assay artificially alters the interaction with the exterior surface in a way analogous to the altered DNA binding paths and kinetics for the R74A and R164A mutants. Therefore, measured increases in unwinding in both studies can be explained by altered DNA binding paths on the exterior surface. It is notable that a variety of recent biophysical techniques monitoring EcDnaB activity and binding have detected an elusive external interaction with the excluded strand [111, 215]. We can now conclude that this SEW interaction and the precise binding path regulates the speed of unwinding for EcDnaB.
2.5.5 Potential In Vivo Roles of Excluded Strand Interactions

Although similar external SEW interactions exist for both the archaeal SF6 3’-5’ MCM helicase and the bacterial SF4 5’-3’ helicase independent of evolution, the effects on unwinding are opposing (Figure 2.27). In both cases, electrostatic residues primarily contribute to binding of the excluded strands, but a comparison of the exact binding paths on \textit{Sso}MCM [97] to \textit{Ec}DnaB is not possible at this time. For each helicase, the motor domain is positioned adjacent to the duplex DNA [119, 220], placing the leading strand on the exterior surface of DnaB and the lagging strand on the exterior of MCM. Electrostatic mutations on the surface of MCM abrogate DNA unwinding [96, 97, 221], while similar mutations on the surface of DnaB shown here increase unwinding. Clearly, interactions with the excluded strand act to control the speed of unwinding. It remains to be seen whether this regulation is because of a greater increased force applied by the motor domain for DnaB that is modulated by the excluded strand or whether discrete external binding paths or polarity dictate the rate of unwinding. For DnaB, the excluded strand interaction may act as a ‘molecular brake’ to control the amount of exposed ssDNA or provide a platform for accessory helicases, \textit{i.e.} Rep, to assemble and rescue stalled forks [188, 222]. Alternatively, coupled DNA synthesis by the leading strand polymerase (Pol III) could sequester the excluded strand from the exterior surface of DnaB and explain the increased rate of unwinding by the coupled replisome [183, 184, 187]. In addition to controlling the unwinding rate, the external DNA binding sites on both helicases are likely to contribute during the loading mechanism for encircling of the translocating strand to maintain strand separation during the action and conformational changes induced by the initiation enzymes.
Figure 2-27: **SEW models for hexameric helicase unwinding.**
Bacterial DnaB and archaeal MCM encircle the lagging or leading strand, respectively, and interact with the excluded strand on the exterior surface to either regulate the unwinding rate using this molecular ratchet or stabilize unwinding in a forward direction.

2.6 **ACKNOWLEDGEMENTS**

pET11b-*EcDnaB* was a generous gift from Dan Kaplan. The plasmid expressing *EcDnaC* was provided by Valerie O’Shea and James Berger. We thank the Walter lab for access to their POKIT analysis program. We are grateful to Dr. Bennett Van Houten (University of Pittsburgh) for the use of his fluorescence spectrophotometer and to Muwen Kong for assistance with the fluorescence anisotropy and *EcDnaB-EcDnaC* unwinding assays. We thank Brian Graham, Grant Schauer, and Valerie O’Shea for helpful discussions.
2.7 FUNDING

This work was supported by the University of Pittsburgh, Baylor University, a Research Scholar Grant (RSG-11-049-01-DMC to M.A.T.) from the American Cancer Society, and the NIH (AI081571 and GM068406).

2.8 CONTRIBUTIONS

Sean Carney, Heather McFarland, Sanford Leuba, and Michael Trakselis designed the experiments. Sean Carney performed the cloning, protein purification, biochemical assays, and single-molecule experiments. Heather McFarland performed the translocation assays.
3.0 MITOCHONDRIAL TWINKLE HELICASE EXCLUDED STRAND WRAPPING

3.1 SUMMARY

Mutations in the c10orf2 gene encoding the human mitochondrial DNA replicative helicase Twinkle are linked to several rare genetic diseases characterized by mitochondrial defects. In this study, we examined the catalytic activity of Twinkle helicase on model replication fork structures. Novel single molecule FRET binding and unwinding assays show an interaction of the excluded strand with Twinkle as well as events corresponding to stepwise unwinding and annealing. These studies shed new light on the catalytic functions of Twinkle on key DNA structures it would encounter during replication or possibly repair of the mitochondrial genome.

3.2 INTRODUCTION

The gene product of c10orf2, also known as Twinkle, is a DNA helicase required for the replication of human mitochondrial DNA (mtDNA) [61]. (1). Mutations in the c10orf2 gene

---

encoding Twinkle helicase lead to mitochondrial deletions in post-mitotic tissues and are responsible for a number of hereditary disorders including adult-onset progressive external ophthalmoplegia, hepatocerebral syndrome with mtDNA depletion syndrome, and infantile-onset spinocerebellar ataxia [60]. Twinkle helicase collaborates with DNA polymerase γ (POLγA) and its associated processivity factor (POLγB), and the mitochondrial single-stranded DNA binding protein (mtSSB) as the minimally reconstituted mtDNA replisome [223]. While there has been much debate and interest in the mechanism(s) underlying mtDNA synthesis, it is generally thought based on experimental evidence that strand displacement DNA synthesis occurs for both the light and heavy guanine-rich strands of the circular double-stranded mitochondrial genome [224]. Because there are two origins of replication for the heavy and light strands, a three-stranded displacement loop structure known as the D-loop is created. Displacement DNA synthesis by POLγ is stimulated by the DNA unwinding activity catalyzed by Twinkle helicase.

A number of biophysical and biochemical studies have provided insight to the molecular and cellular functions of the mitochondrial replicative Twinkle DNA helicase. Twinkle is a member of the Superfamily 4 DNA helicases, which includes the E. coli DnaB replicative helicase as well as the bacteriophage T7 gene 4 helicase [225, 226]. These helicases are known to form ring-like structures [42], and Twinkle’s stoichiometry is that of a 6-subunit or 7-subunit architecture that can be modulated by solution components [62, 227]. Twinkle catalyzes unwinding of duplex DNA with a 5’ to 3’ directionality fueled by the hydrolysis of nucleoside triphosphate [228-230]. Twinkle can load onto single-stranded DNA circles [228, 230] or double-stranded DNA bubble structures [230], and binds both single-stranded and double-stranded DNA [228]. Twinkle anneals complementary single-stranded DNA [228], but the physiological significance of this biochemical activity remains to be understood. Although
Twinkle plays a key role in separating strands at the mitochondrial replication fork, and the enzyme is weakly active on G-quadruplex DNA structures [231], which are likely to form in the guanine-rich strand of the mitochondrial genome [231, 232]. Presumably other proteins, including auxiliary DNA helicases (PIF1, DNA2, SUV3, RECQ4) [233], may either assist or substitute for Twinkle in dealing with certain alternate DNA structures.

Despite the wealth of information that has been gained from molecular studies of Twinkle helicase, its precise role(s) in mtDNA replication remains to be fully characterized. Moreover, it is unclear if Twinkle plays any role in mtDNA repair. In this work, Twinkle was tested on a series of DNA structures associated with the replication fork to ascertain the enzyme’s substrate specificity and activities. The results from these studies provide new insight on Twinkle’s catalytic functions on key DNA structures it would encounter during the replication of the mitochondrial genome and potentially other processes including DNA repair in the organelle.

3.3 MATERIALS AND METHODS

3.3.1 Materials

Recombinant human Twinkle protein (gene product of c10orf2) with a C-terminal hexahistidine tag was expressed in bacterial cells, purified, and protein concentration determined by extinction coefficient as described previously [228]. Oligonucleotides used (Table 3.1) were purchased from IDT (Coralville, IA). Fluorescently labeled DNA was HPLC purified (IDT).
### Table 3-1: DNA Substrate Sequences.

<table>
<thead>
<tr>
<th>Oligo Number</th>
<th>Oligo Name</th>
<th>Length</th>
<th>(nt) Oligo Sequence (5’3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>DNA45</td>
<td>50</td>
<td>(B)TGCGACGGCAAGCAAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>33</td>
<td>DNA44</td>
<td>50</td>
<td>(S)TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>34</td>
<td>DNA116</td>
<td>60</td>
<td>(S)TGCGACGGCAAGCAAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>35</td>
<td>DNA111</td>
<td>70</td>
<td>(S)TGCGACGGCAAGCAAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>36</td>
<td>DNA140</td>
<td>58</td>
<td>(S)ACCTTCTACCATGAGAGCATTTCTAGTAC(5)AACGTCTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>37</td>
<td>DNA141</td>
<td>59</td>
<td>TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
</tbody>
</table>

B (biotin); 3 (Cy3); 5 (Cy5)

#### 3.3.2 Single-molecule FRET—based DNA binding and unwinding measurements.

The oligonucleotides (IDT Corp, Coralville, IA) used to create the fluorescently labeled DNA substrates are listed in Table 3.1. The DNA substrates were immobilized on a pegylated quartz slide via biotin-streptavidin interactions [202, 203]. The smFRET buffer conditions were 20 mM Tris-HCl (pH 7.5), 1 mM DTT, 5 mM AMP-PNP, and 10 mM MgCl₂ as well as an oxygen scavenging system (1 mg/mL glucose oxidase, 0.4 % (w/v) D-glucose, 0.04 mg/mL catalase, and 2 mM trolox). The concentration of Twinkle used in the assays was 25 nM. Unwinding assays were initiated by addition of 5 mM ATP (final concentration). Data were collected on a prism-based total internal reflection microscope at 10 frames per second for 5 or more regions, with each region containing 50-250 single molecules. A 532 nm diode laser was used to excite the Cy3 fluorophore. The resulting Cy3 and Cy5 fluorescence signals were separated using a 610 nm dichroic longpass mirror, a 580/40 bandpass filter, a 660 nm longpass filter, and then imaged by an EM-CCD iXon camera (Andor, Belfast, UK). Regions of signal intensity were fit to 2D Gaussians in order to identify single-molecule FRET pairs, and subsequently corrected for thermal drift and local background intensity [205, 206]. FRET Efficiency, $E_{app}$, was calculated using
\[ E_{\text{app}} = \frac{I_A}{I_A + I_D} \]  

where \( I_A \) and \( I_D \) are the intensities of the acceptor and donor signals, respectively.

3.3.3 Single-molecule FRET data analysis for DNA fork binding and unwinding assays.

Traces were stitched together when the experimental conditions were identical, and the resulting trace was fit to ideal states using vbFRET software package [207]. The number of FRET states was determined by fitting the stitched trace to the maximum number of states where the states are \( E_{\text{app}} = 0.1 \) away from one another and the variation of one state does not overlap with another. Once fit to ideal states, traces were then unstitched and analyzed utilizing the ExPRT (Explicit Probability and Rate Transition) analysis program. The ExPRT Plot analysis produces transition plots from ideal traces where a marker’s position corresponds to a given transition where the initial state corresponds to the x-axis position and the final state corresponds to the y-axis position. The size of the marker corresponds to the fraction of all traces that contain that particular transition, and the color of the marker corresponds to the dwell time of an initial state preceding the transition. The collected dwell times for each transition are fit to a single or double exponential survival curve. If the \( R^2 \) value of the double exponential fit was greater than 0.970 and increased from the \( R^2 \) value of the single exponential fit by more than 0.015, then the rates from the double exponential fit were used to create a concentric marker with two colors, each representing a dwell time based on the two rates. Otherwise, the dwell time was defined as the average of all measured dwell times for a given transition. Only dwell times that were both preceded and followed by transitions were taken into account in this analysis.
For the unwinding assay, each individual trace was fit to ideal states using vbFRET software. Unwinding and rewinding events were then identified, and the number of steps and the time for unwinding or rewinding to occur were collected. The unwinding time is defined as starting from the point in which there is a transition from the high FRET state and concludes when the lowest FRET state is reached. Similarly the rewinding time is defined at that point when the lowest FRET state is departed until the highest FRET state is reached. Histograms were generated from these data and fit with a standard Gaussian equation.

3.4 RESULTS

3.4.1 Twinkle dynamically interacts with both DNA strands.

To test for a direct interaction between the mitochondrial helicase, Twinkle, and the nontranslocating strand of the model DNA fork, single-molecule FRET (smFRET) experiments were performed as described previously [96, 189]. Three FRET-labeled DNA fork substrates (30/30, 40/30, 50/30) were tested that had 30 bases (dT) on the 5’-arm of the fork where the helicase loads, and increasing 30, 40, and 50 bases (dT) on the 3’ excluded arm of the fork. Histograms for the immobilized DNA forks alone produced a single stable low-FRET peak corresponding to the termini of the fork arms not being in close proximity (Figure 3.1A). Increasingly longer 3’ fork arms correspond to lower FRET signals as expected.

When Twinkle was added to each of these forks, there was a concomitant shift to higher FRET states for all three substrates in the histogram (Figure 3.1A). On the 30/30 substrate, Twinkle’s interaction with the fork results in a predominantly high-FRET population with
additional smaller medium FRET populations. The 30/50 fork produced a nearly identical histogram, while the 30/40 substrate had a larger low and medium-FRET populations with a smaller high-FRET population. In the case of all three substrates, we see a significant high-FRET population, which has previously been shown to correspond to a wrapping interaction where the excluded-strand wraps around the outer surface of the helicase [96, 115, 189]. The presence of multiple populations within the histogram data indicates that there are multiple FRET states sampled by each model fork substrate upon Twinkle binding. However, the histogram alone does not reveal any information concerning possible dynamic transitioning between these states.

To probe the dynamics of the interaction, we generated ExPRT plots (Figure 3.1B-D) for each data set. The novel ExPRT analysis quantifies all unique transitions, explicit probabilities, and dwell times on a single plot for all smFRET traces containing two or more transitions. Each marker represents a single transition from an initial FRET state on the x-axis to a final FRET state on the y-axis. The size of the marker corresponds to the fraction of traces analyzed that exhibited that specific transition, and the color of the marker corresponds to the average dwell time spent in the initial state before the transition (Figure 3.1E). The number of FRET states decreases (6 to 5 to 4) as the length of the excluded-strand increases, correlating with greater stabilization of the excluded-strand interaction along its binding path(s). Despite this difference, there are also similarities among the three ExPRT plots. For example, there is significantly greater stability of the excluded-strand interaction corresponding to the highest-FRET state in each ExPRT plot as the length of the 3'-strand increases. Markers indicating the highest initial FRET state have longer dwell times, mostly between five and eight seconds. All other markers representing transitions between alternative FRET states, even those that return to the highest FRET state, have shorter dwell times between one and four seconds. Another similarity exists
between the 30/30 (Figure 3.1B) and 40/30 (Figure 3.1C) datasets in which the largest markers, which are the most commonly seen transitions, occur between adjacent FRET states. This is consistent with multiple points of contact between the excluded-strand and the exterior surface of the helicase, where in most cases there is a ‘stepping’ in the disruption or formation of these contacts such that most transitions are occurring between adjacent states. Transitions occurring between neighboring states are indicative of wrapping the excluded-strand around the hexamer to form the most stably bound high-FRET state. The ‘peeling off’ of this strand, where adjacent contacts are disrupted one at a time, leads to the less stable lower FRET states. The ExPRT plot for Twinkle on the 50/30 fork (Figure 3.1D) shows that the highest initial FRET state can transition into either the second or third highest final FRET states with roughly equal probability. This may be a consequence of the longer excluded-strand arm of the fork being able to occupy distinct pathways on the external surface of the helicase.
Figure 3-1: Twinkle Binding and Unwinding DNA Forks.
Histograms and ExPRT plots of Twinkle bound to DNA forks. A) A histogram of the FRET signals from the 30/30 DNA fork substrate (oligo 32/oligo 33) (blue), 40/30 DNA fork substrate (oligo 33/oligo 34) (green), and 50/30 DNA fork substrate (oligo 44/oligo 111) (red) DNA fork substrates alone (dashed) and after addition of 25 nM Twinkle (solid). The length of the 5’ translocating strand was kept constant at 30 (dT) nucleotides, and the length of the 3’ excluded strand was varied from 30 to 50 (dT) nucleotides. Yellow, blue, and red regions indicate low, medium, and high FRET states, respectively. ExPRT plots for Twinkle bound to the B) 30/30, C) 40/30, and D) 50/30 DNA forks. Each marker on the ExPRT plot represents a transition from the initial FRET state on the x-axis to the final FRET state on the y-axis. E) The size of the marker corresponds to the fraction of analyzed traces that exhibit that particular transition (transition probability), and the color represents the dwell time (sec) of the initial state. Representative traces from smFRET experiments showing F) unidirectional unwinding or I) alternating unwinding and reannealing for a DNA fork (oligo 36/oligo 37). The calculated FRET signal is shown in blue, while the fit to ideal states is overlaid in red. A global quantification of the unwinding time or number of steps for (B & H) unidirectional unwinding or (J & K) alternating unwinding and reannealing. For multistate smFRET traces, the individual unwinding (dark grey) and annealing (light grey) times and steps are quantified individually. The lines correspond to fits to a Gaussian equation.
In order to directly monitor the specific DNA unwinding and reannealing of DNA by Twinkle, we again utilized smFRET. The biotinylated fork DNA substrate was altered to include internal Cy3 and Cy5 dyes six base pairs apart in the duplex region adjacent to the fork arms (substrate 26). Upon addition of ATP, Twinkle displayed single molecule FRET traces consistent with either unidirectional stepwise unwinding or alternating unwinding/reannealing at a 2:1 ratio, respectively, over the acquisition time (100 sec) (Figure 3.1 F and I). It should be noted that the time frame of our experiments is limited by the lifetime of the fluorescent dyes, and that events that take place on the time scale of minutes are lost. Therefore, if an unwinding event takes tens of seconds to occur, a subsequent rewinding event would likely not be captured. The traces shown in Figure 3.1F and Figure 3.1I also demonstrate that Twinkle does not always unwind the fork substrate completely. This could be a result of the biotin-streptavidin junction at the base of the duplex that could act as a block to the helicase as shown previously for the NS3 helicase [234] and the T7 gp41-61 helicase-primase [235]. To be certain that step-wise changes in FRET corresponded to the action of Twinkle, smFRET experiments were compared using ATP or the nonhydrolyzable analog AMP-PNP. Upon addition of AMP-PNP, ~2.5% and 4% of observed molecules display clear unwinding/reannealing and unwinding events, respectively. When ATP is added, ~17% and 8% of the molecules display clear unwinding/reannealing and unwinding events, respectively. These results are consistent with a previous bulk study that showed Twinkle annealing activity can take place in the presence of a non-hydrolyzable nucleotide analogue or in the absence of nucleotide, but rates of annealing increased when a hydrolyzable nucleotide was added [228]. Using smFRET, we can quantify the steps taken by Twinkle for unwinding or reannealing of the fork as well as the time it takes for these events to occur. A
quantification of the unwinding or reannealing times for all traces show that the mean unwinding time is 4-fold greater when unidirectionally unwinding than when molecules are alternating between unwinding and reannealing (Figure 3.1G versus Figure 3.1J). Unidirectional unwinding also tended to have a greater number of steps (2.3 ± 0.7) for the progression on our fork substrate compared with alternating unwinding/reannealing events where the average step numbers were 1.7 ± 1.0 and 1.4 ± 0.8, respectively for each. The cumulative smFRET data suggest that Twinkle on its own may occasionally release its grip on the translocating strand and allow the DNA to reanneal. However, maintaining contact with the excluded strand or reengagement on the translocating strand can reverse the regression to reactivate unwinding in a forward direction.

3.5 DISCUSSION

In this study, we have undertaken a biochemical analysis of the ATP-dependent catalytic strand separation activities of the purified human recombinant mitochondrial DNA helicase Twinkle. This work has provided new insight to Twinkle’s interactions with forked DNA substrates as well as its ATP-dependent catalytic activity on such substrates. A dynamic interaction with the non-translocating strand has been identified, and utilizing smFRET to study helicase unwinding has revealed previously unresolved dynamics as well as validated a previously reported annealing activity.
3.5.1 DNA Unwinding and Reannealing by Twinkle

Consistent with previous work showing that Twinkle is an unwinding and annealing helicase [228], we now show at the single molecule level that during step-wise unwinding, Twinkle will regress and anneal duplex DNA more than half the time. The repetitive release and reengagement of the DNA strand may be important in processing D-loops and three-stranded junctions. The low efficiency in DNA unwinding could be enhanced through binding of mitochondrial SSB, which is known to stimulate Twinkle helicase activity [229]. Alternatively, or in addition, the coupled action of DNA synthesis by mtDNA polymerase γ with Twinkle unwinding may help to stabilize the unwound strands, similar to what was reported for the bacteriophage T7 DNA polymerase and helicase in which polymerase rapidly traps single-stranded bases arising from helicase action at the fork [236]. Our smFRET fork-binding assays show that multiple binding paths and contacts on the helicase exterior are sampled by the non-translocating strand. It is possible that some or all of these binding interactions could play a role in Twinkle’s unwinding and annealing activities.

3.5.2 Interactions Between Twinkle and the Excluded Strand

smFRET measurements of DNA binding strongly suggest that Twinkle makes multiple points of contact between the excluded strand and the exterior surface of the helicase. The number of smFRET states and transitions for Twinkle binding the excluded strand are much greater than that seen for SsoMCM or EcDnaB [96, 115, 189]; however, the dwell times in each state are similar. This implies that Twinkle does not necessarily have greater dynamics for binding the excluded strand; rather, it samples more exterior binding paths, especially for the 30/30 fork. On
the other hand, Twinkle binding to DNA is similar to both EcDnaB and SsoMCM in that the largest dwell times occur at the highest FRET states and the probabilities for transitions between the two highest FRET states are greater than lower, less abundant FRET states. This implies that the most stable surface-excluded strand interaction occurs when the non-translocating strand is tightly bound to the helicase exterior via multiple points of contact. Previously published biochemical data suggest that the sequence- and structure-related T7 gene 4 helicase and EcDnaB make contacts with the displaced strand during DNA unwinding [192, 193, 237].

Further studies that address this proposed mechanism of action by Twinkle in a biological setting will be informative to explain perhaps an unappreciated role of Twinkle to deal with mitochondrial genome replication stress. For example, recent work demonstrated that Twinkle overexpression suppresses oxidative induced replication stalling in cardiomyocyte mitochondria, thereby reducing mtDNA point mutations and rearrangements [238].

3.6 ACKNOWLEDGEMENTS

We thank members of the Laboratory of Molecular Gerontology (NIA-NIH) for helpful discussion. We are grateful to Becky Kucera and Dr. Jurate Bitinaite (New England Biolabs) for the BamHI-E111A-purified protein. We thank Sanford Leuba (Univ. Pittsburgh) for continued access to the smFRET instrumentation.
3.7 AUTHOR CONTRIBUTIONS

Sean Carney and Michael Trakselis designed, performed, and analyzed the experiments shown in Figure 3.1. Irfan Khan, Jack Crouch, Sanjay Bharti, Joshua Sommers, and Robert Brosh purified Twinkle and performed biochemical assays not shown here.
4.0 CHARACTERIZING SSOSSB BINDING AND INTERACTION WITH THE HELICASE-DNA FORK COMPLEX

4.1 SUMMARY

Single-strand binding proteins (SSBs) are ubiquitous across the domains of life and have essential roles in DNA metabolism. The archaeal SSB from *Sulfolobus solfataricus* has emerged as an interesting protein as its functional properties are distinct from the model SSBs that have typically been studied to understand this class of proteins. However, several aspects concerning the protein’s properties and interactions support contradictory models. The oligomeric state has been proposed to be a monomer compared to other groups who reported dimer/tetramer organizations in the absence of DNA. Also, a wide range of binding affinities have been observed, and this has not been directly addressed by any study. There have been observations of both stimulatory and inhibitory effects that *SsoSSB* has on the *SsoMCM* helicase. We have measured the binding strength of *SsoSSB*, and we propose a length-dependent explanation for the wide-ranging affinities that have been reported. We also studied the mode of interaction of *SsoSSB* on ssDNA and found that it interacts in a way that there is little to no disruption of the ssDNA’s conformational flexibility. In a first step to determining the effect that *SsoSSB* has on *SsoMCM* helicase’s unwinding activity, we have shown that the interaction between *SsoSSB* and the DNA-helicase complex may be dependent on the structure of the DNA fork.
4.2 INTRODUCTION

The function of helicases in DNA replication is essential, but they do not function independently. A growing body of evidence suggests that replicative helicases are regulated by various members of the replisome during DNA replication[239]. Polymerases have been shown to increase the rate and processivity of helicase unwinding and translocation [182, 184, 186, 236, 240-243]. Primases are also found to interact tightly with helicases, which localizes them to the site of priming [63, 64, 77, 244-246]. The primase must repetitively synthesize small RNA oligos on the lagging strand, and it has been suggested in some cases that its interaction and activities regulate the progression of the entire replication fork [247]. Another replisome component of interest is the SSB protein, which is thought to stabilize temporary ssDNA segments intermediates prior to duplication. Together, the active helicase complex, the DNA primase, and SSB comprise the unwindosome complex.

SSB proteins interact with ssDNA intermediates during DNA replication and repair. They mediate these nucleic acid metabolic processes by interacting with, displacing, and recruiting other proteins to the sites of interest. SSBs have been shown to interact with and stimulate dsDNA unwinding activity for several model replicative helicases presumably by stabilizing separated DNA strands [229, 248-251]. Archaeal organisms have served as models for understanding DNA replication [252, 253], and recent work from our lab and others has provided insight into the interactions and activities of various archaeal unwindosome enzymes including the MCM helicase [12, 56, 65, 66, 69, 75, 96, 117-119, 174, 175, 189] and the DNA primase [174]. A lesser-understood protein that has been implicated in archaeal DNA replication is the archaeal SSB protein.
An initial identification and characterization of the SSB protein from *Sulfolobus solfataricus* demonstrated that each monomer of this protein binds ~5 nucleotides of ssDNA by the single OB fold contained within each monomer [254]. The interaction between the OB fold and ssDNA is highlighted in Figure 4.1B-C using the SsoSSB-ssDNA structure [255]. Although the binding site size of 5 nucleotides is generally agreed upon, there have been conflicting reports on the oligomeric state of SsoSSB. Some groups have observed dimeric and tetrameric assemblies of SsoSSB [256-258], while other groups observe a strictly monomeric form of SsoSSB [254, 255, 259]. In addition to the OB fold that mediates ssDNA binding, SsoSSB also contains a flexible C-terminal tail with regions of positive and negative charge [254, 260, 261]. A flexible C-terminal region has also been identified in other bacterial, viral, and eukaryotic SSBs as shown in Figure 4.1A [262-269], and these C-terminal regions have been shown to mediate interactions with a variety of other proteins and DNA replication and repair processes [265-267, 270]. Several studies have reported that SsoSSB also interacts with protein partners through its C-terminal region. SsoSSB interacts with several proteins such as reverse gyrase [271], RadA [256], RNA polymerase [272], XPB, topoisomerase, and others [273] involved in nucleic acid maintenance pathways [274]. SsoSSB has also been studied in the context of DNA replication.

The archaeal SSB from *Sulfolobus solfataricus* was previously reported to physically interact with the MCM helicase and stimulate its unwinding activity [275]. However, in contrast to the initial report, a later study was unable to reproduce those results, and showed that SsoSSB was inhibitory to MCM catalyzed unwinding at high concentrations [276]. These contradictory studies in addition to an incomplete understanding of SsoSSB’s function and mechanism necessitates further study of these two proteins and the potential physical and functional
interaction between them. Here, we have undertaken smFRET studies to more fully characterize the properties of SsoSSB and to determine if SsoSSB interacts physically with SsoMCM at the replication fork. We provide insight into the measured binding strength of SsoSSB, and propose a novel mode of SSB binding that enables ssDNA to maintain its conformational flexibility. smFRET experiments also show that in trans interactions between SsoSSB-coated DNA strands occur on model replication fork substrates, and prebound helicase alters this interaction.

Figure 4-1: SSB Protein Schematics.
A. Cartoon schematics of various SSB proteins are shown. Red regions correspond to OB-fold containing DNA binding domains. Flexible C-terminal tails are shown in blue, and charged regions are shown in cyan. Numbers on the schematics correspond to the amino acid sequence of each protein. B and C show the structure of SsoSSB (PDB 2MNA [255]) in the apo and bound forms, respectively. Residues critical for ssDNA binding are highlighted in red [261], and their base stacking partners from ssDNA (displayed in black) are shown in blue in C.
4.3 MATERIALS AND METHODS

4.3.1 DNA Substrates

The oligos shown in Table 4.1 were purchased from IDT (Coralville, IA). Oligos containing a fluorescent or biotin label were HPLC purified by IDT.

Table 4-1: DNA Substrates

<table>
<thead>
<tr>
<th>DNA</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA23</td>
<td>5’-5GCTACTCTCGTCAGCGTACCACATAGCGAG</td>
</tr>
<tr>
<td>DNA43</td>
<td>5’-BTGCGGACGGCGACGGAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>DNA44</td>
<td>5’-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>DNA45</td>
<td>5’-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>DNA54</td>
<td>5’-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>DNA126</td>
<td>5’-BTGCGGACGGCGACGGAGCGCAGCAGAGCAGAG</td>
</tr>
</tbody>
</table>

5 - Cy5, 3 - Cy3, F - fluorescein, B - Biotin

4.3.2 Protein Purifications

4.3.2.1 SsoSSB

Single-strand binding protein from the archaeal Sulfolobus solfataricus was expressed using a pDEST14 plasmid containing the wild-type SsoSSB gene. The plasmid was transformed into Rosetta2 Bl21 DE3 cells and expressed in ZYP media using an auto-induction protocol [201]. The supernatant was then resuspended in buffer A (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 5 mM β-mercaptoethanol), and subsequently lysed, sonicated, and heat treated at 70 °C for 30 minutes. The sample was then loaded onto an SP column (GE Healthsciences) and eluted with a linear gradient of buffer B (buffer A with 1 M NaCl) using an AKTA Prime FPLC (GE Healthsciences). SSB-containing fractions were pooled, concentrated,
and exchanged back into buffer A. The sample was then loaded onto a Heparin HP column (GE Healthsciences) and eluted with a gradient of buffer B. The resulting SSB-containing sample was then concentrated and loaded onto a Superdex 200 gel filtration column (GE Healthsciences) equilibrated with buffer A and 10% glycerol. SsoSSB-containing fractions were then pooled, concentrated, and quantified using a UV-Vis ($\varepsilon = 12,660 \text{ M}^{-1} \text{ cm}^{-1}$).

**4.3.2.2 SsoMCM**

MCM helicase from the archaeal *Sulfolobus solfataricus* was expressed and purified using as previously described [96]. Briefly, the cell lysate was heat treated at 70°C, and flowed through MonoQ and heparin columns. Protein was eluted with elevated salt concentrations in both cases. The last column used in this process was gel filtration. Concentrations are reported as monomer concentration.

**4.3.2.3 E.coli SSB**

Single-strand binding protein from *Escherichia coli* was purified as previously described [277]. Briefly, 0.4% (w/v) polyethyleimine was added to the cell lysate to precipitate EcSSB. Sample is pelleted and then resuspended. 150g/L ammonium sulfate is added to the resuspended solution. The sample is then spun down, resuspended, and ~40% w/v ammonium sulfate is added. The EcSSB precipitate from the previous ammonium sulfate step is spun down and the EcSSB-containing pellet is then resuspended. The concentration was determined using a UV-Vis ($\varepsilon = 26,840 \text{ M}^{-1} \text{ cm}^{-1}$).
4.3.2.4 T4 gp32

T4 gp32 was expressed and purified essentially as previously described [278, 279]. A pET Impact vector containing the gene for T4 gp32 L301G was transformed into Rosetta 2 BL21 DE3 cells, and grown in 2xYT media at 37°C for 16.5 hours after inducing with 0.15 mM IPTG. The L301G mutation allows for more efficient intein cleavage, but maintains the properties and characteristics of the wild-type (Benkovic Lab) [278]. The cells were then spun down, resuspended in buffer A (20 mM Tris-HCl pH 8, 200 mM NaCl, 0.1 mM EDTA), lysed, and sonicated. A column was packed with chitin (New England Biolabs, Ipswich Ma), and the sample was then loaded onto the column equilibrated with buffer A. The column was washed with buffer B (buffer A with 1.25 M NaCl). The column was reequilibrated with buffer A, and 75 mM β-mercaptoethanol was subsequently added. The column was then capped and left to incubate at 4°C for 46 hours. The chitin column was then eluted, and fractions were collected. gp32-containing fractions were then pooled, concentrated, and quantified using a UV-Vis ($\varepsilon = 41,306 \text{ M}^{-1} \text{ cm}^{-1}$).

4.3.3 Atomic Force Microscopy

A 500 nM solution of SsoSSB in 10 mM Tris pH 8.0, 100 mM NaCl, 0.1 mM EDTA pH 8.0 was diluted 10 fold in AFM deposition buffer (25 mM NaOAc, 10 mM Mg(OAc)$_2$, 25 mM Hepes pH 7.5). The diluted sample was then deposited on freshly cleaved mica (SPI Supply) and washed with MilliQ water. The deposited sample was then dried using nitrogen gas. The mica surface was imaged using a MultiModeV microscope (Bruker) with an E scanner in tapping mode in air. Point probe plus non-contact/tapping mode silicon probes (PPP-NCL; Agilent) with spring
constants of ~50 Nm$^{-1}$ and resonance frequencies from 150-200 KHz were used. The mica surface was imaged using a scan size of 1 µM x 1 µM.

Volumes were calculated using the ImageSXM software as previously described in [280]. A standard curve was generated by measuring the volumes of proteins with known molecular weights: Pot1(65 kDa), Pcra monomer (86.4 kDa), Uvra monomer (105 kDa), Taq MutS (181 kDa), UvrA dimer (210 kDa), and Taq MutS tetramer (362 kDa). A linear fit to this data yields the following equation for the standard curve with an R$^2$ value of 0.9886:

$$\text{Volume (nm}^3\text{)} = 1.41 \times \text{MW(kDa)} - 7.294$$ \hspace{1cm} (1)

The standard curve is shown below, and is the same curve used by the Van Houten Lab in [280].

![Figure 4-2: Standard Curve for Calculating Molecular Weights by Atomic Force Microscopy.](image)

A standard curve of volumes measured by AFM vs. known molecular weights is shown. The known sizes of the proteins used are Pot1(65 kDa), Pcra monomer (86.4 kDa), Uvra monomer (105 kDa), Taq MutS (181 kDa), UvrA dimer (210 kDa), and Taq MutS tetramer (362 kDa). A linear fit to the data is shown as Equation 1.
4.3.4 Fluorescence Anisotropy

All fluorescence anisotropy binding assays were performed using a Fluoromax-3 fluorometer (HORIBA Scientific, Edison NJ). The excitation wavelength was set to 650 nm and the emission wavelength was set to 670 nm, and the slit widths were adjusted for optimal signal detection. Data was collected using 2-second integration times and three consecutive readings. The reported data are averages of at least three independent experiments with the standard error of the mean reported and displayed for each titration point. All titrations were performed using DNA23 labeled at the 5’ end with CY5 (Table 4.1) in 25 mM Tris Acetate (pH 7.5), 125 mM potassium acetate, and 10 mM magnesium acetate, and all data sets were background subtracted. The data was analyzed and fit using the Prism software package using a single site binding model as described below:

\[ A = \frac{B_{\text{max}} \times P}{K_d + P} \]  

(2)

where \( A \) is the change in Anisotropy, \( P \) is the protein concentration, \( B_{\text{max}} \) is the saturation point for \( A \), and \( K_d \) is the binding constant. The data for EcSSB and T4gp32 were fit to one site binding models with Hill slopes as described below:

\[ A = \frac{B_{\text{max}} \times P^n}{K_d + P^n} \]  

(3)

where \( n \) is the Hill slope that describes cooperativity.

4.3.5 Single-Molecule FRET Data Collection

DNA substrates (Table 4.1) labeled with Cy3 and Cy5 fluorophores were immobilized on a pegylated quartz slide utilizing biotin-streptavidin interactions [202]. A prism-based total
internal reflection microscope was used to collect all smFRET data [203, 204]. A 532 nm diode laser was used to excite Cy3 fluorophores, and subsequent Cy3 and Cy5 emission signals were separated by a 610 nm dichroic longpass mirror, a 580/40 nm bandpass filter, and a 660 nm longpass filter. An EM-CCD iXon camera (Andor, Belfast, UK) was used to image the signals. Data was acquired at 10 fps for ten or more regions with each region containing 50 – 250 molecules in the presence of an oxygen scavenging solution [1 mg/mL glucose oxidase, 0.4 % (w/v) D-glucose, 0.04 mg/mL catalase, and 2 mM trolox]. Binding assays utilizing the ssDNA substrate were performed in 10 mM Tris pH 8.0, 100 mM NaCl, 0.1 mM EDTA pH 8.0. Experiments utilizing the forked DNA substrates were performed in 25 mM Tris acetate pH 7.5, 125 mM potassium acetate, and 10 mM magnesium acetate as described previously [96].

Single-pair FRET signals were identified by fitting individual regions of signal intensity to a 2D Gaussian and measuring the goodness of fit. These peaks were corrected for thermal drift and local background intensity [205, 206]. The resulting signal was used to calculate the apparent FRET efficiency, \( E_{\text{app}} \), according to

\[
E_{\text{app}} = \frac{I_A}{I_A + I_D}
\]

in which \( I_A \) and \( I_D \) are the intensity of the acceptor and donor signals respectively.

4.3.6 Single-Molecule FRET Data Analysis

For the experiments binding EcSSB and T4 gp32 SSBs to ssDNA substrates, histograms were generated from the raw data. For the DNA fork binding assays and SsoSSB-ssDNA binding assays, single-molecule traces that displayed anti-correlation between the donor and acceptor fluorophore as well as single-step fluorophore bleaching were manually selected in order to perform the downstream ExPRT analysis. Selected traces from identical experimental conditions...
were stitched, and fit to ideal states using the Hidden Markov Modeling program, vbFRET [207]. The number of states fit to the data was selected for each dataset based on those states being more than $E_{\text{app}} = 0.1$ apart from one another and the variance of a state not overlapping with the neighboring state(s). The traces were then unstitched, and analyzed by the ExPRT program. Double exponential fits to a survival analysis of the dwell times were only selected if the $R^2$ value was greater than 0.970 and the increase in that value from a single exponential fit was 0.015 or greater.

4.4 RESULTS

4.4.1 Determining the Oligomeric State of Free $SsoSSB$. Several reports have investigated the oligomeric state of $SsoSSB$ as it exists free in solution as well as bound to ssDNA. Malcom White’s group has consistently observed monomeric $SsoSSB$ in the absence of ssDNA [254, 255, 259], but Haseltine and coworkers have observed mostly tetrameric species in addition to smaller amounts of monomers and dimers in the absence of ssDNA [256, 257]. The gel filtration profile from the last step of the $SsoSSB$ purification in our lab was consistent with the monomer molecular weight [281]. We sought to settle this contradiction and determine the oligomeric state of free $SsoSSB$ by measuring the volume of $SsoSSB$ units by atomic force microscopy. Atomic force microscopy is widely used to determine molecular weights, oligomeric states, and organization based on the measurement of single unit volumes within a sample [282-284]. We measured the volumes of free $SsoSSB$ sample (Figure 4.3A and B) and calculated the corresponding molecular weights of each peak using the method
described in the materials and methods and previously in [284]. A histogram of molecular weights is shown in Figure 4.3C. The majority of peaks yielded volumes corresponding to molecular weights close to the ~16 kDa molecular weight of a single monomer of SsoSSB. This is consistent with reports from Malcom White’s group as well as gel filtration data from other members of our group [281]. Similar AFM experiments to the one discussed above were performed where DNA substrates in both the presence and the absence of SsoSSB were imaged. A 90bp duplex with 10nt 5’ overhangs on each end as well as a substrate with a 18bp duplex with a 50nt 5’ overhang were imaged, but both were too small to reliably measure volumes in the absence of SsoSSB. In the presence of SsoSSB, the data were indistinguishable from the results shown below of SsoSSB alone (Figure 4.3). We were unable to test the oligomeric state of SsoSSB when bound to ssDNA because a suitable ssDNA substrate was not found. It is likely that the concentration of SsoSSB in addition to the length of the DNA substrate will have to be optimized to gain insight on SsoSSB’s binding using AFM.
Figure 4-3: AFM of Free SsoSSB.
2D and 3D images showing volumes of the SsoSSB sample are shown in A and B respectively. C shows a histogram of the calculated molecular weights based on volume measurements in the absence of ssDNA.

4.4.2 Determining the Binding Affinity of SSB Proteins.

The roles that SSB plays within *Sulfolobus solfataricus* are not clear, and there are several contradictory models currently proposed to explain SsoSSB’s function. For this reason, we first characterized the binding of SsoSSB to ssDNA while also measuring the affinity of two well
characterized SSBS, EcSSB and gp32 from T4 bacteriophage, as controls. Each SSB was titrated onto a 28-nucleotide oligo fluorescently labeled with Cy5 (DNA23). As the SSBS bind to the ssDNA, the anisotropy increases as a result of the labeled DNA molecule tumbling more slowly in solution. However, the different binding modes of each produce different binding curves. Each EcSSB subunit consists of an OB fold and a flexible C-terminal tail with acidic regions, but EcSSB acts as a functional tetramer, with four OB folds that can interact with ssDNA in two distinct modes, EcSSB_{35} and EcSSB_{65}, depending on the concentration of salt and the DNA:protein ratio [285-292]. In our binding assay, EcSSB produced a cooperative curve with a dissociation constant of 4.3 \pm 0.1 \text{nM} and a Hill coefficient 3.4 \pm 0.2 as determined by fitting the data to equation (3) (Figure 4.4 A). This is consistent with previous studies that showed EcSSB binds ssDNA with positive cooperativity equivalent to a tetramer and low nanomolar affinity [290, 293]. T4 gp32 also consists of an OB fold and flexible C-terminal tail, and gp32 monomers bind cooperatively to ssDNA [262, 294-297]. Fluorescence anisotropy showed that T4 gp32 also binds cooperatively with a Hill coefficient of 2.1 \pm 0.2 and a dissociation constant of 2.5 \pm 0.01 \text{\mu M} as determined by fitting the data to equation (3) (Figure 4.4 B). This result is similar to previously reported binding data [298]. SsoSSB binding produced a curve with no observed cooperativity and a dissociation constant of 0.20 \pm 0.03 as determined by fitting the data to equation (2) (Figure 4.4 C). Other groups have also reported similar K_d values [255, 259] but significantly lower values have also been reported [255, 259]. Overall, our fluorescence anisotropy data demonstrates that our purified SSB proteins are interacting with ssDNA in ways that are consistent with previous studies.
4.4.3 Probing the Binding Modes of SSBs to ssDNA using smFRET.

In addition to performing fluorescence anisotropy binding assays to determine the binding strength of the SSBs, we also performed single-molecule FRET binding assays. smFRET experiments have previously been used to report on the mode of interaction between SSBs and ssDNA [259, 291, 299, 300]. We proposed to undertake similar studies to study the ssDNA binding mode(s) of SsoSSB, and we performed control experiments on EcSSB and T4 gp32 to
ensure that our ssDNA substrates are reliable for measuring interactions with *SsoSSB*. Figure 4.5A shows a histogram of the data produced from *EcSSB* binding to the ssDNA70 substrate. At low protein concentrations, a high FRET population exists that corresponds to the *EcSSB*$_{65}$ state where a single tetramer of *EcSSB* wraps 65 nt around the four OB folds bringing Cy5 in close proximity to Cy3. As the concentration of *EcSSB* increases, there is a decrease in the high FRET state and simultaneous increase in a medium FRET state (~0.4) that corresponds to the *EcSSB*$_{35}$ binding mode, where only two OB folds within a given tetramer interact with the ssDNA to sequester 35 nt. The same concentration dependence of *EcSSB* binding modes has been shown using this assay previously [291]. The histograms in Figure 4.5A and B both show significant populations ~$E_{app}$=0.1. This is the ‘zero peak’ due to signal from only Cy3 and the emission from Cy3 bleeding over into the Cy5 emission channel. This is a result of visualizing the raw data from these two experiments compared to visualizing the processed data where traces are truncated at Cy5 bleaching events. Binding of gp32 was then examined on the ssDNA30 substrate. Binding of gp32 to DNA has been compared to ‘beads on a string’, where the C-terminal tail region of bound gp32 monomers act to recruit and mediate the binding of neighboring monomers, which is the basis of the observed cooperativity [295, 296, 298, 299]. This results in an extension of the ssDNA molecule as the tightly packed subunits increase the rigidity of the substrate and increase the end-to-end distance, which can be measured by FRET [295, 296, 299]. Using ssDNA30, extension of the ssDNA was observed upon binding of gp32 (Figure 4.5B). The significant decrease in FRET efficiency when bound by gp32 is consistent with the current model of gp32 ssDNA binding [296, 299].
4.4.4 Characterizing ssDNA Binding Mode of SsoSSB.

After confirming the reliability of our substrates to report on SSB binding modes using EcSSB and gp32 as controls, the ssDNA binding mode(s) of SsoSSB was then probed using smFRET. SsoSSB binding was tested by titrating increasing amounts of protein onto ssDNA substrates of various lengths. Interestingly, the distance between the dyes changed very little when SsoSSB was titrated onto the ssDNA30 substrate as shown in Figure 4.6A. Even at micromolar concentrations, the measured FRET and distribution of the signal are not significantly different from ssDNA30 by itself (Figure 4.7A). In contrast, when the length of the ssDNA region was increased to 50 nt (ssDNA50), there was a significant increase in FRET efficiency when SsoSSB was titrated onto that substrate as seen in Figure 4.6B. The increase in FRET is due to a
stabilized interaction state, as the smFRET traces showed little to no dynamic transitioning between states (Figure 4.7B). When the length was increased further to 70 nt (ssDNA70), the distribution develops a tail as increasing concentrations of SsoSSB are titrated (Figure 4.6C). This is due to dynamic departures from the predominant low FRET(≈0.2) state to a slightly higher FRET state (≈0.3) as visualized by fitting Gaussian curves to the data (Figure 4.6D). This behavior can also be seen in the representative smFRET traces shown in Figure 4.7C. We should note that at the lower apparent FRET efficiency we are measuring, large changes in distance may result in minimal changes in FRET due to the non-linear relationship between dye distance and FRET at distances much great than or much less than the R₀ distance for the dye pair. Here, SsoSSB has a ssDNA binding mode distinct from that of the highly studied T4 gp32 and EcSSB proteins. SsoSSB interacts with ssDNA in a way that only alters the end-to-end distance in a minimal length dependent manner significantly different to the other SSBs studied (Figure 4.5).
Figure 4-6: smFRET ssDNA Binding Assays.
Histograms of SsoSSB titrations onto ssDNA30, ssDNA50, and ssDNA70 are shown in A, B, and C respectively along with corresponding cartoon schematics of each substrate. D shows select data from C (ssDNA70) as data (circular markers) as well as the gaussian fits to those data (solid lines).
Figure 4-7: SsoSSB Binding smFRET Traces.
Example smFRET traces of the substrate alone and in the presence of 5000 nM SsoSSB for the ssDNA30, ssDNA50, and ssDNA70 substrates are shown in A, B, and C respectively. Dotted red lines $\sim E_{\text{app}}=0.2$ and $\sim E_{\text{app}}=0.3$ highlight the differences in FRET states between traces shown as well as transitions between these two FRET states seen within a single trace.

4.4.5 Probing the Binding Modes of SsoSSB to fork DNA using smFRET.

In order to determine if SsoSSB interacts differently with fork substrates compared to ssDNA substrates the smFRET experiments were repeated with different DNA substrates containing a 3’-tail in addition to the 5’-tail. SsoSSB was titrated onto DNA fork substrates, 30/30 and 30/50. When titrated onto the 30/30 fork, SsoSSB caused an immediate shift to a higher FRET state ($>0.4$) and increasing concentrations of SsoSSB caused peak broadening that can be seen in Figure 4.8A. The increase in FRET is likely due to an interaction between SsoSSB on each arm of the
30/30 fork bringing the dyes on the ends of each strand closer, since SsoSSB did not alter the conformation of a ssDNA of 30nt alone (Figure 4.6A). Similar trends are also noted when titrating SsoSSB onto the 30/50 fork substrate. Figure 4.8B shows that addition of SsoSSB induces a shift to higher FRET states (~0.4), and increasing concentrations of SsoSSB lead to broadening of the signal. At the highest concentrations, a new population of signal arises as a high FRET state (>0.9). This is likely due to inter-strand contacts similar to what we observed for the 30/30 fork in addition to the strand compaction we see for the ssDNA50 substrate alone (Figure 4.6B). On model fork substrates, SsoSSB brings the arms of the forks closer together in a way that increases the FRET efficiency and increases dynamics as seem by peak broadening of the histogram profiles at higher concentrations of SsoSSB.

Titrations of SsoSSB onto both the 30/30 and 30/50 forks was analyzed using the ExPRT program, which reveals the kinetic and probabilistic details of transitions between FRET states. The first two concentration points (50 and 100 nM) fit to only a single state during analysis, and therefore there were no transitions to be analyzed. This is consistent with the general trend seen in the histogram (Figure 4.8A), where the lowest concentration points have the tightest distribution, and these distributions broaden as the concentration increases. Figure 4.9A-E show the ExPRT plots for 250 – 5000 nM SsoSSB on the 30/30 fork. Each of these datasets fit to either
three or four states, and they all exhibit a predominant reversible transition from $E_{\text{app}=0.25}$ to $E_{\text{app}=0.45}$. Binding of $SsoSSB$ onto ssDNA30 does not alter the end-to-end distance or flexibility of the ssDNA (Figure 4.6A). Here, the increase in FRET upon addition of $SsoSSB$ is likely due to interactions between SSBs bound on each arm of the fork. These interactions would most likely occur near ss-dsDNA junction where the diffusion of the opposing fork arms is most limited. This would provide an opportunity for the $SsoSSB$-mediated interaction of fork arms to occur. However, these interactions are likely weak and transient, as the ExPRT plots display large probabilities for this reversible transition and relatively short dwell times in each state Figure 4.9A-E. ExPRT analysis of the titrations of $SsoSSB$ onto the 30/50 fork are similar and indicate that the low concentration datasets (10-500nM) fit to two states, and exhibited few transitions (Figure 4.10A-E). At higher concentrations (750-5000nM), the datasets fit to 3 states, but the percentage of analyzed traces that exhibited transitions was still small.
Figure 4-9: ExPRT Analysis of SsoSSB Titration onto 30/30 Fork.
A-E show the ExPRT plots corresponding to the titration of SsoSSB onto the 30/30 fork along with the legend for probability and dwell time. The histogram for this data is displayed in Figure 4.8A.
Figure 4-10: ExPRT Plots of SsoSSB Titration onto the 30/50 Fork.
A-I show the ExPRT plots for the titration of SsoSSB onto the 30/50 fork. The legend for probability and dwell time is shown at the bottom. A histogram of this data is displayed in Figure 4.8B.
4.4.6 Identifying a DNA-Helicase-SSB Tertiary Complex

In order to determine if SsoSSB interacts with the archaeal helicase, MCM, or the MCM-DNA fork complex, SsoSSB was titrated onto DNA fork substrates, 30/30 and 30/50, in the presence of pre-bound SsoMCM. As shown previously, binding SsoMCM to DNA fork substrates brings the FRET-labeled fork arms together to give a high FRET signal [96, 189]. This is consistent with that observed in Figure 4.11A and B going from low FRET due to DNA alone (black) to high FRET due to MCM loading onto the 3’ strand and wrapping the 5’ strand on the exterior surface (brown). For the 30/30 fork, there is a decrease in the high FRET population upon adding SsoSSB (Figure 4.11A). There is also a simultaneous appearance of a medium FRET state that is consistent with the FRET signal observed on 30/30 when SsoSSB was titrated in the absence of SsoMCM. However, a significant population of high FRET that corresponds to helicase signal is maintained even at micromolar concentrations of SsoSSB. This suggests that some fraction of the DNA complexes being monitored still have helicase bound while others have only SsoSSB from the helicase being competed off the fork. When the 30/50 fork was tested, the FRET population corresponding to the MCM helicase alone decreases as SsoSSB is added (Figure 4.11B), and a medium FRET population consistent with the signal produced by SsoSSB alone on 30/50 increase with increasing concentration. In contrast to the results for the 30/30 fork shown in Figure 4.11A, the histogram corresponding to the 30/50 fork (Figure 4.11B) shows the high FRET signal corresponding to the helicase binding is completely lost at concentrations above 100 nM SsoSSB. This suggests that SsoSSB interacts with helicase-DNA complexes in a substrate specific manner possibly limited by distinct contacts of the 5’-tail with the exterior surface of SsoMCM. In order to reveal further differences and insights into potential interactions, the titrations of SsoSSB onto SsoMCM-DNA fork was analyzed by ExPRT analysis.
Figure 4-11: Titration of SsoSSB onto Helicase - DNA Complex.
smFRET histograms of titrations of SsoSSB onto the 30/30 (B) and 30/50 forks prebound with SsoMCM. Low, medium, and high FRET states are highlighted by shades of yellow, blue, and red, respectively.

ExPRT analysis was also performed for the titrations of SsoSSB onto SsoMCM-bound fork substrates. Loading SsoMCM onto the 30/30 fork produces a predominantly high FRET population with a ExPRT plot (Figure 4.12A) displaying dynamics distinct from those produced by SsoSSB but consistent with those shown previously for SsoMCM on the same forked substrate [115]. A titration of SsoSSB onto the 30/30-MCM complex from 50-5000 nM (Figure 4.12B-H) shows that the dynamics of the helicase-DNA interaction are disrupted upon introduction of SsoSSB. Generally, throughout this titration there is an increased number of FRET states being sampled, an increase in transitions between low and medium FRET states, and a reduction in transitions between high FRET states consistent with helicase alone. However, even at micromolar concentrations of SsoSSB, the patterns of dynamics never mirror those seen in the titration of SsoSSB onto 30/30 in the absence of MCM, which suggest that either SsoMCM is still bound to a significant fraction of the DNA forks or having the helicase pre-bound to 30/30 allowed SsoSSB to interact with and distort the DNA’s conformation in a distinct way.

When SsoMCM was loaded onto the 30/50 fork, there was a distinct pattern of FRET state transitions on the resulting ExPRT plot (Figure 4.13A). Adding increasing amounts of SsoSSB onto the 30/50-MCM complex does not change transitions between FRET states, but the
probability of these transitions occurring decreases with increasing SsoSSB concentration (Figure 4.13B-D). Once 250 nM SsoSSB is reached, the patterns displayed on the ExPRT plots now mirror those patterns obtained from the titrations without helicase (Figure 4.13E-H). This is consistent with the general trends seen in the histogram (Figure 4.11A and B) as discussed above, and is suggestive of higher concentrations of SsoSSB competing SsoMCM off of the 30/50 fork substrate at concentrations close to the measured Kd of SsoSSB for ssDNA (Figure 4.4 C). The presence of a longer 5’-tail may act as an assembly point for nucleation of SsoSSB binding to ultimately displace SsoMCM from the fork.
Figure 4-12: ExPRT Analysis of SsoSSB onto the SsoMCM - 30/30 Complex.
A-H show the ExPRT plots for the titration points of SsoSSB onto the SsoMCM – 30/30 fork complex. A legend for the ExPRT plots is shown at the bottom.
Figure 4-13: ExPRT Analysis of SsoSSB Titrated onto the SsoMCM - 30/50 Complex.
A-H show the ExPRT plots for the titration of SsoSSB onto the SsoMCM – 30/50 fork complex. A legend for the ExPRT plots is shown at the bottom.
4.5 DISCUSSION

4.5.1 Atomic Force Microscopy Study of SsoSSB Oligmeric State.

*E.coli* SSB acts as a tetramer that wraps ssDNA around itself [291, 292], and T4 gp32 is monomeric and assembles on ssDNA as ‘beads on a string’ [294-296, 298, 299]. The oligomeric state of *SsoSSB* has been reported as both monomeric [254, 255] and as a mixture of mostly tetrameric species along with some dimers and monomers [256, 257]. To determine the oligomeric state(s) of *SsoSSB* as purified by our lab, atomic force microscopy was used to measure the volumes of *SsoSSB* units as they exist in the absence of ssDNA. According to the calculated the molecular weights corresponding to each volume measured the majority of the peaks scanned by the AFM agreed with monomeric (~16 kDa) *SsoSSB* (Figure 4.3). No significant higher molecular weight species were observed that would correspond to dimeric or tetrameric assemblies. This also agrees well with gel filtration data produced by other members of the lab, where they showed that a sample of *SsoSSB* in the absence of DNA eluted as one peak that corresponded to *SsoSSB* monomer [281].

We attempted several AFM experiments where DNA was pre-bound to *SsoSSB*. However, due to *SsoSSB*’s small size and difficulty collecting reliable control data corresponding to DNA alone, we were unable to produce reliable results here. AFM has been used in the past to measure SSB proteins and their interaction with DNA substrates [301-303]. One study imaged both *EcSSB*, gp32, and RPA from yeast [303]. The authors were able to show changes in the apparent contour length depended on whether conditions induced the *EcSSB*$_{35}$ or *EcSSB*$_{65}$ binding mode. They also visualized gp32 and RPA filaments bound to M13 ssDNA. However, they noted the general difficulty in depositing ssDNA-SSB complexes onto mica.
surfaces. The authors propose using spermidine to enable successful deposition onto the mica surface and demonstrated the effectiveness of including this reagent. Similar studies have also been carried out using electron microscopy [304-306]. In all these studies, the DNA substrates used were hundreds to thousands of bases in length, while we attempted to use oligos that were tens of bases long. It is possible that binding SsoSSB onto a longer substrate and altering the conditions to include spermidine may provide the resolution and reliability needed to image SsoSSB bound to DNA.

4.5.2 Determining the Basis of Varying SsoSSB ssDNA Binding Affinities.

We further characterized the DNA binding properties of SSB from Sulfolobus solfataricus and determined that it interacts uniquely with a DNA fork-helicase complex. Fluorescence anisotropy quantified the binding of SsoSSB to DNA and smFRET binding assays revealed the binding mode(s) of SsoSSB compared to SSB proteins from *E. coli* (EcSSB) and T4 bacteriophage (gp32) as controls. The ssDNA binding affinity of EcSSB and gp32 were consistent with previously published results [289, 290, 293, 295, 298]. SsoSSB had a binding affinity (K<sub>d</sub>) of 0.20 +/- 0.03 nM and observed no cooperativity (Figure 4.4A). Several groups have quantified the binding affinity of SsoSSB for ssDNA, and although there is a general consensus on the nucleotide footprint of the protein (5 nucleotides) [254, 255, 257], there is inconsistency in the binding constants that have been reported. Several studies have reported very low nanomolar K<sub>d</sub> values (below 25 nM) [255, 259, 307], while other studies have reported significantly higher values more in line with our measurement of 0.20 +/- 0.03 nM (Figure 4.4A) [255, 259, 261]. This discrepancy has existed for several years, and has not been directly addressed by any group. Throughout the various studies that have reported binding constants, a
trend exists. Figure 4.14 shows that based on the previous studies conducted and values reported in the literature, binding affinity may be length dependent. A minimum dissociation constant value exists at a ssDNA length of 10-15 nucleotides. It should also be noted that a similar trend exists when considering positive binding cooperativity. Those substrates with the lowest measured dissociation constants also produced strong positive cooperativity [259, 261]. This cooperativity may be due to SSB-SSB interactions between bound and unbound monomers that result in occupied neighboring binding sites, or a bound SSB monomer may alter the adjacent binding site on the ssDNA to enable SSB binding more readily [259]. However, significant cooperativity seems to be confined to substrates between 10-15 nucleotides, where two monomers can easily bind and the only available binding sites are directly adjacent to one another. The site size of \( Sso \)SSB is 5 nucleotides, and substrates shorter than 10 nucleotides cannot accommodate two monomers binding simultaneously. On longer substrates, \( Sso \)SSB may bind DNA distributively [254, 255, 257], and the opportunity for a cooperative interaction would be less probable. The length dependence of \( Sso \)SSB’s binding affinity is consistent with the current literature, but will need to be directly tested.
Figure 4-14: Binding Constant vs. DNA Length.
Plot showing the measured binding affinity vs. substrate length used in the assay. Black markers correspond to data from [259], the blue marker corresponds to data from [255], the red marker corresponds to data from [307], the green marker corresponds to data from [261], and the cyan marker corresponds to data from our lab.

4.5.3 The Novel Binding Mode of SsoSSB.

After quantifying the binding affinity, smFRET was utilized to reveal significant distortions in the ssDNA upon SsoSSB binding to different length substrates. These assays have been performed for other SSBs, and we have successfully reproduced the results (Figure 4.5A and B) reported in previous studies of T4 gp32 and EcSSB binding modes [291, 299]. These binding modes are illustrated as cartoon models in Figure 4.15. Even at micromolar concentrations of SsoSSB, ssDNA30 maintains its flexibility as the end-to-end distance as measure by FRET does not change significantly (Figure 4.6A). On 30 nucleotides of ssDNA, a maximum of six monomers could bind. Although these assays do not measure SsoSSB binding directly, the concentration was increased well beyond the range of binding saturation of a 28 nucleotide oligo in our fluorescence anisotropy assay, and it is likely that ssDNA30 is highly saturated with SSB monomers. This suggests that SsoSSB’s binding mode on ssDNA is distinct from well-known
SSB examples like EcSSB and gp32 including much less binding cooperativity or reorganization of a single-strand.

When considering 30 nucleotides of ssDNA, an estimate of the contour length of the oligo can be determined using a similar approach to the one used by Ha and coworkers in [308]. They measured the distance between adjacent phosphorus atoms as observed in five cocrystal structures to produce an average unit length of each base, which was calculated to be 6.3 (+/- 0.8) Å [308]. The five structures used were HCV NS3 helicase bound to poly dU (PDB 1A1V) [48], Rep helicase bound to poly dT (PDB 1UAA) [309], EcSSB bound to poly dT (PDB 1EYG) [285], Fbp bound to a mixed sequence oligo (PDB 1J4W) [310], and human RPA bound to poly dC (PDB 1JMC) [311]. This averaged unit length value is consistent with previous reports [312, 313] as well as the average distance measured from the structure of SsoSSB bound to ssDNA [255], for which the average unit length is 6.7 (+/- 0.8) Å. Using this approximation, a 30 nucleotide oligo would have a contour length of 189 Å. However, according to our FRET measurement the end-to-end distance of the ssDNA region of ssDNA30 is approximated to be 54.5 Å, using Forster distance of 53 Å [314]. This is consistent with the known effect of various ions and buffer conditions on the flexibility of single-stranded nucleic acids, where the polymer relaxes and significantly decreases the end-to-end distance from the contour length [308]. However, it is surprising that the E_app does not change upon the addition of SsoSSB. This suggests that binding does not intrinsically affect the flexibility of the ssDNA oligo. To determine if this mode of interaction is length-independent, we tested substrates of 50 and 70 nucleotide lengths.
Figure 4.6B shows that on the ssDNA50 substrate, addition of SsoSSB causes a shift to a higher FRET state, and the resulting complex is stable over tens of seconds as measured in our assay (Figure 4.7B). However, the measured $E_{\text{app}}$ is consistent with an end-to-end distance of 56.7 Å, while the contour length is approximated to be 315 Å. For the ssDNA70 substrate, there is little to no change in the $E_{\text{app}}$ except at the highest concentrations of SsoSSB tested where a short-lived transition to $E_{\text{app}} \approx 0.4$ is observed (Figure 4.7B). The results again suggest that the binding of SsoSSB to ssDNA has a minimal effect on the oligo’s flexibility. Together, these datasets are consistent with a model where ssDNA is highly populated with bound SsoSSB, but the overall conformation and dynamics are not significantly altered. However, we also note that because of a limiting time resolution, dynamics that occur faster than 100 milliseconds would not be captured by our instrument. It is possible that SsoSSB alters the dynamics of ssDNA, but if
the dynamics of bound ssDNA are faster than our acquisition rate, they will average out to a broad state in our data.

Although dimers and tetramers of SsoSSB have been reported [256, 257], our data as well as others indicate primarily a monomeric species in solution in the absence of ssDNA [254, 255, 259]. However, it is difficult to imagine a mode of interaction that allows the end-to-end distance of an oligo to remain between 50-60 Å despite the contour length being hundreds of angstroms long without the SSB monomers interacting with more than just the adjacent monomers along the oligo as for T4 gp32. However, molecular simulations of single-stranded DNA have shown that the DNA can form a pseudo-globule or partial hairpin structure [315-318]. One possibility is that the subunits of SSB bind without disrupting an already existing partial hairpin. In this case, the nearest-neighbor contacts would likely be the only significant protein-protein interactions. It is also possible that monomers interact with non-adjacent subunits to enable a highly condensed globule conformation of the ssDNA. Several crosslinking studies have reported that higher order oligomers can form upon the addition of ssDNA [254, 255], but these studies did not determine which regions of SsoSSB are responsible for the interaction. Our studies do not reveal the structure and organization of SsoSSB bound to the oligos, however smFRET experiments of SsoSSB on fork substrates are consistent with a trans interaction between SsoSSBs on each strand.

4.5.4 SsoSSB Fork Binding.

Model DNA fork substrates that have 30 nucleotides on the 3’ loading arm and variable length 5’ excluded strand (30 and 50 nt) were tested. When SsoSSB was titrated onto the 30/30 substrate, there is an increase in $E_{app}$, and the distributions generally broaden as the concentration increases.
(Figure 4.8A). This is suggestive of an SsoSSB-mediated interaction between the two arms of the 30/30 fork, possibly occurring at the base of the fork arms where the probability of inter-strand contact would be high due to limited diffusion. A similar increase in FRET is observed when SsoSSB is titrated onto the 30/50 substrate (Figure 4.8B). Although at high concentrations, a high FRET state exists for the 30/50 substrate that is not present in the 30/30 data. On the ssDNA 50 substrate alone, SsoSSB increased the E_{app} upon binding. It is probable that SsoSSB decreases the 5’ terminus-to-duplex length in addition to mediating an interaction between the fork arms that results in the high FRET states observed at the highest concentrations for the 30/50 substrate.

4.5.5 Possible DNA-Helicase-SSB Tertiary Complexes.

There are conflicting reports concerning the effect that SsoSSB has on SsoMCM unwinding [275, 276]. smFRET was utilized to determine whether SsoSSB can interact with a preformed SsoMCM-DNA complex (Figure 4.11). Titrations of SsoSSB onto both substrates in the presence of helicase systematically decreased the high FRET state corresponding to the signal produced by the helicase alone on both forks. Although this signal completely disappears after 250 nM SsoSSB is titrated onto the 30/50 substrate, the signal corresponding to the helicase persists to some degree throughout the entire range of the titration onto the 30/30 fork. This suggests that SsoSSB interacts differently with the 30/30-helicase complex than it does with the 30/50-helicase complex. In order to interrogate these differences, these datasets to were subjected to ExPRT analysis.

The ExPRT plots display each unique transition within a dataset along with its corresponding probability and dwell time(s). On the 30/30 and 30/50 substrates, the helicase
binding to the fork generates unique patterns on the ExPRT plots shown in Figure 4.12A and Figure 4.13A respectively. For the titration of \textit{Sso}SSB onto 30/50 in the presence of \textit{Sso}MCM, the patterns on the ExPRT plots (Figure 4.13) mirror those from the titration onto 30/50 in the absence of \textit{Sso}MCM above 100 nM \textit{Sso}SSB, which is consistent with the dissociation of the helicase. In contrast, the ExPRT plots corresponding to titrations onto the 30/30 substrate (Figure 4.12) show that the ExPRT patterns produced by the titration in the presence of \textit{Sso}MCM never return to mirror those generated from the titration in the absence of \textit{Sso}MCM, even at the highest concentrations of \textit{Sso}SSB. This suggests that the helicase is still bound on the 30/30 substrate despite the presence of high concentrations of \textit{Sso}SSB. Alternatively, the prebound helicase enables \textit{Sso}SSB to interact with 30/30 in a distinct manner. This data suggests that a tertiary complex may be substrate-specific in addition to concentration dependent. This may have implications for \textit{Sso}SSB’s potential role in enhancing or inhibiting \textit{Sso}MCM helicase unwinding should too much ssDNA become available. The initial study that observed a \textit{Sso}SSB-mediated enhancement of \textit{Sso}MCM unwinding utilized substrates with either a 0 or 5 nucleotide overhang on the 5’ non-loading arm of the fork [275]. The subsequent study that observed no \textit{Sso}SSB-mediated helicase stimulation utilized a model fork with 44 nucleotide arms [276], more similar to our 30/50 substrate, where \textit{Sso}SSB can compete \textit{Sso}MCM off the fork substrate at moderate concentrations. It may also be the case that additional DNA replication proteins are required to mediate any functional interaction within a larger complex. Additional studies will be required to determine if a tertiary complex is formed, and if so, whether or not a functional interaction is formed with the helicase.
4.5.6 Fork Binding Assays in the Context of Helicase Function.

Other members of our group have performed a variety of SsoSSB binding assays as well as assays to determine the effect of SsoSSB on SsoMCM helicase activity [281]. Similar to the titration discussed above (Figure 4.11B), the lab also analyzed SsoSSB binding to the helicase – 30/50 DNA fork complex by performing a titration and visualizing each binding reaction using electrophoretic mobility shift assays. As the concentration of SsoSSB increased, SsoMCM was generally competed off the fork by SsoSSB. However, the EMSAs do show that a small fraction of DNA was bound by SsoSSB and SsoMCM simultaneously as determined by the colocalization of their fluorescent signals on the gel. This is largely consistent with the results shown in Figure 4.11A and B, where SsoSSB competes the helicase off the fork with increasing concentrations, but a fraction of DNA forks may bind SsoMCM and SsoSSB to form a ternary complex depending on the fork structure.

Also, helicase unwinding and ATPase assays were performed as increasing concentrations of SsoSSB were titrated onto the 30/30-SsoMCM complex. Decreasing unwinding activity is observed with increasing SsoSSB concentrations, consistent with previous results [276]. The ATPase activity of SsoMCM bound to the 30/30 fork was also tested in the presence of increasing SsoSSB concentrations, and there is little to no variation in the elevated ATPase rates due to DNA binding until ~5 μM of SsoSSB has been reached. These two results are consistent with the smFRET data shown in Figure 4.11A, where a significant fraction of helicase-30/30 complexes are still intact even at low micromolar concentrations of SsoSSB. Collectively, these results suggest a model where SsoMCM can still interact with the 30/30 fork DNA in a way that allows for elevated ATPase rates to be maintained in the presence of high
concentrations of SsoSSB, but SsoSSB interacts with the complex in a way that inhibits helicase-mediated fork unwinding.

4.6 CONCLUSION

SSB proteins are crucial elements in coordinating DNA replication and repair processes across all domains of life. Here, we have further characterized the binding of the SSB protein from the archaeal *Sulfolobus solfataricus* to the highly studied SSB proteins from *E.coli* and T4 bacteriophage. Correlating data from this Chapter with work from others show that the binding affinity of SsoSSB is highly dependent on the length of the oligo. Using smFRET, SsoSSB interacts with ssDNA through a unique binding mode that allows ssDNA to maintain its conformational flexibility. This binding mode may be dependent on the length of the ssDNA as well as well as trans interactions from non-neighbor bound molecules. SsoSSB interacts with helicase-DNA fork complexes in a length-specific manner. SsoSSB can compete SsoMCM off substrates with longer strands (i.e. 30/50 fork) but on shorter forks, either the helicase remained bound to or it enabled SsoSSB to interact with the 30 nucleotides arms in a distinct manner. In all, the characterization of SsoSSB DNA binding provided here reveals new properties of representative member of an emerging class of SSB proteins.
4.7 CONTRIBUTIONS

Sean Carney purified T4 gp32, performed the fluorescence anisotropy assays, and all smFRET experiments and analyses. Brian Graham purified SsoSSB and EcSSB. The AFM data was collected and processed by Muwen Kong in the lab of Ben Van Houten, University of Pittsburgh.

4.8 ACKNOWLEDGEMENTS

We thank James Keck for providing the EcSSB plasmid. We also thank the Stephen Benkovic for providing the T4 gp32 plasmid. We thank Brian Graham and Grant Schauer for assistance with experiments and helpful discussions. We thank Ben Van Houten and Muwen Kong for collaborating to produce the AFM results.
5.0 CONCLUSIONS AND PERSPECTIVES

5.1 SUMMARY

DNA replication is a necessary process across all domains of life. A wide array of protein machinery mediates replication. However, the dynamic assembly, interactions, and activities coordinated and catalyzed by each component of the replisome are not fully understood. It is important to understand DNA replication more fully due to its essential role in all living organisms, its exploitation for cancer diagnostics [98-100] and therapeutics [319], antivirals [106, 107], and biotechnology applications [109, 110]. The work presented in this thesis is focuses on the DNA replicative helicase, particularly, interactions between the excluded strand and the helicase exterior and the effects thesis interactions have on helicase activity.

DNA replication requires the separation of the parental dsDNA before polymerases can synthesize daughter strands against single stranded templates. Hexameric helicases catalyze this dsDNA unwinding at the head of the replication fork, and many of the replisome components interact with and assemble around the helicase after loading and initiation [10, 16, 25, 26, 177]. Current structural, biochemical, and biophysical evidence point towards a steric exclusion model of hexameric helicase unwinding [82, 91, 92]. In this model, the ring-shaped hexamer encircles a single strand of DNA that translocates upon using ATP hydrolysis. The opposing strand is excluded from the central channel during unwinding. However, until recently the excluded strand
was largely ignored, but work from our lab [96, 115, 168, 189, 320] and this thesis suggest that
the excluded strand plays an active role in helicase unwinding. We have proposed the steric
exclusion and wrapping (SEW) model of hexameric helicase unwinding, which expands the
current steric exclusion model of unwinding to include contributions of the excluded strand
physically interacting with the outer surface of the helicase.

5.2 STERIC EXCLUSION AND WRAPPING MODEL FOR OPPOSITE POLARITY
HEXAMERIC HELICASES

The SEW model of unwinding was first proposed for the archaeal SsoMCM helicase. Our group
used smFRET, footprinting, and hydrogen-dueterium exchange couple with mass spectrometry
to show that the SsoMCM encircles a single strand of DNA and the excluded strand physically
interacts with the outer surface of the hexamer [96, 97, 189]. This excluded strand interaction
could be disrupted by mutating surface exposed positively charged residues, and these mutants
were severely deficient in unwinding activity. We propose that the excluded strand interaction
acts to prevent slippage and stabilizes forward unwinding. This interaction might be thought of
as a ratchet-like mechanism, as illustrated in Figure 1.7A.

We then sought to determine if the SEW mechanism is conserved outside of superfamily
6 archaeal MCM helicases, and we used smFRET to detect and characterize excluded strand
interactions for the superfamily 4 E.coli DnaB (EcDnaB) and human mitochondrial Twinkle
helicases, which both have a global ring-like hexameric structure but do not share sequence
homology or translocation polarity with superfamily 6 helicases [14]. smFRET studies showed
that EcDnaB interacts with the excluded strand in a manner very similar to the interaction
observed for SsoMCM [115]. Mutants were generated to disrupt the electrostatic contacts between the helicase exterior and sugar phosphate backbone of DNA. However, unlike SsoMCM, these mutants unwound dsDNA unwound faster than the wild type EcDnaB. This was a surprising result, but when thinking about the excluded strand interaction in the context of the replisome, the results are consistent with the known interactions and events at the replication fork.

5.3 EXCLUDED STRAND INTERACTIONS WITHIN THE CONTEXT OF REPLISOMES

SsoMCM translocates 3’-5’ on the leading strand, and the lagging strand is excluded from the central channel. Based on the in vitro observations that the excluded strand promotes unwinding, it is likely that SsoMCM interacts with the excluded strand during active replication. Also, the archaeal MCM helicase interacts with the primase [174]. The excluded strand interaction may act as a hand-off mechanism to feed the lagging strand to the primase for priming during lagging strand synthesis.

The bacterial DnaB helicase translocates on DNA in the opposite direction (5’-3’). The lagging strand is encircled by the helicase in this case, and the leading strand would be excluded from the central channel during unwinding. The mutants generated to disrupt electrostatic contacts on the helicase exterior unwound dsDNA more efficiently than the wild type EcDnaB [115]. This is consistent with reports that the addition of the leading strand polymerase stimulates the bacterial helicase’s unwinding speed and processivity [186, 187]. This is also true for bacteriophage replication systems like T4 and T7 [180, 181, 183, 184]. When the helicase
and leading strand polymerase are progressing together in the context of an active replication fork, the polymerase would sequester the excluded strand away from the helicase exterior. This would prevent the excluded strand interaction from taking place, which we suggest normally slows down unwinding by acting like an electrostatic brake. This is illustrated in Figure 1.7B. This idea is further supported by a report where T7 gp4 helicase that show removing the charge from the excluded strand by using morpholinos enhances unwinding [122], and another single-molecule study showed the helicase unwinding rate decreased after the leading strand polymerase encountered a lesion and stalled on the excluded strand [182]. Within bacterial and phage DNA replication systems, the excluded strand interaction may act as an electrostatic brake when leading strand replication becomes uncoupled in order to slow down and minimize the production of vulnerable ssDNA by the helicase. A similar ensemble biochemical study demonstrated that placing a lesion on the leading strand cause the replication fork to slow down or stall completely [185]. It likely that the excluded strand interaction plays an important role in helicase function in both archaeal and bacterial systems, but the role it plays depends on the local replisome architecture.

We also showed that the human mitochondrial Twinkle helicase interacts with the excluded strand when loaded onto a model replication fork substrate [168]. However, this interaction was measured to be much more dynamic than those measured for EcDnaB and SsoMCM [96, 115]. Recent reports may provide insight into the basis of the significantly different excluded strand interaction. Twinkle is able to catalyze strand annealing, D-loop migration, and strand exchange [168, 228, 321]. This suggests that Twinkle may have a more intimate role in DNA damage recognition and repair when compared to other replicative helicases, and that the excluded strand interaction plays a role in these processes. This suggests
that the excluded strand interaction has roles in processes other than dsDNA unwinding within an active and progressing replication fork. It is possible that interactions with the excluded strand are important for helicase activities and processes other than those listed above.

## 5.4 EXCLUDED STRAND INTERACTIONS TO OVERCOME OBSTACLES

Several helicases have been shown to transiently break their ring-like structure and open to overcome protein obstacles, DNA damage, or other obstructions to helicase translocation [91, 92, 122]. During this opening, it is likely that the helicase disengages from the ssDNA bound within the central channel, but it must also not dissociate from the substrate completely. An interaction with the excluded strand might keep the helicase localized during transient opening. The hexameric ring-structure must also be opened during helicase loading, where loading proteins assist the helicase in encircling a single strand of DNA within the central channel [10, 14, 16, 197, 198, 322]. Like transient opening during the bypass of an obstacle on DNA, the helicase is opened during loading and must be localized to and engage the DNA substrate. It is not clear what contacts are made between the helicase and DNA prior to engaging the encircled strand and closing the hexameric ring. It is plausible that the excluded strand may interact with the helicase surface during this process to prevent dissociation before loading completion. It will be interesting to determine if an excluded strand interaction has an effect on helicase loading.

In general, one of the next steps that must be taken is testing the effects of the excluded strand interaction in the context of the replisome. As discussed above other components at the replication fork are expected to alter the helicase’s interaction with the excluded strand and alter the activity of the helicase and the entire replication fork. We have shown that the SSB protein
from *Sulfolobus solfataricus* alters the helicase-excluded strand interaction. This is just one component among many that are expected to change the interaction in the context of the assembled replisome. It will be interesting to determine the effect that the excluded strand interaction has on DNA replication *in vivo* as well as the potential exploitation of this effect for therapeutics or biotechnology. In addition to helicases, it is possible that other classes of DNA translocases utilize excluded or non-translocating strand interactions as part of their mechanisms.

### 5.5 EXCLUDED STRAND INTERACTIONS OF POLYMERASES

Polymerases also translocate on ssDNA in order to synthesize a new strand complimentary to the existing template. Structural analyses of polymerases like phi29 bound to DNA substrates have revealed that they use steric mechanisms analogous to the ring-like structures found within helicases and processivity clamps to enhance strand displacement and translocation processivity [323-326]. Despite the ‘tunnels’ and obvious paths that the template strand and exiting duplex occupy, there is little information on the fate of the displaced strand. phi29 is widely utilized in sequencing applications due to its remarkable processivity and strand displacement activity [327-330]. A structural study of phi6 RNA polymerase revealed a significant patch of positive charge on the outer surface of the polymerase adjacent to the downstream tunnel where the template strand is fed into the active site [331]. The authors suggest that the non-template strand is separated and stabilized by electrostatic interactions in this exterior region during translocation. However, to our knowledge there has been no direct characterization of the effects of an excluded or non-template strand interaction in polymerases. It will be interesting to see if properties like translocation rate and processivity are affected by altering such an interaction.
Like helicases, the excluded strand interaction may be of particular interest when designing recombinant proteins for use in biotechnology applications. Recent evidence, including the work presented in this thesis, is beginning to reveal a previously unrecognized role of the excluded strand in translocation and unwinding mechanisms of DNA interacting proteins. Moving forward, we expect this trend will continue, and we anticipate more detailed characterizations of excluded strand interactions and the effects they have on \textit{in vivo} replication and repair processes.
APPENDIX A

HEXAMERIC HELICASES COMPRESS THE ENCIRCLED STRAND

A.1 INTRODUCTION

Helicases are ubiquitous throughout the domains of life, and by unwinding dsDNA provide the single strand templates required by polymerases during DNA replication. Replicative helicases are arranged in hexameric rings, and are thought to unwind dsDNA by encircling one strand of DNA while excluding the other strand from the central channel, termed the steric exclusion model of unwinding [14, 79, 80, 82, 91, 92]. However, several aspects of the unwinding mechanism remain poorly understood. The role of the excluded strand has been investigated [96, 115, 168, 189] and has been discussed in other sections of this thesis. In addition to the excluded strand, interactions between the helicase and the encircled strand play a pivotal role in unwinding, but the dynamics of such an interaction are largely uncharacterized.

Several cocrystal structures show single strand nucleic acid bound within the central channel of the hexameric helicase [52, 68, 70, 332]. Several of these structures reveal a decrease in the rise per base in the bound nucleic acid when compared to B-form DNA [70]. These structures are shown in Figure A.1.
Figure A-1: Cocrystals of hexameric helicases and nucleic acid substrates. Cocrystal structures of Rho[68], E1[52], and DnaB[70] nucleic acid translocases are shown with two of the six hexamer subunits removed and the other subunits colored to distinguish subunit interfaces. Nucleic acid substrates bound within the central channels are shown in grey.

These structures also reveal that nucleic acid binding loops are positioned within the central channels to interact with the substrate in a way that allows for directional translocation [52, 68, 70]. However, the relationship between translocation and the decrease in rise per base and conformation of the bound nucleic acid is unknown. Details of this interaction come from cocrystals, which provide no information on the dynamics. Here, we performed single-molecule FRET experiments in an effort to further characterize the interaction between hexameric helicases and the encircled strand.

A.2 MATERIALS AND METHODS

A.2.1 DNA Oligonucleotides

The oligonucleotides shown in Table A.1 were purchased from IDT (Coralville, IA). Oligos containing a fluorescent or biotin label were HPLC purified by IDT.
### A.2.2 Protein Preparation

Wild-type EcDnaB was expressed in Rosetta 2 cells (EMD Millipore, Billerica, MA) using autoinduction [201]. Cells were pelleted, resuspended in DnaB lysis buffer [10 % sucrose, 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM dithiothreitol (DTT)], and lysed using lysozyme and sonication. Ammonium sulfate was added to the resulting supernatant at 0.2 g/mL, pelleted, and then resuspended in DnaB buffer A [10 % glycerol, 0.1 mM EDTA, 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM DTT]. The supernatant was purified using an AKTA Prime FPLC equipped with a HiTrap MonoQ column (GE Healthsciences, Sunnyvale, CA) and eluted with a stepwise gradient of DnaB buffer A with 500 mM NaCl followed by a similar procedure using a HiTrap Heparin column (GE Healthsciences, Sunnyvale, CA). The purified fractions were combined and applied to a Superdex S-200 26/60 gel filtration column (GE Healthsciences, Sunnyvale, CA) with Buffer C [50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM DTT] to isolate the hexamer. An extinction coefficient (185,000 cm$^{-1}$ M$^{-1}$) was used to quantify the fractions containing purified hexameric DnaB [74]. All concentrations for DnaB are indicated as hexamer throughout.
MCM helicase from the archaeal *Sulfolobus solfataricus* was expressed and purified using as previously described [96]. Briefly, the cell lysate was heat treated at 70°C, and flowed through MonoQ and heparin columns. Protein was eluted with elevated salt concentrations in both cases. The last column used in this process was gel filtration. Concentrations are reported as monomer concentration.

### A.2.3 Single-molecule FRET

DNA substrates (Table A.2) labeled with Cy3 and Cy5 fluorophores were immobilized on a pegylated quartz slide utilizing biotin-streptavidin interactions [202]. A prism-based total internal reflection microscope was used to collect all smFRET data [203, 204]. A 532 nm diode laser was used to excite Cy3 fluorophores, and subsequent Cy3 and Cy5 emission signals were separated by a 610 nm dichroic longpass mirror, a 580/40 nm bandpass filter, and a 660 nm longpass filter. An EM-CCD iXon camera (Andor, Belfast, UK) was used to image the signals. Data was acquired at 10 fps for five or more regions with each region containing 50 – 250 molecules in the presence of an oxygen scavenging solution [1 mg/mL glucose oxidase, 0.4 % (w/v) D-glucose, 0.04 mg/mL catalase, and 2 mM trolox]. Binding assays utilizing the ssDNA substrate were performed in 10 mM Tris pH 8.0, 100 mM NaCl, 0.1 mM EDTA pH 8.0. Experiments were performed in helicase buffer (25 mM Tris acetate pH 7.5, 125 mM potassium acetate, and 10 mM magnesium acetate) unless otherwise noted.

Single-pair FRET signals were identified by fitting individual regions of signal intensity to a 2D Gaussian and measuring the goodness of fit. These peaks were corrected for thermal drift.
and local background intensity [205, 206]. The resulting signal was used to calculate the apparent FRET efficiency, \( E_{app} \), according to

\[
E_{app} = \frac{I_A}{I_A + I_D}
\]

in which \( I_A \) and \( I_D \) are the intensity of the acceptor and donor signals respectively. Single-molecule traces that displayed anti-correlation between the donor and acceptor fluorophore as well as single-step fluorophore bleaching were manually selected and used to produce histograms.

### A.3 RESULTS

Helicases must encircle a single strand of DNA to translocate and unwind a dsDNA substrate. As shown in Figure A.1, several cocrystals reveal the details of the contacts between the helicase and nucleic acid, but the dynamics of this interaction remain uncharacterized. Here, we use single-molecule to measure the encircled strand interaction in real time. The immobilized substrates were designed to have a 18bp duplex and either a 30 or 50nt single strand extension. The dyes were placed at the ds-ss junction and at the terminus of the single strand extension. Figure A.2A shows the FRET signal from the ssDNA30 (MAT44:MAT126 or MAT43:MAT129) and ssDNA50 (MAT45:MAT126 or MAT111:MAT129). Introducing \( EcD\)naB onto either the ssDNA30 or ssDNA50 substrates causes shifts to higher FRET states compared to the signal we see for those substrates in the absence of any helicase (Figure A.2B). Consistent with a decreased length, a higher FRET signal is seen for the ssDNA30-\( EcD\)naB complex compared to the ssDNA50-\( EcD\)naB complex.
Figure A.2C shows an increase in FRET upon introducing SsoMCM onto the ssDNA50 substrate. However, loading SsoMCM onto the ssDNA30 substrate shows little change in the distance between the dyes located at the ss-ds junction and the terminus of the 50-nucleotide single strand extension. In order to better compare the experimental data presented in Figure A.2 with data from the cocrystal structures, we generated a plot based on the rise per base measured by either smFRET or as determined by the published cocrystals. This is shown in Figure A.3. FRET signals in the presence of EcDnaB on each substrate. C) shows the FRET signals produced by introducing SsoMCM onto both ssDNA30 (black) and ssDNA50.
Figure A-2: Helicases induced compression of the encircled strand.
A) shows the FRET signals from ssDNA30 (black) and ssDNA50 (blue). B) shows the increased FRET signals in the presence of EcDnaB on each substrate. C) shows the FRET signals produced by introducing SsoMCM onto both ssDNA30 (black) and ssDNA50.
Figure A-3: Rise per base measured by smFRET and crystallography.
This figure visualizes measured lengths of various free and bound DNA substrates. The left side displays data based on smFRET measurements made in our laboratory. Measurements of ssDNA30 and ssDNA50 are shown in black and blue respectively. The total end-to-end distance as measured by FRET is shown above each dataset, and a corresponding rise per base value was calculated based on the overall length and total number of ssDNA bases. The right side shows the overall length for ssDNA30 as bound by Rho, E1, and DnaB based on the rise per base as reported for each cocrystal [52, 68, 70]. There is also a schematic for B-form DNA as well as single strand DNA based on the calculations of Ha et al in [308].

Here, we see that nucleic acids bound within the central channel of hexameric helicases experience a significant reduction in rise per base. There is a significant difference between the distances measured by crystallography and those measured in our lab using smFRET. The structures of hexameric helicases are bound to short oligos ~8-15 nucleotides in length. The smFRET experiments measure the distance changes for much longer substrates (30 and 50 nucleotides). The decreased lengths measured using smFRET may be a result of the inner channel of the hexamer not being able to fully accommodate the length of ssDNA being used, and the excess ssDNA interacting with regions other than the central channel of the hexamer. Also, it
is known that high salt, magnesium, and potassium, which are all present in the helicase buffer being used, all allow for greater flexibility of unbound ssDNA which results in higher FRET [308].

A.4 DISCUSSION AND CONCLUSION

Contact with the encircled strand of DNA during unwinding by a replicative helicase is critical for translocation and dsDNA unwinding [14]. The connection between the compression or scrunching of the nucleic acid substrate bound within the central channel of a hexameric helicase as seen in several crystal structures (Figure A.1) and the translocation and unwinding mechanisms is not clear. We have shown that loading EcDnaB onto DNA substrates with single strand extensions decreases the end-to-end distance of the single strand extension (Figure A.2B). This was true for both 30 and 50 nucleotide lengths of single strand extensions, and the reduction in Cy3-Cy5 distance is consistent with the decrease in rise per base reported by others [70]. The published cocrystal of Geobacillus DnaB bound to DNA in its central channel shows the single strand of DNA exists in a conformation with a significantly reduced rise per base pair compared to ssDNA or even B form DNA. Ribeck et al have proposed that this compression or scrunching of the encircled strand is a critical aspect of DnaB’s unwinding mechanism [215]. This was based on magnetic tweezer-based unwinding assays where increased force applied to the encircled strand decrease unwinding rates. The dynamics of this encircled strand scrunching remain unknown. Our smFRET binding assays showed little to know variation in the FRET signal over time, which corresponds to no detectable dynamics. It is possible that including ATP
in the buffer conditions may result in a dynamic scrunching or compression that is required for translocation and unwinding.

We also performed the same smFRET ssDNA binding assay for the archaeal SsoMCM hexameric helicase (Figure A.2C). We see a significant increase in FRET for the ssDNA50 substrate, similar to what was observed for EcDnaB. However, there is little to no change in the FRET signal when SsoMCM is loaded onto the ssDNA30 substrate. The footprint of SsoMCM is thought to be ~30nt when ssDNA is bound within the central channel. It is possible that SsoMCM does not significantly compress the encircled strand, in constrast to EcDnaB. In that case, the increase in FRET we see when SsoMCM is bound to the ssDNA50 substrate may be due to the excess 20nt not bound within the central channel being bound around of top rim or outside surface of the hexamer. This would still bring the dyes within closer proximity, and would explain why there is no difference in the measured FRET signal when SsoMCM is loaded onto the ssDNA30 substrate. If this is the case, it is possible that the measured increases in FRET and corresponding decreases in rise per base calculated may also be due to interactions between the helicase and ssDNA that go beyond those contacts within the central channel. Various lengths of the single strand extension, including lengths closer to the EcDnaB footprint of ~20 nucleotides [94, 216, 220], will need to be tested in order to determine if the encircled strand scrunching is the only interaction responsible for the measured increase in FRET.

It will be important to determine the role of encircled strand scrunching in the translocation and unwinding mechanisms of hexameric helicases. Single-molecule methods will likely be helpful in elucidating the details of this interaction. However, the detection of dynamics might only be observed during active translocation. Although poorly characterized in the context of hexameric helicases, the idea of DNA scrunching playing a role in nucleic acid translocation
is not new. Like replicative helicases, viral packaging motors are also ring-shaped ATPases. The current model of dsDNA packaging motors includes a ‘crunching’ of the dsDNA as it passes through the motor during packaging [333-335]. It would not be surprising if other toroidal ATPases such as helicases utilized scrunching of the nucleic acid substrate as part of their translocation mechanism.

A.5 CONTRIBUTIONS AND ACKNOWLEDGEMENTS

Brian Graham purified SsoMCM. Sean Carney purified EcDnaB and performed all smFRET experiments and analyses. We thank Brian Graham and Grant Schauer for helpful discussions.
APPENDIX B

DATA ANALYSIS PROGRAMS

B.1 STITCHING PROGRAM

StitchFRET

Creates a single trace from all by stitching all saved traces together.

Requires input of two cell arrays. The first named 'data' with each cell containing data from a single trace in the form of 2 columns for donor signal(column 1) and acceptor(column 2). A second cell array named 'labels' should contain a trace number or name for each trace.

The program will automatically call and open vbFRET program for fitting. Save the 'Ideal Trace' as .mat once you are satisfied with the fitting in vbFRET. Closing vbFRET will allow StitchFRET to resume.

--Sean Carney(Trakselis Lab) August 2015--

clear all
close all

The user must define below variables

NumFrames=1000; %how many frames collected
TimeCollect=0.1; %Integration time for camera collection in seconds.
% (10 frames per second would be TimeCollect=0.1)
MinTime=2; %Minimum trace length in seconds for traces to be included in analysis.

References


Notes

1 %data read into downstream Programs as an xRows by 5Column Matrix, if the number of datapoints is not divisible by 5, they fill in the extra spaces with 0s, which gives false transitions to 0 in the final plot. Let's truncate the data by 1-4 pts to make sure there are no false transitions.

Stitching Block: Do not alter the code below

display('Please select file')
filename = uigetfile;
display('Please select folder for saving')
folder_name = uigetdir;
workdir = folder_name;
cd(workdir);
load(filename);
LoadData = data;
LoadLabels = labels;
n = length(LoadData);
DataStitch = zeros(n*NumFrames, 2);
StitchMatrix = [];
j = 1;
RawIndex = [];
RawCount = 1;
% Keep track of the number of pts put into matrix so we can truncate later
SizeVector = [];
keeptracecount = 0;
tosstracecount = 0;
for i = 1:n
    MolData = LoadData{1, i};
    AddData = MolData;
    L = length(AddData);
    if L < (MinTime/TimeCollect)
        display(i)
        display('_has been tossed_')
tosstracecount = tostracecount + 1;
    else
        SizeVector(i, 1:2) = size(LoadData{1, i});
        DataStitch(j:j+(L-1), 1:2) = AddData(1:L, 1:2);
        StitchMatrix(1, i) = L;
        % To help plot raw w/ ideal traces @ end
        RawIndex(1, RawCount) = i;
        RawCount = RawCount + 1;
        j = L+j;
        keeptracecount = keeptracecount + 1;
    end
end
% Truncate matrix to actual size
numpts = sum(SizeVector(:, 1));
DataStitchTrunc = DataStitch(1: numpts, 1: 2);
% Remove zero rows
SizeVector( all(~SizeVector, 2), :) = [];
% Remove zero columns
StitchMatrix(:, all(~StitchMatrix, 1)) = [];

% Prepare for vbFRET compatibility
data = DataStitchTrunc;
label = strcat('_Stitched_', filename);
% data and labels must be in cell arrays for vbFRET to recognize them
data = {data};
label = {label};
%Save stitched FRET trace to file
save(label,'data','labels');

Fit Stitched Trace to Ideal States using vbFRET (Ref. 1)

%%%%%%-------- now run vbFRET on the saved file to get idealized traces
%               to acquire the dwell times and transitions.
%Open vbFRET to fit stitched trace to ideal states
%Save 'Idealized Traces' as a .mat file
h=vbFRET
%Wait for program to proceed until vbFRET is closed
waitfor(h)

UnStitching Block

%Select the file containing idealized traces from vbFRET use the .mat file
display('Load Idealized Trace File')
Stitched_File=uiimport;

%Truncate StitchMatrix down to # of molecules that were above the threshold
% and used in stitched vbFRET.
StitchMatrix(StitchMatrix == 0) = [];
ptnum = 0;
oldptnum = 0;
IdealTraces = cell(1,length(StitchMatrix));
StitchTrace = Stitched_File.path{1};

%User selects to go through traces
GoTraces=input('Go through traces? 1=Yes or 2=No');
while GoTraces
    if GoTraces==1 || GoTraces==2
        break
    else
        GoTraces=input('Go through traces? 1=Yes or 2=No');
    end
end

%Unstitching Loop - Ideal Traces will contain the separated idealized traces
for t = 1:length(StitchMatrix)
    ptnum = StitchMatrix(1,t)+oldptnum
    if t ==1
        IdealTraces{t} = StitchTrace(1:ptnum1);
        oldptnum=ptnum
    else
        IdealTraces{t} = StitchTrace(oldptnum+1:ptnum1);
        oldptnum = ptnum
    end
    tstring=num2str(t);
    FName=[ workdir '/' tstring '_unStitchPATH_.dat'
    savetrace=[IdealTraces{t}];
    %Note 1
    remainder=rem(length(savetrace),5);
    savetrace=savetrace(1:length(savetrace)-remainder,1);
    if GoTraces==1
% MAKE A PLOT for each to make sure it the indexing is correct
DA = LoadData{RawIndex(t)};
D  = DA(:,1);
A  = DA(:,2);
FRET = A/(D+A);
x = 1:length(FRET);
x = x*TimeCollect;

subplot(2,1,1)
plot(x,A,'Color','red')
hold on
plot(x,D,'Color','green')

subplot(2,1,2)
plot(x,FRET)

molnum = num2str(t);
title(strcat('Molecule',molnum))
xlabel('Time (seconds)')
ylabel('FRET Efficiency')
hold on
plot(x,IdealTraces{t},'Color','red','LineWidth',2);
xlabel('Time (seconds)')
ylabel('FRET Efficiency')
display('Press Enter to go to Next Trace');

w = waitforbuttonpress;
if w == 1
else
end
close

% Save each traces as a .dat file for subsequent analysis(ExPRT)
save(fname, 'savetrace','-ascii');
display(t);
end

% Below code to generate histogram of raw stitched trace--use to
% compare to original red laser hist.

allhist = [];
for i=1:length(FRET)
allhist(1:length(allhist)+1:length(FRET{1,i})) = FRET{1,i};
end
hist(allhist,61)

Published with MATLAB® R2013a
B.2 EXPRT ANALYSIS PROGRAM

ExPRT Plot Porgram: SMC June 2016

% Calculates the transition probability and rate for each unique transition.
% Size of plot marker depends on probability of transition and color depends
% on ave. dwell time. This program uses the unsitiched ideal path files
% (.dat) from vbFRET as input.
clear all;
close all;

User must alter the below variables according to data collection parameters:  1=yes  2=no

LegendinPlot=2; % Display color and size legend within the plot, no places legend within new window 1=yes  2=no
ShowTextBox=2; % Show text boxes on ExPRT Plots:  1=yes  2=no
FirstLast=2; % Include first and last FRET states in analysis:  1=yes  2=no
RelDwells=2; % Do we visualize the double-exp markers as relative abundance/area of circles  1=yes  2=no
ProbabilityOption=1; % 1=probability based on the fraction of single traces containing the transition at least once.  2=probability of transition across all observed transitions.
TimeCollect=100; % How long did we collect data for? (in seconds), for preallocation
FrameRate=10; % How many frames per second are we collecting
mrksize=180; % Choose marker size for final transition plot
MaxDwell=20; % Max Ave Dwell Time of all data of interest used for max on color gradient
MaxDwellCutOff=20; % Dwell Time used as cutoff for dwell time colorbar. All dwell times above this value have same color (black) - good to use if there are outliers
axesfontsize=25; % Font Size of numerical tick labels on ExPRT plot
labelsfontsize=35; % Font size of labels on ExPRT plots

Don't alter anything below this line

References that may be helpful

% CARNEY, S. M. & TRAKSELIS, M. A. 2016. The excluded DNA strand is SEW important for hexameric helicase unwinding. Methods.

Prompt user to select files for ExPRT Analysis

Input is .dat files containing Ideal traces from vbFRET

display('Select all your ideal trace files')
[FName, DirName] = uigetfile('*.dat','Select all your ideal trace files','Multiselect','on');
% Convert array to cell array if user selects only 1 file
if ~iscell(FName)
    NewCell = cell(1);
NewCell = FName;
FName = NewCell;
end
NumTraces = length(FName);
NumTraces = NumTraces(1,1);
TempName = FName;
TempPathName = DirName;
prompt = 'Enter a filename to save data: ';
name = input(prompt,'s');

Preallocate
AllTrace=zeros((TimeCollect*FrameRate),NumTraces);
TransAll=zeros((TimeCollect*FrameRate),(NumTraces*3));
AllTransDwells=[];
counter=0;
tempAllTransDwells=[];
TraceCount=[];

Loop through each ideal FRET trace file
for n=1:NumTraces
    Dwells=[];
    FileName = strcat(TempPathName,TempName{n});
    fid = fopen(FileName, 'rt');
    rawData = fscanf(fid, ' %e %e %e %e %e');
    fclose(fid);
    TraceData=rawData;
    % ideal FRET traces of each molecule in column vectors within one matrix
    AllTrace(1:length(TraceData),n)=TraceData(:,1);
    % Reorganize idealized traces so that new format is:
    % current data point\FRET state\FRET state @next time step
    OrgTrans=[];
    for l=1:length(TraceData)-1
        OrgTrans(l,1)=l;
        OrgTrans(l,2)=TraceData(l,1);
        OrgTrans(l,3)=TraceData(l+1,1);
    end
    TransAll(1:l,n:n+2) = OrgTrans(1:l,1:3);
    % Finds indices where idealized FRET values change - A FRET transition
    TransLocation=find(OrgTrans(:,2)~=OrgTrans(:,3));
    TF = isempty(AllTransDwells);
    % Fill zeros in trace count if there are no transitions
    if isempty(TransLocation)==1
        TraceCount(:,n+2)=0;
    end
    % If there are transitions
    if ~isempty(TransLocation)
        for t=1:length(TransLocation)
            Dwells(t,1)=TransLocation(t);
            Dwells(t,2:3)=OrgTrans(TransLocation(t),2:3);
        end
    end
end
end
\[ ab, ca = \text{size}(\text{DwellTimes}); \]
\text{for } w = 1: \text{size}(\text{DwellTimes}, 1) \%
\text{Use to exclude first and last state/transition}
\text{if FirstLast == 2}
\text{if } w == 1 \text{ \text{||} } w == \text{size}(\text{DwellTimes}, 1); \%
\text{DwellTime of first and last states}
\text{else}
\text{DwellTimes2}(w, 1) = \text{DwellTimes}(w, 1) - \text{DwellTimes}(w - 1, 1); \%
\text{Dwell time calculated by subtracting the time until the next transition occurs}
\text{end}
\%
\text{Use to include first and last state/transition}
\text{elseif FirstLast == 1}
\text{if } w == 1 \text{ \text{||} } w == \text{size}(\text{DwellTimes(:, 1))}
\text{DwellTimes2}(w, 1) = \text{DwellTimes}(w, 1); \%
\text{include first state}
\text{else}
\text{DwellTimes2}(w, 1) = \text{DwellTimes}(w, 1) - \text{DwellTimes}(w - 1, 1); \%
\text{Dwell time calculated by subtracting the time until the next transition occurs}
\text{end}
\text{if } w == \text{size}(\text{DwellTimes, 1)} \text{ \&\& size}(\text{DwellTimes, 1)) == 1
\text{DwellTimes2}(w, 1) = \text{DwellTimes}(w, 1) - \text{DwellTimes}(w - 1, 1);
\text{end}
\text{end}\]
\text{DwellTimes2(:, 2:3) = DwellTimes(:, 2:3); \%
\text{Time spent in initial state before going to final state, initial state, final state}
\text{P = unique(DwellTimes2(:, 2:3), \text{\texttt{rows}})}; \%
\text{Collect each unique FRET transition in a matrix}
\text{TransDwells[]} = \text{;};
\text{TransDwells} = \text{P};
\text{[g, i] = size(TransDwells)};
\text{[b, c] = size(DwellTimes2)};
\text{for } p = 1: b
\text{Index} = \text{ismember}(\text{P}(; 1: 2), \text{DwellTimes2}(p, 2:3), \text{\texttt{rows}});\%
\text{k = find(Index == 1)};
\text{if } \text{DwellTimes}(p, 1) == 0 \text{ \&\& DwellTimes2(p, 2:3) == P(k, 1:2)}
\text{TransDwells(k, i +1) = DwellTimes2(p, 1); \%
\text{i = i +1; \text{\texttt{\texttt{}}}}}
\text{elseif length(\text{DwellTimes2)) <= 2}
\text{fprintf('No Transitions in this File, Moving Ahead\n')}\text{end}\]
\text{end}\]
\text{[ao, bo] = size(P); \%
\text{Collect which molecules are exhibiting which transitions}
\text{ZA = 0;}
\text{if } n == 1 \text{ \&\& ~isempty(TransLocation)}
\text{TraceCount = P; \%
\text{TraceCount(:, 3) = 1; \text{\texttt{\texttt{}}}}}
\text{end}\]
\text{[PO, PR] = size(TraceCount); \%
\text{if } PO == 1 \text{ \&\& TraceCount(PQ, 1:2) == 0}\text{\texttt{\texttt{}}}
\text{TraceCount = P; \%
\text{TraceCount(:, 3) = 0; \text{\texttt{\texttt{}}}}}
\text{end}\]
\text{if } n > 1
[PO,PR] = size(TraceCount);
for G=1:ao
    ZA=0;
    for E=1:PO
        if P(G,1:2) == TraceCount(E,1:2)
            TraceCount(E,1:2) = P(G,1:2);
            TraceCount(E,n+2) = 1;
        else
            ZA = ZA + 1;
        end
    end
    if ZA == E
        TraceCount(PO+1,1:2) = P(G,1:2);
        TraceCount(PO+1,n+2) = 1;
        [PO,PR] = size(TraceCount);
    end
end
end

% Collect all dwell times for each unique transition
for d=1:ao
    TF = isempty(AllTransDwells);
    K = nnz(TransDwells(d,:));
    N = size(nonzeros(TransDwells(d,:)));
    U = size(TransDwells(d,:));
    N=max(N);
    U=max(U);
    [cx, cy] = size(AllTransDwells);
    if counter > 0
        v = 0;
        for R=1:cy
            if TransDwells(d,1) == AllTransDwells(1,R) &&
                TransDwells(d,2) == AllTransDwells(2,R)
                % fprintf('match ')
                L = find(AllTransDwells(:,R),k,'last');
                L = max(L);
                AllTransDwells(L+1:L+(K-2),R) = nonzeros(TransDwells(d,3:U));
                v = v + 1;
            end
        end
        if v == 0 && R == cy
            AllTransDwells(1:N,cy+1) = nonzeros(TransDwells(d,:));
            AllTransDwells(1:2,cy+1) = TransDwells(d,1:2);
        end
    else
        if TF == 0
            tempAllTransDwells = AllTransDwells;
        end
        counter = counter + 1;
    end
end
end

170
Find Probabilities and Mean Dwell Times

```matlab
%AllTransDwells now filled with data points. Convert to units of seconds
[mj,nj]=size(AllTransDwells);
AllTransDwells(3:jm,:)=((AllTransDwells(3:jm,:))*(1/FrameRate));

%Gives the fraction of single molecule traces that exhibit each transition
TraceTransProb=[];
[GE,GO]=size(TraceCount);
TraceTransProb(:,1:2)=TraceCount(:,1:2);
for H=1:GE
    TraceTransProb(H,3)=nnz(TraceCount(H,3:GO))/(GO-2);
end

%Gives the probability of transition based on all transitions observed
TransProb=[];
TotalTrans=nnz(AllTransDwells(3:cx,1:cy));
TransProb(1:2,1:cy)=AllTransDwells(1:2,1:cy);
for H=1:GE
    TransProb(3,H)=nnz(AllTransDwells(3:cx,H))/TotalTrans;
end

%Calculates Mean Dwell Time for each Transition
MeanDwell=[];
MeanDwell(1:2,1:cy)=AllTransDwells(1:2,1:cy);
for H=1:GE
    if AllTransDwells(3,H)==0 %if there are no transitions between any combination of FRET states
        MeanDwell(3,H)=0;
    else
        MeanDwell(3,H)=mean(nonzeros(AllTransDwells(3:cx,H)));
    end
end

Test single or double exp function fits to the decay function
```

```matlab
[mj,nj]=size(AllTransDwells);
%Collect parameters and data
FitMat=[];
Ratio1=[];
Ratio2=[];
Gaus=[];
for u=1:nj
    [f,x]=ecdf(AllTransDwells(3:mj,u));
    figure(u) % create new figure
    transtr1 = num2str(AllTransDwells(1,u));
    transtr2 = num2str(AllTransDwells(2,u));
    UniqueTransDwells=unique(AllTransDwells(3:mj,u));
    if AllTransDwells(6,u)==0 %if there is less than 5 data pts. (dwell times) MATLAB can not fit
        fitMat(1:2,u)=AllTransDwells(1:2,u); %if there is less than 5 data pts.
    else
        MeanDwell(3,H)=mean(nonzeros(AllTransDwells(3:cx,H))); %if there is less than 5 data pts.
    end
end
```

FitMat(4,u)=0; % R-square fit
FitMat(5,u)=0; % rate of transition
FitMat(6,u)=0; % space for 2nd rate
FitMat(7,u)=MeanDwell(3,u); % Ave Dwell Time
FitMat(8,u)=TraceTransProb(u,3); % Prob of trace containing that transition
FitMat(9,u)=TransProb(3,u); % Prob of transition occurring over all transitions
FitMat(10:9+(length(AllTransDwells(:,u)) - 2),u)=AllTransDwells(3:length(AllTransDwells(:,u)),u);

elseif length(UniqueTransDwells)<5 % need at least 4 unique data pts for exp fit functions
    prompt = 'Press 0: Not enough data for fit exp. decay function  
    result = input(prompt);
    if result==0
        FitMat(1:2,u)=AllTransDwells(1:2,u); % initial and final FRET states of the transition
        FitMat(3,u)=result; % num exp fits
        FitMat(4,u)=0; % R-square fit
        FitMat(5,u)=0; % rate of transition
        FitMat(6,u)=0; % space for 2nd rate
        FitMat(7,u)=MeanDwell(3,u); % Ave Dwell Time
        FitMat(8,u)=TraceTransProb(u,3); % Prob of trace containing that transition
        FitMat(9,u)=TransProb(3,u); % Prob of transition occurring over all transitions
        FitMat(10:9+(length(AllTransDwells(:,u)) - 2),u)=AllTransDwells(3:length(AllTransDwells(:,u)),u);
    else
        % Single exponential fit
        subplot(2,2,1) % first subplot
        plot(x,1-f)
        options=fitoptions('exp1');
        [cf,gof]=fit(x,1-f,'exp1',options);
        % title('Transition Single Exponential')
        x_curve=0:0.01:100;
        fit_low=cf.a.*exp(x_curve'.*cf.b);
        Rsquare_1exp=gof.rsquare;
        hold all; plot(x_curve,fit_low,'r','LineWidth',2); % plot exponential fit
        xlim([0 100]);
        kon_pre_exp=cf.a;
        kon_exp=abs(cf.b);
        text(0.5,0.1, ['kon= ' num2str(kon_exp)]);
        title=['Trans=' transtr1 '-' transtr2 '  kon=' num2str(kon_exp) ', rsquare= ' num2str(gof.rsquare)];
        % Suppress latex interpretation in title
        % Double Exponential fit
        subplot(2,2,2) % second subplot
        plot(x,1-f)
        options=fitoptions('exp2');
        [cf,gof]=fit(x,1-f,'exp2',options);
        x_curve=0:0.01:100;
        fit_low=(cf.a.*exp(x_curve'.*cf.b)) + (cf.c*exp(cf.d*x_curve'));
        Rsquare_2exp=gof.rsquare;
        hold all; plot(x_curve,fit_low,'r','LineWidth',2); % plot exponential fit
        xlim([0 100]);
        kon_pre_exp=cf.a;
        kon_exp=abs(cf.b);
        text(0.5,0.1, ['kon= ' num2str(kon_exp)]);
        title=['Trans=' transtr1 '-' transtr2 '  kon=' num2str(kon_exp) ', Rsquare= ' num2str(gof.rsquare)];
        % Suppress latex interpretation in title
        end

end
kon1_exp=abs(cf.b);
kon2_exp=abs(cf.d);
text(0.5,0.1, ['\text{kon1}=' num2str(kon1_exp)])
text(0.5,0.2, ['\text{kon2}=' num2str(kon2_exp)])
titulo=['\text{Trans}=\ \text{transtr1}-' transtr2 '  \text{kon1}=\ ' num2str(kon1_exp) '  \text{kon2}=\ ' num2str(kon2_exp) ',  \text{rsquare}=\ ' num2str(gof.rsquare)]; title (titulo,'interpreter','none');

prompt = 'Press 1 for single exp fit, Press 2 for double exp fit:  ';
result = input(prompt);

if result==1
    FitMat(1:2,u)=AllTransDwells(1:2,u);  \% initial and final FRET states of the transition
    FitMat(3,u)=result;  \% num exp fits
    FitMat(4,u)=Rsquare_1exp;  \% Rsquare fit
    FitMat(5,u)=kon_exp;  \% rate of transition
    FitMat(6,u)=0;  \%Space for 2nd rate
    FitMat(7,u)=MeanDwell(3,u);  \% Ave Dwell Time
    FitMat(8,u)=TraceTransProb(u,3);  \%Prob of trace containing that transition
    FitMat(9,u)=TransProb(3,u);  \%Prob of transition occurring over all transitions
    FitMat(10:9+(length(AllTransDwells(:,u))-
        2),u)=AllTransDwells(3:length(AllTransDwells(:,u)),u);
    Ratio1(1,u)=0;
    Ratio2(1,u)=0;
end

if result==2
    FitMat(1:2,u)=AllTransDwells(1:2,u);  \% initial and final FRET states of the transition
    FitMat(3,u)=result;  \% num exp fits
    FitMat(4,u)=Rsquare_2exp;  \% Rsquare fit
    FitMat(5,u)=kon1_exp;  \% k1 rate of transition
    FitMat(6,u)=kon2_exp;  \% K2 rate of transition
    FitMat(7,u)=MeanDwell(3,u);  \% Ave Dwell Time
    FitMat(8,u)=TraceTransProb(u,3);  \%Prob of trace containing that transition
    FitMat(9,u)=TransProb(3,u);  \%Prob of transition occurring over all transitions
    FitMat(10:9+(length(AllTransDwells(:,u))-
        2),u)=AllTransDwells(3:length(AllTransDwells(:,u)),u);
    if RelDwells==1
        \% for double exp. fit, let's fit dwell times with 2 gaussians to determine the
        \% relative number of transitions corresponding to each rate
        DwellsHist=nonzeros(AllTransDwells(:,u));  \%Truncate so we don't include trailing
        zeros
        [counts,centers]=hist(DwellsHist,40);
        \%- Important: Starting points for coefficients are better than the eventual
        \% calculated coefficients, so we only use 1 iteration below so starting points don't change too
        \% much. This is not ideal! Fix later~!
        options = fitoptions('gauss2', 'Normalization', 'off', 'MaxIter', 400);
        [ff,gof2] = fit(centers',counts','gauss2',options);
        plot(ff,centers,counts)
        \%Save fit coefficients
        Gausamp1=ff.a1;
        Gausmu1=ff.b1;
        Gaussd1=ff.c1;
        Gausamp2=ff.a2;
Gausmu2=f. b2;
Gaussd2=f. c2;
Gaus(1, u)=Gausamp1;
Gaus(2, u)=Gausmul;
Gaus(3, u)=Gaussd1;
Gaus(4, u)=Gausamp2;
Gaus(5, u)=Gausmu2;
Gaus(6, u)=Gaussd2;
figure
bar (centers, counts)
hold on
xNorm=0: 0.1: TimeCollect;
yNorm1=[];
yNorm2=[];
for lX=1:1:length(xNorm)
    yNorm1(lX,1)=Gausamp1*exp(-((xNorm(lX)-Gausmu1)/Gaussd1)^2);
    yNorm2(lX,1)=Gausamp2*exp(-((xNorm(lX)-Gausmu2)/Gaussd2)^2);
end

% Use below lines to see fits to gaussian distribution of double exp dwell times.
plot(xNorm,yNorm1,'Color','red')
hold on
plot(xNorm,yNorm2,'Color','red')

% Calculate Area under each gaussian
fun1= @(xNorm) Gausamp1*exp(-((xNorm-Gausmu1)/Gaussd1).^2);
fun2= @(xNorm) Gausamp2*exp(-((xNorm-Gausmu2)/Gaussd2).^2);
IntArea1=integral(fun1,0,TimeCollect);
IntArea2=integral(fun2,0,TimeCollect);

% Take this ratio, and use it to size the two markers based on the frequency or abundance of each dwell-time transition
end
end
close(figure(u))
end

% Round FitMat values to display correct sig figs
FitMatr=[];
% FRET values get 4 sig figs
FitMatr(1:2,:)=round(FitMat(1:2,:)/.0001)*.0001;
FitMatr(3,:)=FitMat(3,:);
% Rsq values get 3 sig figs as do teh ave dwell, trans and trace probs
FitMatr(4,:)=round(FitMat(4,:)/.001)*.001;
for y=1:u
    FitMatr(5:9,y)=round(FitMat(5:9,y)/.001)*.001;
FitMatr(10:9+(length(AllTransDwells(:,y))-2),y)=AllTransDwells(3:length(AllTransDwells(:,y)),y);
end

Build custom color gradient matrix

colors=[1,0,1,0]; %Start gradient from magenta, loop takes it through blue - green - red - black
for tt=2:400
    if tt<50 %Magenta to blue
        colors(tt,1)=colors(tt-1,1)-0.02;
        colors(tt,2)=0;
        colors(tt,3)=1;
    end
    if tt==50
        colors(tt,1)=0;
        colors(tt,2)=0;
        colors(tt,3)=1;
    end
    if tt>50 && tt<100 %Blue to Cyan
        colors(tt,1)=0;
        colors(tt,2)=colors(tt-1,2)+0.02;
        colors(tt,3)=1;
    end
    if tt==100
        colors(tt,1)=0;
        colors(tt,2)=1;
        colors(tt,3)=1;
    end
    if tt>100 && tt<150 %Cyan to Green
        colors(tt,1)=0;
        colors(tt,2)=1;
        colors(tt,3)=colors(tt-1,3)-0.01;
    end
    if tt==150
        colors(tt,1)=0;
        colors(tt,2)=1;
        colors(tt,3)=0;
    end
    if tt>150 && tt<200 %Green to yellow
        colors(tt,1)=colors(tt-1,1)+0.02;
        colors(tt,2)=1;
        colors(tt,3)=0;
    end
    if tt==200
        colors(tt,1)=1;
        colors(tt,2)=1;
        colors(tt,3)=0;
    end
    if tt>200 && tt<250 %Yellow to Orange
        colors(tt,1)=1;
        colors(tt,2)=colors(tt-1,2)-0.01;
        colors(tt,3)=colors(tt-1,3)+0.004;
    end
    if tt==250
        colors(tt,1)=1;
        colors(tt,2)=1;
        colors(tt,3)=0;
    end
if tt>250 && tt<300 %Orange to Red
    colors(tt,1) = 1;
    colors(tt,2) = colors(tt-1,2) - 0.01;
    colors(tt,3) = colors(tt-1,3) - 0.004;
end
if tt==300
    colors(tt,1) = 1;
    colors(tt,2) = 0;
    colors(tt,3) = 0;
end
if tt>300 && tt<400 %Red to Black
    colors(tt,1) = colors(tt-1,1) - 0.01;
    colors(tt,2) = 0;
    colors(tt,3) = 0;
end
if tt==400
    colors(tt,1) = 0;
    colors(tt,2) = 0;
    colors(tt,3) = 0;
end

% Index the percentage for later usage in matching gradients
colors(tt,4) = tt/400;

Match Colors to Dwell Times

colorindex=[];
colorindex(1:2,1:cy)=AllTransDwells(1:2,1:cy);
prompt2 = 'Press 1 to use internal MaxDwell Time, Press 2 to use user-defined MaxDwell Time,
Press 3 to use MaxDwell CutOff:   ';
result2 = input(prompt2);
prompt3 = 'Display colorbar? Type 1 for yes OR 2 or No:   ';
result3 = input(prompt3);

% Below used for using the max within each individual dataset
if result2 == 1
    FitCheck = FitMatr(3,:) == 2;
    if any(FitCheck) == 1 % For a case where kon1 or kon2 from double exp. fit are longer than
dwells from mean dwells
        A = FitMatr(5:6,:);
        LongestDwell = min(A(A>0));
        MaxDwell | Ti me = LongestDwell; % 1/ Rate=Dwell
        if MaxDwell | Ti me >= max(meanMeanDwell(3,1:nj)) % Check if Dwells from 2 exp fit are longer than
            MeanDwells
            MaxDwell | Ti me = MaxDwell | Ti me;
        else
            MaxDwell | Ti me = max(meanMeanDwell(3,1:nj));
        end
end
if result2 == 2
    for H=1:GE
        if FitMat(3,H) == 1 || FitMat(3,H) == 0 % 1 state fitting in dwell time analysis
            colorindex(3,H) = MeanDwell(3,H) / MaxDwellTime;
        end
        if FitMat(3,H) == 2 % 2 state fitting in dwell time analysis
            colorindex(3,H) = 1 / FitMatr(5,H) / MaxDwellTime; % 1/Rate=dwell time
            colorindex(4,H) = 1 / FitMatr(6,H) / MaxDwellTime;
            colorindex(3:4,1:GE) = round(colorindex(3:4,1:GE) / 0.01) * 0.01;
        end
        colorindex(3,1:GE) = round(colorindex(3,1:GE) / 0.01) * 0.01;
    end
end
% Use below to use user defined MaxDwell Value
if result2 == 2
    for H=1:GE
        if FitMat(3,H) == 1 || FitMat(3,H) == 0 % 1 state fitting in dwell time analysis
            colorindex(3,H) = MeanDwell(3,H) / MaxDwell;
            if colorindex(3,H) > 0.75
                colorindex(3,H) = 1;
            else
                colorindex(3,H) = colorindex(3,H) * 0.75; %--to put on colorscale of 300(up to red)
            end
        end
        if FitMat(3,H) == 2 % 2 state fitting in dwell time analysis
            colorindex(3,H) = 1 / FitMatr(5,H) / MaxDwellCutOff; % 1/Rate=dwell time
            if colorindex(3,H) > 0.75
                colorindex(3,H) = 1;
            else
                colorindex(3,H) = colorindex(3,H) * 0.75; %--to put on colorscale of 300(up to red)
            end
            colorindex(4,H) = 1 / FitMatr(6,H) / MaxDwellCutOff;
        end
    end
end
% Use below to utilize DwellTimeCutOff so that a values above the user-defined cut off are all black, rest of scale goes to red.
if result2 == 3
    for H=1:GE
        if FitMat(3,H) == 1 || FitMat(3,H) == 0 % 1 state fitting in dwell time analysis
            colorindex(3,H) = MeanDwell(3,H) / MaxDwellCutOff;
            if colorindex(3,H) > 0.75
                colorindex(3,H) = 1;
            else
                colorindex(3,H) = colorindex(3,H) * 0.75; %--to put on colorscale of 300(up to red)
            end
        end
        if FitMat(3,H) == 2 % 2 state fitting in dwell time analysis
            colorindex(3,H) = 1 / FitMatr(5,H) / MaxDwellCutOff; % 1/Rate=dwell time
            if colorindex(3,H) > 0.75
                colorindex(3,H) = 1;
            else
                colorindex(3,H) = colorindex(3,H) * 0.75; %--to put on colorscale of 300(up to red)
            end
            colorindex(4,H) = 1 / FitMatr(6,H) / MaxDwellCutOff;
        end
    end
if colorindex(4,H) > 0.75
    colorindex(4,H) = 1;
else
    colorindex(4,H) = colorindex(4,H) * 0.75; % to put on colorscale of 300 (up to red)
end

colorindex(3:4,1:GE) = round(colorindex(3:4,1:GE)/0.01)*0.01;

end

end

% now match fraction of gradient values and get indices for corresponding color
for H=1:GE
    if FitMat(3,H) == 1 % 1 state fitting in dwell time analysis
        matchvalue = num2str(colorindex(3,H));
        matchvalue = str2num(matchvalue);
        Ind = colors(:,4) == matchvalue;
        I = find(Ind == 1);
        colorindex(4,H) = I;
    end
    if FitMat(3,H) == 2 % 2 state fitting in dwell time analysis
        matchvalue = num2str(colorindex(3,H));
        matchvalue = str2num(matchvalue);
        Ind = colors(:,4) == matchvalue;
        I = find(Ind == 1);
        colorindex(5,H) = I;
        matchvalue = num2str(colorindex(4,H));
        matchvalue = str2num(matchvalue);
        Ind = colors(:,4) == matchvalue;
        I = find(Ind == 1);
        colorindex(6,H) = I;
    end
end

% New figure for transition plot
figure(u+1)
% Create axes
axes1 = axes('Parent', figure(u+1), 'YTickLabel', {'0.0', '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8', '0.9', '1.0'}, 'YTick', [0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1], 'XTickLabel', {'0.0', '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8', '0.9', '1.0'}, 'XTick', [0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1], 'TickDir', 'out', 'Position', [0.13 0.153591160220994 0.775 0.807734806629834], 'FontWeight', 'bold', 'FontSize', axesfontsize);
xlim(axes1, [0 1]);
ylim(axes1, [0 1]);
box(axes1, 'on');
hold(axes1, 'all');
for h=1:nj
    xlabel({'Initial FRET', 'FontSize', 'label sfont size', 'font weight', 'bold d'});
    ylabel({'Final FRET', 'FontSize', 'label sfont size', 'font weight', 'bold d'});
```matlab
x lim([0, 1]);
y lim([0, 1]);
hold on
if FitMat(3, h) == 2
    if RelDwells == 2
        q = plot(FitMat(1, h), FitMat(2, h), 'o', 'MarkerEdgeColor', 'Black', 'LineWidth', 2);
        % Linear marker size
        if ProbabilityOption == 1
            set(q, 'MarkerSize', mrksize*(TraceTransProb(h, 3))); % Size here ---------------
        elseif ProbabilityOption == 2
            set(q, 'MarkerSize', mrksize*(TransProb(3, h))); % Size here -------------------
        end
        set(q, 'MarkerFaceColor', colors(colorindex(5, h), 1:3));
        hold on
        qp = plot(FitMat(1, h), FitMat(2, h), 'o', 'MarkerEdgeColor', 'Black', 'LineWidth', 2);
        % Linear marker size
        if ProbabilityOption == 1
            set(qp, 'MarkerSize', (mrksize*(TraceTransProb(h, 3)))/2, 'MarkerFaceColor', colors(colorindex(6, h), 1:3)); % Size here ---------------
        elseif ProbabilityOption == 2
            set(qp, 'MarkerSize', (mrksize*(TransProb(3, h)))/2, 'MarkerFaceColor', colors(colorindex(6, h), 1:3)); % Size here -------------------
        end
    end
end
if RelDwells == 1
    q = plot(FitMat(1, h), FitMat(2, h), 'o', 'MarkerEdgeColor', 'Black', 'LineWidth', 2);
    % Linear marker size
    AreaMarker = ((pi*(TraceTransProb(h, 3)*mrksize)^2))/4; % Gives area of circle based on current marker size and
    elseif ProbabilityOption == 2
        AreaMarker = ((pi*(TransProb(3, h)*mrksize)^2))/4; % Gives area of circle based on current marker size and
    end
    RelMarkerSize1 = mrksize*(TraceTransProb(h, 3));
    elseif ProbabilityOption == 2
        RelMarkerSize1 = mrksize*(TransProb(3, h));
    end
    RelMarkerSize2 = sqrt((AreaMarker*pi)/4); % Gives relative size of second marker based on relative area, diameter in point number is marker size
    set(q, 'MarkerSize', RelMarkerSize2); % Size here -------------------------------
    set(q, 'MarkerFaceColor', colors(colorindex(5, h), 1:3));
    hold on
    qp = plot(FitMat(1, h), FitMat(2, h), 'o', 'MarkerEdgeColor', 'Black', 'LineWidth', 2);
    % Linear marker size
    set(qp, 'MarkerSize', RelMarkerSize2, 'MarkerFaceColor', colors(colorindex(6, h), 1:3)); % Size here -------------------------------
end
```
end

hold on
else if Dwell Time analysis gives a single exponential decay
hold on
q=plot(FitMat(1,h),FitMat(2,h),'o','MarkerEdgeColor','Black','LineWidth',2);

%%%Linear marker size
if ProbabilityOption==1
    set(q,'MarkerSize',mrksize*(TraceTransProb(h,3)));%Size here------------------------
elseif ProbabilityOption==2
    set(q,'MarkerSize',mrksize*(TransProb(3,h)));%Size here------------------------
end

set(q,'MarkerFaceColor',colors(colorindex(4,h),1:3));
hold on

qp=plot(FitMat(1,h),FitMat(2,h),'o','MarkerEdgeColor','Black','LineWidth',2);

%%%Linear marker size
end

Below lines for insertion of textbox containing probability and kinetic information
if ShowTextBox==1
    if FitMat(3,h)==0
        ratestr='Rate= '; %
        rate=num2str(FitMatr(5,h));
        ratecat=strcat(ratestr,rate);%
        Rsqstr='Rsq= ';%
        Rsq=num2str(FitMatr(4,h));
        Rsqcstr=strcat(Rsqstr,Rsq);
        TraceFract='TraceProb= ';%
        TraceProb=num2str(FitMatr(8,h));
        TraceProbcat=strcat(TraceFract,TraceProb);
        TransProbstr='TransProb= ';%
        TransProbv=num2str(FitMatr(9,h));
        TransProbcat=strcat(TransProbstr,TransProbv);
        MeanDwellstr='AveDwell= ';
        MeanDwelle=num2str(FitMatr(7,h));
        MeanDwellcat=strcat(MeanDwellstr,MeanDwelle);
        str = {ratecat,Rsqcat,MeanDwellcat,TraceProbcat,TransProbcat};
        str = {ratecat,MeanDwellcat};
        annotation('textbox', [FitMat(1,h)-.2,FitMat(2,h)-.2,0.1,0.1], 'String', str);
    end
    if FitMat(3,h)==1
        ratestr='Rate= '; %
        rate=num2str(FitMatr(5,h));
        ratecat=strcat(ratestr,rate);%
        Rsqstr='Rsq= ';%
        Rsq=num2str(FitMatr(4,h));
        Rsqcstr=strcat(Rsqstr,Rsq);
        TraceFract='TraceProb= ';%
        TraceProb=num2str(FitMatr(8,h));
        TraceProbcat=strcat(TraceFract,TraceProb);
        TransProbstr='TransProb= ';%
        TransProbv=num2str(FitMatr(9,h));
        TransProbcat=strcat(TransProbstr,TransProbv);
        MeanDwellstr='AveDwell= ';
        MeanDwelle=num2str(FitMatr(7,h));
        MeanDwellcat=strcat(MeanDwellstr,MeanDwelle);
        str = {ratecat,Rsqcat,MeanDwellcat,TraceProbcat,TransProbcat};
        str = {ratecat,MeanDwellcat};
        annotation('textbox', [FitMat(1,h)-.2,FitMat(2,h)-.2,0.1,0.1], 'String', str);
    end
end
Create a new table with rates/probabilities, numbers

SaveTableValid = FitMatr(1:9,:);
for nj = 1:GE
    SaveTableValid(10,nj) = n; % This is the number of molecules
    SaveTableValid(11,nj) = nnz(TraceCount(nj,3:GO)); % number of transitions for each unique transition
end
Create Marker Legend

if LegendinPlot==2 %o generate new window for legend
    figure(u+2)
    xlim([0,1]);
    ylim([0,1]);
    hold on
end

probs=[.05,.20,.50,.75,1];

%msizes=180;
probs(2,:)=probs(1,:)*msizes;
xmarker=[.1,.1,.1,.1,.1];
ymarker=[.2,.26,.38,.55,.8];
for i=1:5
    q=plot(xmarker(i),ymarker(i), 'o','MarkerEdgeColor','Black','LineWidth',2);
    set(q, 'MarkerSize',(probs(2,i)));
    set(q, 'MarkerFaceColor','Black');
    hold on
    qp=plot(xmarker(i),ymarker(i), 'o','MarkerEdgeColor','Black','LineWidth',2);
    set(qp, 'MarkerFaceColor','White');
    num=probs(1,i)*100;
    num2str(num);
    percent= '%';
cat=strcat(num,percent);
    str = {cat};
    annotation('textbox',[xmarker(i)+.15,ymarker(i),.1,.1],'String', cat);
end

%number of ticks you want on the colorbar
if result2==1
    MaxDwellTime=round(MaxDwellTime/2)*2; %Round MaxDwellTime up to nearest 5
    ticknum=MaxDwellTime/2;
else if result2==2
    MaxDwellTime=round(MaxDwellTime/2)*2; %Round MaxDwellTime up to nearest 5
    ticknum=MaxDwellTime/2;
else if result2==3
    MaxDwellCutOff=round(MaxDwellCutOff/2)*2; %Round MaxDwellCutOff up to nearest 5
    ticknum=MaxDwellCutOff/2;
end

TickPos=[];
TickLabels={};
if result2==1
    for kk=ticknum:-1:1
        tickval=round(MaxDwellTime*(kk/ticknum));
        TickLabels{1,kk}=num2str(tickval);
        TickPos(1,kk)=(kk/ticknum)*400; %400 is length of color matrix
    end
else if result2==2
    for kk=ticknum:-1:1
        tickval=round(MaxDwell*(kk/ticknum));
        TickLabels{1,kk}=num2str(tickval);
        TickPos(1,kk)=(kk/ticknum)*400; %400 is length of color matrix
    end
else if result2==3
    for kk=ticknum:-1:1
        tickval=round(MaxDwellCutOff*(kk/ticknum));
        TickLabels{1,kk}=num2str(tickval);
        TickPos(1,kk)=(kk/ticknum)*300; %300 is length of color matrix here
    end
else if result2==1

%Get colorbar with appropriate ticks/labels onto plot
    colormap(colors(:,1:3));
    cbh = colorbar('location', 'EastOutside','FontSize',16,'FontWeight','bold');
    set(cbh,'YTick',TickPos);
    set(cbh,'YTickLabel',TickLabels);
    set(get(cbh,'title'),'String','Ave Dwell Time(sec)','fontsize',16,'fontweight','bold');
end
end
if result2==2
    for kk=ticknum:-1:1
        tickval=round(MaxDwell*(kk/ticknum));
        TickLabels{1,kk}=num2str(tickval);
        TickPos(1,kk)=(kk/ticknum)*400; %400 is length of color matrix
    end
else if result2==3
    for kk=ticknum:-1:1
        tickval=round(MaxDwellCutOff*(kk/ticknum));
        TickLabels{1,kk}=num2str(tickval);
        TickPos(1,kk)=(kk/ticknum)*300; %300 is length of color matrix here
    end
else if result2==1

%Get colorbar with appropriate ticks/labels onto plot
    colormap(colors(1:300,1:3));
    cbh = colorbar('location', 'EastOutside','FontSize',16,'FontWeight','bold');
    set(cbh,'YTick',TickPos);
B.3 UNWINDING-REWINDING ANALYSIS PROGRAM

Twinkle Unwind/Rewind Analysis Program

Don't alter below code

```matlab
% Create histogram from fname_vbFRET.mat
% File put into /traces after red laser
% Data: col_1=D, col_2=A, we will make col_3=FRET
% --Sean Carney June 2016, Trakselis Lab--

TimeCollect = 0.1; % time/frame 0.1=10 frames/sec

TraceType = zeros(1, 6);
Type1 = []; Type2 = []; Type3 = []; Type4 = []; Type5 = []; Type6 = [];
I deal 1 = []; I deal 2 = []; I deal 3 = []; I deal 4 = []; I deal 5 = []; I deal 6 = [];
Count 1 = 1; Count 2 = 1; Count 3 = 1; Count 4 = 1; Count 5 = 1; Count 6 = 1;
Ti meCollect A = []; Ti meCollect B = []; Ti meCollect C = []; Ti meCollect D = [];
Ti meCollect E = []; Ti meCollect F = []; Ti meCollect All = [];
StepFrom = []; StepTo = []; StepLength = []; Unw ndTi me = []; Unw ndTi me 2 = []; Rewi ndTi me = [];
StepNum1 = []; StepNumRewi nd = []; StepNumUnw nd 2 = [];
Unw ndDat a = []; Unw ndDat a 2 = []; Rewi nd Dat a = []; Unw ndDat Count = 1; Unw ndDat Count 2 = 1; Rewi ndDat Count = 1; NoSaveCount = 0; NoSaveCount 2 = 0;
% Loop through data and collect
```
for i = 1:length(data)
    Data = data{i};
    [nm, mm] = size(Data);
    if isempty(Data) == 1
        continue
    else
        % calculate FRET
        D = (Data(:, 1));
        A = (Data(:, 2));
        E = (A ./ (D + A));
        Ideal = path{i};
        [n, m] = size(Data);
        Time = (1:nm) .* TimeCollect; % get your time matrix for x axis
        NumStr = 'Trace Num_';
        molnum = num2str(i);
        molnum = strcat(NumStr, molnum);
        % get length of trace
        TraceLength = size(path{i});
        TraceLength = TraceLength(1);
        % Lets plot each trace
        figure
        subplot(2, 1, 1);
        plot(Time, D, 'g', Time, A, 'r');
        axis([0, nm * TimeCollect, -.1, 1])
        subplot(2, 1, 2);
        plot(Time, E, 'b', Time, Ideal, 'r');
        axis([0, nm * TimeCollect, -.1, 1])
        % Let sort traces
        Strprompt1 = '_ Enter: e end  0 Skip   1 DNA  2 DNA Cy5 Bleach  3 Unwind  4 Unwind - Rewind
5 Cy3 (and Cy5) Bleach each  6 Other->   ';
        prompt1 = strcat(molnum, Strprompt1);
        assign = input(prompt1, 's');
        strtest = {'0', '1', '2', '3', '4', '5', '6'};
        assignCheck = strcmp(assign, strtest);
        assignCheck = any(assignCheck);
        while assignCheck == 0
            Strprompt1 = '_ Re-Enter: e end  0 Skip   1 DNA  2 DNA Cy5 Bleach each  3 Unwind  4 Unwind - Rewind
5 Cy3 (and Cy5) Bleach each  6 Other->   ';
            prompt1 = strcat(molnum, Strprompt1);
            assign = input(prompt1, 's');
            strtest = {'e', '0', '1', '2', '3', '4', '5', '6'};
            assignCheck = strcmp(assign, strtest);
            assignCheck = any(assignCheck);
        end
        if assign == 'e'
            break
        end
        if assign == '0'
            continue
        end
% or histogram
[nn, mn] = size(FRETHistData);
FRETHistData(nn+1:nn+n, 1) = E(:, 1);
FRETData(:, i) = E(:, 1);

% DNA or High FRET with No Bleaching
if assign == '1'
    TraceType(1, 1) = TraceType(1, 1) + 1;
    Type1(1: nm, 3*Count1 - 2) = D;
    Type1(1: nm, 3*Count1 - 1) = A;
    Type1(1: nm, 3*Count1) = E;
    Ideal1(:, Count1) = Ideal;
    TimeCollectA(1, Count1) = TraceLength;
    Count1 = Count1 + 1;
end

% DNA Cy5 bleaching event
elseif assign == '2'
    TraceType(1, 2) = TraceType(1, 2) + 1;
    Type2(1: nm, 3*Count2 - 2) = D;
    Type2(1: nm, 3*Count2 - 1) = A;
    Type2(1: nm, 3*Count2) = E;
    Ideal2(:, Count2) = Ideal;
    TimeCollectB(1, Count2) = TraceLength;
    % find first occurrence of low FRET state, and trim array
    [~, idx] = min(Ideal(:, any(Ideal)));
    TimeCollectB(1, Count2) = idx - 1;
    Ideal2(:, idx : end) = [];
    Count2 = Count2 + 1;
end

% DNA Unwinding Events
elseif assign == '3'
    TraceType(1, 3) = TraceType(1, 3) + 1;
    Type3(1: nm, 3*Count3 - 2) = D;
    Type3(1: nm, 3*Count3 - 1) = A;
    Type3(1: nm, 3*Count3) = E;
    Ideal3(:, Count3) = Ideal;
    TimeCollectC(1, Count3) = TraceLength;
    Strprompt2 = 'Use this trace for Unwinding Time Analysis? 1=yes  2=no'
    assig2 = input(Strprompt2, 's');
    strtest2 = {'1','2'};
    assignCheck2 = strcmp(assig2, strtest2);
    assignCheck2 = any(assignCheck2);
    while assignCheck2 ~= 1
        Strprompt2 = 'Use this trace for Unwinding Time Analysis? 1=yes  2=no'
        assig2 = input(Strprompt2, 's');
        strtest2 = {'1','2'};
        assignCheck2 = strcmp(assig2, strtest2);
        end
    end
    if assig2 == '1'
        % use cross hairs to select pre-and-post unwound DNA FRET states
        [x, y] = ginput;
        while length(x) ~= 2
display('Re-select Unwind Start and End points - >');
[x, y] = ginput;
end
x=roundn(x, -1); % round and multiply so we can use as indices
x=x*(1/TimeCollect);
% Save Selected Unwind Region
UnwindRegion(:,1,:)=deal(x(1,1):x(2,1),UnwindCount);
UnwindData(1:length(UnwindRegion),UnwindCount)=UnwindRegion; % Save Unwind Region in matrix

% find first occurrence of LowFRET state
[~, First_Low_idx]=min(UnwindRegion(:,any(UnwindRegion)));
% find last occurrence of HighFRET state
Last_High_idx=find(UnwindRegion==max(UnwindRegion),1,'last');
% Calculate and Save Unwinding Time from last high FRET pt. to first Low FRET state
UnwindTime(:,UnwindCount)=First_Low_idx-Last_High_idx;
UnwindTime(:,UnwindCount)=UnwindTime(:,UnwindCount)*TimeCollect;

% sort out indices of steps
[s1, s2]=size(UnwindRegion);
transcount=1;
for j=2:s1
    if UnwindRegion(j,1)~=UnwindRegion(j-1,1)
        StepFrom(transcount,UnwindCount)=UnwindRegion(j-1,1);
        StepTo(transcount,UnwindCount)=UnwindRegion(j,1);
        StepLength(transcount,UnwindCount)=j;
        transcount=transcount+1;
    end
end
StepNum(:,UnwindCount)=transcount; % transcount is # of states for a single unwinding trace
% Number of steps taken to unwind
UnwindCount=UnwindCount+1;

% Assign 2 for traces that show unwinding but we don't want to/can not save the time/steps
if assign2=='2'
    NoSaveCount=NoSaveCount+1;
end
Count3=Count3+1;

% Unwinding - Rewinding Events
elseif assign=='4'
    TraceType(:,4)=TraceType(:,4)+1;
    Type4(:,nm)=Type4(:,nm)+1;
    Type4(:,nm)=Type4(:,nm)+1;
    Type4(:,nm)=Type4(:,nm)+1;
    Ideal4(:,nm)=Ideal4(:,nm)+1;
    TimeCollect(:,nm)=TimeCollect(:,nm)+1;
    Strprompt3 = 'Use this trace for Unwinding Time Analysis? 1=yes  2=no';
    assign3=input(Strprompt3,'s');
    strtest3={'1','2'};
    assignCheck3=0;
assignCheck3 = strcmp(assign3, strtest3);
assignCheck3 = any(assignCheck3);
while assignCheck3 ~= 1
    Strprompt3 = 'Use this trace for Unwinding Time Analysis? 1=yes  2=no';
    assign3 = input(Strprompt3, 's');
    strtest3 = {'1', '2'};
    assignCheck3 = strcmp(assign3, strtest3);
    assignCheck3 = any(assignCheck3);
end
if assign3 == 1'
    %use cross hairs to select pre-and-post unwound DNA FRET states
    [x, y] = ginput;
    checkrem = rem(length(x), 4);
    while checkrem ~= 0 %Check for correct number of points (4 for every unwind-rewind event)
        disp('Re-select Unwind Start and End points - ->');
        [x, y] = ginput;
        checkrem = rem(length(x), 4);
    end
    x = roundn(x, -1); %round and multiply so we can use as indices
    x = x *(1/TimeCollect);
    xlength = length(x);
    %Loop through each set of 4 x, y pairs from clicking unwinding and rewinding events
    for h = 4:4:xlength
        Section for Unwinding Events %%%%%
        %Save Selected Unwind Regions
        UnwindRewindRegion = deal(4*(x(h-3, 1):x(h-2 ,1), Count4));
        UnwindData2(1:length(UnwindRewindRegion), UnwindCount2) = UnwindRewindRegion; %Save
        %find first occurrence of Low FRET state
        [~, First_Low_idx] = min(UnwindRewindRegion(:, any(UnwindRewindRegion)));
        First_Low_idx = find(UnwindRewindRegion == min(UnwindRewindRegion), 1, 'first');
        %find last occurrence of High FRET state
        Last_High_idx = find(UnwindRewindRegion == max(UnwindRewindRegion), 1, 'last');
        %Calculate and Save Unwinding Time from last high FRET pt. to First Low FRET state
        UnwindTime2(1, UnwindCount2) = First_Low_idx - Last_High_idx;
        UnwindTime2(1, UnwindCount2) = UnwindTime2(1, UnwindCount2) * TimeCollect;
        %sort out indices of steps
        [s1, s2] = size(UnwindRewindRegion);
        transcount = 1;
        for j = 2:s1
            if UnwindRewindRegion(j, 1) != UnwindRewindRegion(j-1, 1)
                StepFrom(transcount, UnwindCount2) = UnwindRewindRegion(j-1, 1);
                StepTo(transcount, UnwindCount2) = UnwindRewindRegion(j, 1);
                StepLen = length(transcount, UnwindCount2) - j;
                transcount = transcount + 1;
            else

Section for Unwinding Events %%%%%

%Save Selected Unwind Regions
UnwindRewindRegion = deal(4*(x(h-3, 1):x(h-2 ,1), Count4));
UnwindData2(1:length(UnwindRewindRegion), UnwindCount2) = UnwindRewindRegion; %Save
%find first occurrence of Low FRET state
%~ First_Low_idx = min(UnwindRewindRegion(:, any(UnwindRewindRegion)));
First_Low_idx = find(UnwindRewindRegion == min(UnwindRewindRegion), 1, 'first');
%find last occurrence of High FRET state
Last_High_idx = find(UnwindRewindRegion == max(UnwindRewindRegion), 1, 'last');
%Calculate and Save Unwinding Time from last high FRET pt. to First Low FRET state
UnwindTime2(1, UnwindCount2) = First_Low_idx - Last_High_idx;
UnwindTime2(1, UnwindCount2) = UnwindTime2(1, UnwindCount2) * TimeCollect;
%sort out indices of steps
[s1, s2] = size(UnwindRewindRegion);
transcount = 1;
for j = 2:s1
    if UnwindRewindRegion(j, 1) != UnwindRewindRegion(j-1, 1)
        StepFrom(transcount, UnwindCount2) = UnwindRewindRegion(j-1, 1);
        StepTo(transcount, UnwindCount2) = UnwindRewindRegion(j, 1);
        StepLen = length(transcount, UnwindCount2) - j;
        transcount = transcount + 1;
    else

188
Section for Rewinding Events

% Save Selected Rewinding Regions
RewindRegion = deal(4(x(h-1,1):x(h,1),Count4));
RewindData(1:length(RewindRegion),RewindCount) = RewindRegion; % Save Unwind Region in matrix

% Save Selected Unwind Regions
StepNumUnwind2(1,UnwindCount2) = transcount; % transcount is # of states for a single unwinding trace
UnwindCount2 = UnwindCount2 + 1;

if assign3 == '2' % For traces that show unwinding-rewinding but we don't want to/can not save the time/steps
NoSaveCount2 = NoSaveCount2 + 1;
end
Count4 = Count4 + 1;

% Cy3 (and Cy5) Bleaching Event - loss of signal may be Cy3-labeled strand dissociating
elseif assign == '5'
TraceType(1,5) = TraceType(1,5) + 1;
Type5(1:nm,(3*Count5)-2) = D;
Type5(1:nm,(3*Count5)-1) = A;
Type5(1:nm,3*Count5) = E;
Ideal5(1:nm,Count5) = Ideal;
TimeCollectE(1,Count5) = TraceLength;
end
Count 5 = Count 5 + 1;

% Other types of traits
else if assign == '6'
    TraceType(1, 6) = TraceType(1, 6) + 1;
    Type6(1: nm (3*Count 6) - 2) = D;
    Type6(1: nm (3*Count 6) - 1) = A;
    Type6(1: nm 3*Count 6) = E;
    deal 6(1: nm Count 6) = deal;
    TimeCollectF(1, Count 6) = TraceLength;
    Count 6 = Count 6 + 1;
end
end
close
end

Correct for figures below

UnwindTime = UnwindTime * TimeCollect;
StepNum = StepNum - 1; % StepNum is the number of states or stairs, we want the number of steps taken, which is 1 less
StepNumUnwind2 = StepNumUnwind2 - 1;
StepNumRewind = StepNumRewind - 1;
prompt3 = 'Select Folder to Save Data to: ';
foldername = uigetdir;
cd(foldername);
prompt5 = 'Folder name for saving';
name5 = input(prompt5, 's');
mkdir(foldername, name5)
cd(name5)

Create hist of 'Dye Bleaching Times'

% A-DNA only is fine the way it is
% B-DNA bleach - we take time until lowest FRET State
% C-DNA Unwind - we take time until lowest FRET State
% D-Other - Leave as is for now
TimeCollectAll(1: numel(TimeCollectA), 1) = TimeCollectA;
TimeCollectAll(1: numel(TimeCollectB), 2) = TimeCollectB;
TimeCollectAll(1: numel(TimeCollectC), 3) = TimeCollectC;
TimeCollectAll(1: numel(TimeCollectD), 4) = TimeCollectD;
TimeCollectAll(1: numel(TimeCollectE), 5) = TimeCollectE;
TimeCollectAll(1: numel(TimeCollectF), 6) = TimeCollectF;

% Convert back to seconds
TimeCollectAllSec = TimeCollectAll(1: end, 1: end) * TimeCollect;
[TimeCounts, TimeCenters] = hist(TimeCollectAllSec, 15); % can change bin number here
Get rid of zero values that would be plotted

% Get rid of zero values that would be plotted
TimeCounts = TimeCounts(2:end,:);
TimeCenters = TimeCenters(2:end,:);

figure1 = figure;
% Create axes
axes1 = axes('Parent',figure1);
box(axes1,'on');
hold(axes1,'all')
bar1 = bar(TimeCenters, TimeCounts, 'Parent', axes1);
savefig('Fig1');

Create Pie Chart of Trace Types

PiePercents = TraceType(1, 1:6) / sum(TraceType(1, 1:6));
lab1 = 'DNA:
lab2 = strcat(lab1, num2str(PiePercents(1, 1)));
lab3 = 'Cy5 Bleach:
lab4 = strcat(lab3, num2str(PiePercents(1, 2)));
lab5 = 'Unwinding:
lab6 = 'Other:
lab6 = strcat(lab5, num2str(PiePercents(1, 4)));
lab6 = 'Cy3 Bleach:
lab5 = strcat(lab6, num2str(PiePercents(1, 5)));
lab6 = 'Other:
lab6 = strcat(lab5, num2str(PiePercents(1, 6)));
labels2 = [lab1, lab2, lab3, lab4, lab5, lab6];
figure2 = figure;
pie(TraceType(1, 1:6), labels2);
savefig('Fig2');

For option 3: DNA unwinding via steps

Create histograms of unwinding times and step numbers

figure3 = figure;
[h3count, h3center] = hist(StepNum, 0:1:100);
bar(h3center, h3count);
savefig('Fig3')

figure4 = figure;
[h4count, h4center] = hist(UnwindTime, 0:0.2:100);
bar(h4center, h4count);
savefig('Fig4');

For final histogram of all traces
Save data

For option 4: DNA unwinding and Rewinding
bar(h8centers,h8counts)
savedata.RewindTime=savedata.RewindTime*TimeCollect;
savefig('Fig8')
figure9=figure;
[h9counts,h9centers]=hist(RewindTime,0:0.2:100);
bar(h9centers,h9counts)
savefig('Fig9');
BIBLIOGRAPHY


198


W. Bujalowski and M. J. Jezewiska, "Interactions of *Escherichia coli* primary replicative helicase DnaB protein with single-stranded DNA. The nucleic acid does not wrap around the protein hexamer," *Biochemistry*, vol. 34, pp. 8513-8519, 1995.


[226] T. E. Shutt and M. W. Gray, "Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes," *J Mol Evol*, vol. 62, pp. 588-99, May 2006.


T. A. Brown, C. Cecconi, A. N. Tkachuk, C. Bustamante, and D. A. Clayton, "Replication of mitochondrial DNA occurs by strand displacement with alternative light-


